

Ekbert Hering
Gert Schönfelder *Editors*

Sensors in Science and Technology

Functionality and Application Areas

 Springer



Sensors in Science and Technology

Ekbert Hering • Gert Schönfelder
Editors

Sensors in Science and Technology

Functionality and Application Areas

 Springer

Editors

Ekbert Hering
Heubach, Baden-Württemberg
Germany

Gert Schönfelder
Dresden, Germany

ISBN 978-3-658-34919-6

ISBN 978-3-658-34920-2 (eBook)

<https://doi.org/10.1007/978-3-658-34920-2>

This book is a translation of the original German edition „Sensoren in Wissenschaft und Technik“ by Hering, Ekbert, published by Springer Fachmedien Wiesbaden GmbH in 2018. The translation was done with the help of artificial intelligence (machine translation by the service DeepL.com). A subsequent human revision was done primarily in terms of content, so that the book will read stylistically differently from a conventional translation. Springer Nature works continuously to further the development of tools for the production of books and on the related technologies to support the authors.

© Springer Fachmedien Wiesbaden GmbH, part of Springer Nature 2022

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH, part of Springer Nature.

The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Preface

In the age of the Internet, in which all information is available on the net in more or less good quality and companies provide a lot of information in pictures and text, to write a comprehensive work about sensor technology is a special challenge. We have written a book by practitioners for practitioners and by scientists for the students and prospective scientists of technical, biological and medical faculties. Our goal was to present the field of sensor technology in a comprehensive way and also to deal with its peripheral areas such as interfaces and safety aspects. We have structured the wealth of information and prepared it in a clear, compact and illustrative way. Together with a network of universities and industry, we have brought together the competencies that make this work unique in breadth and depth. In this way, it is a valuable textbook, a comprehensive reference book and an important aid to orientation in the variety of possible applications and uses of sensors for students and practitioners alike.

This work provides a comprehensive overview of sensors of physical, chemical, climatic, meteorological, biological and medical quantities. After an introduction in Chap. 1, the second chapter discusses the physical effects that are used for sensor use. In the following 11 chapters the fields of application of the sensors are presented: The acquisition of geometric quantities such as displacement and angle (Chap. 3), mechanical quantities (Chap. 4), time-based quantities (Chap. 5), thermal sensors (Chap. 6), sensors for electrical and magnetic quantities (Chap. 7), optical and acoustic quantities (Chaps. 8 and 9), sensors for climatic and meteorological purposes (Chap. 10), sensors in chemistry (Chap. 11) as well as in biology and medicine (Chap. 12), measurement of ionizing radiation (Chap. 13) as well as photoelectric sensors in Chap. 14. The electrical output signals must be processed and the sensors must be calibrated in order to measure the sensor signals truthfully (Chap. 15). The interfaces for transmission and further processing as well as the networking of different sensors play a very important role (Chap. 16). Safety aspects of sensors are dealt with in Chap. 17 and measurement errors and measurement accuracies are dealt with in Chap. 18. After each chapter, references to further literature or Internet addresses are given. This enables the readers to obtain more detailed information.

It was both a great pleasure and a challenge to work on this comprehensive work. We would like to take this opportunity to thank Mr. *Reinhard Dapper* of Vieweg+Teubner Verlag. He supported us in every respect and gave us a free hand in many decisions. *Fromm MediaDesign* has done an excellent job of designing such a comprehensive work in an appealing way and presenting the content well. We would like to thank Ms. *Angela Fromm* for her professional and friendly support.

More than 20 companies and many more experts from industry and universities were involved in the creation of the plant. It is thanks to their expertise that we are able to present the reader with a clear, compact and well-structured work. From the industrial side, Mr. *Albert Feinäugle*, Dipl.-Ing. of *Balluff GmbH*, is to be mentioned in particular. He has masterfully succeeded in inspiring the many experts in his company to contribute to this work. We would like to expressly praise and thank him for his special commitment. We would also like to thank the many other co-authors from industry and science.

A very special thanks to our wives and children, who had to forego many wonderful days together. At many times the work has taken over our lives and we had to do everything in our power to ensure the quality of this work. Without the patience of our families, their moral support and their great understanding, this work would not have been possible.

We hope that this book will be a valuable source of information and orientation for students, scientific workers and practitioners in the industry. We are happy to receive criticism and suggestions for improvement.

Heubach, Germany
Dresden, Germany

Eckbert Hering
Gert Schönfelder

Contents

1	Sensor Systems	1
	Ekbert Hering	
1.1	Definition and Mode of Operation	1
1.2	Classification	2
2	Physical Effects of Sensor Use	5
	Ekbert Hering, Karl-Ernst Biel, Ulrich Guth, Martin Liess, and Winfried Vonau	
2.1	Piezoelectric Effect	5
2.1.1	Principle of Operation and Physical Description	5
2.1.2	Materials	7
2.1.3	Applications	8
2.2	Resistive and Piezoresistive Effect	9
2.2.1	Operating Principles and Physical Description	9
2.2.2	Resistive Effect and Its Application by Means of SGs	11
2.2.3	Piezoresistive Effect and Its Application by Silicon Semiconductor Elements	13
2.2.4	Materials	13
2.3	Magnetoresistive Effect	15
2.3.1	Principle of Operation and Physical Description	15
2.3.1.1	AMR (Anisotropic Magneto Resistance)	16
2.3.1.2	GMR (Giant Magneto Resistance)	18
2.3.1.3	CMR (Colossal Magneto Resistance)	19
2.3.1.4	TMR (Tunnel Magneto Resistance)	19
2.3.2	Advantages of XMR Technology	19
2.3.3	Applications of XMR Technology	22
2.4	Magnetostrictive Effect	25
2.4.1	Principle of Operation and Physical Description	25
2.4.2	Advantages of Magnetostrictive Sensor Technology	26
2.4.3	Applications of Magnetostrictive Sensor Technology	27

2.5	Effects of Induction	28
2.5.1	Principle of Operation and Physical Description	28
2.5.1.1	Law of Induction	28
2.5.1.2	Generation of Eddy Currents in Electrically Conductive Materials	33
2.5.1.3	Electromagnetic Oscillating Circuits	33
2.5.2	Advantages of Inductive Sensor Technology	34
2.5.3	Applications of Inductive Sensor Technology	35
2.6	Effects of Capacitance	35
2.6.1	Principle of Operation and Physical Description	35
2.6.1.1	Capacitor and Capacitance	36
2.6.1.2	Capacitance in the Alternating Current Circuit	41
2.6.2	Advantages of Capacitive Sensor Technology	46
2.6.3	Applications of Capacitive Sensor Technology	47
2.7	Gaussian Effect	48
2.7.1	Principle of Operation and Physical Description	48
2.7.2	Application of the Gaussian Effect	50
2.8	Hall Effect	52
2.8.1	Principle of Operation and Physical Description	52
2.8.2	Application of the Hall Effect	54
2.9	Eddy Current Effect	55
2.9.1	Principle of Operation and Physical Description	55
2.9.2	Application of the Eddy Current Effect	56
2.10	Thermoelectric Effect	63
2.11	Thermoresistance Effect	68
2.11.1	Principle of Operation and Physical Description	68
2.11.1.1	Thermal Response Times	69
2.11.1.2	Self-heating and Measuring Current	70
2.11.2	Advantages of Sensor Technology with the Thermoresistance Effect	71
2.11.3	Fields of Application	71
2.12	Temperature Effects in Semiconductors	72
2.12.1	Principle of Operation and Physical Description	72
2.12.2	PTC Thermistors (PTC Resistors)	73
2.12.2.1	Physical Context	74
2.12.2.2	Advantages and Application Areas	75
2.12.3	Thermistors (NTC Resistors)	76
2.12.3.1	Advantages and Application Areas	77
2.13	Pyroelectric Effect	79
2.13.1	Principle of Operation and Physical Description	79
2.13.2	Materials	80

2.13.3	Applications	81
2.14	Photoelectric Effect	84
2.14.1	Operating Principles and Physical Description	84
2.14.1.1	Outer Photoelectric Effect	85
2.14.1.2	Inner Photoelectric Effect: Photoconductor	87
2.14.1.3	Internal Photoelectric Effect: Optocoupler	88
2.14.1.4	Internal Photo Effect: Photovoltaic Effect	88
2.14.1.5	Photoionization	88
2.14.2	Photoelectric Sensor Elements	90
2.14.3	Photoelectric Sensor Elements	90
2.14.3.1	Photomultiplier	90
2.14.3.2	Optocoupler	91
2.14.3.3	Light Barriers	93
2.14.3.4	Photoelectric Switches	95
2.14.3.5	Light Curtain	96
2.14.3.6	Light Measurement	96
2.14.3.7	Colour Recognition	97
2.15	Electro-optical Effect	98
2.15.1	Principle of Operation and Physical Description	98
2.15.2	Materials	100
2.15.3	Applications	100
2.16	Electrochemical Effects	103
2.16.1	Principle of Operation and Classification	103
2.16.2	Potentiometric Sensors	104
2.16.3	Amperometric Sensors	108
2.16.4	Conductometric and Impedimetric Sensors	109
2.16.5	Areas of Application	110
2.17	Chemical Effects	110
2.17.1	Physical-chemical Interactions of Gases with Surfaces	110
2.17.2	Gas Solubility (Absorption)	112
2.17.3	Gas Transport to the Solid Surface	116
2.17.4	Adsorption and Chemisorption	117
2.17.5	Reactions with Adsorbed Species	118
2.17.6	Reaction of the Gas with the Solid	118
2.17.7	The Mixed-phase Disorder	121
2.18	Acoustic Effects	123
2.18.1	Definition and Classification of Sound	123
2.18.2	Characterization of Acoustic Waves	123
2.18.3	Sound Velocity in Ideal Gases	124
2.18.3.1	Dependence on Temperature	124
2.18.3.2	Dependence on the Relative Air Humidity	125

2.18.3.3	Dependence on Pressure	125
2.18.4	Intensity or Sound Intensity	125
2.18.5	Sound Absorption in Air	126
2.18.6	Reflection and Transmission	127
2.19	Optical Effects	128
2.19.1	Physical Effects	128
2.19.2	Design of Optical Sensors	133
2.19.3	Categories of Optical Sensors	135
2.19.4	Application Fields of Optical Sensors	136
2.20	Doppler Effect	137
2.20.1	Principle of Operation and Physical Description	137
2.20.1.1	Observer Moves, Source Rests	137
2.20.1.2	Source Moves, Observer Rests	138
2.20.1.3	Observer and Source Move	138
2.20.1.4	Doppler Effect of Light (Doppler Effect Without Medium)	139
2.20.2	Application Areas	140
2.20.2.1	Astronomy	140
2.20.2.2	Geodesy (Land Surveying)	141
2.20.2.3	Navigation	141
2.20.2.4	Vibration Analysis	142
2.20.2.5	Speed Measurement	142
2.20.2.6	Determination of Chemical Elements	142
2.20.2.7	Medical Technology	142
2.20.2.8	Acoustics	142
	Bibliography	143
3	Geometric Quantities	147
	Ekbert Hering, Gert Schönfelder, Stefan Basler, Karl-Ernst Biehl, Thomas Burkhardt, Thomas Engel, Albert Feinäugle, Sorin Fericean, Alexander Forkl, Carsten Giebeler, Bernhard Hahn, Ernst Halder, Christopher Herfort, Stefan Hubrich, Jürgen Reichenbach, Michael Röbel, and Stefan Sester	
3.1	Displacement and Distance Sensors	148
3.1.1	Inductive Distance and Displacement Sensors	148
3.1.1.1	Functional Principle and Morphological Description of Inductive Sensors	148
3.1.1.2	Non-contact Inductive Distance Sensors (INS)	152
3.1.1.3	Non-contact Inductive Displacement Sensors (IWS)	160
3.1.1.4	Differential Transformers with Sliding Core (LVDT)	163

3.1.1.5	Pulsed Inductive Linear Position Sensor (Micropulse [®] BIW)	168
3.1.1.6	Signal Processing by Phase Measurement (Sagentia)	173
3.1.1.7	PLCD Displacement Sensors (Permanent Linear Contactless Displacement Sensor)	178
3.1.1.8	Non-contact Magnetoinductive Displacement Sensors (Smartsens-BIL)	184
3.1.2	Optoelectronic Distance and Displacement Sensors	193
3.1.2.1	Overview	193
3.1.2.2	Optoelectronic Components	195
3.1.2.3	Optical Principles of Distance Sensors	200
3.1.2.4	Measuring Principle: Triangulation	204
3.1.2.5	Measuring Principle: Pulse Delay Method	206
3.1.2.6	Measuring Principle: Phase or Frequency Delay Method	208
3.1.2.7	Measuring Principle: Photoelectric Scanning	211
3.1.2.8	Measuring Principle: Interferometric Length Measurement	212
3.1.3	Ultrasonic sensors for Distance Measurement and Object Detection	213
3.1.3.1	Operating Principles and Design Touch Operation with Echo Runtime Measurement	213
3.1.3.2	Design of the Ultrasonic Transducer	215
3.1.3.3	Detection Range of an Ultrasonic Sensor	217
3.1.3.4	Deflection of the Ultrasound	219
3.1.3.5	Object and Environmental Influences	219
3.1.3.6	Applications	220
3.1.4	Potentiometric Displacement and Angle Sensors	222
3.1.4.1	Introduction	222
3.1.4.2	Operating Principle and Characteristics of Potentiometric Sensors	225
3.1.4.3	Technology and Construction Techniques	228
3.1.4.4	Products and Applications	233
3.1.5	Magnetostrictive Displacement Sensors	234
3.1.5.1	Operating Principle and Design of Magnetostrictive Displacement Sensors	236
3.1.5.2	Housing Concepts and Applications	239
3.1.6	Displacement Sensors with Magnetically Coded Measuring Standard	244
3.1.6.1	Measuring Principle	244

	3.1.6.2	Design and Function of Incremental and Absolute Measuring Systems	248
	3.1.6.3	Characteristic Values	251
	3.1.6.4	Sensor Types in Comparison	254
	3.1.6.5	Application Examples	255
3.2		Sensors for Angle and Rotation	257
	3.2.1	Optical Encoders	267
	3.2.1.1	Physical Principles	267
	3.2.1.2	Design of Optical Encoders	270
	3.2.1.3	Special Features of Optical Encoders	274
	3.2.2	Magnetically Encoded Rotary Encoder	275
	3.2.2.1	Applications	279
	3.2.3	Rotation-Counting Angle Sensors	281
	3.2.3.1	General Principle of Operation and Morphological Description of Rotation-Counting Angle Sensors	281
	3.2.3.2	Gearbox-Based Revolution-Counting Methods	282
	3.2.3.3	Rotation-Counting Method on Inductive Basis	284
	3.2.3.4	Battery Buffering of the Rotation Information	286
	3.2.3.5	Novel GMR System for Detection and Storage of Rotation Information	287
	3.2.4	Capacitive Encoders	292
	3.2.5	Variable Transformers, Resolvers	295
	3.2.5.1	General Operating Principle of the VT	295
	3.2.5.2	Significant Variants of VT	298
	3.2.5.3	Resolver, a Representative Variant of VT	298
	3.2.6	1Vpp or Sin/Cos Interface	304
	3.2.7	Incremental Encoders	307
	3.2.7.1	Summary of the Properties of Incremental Interfaces	308
3.3		Inclination	308
	3.3.1	Magneto-resistive Inclination Sensors	310
	3.3.2	Compass Sensors	311
	3.3.3	Electrolytic Sensors	311
	3.3.4	Piezoresistive Inclination Sensors/DMS Bending Beam Sensors	313
	3.3.5	MEMS	314
	3.3.6	Servo-Inclinometer	315
	3.3.7	Overview and Selection of Inclination Sensors	316
3.4		Sensors for Object Detection	316

3.4.1	Proximity Switch	316
3.4.1.1	Block Diagram of the NS, Detection Type and Mode of Operation of the NS	319
3.4.1.2	Main Features of the NS	321
3.4.1.3	Switching Distances	321
3.4.1.4	Hysteresis	323
3.4.1.5	Switching Element Function	323
3.4.1.6	Output Type	324
3.4.1.7	Design and Size	324
3.4.1.8	Characteristics of the NS	325
3.4.1.9	Special Designs and Their Applications	326
3.4.1.10	NS with Several Switching Outputs	326
3.4.1.11	NS with Diagnostic Information	328
3.4.1.12	Applications of the NS	328
3.4.2	Object Detection and Distance Measurement with Ultrasound	330
3.4.2.1	Speed of Sound in Air	330
3.4.3	Object Detection with Radar	331
3.4.3.1	Impulse Radar	331
3.4.3.2	CW Radar	333
3.4.3.3	Application	334
3.4.4	Pyroelectric Sensors for Motion and Presence Detection	334
3.4.4.1	Material Properties of Sputtered PZT Films	334
3.4.4.2	Effect on Electronics	335
3.4.5	Object Detection with Laser Scanner	337
3.4.5.1	Application	337
3.4.6	Sensors for Automatic Identification (Auto-Ident)	337
3.4.6.1	Overview	337
3.4.6.2	Barcode Scanner	339
3.4.6.3	Auto-Ident Cameras	346
3.4.6.4	Construction of Auto-Ident Cameras	347
3.4.6.5	Important Fields of Application for Auto-Ident Cameras	348
3.4.6.6	RFID Systems and Readers	348
3.5	Three-Dimensional Measuring Methods (3D Measurement)	356
3.5.1	Palpable 3D Measurement Methods	357
3.5.1.1	Switching Sensor	357
3.5.1.2	2-phase Switching Sensor	359
3.5.2	Optical Probing 3D Measuring Methods	360
3.5.2.1	Optical, Switching Single Point Sensor	360
3.5.2.2	Scanning Sensors	361

	3.5.2.3	Other Probe and Measuring Systems	362
3.5.3		3D Imaging Measuring Methods	363
	3.5.3.1	Optical 3D Measurement (Grid and Line Projection)	363
	3.5.3.2	Measuring Principle and Measuring Arrangement	364
	3.5.3.3	Fringe Projection	365
	3.5.3.4	Limitations of the Procedure	367
	3.5.3.5	Fields of Application	367
3.5.4		Overview of 3D Measuring Methods	368
		Bibliography	369
4		Mechanical Measured Variables	373
		Ekbert Hering, Gert Schönfelder, and Stefan Vinzelberg	
4.1		Mass	373
	4.1.1	Definition	373
	4.1.2	Applications	374
4.2		Force	376
	4.2.1	Definition	376
		4.2.1.1 Weight Force F_W	376
		4.2.1.2 Centripetal Force F_{cp}	376
		4.2.1.3 Elastic Force or Spring Force F_{el}	377
		4.2.1.4 Frictional Force F_F	377
	4.2.2	Effects for the Applications	377
		4.2.2.1 Piezoelectric Effect	378
	4.2.3	Application Areas	381
		4.2.3.1 Process Control	381
4.3		Elongation	384
	4.3.1	Definition	384
	4.3.2	Strain Measurement	385
		4.3.2.1 Bending Beam	387
		4.3.2.2 Fiber Bragg Grating	387
4.4		Pressure	388
	4.4.1	Definition	388
		4.4.1.1 Fields of Application	389
	4.4.2	Measuring Principles	389
		4.4.2.1 Measuring Principles	389
		4.4.2.2 Differential Pressure Measurement	389
		4.4.2.3 Relative Pressure Measurement	390
		4.4.2.4 Absolute Pressure Measurement	390
	4.4.3	Measuring Arrangements	391
4.5		Torque	395

4.5.1	Definition	395
4.5.2	Measuring Principles	396
4.5.3	Application Areas	397
4.6	Hardness	398
4.6.1	Definition	398
4.6.2	Macroscopic Hardness Determination	399
4.6.3	Hardness Determination by Nanoindentation	400
4.6.4	Sensors for Nanohardness Measurement	401
4.6.5	Model and Evaluation	401
4.6.6	Applications	403
5	Time-Based Measurements	405
	Gert Schönfelder and Gerd Stephan	
5.1	Time	405
5.1.1	Definition	405
5.1.2	Measuring Principles	405
5.2	Frequency	406
5.2.1	Definition	406
5.2.2	Measuring Principles	407
5.2.3	Measuring Arrangements for Frequency and Time Measurement	407
5.2.4	Measurement Error of Time-Discrete Measurements	407
5.2.5	Measuring Arrangements	409
5.3	Pulse Width	412
5.3.1	Definition	412
5.3.2	Measuring Principles	412
5.3.3	Analogue Evaluation	414
5.3.4	Digital Evaluation	415
5.3.5	Measuring Arrangements	415
5.4	Phase, Running Time, and Light Running Time	415
5.4.1	Definition	415
5.4.2	Measuring Principles	416
5.4.3	Measuring Arrangements	417
5.4.4	Light Running Time	417
5.4.5	Direct Time-of-Flight Measurement	418
5.4.6	Transit Time Measurement Through Pulse Integration	419
5.4.7	Time of Flight Measurement with Modulated Light	421
5.5	Visual Representation of Measured Variables	422
5.5.1	Measuring Principle	422
5.5.2	Analogue Oscillographs	423
5.5.3	Digital Storage Oscilloscopes (DSOs)	424
5.5.4	Mixed Signal Oscilloscopes (MSOs)	426

5.5.5	The Sampling Principle	426
5.5.6	Measuring Arrangements	427
5.5.7	Voltage and Period Duration Measurement	427
5.5.8	Measurement of Rise Times	428
5.5.9	Measurement of Phases	428
5.5.10	Measurement of Pulse Duration and Pulse Ratios (PWM)	431
5.6	Speed and Angle of Rotation	431
5.6.1	Definition	431
5.6.2	Measuring Principles	433
5.6.3	Measurement Through Event Recording	433
5.6.4	Measurement by Position Detection	433
5.6.5	Speed Determination by Beating (Stroboscope)	433
5.6.6	Measuring Arrangements	434
5.6.6.1	Magnetic and Optical Scanning	434
5.6.7	Digital or A/B Interface (Incremental Encoder)	434
5.7	Speed	436
5.7.1	Definition	436
5.7.2	Measuring Principle Speed	437
5.7.3	Measuring Arrangements for Speed Measurement	437
5.8	Acceleration	439
5.8.1	Definition	439
5.8.2	Fields of Application	440
5.8.3	Measuring Principle Linear Acceleration	440
5.8.4	Measuring Principle Angular Acceleration	443
5.8.5	Measuring Arrangement for Measuring Acceleration	445
5.9	Flow Rate (Mass and Volume)	445
5.9.1	Definition	445
5.9.2	Mass	445
5.9.3	Volume	445
5.9.4	Volume Flow	446
5.9.5	Mass Flow	446
5.9.6	Main Groups	446
5.9.7	Measurement Methods and Application	446
5.9.8	Bulk Flow Meter	447
5.9.9	Baffle Plate Scales	448
5.9.10	Weighfeeders	448
5.9.11	Belt Scales	448
5.9.12	Differential Scales	448
5.9.13	Optical Belt Weigher	449
	Bibliography	450

6	Temperature Measurement	451
	Ekbert Hering, Gert Schönfelder, Martin Liess, and Lothar Michalowski	
6.1	Temperature as Physical State Variable	451
6.2	Measuring Principles and Measuring Ranges	453
6.3	Temperature Dependence of the Electrical Resistance	453
6.3.1	Metals	453
6.3.2	Metals with Defined Additives (Alloys) or Lattice Defects	458
6.3.3	Ion Conducting Materials for High Temperatures	459
6.3.4	Thermistors	460
6.3.5	Constriction Resistor Temperature Sensors (Spreading Resistor)	461
6.3.6	Diodes	463
6.4	Thermoelectricity (Seebeck Effect)	464
6.4.1	Thermocouples for Very High Application Temperatures	469
6.4.2	Metallic Thermocouples for Very High Application Temperatures	469
6.4.3	Inorganic-Non-metallic Thermocouples for Very High Application Temperatures	469
6.5	Thermal Expansion	470
6.5.1	Thermal Expansion of Solid Bodies	470
6.5.2	Thermal Expansion of Liquids	472
6.5.3	Thermal Expansion of Gases	474
6.6	Temperature and Frequency	474
6.7	Thermochromism	475
6.8	Seeger Cone	476
6.9	Non-contact Optical Temperature Measurement	477
6.9.1	Radiation Thermometer (Pyrometer)	477
6.9.2	Fiber Optic Applications	479
6.9.2.1	Intrinsic Sensors, DTS (Distributed Temperature Sensing)	479
6.9.2.2	Extrinsic Sensors	481
	Bibliography	483
7	Electrical and Magnetic Measured Variables	485
	Gert Schönfelder and Andreas Wilde	
7.1	Voltage	485
7.1.1	Definition	485
7.1.1.1	AC Voltage	486
7.1.1.2	Measuring Principles	487
7.1.1.3	Conversion into an Electromagnetic Field	488
7.1.1.4	Conversion to Heat	488

	7.1.1.5	Conversion into a Current	488
	7.1.1.6	Measurement by Comparison with a Standard . . .	488
	7.1.2	Measuring Arrangements	490
	7.1.2.1	Measurement by Energy Extraction	490
	7.1.2.2	Measurement Through Integration	491
	7.1.2.3	Measurement by Comparison	492
	7.1.2.4	Servo Converter	492
	7.1.2.5	The Successive Approximation	492
	7.1.2.6	Measurement of Alternating Voltages	494
7.2	Amperage		495
	7.2.1	Definition	495
	7.2.1.1	Measuring Principles	495
	7.2.1.2	Current Flow Through a Resistor	495
	7.2.1.3	Thermal Effect of Current Flow	496
	7.2.1.4	Magnetic Field Due to Current Flow	496
	7.2.2	Measuring Arrangements	496
7.3	Electrical Charge and Capacity		498
	7.3.1	Definition	498
	7.3.1.1	Measuring Principle Charge	500
	7.3.1.2	Measuring Principle Capacity	501
	7.3.2	Measuring Arrangements	503
	7.3.2.1	Capacity Measurement Through Charge Sharing	503
	7.3.2.2	Capacity Measurement by RC Generators	503
	7.3.2.3	Complex Components for Capacitance Determination	503
7.4	Electrical Conductivity and Specific Electrical Resistance		504
	7.4.1	Definition	504
	7.4.1.1	Measuring Principles for Resistance	506
	7.4.2	Measuring Arrangements	506
	7.4.2.1	Determination of Current and Voltage at the Measured Object	506
	7.4.2.2	Resistance Measurement Through Compensation	508
	7.4.2.3	AC Voltage Resistance Measurement	509
7.5	Electric Field Strength		509
	7.5.1	Definition	509
	7.5.2	Measuring Principles for the Electric Field Strength	510
7.6	Electrical Energy and Power		511
	7.6.1	Definitions	511
	7.6.1.1	Practical Case Accumulator	512

7.6.2	Forms of Power	512
7.6.2.1	Power in DC Circuit	512
7.6.2.2	Power in AC Circuit	513
7.6.3	Measuring Principles	515
7.6.3.1	Electromechanical	515
7.6.3.2	With Analogue Electronics	515
7.6.3.3	With Digital Electronics	516
7.6.3.4	With Statistical Methods	517
7.7	Inductance	517
7.7.1	Definition	517
7.7.2	Measuring Principles	519
7.8	Magnetic Field Strength	520
7.8.1	Definition	520
7.8.2	Measuring Principles of Magnetic Quantities	521
7.8.2.1	Hall Sensor	521
7.8.2.2	GMR Sensor	521
7.8.2.3	Field Plate	521
7.8.2.4	SQUID	521
7.8.3	Measuring Arrangements	522
7.8.3.1	Hall Switch	522
7.8.3.2	Hall Sensors	522
7.8.3.3	Angle Sensors	522
7.8.3.4	Current Sensors	522
7.8.4	Multidimensional Measurements with the Hall Effect	523
7.8.4.1	Basics	523
7.8.4.2	Applications	524
	Bibliography	525
8	Radio and Photometric Quantities	527
	Ekbert Hering and Gert Schönfelder	
8.1	Radiometry	527
8.1.1	Radiometric Quantities	527
8.1.1.1	Energy Density	527
8.1.1.2	Radiant Power Φ_e	528
8.1.1.3	Solid Angle Ω	528
8.1.1.4	Radiant Intensity I_e	529
8.1.1.5	Radiance L_e	530
8.1.2	Measurement of Electromagnetic Radiation	531
8.2	Photometry	531
8.2.1	Photometric Quantities	532
8.2.1.1	Luminous Flux Φ_V	532
8.2.1.2	Light Quantity Q_V	534

8.2.1.3	Luminous Intensity I_v	534
8.2.1.4	Luminance L_v	534
8.2.1.5	Illuminance E_v	535
8.2.1.6	Specific Light Emission M_v	535
8.2.1.7	Luminous Efficacy η	535
8.2.2	Measurement of Photometric Quantities	535
8.3	Application of Brightness Sensors	537
8.4	Colour	540
8.4.1	Colour Perception	540
8.4.2	Colour Models	543
8.4.2.1	Other Color Spaces	544
8.4.3	Color Systems	544
8.4.4	Colour Filters for Sensors	545
8.4.4.1	Bayer Pattern	545
8.4.4.2	Color Correction	546
8.4.4.3	White Balance	546
8.4.5	Colour Sensors	547
	Bibliography	548
9	Acoustic Measured Variables	549
	Ekbert Hering	
9.1	Definition of Important Acoustic Quantities	549
9.2	Human Perception	551
9.2.1	Level	551
9.2.2	Volume	553
9.2.3	Loudness	554
9.3	Transducer	555
9.4	Fields of Application	557
	Bibliography	560
10	Climatic and Meteorological Measured Variables	561
	Gert Schönfelder, Robert Krah, Gerd Stephan, and Roland Wernecke	
10.1	Moisture in Gases	561
10.1.1	Definitions and Equations	561
10.1.1.1	Humidity	561
10.1.1.2	Water Vapour Partial Pressure	561
10.1.1.3	Absolute and Specific Humidity	562
10.1.1.4	Saturation Humidity	562
10.1.1.5	Relative Humidity φ	562
10.1.1.6	Dew Point	563
10.1.1.7	Enthalpy	564
10.1.1.8	Mollier Diagram (h-x Diagram)	564

10.1.2	Humidity Measurements in Gases	566
10.1.2.1	Psychrometers, Design and Function	566
10.1.2.2	Design Types and Areas of Application	568
10.1.2.3	Aspiration Psychrometer	568
10.1.2.4	Slingshot Psychrometer	568
10.1.2.5	Dew Point Mirror	569
10.1.2.6	Capacitive Humidity Measurement	572
10.1.2.7	Integrated Capacitive Humidity Sensors with Bus Output	574
10.2	Moisture Analysis in Solid and Liquid Substances	574
10.2.1	Direct Methods for the Determination of Material Moisture	576
10.2.1.1	Percentage Water Content of a Material Sample	576
10.2.1.2	Water Activity of a Material Sample	577
10.2.1.3	Karl Fischer Titration	577
10.2.1.4	Calcium Carbide Method	579
10.2.1.5	Calcium Hydride Method	579
10.2.2	Indirect Measuring Methods for Determining the Moisture Content of Materials	579
10.2.2.1	Measurement of the Electrical Properties	579
10.2.2.2	Detection of the Optical Properties of Water and Water Vapour	580
10.2.2.3	Measurement of the Suction Pressure in Moist Materials (Tensiometry)	582
10.2.2.4	Measurement of Atomic Properties	582
10.2.2.5	Nuclear Magnetic Resonance (NMR) Method	583
10.2.2.6	Measurement of Thermal Conductivity	584
10.3	Measurement of Precipitation in Outdoor Climate	585
10.3.1	Measurement of the Relative Air Humidity	585
10.3.2	Precipitation Measurement	585
10.3.3	Condensation Measurement	586
10.4	Humidity Measurement in Closed Rooms	587
10.4.1	Measurement of the Climate in Homes and at Workplaces	587
10.4.2	Climate in Museums and Exhibition Rooms	588
10.4.3	Climate in Electrical Installations	591
10.4.4	Influencing the Indoor Climate	591
10.4.4.1	Air Humidification	591
10.4.4.2	Evaporator	591
10.4.4.3	Steam Humidifier	592
10.4.4.4	Nebulizer	592

	10.4.4.5	Dehumidifying Rooms	593
	10.4.4.6	Condensation Dehumidification	593
	10.4.4.7	Adsorption Dehumidifier	593
	10.4.4.8	Absorption Dehumidifier	593
10.5		Air Pressure	593
	10.5.1	Fields of Application	593
	10.5.2	Measuring Principles	594
	10.5.3	Definitions	594
10.6		Wind and Air Flow	595
	10.6.1	Definition	595
	10.6.2	Methods for Wind Measurement	595
10.7		Water Flow	600
	10.7.1	Definition	600
	10.7.2	Direct and Indirect Flow Measurement	601
		10.7.2.1 Ultrasonic Measurement	601
		10.7.2.2 Term Principle	601
		10.7.2.3 Doppler Principle	603
		10.7.2.4 Electromagnetic Measuring Method	604
		10.7.2.5 Hydrometric Measuring Blade	604
		10.7.2.6 Measurement with Markers (Tracers)	605
		Bibliography	605
11		Selected Chemical Parameters	607
		Winfried Vonau	
11.1		Redox Potential	607
	11.1.1	General	607
	11.1.2	Precious Metal Redox Electrodes	610
	11.1.3	Redox Glass Electrodes	611
	11.1.4	Reference Electrodes	613
11.2		Ions Including Hydronium Ions	617
	11.2.1	General Information	617
	11.2.2	pH Measurement	618
	11.2.3	Other Ions	623
11.3		Gases	627
	11.3.1	General Information	627
	11.3.2	Gases in a Physically Dissolved State or at Normal Temperature	628
		11.3.2.1 Solid Electrolyte Sensors	630
	11.3.3	Semiconductor Gas Sensors: Metal Oxide Semiconductor Sensors (MOS)	641
		11.3.3.1 Surface Conductivity	641
		11.3.3.2 Volume Conductivity	642
	11.3.4	Pellistors	643

11.4	Electrolytic Conductivity	644
11.4.1	General Information	644
11.4.2	Kohlrausch Measuring Cells	644
11.4.3	Multi-Electrode Measuring Cells	645
11.4.4	Electrodeless Conductivity Measuring Cells	646
11.4.5	Examples for the Application of Conductivity Sensors	647
12	Biological and Medical Sensors	651
	Elfriede Simon	
12.1	Biological Sensor Technology	651
12.1.1	Biosensor Technology	651
12.1.2	Real Biological Sensors	653
12.2	Functional Principles of Biosensors	655
12.2.1	Calorimetric Sensors	655
12.2.2	Microgravimetric Sensors	657
12.2.3	Optical Sensors	658
12.2.4	Electrochemical Sensors	660
12.2.5	Immobilization Methods	663
12.3	Physical and Chemical Sensors in Medicine	664
12.3.1	Physical-Chemical Blood Analyses	665
12.3.2	Clinical-Chemical Blood Analyses	668
12.4	Enzymatic Methods: Enzyme Sensors	669
12.4.1	Enzyme-Based Analyte Detection	672
12.4.2	Determination of Enzyme Activity	673
12.4.3	Application Fields of Enzymatic Tests	674
12.5	Immunological Methods: Immunosensors	675
12.5.1	Direct Immunosensors	679
12.5.2	Indirect Immunosensors	679
12.5.3	Application Fields of Immunosensors	681
12.6	DNA-Based Sensors	681
12.6.1	Hybridization Diagnostics	682
12.6.2	Application and Use of DNA Sensors	684
12.7	Cell-Based Sensor Technology	686
12.7.1	Metabolic Cell Chip	687
12.7.2	Neuro-chip	687
	Bibliography	689
13	Measured Quantities for Ionizing Radiation	691
	Hartmut Bärwolff	
13.1	Introduction and Physical Quantities	691
13.2	Interaction of Ionising Radiation with Matter	696

13.3	Classification of Sensors	700
13.4	Gas-Filled Radiation Sensors	703
13.5	Radiation Sensors According to the Excitation Principle	708
13.6	Semiconductor Sensors	710
	Bibliography	720
14	Photoelectric Sensors	721
	Gert Schönfelder and Martin Liess	
14.1	Radiation	721
14.2	Scintillators	722
14.3	Outer Photoelectric Effect	723
	14.3.1 Photomultiplier	723
	14.3.2 Channel Photomultiplier	724
	14.3.3 Image Pickup Tubes	725
14.4	Internal Photoelectric Effect	726
	14.4.1 Photoconductor	726
	14.4.2 Photodiodes	728
	14.4.3 Phototransistor, Photothyristor and Photo-FET	731
	14.4.4 CMOS Image Sensors	732
	14.4.5 High Dynamic CMOS Image Sensors	732
14.5	CCD Sensors	734
	14.5.1 Line Sensors	735
	14.5.2 CCD Matrix Sensors	735
14.6	Quantum Well Infrared Photodetector QWIP	736
14.7	Thermal Optical Detectors	738
	14.7.1 Thermal Piles	739
	14.7.2 Pyroelectric Detectors	741
	14.7.3 Bolometer	742
15	Signal Processing and Calibration	745
	Gert Schönfelder	
15.1	Signal Processing	745
	15.1.1 Analog (Discrete) Signal Conditioning	746
	15.1.2 Signal Conditioning with System Circuits	746
	15.1.3 Signal Conditioning with ASICs	746
	15.1.4 Signal Conditioning with Microcontrollers	748
15.2	Sensor Calibration	748
	15.2.1 Passive Compensation	749
	15.2.2 Adjustment with Analog Signal Processing	750
	15.2.3 Adjustment with Digital Signal Processing	751
15.3	Energy Management for Sensors	753
	Bibliography	755

16 Interface	757
Gert Schönfelder	
16.1 Analogue Interfaces	757
16.1.1 Voltage Output	758
16.1.2 Ratiometric Voltage Output	759
16.1.3 Current Output	759
16.1.4 Frequency Output and Pulse Width Modulation	760
16.1.5 4-/6-Wire Interface	762
16.2 Digital Interfaces	762
16.2.1 CAN Group	765
16.2.2 LON	767
16.2.3 HART	767
16.2.4 RS485	768
16.2.5 IO-Link	768
16.2.6 Profibus	770
16.2.7 I2C	771
16.2.8 SPI	772
16.2.9 IEEE 1451	773
Bibliography	776
17 Safety Aspects for Sensors	777
Gert Schönfelder and Sorin Fericean	
17.1 Features for Function Monitoring	777
17.2 Electromagnetic Compatibility (EMC)	782
17.3 Functional Safety (SIL)	785
17.4 Sensors in Explosive Environments (ATEX)	787
17.4.1 Basic Principles of ATEX	788
17.4.2 Ignition Protection Type Intrinsic Safety	789
17.4.3 Type of Protection Flameproof Enclosure	791
Bibliography	791
18 Measurement Errors, Measurement Accuracy and Measurement Parameters	793
Gert Schönfelder	
18.1 Classification of Measurement Errors According to Their Cause	793
18.2 Display of Measurement Errors	795
18.2.1 Arithmetic Mean, Error Sum and Standard Deviation	795
18.2.2 Absolute Error	796
18.2.3 Relative Error	796
18.3 Measurement Parameters	799
18.3.1 Scattering of Measured Values	800
18.3.2 Resolution of Measured Values	801
18.3.3 Signal-to-Noise Ratio and Dynamics of Measured Values	801
Bibliography	802
Index	803

List of Contributors

Hartmut Bärwolff Hochschule Köln, Köln, Germany

Stefan Basler ehemals SICK STEGMANN GmbH, Donaueschingen, Germany

Karl-Ernst Biehl PIL Sensoren GmbH, Erlensee, Germany

Thomas Burkhardt Balluff GmbH, Neuhausen a.d.F., Germany

Thomas Engel Carl Zeiss IMT GmbH, Oberkochen, Germany

Albert Feinäugle Balluff GmbH, Neuhausen a.d.F., Germany

Sorin Fericean Balluff GmbH, Neuhausen a.d.F., Germany

Alexander Forkl Balluff GmbH, Neuhausen a.d.F., Germany

Carsten Giebeler Pyreos Ltd, Edinburgh, UK

Ulrich Guth Kurt-Schwabe-Institut für Messtechnik, Meinsberg, Waldheim, Germany

Bernhard Hahn Balluff GmbH, Neuhausen a.d.F., Germany

Ernst Halder Novotechnik Messaufnehmer OHG, Ostfildern, Germany

Christopher Herfort PIL Sensoren GmbH, Erlensee, Germany

Eckbert Hering Hochschule Aalen (Rektor i. R.), Aalen, Germany

Stefan Hubrich Novotechnik Messaufnehmer OHG, Ostfildern, Germany

Robert Krah Krah&Grote Messtechnik, Otterfing, Germany

Thomas Kubasta Balluff GmbH, Neuhausen a.d.F., Germany

Martin Liess Hochschule RheinMain, Rüsselsheim, Germany

Lothar Michalowski Universität Göttingen, Göttingen, Germany

Jürgen Reichenbach SICK AG, Freiburg, Germany

Michael Röbel Deutz AG, Köln, Germany

Gert Schönfelder Prignitz Mikrosystemtechnik, Wittenberge, Germany

Stefan Sester Novotechnik Messaufnehmer OHG, Ostfildern, Germany

Elfriede Simon Siemens AG, München, Germany

Gerd Stephan Quantum Hydrometrie Gesellschaft für Mess- und Systemtechnik, Berlin, Germany

Stefan Vinzelberg Atomic Force F&E GmbH, Mannheim, Germany

Winfried Vonau Kurt-Schwabe-Institut für Messtechnik, Meinsberg, Waldheim, Germany

Roland Wernecke Dr. Wernecke Feuchtemesstechnik GmbH, Potsdam, Germany

Andreas Wilde Fraunhofer Institut für Integrierte Schaltungen, Dresden, Germany

Frank Winkens Balluff GmbH, Neuhausen a.d.F., Germany



Ekbert Hering

In order to master daily life with his tasks, man must be able to influence the processes surrounding him, that is, he must be able to control them according to his goals. To make this possible, the current actual state must be recorded, the information from this must be evaluated, and the measures must be taken to achieve the goals. The elements that make it possible to record the measured variables of the environment are the sensors. They thus form the basis for all changes and developments in the natural or technical environment of human beings. Different effects in physics, chemistry, biology, and medicine enable a variety of sensors and applications. These are presented comprehensively in this work, but without claiming to be complete.

1.1 Definition and Mode of Operation

The word “sensor” comes from Latin (*sensus*: sense) and means feeler. Figure 1.1 shows the mode of operation of sensors.

A sensor is used for quantitative and qualitative measurement of physical, chemical, climatic, biological, and medical parameters. As Fig. 1.1 shows, the sensor consists of two parts: the *sensor element* and the *evaluation electronics*. The *non-electrical input variables* to be measured are converted into an *electrical output signal* in the sensor element by natural scientific laws. In an *electronic evaluation unit*, these output signals are processed by circuit electronics or software programs to produce a sensor output signal that is available for control or evaluation purposes. The external disturbance variables that

E. Hering (✉)
Hochschule Aalen (Rektor i. R.), Aalen, Germany
e-mail: Ekbert.Hering@hs-aalen.de

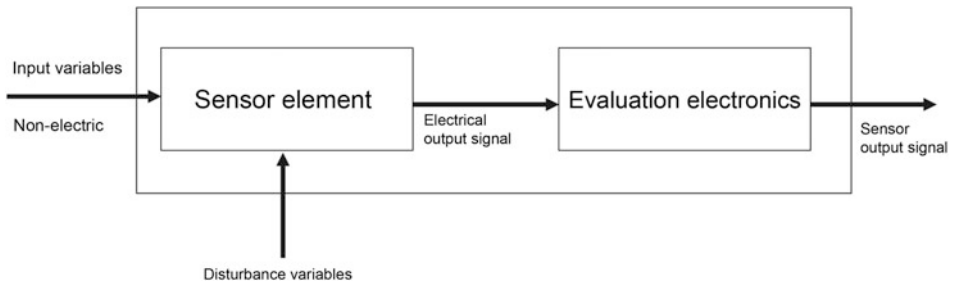


Fig. 1.1 Operating principle of sensors

influence a sensor element can be taken into account by calculation (e.g., consideration of temperature dependency or linearization of non-linear relationships). This is usually done by a microprocessor. Progressive miniaturization increasingly allows both parts, the sensor element and the evaluation electronics, to be accommodated in a single sensor. These intelligent sensors are also known as *smart sensors*.

1.2 Classification

If the measured variable is converted into an electrical variable without external auxiliary voltage, these are *active sensors* (e.g., in the case of the piezoelectric effect, the pressure is converted directly into an electrical variable, see Sect. 2.1). *Passive sensors*, on the other hand, require an external auxiliary voltage for conversion (e.g., when measuring distances with ultrasonic sensors).

A further classification can be made by the laws of science, which convert the input variable into an electrical output signal. These laws are described in detail in Chap. 2 of this book.

However, a classification can also be made by the quantities to be measured themselves (e.g., measuring geometric quantities such as length or time-based quantities such as frequencies). This can be read in Chaps. 3–14.

Digital technology is playing an increasingly important role in the further processing of the sensor output signals. Therefore, *digital* sensor output signals are often expected. This is often realized by integrating analog-to-digital converters (A/D converters) into the evaluation electronics of the sensor system.

If the desired measured variables are calculated from the measured values of real sensor elements by means of software, these are called *virtual sensors*. These are based on an extensive *mathematical calculation model* or *empirically recorded correlations* (e.g., subjective hearing sensation). Since such computers today can be accommodated in a very space-saving manner, manufactured at low cost, and can be used without problems in

harsh industrial environments, virtual sensors are used wherever costs need to be saved, where scientific relationships are only empirically available or in applications where real sensors would be destroyed or wear out too quickly (e.g., in nuclear power plants).



Eckbert Hering, Karl-Ernst Biel, Ulrich Guth, Martin Liess,
and Winfried Vonau

2.1 Piezoelectric Effect

2.1.1 Principle of Operation and Physical Description

If certain materials are *deformed* by the action of external forces or pressures, an *electrical voltage* is generated. As Fig. 2.1 shows, the force or pressure displaces the charges inside the material. The centers of gravity of the positive and negative charges no longer coincide. This results in an *electric polarization P*. Charges accumulate on the surface of the materials so that an *electrical voltage* can be measured.

Depending on the direction of the vectors force F , the polarization P and the surface normal vector \mathbf{n} relative to each other, there are three categories of the piezoelectric effect (Fig. 2.2):

In the opposite case (*inverse piezoelectric effect*), the crystal is deformed by applying an electrical voltage. If these are alternating voltages, piezoelectric bodies perform *mechanical oscillations*. These can be measured in the direction of stress (longitudinal oscillations,

E. Hering (✉)

Hochschule Aalen (Rektor i. R.), Aalen, Germany

e-mail: Eckbert.Hering@hs-aalen.de

K.-E. Biel

PIL Sensoren GmbH, Erlensee, Germany

U. Guth · W. Vonau

Kurt-Schwabe-Institut für Messtechnik, Meinsberg, Waldheim, Germany

e-mail: guth@ksi-meinsberg.de; vonau@ksi-meinsberg.de

M. Liess

Hochschule RheinMain, Rüsselsheim, Germany

e-mail: Martin.Liess@hs-rm.de

Fig. 2.1 Principle of the longitudinal, piezoelectric effect

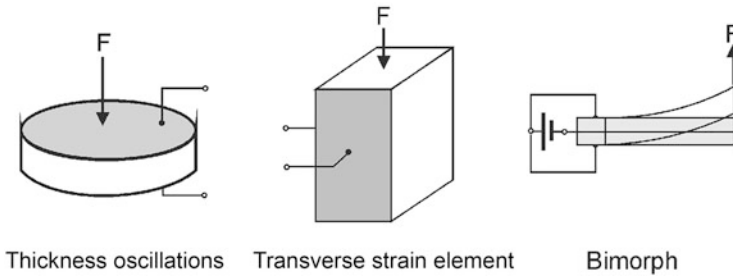
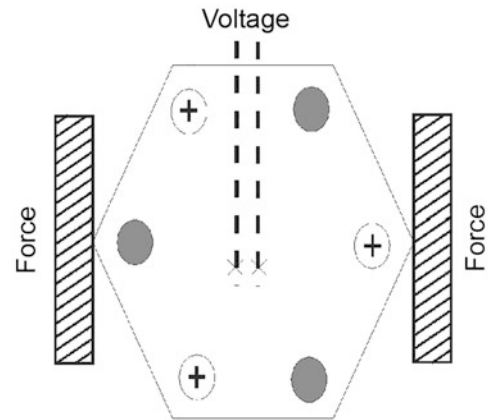


Fig. 2.2 The three principles of the piezoelectric effect

thickness transducer) or perpendicular to the stress direction (transverse oscillations, *transverse strain element*). In the following, only the *direct piezoelectric effect* will be discussed.

Figure 2.3 describes the piezoelectric effect in space. A force F causes a charge displacement in the longitudinal direction (1: 1, 2, 3) and at an angle around the axes (1: 4, 5, 6). The polarization directions are k : 1, 2, 3. The force F then causes a charge shift Q to which applies:

$$Q = d_{kl} F,$$

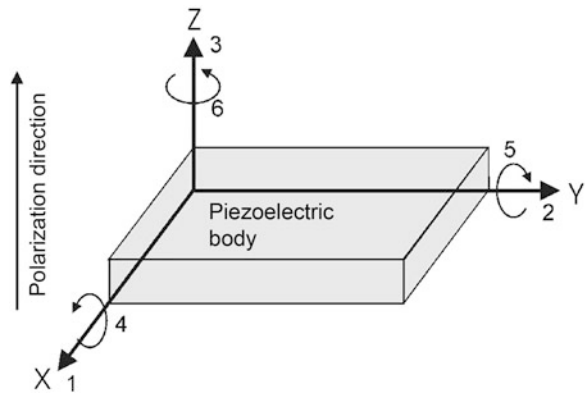
where d_{kl} is the *piezoelectric coefficient*.

Between polarization P and the mechanical tension σ applies:

$$P = e \sigma,$$

where e is the *piezoelectric elongation coefficient* and applies to σ :

Fig. 2.3 Coordinate system for describing the piezoelectric effect



$$\sigma_{ij} = \frac{F_i}{A_j},$$

where A is the area and F denotes the force. The direction of the force is denoted by the indices i (1,2,3) and the direction of the area by the indices j (1,2,3). The piezoelectric effect is *temperature-dependent* and partly strongly nonlinear.

A constant force, which does not cause any additional geometrical change of the body, does not cause a piezoelectric effect. With the piezoelectric effect, therefore, only the *temporal change* of the force is proportional to the measured current. If integration is carried out over this current after time, the shifted charge ($Q = \int I(t)dt$) is obtained, which is proportional to the change in force according to the above equation. By means of special circuit electronics, however, quasi-stationary processes of a few minutes can also be measured.

2.1.2 Materials

Only in *nonconductive materials*, the piezoelectric effect occurs. These materials must not have a center of symmetry because otherwise, no shift of the charges is possible. (In a point mirroring, the crystal is transferred into itself.) A distinction is made between the materials:

- *Piezoelectric crystals*

These include the α crystal (SiO_2), tourmaline, lithium niobate (LiNbO_3), lithium tantalate (LiTaO_3), gallium orthophosphate (GaPO_4), barium titanate (BaTiO_3 ; BTO) and lead zirconate titanate (Pb,Zr,TiO_3 ; PZT). These crystals have very low piezoelectric coefficients; however, they exhibit higher temperature stability, lower losses and a smaller hysteresis curve.

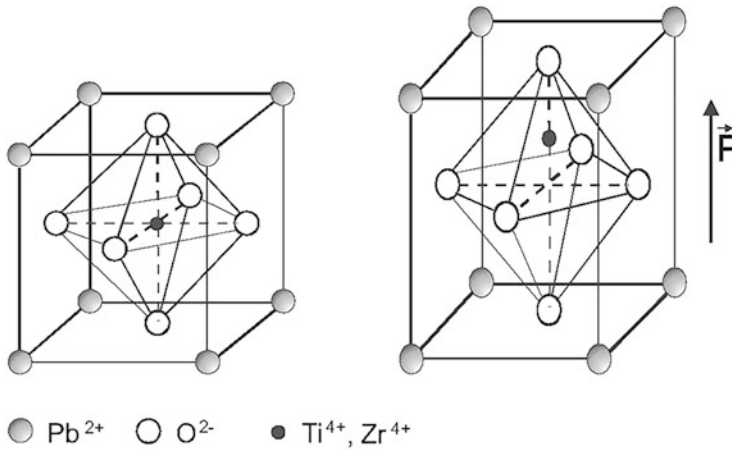


Fig. 2.4 Perovskite structure of piezoelectric ceramics

- *Piezoelectric ceramics*

Most piezoelectric materials are manufactured synthetically. Typical representatives are lead zirconate titanates (PZT). They crystallize, as shown in Fig. 2.4, in a *perovskite crystal structure*.

- *Other piezoelectric materials*

It is possible to deposit piezoelectric thin films, such as zinc oxide (ZnO) or aluminum nitride (AlN), as thin films on silicon using semiconductor technology. Polyvinylidene fluoride (PVDF) is also used.

Table 2.1 shows the piezoelectric coefficient of the selected materials.

2.1.3 Applications

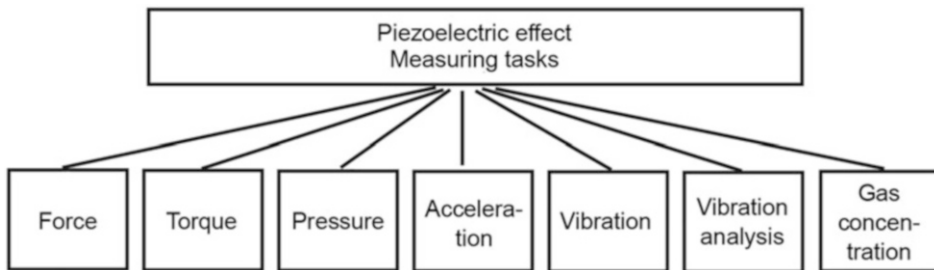
Figure 2.5 shows the most important fields of application. The specific applications are described in detail in Chap. 3.

Piezoelectric materials are also used in many special applications, of which only a few can be mentioned here:

- Injection nozzles for diesel engines,
- Print heads in inkjet printers,
- Micro and nanopositioning systems and
- Microscopy (scanning electron, atomic force and scanning tunneling microscopes).

Table 2.1 Piezoelectric coefficient of selected materials

Material	Piezoelectric coefficient d_{ij}	Value in pC/N
Lithium niobate	d_{22}	0.67
Tourmaline	d_{33}	1.83
Quartz	d_{11}	2.3
Lithium tantalate	d_{33}	9.2
PVDF film	d_{13}	23
PZT ceramics	d_{33}	593

**Fig. 2.5** Application fields of piezoelectric materials

2.2 Resistive and Piezoresistive Effect

2.2.1 Operating Principles and Physical Description

The *resistive* effect describes the dependence of the specific electrical resistance ρ of a conductor from the mechanical stress σ ($\sigma = F/A$ where F is the force acting on the vertical surface A). For metals, the *specific electrical resistance* ρ is *independent of voltage* σ , as shown in Fig. 2.6.

The electrical resistance R of a metallic conductor of the length l_0 and the cross section A is:

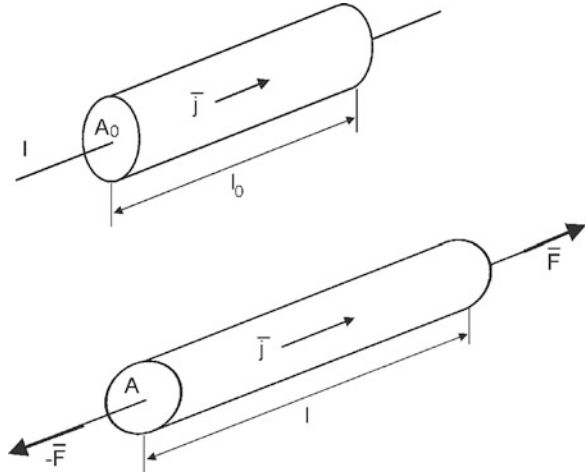
$$R = \rho \frac{l_0}{A_0}.$$

That is to say:

- The resistance increases as the length increases (R proportional to l_0).
- The resistance decreases as the cross section increases (R proportional to $1/A_0$).

This means that if an electrical conductor is stretched, that is, if its length increases and at the same time its cross section decreases, then the electrical resistance increases. If the

Fig. 2.6 Conductor under mechanical stress



material is stressed in the elastic range, this also applies in reverse: if an electrical conductor is compressed (shortening of the length and increasing of the cross section), the electrical resistance decreases. These effects are observed with *strain gage* (SG; Sect. 4.3).

If mechanical stress acts on the metallic conductor, the following applies:

$$\sigma = E \frac{\Delta l}{l} = E \varepsilon,$$

where E is the modulus of elasticity and $\Delta l/l$ is the relative length change (*elongation* ε).

If a force F acts on the conductor area A , the same applies to the resistance R_F :

$$R_F = \rho \frac{l_F}{A_F}.$$

The change in resistance is then calculated:

$$\frac{\Delta R}{R} = k \frac{\Delta l}{l} = k \varepsilon,$$

where for the same volume k is about 2 (the factor k stands for *sensitivity*). This effect is called the *resistive effect* and depends only on the change of the geometrical quantities length and cross section. It finds its application in *SGs* (Sect. 4.3). In semiconductors, such as silicon, the specific electrical resistance ρ changes when a mechanical stress σ is applied (*piezoresistive effect*, Sect. 2.2.3). The reason is that mechanical stresses influence the mobility of the charge carriers and the occupation probabilities of the conduction and valence bands.

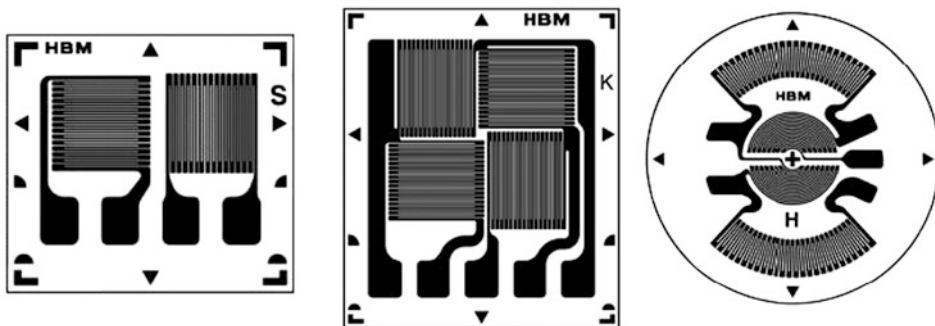


Fig. 2.7 Foil strain gages (SGs) of different geometries (Factory photo: HBM)

2.2.2 Resistive Effect and Its Application by Means of SGs

Foil SGs are used most frequently. On a thin plastic foil (usually made of polyimide), a wafer-thin resistance wire (3 μm to 8 μm thick) is meandered as a *measuring grid* and provided with electrical connections (Fig. 2.7). In most cases, the SGs also have a plastic film on their surface to protect the measuring grid mechanically. A wide variety of geometries can be used for a wide range of applications (Fig. 2.7). The SGs are optimized accordingly for each area of application.

Table 2.2 shows the k-factors of different SG materials.

The resistance change ΔR of a commercially available SG is relatively small. For an SG of 120 Ω and an elongation of 1/1000, the change in resistance $\Delta R = 0.24 \Omega$. These changes can be easily determined with appropriate measuring devices. An SG is attached to the application either *spot welding* (rather rarely) or by *gluing*. Depending on the temperature, *hot-* or *cold-hardening adhesives* are used. These mainly include cyanoacrylates, methyl methacrylates and epoxy resins. The field of application is often determined by the temperature resistance of the adhesives.

The resistance change ΔR is usually determined by a *Wheatstone bridge circuit* (Fig. 2.8).

The four branches of the bridge circuit contain the resistors R_1 to R_4 . At points 2 and 3, a *bridge supply voltage* U_B is applied, which leads to a *bridge output voltage* U_A . The voltages U_A and U_B are divided into the two bridge sections R_1 and R_2 and R_3 and R_4 as follows:

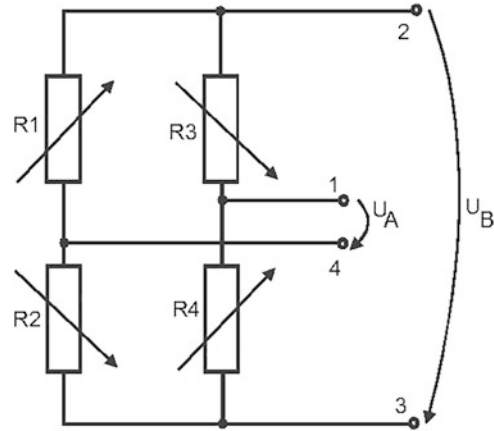
$$\frac{U_A}{U_B} = \frac{R_1}{R_1 + R_2} - \frac{R_4}{R_3 + R_4}.$$

If all resistors have the same size ($R_1 = R_2 = R_3 = R_4$), the bridge is *synchronized*, that is, $U_A = 0$, and it applies: $R_1/R_2 = R_4/R_3$.

Table 2.2 Sensitivities (k-factors) of different strain gage materials

Designation	Composition	k-factor
Constantan	54Cu, 45Ni, 1Mn	2.05
Karma	73Ni, 20Cr, rest Fe and Al	2.1
Nichrome V	80Ni, 20Cr	2.2
Chromol C	65Ni, 20Fe, 15Cr	2.5
Platinum tungsten	92Pt, 8 W	4.0
Platinum	100 Pt	6.0

Fig. 2.8 Wheatstone bridge circuit



If the resistances change, an output voltage U_A occurs. In addition, if the change is relatively small ($\Delta R_i \ll R_i$), as is the case with the SG, the following applies:

$$\frac{U_A}{U_B} = \frac{1}{4} \left(\frac{\Delta R_1}{R_1} - \frac{\Delta R_2}{R_2} + \frac{\Delta R_3}{R_3} - \frac{\Delta R_4}{R_4} \right).$$

With $\frac{\Delta R}{R} = k\varepsilon$ applies: $\frac{U_A}{U_B} = \frac{k}{4} (\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4)$.

This means: The change contributions of adjacent SGs are subtracted if they have the same sign. If they have different signs, they add up.

Applications

According to Fig. 2.9, two main areas of application can be distinguished:

- Transducers for the *mechanical measured variables*: mass, force, torque and pressure (Chap. 4) and
- Possibilities of *experimental analysis*. These analyses are used both in research and development for the optimization of materials and their fields of application and in practice in the automotive, aerospace, rail and construction industries (e.g. settlement or



Fig. 2.9 Areas of application for SG bending beams (Factory photo: HBM)

geodynamic control measurements in tunnel construction). Frequently, stress analyses are used to verify stress simulations with finite elements experimentally.

2.2.3 Piezoresistive Effect and Its Application by Silicon Semiconductor Elements

Due to the anisotropy of the Si crystal, the changes in the specific electrical resistance are also strongly dependent on direction. This is described by the *tensor* of the *piezoresistive constants*:

$$\begin{Bmatrix} r_1 \\ r_2 \\ r_3 \end{Bmatrix} = \begin{Bmatrix} \pi_{11} & \pi_{12} & \pi_{12} \\ \pi_{12} & \pi_{11} & \pi_{12} \\ \pi_{12} & \pi_{12} & \pi_{11} \end{Bmatrix} \cdot \begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{Bmatrix}.$$

They are r_i the changes of the specific electrical resistance: $r_i = \Delta\rho_i/\rho_i$ and π_{ij} the *piezoresistive constants*. These depend on the crystal direction, the doping (n-type or p-type) and the temperature. Table 2.3 shows the values for room temperature.

As a rough guide value for silicon, one can expect that at a mechanical stress σ of 1 GPa, the change of the specific electrical resistance $\Delta\rho/\rho$ is about 10%.

2.2.4 Materials

The most important material is silicon. The following parameters determine, as already mentioned, the properties of the piezoelectric effect:

- type of doping (p-type or n-type doping),
- doping density and

Table 2.3 Piezoresistive constants as a function of direction and doping (at room temperature: $T = 300$ K)

	ρ in Ωcm	π_{11} in 10^{-11} Pa^{-1}	π_{12} in 10^{-11} Pa^{-1}
p-Si	7.8	6.6	-1.1
n-Si	11.7	-102	53.4

- crystal direction and temperature.

The load cases tension and compression, bending and torsion as well as their superpositions result in load cases that can be represented in Cartesian coordinates as tension and shear stresses (Sect. 4.3).

The measuring cells essentially consist of many resistors mounted on a silicon chip. Their application range is between -50 and $+150$ °C. At higher temperatures, the *SOS technology* (SOS: silicon on sapphire) is used in which the resistors are applied to an insulator (in this case a sapphire) and are not diffused into the silicon substrate.

If the tensor of the piezoresistive constant is known, the *membrane* can be integrated into the chip. Location and direction of the resistors and membranes relative to the crystal direction in the silicon chip form the sensor. The membrane is deflected by pressure or tension, resulting in a change in resistance. Because the semiconductor resistors are strongly temperature-dependent, a *Wheatstone bridge* circuit is usually used (Fig. 2.8). Figure 2.10 shows how the influence of temperature can be compensated (Fig. 2.10a) and how an external pressure can be measured by the sensor as a change in resistance.

The bridge circuit offers the following possibilities, which are frequently used in practice:

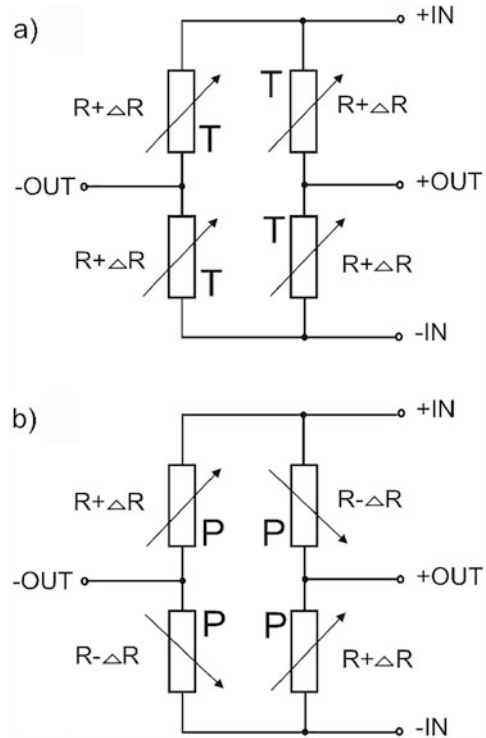
- four resistors in the border area of the membrane (e.g. measure two radial voltages and two tangential voltages),
- four resistors in the center of the membrane and
- two resistors each in the border area and two in the center of the membrane.

Applications

The piezoresistive effect is used to measure *pressure, force, torque* and *elongation* (Chap. 4) is used. Silicon sensors have the following special features:

- *higher accuracy* than other materials and methods, in particular those of SGs (Sect. 2.4),
- very cost-effective manufacturing method (about 10 times cheaper than thin-film sensors or SGs)
- limiting the maximum measurable force to 1 GN and
- limited formability, as silicon is a brittle material.

Fig. 2.10 Wheatstone bridge circuit (a) to compensate for temperature, (b) to measure pressure



2.3 Magnetostrictive Effect

2.3.1 Principle of Operation and Physical Description

If an external magnetic field is applied to a material through which a current flows and the electrical resistance changes as a result, a magnetostrictive effect is present. The strength of the magnetostrictive effect $\Delta R/R$ is determined by the quotient of the change in resistance ($R(H) - R(0)$) and the resistance without magnetic field $R(0)$ is described:

$$\frac{\Delta R}{R} = \frac{R(H) - R(0)}{R(0)}$$

$R(H)$: electrical resistance in a magnetic field H .

Figure 2.11 shows an overview. Magnetostrictive effects are found in nonmagnetic and magnetic materials.

In nonmagnetic materials, electrons must be mobile. If an external magnetic field $B = \mu_0 H$ (μ_0 : magnetic field constant = $4\pi \cdot 10^{-7}$ (Vs)/(Am), H : magnetic field) acts on the electric charge Q , which can change at the speed v , then the *Lorentz force* F_L is created.

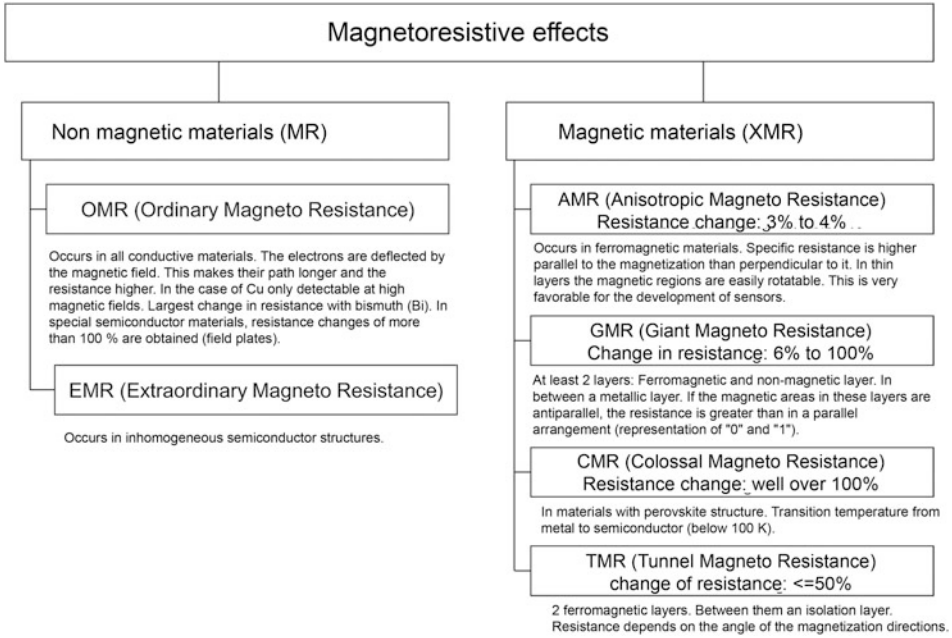


Fig. 2.11 Overview of magnetoresistive effects

This force acts perpendicular to the vector of velocity \mathbf{v} and perpendicular to the magnetic flux vector \mathbf{B} so that the following applies:

$$\mathbf{F}_L = Q(\mathbf{v} \times \mathbf{B}).$$

This Lorentz force causes the electrons to be deflected. Therefore, their path through the material becomes longer and thus the resistance increases. With metals, the change in resistance is only noticeable at high magnetic fields; with semiconductors, because of the greater mobility of the electrons, it is already noticeable at low magnetic fields. Different effects occur with magnetic materials, as shown in Fig. 2.12.

2.3.1.1 AMR (Anisotropic Magneto Resistance)

In ferromagnetic materials, there are crystal regions with the same magnetization \mathbf{M} , the *Weiss domains*. If a current \mathbf{I} flows through a body with the length l , the width b and the thickness d , the following relationships result (in case the electric field \mathbf{E} is parallel to the current density \mathbf{j} (Fig. 2.12)):

$$E = E_p \cos \beta + E_s \sin \beta \text{ as well as } j_p = j \cos \beta \text{ and } j_s = j \sin \beta.$$

(index p: parallel and index s: vertical)

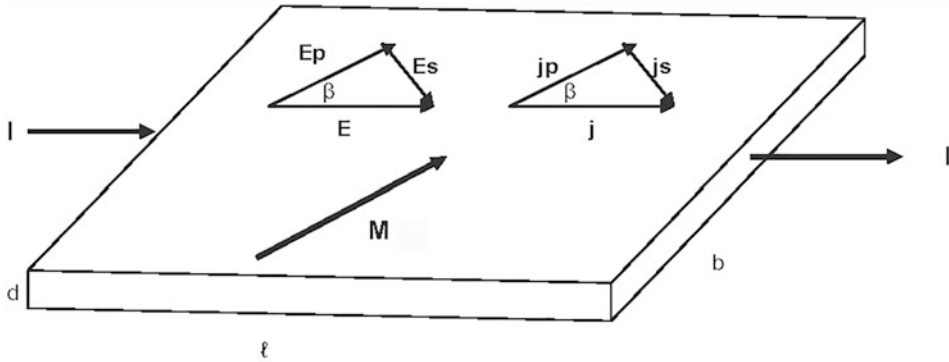


Fig. 2.12 Dependence of the resistance on the angle between magnetization and current flow

According to Ohm's law:

$$E_p = \rho_p j_p \text{ and } E_s = \rho_s j_s \text{ (} \rho \text{ : specific electrical resistance).}$$

If the above correlations are taken into account, the result is

$$E(\beta) = j\rho_s \left(1 + \frac{\rho_p - \rho_s}{\rho_s} \cos^2 \beta \right).$$

The angle dependence of the resistance $R(\beta)$ is then calculated as follows:

$$R(\beta) = \frac{l}{bd} \rho_s + \frac{l}{bd} (\rho_p - \rho_s) \cos^2 \beta.$$

Through further calculations and taking into account that the y-component of the magnetization M_y has the same direction as the external magnetic field H (vector $\mathbf{H} = 0, H_y, 0$), the dependence of the longitudinal resistance $R(H_y)$ on the applied magnetic field H_y :

$$R(H_y) = \frac{l}{bd} \rho_s \left(1 + \frac{\rho_p - \rho_s}{\rho_s} \left[1 - \left(\frac{H_y}{H_k} \right)^2 \right] \right) \text{ for } |H_y| \leq H_k \text{ and}$$

$$R(H_y) = \frac{l}{bd} \rho_s \text{ for } |H_y| > H_k.$$

(H_k : critical magnetic field, that is, all Weiss domains are oriented in the direction of the external magnetic field (turned upside down))

The characteristic curve of the resistance as a function of the magnetic field is quadratic, as shown in the above formula. In order to linearize this characteristic curve, metallic layers

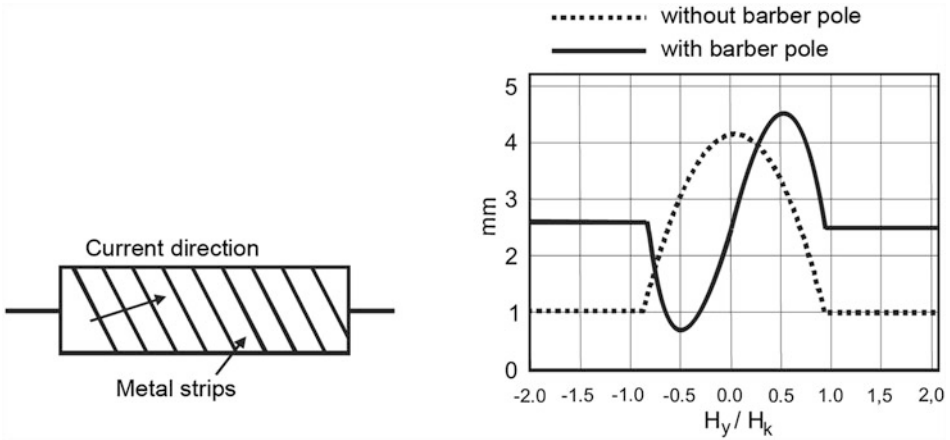


Fig. 2.13 AMR sensor in barber pole structure (left). Characteristic curve without barber pole (solid) and with barber pole (dashed)

(e.g. Al on permalloy) are deposited on the material at an angle of 45° (to the current direction I , which is usually parallel to the longitudinal edge of the material) (*barber pole arrangement*). Thus, the current direction I turned by 45° against the resistance axis. This characteristic curve is then linear for $H_y \ll H_k$ and shows the following relationship:

$$R(H) = R_0 + \Delta R \frac{H_y}{H_k} \sqrt{1 - \left(\frac{H_y}{H_k}\right)^2}.$$

Figure 2.13 shows the relationship. The barber pole arrangement is shown on the left. In the right picture, the continuous characteristic describes the quadratic dependence of the resistance. The dashed curve is linear (within the limits given above) due to the barber pole arrangement.

2.3.1.2 GMR (Giant Magneto Resistance)

The GMR effect is a quantum mechanical effect between thin ferromagnetic and nonmagnetic layers (Fig. 2.14). The electrons rotate around themselves. Therefore, they have an intrinsic angular momentum, the *spin*. This acts like a small magnet. In ferromagnetic materials, the spin magnets are magnetically active outward. In layers with the parallel magnetization of the ferromagnetic material, the electrons move almost unhindered across the layers (*lower resistance*). If the layers have *antiparallel* aligned spins, the electrons are strongly scattered during their movement through the layers (*greater resistance*). Figure 2.15 shows this effect.

The structure of a GMR component, for example, consists of three thin layers of about 2 nm thickness. Between two ferromagnetic layers (e.g. Fe), there is a metallic layer called *spacer* (e.g. Cr). Figure 2.15 shows the path of an electron through the layers as a dashed

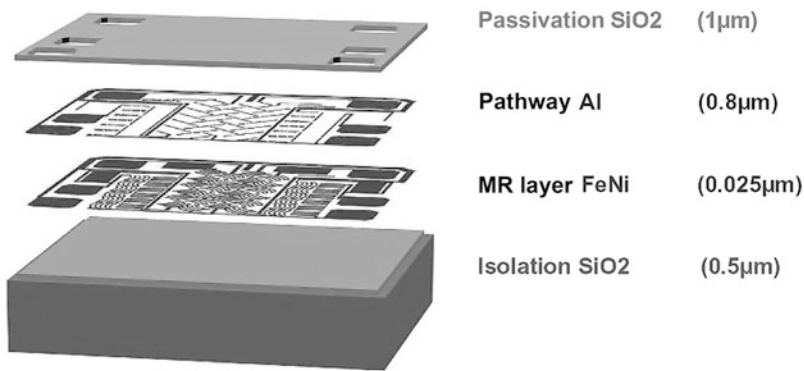


Fig. 2.14 Typical structure of an AMR sensor (Source: Sensitec)

line. In the antiparallel arrangement, the electron is scattered much more and travels a longer path. This results in a greater change in resistance. For the Fe/Cr-GMR sensor, the resistance ratio $\Delta R/R$ is about 20% higher at room temperature and about 80% higher at helium temperature (4.2 K).

Higher resistance changes are achieved by the fact that the two ferromagnetic layers are different (*spin-valve principle*): one layer is *magnetically soft* (easy alignment of the elementary magnets in a magnetic field) and *another* is *hard magnetic* (alignment of the elementary magnets only with higher magnetic field strengths). The layer systems based on the spin-valve principle work like *spin valves*: They allow the electrons to pass faster in parallel magnetization (valve is open) than in antiparallel magnetization (valve is closed). A higher resistance change can also be achieved by a multilayer arrangement (Fig. 2.16).

Table 2.4 shows the main differences in the properties of AMR and GMR sensors.

2.3.1.3 CMR (Colossal Magneto Resistance)

Special materials with perovskite structure create transitions from metal to semiconductor. This enables very large changes in resistance (up to over 1000).

2.3.1.4 TMR (Tunnel Magneto Resistance)

Two ferromagnetic layers are separated by an *insulating layer*. Through this insulating layer, the electrons can *tunnel*, while their spin remains constant in magnitude and direction (Fig. 2.17). The layer thickness is usually about 100 nm. Normally, resistance changes of up to 50% occur. Resistance changes of up to 1000 have already been observed in the laboratory. However, the applications are not yet ready for the market.

2.3.2 Advantages of XMR Technology

Table 2.5 provides an overview of the advantages of XMR technology.

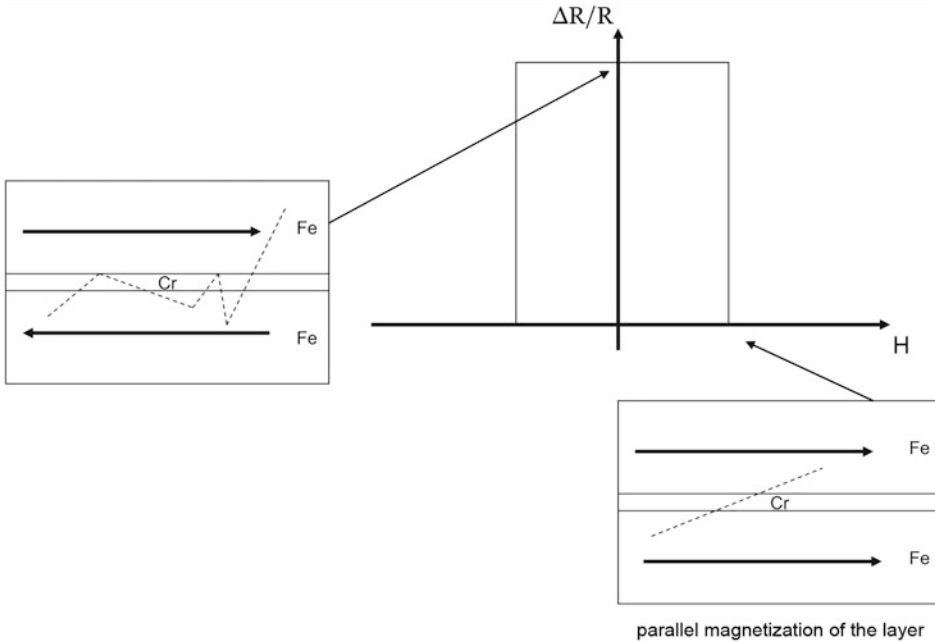


Fig. 2.15 Structure of a GMR sensor and mode of action of the GMR effect

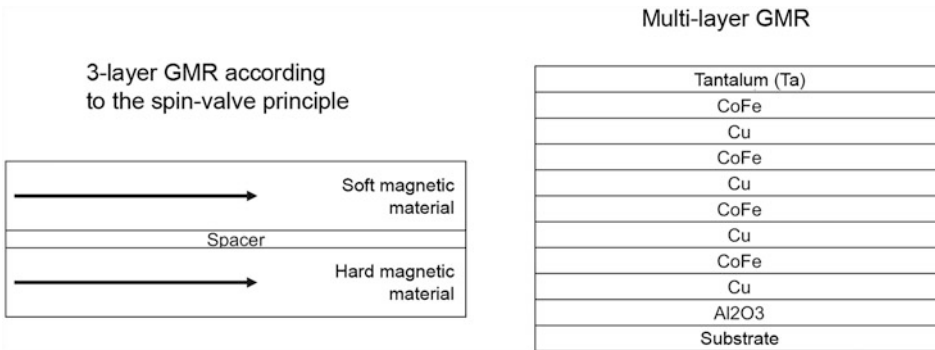


Fig. 2.16 Three-layer GMR based on the spin-valve principle (left) and multilayer GMR (right)

Table 2.4 Differences between AMR and GMR sensors

Property	GMR	AMR
Size	Very small	Small
Costs	High	Low
Sensitivity	High	High
Output signal	Very good	Good
Temperature stability	Very good	Good
Power consumption	Very low	Medium

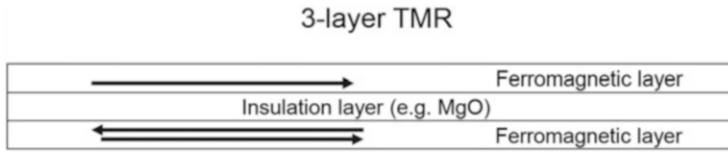


Fig. 2.17 Principle structure of a TMR sensor (in the lower layer the electron and its spin is tunneled)

Table 2.5 Advantages of XMR technology

Areas	Advantages
Possible applications	<p>High sensor resistances ($> 1 \text{ k}\Omega$):</p> <ul style="list-style-type: none"> • Use in battery-powered systems • Use in low power systems <p>Read heads in magnetic storage devices</p> <p>Automotive industry:</p> <ul style="list-style-type: none"> • Safety (e.g. ESP, ESR, ARS, ABS, airbag) • Driver assistance systems • Engine management (e.g. NO_2 sensor, λ probe) • Comfort functions (e.g. temperature, humidity)
Contactless measurement procedure	<ul style="list-style-type: none"> • No mechanical wear of the measuring system • No encapsulation necessary due to environmental influences
Thin layers	<ul style="list-style-type: none"> • Use at higher temperatures • Advantages in space projects (shielding of cosmic radiation; miniaturization, low weight) • Production on Si wafers. Thus, very small components and integration of the system electronics
Minor errors	<ul style="list-style-type: none"> • Suppression of homogeneous foreign fields. • Compensation of nonlinearities of the magnetic fields. • High reproducibility. • Self-diagnosis of the sensor (monitoring of the output amplitudes by the sum formula: $\sin^2\alpha + \cos^2\alpha = 1$). • No temperature dependence of the amplitude. • High accuracy
Costs	<ul style="list-style-type: none"> • Favorable manufacturing costs of the mechanical components and the guides • Favorable manufacturing costs for mass production in connection with Si wafers
Flexibility	<ul style="list-style-type: none"> • Development and adaptation for certain applications through the choice of layer structures (“spin engineering”)
Robust	<ul style="list-style-type: none"> • Insensitive to dust, moisture and oil. • Easy adjustment

Table 2.6 Application areas of XMR technology (sorted by most common use)

Areas	Measuring task
Geometric (Section 3)	1. Angle 2. Inclination 3. Filling level 4. Length and position measurement
Dynamic (Section 5)	1. Speed 2. Flow 3. Flow rate
Thermal (Section 6)	1. Temperature 2. Filling level 3. Energy
Electrical (Section 7)	1. Electric current 2. Magnetic field
Hard disk drives	Read-write head systems

2.3.3 Applications of XMR Technology

Table 2.6 provides an overview of the main fields of application.

Some special applications are presented below.

Angle Measurement (Fig. 2.18)

Angle measurement

MR angle sensor with incremental output

Arrangement on the shaft circumference in a motor

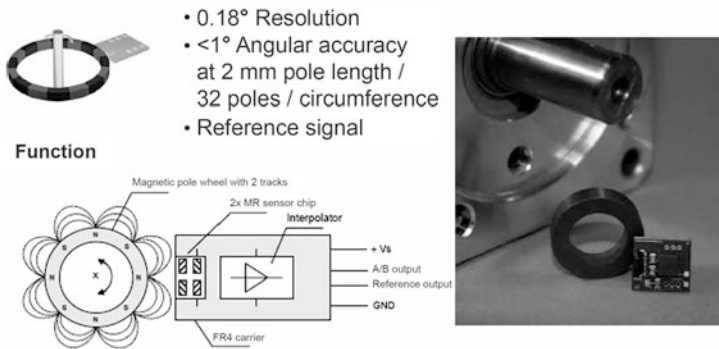


Fig. 2.18 Angle sensor (Factory photo: Sensitec)

Length Measurement (Fig. 2.19)

Length measurement

Measuring principle of an AMR sensor for length measurement



By shifting the resistors in the magnetic field, their resistance values change according to the field direction.

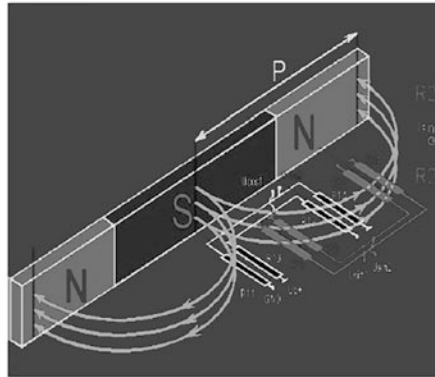
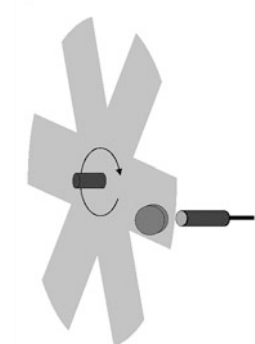


Fig. 2.19 Length sensor (Factory photo: Sensitec)

Speed Measurement (Fig. 2.20)

Fig. 2.20 Speed sensor
(Factory photo: Sensitec)

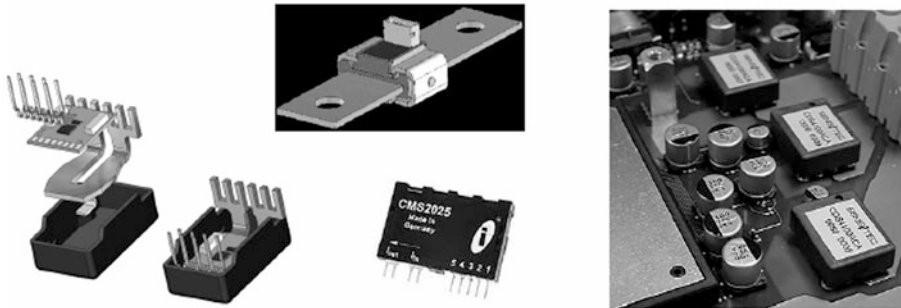
Speed measurement



Current Measurement (Fig. 2.21)

Current measurement

Measuring task: Potential-free current measurement from 1A to kA

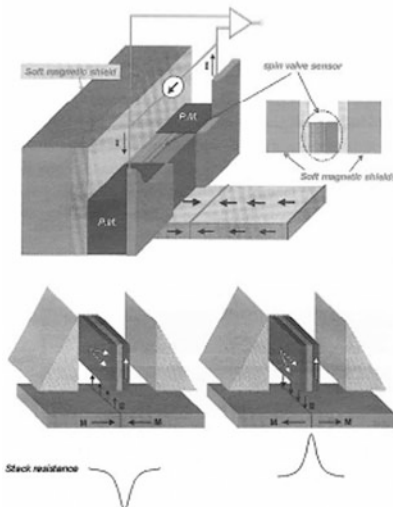


(Source: Sensitec)

Fig. 2.21 Current sensor (Factory photo: Sensitec)

Read-write Heads with the GMR Effect (Fig. 2.22)

Reading process with GMR sensors



- GMR sensor is perpendicular to the surface of the HD disc and the scan direction
- A magnetic field is seen only at the border between two adjacent oppositely magnetized areas of the HD disk
- Magnetization of the soft magnetic layer of the GMR sensors is parallel to the disk surface for magnetic field-free operation

Fig. 2.22 GMR effect for the read-write heads (Source: IBM)

2.4 Magnetostrictive Effect

2.4.1 Principle of Operation and Physical Description

If a linear magnetic field is applied in ferromagnetic materials, length changes $\Delta l/l$ occur. This effect is called *magnetostriction*. An *increase in length* results in *positive* magnetostriction, decrease in length in *negative* magnetostriction. The length changes are generally between $-3 \cdot 10^{-5}$ for nickel and $+5 \cdot 10^{-5}$ (this corresponds to an increase in the length of $50 \mu\text{m}$ for a 1 m long rod) for iron. As Fig. 2.23 shows, the change in length is caused by the ferromagnetic regions (*Weiss domains*) rotating in the direction of the externally applied magnetic field. The volume of the body remains constant. It applies to the *elastic deformation* ε :

$$\varepsilon = \Delta l/l = \kappa^* \cdot H/E,$$

where κ^* is the magnetostriction constant, H is the magnetic field strength and E is the modulus of elasticity of the rod.

Figure 2.24 shows other effects of magnetostriction. A linear magnetic field causes a change in length or volume, as shown in Fig. 2.23. After its discoverer, this effect is also called the *Joule effect*. If the magnetostrictive materials are in a coil through which an alternating current flows, the length changes with the frequency of the alternating current. This allows *ultrasonic waves* to be generated. In a reversal of the Joule effect, magnetic fields can be altered by changes in length or volume (*Villari effect*). In a *spiral-shaped* magnetic field, a *torsional motion* takes place (*Wiedemann effect*). This *torsional wave*

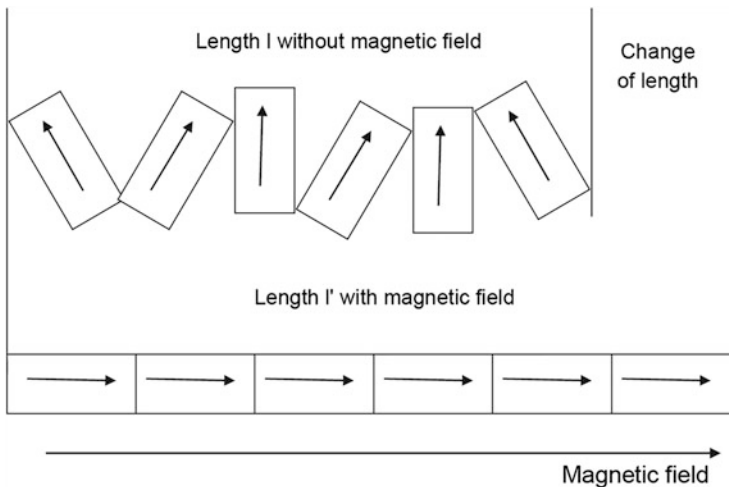


Fig. 2.23 Explanation of magnetostriction by the change in length

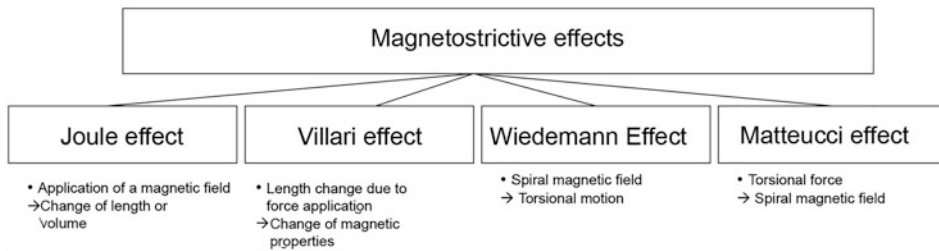


Fig. 2.24 Effects of magnetostriction

Table 2.7 Magnetostriction of different materials

Material	Maximum length change $\Delta l/l \cdot 10^{-6}$	Effect
Fe ₉₀ Al ₄	24	Joule
Fe ₈₄ Al ₁₆	86	Joule
Fe ₆₀ Co ₄₀	147	Joule
Fe ₈₃ Ga ₁₇	207	Joule
FeTb _{0.3} Dy _{0.7} Fe ₂ (Terfenol-D)	1600	Joule
Fe ₅₅ Pd ₄₅	148	Wiedemann
Fe ₇₀ Pd ₃₀	12,000	Wiedemann
NiMnGa	50,000	Wiedemann

propagates at a speed of 2800 m/s. In the opposite case, a torsional movement causes a spiral-shaped magnetic field (*Matteucci effect*).

Table 2.7 lists the materials and their magnetostrictive effects.

2.4.2 Advantages of Magnetostrictive Sensor Technology

The use of magnetostrictive sensors has the following advantages:

- Contactless and therefore wear- and maintenance-free and long service life.
- Absolute measurement; that is, no approach to reference marks.
- Highly accurate measurement (up to 1 μm).
- Measurement in real-time.
- Many positions can be measured simultaneously and independently (up to 20 positions).
- High repeatability (up to 0.001% or 4 μm).
- Very good interference resistance (shock resistance, insensitive to vibrations).
- Insensitive to extreme temperatures.
- Insensitive to dirt and moisture.
- Quick installation without calibration.

- Wide range of applications (medical technology, mechanical and plant engineering, mobile machinery, means of transport).

2.4.3 Applications of Magnetostrictive Sensor Technology

Figure 2.25 shows which physical quantities can be measured with magnetostrictive sensors.

For displacement, position and level measurement, the *Wiedemann effect* is used: A spiral-shaped magnetic field generates a torsional wave that travels through the rod at a speed of 2800 m/s. The time between the transmission and reception of the wave is measured. According to the relation: distance = speed · time, the distance is calculated. Figure 2.26 shows the principle arrangement. The sensor consists of the following four essential components:

- A waveguide as a measuring element,
- Position-setting permanent magnet,
- Torsional pulse converter and
- Sensor electronics.

Speeds can also be measured with these sensors.

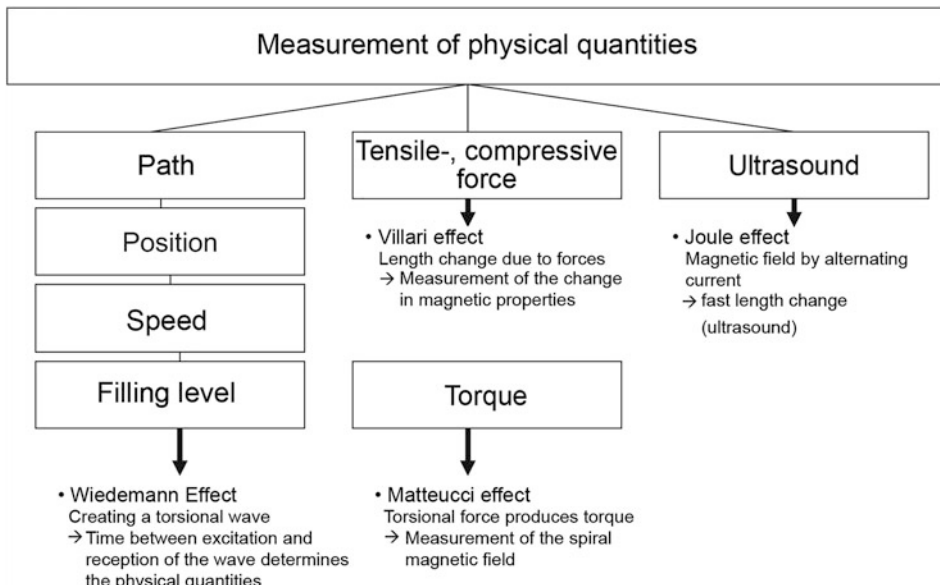


Fig. 2.25 Measurement of physical quantities with magnetostrictive sensors

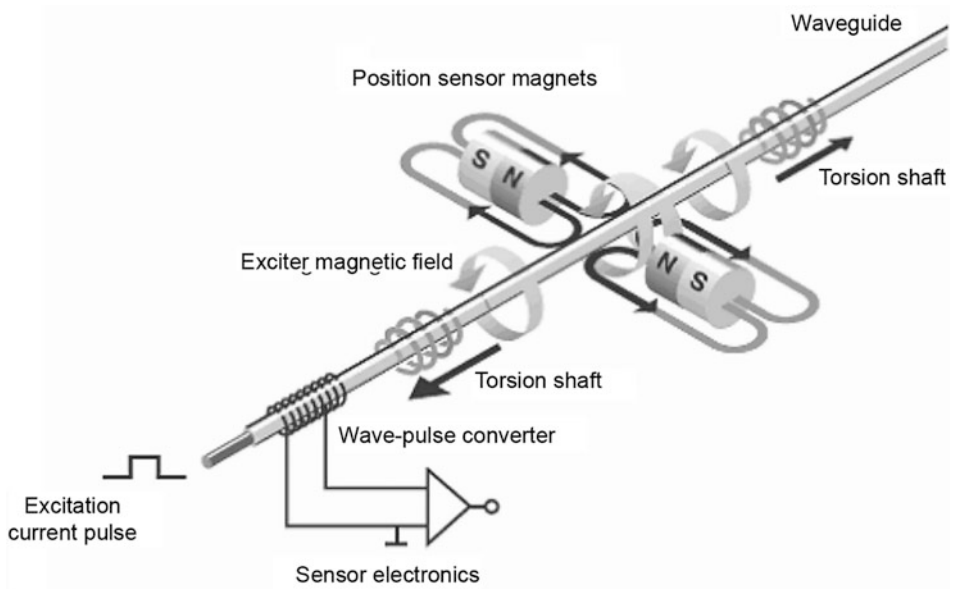


Fig. 2.26 Design of a magnetostrictive sensor for position determination (Factory photo: Balluff GmbH)

Tensile and *compressive forces* can be measured with the *Villari effect* (Sects. 4.2 and 4.4). The forces cause a change in length, which in turn changes the magnetic field. The *Matteucci effect* is used to measure *torque* (Sect. 4.5). The torsional force causes a spiral-shaped magnetic field, which is used for evaluation. With the Joule effect, measurements can be made with ultrasound. Table 2.8 shows some areas of application.

2.5 Effects of Induction

2.5.1 Principle of Operation and Physical Description

The following inductive effects play an important role in the sensors:

- Law of induction and
- Behavior of the inductance of coils in an alternating current circuit.

2.5.1.1 Law of Induction

The law of induction is:

Table 2.8 Application areas of magnetostrictive sensors

Areas	Measuring task
Forming technology	<ul style="list-style-type: none"> • Used in forging presses as it is shock-resistant • Stroke monitoring for hammer presses
Process engineering	<ul style="list-style-type: none"> • Recording of path and speed of piston dosing pumps
Handling technology	<ul style="list-style-type: none"> • Monitoring of the clamping stroke of clamping systems • Monitoring of the welding spot thickness of welding guns
Plastics technology	<ul style="list-style-type: none"> • Multi-position measurement for clamping units of injection molding machines
Mobile hydraulics	<ul style="list-style-type: none"> • Positioning of booms of excavators and crawlers • Electrohydraulic switch adjustments of the rail network without service work • Steering of pleasure boats
Medical technology	<ul style="list-style-type: none"> • Recording the stroke of an injection unit for the finest dosage in minimally invasive surgery

$$U_{\text{ind}} = - \frac{d\Phi}{dt}.$$

This means: Any change in the magnetic flux Φ over time through an area A , which is separated by a border line s , induces a voltage. The magnetic flux Φ through an area A changes if

- the value (magnitude or direction) of magnetic field penetrating the area A changes over time or
- the area A permeated by the magnetic field changes over time.

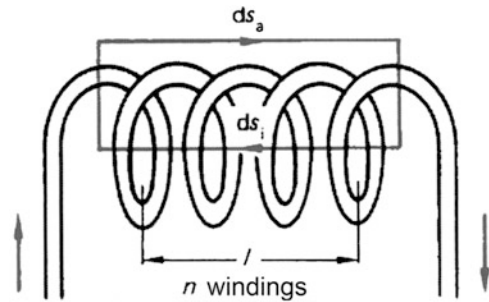
The magnetic flux is the sum of all the magnetic field lines passing through an area dA , which are determined by the magnetic flux density B . Thus, the following applies:

$$\Phi = \int B dA.$$

The *coil* plays an important role as a structural element. It is a circularly wound electrically conductive wire through which a current is passed (Fig. 2.27). This results in a magnetic flux density B , which is constant over the entire cross section of a long coil and can be described as follows:

$$B = \mu_0 \mu_r \frac{NI}{\ell}.$$

Fig. 2.27 Magnetic field of a current-carrying coil



Here N is the number of turns of the coil, I is the current flowing through the coil, ℓ is the length of the coil, μ_0 is the magnetic field constant ($\mu_0 = 4\pi \cdot 10^{-7} \text{ (Vs)/(Am)} \approx 1.257 \cdot 10^{-6} \text{ (Vs)/(Am)}$) and μ_r is the relative permeability (indicates the contribution of matter to the magnetic field).

With this formula, the magnetic flux within the coil is

$$\Phi = \int \mathbf{B} d\mathbf{A} = B \cdot A = \mu_0 \mu_r \frac{NI}{\ell} A = L \frac{1}{N}.$$

The term $\mu_0 \mu_r \frac{N^2}{\ell} A$ is referred to as *inductance* L of a coil.

It therefore applies: $L = \mu_0 \mu_r \frac{N^2}{R_{\text{magn}}} A = \frac{N^2}{R_{\text{magn}}}$.

The magnetic resistance R_{magn} is a measure of the permeability of a material to magnetic flux and is defined as:

$$R_{\text{magn}} = \frac{\ell_m}{\mu_0 \mu_r A}.$$

Figure 2.28 shows a current-carrying coil, which magnetizes a U-shaped iron core. A movable, electrically conductive plate (yoke) is located at a distance s (air gap) of the magnet.

The magnetic resistance R_m has two components, according to Fig. 2.28, the magnetic resistance when passing through the *iron core* R_{m1} and the magnetic resistance in the *air gap* R_{m2} . It applies:

$$R_m = R_{m1} + R_{m2} = \frac{\ell_{\text{Fe}}}{\mu_0 \mu_{\text{Fe}} A} + \frac{2s}{\mu_0 \mu_{\text{Air}} A}.$$

The relative permeability μ_{Fe} is very large and μ_{Air} is about 1. Therefore, the first expression can be neglected and it results for the inductance L :

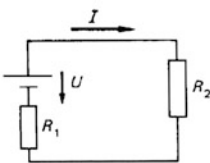
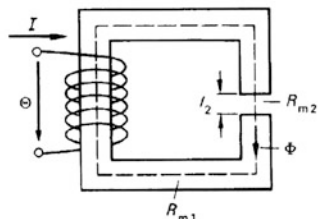
	electrical circuit	magnetic circuit
Sketch		
Cause	electric voltage U in V	magnetic voltage $\Theta = \oint \mathbf{H} ds = NI$ in A
Effect	electric current I in A $I = \frac{U}{R_1 + R_2} = \frac{U}{R_{\text{ges}}}$	magnetic flux φ in Wb $\varphi = \frac{NI}{R_{m1} + R_{m2}} = \frac{\Theta}{R_{\text{mtot}}}$ (4.238)
Ohm's law	$R = \frac{U}{I}$	$R_m = \frac{\Theta}{\varphi}$ (4.239)
Resistance	$R = \frac{l}{\chi A}$ in Ω	$R_m = \frac{l}{\mu_0 \mu_r A}$ in $\frac{\text{A}}{\text{Wb}}$ (4.240)
Conductivity	χ in $\frac{\text{A}}{\text{Vm}}$	$\mu_0 \mu_r$ in $\frac{\text{Wb}}{\text{Am}}$

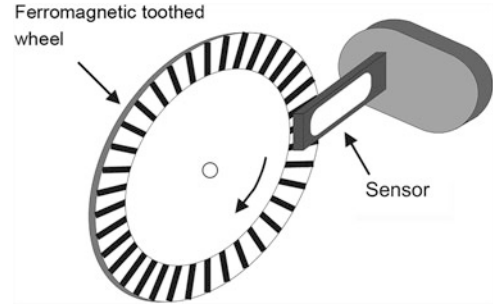
Fig. 2.28 Comparison of the electric and magnetic circuits (Source: Hering, Martin, Stohrer, physics for enineers, 13 edition)

$$L = \frac{\mu_0 AN^2}{2s}.$$

However, this means that the inductance is not linear to the air gap width. A linearization is achieved by operating two coils in a *differential circuit*. Then the change of path Δs becomes proportional to the voltage change ΔU . This circuit is used to measure displacements inductively (displacement sensors, Sect. 3.1.1). These sensors are also called *LVDT* (linear variable differential transformer, Sect. 3.1.1.4) sensors.

As already mentioned above, a voltage is induced when a magnetic field changes. This can be used to measure the rotational speed n (according to the generator principle). At constant angular velocity ω ($\omega = d\varphi/dt$; where φ is the angle of rotation), the armature is penetrated sinusoidally by the field of the permanent magnet with the magnetic flux density B . This induces the following coil voltage u_S :

Fig. 2.29 Ferromagnetic toothed wheel in the field of a coil



$$u_S = K_G(B \sin(\omega t)) \frac{d\phi}{dt}.$$

K_G is a constant that depends on the generator.

For the peak value of the sinusoidal voltage \hat{u}_S , applies with $\omega = 2\pi n$:

$$\hat{u}_S = 2\pi K_G B n.$$

Thus, the peak value of the induced voltage can be used to determine *rotational speed* (Sect. 5.6).

Another way to determine rotational speeds or frequencies using the induction principle is to move a ferromagnetic toothed wheel along a coil or permanent magnet. As Fig. 2.29 shows, a ferromagnetic toothed wheel with Z teeth rotates at an angular velocity ω . Above this toothed wheel, there is a permanent magnet in a coil with N turns. If the tip of the tooth comes close to the magnet, the magnetic flux density becomes maximum (B_{\max}); between the teeth, it is minimal (B_{\min}). The magnetic flux through the coil is approximately sinusoidal. In this case, the following applies to the magnetic flux B :

$$B = \left(\frac{B_{\max} + B_{\min}}{2} \right) + \left(\frac{B_{\max} - B_{\min}}{2} \right) \sin(Z\omega t) = B_0 + \hat{B} \sin(Z\omega t).$$

The induced voltage u_{ind} in the coil with the cross section A is then according to the law of induction:

$$u_{\text{ind}} = N \hat{B} A Z \omega \cos(Z\omega t).$$

The induced voltage is proportional to the speed in its amplitude and frequency. This arrangement serves as a sensor for measuring rotational speed or frequency. This sensor requires no auxiliary power and is extremely robust and insensitive to dirt and temperature fluctuations.

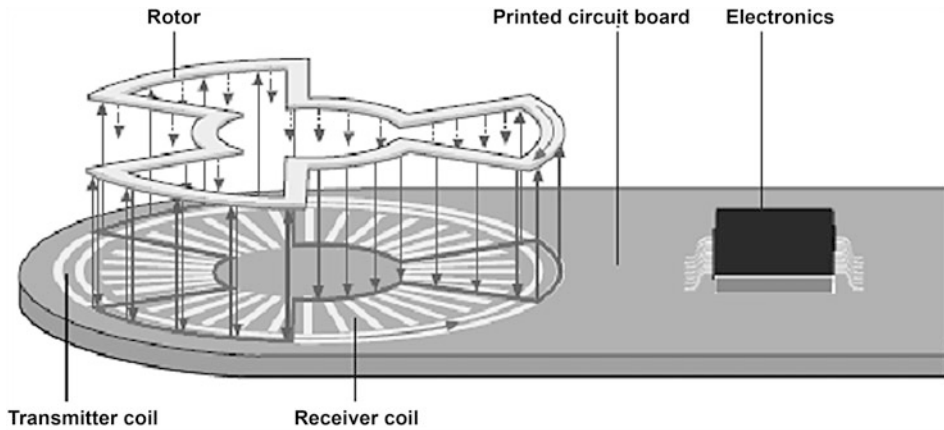


Fig. 2.30 Contactless position sensor (Factory photo: Hella)

2.5.1.2 Generation of Eddy Currents in Electrically Conductive Materials

With this effect, a contactless position sensor can be built. Figure 2.30 shows construction. The stator consists of a transmitter and receiver coil. The rotor is a conductor loop with a certain geometry. It can be realized by a punched part made of an electrically conductive material or by a printed circuit board element. An alternating current flows through the transmitter coils. This generates an electromagnetic field, which influences the rotor. An alternating current is also generated there, which creates an electromagnetic field. This induces voltages in the receiver coils, which depend on the position of the rotor. An electronic system evaluates these voltages and can determine both the position and the angle.

2.5.1.3 Electromagnetic Oscillating Circuits

In this method, the energy losses due to induced eddy currents are measured. The sensor consists of an *LC resonant circuit* (Fig. 2.32 left), which is stimulated by an oscillator. The coil of the sensor element generates a high-frequency electromagnetic field, which exits at the *active surface* of the ferrite core. This electromagnetic field acts over a spatially limited area (*active switching zone*). If a metallic object (*switching lug*) approaches the active switching zone, eddy currents are induced there, which strongly dampen the vibrations. If the vibration amplitude falls below a certain value due to this damping, a comparator responds and outputs an output signal via the output stage. Figure 2.31 shows the design of the sensor.

In this design, the sensor has only two states and therefore acts as an on-off switch:

- undamped oscillation with high amplitude and
- damped vibration with very low or no amplitude.

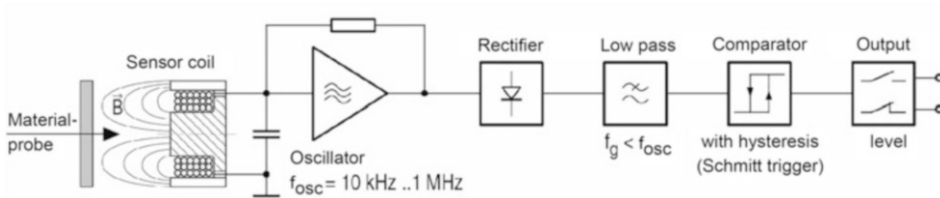


Fig. 2.31 Design of an inductive sensor (teaching material Fachhochschule Düsseldorf)

These operating states can easily be converted into an electrically evaluable signal (current, voltage, frequency).

These sensors are often used in practice as *proximity* or *limit switch* (Sect. 3.4.1). The sensors are often integrated into screws for mechanical engineering applications. The advantages of this sensor are that it works *contactlessly* and after one-time adjustment *wear-free* in *harsh industrial environment* reliably with *high reproducibility*. Another advantage is that the high-frequency electromagnetic field does not cause any measurable heating or magnetic influence.

In the standard DIN (Deutsches Institut für Normung) EN 50010, the parameters are defined. Especially, the distances between the active surface and the standard plate are important. According to DIN EN 50010, the following was defined (Fig. 2.32):

- The *switching distance* s is the distance between the *active surface* of the proximity switch and the *standard measuring plate* (defined in DIN EN 50010).
- The *rated switching distance* s_n is the characteristic value of the sensor. The following variables are derived from this:
 - *Real switching distance* s_r . The following applies: $0.9 s_n \leq s_r \leq 1.1 s_n$.
 - *Usable switching distance* s_u . The following applies: $0.81 s_n \leq s_u \leq 1.21 s_n$.
 - *Assured switching distance* s_a . The following applies: $0 \leq s_a \leq 0.81 s_n$.

2.5.2 Advantages of Inductive Sensor Technology

Inductive sensors offer the following advantages:

- contactless,
- wear- and maintenance-free,
- sweatproof,
- high switching frequencies (up to 18,000 Hz),
- high precision of the circuits,
- high repeat accuracy ($< 1\%$)
- low temperature drift ($< 0.06\%/^{\circ}\text{C}$)

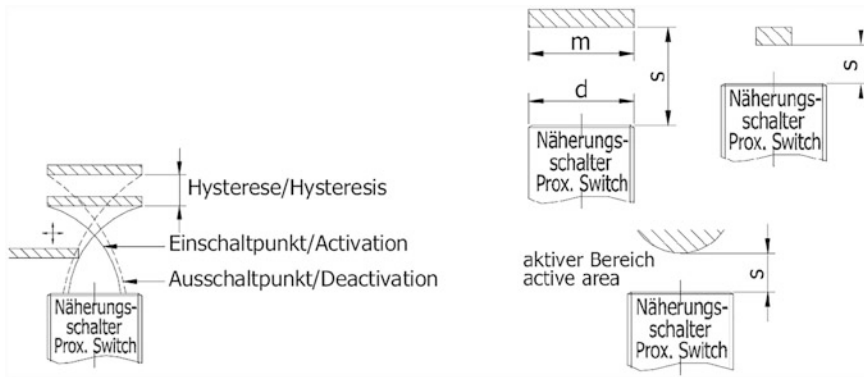


Fig. 2.32 Switching distances according to DIN EN 50010

- high operational reliability,
- large measuring range (from 0.5 mm to 1.1 m),
- high linearity (0.3% of full scale),
- high-temperature resistance ($-60\text{ }^{\circ}\text{C}$ to $+200\text{ }^{\circ}\text{C}$),
- high EMC resistance,
- long service life,
- low power consumption,
- adaptive and programmable,
- cost-effective and
- insensitive in an industrial environment (vibration, dust, humidity, cleaning agents, cooling lubricants, cutting and grinding oils).

2.5.3 Applications of Inductive Sensor Technology

Table 2.9 provides an overview of the main fields of application. In automation technology, inductive sensors are used to reduce manufacturing process costs and increase the availability of the systems. This increases *productivity* in manufacturing.

Figure 2.33 shows installation examples for screws with inductive sensors.

2.6 Effects of Capacitance

2.6.1 Principle of Operation and Physical Description

The following capacitive effects play an important role in the sensors:

Table 2.9 Application areas of inductive sensors

Areas	Measuring task
Automotive industry	<ul style="list-style-type: none"> • Speed measurement (e.g. on crankshaft and gearbox, from 0 revolutions/minute (rpm) to 3000 rpm) • Position of the throttle valve • Pulse generator for ignition • Level sensor (e.g. chassis control, headlight range control) • Electronic driving assistants <ul style="list-style-type: none"> – Steering angle (steer by wire) – Accelerator pedal sensor (E-gas, driver request) – Brake pedal sensor (E-brake)
Mechanical engineering/automation technology	<ul style="list-style-type: none"> • Detection of different materials • Recognition of workpieces of different sizes • Positioning and position control of workpieces or parts (e.g. chip assembly) • Distance measurement (e.g. of tools) • Width and thickness measurement of films and rolls • Gap and distance measurement • Measuring the deformation of a casting mold • Tension control • Measurement of abnormal vibrations • Control of opening and closing (e.g. of press molds) • High-speed measurements (40,000 samples/second) in the production line (e.g. eccentricity of a roller, parallelism of a hard disk drive) • Position detection and control of robot arms
Food/pharmaceutical industry	<ul style="list-style-type: none"> • Checking the can vacuum • Measurement of the vibrations of a filling machine • Detection of fed paper and plastic bags
Process engineering	<ul style="list-style-type: none"> • Conductivity measurement • Flow measurement
Special applications	<ul style="list-style-type: none"> • Underwater use (up to 500-m water depth) • Use at high pressures, extreme ambient conditions, high mechanical stress • Sensors resistant to magnetic fields

- Relationship of the capacitance with electrical and geometrical quantities in the case of direct current and
- Behavior of the capacitance in the alternating current circuit.

2.6.1.1 Capacitor and Capacitance

Capacitors are two mutually insulated, oppositely charged conductor surfaces of any geometry, between which a voltage U exists. The *capacitance* C indicates the charge quantity Q that can be stored at the capacitor surfaces when a voltage U is applied. It therefore applies:

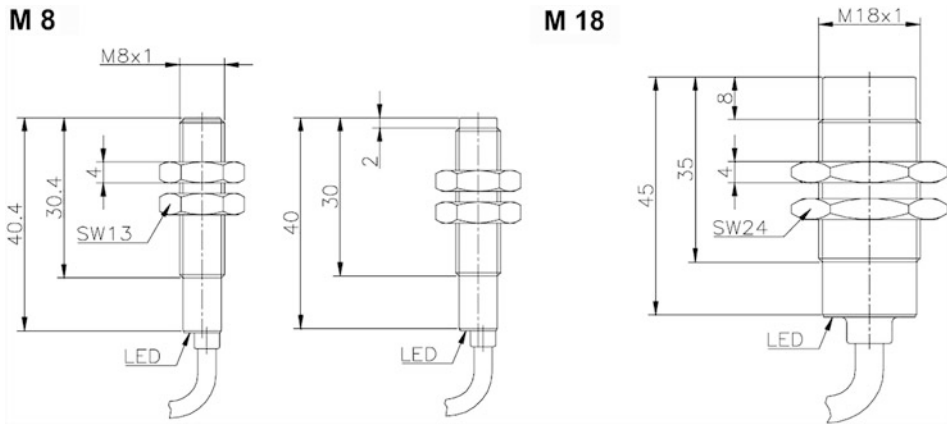


Fig. 2.33 Examples for inductive sensors as screws (Source: Dietz Sensortechnik catalog)

$$C = \frac{Q}{U}. \text{The unit is Farad [F].}$$

If the two-conductor surfaces are plates with the area A at a distance d , then a *plate capacitor* exists. Its capacitance is calculated:

$$C = \epsilon_0 \epsilon_r \frac{A}{d},$$

where ϵ_0 is the electric field constant ($\epsilon_0 = 8.854 \cdot 10^{-12}$ (A s)/(V m)) and ϵ_r is the material-dependent permittivity number (dielectric number, dielectric coefficient). Table 2.10 lists the permittivity numbers of different materials.

According to the formula for a plate capacitor, the capacitance C can be influenced by the following actions (Fig. 2.34):

1. Change of the *plate distance* d . The relationship to capacitance is nonlinear (Fig. 2.34a). It can be linearized by suitable circuits (e.g. differential capacitor) (Fig. 2.34b).
2. Change of *active area* A (the area where the electric field lines end). This relationship with capacitance is linear (Fig. 2.34c).
3. Change of the *dielectric*, that is, the permittivity number ϵ_r . The capacitance of the capacitor is linearly dependent on this (Fig. 2.34d). The change in permittivity number can be linked to a change in length. In this case, an additional medium is inserted between the capacitor plates. Depending on the position, this results in an average permittivity number. The capacitance is linear to this. This is how *filling levels* are determined. If the liquids are not conductive, two electrodes can be inserted. The measured permittivity number is then dependent on the proportion of liquid and air in the capacitor gap.

Table 2.10 Permittivity numbers of different materials

Material	Permittivity number ϵ_r
Vacuum, air	1
Wood (depending on humidity)	2 to 7
Paraffin, petroleum, turpentine oil, transformer oil	2.2
Paper, polyethylene, polypropylene	2.3
Soft rubber, polystyrene	2.5
Celluloid	3
Polyester, Plexiglass	3.3
Quartz glass	3.7
Hard rubber, oiled paper, pressboard	4
Porcelain, hard paper, quartz sand	4.5
Glass, polyamide	5
Marble	8
Al_2O_3	12
Alcohol	26
Ta_2O_5	27
Water	81
Ceramics (LDC: low dielectric coefficient)	10 to 200
Ceramics (HDC: high dielectric coefficient)	10_3 to 10_4

With the principle of the *differential capacitor* (Fig. 2.34b), the influence of the distance on the capacitance can be *linearized*. It can also detect *accelerations*: Between the two capacitor plates, there is a movable electrode in the middle, which represents the seismic mass. During acceleration, the distances change by Δd ($d_0 + \Delta d$ or $d_0 - \Delta d$). This also changes the two capacitances C_1 and C_2 in opposite directions. This capacitance change is detected and measures the *acceleration* (Fig. 2.44).

Sensors based on the change in permittivity ϵ_r may determine the following:

- Type of material (e.g. paper, wood, metal, plastic, ceramics),
- Counting the number of objects of different materials,
- Determination of the moisture content of materials (e.g. of air: a polymer dielectric absorbs the water in the air and thus changes the permittivity number),
- Measurement of the layer thickness of materials (series connection of the capacitors according to Fig. 2.34d),
- Level measurement of solid, liquid and powdery substances. Depending on the level, the permittivity number changes (parallel connection of the capacitors according to Fig. 2.34d).

For the *level measurement*, it is important to distinguish between a *non-conductive* and *conductive* filling material. Figure 2.35 shows the scheme for the noncontact measurement

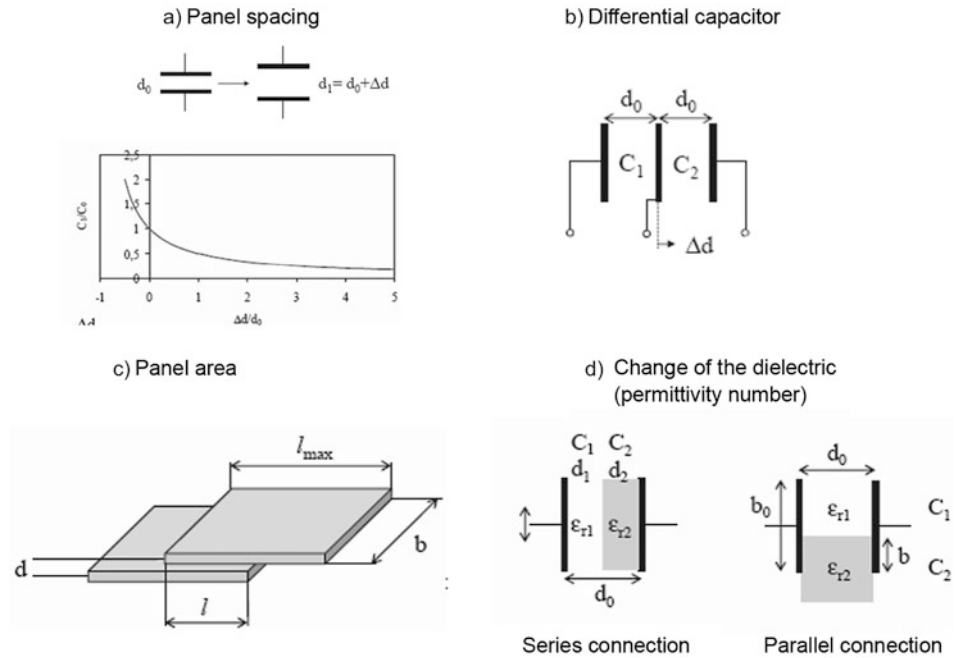


Fig. 2.34 Capacitance of a plate capacitor (a) change of the plate spacing; (b) circuit as a differential capacitor; (c) change of the area; (d) change of the dielectric

of nonconductive filling material. The electrodes consist of a measuring electrode (*active electrode*) and a second electrode to ground potential (*ground electrode*). The ground electrode is arranged in a circle around the active electrode. As shown in Fig. 2.35, part of the electric field (shown as dashed lines) is in the dielectric of the container wall and part is inside the container. When the filling material has reached both electrodes, the higher permittivity number ϵ_r becomes noticeable through the filling material. The capacitance increases and the previously set threshold value is exceeded. The sensor triggers a signal, which indicates that the level has been reached.

The prerequisite for this measurement is that the filling material has an ϵ_r value >2 and the threshold value of the capacitance for triggering the sensor signal is preset. The capacitive changes are usually very small (10^{-14} F to 10^{-13} F). For this reason, a large active area is often required, which is given from design M18 upward. Figure 2.36 shows a level sensor as it is typically used in the chemical and food industry as an immersion sensor.

If the container contains *electroconductive* liquids, the conditions are as in Fig. 2.37.

In this case, the *conductive filling material* represents the *second capacitor plate*. The container wall is the dielectric. When empty, the second capacitor plate is so far away from the measuring electrode that practically no basic capacitance is measured. In full condition, an ideal plate capacitor is created. The capacitance is considerably higher than when empty ($0.5 \cdot 10^{-12}$ F to $3 \cdot 10^{-12}$ F). As a result, wall thicknesses of up to 15 mm are also possible

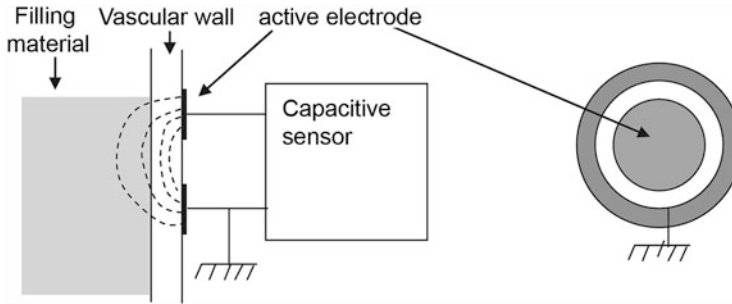


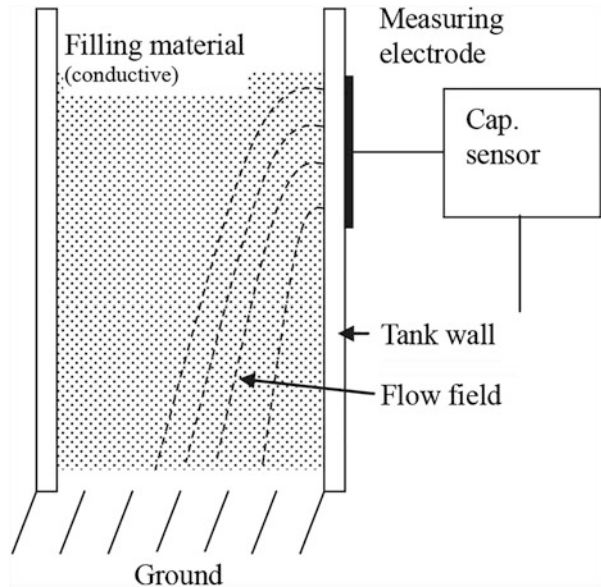
Fig. 2.35 Scheme for level measurement of a nonconductive filling material

Fig. 2.36 Level sensor (Factory photo: EasyTeach)



(for nonconductive materials only up to 5 mm). The big problem is the measurement of the level of electrically conductive materials if residues of the filling material remain on the inner wall of the container as a film or other adhesions during emptying of the filling material. In this case, capacitances are measured that do not correctly reflect the filling levels. In order to measure the filling levels correctly nevertheless, the *conductivity* is measured in addition to the capacitance. The conductivity is high when filled with the whole liquid. If there are only residues (e.g. on the outer walls), the conductance drops significantly. Through the combination of capacitance and conductivity measurement, real

Fig. 2.37 Scheme for level measurement of a conductive filling material



filling conditions can be easily distinguished from buildup on the container wall. The respective measurements are performed with oscillatory or externally controlled measuring circuits (Sect. 2.6.1.2).

2.6.1.2 Capacitance in the Alternating Current Circuit

If there is a capacitance C in the alternating current circuit with the frequency ω , then an imaginary reactance X_C results:

$$X_C = -j \frac{1}{\omega C}.$$

with $C = \epsilon_0 \epsilon_r \frac{A}{d}$ results in :

$$X_C = -j \frac{d}{\omega \epsilon_0 \epsilon_r A} = \text{constant} \cdot d.$$

This means that the reactance changes in proportion to the distance d .

The sensor consists of an RC oscillator, an electrode arrangement as a transducer, a comparator and an output stage (Fig. 2.38).

If metals or non-metallic materials approach the active zone (the area where the high-frequency sensor field is active), this causes a change in capacitance and the RC oscillator begins to *oscillate*. A comparator detects this and sends a signal to the output stage. If the sensor is used as a proximity switch, the result is a simple switch: at a long distance the

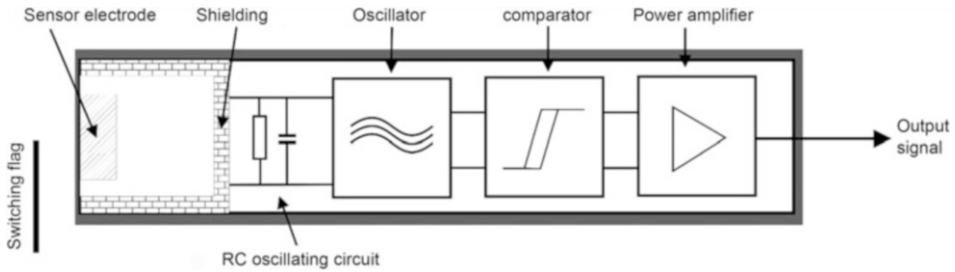


Fig. 2.38 Structure of a capacitive sensor in an alternating current circuit

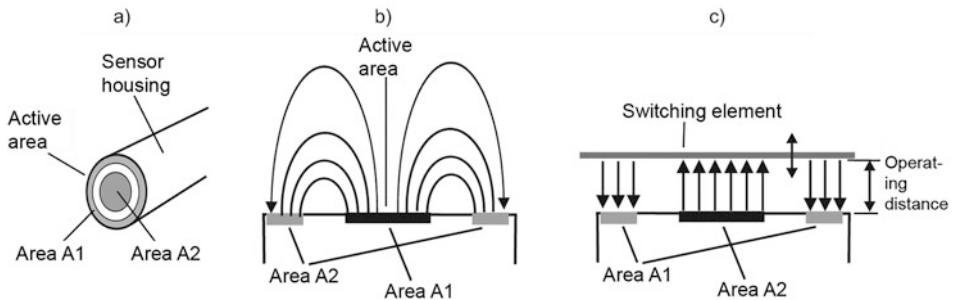


Fig. 2.39 Mode of operation of a capacitive sensor (a) mechanical construction; (b) open capacitor; (c) capacitor with switching lug: switching operation in the presence of an actuating element (Source: Hering, Gutekunst, Martin, Kempkes: Electrical engineering and electronics for mechanical engineers, eighth edition, Springer)

amplitude is low (no oscillation) and at a short distance the amplitude is high (oscillating state). The switching frequencies f are between 10 Hz and 100 Hz. Figure 2.39 shows the mechanical structure of a capacitive sensor element (Fig. 2.39 a). If there is no object (switching lug) near the capacitor, the capacitor is *open* (Fig. 2.39 b)). If an object (switching lug) comes close to it, the capacitance changes proportionally to ϵ_r and reverses proportionally to the distance to the active surface (Fig. 2.39 c)). The distance at which the RC oscillator oscillates, that is, the sensor switches, is called *rated switching distance* s_n .

Non-metallic materials always result in a smaller rated operating distance. This is called the *safe operating distance* s_a and is determined by empirically determined correction factors ($s_a = \text{correction factor} \cdot s_n$). Table 2.11 shows some values.

The safe switching distance of capacitive sensors depends on the following factors:

- *Diameter* of the sensor.
- *Material* of the approximated body. The material properties must be determined in advance. A potentiometer is used to set the corresponding sensitivity.
- *Mass* of the approximated body.
- *Installation type* (flush or non-flush).

Table 2.11 Correction factors for non-metallic switching elements

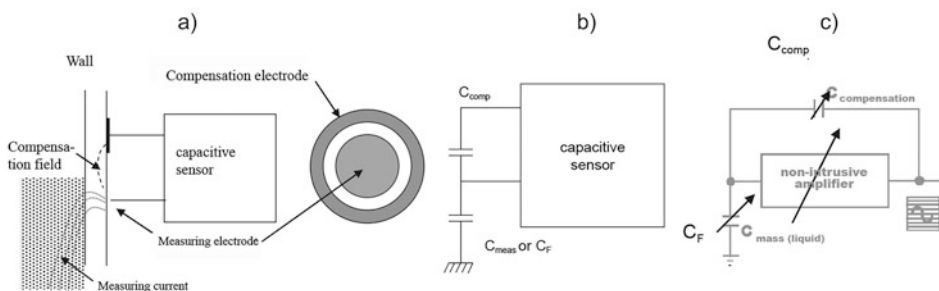
Material	Correction factor
Metal, water	1
Wood (depending on humidity)	0.2 to 0.7
PVC	0.6
Glass	0.5
Oil	0.1

With capacitive as well as inductive sensors, the switch-on point when approaching p_a is different from the switch-off point when removing p_e . This is called *switching hysteresis* H and is given as a percentage of the nominal switching distance s_n . It applies to the switching hysteresis H :

$$H = \frac{p_a - p_e}{s_n} \cdot 100 \text{ in}\%.$$

The problem mentioned in the previous section of distinguishing between real filling conditions of conductive materials and residues or adhesions on walls has been developed by so-called *smart level technologies*. The sensor makes it possible to detect the difference between the electrical resistance of the filling level and the electrical resistance of the buildup without contact. The buildup has a higher electrical resistance than the product. Figure 2.40 shows the arrangement for *Smart Level 15* technology (measurement of the specific conductivity up to 15 mS/cm).

In this circuit, two electrodes (measuring and compensation electrode) and two electric fields (active field and compensation field) must be considered (Fig. 2.40a)). The active field measures the capacitance C_F between the active electrode and ground, that is, the level. The electronically generated compensation field acts as a counter field (Fig. 2.40 b)). This compensates for example the non-conductive vessel wall. The circuit is shown in Fig. 2.40 c). The amplitude of the oscillations is influenced by the two capacitors C_{komp} and C_F in opposite directions. The output signal vibration (level empty) or zero signal (level

**Fig. 2.40** Arrangement for measuring the electrical resistance of the build-up and the medium

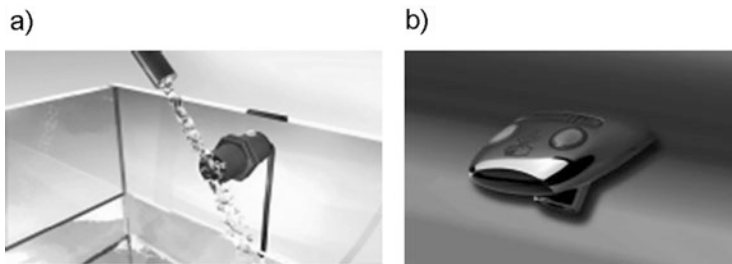


Fig. 2.41 (a) Level measurement independent of external disturbance influences; (b) SmartLevel sensor in a mixer tap in the toilets of the Airbus 380 (Factory photo: Balluff GmbH)

full) are evaluated and in this way, the filling is indicated exactly. Condensation effects, buildup and liquid films, as well as contamination, are faded out. The operating frequency is approximately 6 MHz and is thus approx. 6 times higher than with a capacitive standard measurement (1 MHz).

Figure 2.41 shows the independence of the level measurement from external interference (Fig. 2.41 a) and the SmartLevel sensor for temperature adjustment in the mixer tap in the washrooms of the Airbus A 380.

The *Smart-Level 50 technology* is a consistent further development of the Smart-Level 15 technology and extends the range of applications to more conductive media up to 50 mS/cm (therefore the name Smart-Level 50). This makes it possible for the first time to detect extracorporeal blood without contact.

This method works with *steep signal edges*. The measuring capacitance C_f is discharged by a steep voltage drop in a few ns. The steep voltage change dU/dt results in high, short-term pulse currents even in the smallest measuring capacitances. Even the smallest resistance increases R_f in the measuring path significantly reduce the pulse current. Conductance differences between compact product and adhesive film can therefore be distinguished even better. Figure 2.42 shows the block diagram for Smart Level 50 technology.

The high pulse currents in the measuring path cause corresponding voltage pulses at the shunt resistor R . These are raised by a fast amplifier and converted to a proportional DC voltage in a peak-to-centre rectification. If this voltage falls below a *preset threshold value* U_{ref} (sensitivity adjustment), the sensor emits a switching signal.

The repetition frequency of the pulses is low (about 500 kHz). This reduces interference emission (important for medical technology) while at the same time increasing immunity to interference (EMC). The steep edges contain the highest frequency components. As a result, residues of liquids with a conductance three times higher than that of the Smart-Level 15 technology can still be distinguished from the true full state. Figure 2.43 shows the capacitive sensors and their properties.

Of the many applications, Fig. 2.44 shows the capacitive measurement of accelerations (Sect. 5.8).

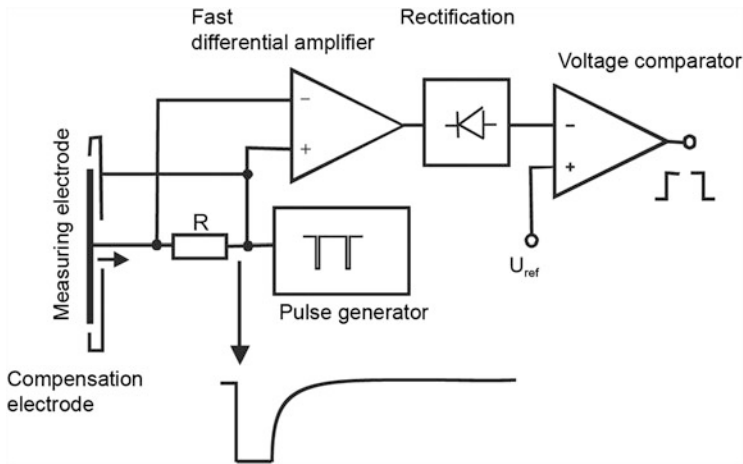


Fig. 2.42 Block diagram Smart-Level 50

Electrical data		general data	
Rated operating voltage	24 DC V	operating voltage display	no
Rated operating current	250 mA	functional display	yes
Ripple max. of Ue	<15% of Ue	short circuit protection	yes
Switching output	PNP	protection class IP	IP 67
Switching function	Normally open contact (NO)	protection class	II
connection	Connectors	recommended connector	BKS-S 19-3-05
idle current vaporized	15 mA	reverse polarity protected	yes
operating voltage	10... 35V	approval	CE
Rated insulation voltage	250 AC V		
company default	17 ms		
electrical design	DC, direct voltage	connection diagram	connector diagram
smoke category	DC 13		
switching frequency	50 Hz		
voltage drop static max.	2.5 V		
mechanical data			
diameter	M18 x1.0		
ambient temperature min.	-25 °C		
pollution degree	3		
material of the active area	PBT		
Material housing	PBT		
Mechanical installation conditions	non-flush		
ambient temperature range max.	80 °C		
ambient temperature range	-25... 80 °C		
repeat accuracy max.	10% of Sr		
workspace max.	6.5 mm		
workspace min.	0 mm		
number of leaders	3		
Rated operating distance	1...8 mm		

Fig. 2.43 Typical properties of capacitive sensors (source: Balluff GmbH)

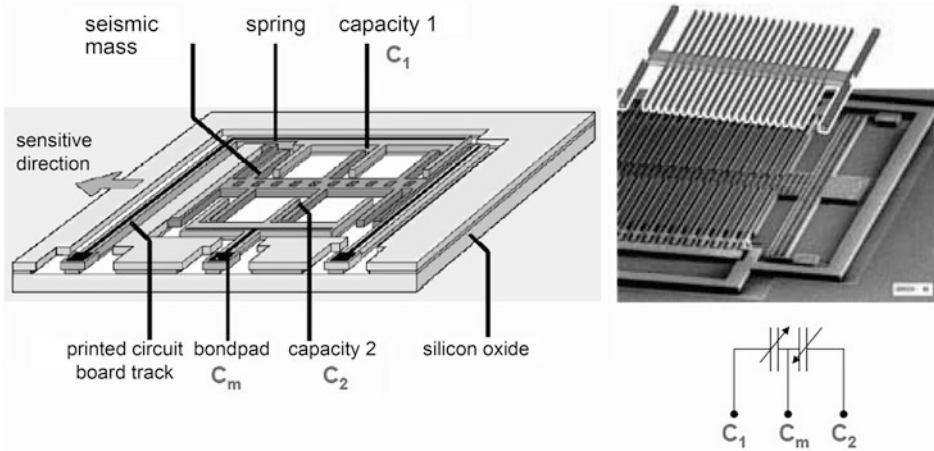


Fig. 2.44 Capacitive measurement of acceleration (Factory photo: BOSCH)

2.6.2 Advantages of Capacitive Sensor Technology

Capacitive sensors offer the following advantages:

- contactless,
- wear- and maintenance-free,
- non-reactive,
- robust and reliable,
- the bounce-free output signal,
- sweat-proof,
- high repeatability (2% to 5% of real operating distance),
- small temperature deviation ($\pm 0,02 \text{ mm}/^\circ\text{C}$),
- high-temperature stability ($5 \cdot 10^{-6} /\text{K}$),
- high operational safety due to contamination compensation,
- high linearity (up to 0.01% of full scale),
- highest resolution in the nanometer range (0.01 nm),
- high EMC resistance,
- long service life,
- low power consumption,
- adaptive and programmable and
- cost-effective due to simple construction.

However, capacitive sensors also have the following disadvantages:

- Impairment in industrial use due to dust and moisture (but can usually be compensated) and
- Small switching distances (15 mm to maximum 50 mm).

Table 2.12 Application areas of capacitive sensors

Areas	Measuring task
Mechanical engineering/ automation technology	<ul style="list-style-type: none"> • Detection of different materials (conductive and non-conductive materials) through glass or plastic walls • Detection of workpieces of different sizes • Recognition of stack heights (e.g. paper, CD) • Positioning and position control of workpieces or parts (e.g. chip assembly) • Detection of missing parts (presence check) • Distance measurement (e.g. of tools) • Path measurement • Monitoring of material thicknesses • Control of the width of punched parts • Measurement of deviations from roundness • Gap and distance measurement • Measurement of accelerations • Measurement of vibrations • Control of opening and closing (e.g. of press moulds) • Counting tasks for metals and non-metals
Food/pharmaceutical and chemical industry/ medical technology	<ul style="list-style-type: none"> • Filling levels of solid (e.g. pellets, plastic granules, PVC, glass, powder, graphite, bulk material) and liquid substances (hydraulic oils, water, acids, alkalis, disinfectants, blood, highly viscous media such as glues, pastes and adhesives) • Measurement of the vibrations of a filling machine • Pressure in containers
Semiconductor industry	<ul style="list-style-type: none"> • Control of acid levels during the processing of wafers and solar cells • Wafer presence check
Special applications	<ul style="list-style-type: none"> • PC touchpads • Portable media players • Mobile phones • Humidity

2.6.3 Applications of Capacitive Sensor Technology

The standard applications essentially relate to the following four areas:

1. Level control,
2. Stack height control,
3. Presence monitoring and object detection
4. Content review.

Table 2.12 provides an overview of some fields and applications.

Figure 2.45 shows capacitive sensors as used for an ultra-precise 6D nanopositioning system in scanning microscopy.

Fig. 2.45 Capacitive position sensors (Factory photo: Physik Instrumente (PI))



2.7 Gaussian Effect

2.7.1 Principle of Operation and Physical Description

The *Gauss effect* belongs to the *galvanomagnetic effects*. These occur when *current-carrying* platelets of conductors or semiconductors are in a *magnetic field*. This magnetic field must not be parallel to the direction of the current but must have vertical components. With the Gaussian effect, the electrical resistance increases with increasing magnetic flux density \mathbf{B} . This is due to the fact that a *Lorentz force* acts which deflects the electrons and they must therefore travel a greater distance through the material. Figure 2.46 a) shows that the electrons pass straight through the conductor without a magnetic field. If a magnetic field with the magnetic flux density \mathbf{B} acts, the path of the electrons runs in a zigzag pattern. The path of the electrons becomes longer and thus the resistance increases.

The increase in resistance is determined by the equations for the *Lorentz force* \mathbf{F}_L and *Ohm's law*.

The Lorentz force is the force acting on a charge q when it moves with a drift velocity \mathbf{v} through a conductor exposed to a magnetic flux \mathbf{B} . The following applies to the Lorentz force \mathbf{F}_L

$$\mathbf{F}_L = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

Here \mathbf{E} is the electric field strength.

Ohm's law is formulated in the generally valid way:

$$\mathbf{j} = \sigma \mathbf{E}.$$

Where \mathbf{j} is the electrical current density and σ is the electrical conductivity. For these, it is also possible to write for the microscopic range:

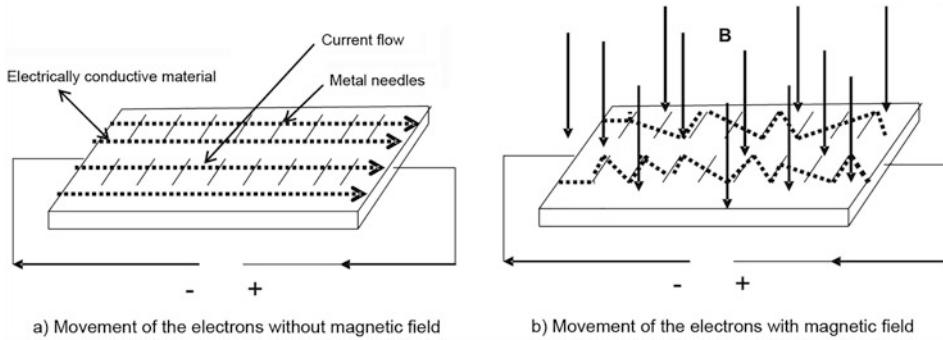


Fig. 2.46 Origin of the Gaussian effect (a) without magnetic field; (b) with the magnetic field

$\sigma = q n \mu$, where q describes the charge of the particles, n the particle density and μ the mobility of the particles. The charge density ρ is the product of the charge q and the particle density n so that $\rho = q n$. This results in σ for electrical conductivity:

$$\sigma = \rho \mu.$$

The electric current density \mathbf{j} can also be written:

$$\mathbf{j} = \rho \mathbf{v}.$$

Ohm's law then applies:

$$\rho \mathbf{v} = \rho \mu \mathbf{E}.$$

Thus the drift velocity \mathbf{v} :

$$\mathbf{v} = \mu \mathbf{E}.$$

The force \mathbf{F} acting on a charge q in an electric field \mathbf{E} is: $\mathbf{F} = q \mathbf{E}$ or $\mathbf{E} = \mathbf{F}/q$. This gives the particle velocity \mathbf{v} :

$$\mathbf{v} = \mu \mathbf{F}/q.$$

This results in \mathbf{F}_L for the Lorentz force in coordinate notation:

$$\mathbf{F}_L = q \left[\begin{pmatrix} E_x \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} v_x \\ v_y \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ 0 \\ B_z \end{pmatrix} \right] = q \begin{pmatrix} E + v_y B \\ -v_x B \\ 0 \end{pmatrix}.$$

If these results are inserted into the formula for the drift velocity \mathbf{v} , the drift velocity in x- and y-direction is obtained:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \mu \begin{pmatrix} E + v_y B \\ -v_x B \end{pmatrix}.$$

It is $v_x = \mu E + \mu v_y B$ and $v_y = -\mu v_x B$.

This results in the drift speed in x-direction:

$$v_x = \mu E - v_x \mu^2 B^2 \text{ oder}$$

$$v_x = \frac{\mu E}{1 + \mu^2 B^2}.$$

If the drift velocity decreases, the current density must decrease and, with the electric field \mathbf{E} remaining constant, the electrical conductivity σ must also decrease. However, since the electrical resistance R is reciprocal to the electrical conductivity σ ($R = 1/\sigma$), the electrical resistance R increases:

$$R(B) = R_0 (1 + \mu^2 B^2).$$

The resistance thus increases square to the B -field (Fig. 2.47).

2.7.2 Application of the Gaussian Effect

The Gaussian effect finds its applications on the one hand in the determination of *magnetic fields* and the *field plate*. It is one of the most common applications. The electrical resistance R depends on the magnetic flux \mathbf{B} , as shown above. Therefore the electrical resistance can be controlled by a magnetic field. An *MDR* (magnetic dependent resistor) is available. Figure 2.48 shows the structure and function of a field plate.

The field plate consists of a ceramic plate on which a thin semiconductor layer of highly conductive indium antimonide (InSb) is deposited in a meandering pattern. Needle-shaped areas of metallic nickel antimonide (NiSb) are embedded in this semiconductor. These needles are aligned parallel to the terminal edges of the field plate (Fig. 2.48a)).

If a voltage is applied to the field plate, a rectilinear current flows, which depends on the resistance. The metallic needles embedded parallel to the current direction have no influence. The resistance range is between 10 Ω and 1 k Ω . If a magnetic field with the

Fig. 2.47 Change in electrical resistance as a function of magnetic flux B

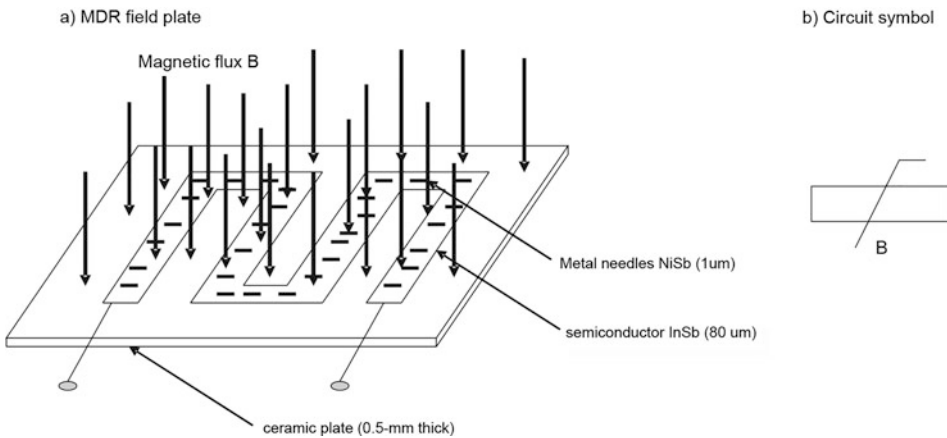
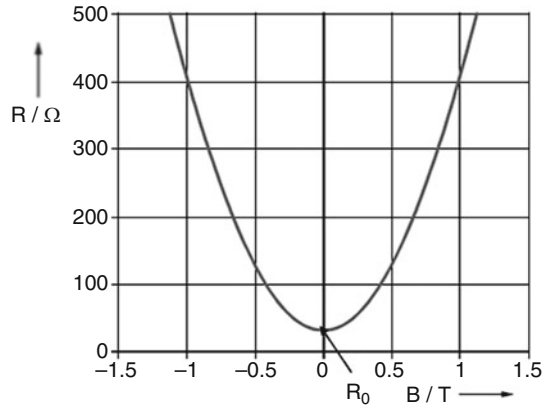


Fig. 2.48 Field plate (MDR) (a) Construction; (b) Circuit symbol

magnetic flux density B acts perpendicular to the current flow, the charge carriers (in this case the electrons of InSb) are deflected laterally due to the Lorentz force. The deflection angle is called the *Hall angle* and is about 80° for InSb at a magnetic flux density of 1 Tesla. The deflected charge carriers hit the metallic conducting needles. These act like a short circuit and the charge carriers are deflected in the direction of conduction. This causes the charge carriers to move through the material in a zigzag pattern (Fig. 2.46 b)). This path extension results in an increased resistance (increase up to twenty times).

Semiconductors with *high charge carrier mobility* are used as the base material. In addition to InSb, these are mainly InAs, Si and GaAs. The electrical resistance of the semiconductor materials is also temperature-dependent. The base materials mentioned have a *negative temperature coefficient*. Often the semiconductor materials are still doped with tellurium in order to adapt the change in resistance to the application. Table 2.13 lists some characteristic values for field plates.

Table 2.13 Limit and characteristic values of field plates

Size	Value
Output resistance R_o ($B = 0$)	10 Ω to 10 k Ω
Temperature coefficient	$\alpha \approx -0.004 \text{ K}^{-1}$
Maximum operating temperature	$T_{\max} \approx 95 \text{ }^\circ\text{C}$
Maximum permissible load	$P_{\max} \approx 0,5 \text{ W}$

The Gaussian effect can be used to build sensors that determine magnetic fields, currents, angles, positions and rotational speeds (as with Hall sensors, Sect. 2.8).

2.8 Hall Effect

2.8.1 Principle of Operation and Physical Description

The *Hall effect*, like the Gaussian effect, belongs to the *galvanomagnetic effects*. These occur when current-carrying platelets of conductors or semiconductors are in a magnetic field. This magnetic field must not be parallel to the direction of the current but must have vertical components. Figure 2.49 shows the origin of the Hall effect. A current I_x flows through a conductive plate of width b and thickness d in the x-direction. A magnetic field B_z acts perpendicular to this. A *Lorentz force* F_L then acts on each electron:

$$F_L = -qv_x B_z$$

(q : charge, v_x : velocity of the charge carriers in the x-direction).

The Lorentz force shifts the charge carriers in the y-direction. As Fig. 2.49 shows, in this case, there is an electron surplus (negative charges) at the left end and an electron shortage (positive charges) at the right end. Thus an electric opposite field is built up:

$$F_{\text{el}} = -qE_y.$$

(E_y : Electric field strength in the y-direction).

The Lorentz force counteracts the electrical force until equilibrium is reached. It holds:

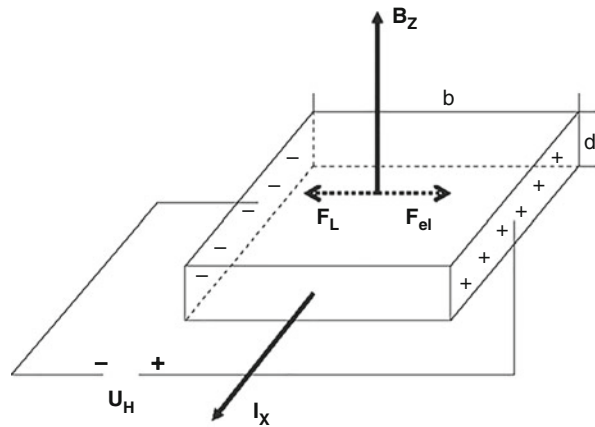
$$-qE_y = -q v_x B_z.$$

It's $E_y = U_y / b$. In y-direction, a voltage U_y is generated, which is called *Hall voltage* U_H . The following applies to this Hall voltage

$$U_H = v_x B_z b.$$

The velocity of the charge carriers v_x depends on the current density j_x as follows:

Fig. 2.49 Origin of the Hall effect



$$j_x = ne v_x$$

(n : number of electrons per volume; e : elementary charge: $e = 1.602 \cdot 10^{-19}$ As). It is then:

$$U_H = \frac{1}{ne} j_x B_z b.$$

The factor $1/(ne)$ is called Hall coefficient A_H . So it is:

$$A_H = \frac{1}{ne}.$$

This allows you to write:

$$U_H = A_H j_x B_z b.$$

The following applies to the current density: $j_x = I_x/(b \cdot d)$. Thus the above equation:

$$U_H = \frac{A_H B_z}{d} I_x = R_H I_x.$$

The expression $(A_H \cdot B_z)/d$ is the *Hall resistance* R_H . This is by no means the measured electrical resistance of a Hall element. It is rather the ratio of the transverse voltage (Hall voltage) U_H to the electric current I_x .

The Hall voltage U_H increases *linearly* with the magnetic flux B_z , it is *inversely proportional* to the signed *charge carrier density* n and is *independent of the specific electrical resistance*. Table 2.14 shows the values of the Hall coefficients of some materials.

Table 2.14 Hall coefficients A_H of some materials

Material	A_H in $10^{-11} \text{ m}^3/\text{C}$
Electron conductor	
Copper (Cu)	-5.5
Gold (Au)	-7.5
Silver (Ag)	-8.4
Sodium (Na)	-25
Cesium (Cs)	-28
Hole conductor	
Cadmium (Cd)	+6
Tin (Sn)	+14
Beryllium	+24.4
Semiconductor	
Bismuth (Bi)	$-5 \cdot 10^4$
Silicon (Si) and germanium (Ge)	10^8 to 10^{10}
Indium antimonide (InSb)	$-2.4 \cdot 10^7$
Indium Arsenide (InAs)	-10^7

This table shows that the *semiconductors* have a comparatively *high Hall coefficient*. This is due to the fact that the mobility of the charge carriers is higher. For this reason, the Hall effect is also very often realized with semiconductor components in CMOS technology. With these elements, temperature dependencies can also be compensated and the signals digitally processed and evaluated.

2.8.2 Application of the Hall Effect

If a current flows through a Hall sensor and a magnetic field acts perpendicularly on it, the Hall voltage U_H results. This is proportional to the product of the current I_x and the magnetic flux B_z . This means that the magnetic flux can be determined if the current is known. If the magnetic field is generated by a current flowing through a coil, the current strength in this coil can be determined potential-free. A high Hall voltage U_H occurs when the charge carrier density in a semiconductor is low and therefore the speed of the charge carriers is high. Hall sensors *can be integrated* into *chips*. Temperature compensation and signal amplification can take place there. These chips with integrated Hall sensors are used, among other things, to control electrical drives (e.g. the drive of floppy disk drives). Typical designs are used:

- Rectangular shape,
- cross shape and
- butterfly shape.

The following applications are important:

- Magnetic field measurement,
- measurement of electric current,
- contactless and non-contact signal transmitters,
- determination of the position of moving parts (position),
- detection of movements (speed and acceleration) and
- measurement of coating thicknesses.

Hall sensors are used in many different ways in the automotive industry. Examples are:

- Determination of the position of pedals (e.g. the brake pedal),
- control of the ignition timing,
- detection of the position of a belt buckle or a hand brake lever or
- measurement of the rotational speed of motors.

Hall sensors are relatively insensitive to liquids and dirt, but not to external magnetic fields. Compared to inductive sensors, Hall sensors have the following advantages:

- Square-wave signals are sent which can be directly evaluated electronically. They do not have to be processed first, as with inductive sensors.
- The signal voltage is independent of the rotational speed. Therefore very slow and even static conditions can be detected.

The following pictures show some applications (Figs. 2.50, 2.51, and 2.52).

Figure 2.53 shows various designs as they are typically used for status detection (e.g. open - closed in the case of window lifters), proximity signals, access controls or level measurements (Fig. 2.54).

For the selection and correct dimensioning of Hall sensor systems, *3D field simulations* can be carried out using the *Finite Element Method (FEM)*. These allow the precise calculation of electrical and magnetic quantities for the optimal design of appropriate Hall sensors. Figure 2.55 shows the results of a simulation for the design of Hall sensor systems for rotational speed measurement with a ferromagnetic toothed wheel.

2.9 Eddy Current Effect

2.9.1 Principle of Operation and Physical Description

If a coil is fed by an alternating electric current, an *alternating magnetic field* is created. If this field penetrates an *electrically conductive* material, a *current* is created according to the *law of induction* (Fig. 2.56). These closed, circular streamlines are called *eddy current* because the induction streamlines are closed in themselves like eddies. The eddy currents in turn generate a magnetic field which, according to *Lenz's rule*, counteracts the original magnetic field of the coil through which the current flows and inhibits the movement.



Fig. 2.50 Apparatus for measuring (a) magnetic fields and (b) currents (Factory photos: Cunz GmbH & Co)

a)



b)

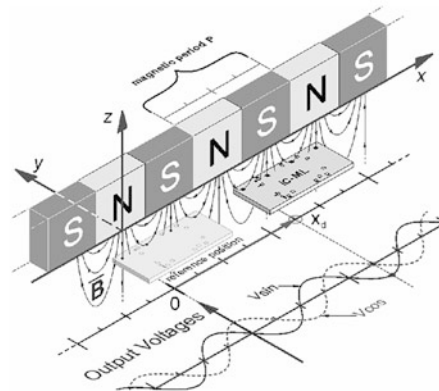


Fig. 2.51 Principle of the toothed wheel sensors (a) Position sensor of camshaft and crankshaft, wheel rotational speed sensor and milometer; (b) Generation of the Hall voltage signals (Factory photo: ICHaus)

The eddy current effect occurs not only when the magnetic field is an alternating field (as shown in Fig. 2.56 and common for sensor applications), but also when a *magnet moves* or the *conductive material is moved*.

2.9.2 Application of the Eddy Current Effect

The following applications are important:

- Geometric quantities (Chap. 3)

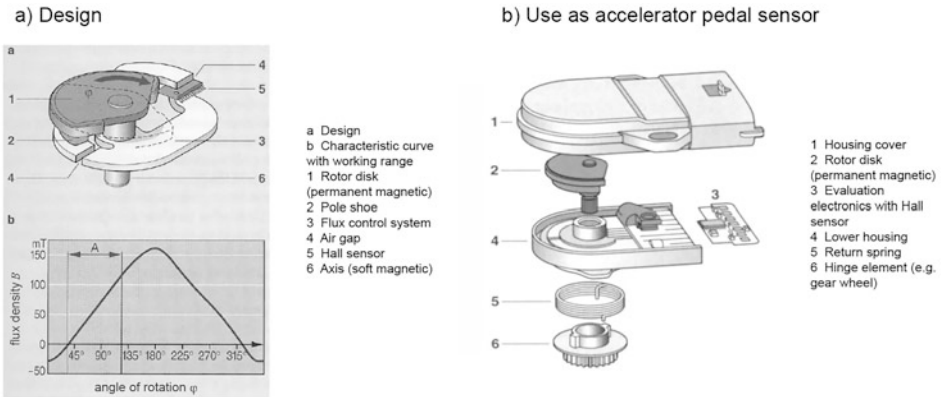


Fig. 2.52 Hall angle sensor (a) Basic design; (b) Use as accelerator pedal sensor (Factory photo: BOSCH)



Fig. 2.53 Typical designs of Hall sensors

- Position, path, displacement, dimension, expansion, deflection, backlash, lubrication gap, gaps (air gap, lubrication gap), pitch (e.g. profiles of toothed wheels).
- Dynamic, time-based quantities (Chap. 5)
 - Vibrations (e.g. bearing, shaft), speed (up to 400,000 rpm), collector concentricity, the position of moving parts.
- Miscellaneous
 - Impact, wear, layer thickness, conductivity,
 - corrosion,
 - crack test,

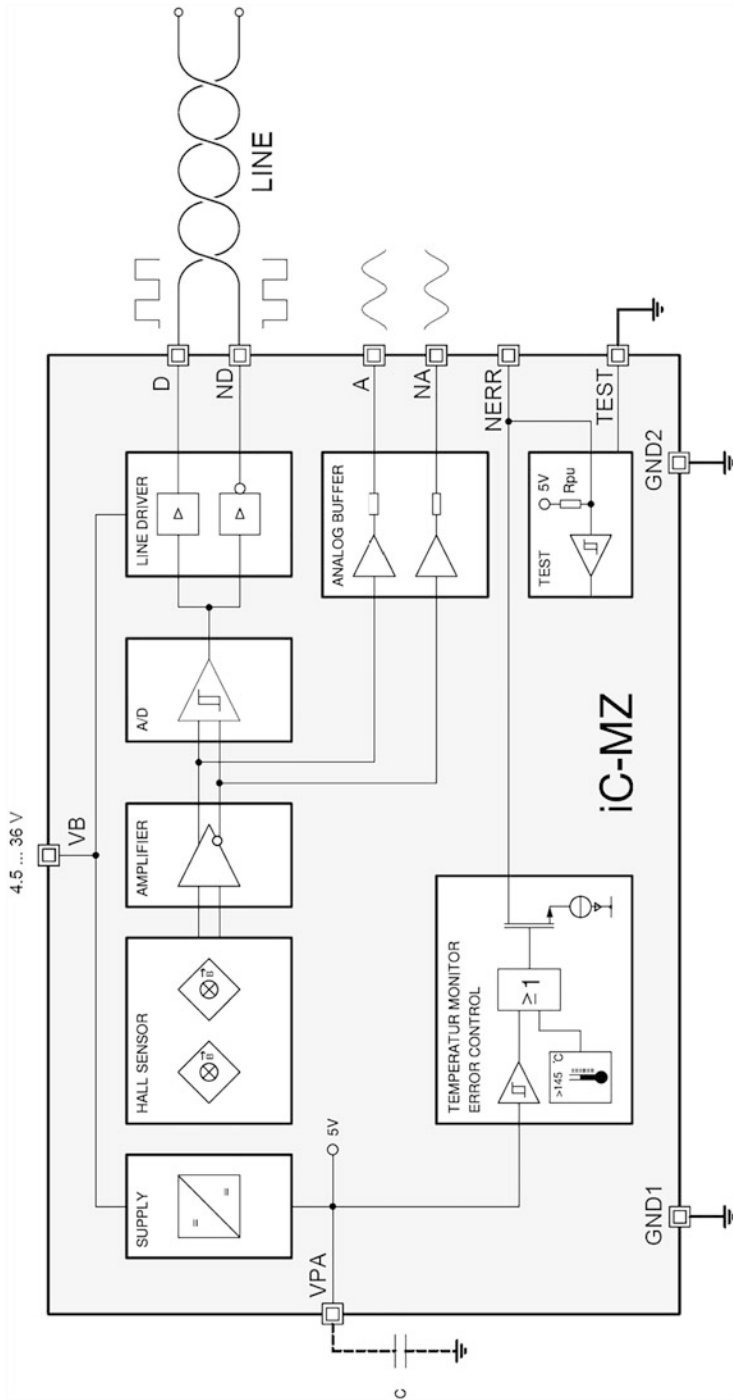


Fig. 2.54 Hall sensors as ICs: Chip assignment and block diagram (Factory photo: iC-MZ from IC-Haus)

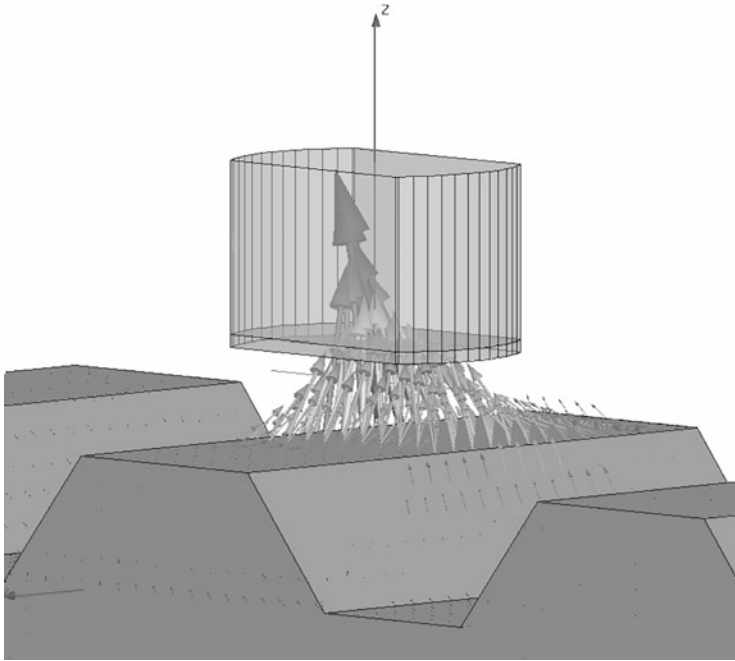
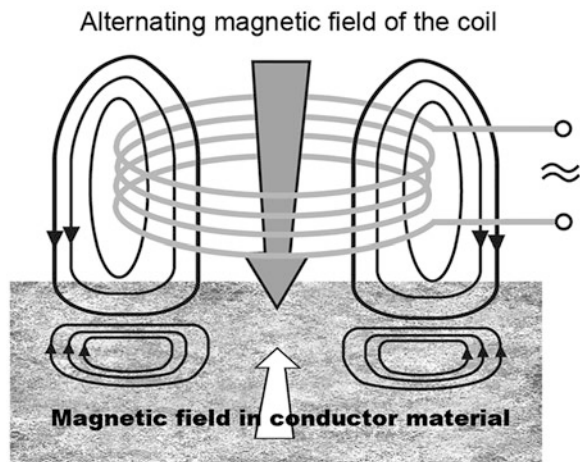


Fig. 2.55 3D simulation for the optimal design of Hall sensor systems (Source: SSG Semiconductor Systems GmbH)

Fig. 2.56 Origin of the eddy current effect



- sorting according to material properties and
- determination of the ferrite content.

Eddy current processes are successfully used in many industries, as examples:

- Mechanical engineering,
- automotive industry,
- semi-finished metal products industry,
- aircraft industry,
- chemical industry, boiler construction and power plant industry,
- foundry and forging industry and
- maintenance industry.

Eddy current sensors have the following advantages:

- Measurement on static and rotating objects;
- non-contact and wear-free measurement, often used for online quality assurance;
- insensitive to liquids and dirt, that is, suitable for harsh industrial environments
- unaffected by non-metallic media;
- fast motion changes (up to 35 kHz) and
- high resolutions down to 100 nm.

The following pictures show some applications (Fig. 2.57).

With the arrangement according to Fig. 2.58, relative movements between shaft and sensor can be measured. If two sensors are arranged at a 90° offset, the out-of-roundness and radial runout can be measured.

Figure 2.58 b) shows the possibilities of measuring gear teeth or splines or providing information about failed teeth or wear.

Eddy current sensors can be used to monitor the condition of continuously running machines. This is important to detect wear and tear and to determine possible maintenance or repair times.

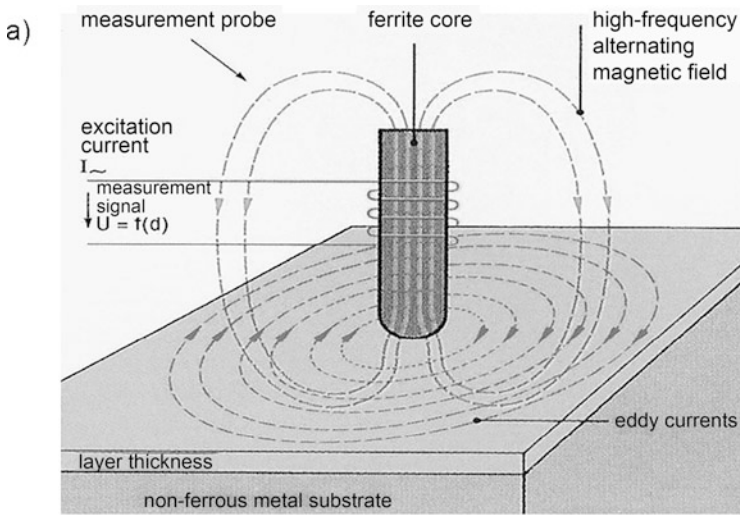
A high-frequency current flows through a measuring coil. This creates eddy currents on the segment discs, which are mounted on a shaft. A torque causes the overlapping of the disks to change, resulting in a change in the impedance of the coil, which is a measure of the torque (Fig. 2.59). The impedance Z is calculated using the following formula:

$$Z = \frac{\hat{U}}{\hat{I}} = \frac{U_{\text{eff}}}{I_{\text{eff}}} = \sqrt{R^2 + X_L^2}.$$

Here, the \hat{U} and \hat{I} amplitudes of the alternating voltage $u(t)$ and the alternating current $i(t)$, U_{eff} and I_{eff} are the effective values, R is the ohmic part of the alternating current resistance and X_L is the *reactance*, that is, the reactive part of the alternating current resistance of the coil.

The torque measuring ranges are from 1 Nm to 1 kNm with an accuracy of 0,5% (Sect. 4.5).

An excitation coil and a receiver coil are mounted perpendicular to its axis of rotation on a cylindrical shaft made of a highly conductive, non-magnetic material (usually copper or aluminium). These coils are offset by 90° to each other (Fig. 2.60 a)). A constant voltage (5 V to 100 V) with constant frequency (50 Hz to 500 Hz) is applied to the excitation coil.



b)

type of construction	designation	measuring range
	ETA3.3H	0-1,200 μm
	EAW3.3	0-1,200 μm
	EAI3.3-150	0-800 μm
	EA9	0-3.5 mm
	EA30	0-20 mm
	ETD3.3	0-800 μm

Fig. 2.57 Coating thickness measurement (a) principle; (b) probe designs (Source: Helmut Fischer)

This creates magnetic fields that induce eddy currents on the surface of the rotor. A rotary motion produces a transformer coupling between the excitation and receiver coils. The sinusoidal measuring voltage at the receiver coil is proportional to the rotational speed.

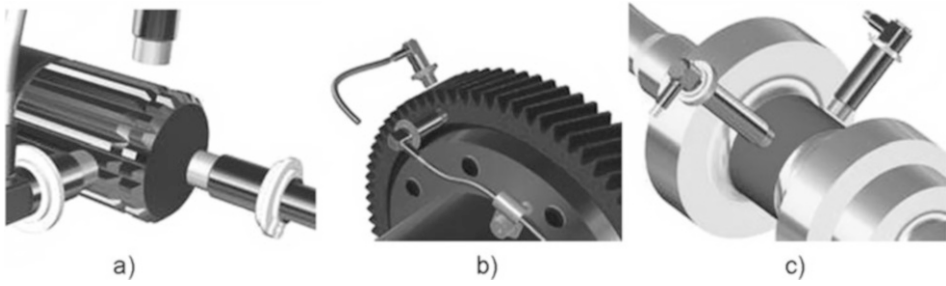


Fig. 2.58 (a) Testing rotating shafts for imbalance, out-of-roundness, radial or axial runout and vibration (source: Waycon); (b) Toothed wheel inspection (Source: Waycon); (c) Machine monitoring (Source: Waycon)

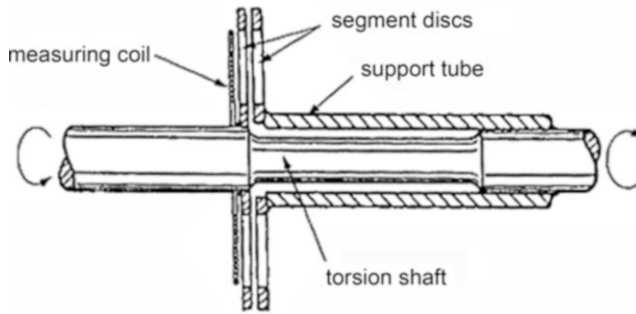


Fig. 2.59 Torque meter (Source: Tschan)

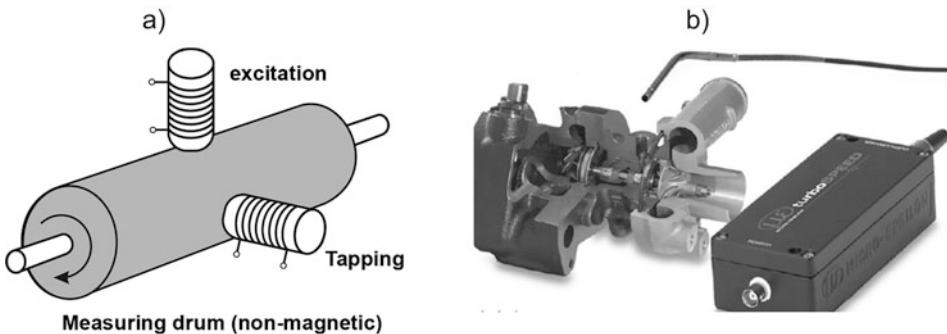


Fig. 2.60 Principles of tachometers (a) Measuring voltage proportional to speed (Source: Tränkler); (b) Counting the blade pulses (Source: Mikro-Epsilon)

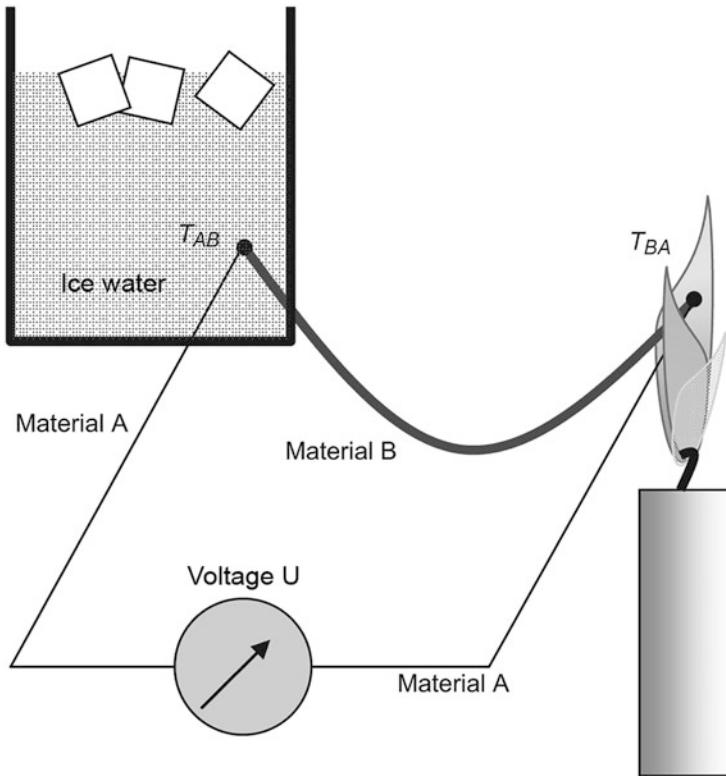


Fig. 2.61 Using a thermocouple to measure temperature

In the second case (Fig. 2.60 b)), a high-frequency alternating current flows through a coil encapsulated in the sensor housing. When a turbocharger blade approaches, an electric pulse is generated due to the change in the eddy current. A controller counts these impulses and calculates the speed on the basis of the number of blades.

2.10 Thermoelectric Effect

If different conductors are electrically connected at both ends (e.g. by welding) and these ends are exposed to different temperatures (T_{AB} , T_{BA}) (Fig. 2.61), then a *thermoelectric voltage* U_{th} can be measured which is proportional to the temperature difference $\Delta T = T_{BA} - T_{AB}$. This effect is called the *Seebeck effect*. It has the following cause in metals: The thermal energy of a conductor is stored to a certain extent in the oscillation energy of the atoms and another extent in the kinetic energy of the free electron gas. The *kinetic energy* of the electron gas leads to *thermal diffusion* of the electrons in the conductor. If there is a temperature gradient, the electron density in equilibrium is lower

on the warm side than on the cold side. The resulting *potential difference* is different for different conductors and gives the *thermoelectric voltage* of a contact pair. Relevant to the function of thermocouples is therefore the fact that the two electrically connected conductors have a *different thermo-diffusion*. Classically, the theoretical *Seebeck coefficient* of a S'_A (material A):

$$S'_A = -\frac{1}{3 \cdot e} \cdot \frac{d}{dT} \cdot \left\langle \frac{1}{2}mv^2 \right\rangle \quad (2.1)$$

$\left\langle \frac{1}{2}mv^2 \right\rangle$: Average kinetic energy of the electrons

S'_A : Theoretical Seebeck coefficient of material A

e: Elementary charge ($e = 1.6021 \cdot 10^{-19}$ As)

T: Temperature.

The Seebeck coefficient requires two differently electrically conductive materials. *Platinum* is taken as a reference material.

Electrons behave as *Fermi particles*, but not in a classical way. Therefore they have a large part of temperature-independent kinetic energy, while some electrons with kinetic energies close to the Fermi energy have temperature-dependent thermal energy. In the simplest case, the Seebeck coefficient is valid for an energy-independent free path length:

$$S'_A = -\frac{\pi^2 k_B}{2e} \cdot \frac{k_B T}{E_F} \quad (2.2)$$

k_B : Boltzmann constant ($k_B = 1.38 \cdot 10^{-23}$ J/K)

E_F : Fermi energy.

It becomes clear that the Seebeck coefficient is material-dependent.

If the two contact points in the circuit have different temperatures, a *diffusion current* is created between the metals for a short time, which builds up an oppositely acting *thermoelectric voltage* U_{th} . Figure 2.61 shows the test setup schematically. The following applies to the thermoelectric voltage

$$U_{th} = \int_{T_{AB}}^{T_{BA}} S_{AB}(T) dT. \quad (2.3)$$

It's in:

T_{AB} : Lower temperature

T_{BA} : Higher temperature ($T_{AB} < T_{BA}$)

$S_{AB}(T)$: Temperature-dependent Seebeck coefficient between material A and material B.

The temperature-dependent proportionality constant $S_{AB}(T)$ is called *the Seebeck coefficient, differential thermal force* or *thermoelectric voltage* and depends on the two materials. It is positive if the thermo current (conventional current direction) flows from conductor A to conductor B at the colder contact point in the circuit. It is determined experimentally from the change in the integral thermoelectric voltage with temperature:

$$S_{AB} = \frac{dU(T_{AB})}{dT_{AB}}. \quad (2.4)$$

Here dT corresponds to infinitesimal small heating of the contact point AB at a constant temperature of the other contact points in the circuit. The differential thermoelectric voltage is composed of two virtual individual voltages, each of which results from the amount of *thermal diffusion* in the materials:

$$S_{AB}(T) = S_A(T) - S_B(T). \quad (2.5)$$

In p-semiconductors it is positive and in n-semiconductors it is negative, and the lower the carrier density in the material, the greater the magnitude. By definition, the differential Seebeck coefficient for platinum is zero. Therefore, the thermoelectric voltage of the material to be described can be determined by means of two contact points between a platinum wire and a wire to be described. The following list shows examples of the thermoelectric voltages of different materials in relation to platinum at contact points of 0 °C and 100 °C:

- Platinum: 0.0 mV
- Constantan: -3.2 mV
- Nickel: -1.9 mV
- Tungsten: 0.7 mV
- Copper: 0.7 mV
- Iron: 1.9 mV
- Nickel chrome: 2.2 mV

Two thermocouples made of copper constantan (100 Cu-45Ni55Cu) thus supply a voltage of $0.7 \text{ mV} + 3.2 \text{ mV} = 3.9 \text{ mV}$ when one contact is at 0 °C and the other contact is at 100 °C. Thermoelectric voltages for certain thermocouples are standardized according to DIN EN 60584. Table 2.15 shows the areas of application and properties of standard alloys for thermocouples.

The differential thermoelectric voltage shows a *linear dependence* at low temperatures and has the value zero at absolute zero (Fig. 2.62). At higher temperatures, the differential thermoelectric voltage decreases more and more. To determine the thermoelectric force according to Eq. (2.3), it must then be integrated over the desired temperature range.

Table 2.15 Standard alloys for thermocouples (source: DIN EN 60584–1)

Type	Materials	Temperature range °C for continuous operation	Temperature range °C for short time load	Remarks
K	Nickel-Chromium / Nickel-Aluminium “Chromel / Alumel	0 to +1100	–180 to +1300	Most common sensor. Well suited for oxidizing atmosphere.
J	Iron / copper-nickel “Iron / Constantan”	0 to +700	–180 to +800	Well suited for dry and reducing atmosphere.
N	Nickel-14.2% Chromium-1.4% Silicon / Nickel –4.4% Silicon-0.1% Magnesium “Nicrosil / Nisil	0 to +1100	–270 to +1300	Relatively new. High stability.
R	Platinum-13% Rhodium / Platinum	0 to +1600	–50 to +1700	High temperatures. Used in the ceramic cladding tube.
S	Platinum-10% Rhodium / Platinum	0 to +1600	–50 to +1750	High temperatures. Used in the ceramic cladding tube.
B	Platinum-30% Rhodium / Platinum-6% Rhodium	+200 to +1700	0 to +1820	Very high temperatures. Always used in high purity ceramic cladding tube.
T	Copper / Copper-Nickel “Copper / Constantan”	–185 to +300	250 to +400	Sensor for low to cryogenic temperatures. Stability in the presence of water.
E	Nickel-chromium / copper-nickel („Chromel / Constantan“)	0 to +800	–40 to +900	High signal.

On the one hand, the Seebeck effect is suitable for measuring *temperature differences* between two points, or for measuring *temperatures* if a known reference temperature is available. On the other hand, it is also suitable for measuring *material properties* if a temperature difference is given. For example, the sign of the thermoelectric voltage can be used to quickly determine whether a semiconductor has electron conduction or hole conduction by simple means.

One of the few examples where the Seebeck effect is not used to measure temperature but to measure a material property is as follows: A method for determining airborne contamination is based on electrically conductive, chemically sensitive layers. The

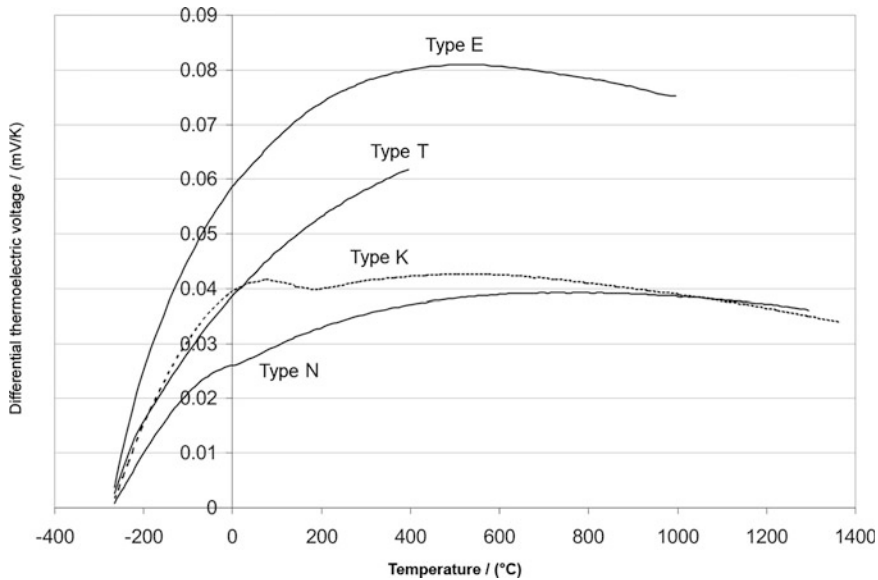


Fig. 2.62 Differential thermoelectric voltage as a function of temperature for different thermocouples

chemisorption (adhesion of molecules to a surface due to a (partial) charge exchange between molecule and surface) of the impurities influences the charge carrier concentration in the material. This modifies the Fermi energy and thus the differential thermoelectric voltage of the material (Eq. 2.3). The *concentration of the impurities* can therefore be determined by measuring them. A known temperature difference is given to a thermocouple with at least one chemically sensitive conductor and the thermoelectric voltage is measured. For a given fixed temperature difference, this depends on the concentration of the gas molecules to be detected. In practical applications, thermocouples are usually relatively low impedance. In general, very low voltages are obtained. Therefore, amplifiers with low offset voltages are used for the pre-amplification of such voltages. *Chopper* or *chopper-stabilized* amplifiers, which are offered as integrated circuits, are well suited.

One way to increase the output voltage of the sensor and thus to measure temperature differences more sensitively is to connect several thermocouples in series to form a *thermopile*. Figure 2.63 shows a sensor that operates as a thermopile. This measures the temperature difference on the surface of the component between the center and the periphery. The individual thermocouples (almost 100) are visible in the periphery.

The sensor for *heat radiation* as shown in Fig. 2.63 has a thin membrane in the middle of the chip, which is coated with an infrared absorber. This heats up by heat radiation. A chain of thermocouples connected in series (thermopile) measures the temperature gradient between the membrane in the middle and the periphery of the chip and thus determines

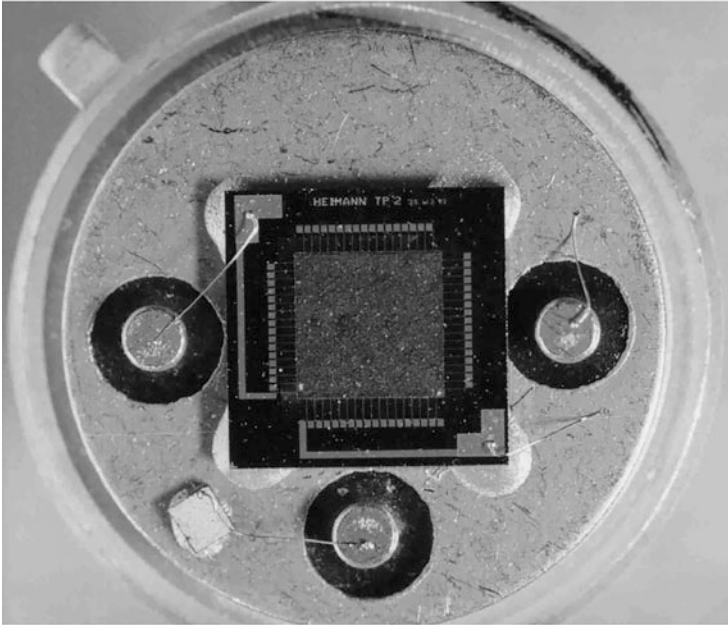


Fig. 2.63 Application of the thermoelectric effect in a heat radiation sensor (Factory photo: PerkinElmer Optoelectronics)

the received thermal radiation. The peripheral thermocouples are visible in the picture. The central elements are hidden by the absorber.

2.11 Thermoresistance Effect

2.11.1 Principle of Operation and Physical Description

The electrical resistance of a material depends on the composition of the material, on the homogeneity of the crystalline state and above all on the temperature. If a voltage is applied to *isotropic* materials, electric field strength is created which accelerates the electrons. The scattering of the electrons at the vibrating atoms, the impurities in the material and at their grain boundaries leads to a resistance that limits the current. For *metals* and *metal alloys*, the following non-linear relationship applies between the electrical resistance R and the temperature T

$$R(T) = R_0 \left(1 + a(T - T_0) + b(T - T_0)^2 \right)$$

with R_0 Resistance at the reference temperature T_0 (usually 20 °C, but also 0 °C)

Table 2.16 Temperature dependence of the temperature coefficient a for Pt 100

Temperature in °C	Resistance R in Ω	Temperature coefficient a in $10^{-3}/K$
-40	84.21	3.96
-20	93.13	3.95
0	100.00	3.90,802
20	107.80	3.89
40	115.54	3.86
100	138.50	3.80

a , b : Material coefficients.

If $b = 0$, the *gradient of the resistance characteristic* describes the *temperature coefficient* a . The following applies: $dR(T)/dT = a R_0$, which results in a :

$$a = R_0(dR(T)/dT).$$

The temperature coefficient a is strongly dependent on the type of metal or metal alloy ($a = 3.89 \cdot 10^{-3} K^{-1}$ for platinum and $a = 0.02 \cdot 10^{-3} K^{-1}$ to $0.05 \cdot 10^{-3} K^{-1}$ for chromium nickel). However, the temperature coefficient a also depends on the temperature T , as Table 2.16 shows for Pt 100.

The coefficient b only plays a role at temperatures above 100 °C and can be taken from the datasheets for the sensors.

Platinum (Pt) plays a special role in temperature measurement with the thermoresistance effect (Sect. 6.1). Pt is very *resistant* to chemically aggressive substances, has a *high melting point* ($T = 1772$ °C) and a *very high specific electrical resistance* ($\rho = 9.81 \cdot 10^{-6} \Omega\text{cm}$). The number behind the designation Pt indicates the value of the resistance in Ω at $T = 0$ °C. So means:

Pt 100: resistance of 100 Ω at 0 °C or

Pt 1.000: resistance of 1.000 Ω at 0 °C.

Common Pt resistors are: Pt 100, Pt 200, Pt 500, Pt 1000 and Pt 10,000.

The Pt resistors are manufactured either as wires or as thin layers. Figure 2.64 middle shows the structure of the resistor track. The right component measures 2.1 mm \times 1.3 mm (Type 0805).

Resistance thermometers can be used in different temperature ranges, as Table 2.17 shows.

2.11.1.1 Thermal Response Times

Thermal response times play an important role in the use of thermal resistance sensors in thin-film technology. This is the time that elapses until a platinum sensor has reacted to a step-like temperature change with a change in resistance that corresponds to a certain percentage of the temperature change. DIN EN 60751 recommends changing the times for

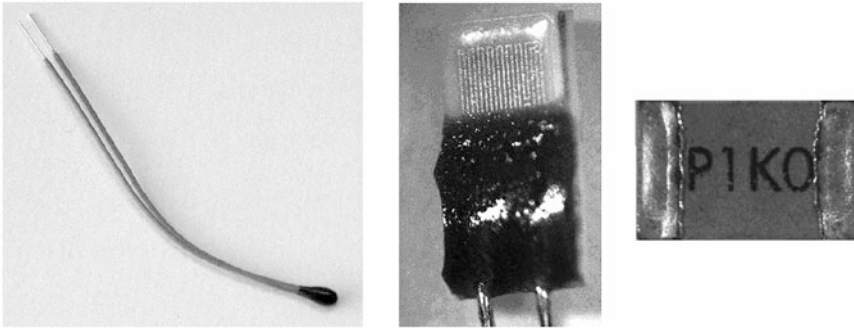


Fig. 2.64 Designs of common Pt resistors

Table 2.17 Preferred temperature application ranges of resistance thermometers

Material	Temperature application range
Platinum (Pt)	−220 °C to +1000 °C
Nickel (Ni)	−60 °C to +180 °C
Copper (Cu)	−50 °C to +150 °C

Table 2.18 Measuring currents for Pt sensors (recommendation: Heraeus)

Measuring range in Ω	Range of measuring currents
100 Ω	0.3 mA to max. 1.0 mA
500 Ω	0.1 mA to max. 0.7 mA
1000 Ω	0.1 mA to max. 0.3 mA
2000 Ω	0.1 mA to max. 0.3 mA
10,000 Ω	0.1 mA to max. 0.25 mA

a 50% ($t_{0,5}$) or 90% ($t_{0,9}$) change for water and air flows from 0.4 m/s to 2.0 m/s. The values for other media can be converted using the VDI/VDE manual 3522. Exact values for the thermal response times can be found in the datasheets of the sensor manufacturers.

2.11.1.2 Self-heating and Measuring Current

The current flowing through heats the platinum thin-film sensor. The measurement error of the temperature ΔT results from $\Delta T = P \cdot S$. Where P is the power loss ($P = I^2 R$) and S is the *self-heating coefficient* in K/mW. The values for S can be found in the datasheets of the sensor manufacturers. In addition to the thermal contact between the sensor and the surrounding medium, the measurement error depends very much on the current flow through the sensor. Depending on the measuring range, the manufacturers recommend the following current values (Table 2.18):

2.11.2 Advantages of Sensor Technology with the Thermoresistance Effect

The Pt resistors are used in many areas of technology because they

- are very precise,
- have high stability in thermal, mechanical and climatic terms (typically the deviation is 0,04% after 5 years of operation at 200 °C),
- are vibration and shock resistant (vibrations: 10 Hz to 2 kHz and accelerating shocks up to 100 g),
- have a high voltage resistance (< 1000 V at 20 °C and > 25 V at 500 °C),
- standardized characteristic curves,
- are resistant to environmental influences (climate, humidity, aggressive chemicals)
- have high long-term stability (several thousand measuring cycles) and
- can process their signals quickly and without errors.

2.11.3 Fields of Application

For better identification of different sensor families and their areas of application, *color codes* are often used for Pt thin-film sensors, as shown in Table 2.19 for an example.

Figure 2.65 shows an example of sensor coding.

Figure 2.66 shows the diverse use of Pt thin-film sensors and some important areas of application.

Table 2.19 Colour coding (Source: Heraeus Sensor Technology GmbH)

Area	Temperature range in °C	Colour
Fixation	Cryo (C) (from –200 °C)	Light blue
	Low (L) (up to 400 °C)	Light blue
	Medium (M) (up to 600 °C)	Blue
	High (H) (from 600 °C)	White
	Resistance value in Ω	
Meander cover	Pt 100	Transparent
	Pt 500	Pink
	Pt 1000	Blue
	Pt 10,000	Opaque

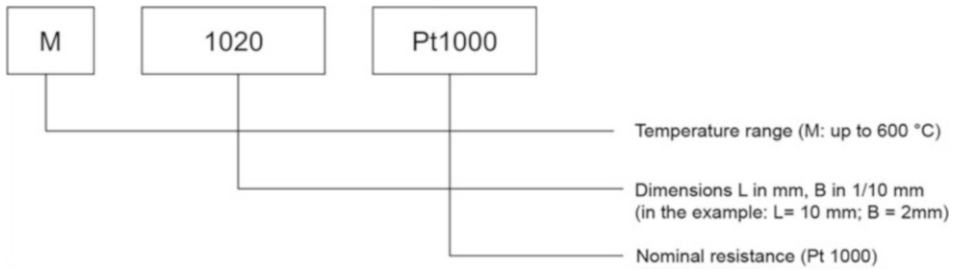


Fig. 2.65 Coding of sensors (source: Heraeus Sensor Technology GmbH)

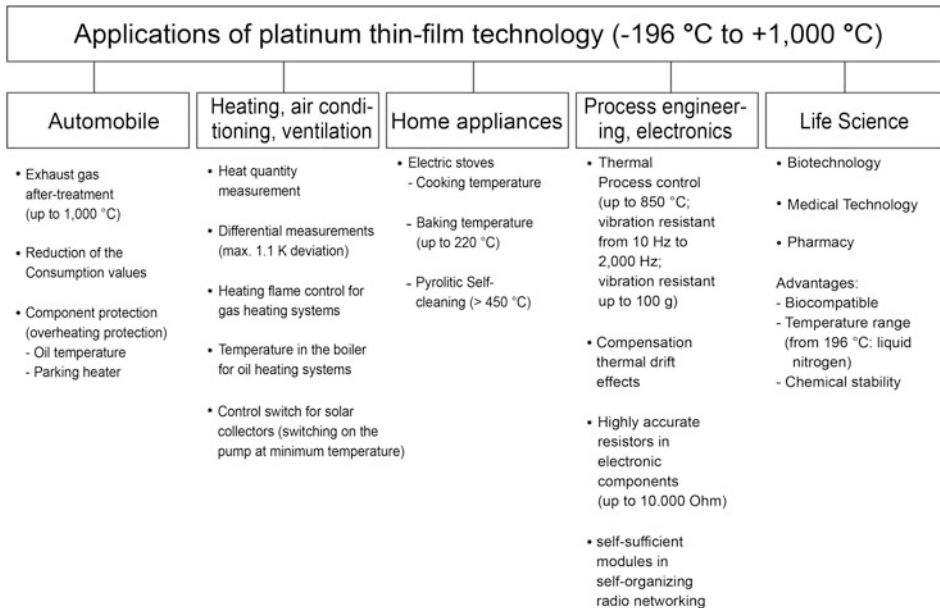


Fig. 2.66 Overview of the fields of application of Pt thin-film sensors

2.12 Temperature Effects in Semiconductors

2.12.1 Principle of Operation and Physical Description

In addition to the metallic sensors for temperature, there are also sensors for temperature measurement made of *semiconducting, polycrystalline ceramics*. These sensors are called *thermistors* (Sect. 6.1). The effects are determined by the electrical and magnetic properties of the materials. The *Curie temperature* plays an important role here. *Below* this temperature, the *magnetic properties* of the materials are decisive. If the Curie temperature is

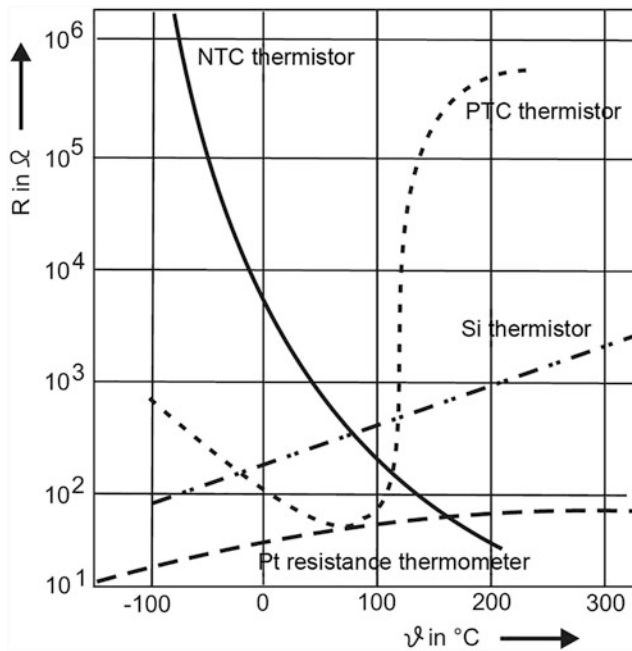


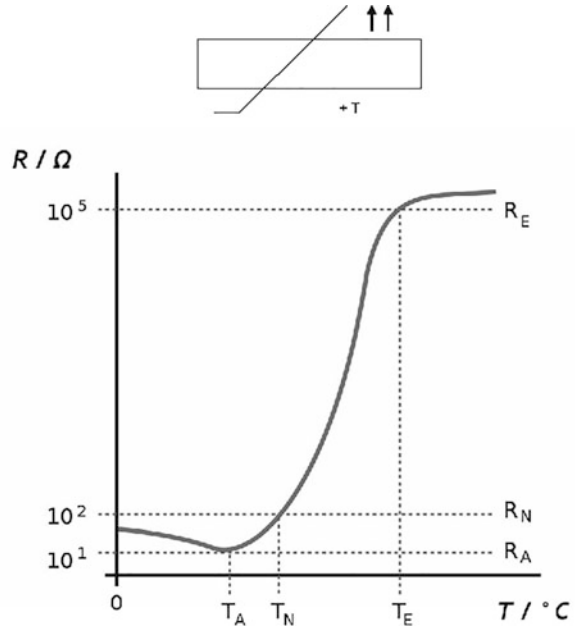
Fig. 2.67 Resistance-temperature characteristics of various resistors (Source: Hesse, Schnell: Sensors for process and factory automation, Vieweg+Teubner, 2009)

exceeded, the magnetic effects hardly play any role. If, as in the case of metals (Sect. 2.11), the resistance in these materials increases as the temperature rises, these are materials with a *positive temperature coefficient*. They are called *PTC thermistors* or *PTC resistors* (PTC: *Positive Temperature Coefficient*). Materials in which the resistance decreases with increasing temperature have a *negative temperature coefficient*. These materials are called *thermistors* or *NTC resistors* (NTC: *Negative Temperature Coefficient*). Figure 2.67 shows the characteristic curves of the various resistors. It can be seen that Pt and Si have approximately linear characteristics with a positive temperature coefficient. The *semiconducting ceramic resistors* are generally much more temperature-dependent, that is, the temperature coefficient A is *much higher* (up to 50 times higher than for metals). The PTC resistor exhibits an *S-shaped* behavior in a certain temperature range. In the area of the inflection point, the characteristic curve can therefore be *linearized* with good approximation. With the NTC resistor, the resistance drops exponentially as the temperature rises.

2.12.2 PTC Thermistors (PTC Resistors)

PTC resistors consist of *titanate ceramic*. The most common materials are barium titanate (BaTiO_3) or strontium titanate (SrTiO_3). To produce them, barium carbonate and titanium

Fig. 2.68 Circuit symbol (top) and the characteristic curve of PTC resistors to DIN 44080 (bottom)



(IV) oxide are ground, mixed and sintered at high temperatures (1000 °C to 1400 °C) together with other materials that determine the desired electrical and thermal properties. Figure 2.68 shows the circuit symbol and the typical characteristic curve. Further properties are described in DIN 44080. The circuit symbol shows the resistance as a rectangle. The diagonal line means that the resistance is variable and the straight line in the continuation indicates that the resistance curve is not linear. The resistance is, as the designation “+T” expresses, positively dependent on the temperature. In the upper right corner, you can see two arrows pointing in the same direction upwards. They indicate that with rising temperature the resistance also increases.

2.12.2.1 Physical Context

The characteristic curve has two sections. At low temperatures *below* the *Curie temperature* (the magnetic properties are decisive) the charge carrier density increases as the temperature rises, that is, the *resistance decreases* (as with NTC resistors). *Above* the Curie temperature (magnetic properties no longer play a significant role), the many grain boundaries of the sintered material, which have an insulating effect, *increase the resistance* with increasing temperature. The following relationship between temperature and resistance $R(T)$ applies to this range:

$$R(T) = R_N \cdot \exp(A(T - T_N)).$$

It's in:

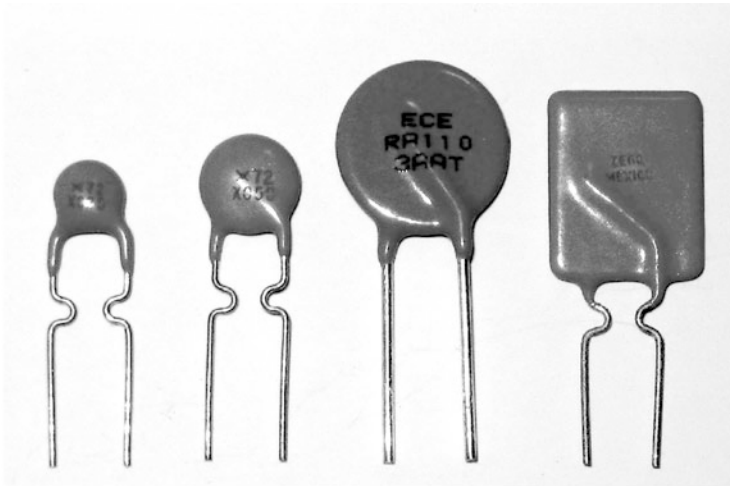


Fig. 2.69 Designs of PTC resistors (Factory photo: Polyswitch)

R_N : Nominal resistance at nominal temperature T_N

T_0 : Reference temperature (usually 25 °C)

A : Temperature coefficient in K^{-1}

$Exp(x)$ stands for e^x .

The temperature at the inflection point of the steeply rising characteristic curve is called the *nominal response ramp*. The value for the temperature coefficient at this point is, for example, $A = 0.16 K^{-1}$, that is, its value is about 40 times greater than with metallic resistors.

2.12.2.2 Advantages and Application Areas

These resistors have the advantages that they are cheap to manufacture, have small dimensions and react very quickly to temperature changes (much higher temperature coefficient than metals). Figure 2.69 shows the usual designs of PTC resistors.

The PTC sensors are mostly used for the following tasks:

- Temperature compensation,
- temperature monitor, that is, protection against thermal overload (from 60 °C to 190 °C in steps of 10 °C each),
- overcurrent protection (a high current generates a higher resistance in the PTC thermistor, which in turn limits the current).

2.12.3 Thermistors (NTC Resistors)

Thermistors are semiconducting, polycrystalline ceramics in which the *resistance decreases with increasing temperature*. Figure 2.70 shows the circuit symbol and the characteristic curves of the NTC resistors. The circuit symbol is similar to the PTC resistor with the following differences: As the designation “-T” indicates, the resistance is negatively dependent on the temperature, that is, the resistance decreases as the temperature rises. This is also shown by the two opposing arrows at the top right. Further properties are described in DIN 44070.

Many NTC resistors have a so-called *spinel structure* and are designed according to the following structural formula: $A^{2+} B^{3+}_2 O^{2-}_4$, where A is divalent and B is trivalent metals. Sintered ceramics made of mixed oxides usually consist of iron oxide (Fe_2O_3), magnesium dichromate ($MgCr_2O_4$) and barium nitrate. The negative resistance at increasing temperature is caused as follows: At higher temperature, more charge carriers are released, which causes the resistance to decrease.

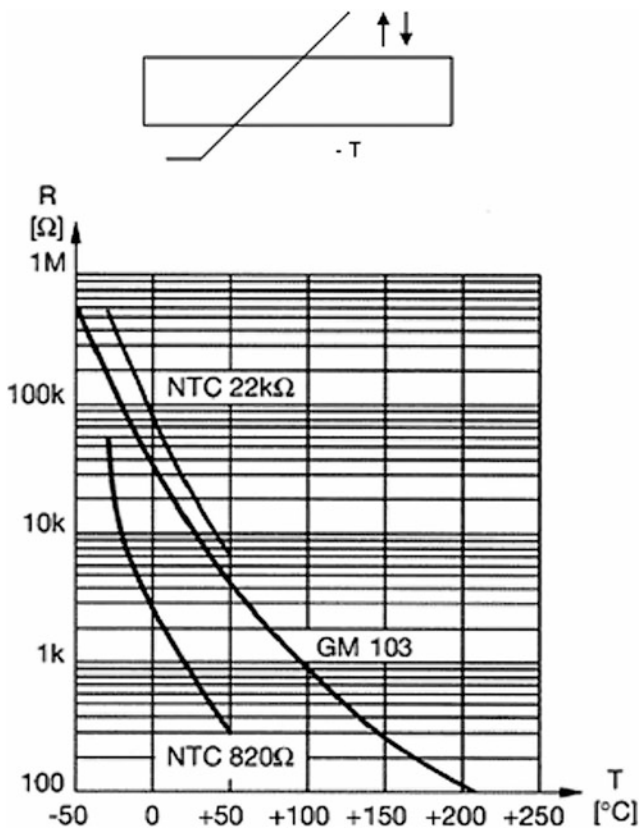


Fig. 2.70 Circuit symbol (top) and the characteristic curve of NTC resistors to DIN 44070 (bottom)

Table 2.20 Temperature dependence of the temperature coefficient A for the NTC resistor B3920/2 K with $R_0 = 2 \text{ k}\Omega$, $T_0 = 293 \text{ K}$ and $B = 2600 \text{ K}$

Temperature in $^{\circ}\text{C}$	Resistance R in Ω	Temperature coefficient A in $10^{-3}/\text{K}$
-40	19,651	-48
-20	8134	-41
0	3831	-35
20	2000	-30
40	1134	-26
60	688.8	-23

The resistance $R(T)$ depends on the temperature T as follows:

$$R(T) = R_N e^{B \left(\frac{1}{T} - \frac{1}{T_N} \right)}$$

There are:

R_N : Nominal resistance at the nominal temperature T_N (mostly 25°C)

B : Material constant (from 2900 K to 3950 K).

From the above formula, the slope of the temperature coefficient A for the temperature-resistance characteristic curve is: $A = - (B/T^2)$. Table 2.20 shows some values.

It should be noted that the reference temperature for the Pt resistor (Table 2.16) is 0°C , while the reference temperature for the NTC resistor and the PTC resistor is 25°C . The temperature-resistance curve can also be described approximately using the *Steinhart-Hart equation* (Chap. 6). According to this, the following applies:

$$1/T = a + b \ln(R) + c \ln^3(R).$$

Here a, b and c are coefficients that describe the behavior of the NTC resistor.

2.12.3.1 Advantages and Application Areas

The advantages are the same as with PTC resistors. Figure 2.71 shows the same designs as for the PTC resistors. In practice, NTC resistors play a much greater role than PTC resistors. The areas of application are very diverse. In order to be able to classify the components quickly and reliably, the first letter indicates the application. It means:

- F: Externally heated NTC resistors,
- K: Temperature compensation,
- M: Temperature measurement,
- R: Regulation of voltages.

Fig. 2.71 Designs of NTC resistors

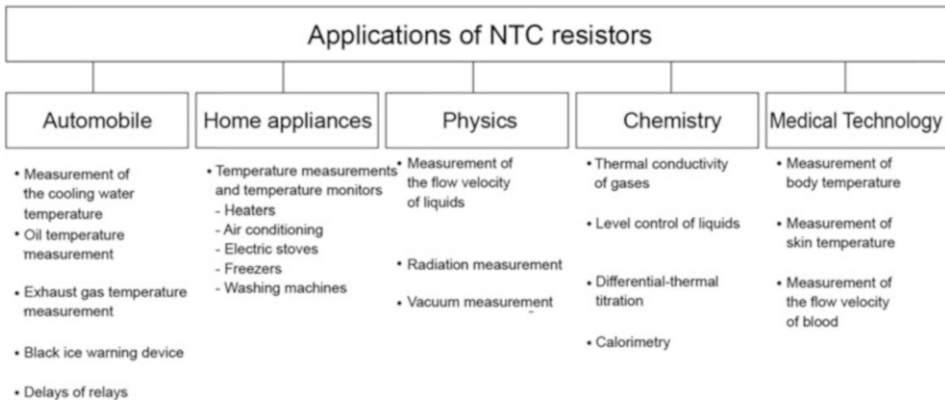
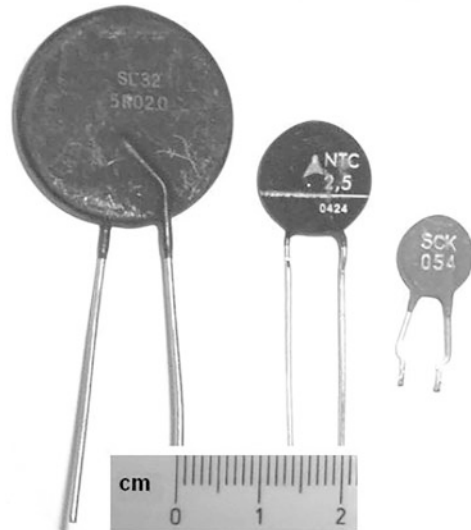


Fig. 2.72 Most important fields of application of NTC resistors

NTC resistors are mostly used for protection and compensation tasks. Figure 2.72 shows the most important areas of application.

2.13 Pyroelectric Effect

2.13.1 Principle of Operation and Physical Description

In pyroelectric materials, the *electrical polarization* P_{el} changes with *temperature changes* (cooling or heating). This produces *surface charges*, which is why a positive electric pole (*analog pole*) and a negative electric pole (*antilog pole*) are generated on the opposite surfaces (Fig. 2.73). The resulting voltage can be tapped. The pyroelectric effect occurs because the *distance between the grid ions* changes due to the change in temperature (similar to the piezoelectric effect, where the cause is forces or pressures, Sect. 2.1). This change in the distance between the lattice ions has two effects (prerequisite: the direction of the electrical polarization P points in the direction of the crystal axis):

1. *Change in length* (an expansion when heated) in the axis of the pyroelectric crystals, which has the same direction as the electrical polarization.
2. Change of *electrical polarization* with temperature.

Both effects act in the same direction and therefore reinforce each other.

The change of the electrical polarization ΔP with the temperature change ΔT can be described as follows:

$$\Delta P = k_p \cdot \Delta T,$$

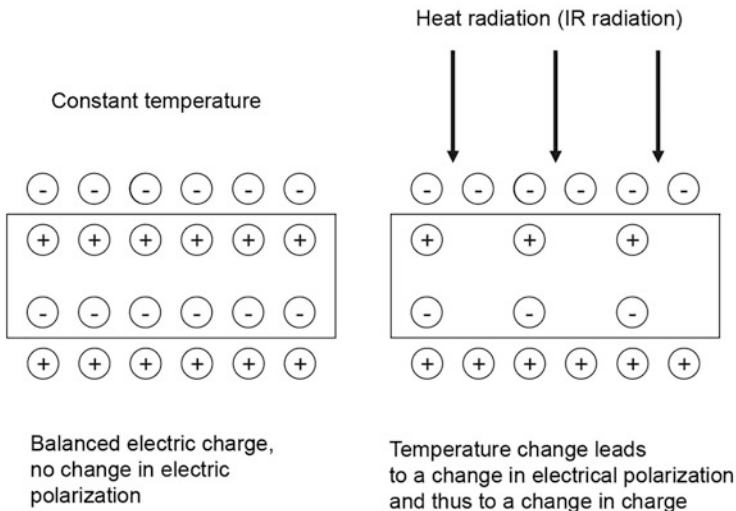


Fig. 2.73 Principle of the pyroelectric effect

where k_p is the *pyroelectric coefficient*. It is measured in $\text{C K}^{-1} \text{m}^{-2}$ and describes the *pyroelectric sensitivity*.

A pyroelectric sensor consists of a pyroelectric plate with a metallized upper and lower surface A , to which the electrical connections are attached. If the platelet is heated by the temperature ΔT , a charge difference ΔQ is created on the opposite surfaces. It applies:

$$\Delta Q = k_p \cdot A \cdot \Delta T.$$

Using electrical circuits analogous to the piezoelectric effect (Sect. 2.1), the pyroelectric capacitance C_p is calculated for a platelet with the thickness d , the area A and the permittivity number ϵ_r :

$$C_p = \epsilon_0 \cdot \epsilon_r \cdot A/d \text{ (electrical field constant } \epsilon_0 = 8.854187817 \text{ C V}^{-1} \text{ m}^{-1}\text{)}.$$

Thus, the voltage is ΔU :

$$\Delta U = \frac{\Delta Q}{C_p} = k_p \cdot A \cdot \Delta T \cdot \frac{d}{\epsilon_0 \epsilon_r A} = \frac{k_p}{\epsilon_0 \epsilon_r} \cdot d \cdot \Delta T.$$

For a sensor platelet of thickness $d = 15 \mu\text{m}$ made of lithium tantalate ((LiTaO₃)) with a permittivity number ϵ_r of 45 and a pyroelectric coefficient of $k_p = 2 \cdot 10^{-4} \text{ CK}^{-1} \text{m}^{-2}$, a temperature change of $\Delta T = 1 \text{ K}$ results in a voltage of 7.5 V.

The resulting surface charges are normally neutralized by corresponding charge carriers from the environment. The pyro sensors are therefore particularly sensitive to changes in radiation intensity over time and space. Preferred areas of application are therefore motion or fire detectors (Sect. 2.14).

Since pyroelectric materials also have piezoelectric properties, they are particularly sensitive to mechanical stresses or vibrations. Here the pyrosensor receives *piezoelectric interference signals*. These interferences are usually compensated by electrically connecting two mechanically coupled sensors against each other. A much better, but much more complex possibility is to apply several pyroelectric layers (2n layers) in a stacked arrangement. The electrical polarization is perpendicular to the layer plane and the polarization direction differs from layer to layer by 180 degrees. There are 2n + 1 electrodes between the layers and on the outside of the stack. This electrically symmetrical but mechanically asymmetrical arrangement of the layers also compensates for the piezoelectric interference signals, which are not linear.

2.13.2 Materials

Piezoelectric materials must have permanent electrical dipoles which are influenced by temperature changes. Possible materials are solid and liquid crystals, ceramics and polymeric plastics. All pyroelectric materials also exhibit the piezoelectric effect (Sect. 2.1), but not all piezoelectric materials exhibit the pyroelectric effect (e.g. quartz is not pyroelectric).

A distinction is made between the materials:

- *Pyroelectric crystals*

These include tourmaline, lithium tantalate (LiTaO_3), barium titanate (BaTiO_3 ; BTO) and lead zirconate titanate ($\text{Pb},\text{O},\text{Ti}/\text{Zr}$: PZT).

- *Pyroelectric ceramics*

Most pyroelectric materials are produced synthetically. Typical representatives are lead zirconate titanates (PZT).

- *Pyroelectric, polymeric plastics*

The most commonly used material is polyvinylidene fluoride (PVDF).

Table 2.21 shows the characteristic values of some pyroelectric materials.

2.13.3 Applications

Figure 2.74 shows the fields of application of pyroelectric sensors.

The pyroelectric effect currently finds its main application in the household and security sector in *motion detectors*. According to the Vienna displacement law ($\lambda_{\text{max}} \cdot T = \text{constant} = 2898 \mu\text{m K}$), a person with a body temperature of 36°C produces heat radiation with a wavelength of about $10 \mu\text{m}$ when the maximum energy radiation of a black body is reached. Figure 2.75 a) and a finished component of a motion sensor Fig. 2.75 b) show the schematic structure. In motion detectors, the *thermal radiation* (IR radiation) of objects in the range from -20°C to $+200^\circ\text{C}$ is measured without contact. The measurements are very fast (in the range ms) so that even very fast temperature changes can be detected. The pyro sensors can also be used as IR fire detectors. These pyro sensors are also called *passive*

Table 2.21 Characteristic values of selected pyroelectric materials at room temperature (20°C)

Material	Pyroelectric coefficient k_p in $10^{-4} \text{ C K}^{-1} \text{ m}^{-2}$	Permittivity number ϵ_r	Curie temperature in $^\circ\text{C}$
Triglycerine sulphate (TGS)	3.5	30	49
Lithium Tantalate (LiTaO_3)	2	45	618
Barium titanate (BaTiO_3)	4	1000	120
Lead zirconate titanate ($\text{Pb}(\text{Zr},\text{Ti})\text{O}_3$)	4.2	1600	340
Lead titanate (PbTiO_3)	2.3	200	470
Polyvinylidene fluoride film (PVDF)	0.4	12	80

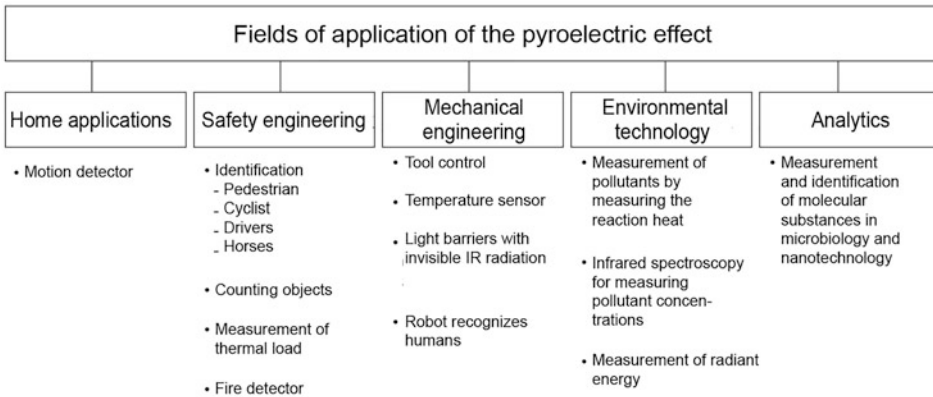


Fig. 2.74 Fields of application of pyroelectric sensors

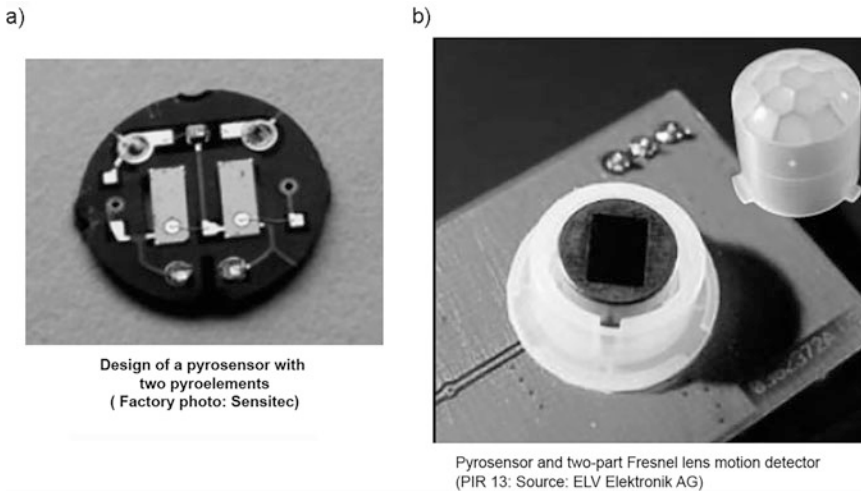


Fig. 2.75 Pyro sensors as motion detectors (a) Structure of a pyro sensor with two pyro elements (Factory photo: Sensitec); (b) Pyro sensor with upstream Fresnel lens as a motion detector (Factory photo: PIR 13 from ELV Elektronik AG)

infrared sensors (PIR). The word “passive” means that only radiated thermal energy is “received” but not emitted.

A *two-element sensor* according to Fig. 2.75 a) is particularly suitable for *motion detection*. The fact that they are located next to each other makes it possible to distinguish particularly well between background and passing people. An effect is only detected if the sensor elements absorb deviating heat radiation. The heat radiated by the object is focused by a Fresnel lens on the sensor element, where it is compared with a reference heat source and detected and amplified accordingly. In order to increase the observation range of the

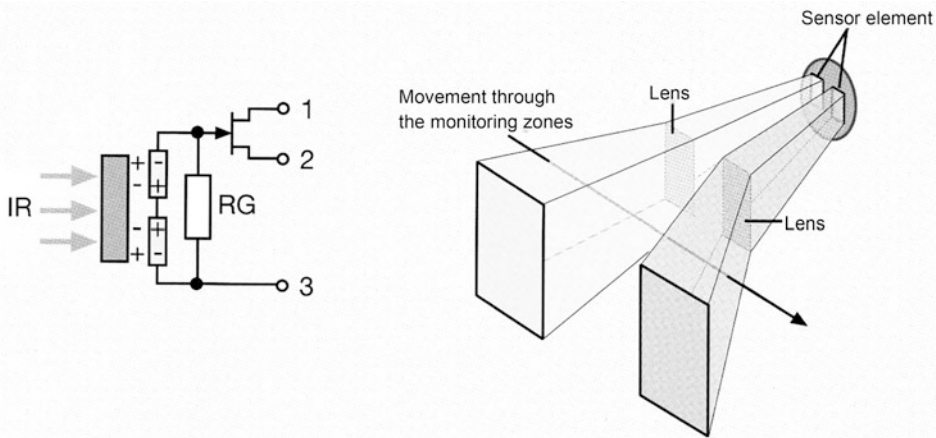


Fig. 2.76 Structure of a two-element pyrosensor and the effect of upstream lenses (Factory photo: PIR 13 from ELV Elektronik AG)

sensor, several lenses are connected upstream, as shown in Fig. 2.76. The lenses are arranged so that they can detect light from different zones.

Figure 2.75 b) shows an upstream Fresnel lens consisting of 17 individual lenses. This enables a detection range of at least 90 degrees. If a person moves through the individual zones, different charge differences are generated on the two sensor surfaces. It becomes possible to detect movement over a large area and at a relatively long distance (about 4 m). The lenses must have image properties in the area of thermal radiation. Modern pyrosensors achieve a bandwidth of 3.3 kHz with a spectral range of 1.8 μm to 23 μm .

Pyrosensors are *thermal detectors* and are therefore also suitable for measuring thermal energy independently of the wavelength of the incident radiation. The coaxial design of thermal detectors makes them largely insensitive to electromagnetic radiation (Fig. 2.77). The black absorption layer enables uniform absorption in the wavelength range from 185 nm to 25 μm . Repeat frequencies above 100 Hz are possible. Depending on the size of the absorption layer and sensor material, certain maximum thermal loads are possible. These thermal detectors allow energy measurements between 1 μJ and 2 J, at a maximum power density of 8 MW/cm^2 for a pulse of 10 ns.

Chemical reactions can cause *reaction heat*. Measuring this amount of heat allows clear conclusions to be drawn about the substances involved in the reaction and their concentrations. This method is successfully used to measure *concentrations* of, for example, oxygen and hydrogen as well as NO_x concentrations and other pollutants.

Artificial nanostructures that can be applied to biologically based surfaces (S-Layer: Surface Layer) are particularly promising. Bacterial membrane proteins are ideal for this purpose as the basic layer. They can be used to deposit defined, regularly arranged, nanometre-sized metals (e.g. platinum). Figure 2.78 shows a possible structure. The

Fig. 2.77 Energy measuring heads (Factory photo: Sensor- and laser technology Dr. W. Bohmeyer)



functional areas are in the order of $100\ \mu\text{m} \times 100\ \mu\text{m}$. The applied pyroelectric detector structures have a diameter of 2 nm to 30 nm with an extremely high packing density of more than 10^{12} per cm^2 .

2.14 Photoelectric Effect

2.14.1 Operating Principles and Physical Description

The photoelectric effect describes the following phenomenon: a *quantum of light*, that is, a photon is *absorbed* by an *electron*. This causes the electron to be *released* from its *bond*. The *kinetic energy* of the photon must be at least as great as the *binding energy* of this electron. As Fig. 2.79 shows, the following types of photoelectric effect are distinguished:

- *Outer photoelectric effect*

Electrons are released from metal surfaces by irradiation with short-wave electromagnetic radiation (e.g. light). This creates a *photocurrent*.

- *Inner photoelectric effect*

In a semiconductor, electrons are lifted from the valence band into the *conduction band* when exposed to light. This increases the electrical conductivity. A distinction is made between two effects:

- *Photoconductivity* (electron is released from its bond).
- *Optocoupler* (optical transmission of an electrical signal between two electrically isolated circuits).
- *Photovoltaic effect* (charge separation).
- *Photoionization*

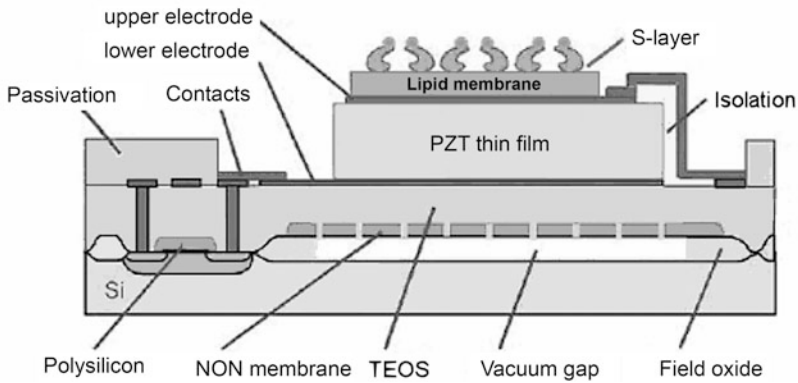


Fig. 2.78 Pyroelectric detector array based on microbiological substances. NON: nitric oxide nitride; TEOS: tetraoxysilane; PZT: Lead zirconium titanate (Source: Wolfgang Pompe, Institut für Werkstoffwissenschaft TU Dresden)

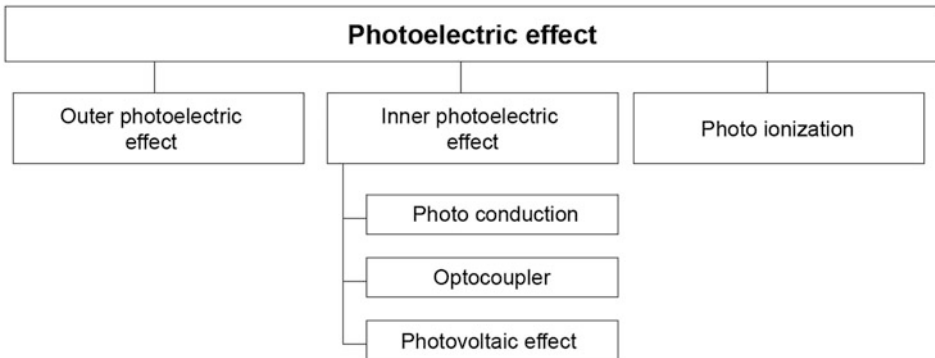


Fig. 2.79 Types of photoelectric effect

2.14.1.1 Outer Photoelectric Effect

Figure 2.80 shows the effect. The energy of a light source is transported in individual light quanta, the *photons*. Each emitted electron is triggered by a photon (Fig. 2.80 a)). This photon gives off its energy to the emitted electron. The individual photons can be determined with a *photomultiplier* (Sect. 14.3.1, Fig. 14.4).

The dependence of the kinetic energy E_{kin} of the emitted photoelectrons on the frequency f of the incident light represents a straight line whose equation is as follows:

$$E_{kin} = h \cdot f - W_A.$$

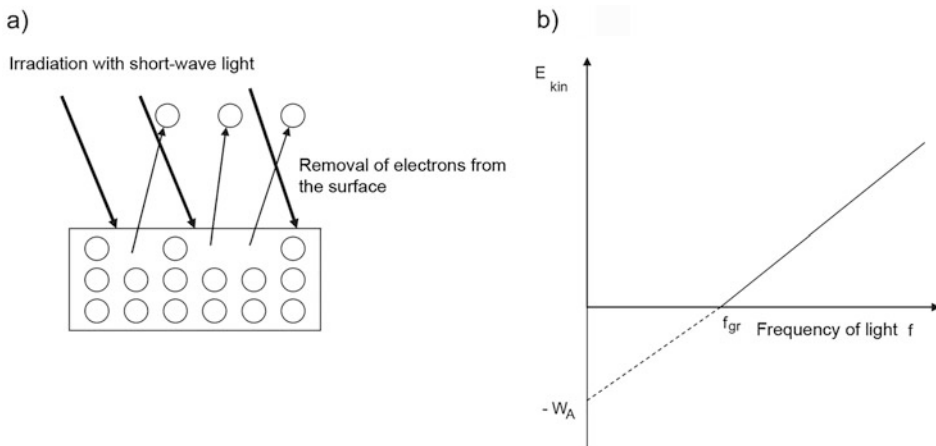


Fig. 2.80 External photoelectric effects (a) Principle (b) Kinetic energy of the emitted photoelectrons as a function of the frequency of the irradiated light

As Fig. 2.80 b) shows, h is the slope of the straight line, $-W_A$ is the intercept and the zero-crossing corresponds to the cut-off frequency f_{gr} :

The energy of the photon E_{ph} is $h f$. Here h is *Planck's quantum of action* ($h = 6.626 \cdot 10^{-34} \text{ Js} = 4.136 \cdot 10^{-15} \text{ eVs}$). It is the slope of the straight line in Fig. 2.80 b). In order for an electron to exit, an *exit energy* W_A must be applied. Then the kinetic energy E_{kin} mentioned above is available for the electron and it is:

$$E_{kin} = E_{ph} - W_A.$$

This also explains the *cut-off frequency* f_{gr} . The electron will only emit if the photon energy E_{ph} is greater than the required work function W_A . In the limiting case applies:

$$h f_{gr} = W_A.$$

These relationships mean that

- the *kinetic energy* of the emitted electrons (photoelectrons) does not depend on the intensity of the irradiated light, but only on the *frequency*.
- The photoemission comes to a standstill when the light frequency reaches a lower limit f_{gr} .
- If the intensity of the light is increased, the current of the emitted photoelectrons increases, but not their kinetic energy.

Since the energy of the photons E_{ph} is proportional to the frequency f of light, it must be inversely proportional to the wavelength λ , so that applies:

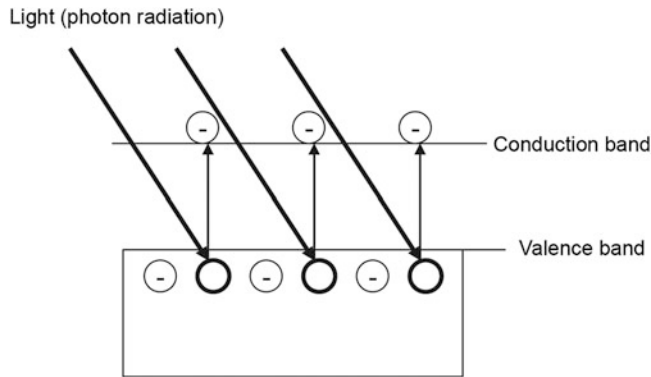


Fig. 2.81 Explanation of the photoconductivity with the band model

$$E_{\text{ph}} = h \cdot f = (h \cdot c) / \lambda.$$

For the product of the two natural constants one usually writes $h' = h \cdot c$ ($h' = 1.24 \text{ eV}\mu\text{m}$), so that for the photon energy applies:

$$E_{\text{ph}} = h' / \lambda.$$

Single photons can be determined with the photomultiplier (Sect. 14.3, Fig. 14.4).

2.14.1.2 Inner Photoelectric Effect: Photoconductor

The photoconductivity can be observed in *semiconductors*. An impinging photon releases an electron from its bond. As a result, it becomes mobile in the material, but it does not leave the material. Photoconductivity in this case means that the light energy of the photons increases the electrical conductivity. This effect can be explained in the *band model* as follows: The absorption of a photon causes an electron to be lifted from the valence band into the energetically higher conduction band, resulting in an *electron-hole pair* with the effect that more electrical charge carriers are available (Fig. 2.81). The energy of the photons must be at least as large as the energy gap E_g between the conduction band and the valence band:

$$E_{\text{ph}} \geq E_g.$$

With the above connection of the photon energy E_{ph} with the wavelength λ ($E_{\text{ph}} = (hc) / \lambda = 1.24 \text{ eV}\mu\text{m} / \lambda$) for the wavelength λ :

$$\lambda \leq (h c) / E_g \leq 1,24 \mu\text{meV} / E_g.$$

This means that the electrical conductivity increases significantly from the bandgap energy E_g .

2.14.1.3 Internal Photoelectric Effect: Optocoupler

An optocoupler transmits an *electrical signal* between *two electrically isolated* circuits. For this purpose, an optocoupler contains two components in one housing: a *light-emitting* element (usually a light-emitting diode, LED) and a *light-receiving* element (usually a phototransistor). For further versions see Sect. 14.4.

2.14.1.4 Internal Photo Effect: Photovoltaic Effect

If light shines on the pn junction of a semiconductor, charge separation takes place. The electrons migrate to the n-contact and the holes to the p-contact. The *solar cell* has a grid of thin contact fingers on its surface which conduct the generated photocurrent. Figure 2.82 shows the structure of a solar cell. Because of the high degree of reflection of the semiconductors, the surface of the solar cell always has an anti-reflection layer.

In short-circuit operation, a photocurrent I_K flows through the solar cell, which is proportional to the irradiated power Φ_e , so that the following applies:

$$I_K \sim \Phi_e = E_e A \quad (E_e: \text{Irradiated photon energy}; A: \text{area}).$$

2.14.1.5 Photoionization

If a light beam (photon stream) strikes atoms or molecules, one or more electrons can be knocked out of the atomic or molecular structure so that the remaining atoms or molecules are ions (charged particles). If a photon gives off all its energy to an electron, this is called

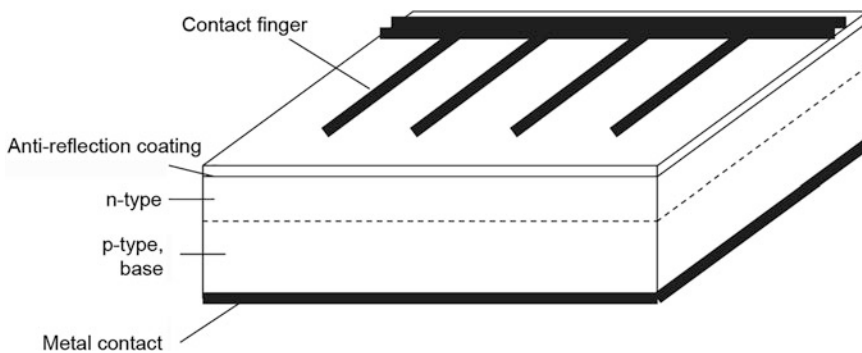


Fig. 2.82 Construction of a solar cell for the photovoltaic effect

the photoelectric effect in nuclear physics. If the electron absorbs only a part of the photon energy and the remaining energy is re-emitted as a photon of lower frequency or longer wavelength, then this is the *Compton effect*.

The *effective cross section* σ for photoionization is dependent on the atomic number Z and the photon energy E_{ph} as follows

$$\sigma \sim Z^5 E_{\text{ph}}^{-7/2}.$$

Because the effective cross section σ is proportional to the fifth power of the atomic number Z , materials with high atomic numbers absorb X-rays and γ radiation particularly well (e.g. lead with $Z = 82$).

As the negative power of photon energy shows, the effective cross section σ decreases with increasing photon energy E_{ph} . This, however, only as long as electrons are available for ionization. As soon as the photon energy reaches the binding energy of the next higher electron shell, the effective cross section increases abruptly, which then decreases again with increasing photon energy until the next electron shell is reached. Figure 2.83 shows this *comb-like* course of the effective cross section σ of the photon energy E_{ph} .

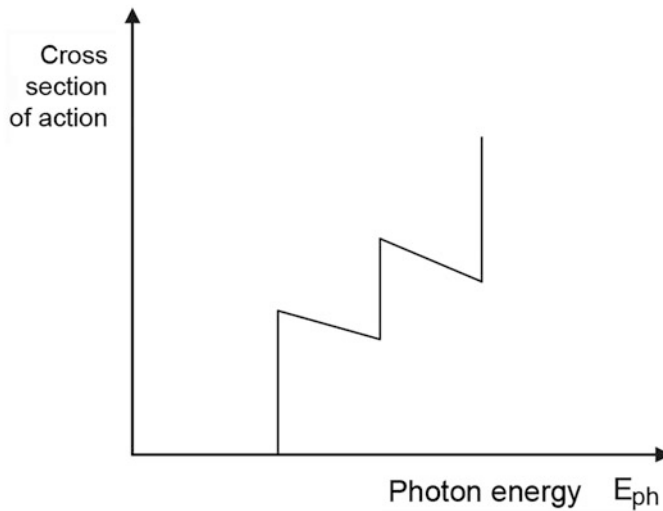


Fig. 2.83 Effective cross section σ as a function of the photon energy E_{ph}

2.14.2 Photoelectric Sensor Elements

The internal photoelectric effect offers the most common sensory applications (Table 2.22).

Figure 2.84 shows the different semiconducting photoelectric elements and the dependence of the electrical properties on the illuminance.

2.14.3 Photoelectric Sensor Elements

Figure 2.85 shows the application areas of photoelectric sensors.

2.14.3.1 Photomultiplier

It consists of a *photocathode* and a downstream *secondary electron multiplier* (SEV, Sect. 14.3.1). The photons hit the photocathode and release electrons. These hit further electrodes (dynodes), where electrons are also released. For this reason, the number of electrons released increases *avalanche-like* from dynode to dynode. To achieve this, a high voltage is applied, which is distributed by voltage dividers in such a way that the electrons are accelerated from dynode to dynode. Finally, the electrons hit an anode and generate a voltage drop in a resistor, which represents the measurement signal. The amplification factor increases exponentially with the number of dynodes. Typical photomultipliers have about 10 dynodes. If 5 more are knocked out at each dynode for every electron that hits it, 5^{10} electrons (about ten million) are finally produced. Figure 2.86 b) shows an 80 mm long photomultiplier. On the left side is the entrance window and in the middle are the dynodes attached to insulating bodies.

Table 2.22 Photoelectric elements based on the internal photoelectric effect

Inner photoelectric effect	Description
Photo element	Active two-pole (does not require a voltage source), delivers a voltage when irradiated (depends logarithmically on irradiance).
Photodiode	Requires auxiliary voltage, supplies a current linear to the illuminance when irradiated.
Photo winch diode (Avalanche photodiode, APD)	A high reverse voltage (between 6 V and 10 V) causes an impact ionization in the barrier layer. This results in an avalanche-like multiplication of the conduction electrons. The amplification is $1:10^5$, which is why the current intensity increases steeply. It is possible to detect very low power (10^{-12} W) and a frequency resolution up to GHz (10^9 Hz).
Photoresistor	The ohmic resistance changes with irradiation.
Phototransistor	Requires auxiliary voltage, supplies a current linear to the illuminance when irradiated. Sensitivity 100 to 500 times greater than that of the photodiode.
Photothyristor	Switching voltages are much higher than with the photoresistor, photodiode and phototransistor.

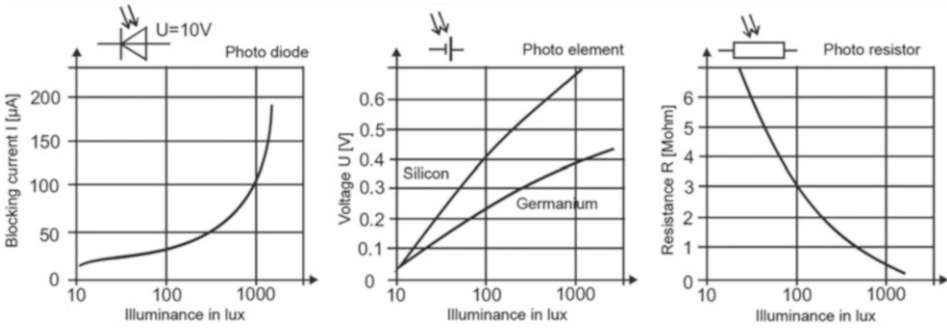


Fig. 2.84 Semiconducting photoelectric elements and the dependence of their electrical properties on illuminance (Source: Hesse, Schnell: Sensors for process and factory automation, Sensoren für die Prozess- und Fabrikautomation, edition 4, Vieweg+Teubner 2009; Fig. 2.66)

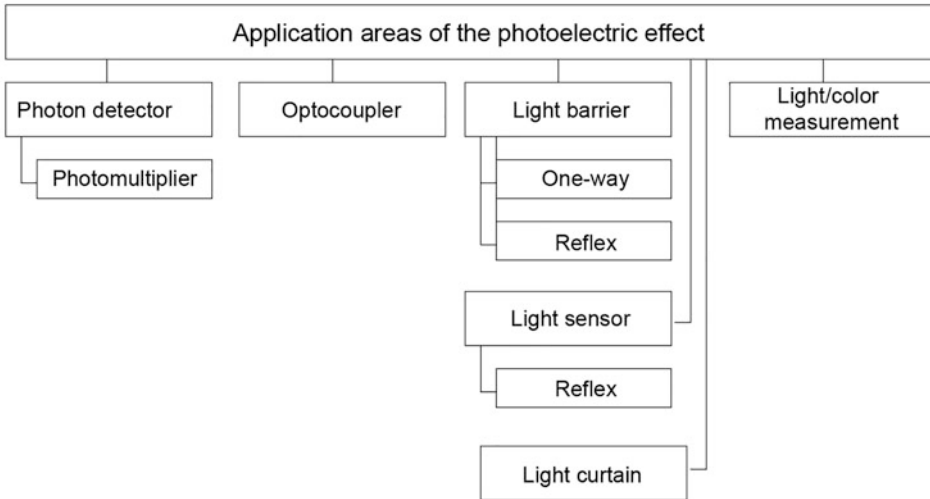


Fig. 2.85 Application fields of electro-optical sensors

Photomultipliers are used to measure weak light signals, down to single photons. They are used for the detection of elementary particles or for measuring cosmic radiation.

2.14.3.2 Optocoupler

As explained in detail in Chap. 14, the optocoupler transmits an electrical signal when input and output circuits are electrically isolated. The actuator (e.g. LED) and detector (e.g. phototransistor) are located in one component. Figure 2.87 a) shows the circuit symbol and the mode of operation. If the switch is closed in the upper left corner, the LED lights up. This light signal is registered by a phototransistor and passed on as an electrical signal.

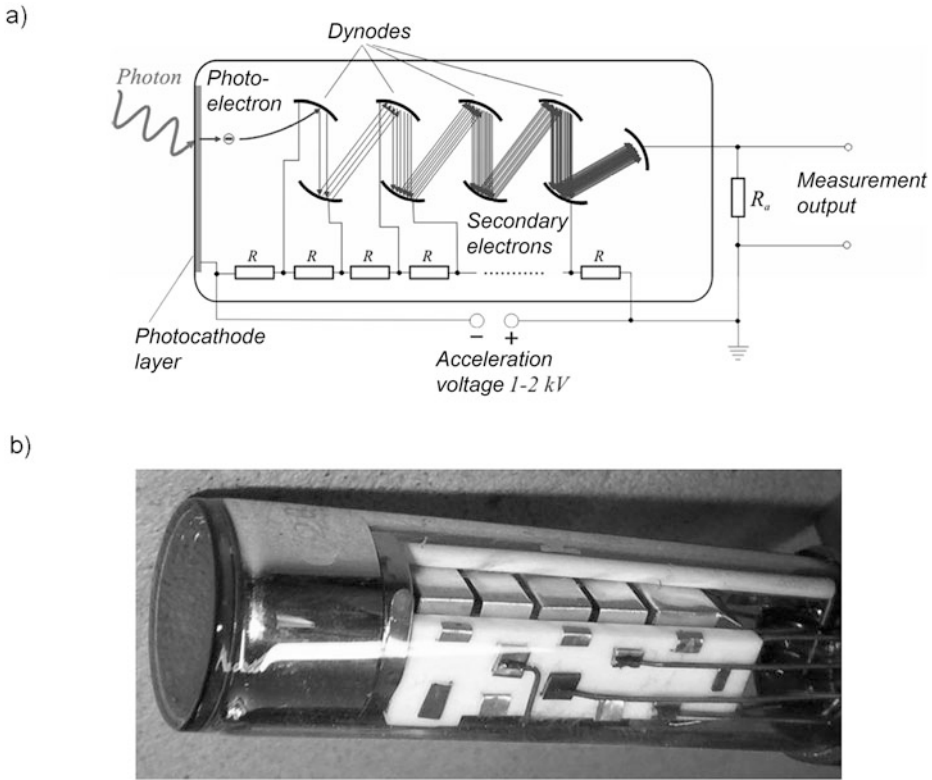


Fig. 2.86 Photomultiplier (a) mode of operation; (b) component (Photo: Ulf Seifert)

Figure 2.87 b) shows a closed component as Solid State Relay (SSR) and Fig. 2.87 c) with an open optical beam path. The left component comes from a video recorder and the right component from a printer.

The following characteristic values are typical for optocouplers:

- *Insulation voltage* (usually: 1.4 to 4 kV; up to a maximum of 25 kV),
- *Insulation resistance* (very high, up to $10^{13} \Omega$),
- *Blocking voltages* (transmitter diode: 5 V; receiver: phototransistor: 30 V to 50 V; photothyristors: up to 400 V),
- *Cut-off frequency* (highest operating frequency: 50 kHz to 200 kHz; photodiode up to 10 MHz) and
- *Switching times* (phototransistor slowest: in the range ms).

Optocouplers have the following advantages: *Digital* and *analog* signals can be transmitted, *small dimensions*, *no interfering inductors* (magnetic fields), *low coupling capacitances* and *no mechanical wear* (i.e. extremely many switching cycles).

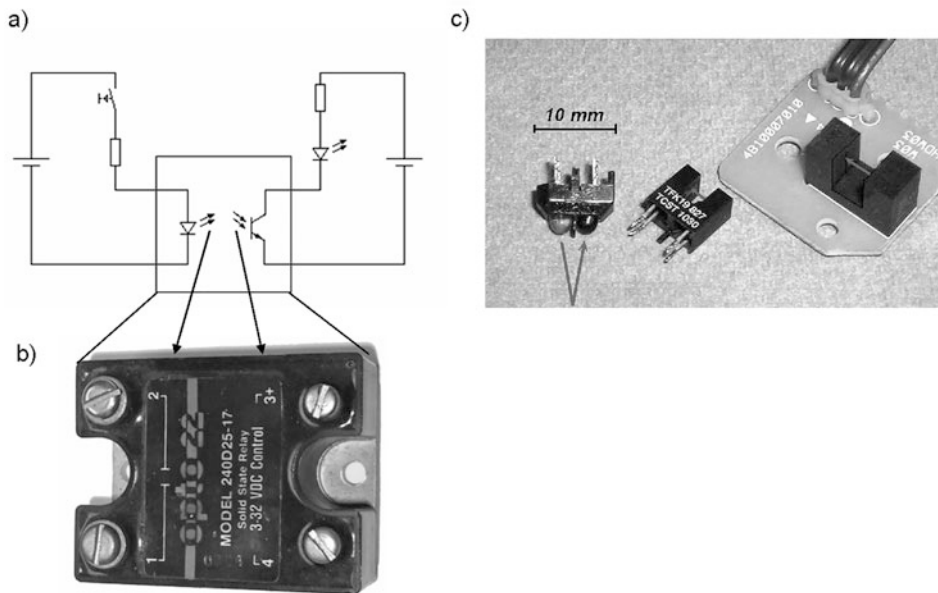


Fig. 2.87 Optocouplers (a) Mode of operation (b) Closed component (source: Martin Broz); (c) Components with open optical beam path (fork couplers and reflex couplers; source: Ulf Bastel)

Optocouplers are mostly used where circuits must be electrically isolated from each other. Typical areas of application are, therefore:

- *Network cards* and *interface cards* in computers (connected devices have different voltages and must therefore be electrically separated).
- *Protective measures* in modules against overvoltages and interference pulses (e.g. in industrial controls, PLCs or medical equipment); if the input section (LED) is destroyed, the output section remains protected and only the optocoupler needs to be replaced.
- In *switching power supplies* the transmission of control information from the secondary to the primary.
- *Driving* of circuit parts that have different voltages.

2.14.3.3 Light Barriers

With *through-beam photoelectric sensors* (Sect. 3.4, Table 3.30), the *transmitter* and *receiver* are spatially and optically *separated* (DIN 44030). If the light beam between the transmitter and receiver is interrupted by an object, the output switches. An advantage is that no minimum distance between transmitter and receiver is required. A high resolution is possible with laser light (object up to 0.1 mm in size). However, in this case, the through-beam photoelectric sensor is also sensitive to small interfering objects (e.g. if a fly flies into

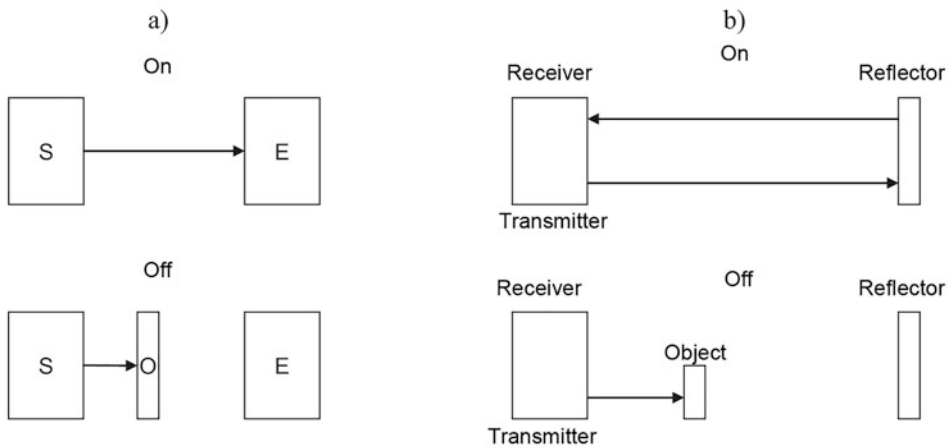


Fig. 2.88 Light barrier (a) Throughbeam light barrier; (b) Reflector light barrier

the light beam). A further advantage is the very high switching speed. Use is possible up to 100 m; however, distances of up to 10 m are common in practice.

In the case of *reflex light barriers*, the transmitter and receiver are accommodated in *one housing* (Fig. 2.88 b)). A reflector reflects the emitted light. If the light beam between transmitter and receiver is interrupted by an object, the sensor switches. The retro-reflective photoelectric sensor can be adjusted better than the throughbeam photoelectric sensor because the reflector can be designed accordingly large. Usually, the distances between the receiver are up to 5 m. Special polarisation filters can be used to ensure that even bare metal parts and glass surfaces are detected as objects.

With a *position-sensitive photodiode* (PSD: position sensitive device), *distances* can also be measured according to the principle of *triangulation*. Figure 2.89 a) shows the principle. It is a calculation in a triangle (therefore: triangulation). The measuring point P is targeted from two different positions. If the distance d between the positions is known, the coordinates of the measurement point P can be calculated from the angles α and β . *The following* generally applies in a triangle: From two points of a straight line whose distance is known, the coordinates of any points in space can be determined by angle measurements. If a light beam is emitted at an angle and reflected by an object, then, depending on the position of the object, the reflected beam hits a different point on the detector (Fig. 2.89b)). The coordinates of the distance of the object are calculated from this point. A prerequisite for this is that a *PSD* is used which measures the photocurrent exactly at the point of incidence of the reflected beam. As Fig. 2.89 c) shows, reflex light barriers can be used to distinguish between different objects on a conveyor belt using the triangulation principle.

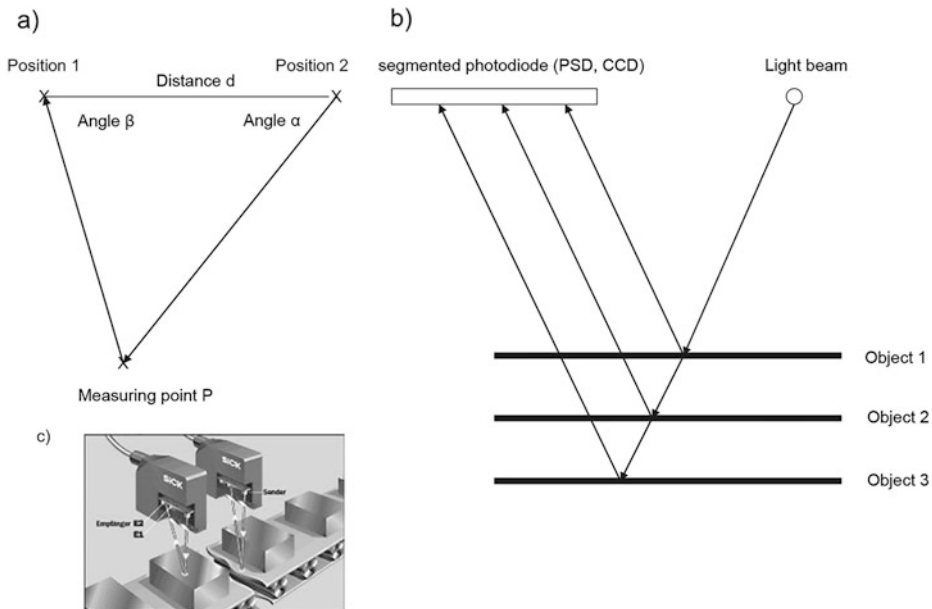


Fig. 2.89 Triangulation (a) Principle; (b) Distance measurement with a PSD; (c) Sensor for differentiating objects on a conveyor belt (Factory photo: Sick)

2.14.3.4 Photoelectric Switches

The transmitter and receiver are housed in one housing. The object acts as a reflector. When the object reaches a defined distance, the output switches. These sensors also work according to the principle of triangulation. The measurement is only based on the distance and not on the reflection properties. Therefore, the shape and surface quality of the object have no influence on the sensing distance. Figure 2.90 shows the mode of operation. If a 2-segment photodiode is used, reliable object detection becomes possible.

In *triangulation*, the angle between the emitted and reflected light beam is measured. An object to be detected and background reflect the light at different angles when they are spatially separated. The probe is adjusted to the distance of the object to be scanned. Then all parts which are behind the set threshold value for the scanning distance of the object are safely blanked. This is the principle of *background suppression HGA*.

If the reflected light beam falls on *Segment A*, the sensor *detects* the object. With segment B, the object is not detected (Fig. 2.90 a)).

If the object and background are very close together or if the object is shiny or uneven, then *foreground suppression (VGA)* is suitable. If the reflected light beam falls on segment B, the sensor registers "no object present" (Fig. 2.90 b)). Therefore, foreground suppression can only function properly if the background is present.

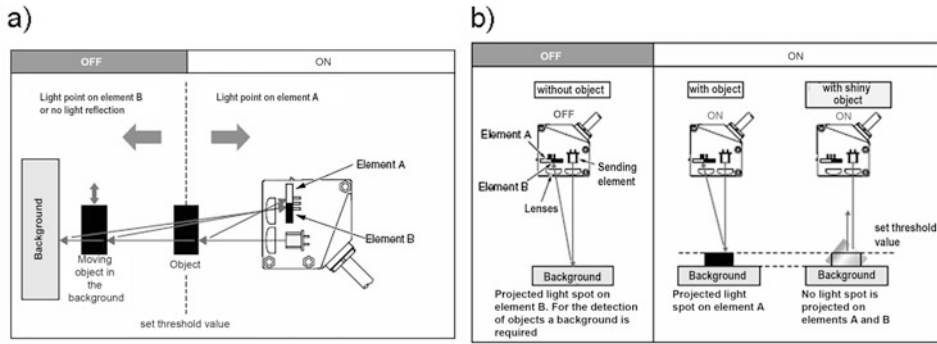


Fig. 2.90 Photoelectric proximity switches (a) with background suppression; (b) with foreground suppression (Factory photos: Samsung)

With photoelectric switches and photoelectric proximity switches, the light can also be guided to the light receiver or reflector via *optical fibres*. This is very interesting for applications in difficult spatial conditions or harsh industrial environments.

2.14.3.5 Light Curtain

If individual light barriers are arranged in rows, *light grids* or *light curtains* are created. Figure 2.91 shows the components of a light curtain (Fig. 2.91 a)) and a safety application (Fig. 2.91 b)). The international safety standards essentially apply for use in the protected area: EN ISO 13849, EN ISO 13855 and EN ISO 13857 as well as the European standards EN 292, EN 294, EN 954–1, EN 1050 and EN 61496. The following points must be taken into account for the design:

- *Resolution*. It is 14 mm for *finger protection* and 30 mm for *hand protection*.
- *Protective field height*. It must not be possible to reach over or under it. More detailed regulations are laid down in the EN 294 standard.
- *Range*. It depends on the resolution. For finger protection, it is usually up to 9 m and for hand protection up to 30 m.
- *Response time*. It depends on the resolution. The normal response time is 60 ms.

Measurements can also be made with light curtains. The *angle-dependent reflections* allow conclusions to be drawn about the course of *contours*.

2.14.3.6 Light Measurement

The *brightness sensation* of the eye (depending on the wavelength) plays an important role in light measurement. A measure of the impression of brightness is the *luminous flux*. It is measured in lm (lumen) (Sect. 8.2). For practical applications, the *illuminance* plays an important role. It is a measure of the luminous flux per m² and is measured in lx (lux: 1 lx = 1 lm/m²). The technical lighting regulations for offices, factory buildings, schools,

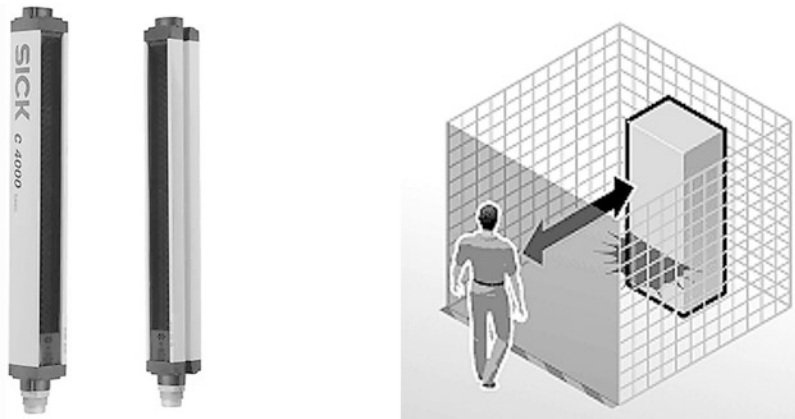


Fig. 2.91 Light curtain (a) Components; (b) Application (Factory photos: Sick)

VDU workplaces and similar areas are drawn up by the Lighting Technology Standards Committee. European regulations are drawn up by CEN/TC (CEN: European Committee of Standardization; TC: Technical Committee). The standard CEN/TC 169 applies to the field of “light and lighting”, while DIN 50356:2006–11 applies to artificial lighting (lighting with artificial light; Part 6: Measurement and assessment). The most important measuring instruments are lux meters (Fig. 2.92).

2.14.3.7 Colour Recognition

The color sensors usually work according to the *three-range method*. They send light (red, blue, green) onto the objects to be inspected and determine the colour components from the reflected beams. If the colour components are within a defined tolerance, the colour value is detected and a circuit is triggered. Figure 2.93 a) shows a color sensor during a component inspection and Fig. 2.93 b) the application during the positioning of a can. Detailed information can be found in Sect. 8.4.

Figure 2.94 shows the areas of application of the photoelectric effect. Photoelectric sensors play an important role, especially in mechanical engineering, automation technology and robotics, as well as in safety engineering, but also the determination of elementary particles or cosmic radiation.

Fig. 2.92 Light measurement with luxmeter (Factory photo: Beha Amprobe)

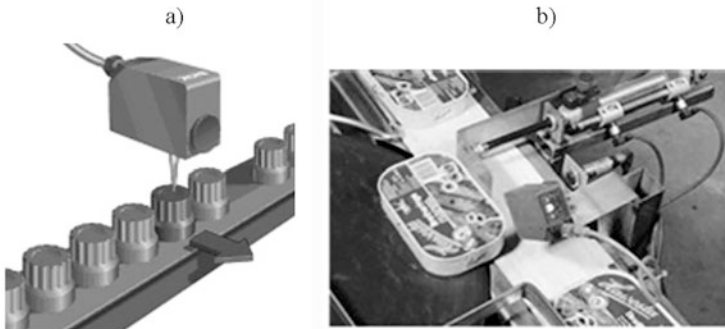


Fig. 2.93 Color sensors (a) Components; (b) Application for positioning (Factory photos: SICK)

2.15 Electro-optical Effect

2.15.1 Principle of Operation and Physical Description

Electro-optical effects generally occur when *electrical quantities* affect *optical properties* (e.g. conversion of electrical energy into light or vice versa). In the special case at hand, the electro-optical effect is understood to be the *change in the refractive index n* in the presence of an electric field E . The following relationship then applies:

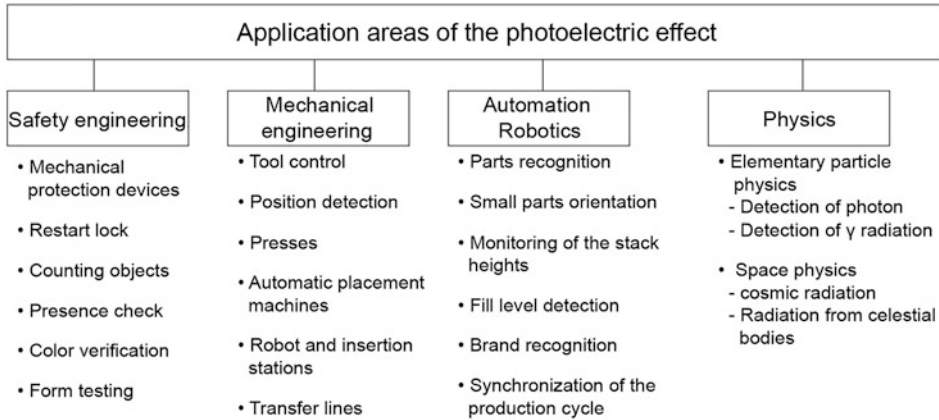


Fig. 2.94 Applications of the photoelectric effect

$$n(\mathbf{E}) = n_0 + r \cdot \mathbf{E} + K \cdot \mathbf{E}^2,$$

whereby the following applies:

- n_0 : Refractive index without electric field
- r : Electro-optical constant
- K : Kerr constant.

It should be noted that the constants r and K are generally *tensors*, that is, they have different values in different directions in space. This is due to the symmetry of the crystals. However, there is always a preferred direction. For this direction the constants are decisive.

The above equation describes two effects:

$$\text{Pockels effect : } n(\mathbf{E}) = n_0 + r \cdot \mathbf{E}.$$

It is $K = 0$, that is, the effect is *linearly* dependent on the electric field strength \mathbf{E} .

$$\text{Kerr effect : } n(\mathbf{E}) = n_0 + K \cdot \mathbf{E}^2.$$

It is $r = 0$. The Kerr effect shows a *quadratic* dependence of the refractive index on the electric field strength \mathbf{E} .

Figure 2.95 shows the light modulation of a Pockels cell.

The Pockels cell according to Fig. 2.95 consists of a crystal on whose front sides a transparent metal film is applied. In front of it is a polarizer P and behind it a polarization analyzer A, which allows a polarization of 90 degrees to pass through completely. If a voltage U and thus an electric field \mathbf{E} is applied in a longitudinal direction, the crystal

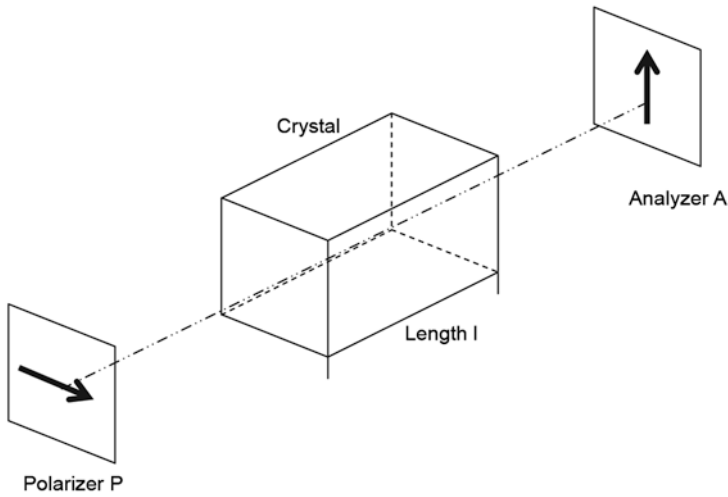


Fig. 2.95 Principle of light modulation in a Pockels cell

becomes *birefringent*. As a result, the ordinary and extraordinary waves, whose direction of oscillation is perpendicular to each other, pass through the crystal at *different speeds*. At the end of the crystal therefore two waves arrive with a *path difference* Δ . The superposition results in *elliptically polarized* light which cannot be completely retained by analyzer A. Depending on the direction of polarization, the light is more or less transmitted. With a path difference of half a wavelength ($\Delta = \lambda/2$), linearly polarized light is obtained again. It is rotated by 90 degrees compared to the original polarization direction and is therefore completely transmitted through the analyzer.

Phase and *amplitude modulations* can be made using the Pockels and Kerr effects. Thus, it is also possible to transmit information through light with the electro-optical effect.

2.15.2 Materials

Table 2.23 shows the differences between the Pockels and Kerr effect. Table 2.24 shows the materials and their properties of Pockels cells, Table 2.25 of Kerr cells. It must be noted that the corresponding constants have characteristics in different spatial directions, that is, they are tensors. Only the most important direction is selected in the tables.

2.15.3 Applications

Figure 2.96 shows the application areas of electro-optical sensors. Electrooptical cells can switch light *without inertia*. As a result, applications in *high-speed photography* (e.g. real-

Table 2.23 Comparison of the Pockels and Kerr Effect (Source: Hering, Martin, Stohrer, Physics for Engineers, edition 13, Springer)

Features	Pockels Effect	Kerr Effect
Declaration	Piezoelectric crystals without a symmetry center are birefringent in the electric field	Optically isotropic material becomes birefringent in a transverse electric field
Dependence on the electric field E	$ n_o - n_e = r \cdot E$	$ n_o - n_e = K \cdot E^2$
Aisle difference after Δ passing through the length ℓ	$\Delta = r \cdot \ell \cdot E \cdot n_o^3$	$\Delta = K \cdot \ell \cdot E^2 \cdot \lambda$
Geometry	Electric field E mostly in the longitudinal direction (also transversal possible)	Electric field E perpendicular to the direction of propagation of light
Typical field strength for the gear difference $\Delta = \lambda/2$	$U \approx 4 \text{ kV}$ (longitudinal cell of KD-P (deuterated potassium hydrogen phosphate))	$E \approx 10^6 \text{ V/m}$
Modulation frequency	Modulated up to over 1 GHz	Up to about 200 MHz

Table 2.24 Refractive index n_0 and electrooptical constant r in materials with the Pockels effect

Material	Refractive index n_0	Electrooptical constant r in 10^{-12} m/V
Quartz	1.54	1.4
Gallium arsenide (GaAs) ($\lambda = 10.6 \text{ }\mu\text{m}$)	3.3	1.51
Zinc sulfide (ZnS) ($\lambda = 600 \text{ nm}$)	2.36	2.1
Ammonium dihydrogen phosphate $(\text{NH}_4)_2\text{H}_2\text{PO}_4$ (ADP) ($\lambda = 546 \text{ nm}$)	1.48	8.56
Potassium dihydrogen phosphate KH_2PO_4 (KDP)	1.51	11
Potassium dihydrogen phosphate deuterised KH_2PO_4 (KD*P)	1.51	24.1
Lithium tantalate (LiTaO_3)	2.18	30.5
Lithium niobate (LiNbO_3)	2.29	30.9

Table 2.25 Kerr constant in selected materials

Material	Kerr constant K in 10^{-12} m/V^2
(20 °C; $\lambda = 589 \text{ nm}$)	
Nitrogen (N_2) under normal conditions	$4 \cdot 10^{-6}$
Other gases	10^{-3}
Carbon disulfide (CS_2)	0.036
Water (H_2O)	0.052
Nitrotoluene ($\text{C}_7\text{H}_7\text{NO}_2$)	1.4
Nitrobenzene ($\text{C}_6\text{H}_5\text{NO}_2$)	2.4

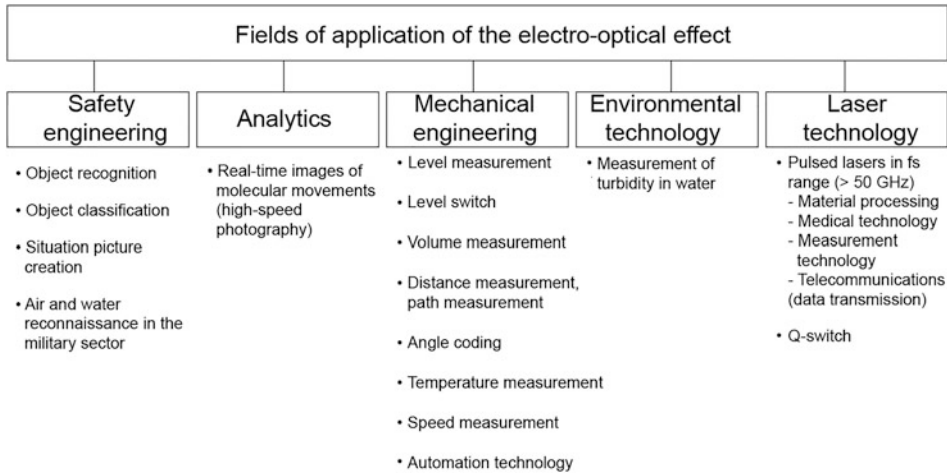


Fig. 2.96 Fields of application of electro-optical sensors

time recordings of molecular movements or atomic vibrations) or light modulation in sound film and image radio are of particular interest. In the military sector, fast and reliable *object detection* (e.g. missiles) and *object classification* (friend/enemy; missile, aircraft, submarine) are possible. This sensor type, therefore, plays an important role in reconnaissance on water and the air.

A wide range of applications is also conceivable in mechanical engineering, especially in automation and production technology (e.g. processing with ultrashort laser pulses). Figure 2.97 shows a level measurement with an electro-optical sensor. This consists of an infrared LED and a light receiver. As long as the tip is not immersed in the liquid, the light within the sensor is reflected to the receiver. If the liquid level rises and surrounds the tip, light in the liquid is refracted so that only part of the light reaches the receiver. This incidence of light is detected and electronically evaluated in order to initiate appropriate measures (e.g. alarm or shutdown).

In order to obtain extremely intense laser pulses in the femtosecond range ($f_s = 10^{-15}$ s), laser pulses with a higher intensity must be amplified. Figure 2.98 shows such an arrangement. A laser beam has a *Gaussian-shaped intensity distribution*, that is, the radiation intensity is highest in the middle range and then decreases at the edges. The refractive index of a Kerr element depends on the light intensity. Therefore, intense light is focused more strongly than non-intensive light. The intense laser pulses are also much shorter. The following effect occurs: The *intense laser light* in the middle is *focused more strongly*. An aperture ensures that only this more intense light is transmitted and reflected back by a mirror. In the next runs, it is further focused so that *increasingly intense and shorter laser pulses* are generated. In this way, laser pulses in the fs range can be generated. The amplification has an end when the material (aperture, mirror, Kerr lens) is destroyed.

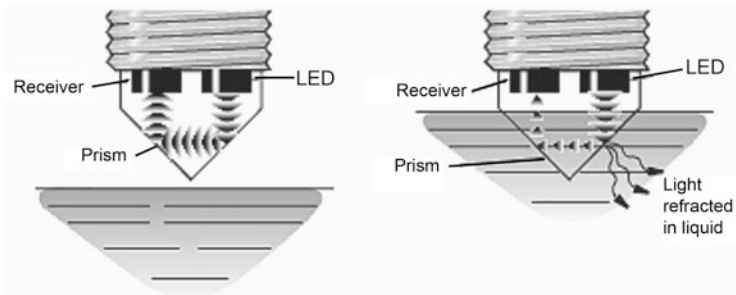


Fig. 2.97 Level measurement with an electro-optical sensor (Factory photo: Meyer Electronics in industry– Meyle)

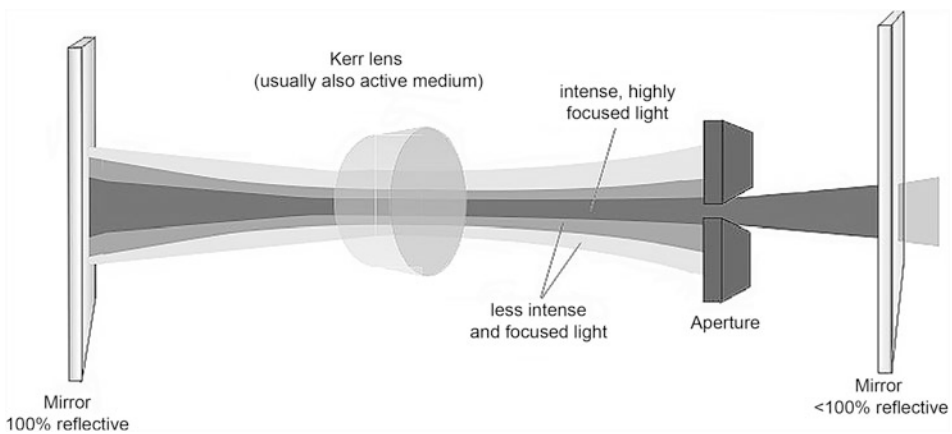


Fig. 2.98 Self-focusing of laser beams through a Kerr lens (Photo: Research Center Dresden-Rossendorf)

2.16 Electrochemical Effects

2.16.1 Principle of Operation and Classification

If a *chemical reaction* is linked to an *electric current* (a combination of mass and charge transport), this is an *electrochemical process* that can be used for *electrochemical sensors*. One often differentiates between *potentiometric*, *amperometric* and *conductometric* or *impedimetric* sensors. Sometimes measuring arrangements *working on the principles of coulometry and voltammetry* are also included. Recently, it has also been recommended that a subdivision be made between voltammetric and potentiometric sensors, as well as chemically sensitized field-effect transistors and potentiometric solid-state electrolyte gas sensors. Figure 2.99 shows a frequently used overview of electrochemical sensors. In

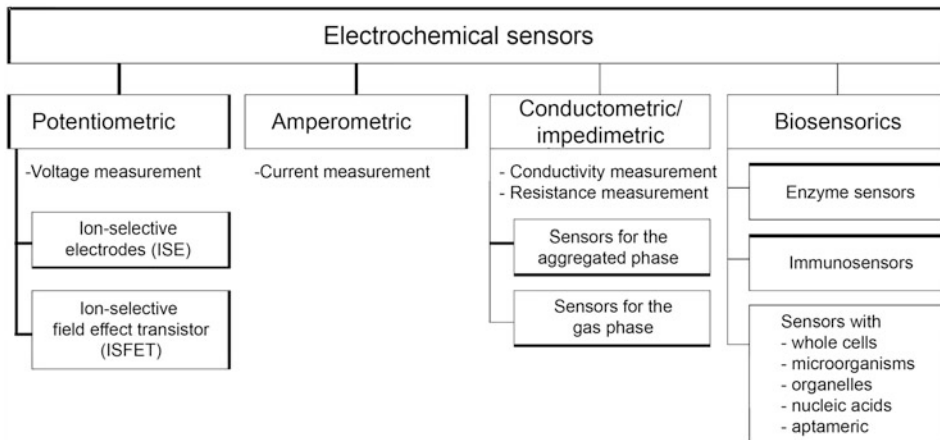


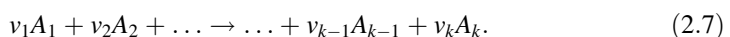
Fig. 2.99 Overview of electrochemical sensors

particular, potentiometry and amperometry play an essential role in a number of *biosensors* (Chap. 12) as basic elements. Potentiometric sensors can be used to determine the activities of special ions, among other things. In these cases, one speaks of *ion-selective* electrodes, which can also be realized as *field-effect transistors* (ISFET: ion-sensitive field-effect transistor).

2.16.2 Potentiometric Sensors

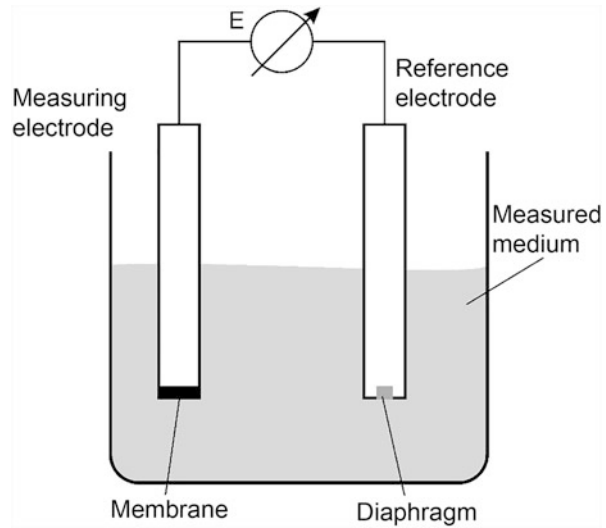
Potentiometric sensor technology, which involves measuring a *potential difference* between measuring and a reference electrode, is based on the *Nernst equation*, which is fundamental to physical chemistry and is formulated in its general form in eq. (2.6). $\Delta_R G$ and $\Delta_R G^\ominus$ represent the change in free enthalpy or its value for the standard state with respect to the reaction given by Eq. (2.7), in which k substances $A_1 \dots A_k$ are involved. R is the general gas constant ($R = 8314 \text{ Ws mol}^{-1} \text{ K}^{-1}$), T the temperature, $\{a_i\}$ the activity of substance i related to the standard activity and ν_i the stoichiometric coefficient of substance i in the reaction equation.

$$\Delta_R G = \Delta_R G^\ominus + RT \ln \prod_{i=1}^k \{a_i\}^{\nu_i} \quad (2.6)$$



If the so-called Eq. (2.8) containing the *electromotive force* E is included in the considerations, which expresses that the electrical energy is a multiple (z) of the amount

Fig. 2.100 Diagram of a potentiometric electrode



of charge corresponding to one equivalent of 96.493 coulomb/val ($= 1 F$), then, assuming $U = -E$ (U : cell voltage), the eq. (2.9) used for the special case of potentiometric sensor technology is arrived at with the standard voltage U^\ominus (voltage resulting for 1 mol/l of the species to be analyzed) and the activity of the chemical component to be determined electrochemically, a_x .

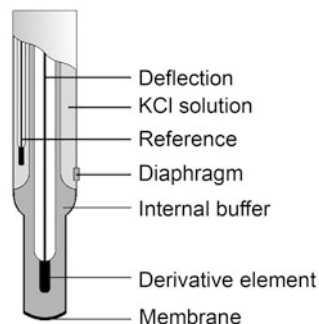
$$\Delta_R G = zFE \quad (2.8)$$

$$U = U^\ominus + \frac{RT}{zF} \ln a_x. \quad (2.9)$$

Potentiometric sensors are used for a number of analytes, both in the normal and high-temperature range, for which special versions are given in Chap. 11. *Ion analysis* is a focal point. From the schematic representation of a potentiometric electrode for this application shown in Fig. 2.100, it is clear that the ion-selective membrane is the core of this type of sensor. The potentiometric measurements must be carried out with amplifiers that have extremely high input resistances ($R_E > 10^{12} \Omega$) and the lowest possible current load ($I < 10^{-12} A$).

A distinction is made between *solid-state* and *liquid* or *gel membranes*. The former are composed of amorphous materials or poorly soluble salts in the form of mono- and polycrystals. The latter are based on organic compounds, whereby there is a close methodological relationship between liquid and gel membranes. In both cases the same ion-active solutions are used; to improve the handling, only suitable *organic high polymers* are added to form the *gel membrane*. Thus, compared to liquid membranes, a mechanically more stable membrane/sample interface is advantageously obtained, the potential of the

Fig. 2.101 Scheme of a glass electrode combination electrode consisting of an indicator electrode in a constructive unit with the (outer) reference electrode (here: Ag/AgCl, Cl^-)



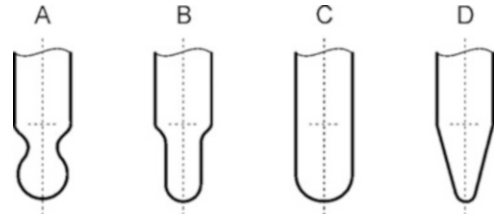
measuring electrode is less flow-dependent and electrode maintenance is easier. In contrast, the supply of ion-active substance is lower, which limits the life of the sensor. Furthermore, there is no regeneration capability of the membrane in the event of “poisoning” with interfering ions in high concentrations. Interfering *ions* generally pose a problem when measuring with ion-selective electrodes. For example, there is practically no potentiometric sensor at all that reacts completely selectively to only one single species, although there are significant differences between the individual membrane materials with regard to their cross-sensitivity to chemically similar components. This is supported by the *Nikolsky equation* (Eq. (2.10)) Calculation which represents an extension of the *Nernst equation* and describes the potential-forming influence of the activities of the ion i to be detected with the valence z and of the interfering ion j with the valence m . The quantity k_i represents the selectivity coefficient.

$$U = U^\ominus + \frac{RT}{zF} \ln a_i + k_i a_j^{\frac{z}{m}} \quad (2.10)$$

For some time now, the potentiometric sensors for the normal temperature range have been dominated by the *pH glass electrode* described for the first time in 1906 in the version shown in Fig. 2.101. For this reason, its mode of operation as a potentiometric sensor will be described in more detail. As can be seen from the schematic structure in Fig. 2.101, a thin (0.1 mm to 0.2 mm thick) oxide glass membrane serves as the sensitive element, which in practice can be designed in different geometries depending on the application purpose (Fig. 2.102). A possible glass composition includes 72 Ma % SiO_2 , 22 Ma % Na_2O and 6 Ma % CaO .

Inside the glass electrode, there is a solution with a known H_3O^+ concentration (*buffer solution*). The inner reference electrode is immersed in this solution as a discharge element; the glass electrode itself is immersed in the measuring solution. The outer reference electrode is connected to the sample solution via a diaphragm (e.g. porous ceramic, twisted platinum wires or ground glass) and completes the potentiometric electrode. Swelling layers form on the surfaces of the glass diaphragm, in which H_3O^+ ions can settle in and out. In the ideal case, if the internal buffer and measuring solution are identical, the

Fig. 2.102 Designs of glass membranes



conditions at the phase interfaces between the two swelling layers and the corresponding solutions are completely identical. For the then existing system *inner reference electrode* (B_i)/*inner buffer/source layer-glass source layer/measuring solution*//*outer reference electrode* (B_a), a potential of 0 volts should be measured in the case of a symmetrical electrode construction ($B_i = B_a$) according to the eq. (2.11) containing the *pH definition* ($pH = -\lg a_{\text{H}_3\text{O}^+}$). In reality, however, the outer and inner swelling layer are usually not completely identical. Reasons for this are, for example, the *alkali evaporation* on the outer glass membrane surface during the glass-blowing electrode manufacture, the different history of the *leaching layers*, large differences in the *osmotic pressures* between the measuring solution and the inner buffer as well as *electrochemical blockages* on glass surfaces caused by adsorbates (e.g. corrosion products of the glass). This leads to the formation of *counter potentials* and is expressed in the occurrence of an *asymmetry voltage* (U_{as}'), which is taken into account in eq. (2.12) (k : the slope of the glass electrode).

$$U = U(B_i) - U(B_a) + \frac{2.3RT}{F} pH_i - \frac{2.3RT}{F} pH_a \quad (2.11)$$

$$U = k \frac{2.3RT}{F} (pH_i - pH_a) + U_{as}' \quad (2.12)$$

For practical application, this means that *calibration* must always be carried out before the sensors are used in order to compensate for deviations from the theoretical behavior of the device.

Potentiometric sensors also include the *ion-selective field-effect transistors (ISFET)* based on the *field-effect transistor* principle. The change in the size of the space charge zone between source S and drain D, which is caused by the ion concentration of the measuring solution, is measured. Figure 2.103 shows the schematic structure.

Instead of the electrical contact at the gate, an ion-sensitive layer is applied, which in the case of pH measurement consists of Si_3N_4 , Al_2O_3 or Ta_2O_5 , for example. It comes into direct contact with the measuring solution. A voltage is applied to the reference electrode, which is also immersed in the measuring solution. The ion concentration in the measuring solution creates an additional potential (U_{GS} : gate voltage, U_{DS} : drain voltage) at the contact between the measuring solution and the ion-sensitive layer according to the Nernst equation. This influences the space charge zone between source and drain. This results in a

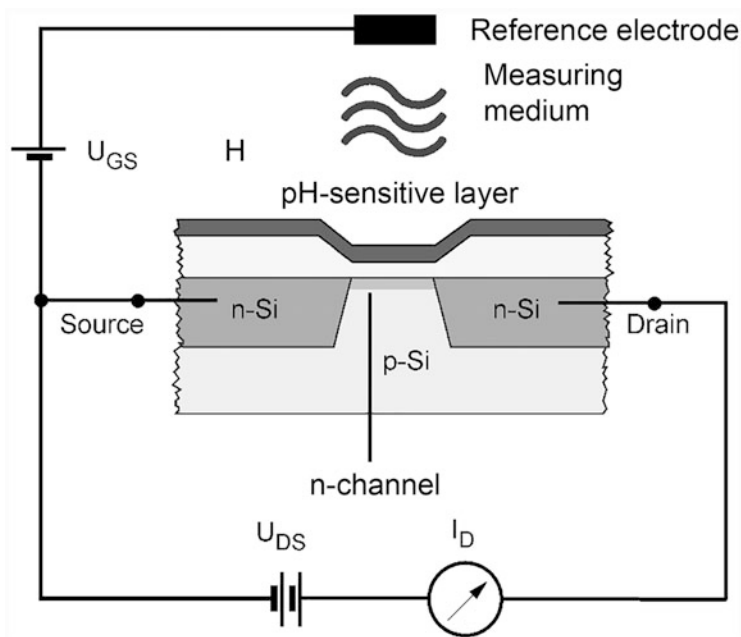


Fig. 2.103 Structure of an ion-sensitive field-effect transistor (ISFET)

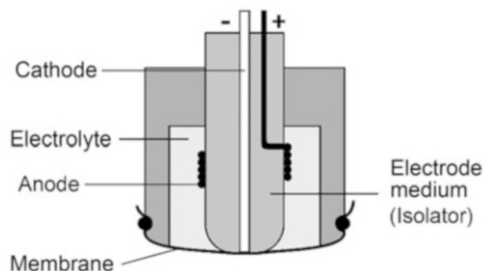
drain current I_D , which is measured and is proportional to the ion concentration of the measuring solution. The advantages of these sensors are a *small size*, the *easy changeability* of the *sensitivity*, the *mechanical stability* and the *low costs* due to the high quantities. The biochemical applications are discussed in Chap. 12.

2.16.3 Amperometric Sensors

Amperometric sensors, usually in the form of 2- or 3-electrode arrangements, measure the *current* that is generated by an electrochemical reaction of the analyte. For this purpose, a constant voltage (polarisation voltage) is applied to the sensor and this current is measured. In a 2-electrode arrangement, the polarisation voltage effective at the measuring electrode is composed of the external voltage and the voltage of the counter-electrode. In a 3-electrode arrangement, the contribution of the counter-electrode is omitted. As the polarization voltage to be realized via an amperometric measuring instrument or via a potentiostat, a value is selected that is within the range of the plateau of the diffusion limit current for the specific system. Figure 2.104 shows the scheme of an oxygen sensor.

The current is a measure of the speed of the reactions in the electrode. It is directly proportional to the *concentration* and has a *short response time* because the electrode does not have to be in thermodynamic equilibrium with the analyte. In addition, this type of

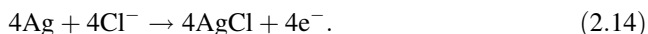
Fig. 2.104 Amperometric sensor in 2-electrode version



sensor can be *miniaturized* relatively easily and thus also cost-effectively. In principle, amperometric sensors are *not very selective*, since different electrochemical reactions can often be triggered by one and the same polarization voltage. In most cases, therefore, selectivity is achieved by two additional measures listed below:

1. Pre-connection of filter layers, which only allow certain species to reach the electrode surface in order to be converted there (*permselective membranes* or *monolayers* with *intermolecular channels*)
2. Application of *selectively* acting *catalysts* on the surface of the electrodes, which among several possible chemical reactions prefer only the desired one.

Figure 2.104 shows the schematic structure of a 2-electrode arrangement of a (membrane-covered) amperometric sensor, which was proposed by *Clark* in particular for oxygen measurement. It consists of a platinum cathode, a silver anode and a KOH solution as the electrolyte. During the chemical reaction, oxygen is reduced to OH^- at the cathode and silver ions (Eq. (2.13) and (2.14)).



The decisive factor is a *gas-permeable membrane*, which often consists of a PTFE film. This material is highly hydrophobic, so that water cannot penetrate the pores, but only the gas. The membrane separates the sensor from the measuring solution and is located directly in front of the working electrode. Figure 2.105 shows a commercial *Clark sensor*.

2.16.4 Conductometric and Impedimetric Sensors

Conductometry is concerned with the measurement of the *conductivity* of media. It is achieved by substances with ionogenic or strongly polar bonds forming cations and anions in suitable solvents through electrolytic dissociation. These are able to conduct the electric current. The conductivity is composed of the individual conductivities of all ions present in



Fig. 2.105 Clark amperometric oxygen sensor

the solution. In such cases, it is therefore not possible to identify a specific type of ion by means of conductivity measurement. The resistance of the solution is determined.

The *conductivity* or the *impedance* of a (sensor) material can also be the measured variables of probes (Taguchi or Figaro sensor), which is discussed in more detail in Sect. 11.3.4.

2.16.5 Areas of Application

Electrochemical sensors are able to measure *concentrations* of substances in *soil, water* and *air*. They, therefore, play an important role in the following fields of application:

- Environmental diagnostics,
- process monitoring and control and
- safety in the workplace.

The sensor type is also frequently used in food technology and medical technology. It is also possible in many cases to monitor the legally prescribed *workplace limits* (AGW) and the *biological limits* (BGW), which replace the former *MAK limits* (MAK: maximum workplace concentration).

As an example, Table 2.26 shows gas sensors used for this purpose, their typical measuring ranges and their areas of application.

2.17 Chemical Effects

2.17.1 Physical-chemical Interactions of Gases with Surfaces

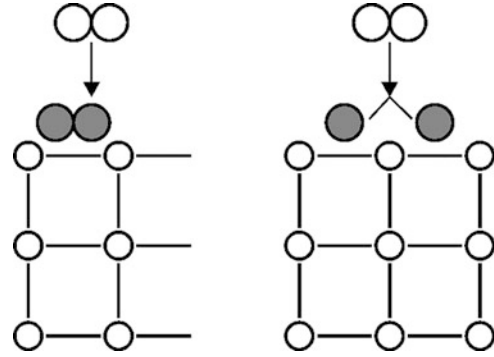
Gases *interact* considerably with condensed phases (liquids and solids). Through *diffusion* and *flow*, they move from the gas volume to the surface of condensed phases, can dissolve in them or are more or less strongly bound to their surfaces. Depending on the type of forces acting between the gas-particle and the surface, a distinction is made between *physisorption* and *chemisorption*.

Which of the mechanisms is effective depends on the type of gas and solid. For example, oxygen is mainly physisorbed on gold surfaces (Fig. 2.106, right)

Table 2.26 Important gas sensors, measuring ranges and fields of application (Source: Honold Umweltmesstechnik)

Sensor	Measuring range	Field of application
AsH ₃ (Arsine)	0 ppm to 1 ppm	AGW monitoring, semiconductor industry
C ₄ H ₁₀ (Butan)	0 ppm to 1000 ppm	AGW monitoring, paint shop
Cl ₂	0; 5 ppm to 50 ppm	AGW monitoring, gas scrubbers, water treatment plants, swimming pools, continuous monitoring at high Cl ₂ concentrations
ClO ₂	0 ppm to 1 ppm	AGW monitoring in paper production, swimming pools and wastewater treatment plants
CO	0; 500 ppm to 1000 ppm	AGW monitoring in tunnels, mines, car parks
CO ₂	0 ppm to 5000 ppm	AGW monitoring, food technology, beverage technology
COCl ₂	0 ppm to 1 ppm	Personal AGW monitoring, stationary monitoring in the chemical industry
F ₂	0 ppm to 1 ppm	Leak detection
GeH ₄	0 ppm to 50 ppm	Semiconductor industry
H ₂	0; 1% to 4%	AGW monitoring, fuel cells
H ₂ S	0; 30 ppm to 5000 ppm	AGW monitoring, soil air measurement, landfill gas monitoring, no reaction of H ₂ with unsaturated hydrocarbons at 40 °C to 60 °C
HBr	0 ppm to 30 ppm	AGW monitoring, leak detection, pharmaceutical industry
HCl	0 ppm to 30 ppm	AGW monitoring, leak detection, pharmaceutical industry
HCN	0 ppm to 30 ppm	AGW monitoring, gold mine monitoring
HF	0 ppm to 10 ppm	AGW monitoring, leak detection, semiconductor industry
NH ₃	0; 100 ppm to 1000 ppm	AGW monitoring, leak detection, cold stores, compressors
NO	0; 100 ppm to 500 ppm	AGW monitoring, medical applications
NO ₂	0 ppm to 50 ppm	AGW monitoring, medical applications
O ₂	0; 1%vol to 100%vol	AGW monitoring, soil air measurement, landfill gas monitoring
O ₃ (Ozone)	0 ppm to 1 ppm	AGW monitoring
SO ₂ (Silane)	0 ppm to 2 ppm	AGW monitoring, food technology, textile industry
SiH ₄	0 ppm to 50 ppm	AGW monitoring, semiconductor industry

Fig. 2.106 Chemisorption (left) and physisorption (right) of molecular oxygen on a solid surface



while under the same conditions at a temperature of $\theta > 500^\circ\text{C}$ on platinum, the chemical bond in the oxygen molecule is broken and oxygen is thus chemisorbed (Fig. 2.106, left):



In physisorption, relatively *weak dispersion forces*, the *van der Waals forces*, are effective, *while* in chemisorption real *chemical bonds* are effective due to *electrostatic interactions*. Sorbed (both adsorbed and chemisorbed) particles can react chemically with each other or with other gas particles; the resulting reaction products can be partially or completely desorbed. At high temperatures, a chemical reaction of the gas with ions formed from it in the solid is observed.

In the following, the individual phenomena necessary for a deeper understanding of sensor mechanisms are treated phenomenologically and the physical-chemical laws used to describe them are listed. An overview of the interactions discussed here, their classification and the laws describing them is given in Table 2.27.

2.17.2 Gas Solubility (Absorption)

Gases dissolve in liquids, but also in plastics according to the gas partial pressure. The higher the partial pressure, the higher the concentration of the dissolved gas. The *Nernst equation* in the following formulation applies for constant temperature:

$$a_{\text{xb}} = K_{\text{H}} p_{\text{gas}}, \quad (2.17)$$

Table 2.27 Interactions gas - condensed phase (MWG: the law of mass action)

Type of condensed phase	Gas-liquid		Gas - solid state		
	Volume	Reaction with liquid	Surface	Chemisorption	Volume
Spatial expansion	No reaction with liquid	Reaction with liquid	Adsorption		Solution in solid-state gas in a polymer
Description					Absorption and release of gas with defect formation
Act	HENRY'S LAW	HENRY'S LAW, MWG	LANGMUIR, FREUNDLICH-BET isotherm	LANGMUIR-FREUNDLICH- Isotherm	HENRY'S LAW MWG
Example	O ₂ /H ₂ O	CO ₂ /H ₂ O	O ₂ /Au	O/Pt	O ₂ in polyethylene
Energetic effect in kJ/mol	20 to 35		< 50	70 to 100	30 to 50 100

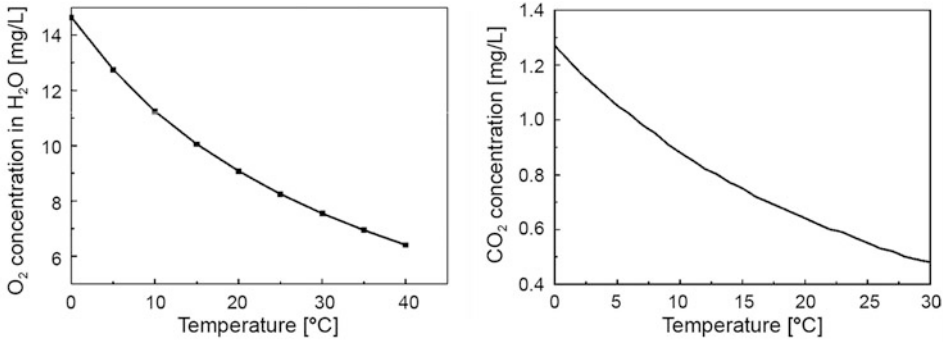


Fig. 2.107 Temperature dependence of the saturation concentration of oxygen [$p(\text{O}_2) = 21,177 \text{ Pa}$, left] and of CO_2 [$p(\text{CO}_2) = 38 \text{ Pa}$, right] in the water-air system

where a_{xb} represents the mole fraction activity ($a_{xb} = x_b f_x$). The index b stands for the solute. K_H is the *Henry constant* and f_x the concentration-dependent *activity coefficient*. In ideal solutions, in which the dissolved particles are present in high dilution, for example, permanent gases like N_2 or Ar in water, one can approximately equate the activity (the thermodynamically effective concentration) to the concentration and one obtains with $a_{xb} \gg x_b$:

$$x_b = K_H p_{\text{gas}}. \quad (2.18)$$

The mole fraction of the solute is proportional to its partial pressure in the gas phase above. For the Henry constant K_H , a constant value results at a constant temperature.

$$\frac{x_b}{p_{\text{gas}}} = \text{const} = K_H(T). \quad (2.19)$$

This solubility constant K_H is given in different units and under different conditions, for example in $\text{cm}^3 \text{g}^{-1} \text{atm}^{-1}$ or mg l^{-1} . The solubility of permanent gases is low and decreases with increasing temperature. The saturation concentration of oxygen in water, for example, is $9.08 \text{ mg O}_2/\text{l}$ at ambient air and $20 \text{ }^\circ\text{C}$ under normal pressure ($101,325 \text{ Pa}$). The temperature dependence of the saturation concentration of atmospheric oxygen in water is shown in Fig. 2.107, left.

In contrast, easily soluble gases such as the anhydrides CO_2 , SO_2 , H_2S or NH_3 react chemically with water. The CO_2 concentration is only 1/10 of the oxygen concentration, although the oxygen partial pressure is about 557 times ($21,177 \text{ Pa}/38 \text{ Pa}$) higher than that of CO_2 . When CO_2 is dissolved, acid-base equilibria can occur, which are described by the law of mass action (MWG). Henry's law applies only to a limited extent in this case. Using the example of the system $\text{CO}_2 - \text{H}_2\text{O}$ one obtains

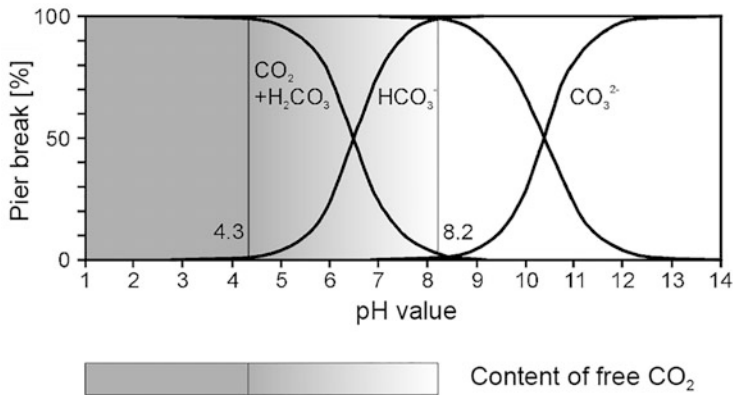
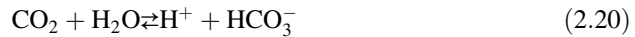


Fig. 2.108 Free and bound CO_2 in water as a function of pH value



$$K_1 = \frac{[\text{H}^+][\text{HCO}_3^-]}{[\text{CO}_2]} \quad (2.22)$$

$$K_2 = \frac{[\text{H}^+][\text{CO}_3^{2-}]}{[\text{HCO}_3^-]} \quad (2.23)$$

The equilibrium constants and their negative decadic logarithms (pK values) are for $K_1 = 4.3 \cdot 10^{-7} \text{ mol/l}$ ($\text{p}K_1 = 6.36$) and for $K_2 = 5.61 \cdot 10^{-11} \text{ mol/l}$ ($\text{p}K_2 = 10.25$) at 25°C . By means of the quantitative treatment of the equilibria, it can be shown that unbound “free” CO_2 in solution is only present at a pH value < 4.3 (Fig. 2.108).

In these cases, the indication of the gas concentration only makes sense together with the pH value. It is easy to imagine that CO_2 can react in real systems such as surface and seawater with a variety of naturally occurring acid-base systems, which then act as sinks or sources for this gas. Therefore, the interpretation of measurements of dissolved CO_2 is difficult. Surface water usually contains less than 10 mg/l of CO_2 , which comes from the atmosphere and biological processes. Groundwater can have concentrations of up to 100 mg/l CO_2 . For the temperature dependence of the saturation concentration of air CO_2 in the water a falling curve is obtained (Fig. 2.107, right).

2.17.3 Gas Transport to the Solid Surface

To understand gas reactions at the surfaces of solids, knowledge of the laws of gas transport from a gas volume to the surface and through porous bodies is necessary. Transport equations can generally be formulated as follows:

$$\text{Current} = \text{Empirical coefficient} \cdot \text{area} \cdot \text{driving force}.$$

The following applies in the case of material flow:

$$\text{Material flow} = \text{Empirical coefficient} \cdot \text{area} \cdot \text{gradient of the driving force}.$$

The diffusion is described by *Fick's laws*. For the stationary case, for a temporally and spatially constant gradient the *first Fick's law* applies:

$$j_b = -\frac{dn_b}{dt} - D_b \cdot A \cdot \text{grad}\mu_b. \quad (2.24)$$

The material flow j_A is determined by the gradient of the chemical potential of the substance b. D is the diffusion coefficient. The negative sign results from the fact that mass transport is opposite to the gradient, that is, degrades it. Since the chemical potential depends on the partial pressure or the concentration (exactly the activity) of the respective substance,

$$\mu_b = \mu_b^0 + RT \ln p_b \text{ or} \quad (2.25)$$

$$\mu_b = \mu_b^0 + RT \ln c_b, \quad (2.26)$$

you can also use the gradient of the partial pressure or concentration and obtain for the case of diffusion in one spatial direction:

$$j_b = -D_b \cdot A \cdot \frac{dp_b}{dx}. \quad (2.27)$$

Approximately, the differential quantities can be replaced by differences:

$$j_b = -D_b \cdot A \cdot \frac{\Delta p_b}{l}. \quad (2.28)$$

A linear and constant partial pressure difference in the diffusion length l is assumed. If, on the other hand, diffusion takes place in all three spatial directions, then

$$\mathbf{j} = -D\nabla c = \text{grad}c$$

$\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$ ist he Nabla operator.

For most diffusion processes the gradient is neither constant in time nor space. For this approach the *second Fick's law* applies:

$$\frac{dc}{dt} = D \frac{d^2c}{dx^2}. \quad (2.29)$$

In this, the change of the gradient (second derivation) is taken into account. As a result of mass transport, the difference in concentration is reduced more and more; the gradient is, therefore, smaller and smaller. The *second Fick's law* is only analytically not always easy to solve under certain boundary conditions.

For practical concerns, however, one tries to create conditions which are to be dealt with by the first Fick's law. Of importance is the calculation of the *diffusion current* at an electrode through which a current flows, assuming that every electrochemical species that reaches the electrode is also converted. Then one can write for the *limit current density* $i_{d, \text{lim}}$, which is independent of the voltage in a wide range:

$$i_{d, \text{lim}} = -z_b D_b \frac{c_b}{\delta}. \quad (2.30)$$

Here z_b is the number of charges converted, D_b is the diffusion coefficient, c_b is the concentration in the solution and δ is the diffusion length within which the concentration decreases linearly Eq. (2.30).

2.17.4 Adsorption and Chemisorption

To describe adsorption on surfaces, a dynamic process between the molecules that are adsorbed and those that are desorbed is assumed. In equilibrium, the speeds of adsorption and desorption are equal. How much is adsorbed on the surface depends on the partial pressure. At low partial pressures, a proportional increase is initially observed, which at higher pressures, when a monomolecular adsorption layer increasingly forms, leads to a pressure-independent range. This behaviour is expressed by the *Langmuir adsorption isotherm*:

$$\theta = \frac{v_B}{v_m} = \frac{k_1 p_b}{1 + k_2 p_b}. \quad (2.31)$$

θ is the degree of coverage - the ratio of the adsorbed gas volume to the gas volume necessary for the formation of total coverage. If the pressure is small compared to 1, that is,

$k_2 p_b < 1$; then the denominator becomes 1 and the function describes the linear increase of the surface concentration with the pressure. If, on the other hand, $k_2 p_b > 1$ applies, the 1 in the denominator can be neglected and a constant value independent of pressure is obtained. The behaviour between the two limit states is only insufficiently represented by the Langmuir isotherm. The isotherm according to the *Freundlich equation* (2.27) reflects just this range well.

$$v = kp^{1/n} \quad n > 1. \quad (2.32)$$

For the description of physisorption in a multilayer, a condensate, the isotherm according to *Brunauer, Emmet and Teller (Bet)* is used:

$$\frac{p}{v(p_0 - p)} = \frac{1}{v_m c} + \frac{c - 1}{v_m c} \frac{p}{p_0}. \quad (2.33)$$

Where p_0 is the vapour pressure of the liquid adsorbent and p is the vapour pressure at equilibrium. This isotherm is the basis of the very frequently used instruments for determining the surface area of powders (*BET method*).

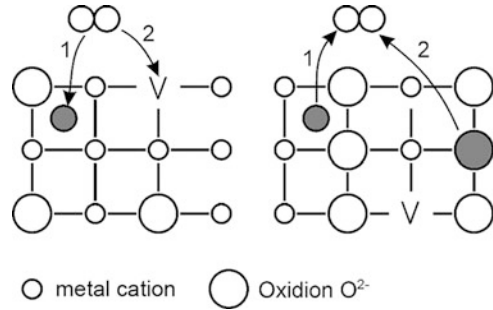
2.17.5 Reactions with Adsorbed Species

Chemical reactions can also take place between the physisorbed and chemisorbed ponds on the solid surface and gases, depending on the catalytic activity and temperature. With the help of kinetic investigations, *two mechanisms* can be distinguished. If *both reaction partners* are first adsorbed on the surface and the reaction then takes place between them, this is called the *Langmuir-Hinshelwood mechanism*. If, on the other hand, the reaction partner is a *gas* that reacts with an adsorbed particle, this is called the *Eley-Rideal mechanism*. These considerations play a major role in understanding the processes at hot sensor surfaces of, for example, semiconductor gas sensors operated in a multi-gas component system.

2.17.6 Reaction of the Gas with the Solid

In addition to the pure physical solution, gases can also react chemically with the solid and change the composition of the surface and the volume of the solid selectively while maintaining the crystal structure. This results in defects compared to the ordered lattice, which considerably influence the electrical properties of the solid. These phenomena occur at higher temperatures ($\vartheta > 450 \text{ }^\circ\text{C}$) and can lead to chemical equilibria between gas and solid.

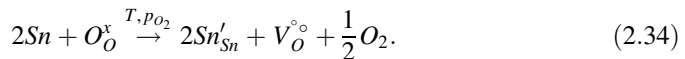
Fig. 2.109 Installation (left) and removal (right) of gaseous oxygen (schematic). Installation and removal to or from interstitial site 1 and to or from regular oxide sites 2



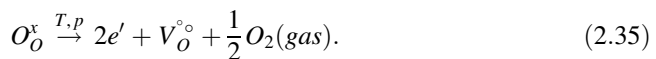
In order to understand the functioning of solid-state sensors, a short introduction to the disorder of solids is useful. With *disorder models*, the electrical properties, but also the diffusion in and through the solid can be explained. Point disorder plays the most important role.

As an example, the interaction between oxide and the oxygen from the surrounding gas phase is described in more detail.

If the oxygen partial pressure above the oxide is small, oxygen is removed from the solid with the formation of excess electrons e' (Fig. 2.109, right). In the other case, oxygen incorporation can be observed with the formation of defect electrons h (from holes) (Fig. 2.109, left). In tin oxide, the tin oxide is partially reduced at temperatures of $>450\text{ }^\circ\text{C}$ and low oxygen partial pressure. While maintaining the SnO_2 lattice (rutile structure), only a little oxygen is expanded and excess electrons are produced, which can be locally imagined as Sn^{3+} . According to a relative notation introduced by *Kröger* and *Vink*, one can write for also $\text{Sn}_{\text{Sn}^{4+}}^{3+} = \text{Sn}_{\text{Sn}'} e'$. The index is the ion that was originally intended to be on this lattice site, here Sn , and the exponent is the charge. The reduced Sn^{3+} has one less positive charge than the original Sn^{4+} . Relative to the undisturbed oxide, this represents a negative charge. Due to the oxygen build-up, oxygen or oxide ion vacancies are formed, which are called V_{O} (*vacancies*). Since two negative charges are missing, these are positively charged twice ($V_{\text{O}}^{\circ\circ}$).

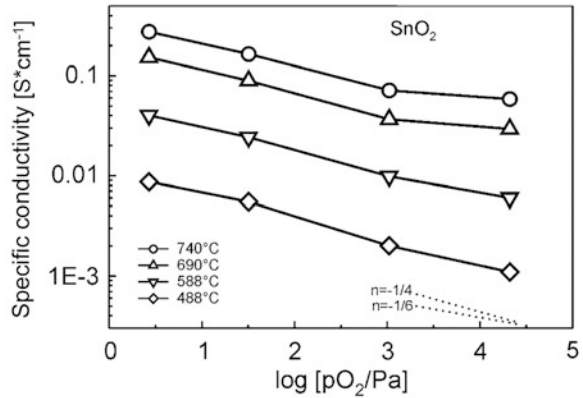


Reduced only to the oxide ions $\text{O}^{2-} = \text{O}_O^x$ (x means: no change in charge compared to the normal state; x is usually omitted) can be formulated:



With the law of mass action you get:

Fig. 2.110 Dependence of the electrical conductivity of SnO_2 on the oxygen partial pressure (Prof. Dr. U. Guth, KSI Meinsberg)



$$K_{p,c,x} = \frac{[e']^2 [V_{O^\circ}^\circ] p_{O_2}^{1/2}}{x_{O_2}}. \quad (2.36)$$

If the mole fraction for the normal oxide ions is considered to be constant (the oxygen evolution at 1000 K is only about 0.001) and the concentration of the oxide ion vacancies is also constant, then one obtains

$$[e'] \propto \sigma \propto \text{const} \cdot p_{O_2}^{-1/4}. \quad (2.37)$$

In the case of variable vacancy concentration, an equivalence consideration of the charge carriers leads to the fact that the concentration of the oxide ion vacancies is half as large as that of the excess electrons

$$[V_{O^\circ}^\circ] = \frac{[e']}{2} \text{ ist, zu}$$

$$[e'] \propto \sigma \propto \text{const} \cdot p_{O_2}^{-1/6}. \quad (2.38)$$

The electrical conductivity σ is the product of the elementary charge e , the concentration of charge carriers c and their mobility u :

$$\sigma = e c u. \quad (2.39)$$

The electrical conductivity, which is proportional to the concentration of excess electrons, thus decreases with the fourth or sixth root of oxygen or, the smaller the oxygen partial pressure, the greater the conductivity. This is in accordance with the idea of partial reduction of SnO_2 . Experimentally both dependencies are found, as can be seen in Fig. 2.110.

Conversely, the incorporation of oxygen into an oxide, for example at interstitial sites (symbol i), produces defect electrons h° (from *holes*):



$$K_{p,c,x} = \frac{[h^\circ]^2 [O_O^\circ]}{xV_i p_{O_2}^{1/2}}. \quad (2.41)$$

Since concentrations of free interstitial sites practically do not change, at a constant concentration of oxide ions on the interstitial sites one obtains

$$[h^\circ] \propto \sigma \propto \text{const} \cdot p_{O_2}^{\frac{1}{4}}. \quad (2.42)$$

If the latter are not constant, the charge equivalence $\frac{[h^\circ]}{2} = [O_i'']$ results in:

$$[h^\circ] \propto \sigma \propto \text{const} \cdot p_{O_2}^{\frac{1}{6}}. \quad (2.43)$$

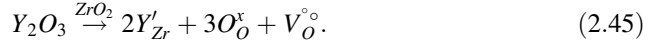
The generation of electronic defects can also be understood with the help of the band model, assuming delocalized excess or defect electrons. Apart from the electrical effects described here, lattice defects cause distortions in their immediate vicinity, which can be explained by a changed ion radius and the effect of the changed charge on the neighbouring ions.

2.17.7 The Mixed-phase Disorder

In addition to the disordered structures discussed here, which arise due to the thermodynamic conditions of temperature and partial pressure and are therefore often referred to as *intrinsic disordered structures*, those generated by the specific selection of the compositions of mixed phases are also of great importance. The *mixed-phase disorder* is also called *an extrinsic disorder* because it is generated from the outside by the choice of the chemical composition.

This type of disorder will be explained in more detail using the example of stabilized ZrO_2 , which is used as a solid electrolyte in *lambda sensors*. If an oxide of a bivalent or trivalent metal such as CaO, MgO or Y_2O_3 is added to the zirconium oxide and this mixture is annealed at temperatures $>1200^\circ\text{C}$, the bivalent or trivalent metal ion is incorporated into the solid in place of the tetravalent Zr ion within the limits of solubility. The Ca^{2+} or

trivalent Y^{3+} ion, which is divalent compared to the tetravalent Zr^{4+} ion, does not require two oxide ions ($2O^{2-}$), but in the case of Ca^{2+} one and for Y^{3+} only $3/2$ oxide ions ($3/2 O^{2-}$). O^{2-} positions, therefore, remain unoccupied in the case of the same lattice structure. The incorporation equations which lead to the chemical disorder described here can be formulated as follows:



Ca''_{Zr} means: A divalent Ca^{2+} ion sits on a Zr^{4+} site. Since the former has only two positive charges, that is, has two less than the Zr^{4+} , two positive charges are missing. So this defect is formally doubly negatively charged.

Chemically, the composition of such phases can be determined by the formulas

$$(ZrO_2)_{0.84}(CaO)_{0.16} = Zr_{0.84}Ca_{0.16}O_{1.84} \text{ and} \quad (2.46)$$

$$(ZrO_2)_{0.84}(YO_{1.5})_{0.16} = Zr_{0.84}Y_{0.16}O_{1.92}. \quad (2.47)$$

In both cases the cation lattice is complete ($0.84 + 0.16 = 1$), the anion (oxygen ion) lattice, however, is incomplete. In the first case, the amount of defects is $2 - 1.84 = 0.16$ mol and in the second $2 - 1.92 = 0.08$ mol. Overall, electro-neutrality must be guaranteed, that is, the charge balance must contain the same number of positive and negative charges. Because these admixtures make ZrO_2 , which undergoes several phase transitions (monoclinic-tetragonal-cubic) when heated in pure form, phase stable (cubic fluorite phase) over a wide temperature range (from room temperature to $1500^\circ C$), these mixed phases are also known as *stabilized zirconium oxide* (abbreviated: YSZ for *yttria-stabilized zirconia*; the suffix -ia means oxide of the respective metal).

The conductivity of stabilized ZrO_2 is proportional to the concentration of the oxygen vacancies and the mobility of the oxide ions over these vacancies:

$$\sigma = zF c_{V_O^{\circ\circ}} u_{O_O}. \quad (2.48)$$

Since mobility depends exponentially on temperature, the electrical (ionic) conductivity can also be represented by an exponential function:

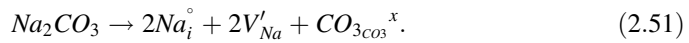
$$\sigma_{ion} = \text{Const} \cdot \exp \left[-\frac{E_a}{RT} \right]. \quad (2.49)$$

With the determined values for ZrO_2 with doping of 8 mol- % Y_2O_3 is obtained:

$$\sigma_{\text{ion}} = 1.63 \cdot 10^2 \exp \left[-\frac{0.79eV}{kT} \right]. \quad (2.50)$$

At 1000 °C, the ionic conductivity is thus about $0.1 \text{ Ohm}^{-1} \text{ cm}^{-1} = 0.1 \text{ S/cm}$ and is thus in the order of magnitude of the conductivity of diluted inorganic acids at 25 °C.

Intrinsic disorder, in contrast, is caused by *entropic effects*. With increasing temperature, particles leave their lattice sites and migrate to interstitial sites (*Frenkel's disorder*) or the surface of the solid (*Schottky's disorder*). For example, solid Na_2CO_3 conducts electricity at temperatures $>300 \text{ °C}$ because sodium ions occupy interstitial sites.



At higher temperatures ($> 350 \text{ °C}$), solid carbonates are solid electrolytes with which galvanic CO_2 sensors can be constructed.

2.18 Acoustic Effects

2.18.1 Definition and Classification of Sound

Sound is the propagation of *local pressure fluctuations*. Unlike electromagnetic waves, these can only propagate in the matter and are *longitudinal waves* (speed of sound is in the direction of the pressure fluctuations). Depending on the sound frequency, the sound ranges are classified as follows:

- *Infrasound*: inaudible range up to 16 Hz,
- *Audible sound*: from 16 Hz to 20 kHz,
- *Ultrasound*: 20 kHz to 1 GHz and
- *Hypersonic*: from 1 GHz.

2.18.2 Characterization of Acoustic Waves

The acoustic waves propagate at the speed of sound c . Their wavelength λ results from the product of the frequency f and the speed of sound c so that the following applies

$$\lambda = c \cdot f.$$

The sound waves can be described as *periodic changes in the velocity and pressure of the particles*. Particle velocity u_T and pressure p_T are composed of the static part u_S, p_S and the periodically changing part u and p . u is also called *sound velocity*. Therefore applies:

$$u_T = u_S + u \text{ and}$$

$$p_T = p_S + p.$$

The sound velocity $u(x,t)$ and the periodic pressure changes $p(x,t)$ can be described by harmonic waves as a function of the location x and the time t as follows

$$u(x,t) = u_0 \sin(kx - \omega t) \text{ and}$$

$$p(x,t) = p_0 \sin(kx - \omega t)$$

with u_0 : Amplitude of the sound particle velocity

p_0 : pressure amplitude

ω : angular frequency: $\omega = 2\pi / T = 2\pi f$

k : wave number: $k = 2\pi / \lambda = \omega / c$

T : period duration: $T = 1 / f$.

The quotient of sound pressure and sound velocity is the *acoustic impedance* Z_a . It is a constant in a plane wave:

$$Z_a = p/u = \rho c.$$

2.18.3 Sound Velocity in Ideal Gases

Distance measurement with ultrasound is based on the *transit time measurement* of a sound pulse. Exact knowledge of the speed of sound is indispensable for the measurements. The speed of sound in ideal gases depends on

- the temperature T ,
- the relative air humidity and
- pressure (it is not dependent on static pressure).

2.18.3.1 Dependence on Temperature

For the speed of sound for ideal gases applies

$$c = c_0(1 + T/273.15^\circ C)^{1/2}.$$

The speed of sound c is proportional to the *root of the absolute temperature* T . The speed of the sound of air according to this equation is: 331.5 m/s. Approximately applies then:

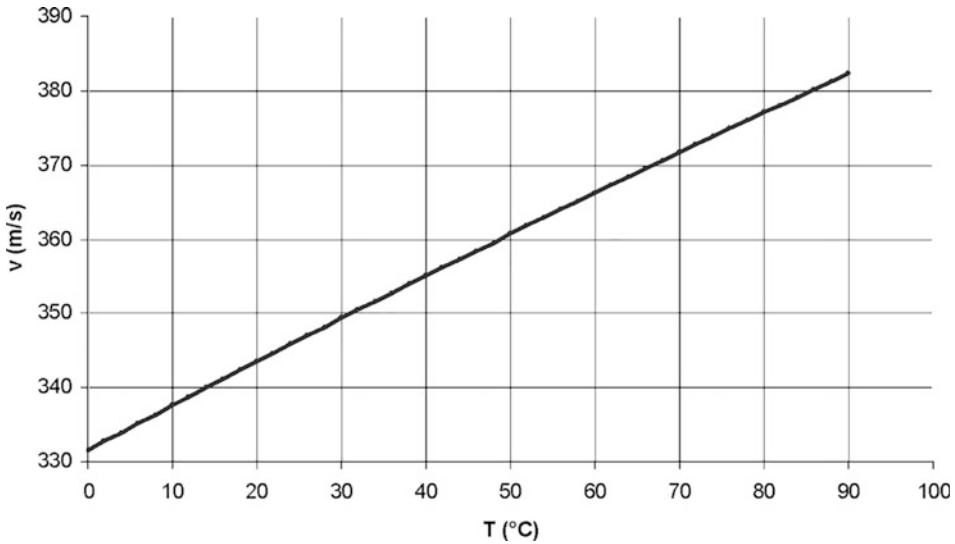


Fig. 2.111 Dependence of sound velocity on temperature for air

$$c \approx (331.5 + 0.6 \delta / ^\circ\text{C}) \text{ m/s} \quad (\text{with } \delta \text{ as temperature in } ^\circ\text{C} : \delta = T - 273.15 \text{ K}).$$

Figure 2.111 shows the relationship.

Table 2.28 lists the sound velocities (calculated and measured from the formula) of other gases.

2.18.3.2 Dependence on the Relative Air Humidity

The speed of sound *increases* with increasing air humidity. This dependence is shown in Fig. 2.112.

2.18.3.3 Dependence on Pressure

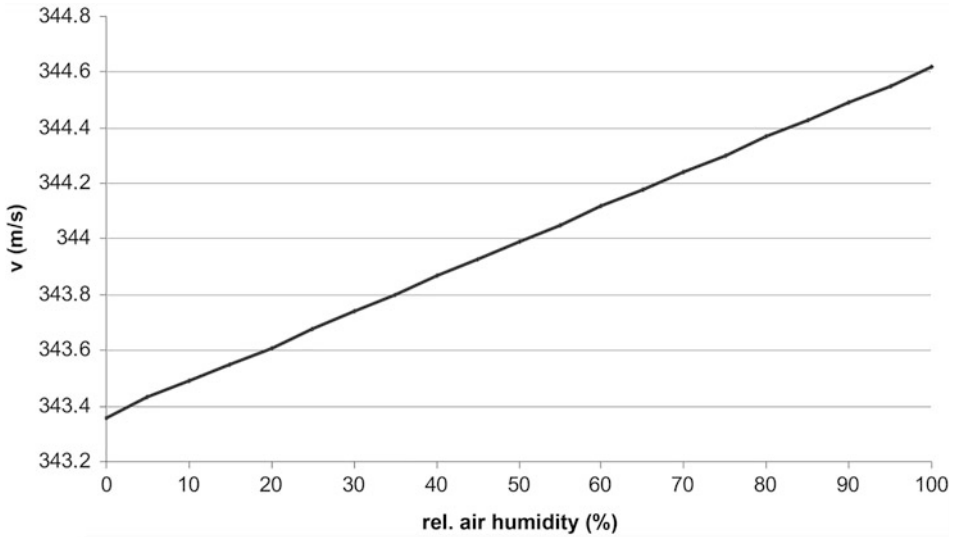
The speed of sound in ideal gases is *not* dependent on static pressure. The air pressure p and the density ρ are proportional to each other at the same temperature: The quotient p/ρ is a constant.

2.18.4 Intensity or Sound Intensity

The intensity I is the average power $\langle P \rangle$ radiated onto a surface A perpendicular to the direction of propagation of the wave. Then applies:

Table 2.28 Sound velocity in different gases

Gas	c at 20 °C in m/s	c measurements in m/s
Air	344	343.2
Hydrogen	1307	1280
Oxygen	308	316

**Fig. 2.112** Dependence of the speed of sound on the relative humidity at 20 °C

$$I = \langle P \rangle / A.$$

Further calculations result:

$$I = \frac{1}{2} p_0^2 / Z_A = \frac{1}{2} p_0 u_0 = \frac{1}{2} Z_A u_0^2$$

with p_0 : Amplitude of the pressure

Z_A : acoustic impedance.

2.18.5 Sound Absorption in Air

The intensity I of a sound wave decreases with the distance x travelled. This *attenuation* depends on the sound frequency used. The *higher* the frequencies of the sound waves are,

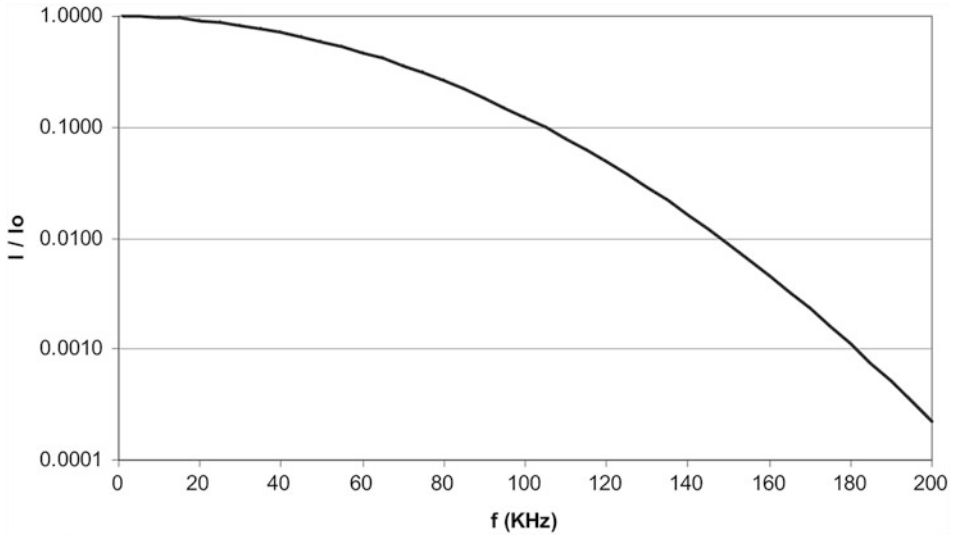


Fig. 2.113 Dependence of the intensity I on the frequency over a distance of one meter

the *more* their range is *shortened*. The following applies to the intensity I as a function of the distance x of the sound source

$$I(x) = I_0 \cdot e^{-\alpha x} = I_0 \cdot \exp(-\alpha f^2 x).$$

The absorption coefficient α is proportional to the square of the sound frequency f^2 , so that

$$\alpha = a f^2 \quad \text{with the proportionality factor } a.$$

Figure 2.113 shows this behavior as a function of frequency over a distance of one meter. The intensity I at the sound source has the value: I_0 , the intensity at a distance of one meter I .

The damping of sound in air is strongly dependent on the *air pressure*. With higher air pressure, the sound absorption decreases and the range of the sensors increases.

2.18.6 Reflection and Transmission

When a sound wave propagating in the air hits a boundary layer to another medium, part of the wave passes through this layer into the other medium (is *transmitted*) and the other part is *reflected*. It applies:

$$u_r = r u \text{ and}$$

$$u_t = t u.$$

with u_r : Sound velocity of the reflected wave

u_t : sound velocity of the sound wave in the other medium

r: reflection factor

t: transmission factor

Since no energy and thus intensity can be lost, the following must apply:

$$r + t = 1.$$

The following applies to the reflection factor r:

$$r = (Z_L - Z_M)/(Z_L + Z_M).$$

Thus the ratio of the intensity of the reflected sound wave I_r to the sound wave I impinging on the interface is valid:

$$I_r/I = ((Z_L - Z_M)/(Z_L + Z_M))^2.$$

Table 2.29 shows acoustic impedances and the proportion of the intensity of the reflected waves from air hitting the interface calculated from them. Table 2.29 shows that - except in air - almost the entire intensity is reflected at the interfaces in water and solids. In these cases, the proportion that passes through the interface into the other medium is negligible.

2.19 Optical Effects

2.19.1 Physical Effects

Optics deals with the properties of light as perceived by the eye. On the one hand, light can be understood as an *electromagnetic wave* and on the other hand, it can be interpreted as a *particle stream of photons*. Figure 2.114 shows the classification of the ranges of electromagnetic waves.

Visible light is a visible electromagnetic wave in the wavelength range from $\lambda = 380$ nm to $\lambda = 780$ nm. This corresponds to a frequency range from $f = 3.84 \cdot 10^{14}$ Hz to $f = 7.89 \cdot 10^{14}$ Hz. Light propagates in a vacuum at a constant speed, the *speed of light* c

Table 2.29 Intensity of reflected sound waves

Medium	Acoustic Impedance Z_a kg /s m ²	Reflected intensity I_r Air - Medium
Air 20 °C	$4.142 \cdot 10^2$	0.000%
Water 20°	$1.484 \cdot 10^6$	99.888%
Wood Spruce	$2.860 \cdot 10^6$	99.942%
Aluminium	$1.377 \cdot 10^7$	99.988%
Iron	$3.866 \cdot 10^7$	99.996%
Piezoceramics	$1.865 \cdot 10^7$	99.991%

($c = 299.792458 \cdot 10^6$ m/s, about 300,000 km/s). For the relationship between the speed of light c , the wavelength λ and the frequency f applies:

$$c = \lambda \cdot f.$$

$$c = \lambda \cdot f.$$

Electromagnetic radiation is subdivided into the following further areas according to Fig. 2.114:

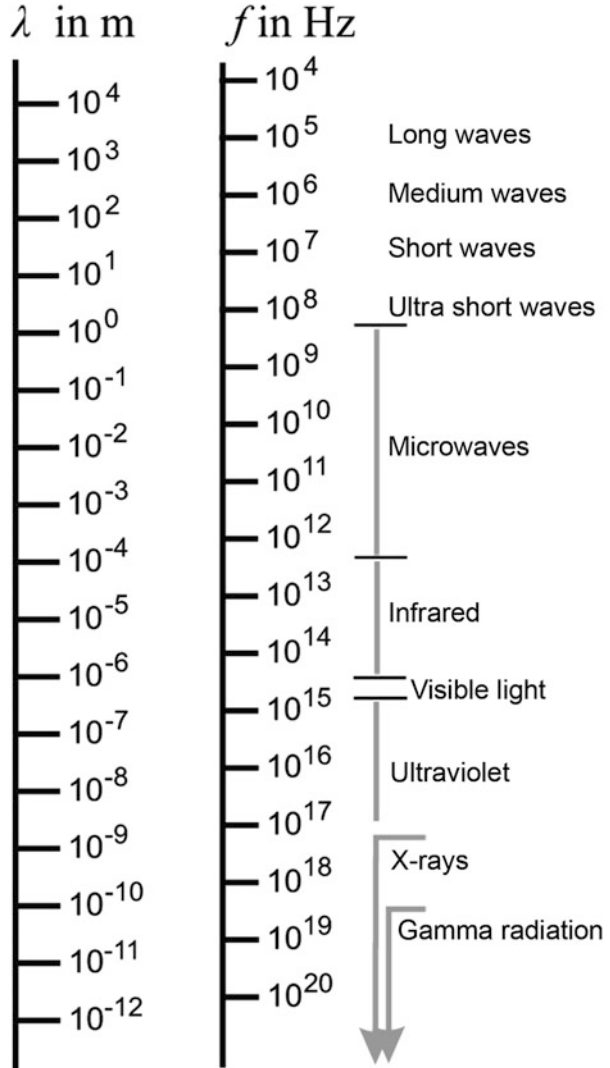
- Wavelengths smaller than visible light (or higher frequencies)
 - *UV light* (ultraviolet light: between $f = 10^{15}$ Hz and $f = 10^{17}$ Hz),
 - *X-ray radiation* (frequencies $f > 10^{17}$ Hz) and
 - *γ radiation* (frequency $f > 10^{19}$ Hz).
- Wavelengths greater than visible light (or lower frequencies)
 - *Infrared* (thermal radiation: frequencies f to 10^{13} Hz),
 - *Microwave radiation* (frequencies f up to 10^9 Hz) and
 - *Radio frequencies* (frequencies f from 10^9 Hz to $f = 10^5$ Hz).

Figure 2.115 shows the properties for light as a *wave* and light as a particle stream (*photon stream*).

An electromagnetic wave can be *absorbed* or *reflected*. When absorbed, the light disappears completely and darkness reigns. The light beam is *reflected* on a *reflecting surface*. The incident beam, the incidence plumb line (surface normal at the point of impact) and the reflected beam are in one plane and the following applies: *angle of incidence = angle of reflection*. This property is an important principle in *geometric optics* and explains the formation of images on curved surfaces (e.g. parabolic mirrors). If the angle of incidence is greater than 90°, the light is totally reflected, that is, the incident beam is held in this medium. This is the principle of *optical fibers* (OF). It consists of many glass fibres that allow a large angle of incidence. The light is totally reflected up to 20,000 times per meter of glass fiber and is thus transported through the glass fiber.

Figure 2.116 shows the effect of *refraction*.

Fig. 2.114 Classification of electromagnetic radiation (source: Hering, Martin, Stohrer: Physics for engineers, 13th edition, Springer)



If a light beam hits an interface between an optically thinner and an optically denser medium at an angle ϵ , there is not only a reflecting beam (dotted with dashes in Fig. 2.116) with the same angle of incidence ϵ but also a *refracted* beam that includes an angle β to the perpendicular. The following applies:

$$\sin(\epsilon) / \sin(\beta) = \text{konstant.}$$

The speed of light c' in the matter is always lower than the speed of light in vacuum c . The law of refraction is then

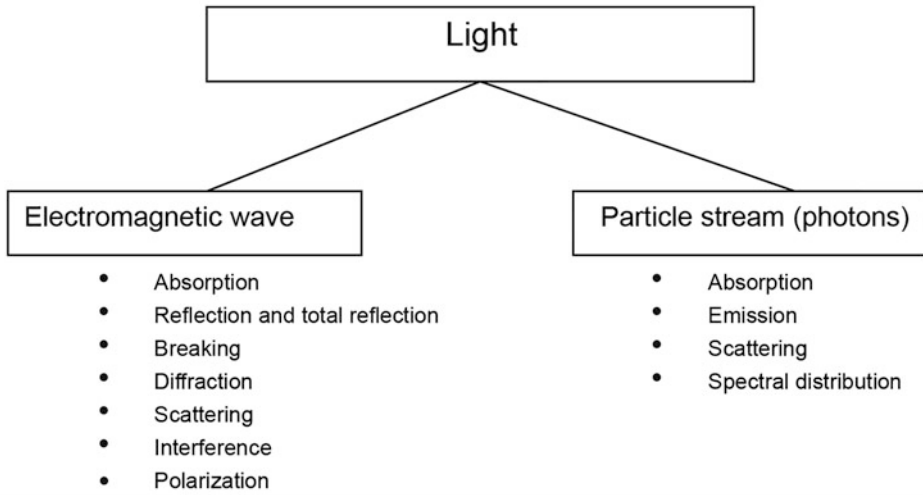
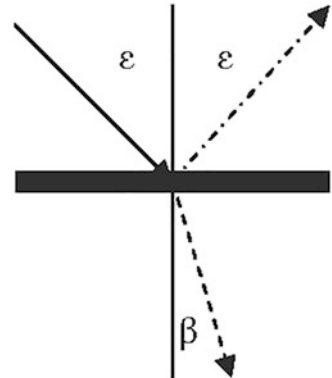


Fig. 2.115 Properties of light as waves and particles

Fig. 2.116 Refraction of a light beam



$$\sin(\varepsilon) / \sin(\beta) = c/c'$$

This means that the ratio of the sine values of angle of incidence and angle of refraction is equal to the ratio of the two speeds of light. The *refractive index* n of a material is the quotient of the speed of light in a vacuum c_0 and the speed of light c in the material concerned. It applies:

$$n = c_0/c.$$

In the above equation, the *Snellius' law of refraction* is given:

Table 2.30 Refractive indices of gaseous, liquid and solid materials

Material	Refractive index n
Air	1.0003
Carbon dioxide (CO ₂)	1.0045
Water	1.333
Ethyl alcohol	1.362
Benzene	1.501
Ice	1.310
Quartz glass	1.459
Flint glass F 3	1.613
Barium oxide	1.980
Diamond	2.417

$$\sin(\epsilon) / \sin(\beta) = n' / n.$$

The refractive index depends on the wavelength (colour) of the light and the temperature. Table 2.30 shows the refractive indices of some materials.

The refraction of light is the basis for all optical images that are realized with lenses. The instruments used for this are mainly cameras, telescopes and microscopes.

The wave character of light is also evident in the *interference*. If two waves (e.g. sine waves) oscillate in the *same phase*, the wave troughs and wave crests are *amplified*. If the two waves are *out of phase* by 180°, the wave troughs and wave crests *cancel* each other out. In this way, patterns of light and dark stripes are created, the *interference patterns*. From the distances between the patterns, *interferometers* can be used to determine physical quantities (e.g. atomic or lattice spacing, angle, length and refractive index). In Sect. 3.1.2.8 the interferometric length measurement is explained in detail. Interference patterns also play an important role in optical encoders (Sect. 3.2.1).

Light is an *electromagnetic wave*. The *electric field strength* E and the *magnetic field strength* H propagate at the speed of light. The two field vectors are perpendicular to each other. In natural light, short wave trains propagate in different directions. Since every direction of oscillation is represented in the statistical average, no direction dependence of the oscillation can be determined. For example, a *polarizer* can be used to force the electric field strength to oscillate in only one direction. If you place an *analyzer* perpendicular to it, no light passes through it. Polarized light can be used to study material surfaces and their properties (e.g. grain sizes of certain inclusions can be determined).

As Fig. 2.115 shows, light can also be described as a particle stream, the *photon stream*. Each photon has an energy of $E_{\text{ph}} = h \cdot f$. Here, h is the *Planckian quantum of action* ($h = 6.626 \cdot 10^{-34}$ Js) and f the frequency. If a photon beam hits a material surface, either *electrons* can be *released* (*external photoelectric effect*) or not. In this case, the photon energy is absorbed and the electron states in the material are energetically raised (*internal*

Table 2.31 Work function and limiting wavelength of some materials

Material	Leaving work W_A in eV	Limiting wavelength λ_{gr} in nm
Cs ₃ Sb	0.45	2.755
GaAs-Cs	0.55	2.254
Cs	2.14	2.254
Ru	2.16	574
Ka	2.30	539.1
Na	2.75	450.9
Li	2.90	427.5

photoelectric effect). Which effect occurs depends on the *work function* W_A . The following relationships apply:

- *External photoelectric effect*: $h \cdot f \geq W_A$ (the photon energy is greater than the work function).
- *Internal photoelectric effect*: $h \cdot f < W_A$ (the photon energy is smaller than the work function).

Table 2.31 shows the magnitude of the work function W_A and the corresponding *limiting wavelengths* λ_{gr} of some materials.

Section 2.14 deals in detail with the interrelationships and their applications in sensor technology, in particular the mode of operation of photoelements such as photodiode, phototransistor and photothyristor (Table 2.22 and Sect. 14.4.3). The use of semiconductor laser diodes for distance and displacement measurement is explained in detail in Sect. 3.1.2. A particularly important application of a diode is the determination of position (PSD: Position-Sensitive Device, Sect. 3.1.2.2, Fig. 3.52).

2.19.2 Design of Optical Sensors

The optical sensors basically consist of a *transmitter* and a *receiver*. Both are connected to each other via a *transmission medium*. Table 2.32 shows the different possibilities in this area.

Lasers are playing an increasingly important role as transmitters. Figure 2.117 shows the mode of operation of a laser. Figure 2.117 a) shows how a photon with the energy $E_{ph} = h \cdot f$ is absorbed. This lifts an electron from energy state E_1 to energy state E_2 . In *spontaneous* emission, an electron falls from the energy state E_2 to the energy state E_1 . Thereby the photon energy $E_{ph} = h \cdot f$ is released. In *stimulated* or *induced* emission, a photon with the energy $E_{ph} = h \cdot f$ stimulates an electron to a transition from E_2 to E_1 . Thereby a photon is generated, which amplifies the first one.

A laser works according to this principle (LASER: Light Amplification by Stimulated Emission of Radiation). In the real laser, three or four energy levels are involved, as

Table 2.32 Design of optical sensors

Transmitter (light sources)	Transmission medium	Recipient
White light (sun, light bulb)	Vacuum	Photodiode (point diode, area diode, quadrant diode)
Laser	Air	Phototransistor
Light-emitting diodes (LED)	Fluid	Photthyristor
	Solid State	CMOS image sensors
	Glass	CCD array (lines, matrix)
	Fibre optic cable	

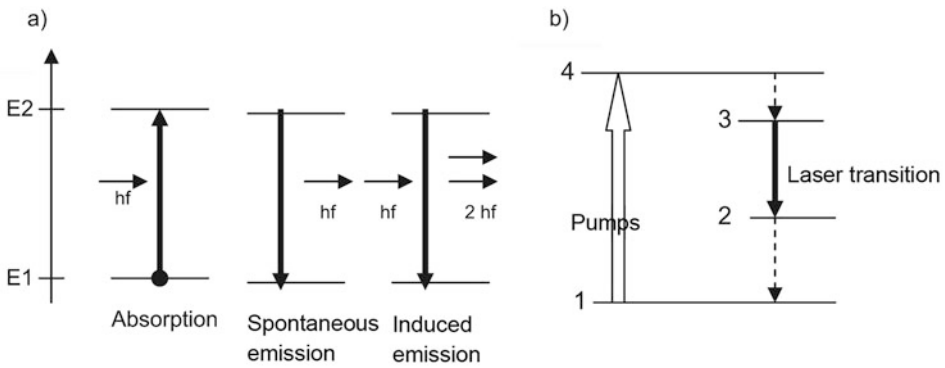
**Fig. 2.117** Interaction between photons and energy states (a) absorption and emission; (b) laser principle

Fig. 2.117 b) shows. The dotted lines are transitions without radiation. The laser has the property that it is a light source with *coherent radiation*. In order for the laser to shine continuously, photons must always be pumped up.

Figure 2.118 shows some important laser types. In logistics, lasers play an important role, especially in automatic identification (e.g. with barcodes or laser scanners) (Sects. 3.4.5 and 3.4.6).

The photoelectric effect, as an important physical effect that converts optical signals into electrical signals, has already been discussed in Sect. 2.19.1. Important detectors of optical signals are the *CMOS image sensors* and the *CCDs* (CCD: Charged Coupled Device).

In CMOS image sensors, the photodiodes are arranged in a row or a matrix and can be addressed as *pixels*. There, corresponding information can be read in or out (Sects. 14.4 and 14.5).

The charge distribution on a surface can be imaged and read out with the CCD sensors (Sect. 14.5).

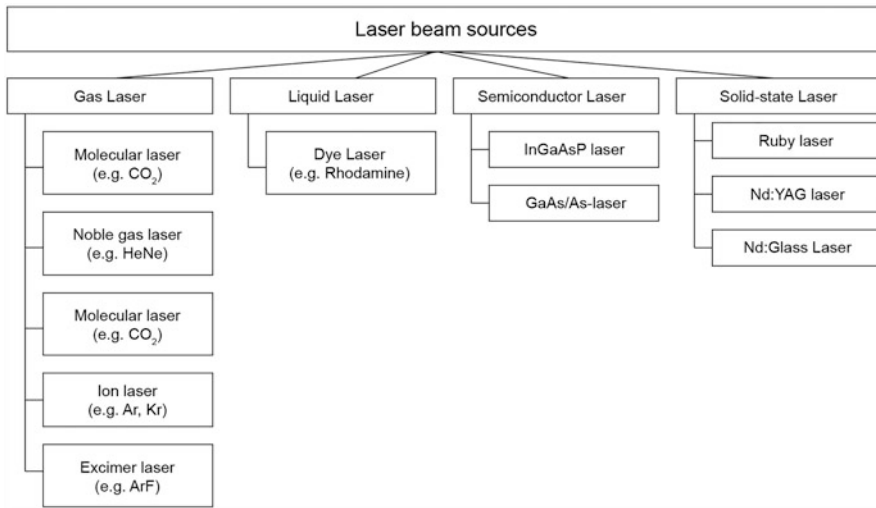


Fig. 2.118 Common laser types (Source: Hering, Martin: Photonic)

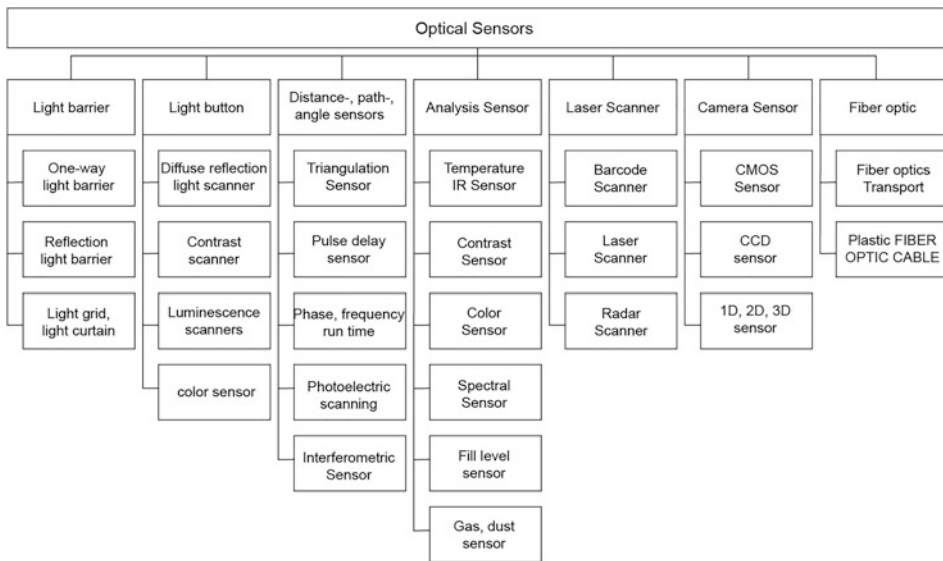


Fig. 2.119 Classification of optical sensors

2.19.3 Categories of Optical Sensors

Figure 2.119 shows a classification of the optical sensors.

The most important sensor types are:

- *Through-beam photoelectric sensor*: The transmitter and receiver are separate. A light beam is emitted and usually registered by a photodiode. The photoelectric sensor switches when the light beam is interrupted by an object (Sect. 2.14).
- *Retro-reflective photoelectric sensor*: transmitter and receiver are in one housing. The reflected beam is detected. The ranges are shorter than with through-beam photoelectric sensors. On the other hand, these photoelectric sensors are much more cost-effective due to the lower cabling and adjustment effort (Sect. 2.14).
- *Light grid, light curtain*: In a light grid, several through-beam photoelectric sensors are arranged in a row. All transmitters and all receivers are housed in a common housing (Sect. 2.14).
- *Light sensor*: The light beam is not reflected by a reflector, but by an object. The beam reflected by this object is measured and evaluated. *Luminescence scanners* are an important field of application. Luminescent dyes can be applied in adhesives, chemical substances or on solids (e.g. as labels). These markings are not visible to the naked eye. When irradiated with UV light, the sensor detects the marking. In the case of high-quality branded articles, such an identification can be applied to detect the authenticity of the objects.
- *Distance, displacement and angle sensors*: The functional principles and applications of these sensors are described in detail in Sects. 3.1 and 3.2.
- *Analysis sensor*: Optical sensors can be used to solve many measurement and identification tasks. A large field of application is the measurement of temperatures and temperature fields by *infrared radiation* (Chap. 6). *Spectral sensors* can be used to detect harmful chemicals or dust. This is particularly important in *environmental analysis*. The analysis sensors are also used for color recognition or as level indicators (Sect. 2.14, Fig. 2.81).
- *Laser scanners* are used to identify objects (Sect. 3.6).
- *Camera sensors*: They are used to capture and evaluate image information and are based on CMOS or CCD technology (Sects. 14.4 and 14.5). These sensors are also able to identify and measure positions (1D), surfaces (2D) or spatially extended objects (3D) (Sects. 3.4.5 and 3.4.6).
- *Fiber Optics*: The optical information can be transported with glass fiber optic cables (FO) or plastic fiber optic cables to the place where it can be evaluated easily and safely.

2.19.4 Application Fields of Optical Sensors

The fields of application of optical sensors are very diverse, as Fig. 2.120 shows. They play a major role above all in safety technology, in the field of quality assurance, logistics, measurement technology as well as chemical and environmental analysis.

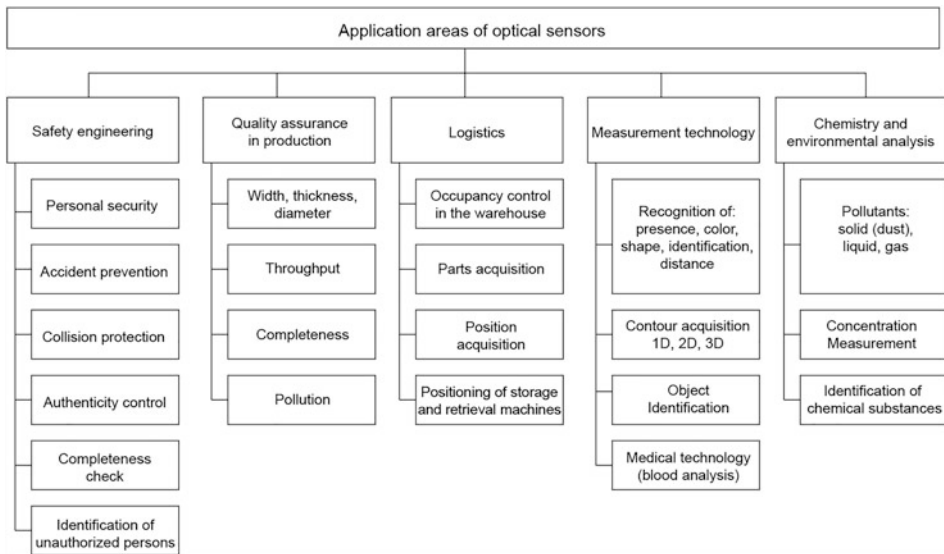


Fig. 2.120 Application areas of optical sensors

2.20 Doppler Effect

2.20.1 Principle of Operation and Physical Description

If a source that emits a wave with the frequency f_Q moves relative to an observer, the observer receives a frequency f_B that is different from f_Q . This *frequency shift* is called the *Doppler effect*. The following motion constellations exist:

- Observer moves, source rests,
- Source moves, the observer rests and
- Observer and source are moving.

2.20.1.1 Observer Moves, Source Rests

Figure 2.121 a shows the relationships. If an observer moves towards the source at the velocity v_B , he perceives the successive wavelike condensations in faster succession. The time interval T_B of the impinging waves is shortened. It applies:

$$T_B = \lambda / (c + v_B)$$

with λ : Wavelength of the wave,

c : Speed of the wave,

v_B : Speed of the observer towards the source.

The observer then perceives the frequency f_B :

$$f_B = (c + v_B)/\lambda.$$

Since: $c = \lambda \cdot f_Q$, the viewer will perceive the following *higher* frequency f_B :

$$f_B = f_Q(1 + v_B/c).$$

If the observer moves away from the source at the velocity v_B , the frequency decreases towards the source:

$$f_B = f_Q(1 - v_B/c).$$

2.20.1.2 Source Moves, Observer Rests

Figure 2.121 b) shows how the wave field of a sound source moving to the right changes. The distance between the wave crests becomes smaller on the front side (the wavelength λ_B becomes smaller and larger on the backside.

If the wave flows towards the observer, then the following applies:

$$f_B = f_Q/(1 - v_Q/c).$$

If the wave moves away from the observer, the following applies:

$$f_B = f_Q/(1 + v_Q/c).$$

2.20.1.3 Observer and Source Move

If both the observer and the source are moving, there are several possibilities of frequency shifting, depending on the direction of movement (Table 2.33). The formulas given only apply if the observer moves *radially* towards or away from the source. If the movement takes place on a circle concentric to the source, no Doppler effect takes place. For arbitrary movements, the corresponding *radial components* of speed v_B must be inserted into the above formulas.

The Doppler frequency f_B can generally be formulated as follows:

Table 2.33 Frequency shifts

Source	Observer	Observed frequency
← •	← ← •	$f_B = f_Q \left(1 + \frac{v_B}{c}\right)$
← •	← • →	$f_B = f_Q \left(1 - \frac{v_B}{c}\right)$
← • →	← •	$f_B = \frac{f_Q}{1 - \frac{v_Q}{c}}$
← ← •	← •	$f_B = \frac{f_Q}{1 + \frac{v_Q}{c}}$
← • →	← ← •	$f_B = f_Q \frac{c + v_B}{c - v_Q}$
← ← •	← • →	$f_B = f_Q \frac{c - v_B}{c + v_Q}$
← ← •	← ← •	$f_B = f_Q \frac{c + v_B}{c + v_Q}$
← • →	← • →	$f_B = f_Q \frac{c - v_B}{c - v_Q}$

$$f_B = f_Q \left(\frac{c \pm v_B}{c \mp v_Q} \right)$$

It is v_B the velocity of the observer and v_Q the velocity of the sound wave. The upper sign applies to the movement towards each other and the lower sign to the distance between the observer and the transmitter.

The above formula shows that there is no Doppler effect when $v_B = -v_Q$. This is the case when observer and transmitter move in the same direction at the same speed relative to the medium, or when the medium is moving but observer and transmitter are at rest. This is the case when a wind is blowing. No Doppler effect occurs.

2.20.1.4 Doppler Effect of Light (Doppler Effect Without Medium)

The propagation of light as an *electromagnetic wave* requires *no transmission medium*, so the above formulas do not apply. Only the *relative velocities* v of observer and source are decisive for the Doppler frequency shift.

In the case of an approximation:

$$f_B = f_Q \sqrt{\frac{c + v}{c - v}}$$

Here c is the vacuum speed of light ($c = 2.99792458 \cdot 10^8$ m/s). If the observer and the source move away, the following applies

$$f_B = f_Q \sqrt{\frac{c - v}{c + v}}$$

If the observer and source move at an angle α to each other, the frequency for the observer is f_B :

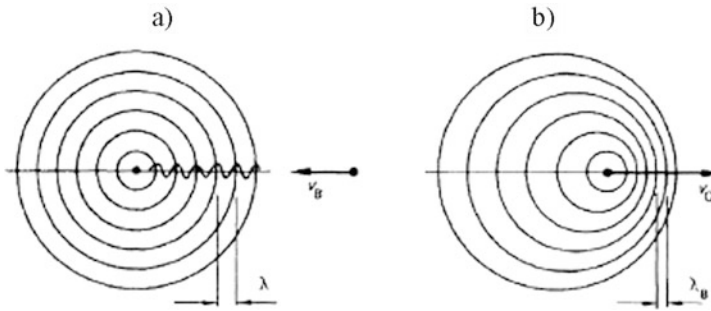


Fig. 2.121 Principle of the Doppler effect (a) Stationary source, moving observer (b) Moving source, stationary observer (Source: Hering, Martin, Stohrer: Physics for engineers, 13 edition, Springer)

$$f_B = f_Q \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{v}{c} \cos \alpha}$$

If one sets for $\alpha = 0^\circ$, one gets the above formulas for the *longitudinal* Doppler effect. For $\alpha = 90^\circ$, $\cos \alpha = 0$, so that the following formula applies to the *transverse* Doppler effect:

$$f_B = f_Q \sqrt{1 - \frac{v^2}{c^2}}$$

The transverse Doppler effect can be neglected at speeds far below the speed of light c .

2.20.2 Application Areas

The Doppler effect can be observed in echoes of acoustic or electromagnetic signals. Figure 2.122 shows an overview of the applications. They are briefly described below.

2.20.2.1 Astronomy

The Doppler effect is observed in astronomy as *redshift*. There are two types:

- *Gravitational redshift*

According to Einstein's general and special theory of relativity, *gravity* influences the transmission frequency. If the light is emitted away from the centre of gravity, that is, upwards, a *lower* frequency is measured. In other words, the *time interval* between the beginning and the end of the signal becomes longer. This describes the *time dilation* in the

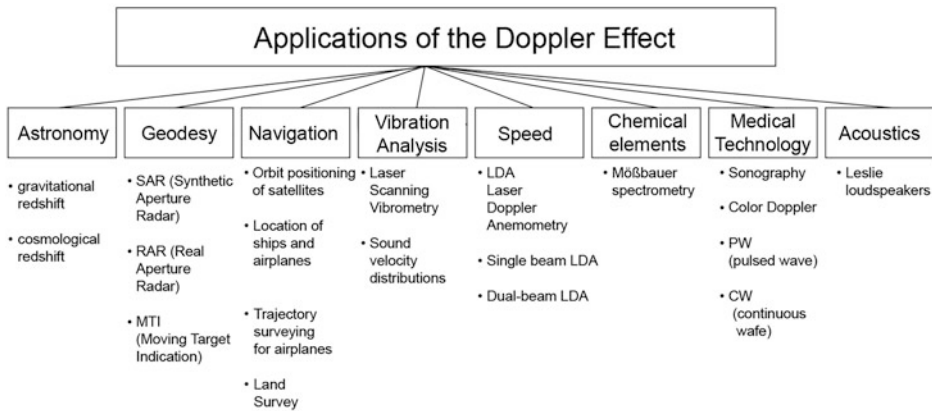


Fig. 2.122 Overview of the applications of the Doppler effect

theory of relativity. This has consequences for clocks. The time on satellites and the time on earth are different. This must be taken into account accordingly.

- *Cosmological redshift*

The light of galaxies is shifted to red in most cases. This means that the galaxies are *moving away from the earth*, that is, the universe is expanding. The speed of the expansion can be calculated using the Doppler effect.

2.20.2.2 Geodesy (Land Surveying)

Land surveying from airplanes, space shuttles or satellites is done by *imaging radar measurements*. A distinction is made between *SAR (Synthetic Aperture Radar)* and *RAR (Real Aperture Radar)*. Both radars have the following advantages over optical measurements: They are independent of day and night as well as of weather influences such as fog, rain or snow. The SAR is preferable to the RAR because it has a higher resolution.

In SAR, an antenna whose position is known moves perpendicular to the direction of the beam. In the *synthetic aperture*, many small antennas replace a large antenna. The target area is illuminated from different angles. From the echo values of the Doppler effect, images are generated from which an image of the landscape can be calculated.

With the Doppler effect of the radar beams, even moving objects in the landscape can be detected and their direction of movement and speed determined. This method is called *MIT (Moving Target Indicator)*.

2.20.2.3 Navigation

Using so-called *Doppler navigation satellites*, the locations can be determined by other satellites, aircraft and ships. The accuracy is within a few meters. They can also be used as landing aids in airports or for land surveying.

2.20.2.4 Vibration Analysis

A laser beam strikes an oscillating area. The backscattered light exhibits a Doppler effect. From the Doppler frequency, the path and speed of the oscillations can be calculated *without contact*. So-called *laser Doppler vibrometers* are used. The oscillation conditions can be analyzed both in microstructures and in fast running processes. These devices are used in production and quality assurance, but also for vibration analysis in mechanical engineering, automotive and aircraft construction.

2.20.2.5 Speed Measurement

Components of the velocities of particles in liquids and gases can be measured contact-free by *LDA (Laser Doppler Anemometry)*. For this purpose, the laser beam is split into two partial beams. A particle that moves generates scattered light and thus a Doppler effect. If both beams are rejoined, a scattered light signal is generated. According to the Doppler effect, its frequency is proportional to its speed. If three laser Doppler devices are used, all three spatial components of velocity can be determined. *Single beam laser Doppler systems* and *double beam laser Doppler systems* are used. The single-beam systems are used in medical technology (e.g. for determining the velocity of the blood flow), whereas only double-beam systems are used for flow measurement technology. These do not need to be calibrated because they are based on differential measurement.

2.20.2.6 Determination of Chemical Elements

In addition to the Mößbauer effect, *Mößbauer spectroscopy* also uses the Doppler effect. A gamma emitter is moved mechanically and thus a *hyper-fine* modulation is generated. With this modulated gamma radiation it is possible to analyze the elements contained in a sample and their electronic properties. Water on Mars was detected with such a miniature probe.

2.20.2.7 Medical Technology

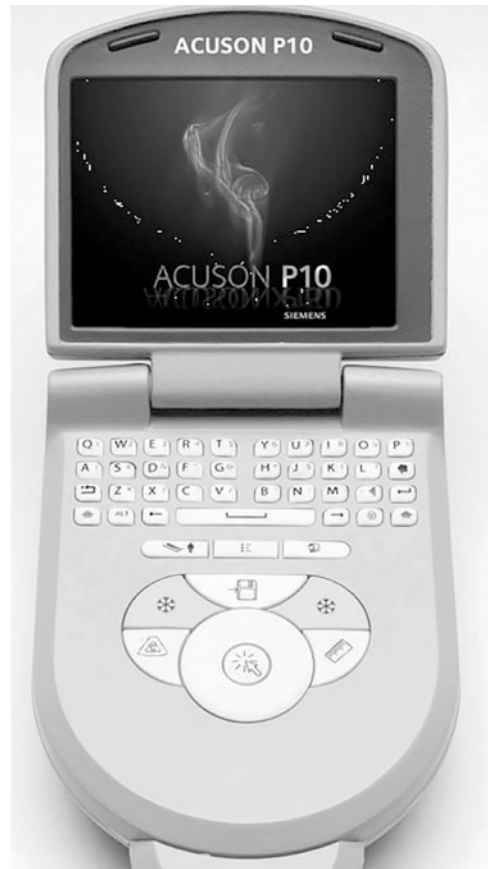
In medical technology, the Doppler effect in the ultrasonic range plays an important role in *sonography*. The flow of blood can be measured with a *colour Doppler*. The colour “red” indicates that the blood is moving towards the sound probe. The color “blue” indicates that the blood is moving away from the probe. With *PW-Doppler* (PW: pulse wave) the sound probe works *alternately* as transmitter and receiver. It is used for vascular examinations. Figure 2.123 shows a portable sonography device.

A CW Doppler (CW: Continuous Wave) has a transducer that acts as both a transmitter and receiver. It is mainly used for cardiac examinations (e.g. examination of the function of the heart valves) because it can also measure high speeds.

2.20.2.8 Acoustics

Moving speakers (*Leslie speakers*) produce a *modulation of the pitch* according to the Doppler effect. The ear perceives this beat as *vibrato*. In a Hammond organ, such rotating Leslie loudspeakers are built into the organ. They give the instrument its unmistakable sound.

Fig. 2.123 Portable ultrasound sonograph (Factory photograph: SIEMENS)



Bibliography

1. Albrecht, H. E.; Borys, M.; Damaschke, N.; Tropea, C.: Laser Doppler and Phase Doppler Measurement Techniques, Springer Verlag, 2003
2. Ashcroft, N. W.; Mermin, N. D.: Solid State Physics, International Edition, Saunders College, Philadelphia, 1976 (Kap. 1, S. 24–25 und Kap. 13, S. 253–261)
3. Atkins, P. W.; de Paula, J.; Bär, M.; Schleitzer, A.: Pyhsikalische Chemie, 4. Auflage Wiley-VCH-Verlag 2006
4. Ayadi-Mießén, A.: Doppler-Effekte in UWB-basierten Rotortelemetrie-Systeme, Leibniz Universität Hannover, 2011
5. Baumann, J.: Einspritzmengenkorrektur in Common Rail Systemen mit Hilfe magnetoelastischer Drucksensoren, KIT Scientific Publishing 2006
6. Bergveld, P.: IEEE Trans. Biomed. Eng. **19** (1972) 340
7. Bouwens, R. J. et al.: A candidate redshift $z \approx 10$ galaxy and rapid changes in that population at an age of 500 Myr. In: Nature 469, S. 504–507, 2011
8. Bucher-Gruppe: Geoinformatik, 2010

9. Camman, C.; Galster, H.: Das Arbeiten mit ionenselektiven Elektroden, Springer Verlag, Berlin 1995
10. Chefki, M.: Untersuchungen von Eu-Verbindungen mit ungewöhnlichen Valenzzuständen mit Hilfe der 151Eu-Hochdruck-Mößbauerspektroskopie, Shaker Verlag, 1998 Cremer, M.: Z. Biol. **47** (1906) 562
11. DIN EN 60584-1 Oktober. 1996 Thermopaare – Teil 1: Grundwerte der Thermospannungen
12. Engel, T.; Reid, P.: Physikalische Chemie, Pearson Studium Verlag, 2009
13. Galaxienentfernungsrekord erneut gebrochen. SpektrumDirekt, 26. Januar 2011
14. Gevatter, H.-J.: Automatisierungstechnik 1. Mess- und Regeltechnik, Springer Verlag, 2000
15. Gunther B.; Hansen K. H.; Veith I.: Technische Akustik – Ausgewählte Kapitel, Grundlagen, aktuelle Probleme und Messtechnik, 8. Auflage, Expert Verlag, Renningen 2010
16. Hering, E.; Gutekunst, J.; Martin, R.; Kempkes, J.: Elektrotechnik und Elektronik für Maschinenbauer, Springer Verlag, 8. Auflage 2021
17. Hering, E.; Martin, R.: Photonik, Grundlagen, Technologie und Anwendung, Springer Verlag 2005
18. Hering, E.; Martin, R.; Stohrer, M.: Physik für Ingenieure, Springer Verlag, 13. Auflage 2021
19. Hesse, S.; Schnell, G.: Sensoren für die Prozess- und Fabrikautomation: Funktion – Ausführung – Anwendung, 4. Auflage, Vieweg+Teubner Verlag, Wiesbaden 2009
20. Hilleringmann, U.: Mikrosystemtechnik: Prozessschritte, Technologien, Anwendungen, Vieweg +Teubner Verlag, Wiesbaden 2006
21. Hofer, M.: Sono Grundkurs, Thieme Verlag, 2009
22. Hoffmann, J.: Taschenbuch der Messtechnik, Carl Hanser Verlag, 2010
23. Honold, F.; Honold, B.: Ionenselektive Elektroden, Birkhäuser Verlag, Basel-Boston-Berlin 1991
24. Hütter, L. A.: Wasser und Wasseruntersuchung, Frankfurt am Main, Otto Salle Verlag, 1990, S. 92
25. Janocha, H.: Unkonventionelle Aktoren: Eine Einführung, Oldenbourg Verlag 2010
26. Krenn, H.: Magnetische Nanostrukturen. Institut für Experimentalphysik, Universität Graz, 2003
27. Langmann, R.: Taschenbuch der Automatisierung, Carl Hanser Verlag, 2010
28. Lauber, R.; Göhner, P.: Prozessautomatisierung, Springer Verlag, 1999
29. Lehnert, M. D. et al.: Spectroscopic confirmation of a galaxy at redshift $z = 8.6$ In: Nature **467**, S. 940–942, 2010
30. Lerch R.; Sessler G.; Wolf, D.: Technische Akustik Grundlagen und Anwendungen, Springer Berlin 2009
31. Liess, M.; Steffes, H.: The modulation of thermoelectric power by chemisorption – A new detection principle for microchip chemical sensors, Journal of the Electrochemical Society (2000), **147**, (8), pp. 3151–3153
32. Maier, J.: Festkörper – Fehler und Funktionen, Teubner Verlag, Stuttgart und Leipzig 2000
33. Mc Innes, D. A.; Dole, M.: Ind. Engng. Chem. Analyt. Edn. **1** (1929) 57
34. Meyer E., Neumann E.-G.: Physikalische und Technische Akustik. Vieweg Verlag, Braunschweig 1967
35. Möller, C.: Magentosresistive Sensoren zur Bewegungssteuerung und -kontrolle, TTN-Jahrestagung 2007
36. Möser, M.: Technische Akustik, 6. erweiterte und aktualisierte Auflage. Springer Verlag, Berlin 2005
37. Pedrotti, F.; Pedrotti, L.; Bausch, W.; Schmidt, H.: Optik für Ingenieure, Springer Verlag, 2. Auflage 2002
38. Pelster1, R.; Pieper, R; Hüttl, I.: Thermospannungen – viel genutzt und fast immer falsch erklärt. In: Physik und Didaktik in Schule und Hochschule 1/4 (2005), S. 10–22

39. Raith, W.: Bermann-Schaefer, Bd. 2, Elektromagnetismus, 8. Auflage, Walther de Gruyter, 1999 (Kap. 8.3.2)
40. Reif, K.: Automobilelektronik: Eine Einführung für Ingenieure. Vieweg+Teubner Verlag, Wiesbaden 2008
41. Reif, K.: Sensoren im Kraftfahrzeug, 1. Auflage, Vieweg+Teubner Verlag, Wiesbaden 2010
42. Rühl, J.: Nanotechnologie als Basis präziser und robuster Sensorik für Automobil und Automation, Lust Antriebstechnik 2007
43. Schanz, G. W.: Sensoren: Sensortechnik für Praktiker, Hüthig Verlag, 5. Auflage 2004
44. Schrüfer, E.: Elektrische Messtechnik: Messung elektrischer und nichtelektrischer Größen, Carl Hanser Verlag, 2007
45. Stimson, G. W.: Introduction to Airborne Radar; Hughes Aircraft Comp., El Segundo, California, 1983
46. Turk: Induktive Sensoren, Turk Industrial Automation, Firmenbroschüre S1015, 2006
47. US 2913 386 (1959)
48. Wecker, J.; Kinder, R.; Richter, R.: Sandwiches mit riesigem Magnetwiderstand, Physik in unserer Zeit, Heft Nr. 5, 2002
49. Wedler, G.: Lehrbuch der physikalischen Chemie, Wiley-VCH Verlag 2004



Geometric Quantities

3

Ekbert Hering, Gert Schönfelder, Stefan Basler, Karl-Ernst Biehl, Thomas Burkhardt, Thomas Engel, Albert Feinäugle, Sorin Fericean, Alexander Forkl, Carsten Giebeler, Bernhard Hahn, Ernst Halder, Christopher Herfort, Stefan Hubrich, Jürgen Reichenbach, Michael Röbel, and Stefan Sester

E. Hering (✉)

Hochschule Aalen (Rektor i. R.), Aalen, Germany

e-mail: Ekbert.Hering@hs-aalen.de

G. Schönfelder

Prignitz Mikrosystemtechnik, Wittenberge, Germany

S. Basler

ehemals SICK STEGMANN GmbH, Donaueschingen, Germany

K.-E. Biehl · C. Herfort

PIL Sensoren GmbH, Erlensee, Germany; CHerfort@pil.de

T. Burkhardt · A. Feinäugle · S. Fericean · A. Forkl · B. Hahn

Balluff GmbH, Neuhausen auf den Fildern, Germany

e-mail: thomas.burkhardt@balluff.de; Albert.Feinaeugle@balluff.de; Sorin.Fericean@balluff.de; alexander.forkl@balluff.de; bernhard.hahn@balluff.de

T. Engel

Carl Zeiss IMT GmbH, Oberkochen, Germany

e-mail: kontakt@ennovare.de

C. Giebeler

Pyreos Ltd, Edinburgh, UK

e-mail: carsten.giebeler@pyreos.com

E. Halder · S. Hubrich · S. Sester

Novotechnik Messaufnehmer OHG, Ostfildern, Germany

e-mail: halder@novotechnik.de; sester@novotechnik.de

J. Reichenbach

SICK AG, Freiburg im Breisgau, Germany

e-mail: juergen.reichenbach@sick.de

M. Röbel

Deutz AG, Köln, Germany

e-mail: info@roebel.com

3.1 Displacement and Distance Sensors

In the following sections, displacement and distance sensors with different physical operating principles are presented. *Displacement sensors* differ from *distance sensors* in that they have a *position-determining element*.

This can be, for example, a magnet that is firmly connected to the moving part and whose linear displacement is to be measured. *Distance sensors*, on the other hand, measure the *distance* between the *active surface* of the sensor and a *target*, which can be of any nature with certain restrictions. This is called a *non-cooperative target*. Examples can be found in workpiece handling by robots. By measuring the distance to the part to be gripped, cycle times can be drastically reduced. Exact knowledge of the distance allows the robot to approach at maximum speed and to brake selectively just before the part.

If a *cooperative target* can be measured, this is usually associated with an increase in performance. For example, the range of an optical distance sensor can be significantly increased if the measurement does not have to be made on an object of any nature, but can be made on a reflector with defined and optimized reflection properties. During the measurement, the reflector *cooperates* with the sensor to a certain extent, from which the term *cooperative target* is derived. A similar situation arises in the case of distance measurement using microwaves. Here, too, the measurement can be performed on any object or on a correspondingly shaped reflecting body with optimized reflection properties.

Table 3.1 gives an overview of the characteristic values of the sensors that are usually used to select a suitable sensor principle. This table shows typical values for measuring sensors. It is based on commercially available sensors, which account for about 80% of the market.

There is a wide variety of measuring products on the market, from simple sensors to complex, expensive measuring devices. Table 3.2 lists the classification of the accuracy classes.

3.1.1 Inductive Distance and Displacement Sensors

3.1.1.1 Functional Principle and Morphological Description of Inductive Sensors

In terms of the general definition of the sensor, the sensor element of inductive sensors consists of at least one reactive component (classic example: the *coil*).

The *reactance* of this component is influenced by a position sensor (*target*) in inductive sensors (IS) for path, distance or position detection. The result is a detectable change in either a main parameter of the sensor element (*inductance* L , *loss resistance*—in the case of a coil) or a derived variable (usually: *coil impedance* or $\underline{Z}_{\text{coil}}$ *quality* QL):

$$\underline{Z} = R_S + j\omega L, \quad (3.1)$$

Table 3.1 Characteristic values of selected displacement and distance sensors

		Measuring section							
Sensor principle	Path	Distance	Absolute/incremental	Measuring distance maximum	Typical market requirement*	Accuracy class	Interfaces	Typical applications	
Inductive (LVDT)	Yes		a	2 mm to 1000 mm	4 mm to 1000 mm	2–3	Analog	Probes, hydraulic valve	
Inductive		Yes	a	0.1 mm to 100 mm	0.1 mm to 50 mm	4–3	Analog	Clamping path control in machine tools	
Capacitive		Yes	a	0 mm to 20 mm	0 mm to 10 mm	4–3	Analog	Height control, material selection	
Optical triangulation		Yes	a	20 mm to 5000 mm	20 mm to 2000 mm	7–5	Analog Digital Bus	Object detection, handling, mounting	
Optical light propagation time		Yes	a	0 mm to 10,000 mm	0 mm to 4000 mm	7	Analog Digital Bus	Robot handling, storage, logistics	
Optical interferometry	Yes		i	0 mm to	0 mm to 4000 mm	1	Digital Bus	Measuring machine, quality control	
Photoelectric scanning	Yes		i, a	0 mm to 10,000 mm	0 mm to 1000 mm	1–2	Analog 1 Vpp Digital Bus	Controlled feed axis, probes, linear drives	
Ultrasound		Yes	a	25 mm to 8000 mm	250 mm to 4000 mm	5–6	Analog Digital	Level control, distance control, height control	
Magneto-Inductive	Yes		a	25 mm to 300 mm	0 mm to 150 mm	4	Analog	Clamping travel control	
Magneto-Strictive	Yes		a	25 mm to 8000 mm	25 mm to 2000 mm	3–2	Analog Digital Bus	Hydraulic axes, automation and handling, level control	

(continued)

Table 3.1 (continued)

Sensor principle	Measuring section						Typical applications	
	Path	Distance	Absolute/incremental	Measuring distance maximum	Typical market requirement*	Accuracy class		Interfaces
Magneto-electric scanning	Yes		i, a	0 mm to 5000 mm	10 mm to 10,000 mm	2-1	Analog 1 V _{ss} Digital Bus	Linear drives, automation and handling, linear guides
Potentiometric	Yes		a	0 mm to 5000 mm	0 mm to 1000 mm	4-3	Analog	Automotive engineering, probes, hydraulic axes



WE BRING INDUSTRY 4.0 ON THE ROAD.

THIS IS SICK

Sensor Intelligence.

The information age has just begun for industry. Intelligent, robust and reliable sensor technology is indispensable for challenges such as safe human-machine interaction, increasingly individual customer requirements, high variance and the control of short-term fluctuations in demand. We show you what is already possible today. Join us on the path to a more efficient future. www.sick.de/i40



Table 3.2 Accuracy classes of selected displacement sensors

Accuracy	<1 μm	<10 μm	<50 μm	<100 μm	<500 μm	<1 mm	<5 mm
Class	1	2	3	4	5	6	7

$$Q_L = \frac{\text{Im}(\underline{Z})}{\text{Re}(\underline{Z})} = \frac{\omega \cdot L}{R_S}. \quad (3.2)$$

$\text{Im}(\underline{Z})$ is the imaginary part and $\text{Re}(\underline{Z})$ is the real part of the coil impedance \underline{Z} , ω is the angular frequency of the coil excitation current and R_S is the coil series loss resistance, which summarizes all losses in the coil. A systematic classification of inductive sensors essentially contains the following main classes:

- *Eddy current inductive sensors* (Eddy/Foucault Current Sensors).
- Inductive sensors with *variable reluctance* (magnetic resistance).
- *LVDT differential transformer sensors* (Linear Variable Differential Transformer).
- *M-In-Track* and *R-In-Track*® Sagentia technologies.
- *Variable transformers*: resolver, synchro, inductosyn, microsyn (Sect. 3.2.5).

For a systematic presentation of different variants, Table 3.3 shows the characteristic values of the inductive sensors and the different versions of these values. In this morphological modular system, each sensor type can be described functionally and in terms of content by a “realization path”.

3.1.1.2 Non-contact Inductive Distance Sensors (INS)

They are also known as “*Inductive Proximity Sensors*” (*INS*) and belong to the most widely used IS. Their main specific characteristics are reliability, robustness, insensitivity to contamination, compact design, electrical stability and a very favourable price/performance ratio. Therefore, they are preferred in all industrial areas.

The morphological realization path of the INS is illustrated with the *checkered line* in Table 3.3. Accordingly, the INS perform a *contactless distance measurement* to a non-cooperative position sensor (any metal object). The target can be the metal object to be detected directly. The operation of INS is based on the interaction of a coil, which belongs to the sensing element and generates an electromagnetic field, with the metal object placed within the detection range of the coil.

Sensor Element and Block Diagram of the INS

The one-piece INS consist of the *sensor element* with the sensor coil and of *sensor electronics* integrated in the sensor (Fig. 3.1). According to its main functions, the sensor electronics is divided into two functional areas: In the *front end*, the coil excitation, the measurement of the coil quality and a first signal processing (e.g. linearization and

temperature compensation) are carried out; in the *back end*, the signal processing, the generation of standardized sensor output signals and the execution of the numerous protective functions required by the standards are carried out.

Figure 3.2 shows a perspective sectional view of the structure of a typical sensor element for cylindrical sensor designs in a metal tube (common design). The essential component is the round, rotationally symmetrical coil in the form of a winding (1) on a winding body (2). The coil installed in a shell core (3) with the external diameter D is protected by the metal tube (4) and, in the sensor viewing direction, by a plastic cap (5); these components do not play a significant role in distance measurement, but do influence the sensor performance.

With this sensor element topology, the standardized cylindrical designs can be implemented in threaded metal tubes M8 to M30 (flush or non-flush mountable versions) (Fig. 3.3).

In addition to these variants, there are also very compact cylindrical or cuboidal versions of the INS. A characteristic trend for all versions is the high degree of miniaturization, which is reflected in the standardized designs in short sensor lengths (up to the minimum permissible length specified in the standard). The smallest INS have diameters from 3 mm to 8 mm, which are comparable to the diameter of the connection cable and have lengths starting at about 30 mm.

Measuring Principle of the INS

Basically the INS are *eddy current IS*. From an electrotechnical point of view, the sensor element of the INS is a coil with losses caused by the object to be detected, but also by the components winding, core and metal tube (inherent losses). This coil is excited by the sensor electronics with high-frequency current and generates a *magnetic field* whose distribution and strength depend on the coil design (geometry, number of turns) or the current parameters (amplitude, frequency) (Fig. 3.4). This field induces *eddy currents* in the object to be detected (Sect. 2.9). The *electrical losses* in the object depend on the strength of the magnetic field, the material properties of the object (electrical or magnetic conductivity) and, above all, the distance between the object and the sensor coil. They lead to a change in the electrical parameters of the coil, which, however, depend on the selected coil replacement circuit. A preferred electrical equivalent circuit of the coil is the *Jordan series equivalent circuit*. This simplified and pragmatic representation consists of the coil inductance L in series with the loss or series resistance R_S (Fig. 3.5).

The functional principle of the INS is usually not yet fully described by the eddy current effects, because most of them have a coil with a core. The geometry of the core is always chosen so that the INS has a preferred “viewing direction”. In the case of a ferromagnetic object, the core forms a *closed magnetic* circuit with the object to be detected, whose magnetic resistance R_M depends on the distance. The variation of this resistance results in an additional, concise change in the coil inductance L .

The components of this equivalent circuit can be determined very easily by measurements and can be inserted into Eq. (3.2) to calculate the quality. Figure 3.6 shows the dependencies of the parameters L and R_s on the distance s to a standardised

Table 3.3 Morphological modular system of inductive sensors (realization paths of the INS with caro line, the IWS with spot line, the LVDT with semicolon caro line and the BIW sensors with semicolon spot line)

Characteristics, function etc.	Known designs, realizations				
1. Geometric measured variable	◆ Distance	◆ Path	◆ Position		
2. Active principle	◆ Electromagnetic damping (passive/active)	◆ Coupling factor influence (passive/active)	◆ Air gap change	◆ Core shift	◆ Influence on core properties
3. Position sensor (target)	◆ Non-cooperative	◆ Cooperative	◆ Magnetic yoke	◆ Diver core	
4. Mechanical connection: target - sensor	◆ Contactless (without connection)	◆ Fixed connection			
5. Electrical connection: position sensor	◆ Contactless	◆ Contact			
6. Sensor element structure	◆ Single coil	◆ Multiple coils			
7. Sensor element coil(s)	◆ With ferritic core	◆ With ferromagnetic core	◆ Coreless		
8. Topology sensor element coil(s)	◆ Volumetric (three-dimensional)	◆ Planar (two-dimensional)			
9. Sensor element replacement circuit	◆ Closed magnetic circuit	◆ Open magnetic circuit			
10. Sensor output signal SAS	◆ Binary	◆ Analog	◆ Digitally coded	◆ Bus interface	

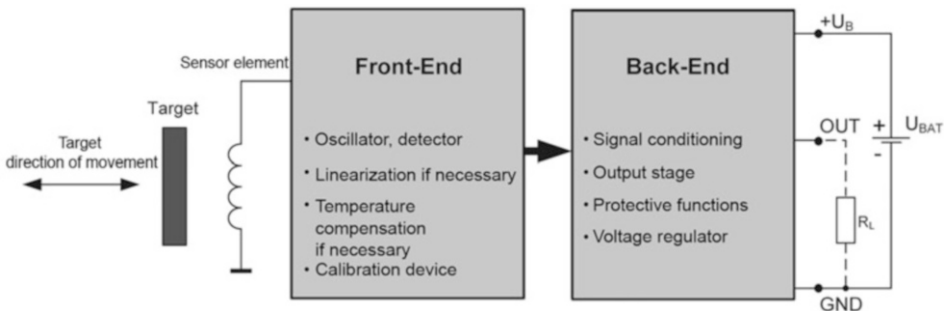
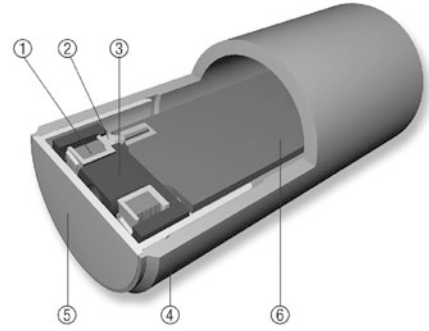


Fig. 3.1 Block diagram of an INS

Fig. 3.2 Perspective view of the sensor element of a standardised INS (Factory photo: Balluff GmbH)



ferromagnetic test object. D is the core diameter and corresponds approximately to the sensor diameter (Fig. 3.2). If one inserts these curves into Eq. (3.2), one obtains the curve of the quality Q_L of the sensor element (Fig. 3.6, right).

A purely theoretical alternative for the quality calculation consists of the evaluation of the FEM simulation results (Fig. 3.4). In this case, the following applies to the coil quality

$$Q_L = \frac{\omega \cdot \sum_{i=1}^n \mu_i \int_{dv_i} H_i(p) \cdot H_i^*(p) dv}{\sum_{i=1}^n \frac{1}{\sigma_i} \int_{dv_i} J_i(p) \cdot J_i^*(p) dv}. \quad (3.3)$$

The vectors \mathbf{H} and \mathbf{J} together with the complex conjugated vectors \mathbf{H}^* and \mathbf{J}^* represent the magnetic field strength and the electric current density, respectively, as a function of the position vector \mathbf{p} and the finite element i ($i \in [1, n]$). The scalar quantities μ_i and σ_i are the magnetic permeability and the electrical conductivity of element i , respectively.

Functionality and Sensor Electronics of the INS

For the physical conversion of the distance to be measured into an electrical quantity, the INS basically carries out a continuous measurement of the coil quality Q_L .

The limitations and challenges of this alternative of quality measurement are shown in Fig. 3.6, right. Through systematic optimization and adequate design of the sensor element, distances can be measured without contact in an area that is practically limited by the coil core diameter. For the case $s \approx D$, the stroke of the quality factor related to the value in the non-actuated state ($s = \infty$) is only 1–2%.

The precise evaluation of this minimal variation in industrial environments and with large temperature fluctuations is a great challenge for the sensor electronics. Despite all optimizations, the *quality* of the sensor element always exhibits a *non-linear distance dependence*, which must be linearized for analog INS versions in the sensor electronics.

The most commonly used method for measuring the coil quality Q_L and evaluating its distance dependency is based on installing the sensor element coil in a parallel resonant

Fig. 3.3 Standardised INS in cylindrical metal housings
(Factory photo: Balluff GmbH)

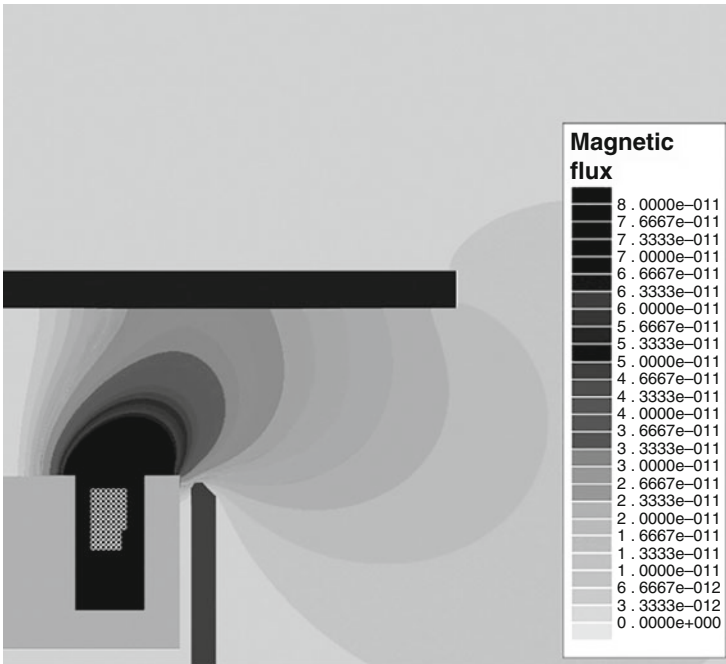


Fig. 3.4 Magnetic field of the rotationally symmetric coil—right half—of an INS (two-dimensional spectral representation of simulation results of the finite element method). Measuring unit of the magnetic flux: [Vs]

circuit whose resonant capacitor is a high-quality component with minimal losses (Fig. 3.5). This resonant circuit is supplied by an oscillator at its own resonant frequency (*sensor operating frequency*) to compensate for the losses caused by the target and to maintain the oscillations.

An example for newer implementations is the special oscillator with negative input resistance. This circuit is very adequate for monolithic integration in application specific integrated circuits (short ASIC: application specific integrated circuit). Characteristic for this oscillator is the realization of the coupling by electronic means, so that the coil tapping

Fig. 3.5 Representation of the coil of the sensor element by the Jordan series replacement circuit $L - R_S$ and the excitation of the sensor element by an oscillator

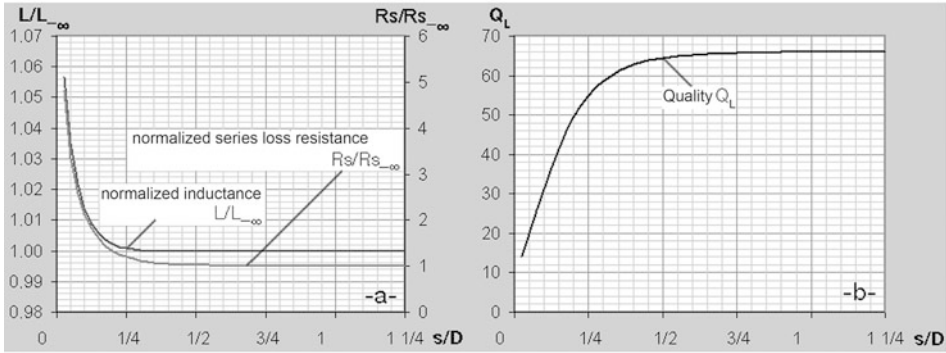
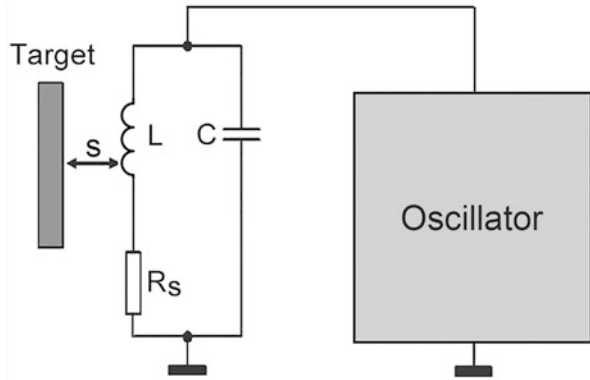


Fig. 3.6 Left: Normalized inductance and normalized loss resistance as a function of the normalized distance ($L_{-\infty}, R_{S-\infty}$ are the inductance and loss resistance in the non-actuated state: $s = \infty$). Right: Quality of the sensing element as a function of the normalized distance

can be omitted. This leads to an advantageous two-wire connection between oscillator and resonant circuit: connections LC and GND in Fig. 3.7.

The influence of the LC input is negligible with this oscillator, so that the oscillating circuit oscillates unloaded and free. The resulting oscillation frequency f_R depends, in addition to the capacitance C , only on the coil parameters L and R_S and is slightly influenced by the target distance:

$$f_R = \frac{1}{2\pi} \sqrt{\frac{L - C \cdot R_S^2}{L^2 \cdot C}} \tag{3.4}$$

The evaluation of the quality Q_L in the INS is performed indirectly by measuring the amplitude \hat{u}_{LC} of the resonant circuit voltage $u_{LC}(t)$ and is based on the following relationships:

The voltage depends on the impedance of the resonant circuit, whereby this impedance corresponds to the real part at its own resonant frequency. For the real part R_p results:

$$u_{LC}(t) = R_p I_E \sin(2\pi f_R \cdot t), \quad (3.5)$$

where I_E is the excitation current supplied by the oscillator.

The general equation of the real part R_p is

$$R_p = \frac{Q_L}{2\pi \cdot f_R \cdot C}. \quad (3.6)$$

The real part R_p shows a proportionality to the quality Q_L , if C and f_R are constant.

The sinusoidal oscillator output signal $u_{LC}(t)$ is fed to a peak value precision rectifier. This stage converts the oscillator output voltage, which depends on the quality of the sensing element, into a DC voltage whose value is equal to the amplitude \hat{u}_{LC} . This unconditioned signal, which contains the distance information, represents the front-end output. Its conditioning, i.e. its conversion into the standardized sensor output signal, takes place in the sensor output stage of the back end (Fig. 3.1). The INS described here have either a voltage output 0 V to 10 V or a current output 0 mA to 20 mA or 4 mA to 20 mA. The load resistor R_L in Fig. 3.1 symbolizes the application. In the front-end part, optional additional functions such as linearization of the distance output characteristic of the INS and/or temperature compensation can be carried out. Suitable temperature compensation ensures a minimum drift of the distance output characteristic and thus allows the INS to be used in a wider temperature range.

Calibration of the INS

An exact calibration of the INS and thus a low specimen scatter of the distance-output characteristic can be achieved by adjusting the sensors in the fully assembled state by means of a *teach-in procedure*. For this purpose, the sensor contains a calibration device in its front end (Fig. 3.1). This is a programmable resistor network connected to the oscillator connection RDIST (Fig. 3.7) and defines the distance-output characteristic of the INS.

When a calibration is initiated, the amplitude \hat{u}_{LC} of the oscillator output signal is varied by the resistor network, the value of which varies continuously after the start of the calibration process, until the sensor output signal reaches the required limit. The determined resistance value is stored in the non-volatile memory (EEPROM: electrically erasable programmable memory) of the calibration device. By means of this teach-in procedure, all scattering variables that influence the oscillator output signal are taken into account and compensated for on an example-specific basis.

The diagram in Fig. 3.8 shows as an example the adjusted distance-output characteristic and the course of the linearity error for an INS with the measuring range: 2 mm to 8 mm. The INS are increasingly being installed as *miniaturized sensors*. Figure 3.9 shows the inner workings of such sensors.

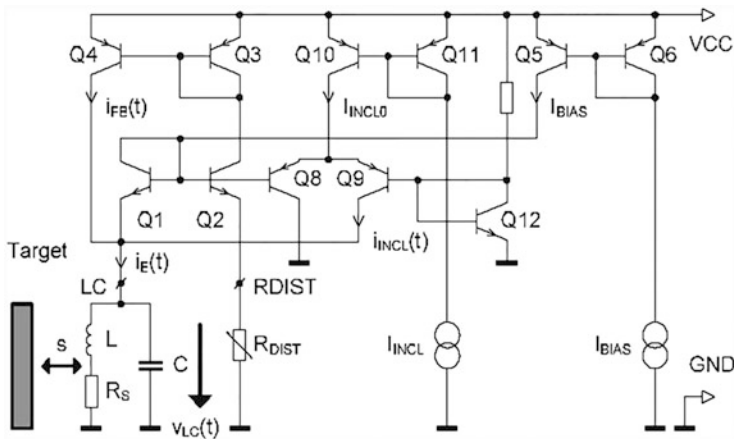


Fig. 3.7 Typical representation of the oscillator with negative input resistance in a bipolar version

Applications and Main Characteristics of the INS

The INS are originally designed and therefore predestined for the *frontal detection* of an object moving along the coil symmetry axis, perpendicular to the active surface. They are calibrated and measured for such an *axial approach* with the standardized measuring plate.

Nevertheless, a *radial movement* (past the active sensor surface) of the object is also possible. With this *lateral approach*, the sensor output signal also depends on the distance of the object to the active sensor surface and on the overlap, so that the sensor characteristics can be represented by a set of curves with these parameters.

Experience shows that the range of applications of the INS is constantly being expanded. Some examples of the wide range of industrial applications can be seen in Fig. 3.10:

- (a) *Thickness measurement* of any object with a metal tag on top.
- (b) *Detection of inhomogeneous zones* of flat metal surfaces (e.g. gap, groove, web).
- (c) *Position detection* and *contour tracking* of small parts during workpiece inspection.
- (d) *Scanning of rotating objects* (e.g. eccentrics, cams, imbalances). If the object is rotationally symmetrical, it can be centred with two distance sensors offset by 90° .
- (e) *Detection of different materials*; because the sensor sensitivity depends, among other things, on the physical properties of the object to be detected. If the distance is kept constant, the sensor output signal is basically determined by the object material.
- (f) *Lateral approach to an inclined plane*. This is a classic application example for the acquisition of larger distances with INS. If the sensor is arranged perpendicular to the base of the inclined plane, the approach path (w) depends on the distance (s) and angle of inclination (β) of the inclined plane as follows

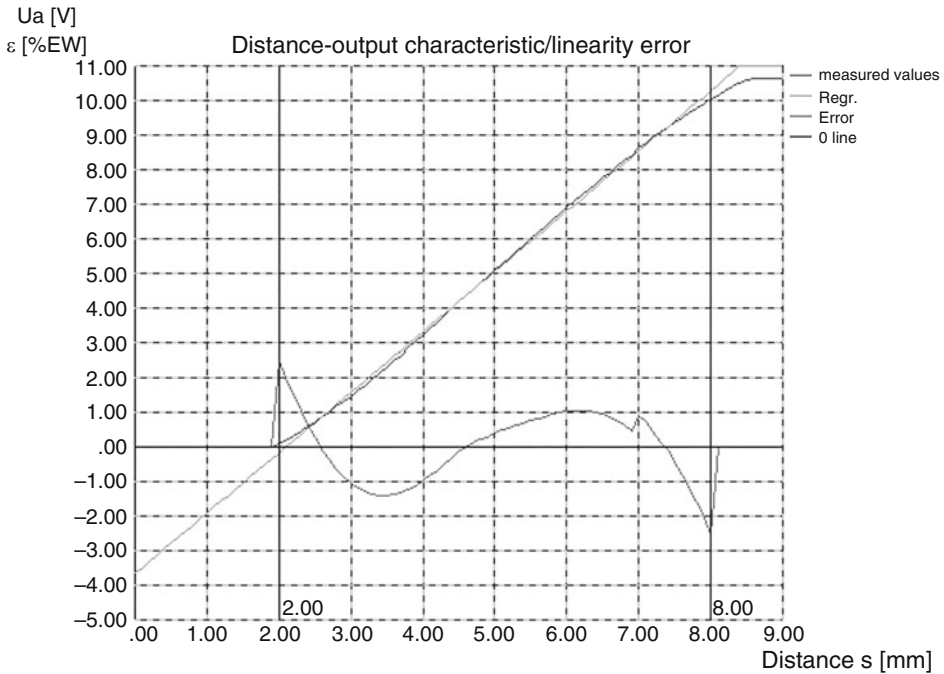


Fig. 3.8 Distance-output characteristic and course of the linearity error after calibration (Factory photo: Balluff GmbH)

$$\Delta w = \frac{\Delta s}{\tan \beta}. \quad (3.7)$$

This translation function enables significant increases in the detection range, for example by a factor of 10 at $\beta = 6.34^\circ$. This is the basis for the traditional application of the INS for monitoring the clamping path on tool spindles or workpiece clamping cylinders.

The *compact, one-piece* inductive distance sensors INS are characterized by *very good characteristics* and *cost-benefit ratio* and are ideally suited for industrial applications. Their parameters are basically dependent on the design. Table 3.4 summarizes the main characteristics of the INS.

3.1.1.3 Non-contact Inductive Displacement Sensors (IWS)

Measuring Principle of the IWS

The morphological realization path of the IWS is shown in Table 3.3 with a *dotted line*. The principle of action, target properties and the mechanical and electrical connections remain basically unchanged compared to INS.

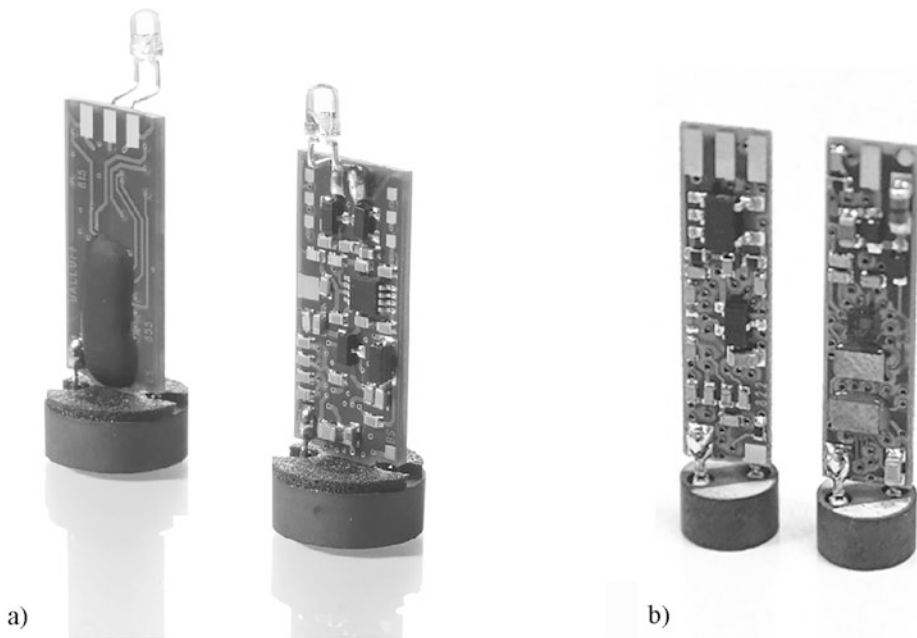


Fig. 3.9 Partially assembled INS (front and rear side, respectively), consisting of a pot core coil and the sensor electronics (Factory photo: Balluff GmbH). **(a)** ASICs wired under the globtop by wire bonding. **(b)** ASICs contacted by flip-chip technology (without protective cover); (Factory photo: Balluff GmbH)

Sensor Element and Block Diagram of the IWS

The IWS operates contact-free and detects the position of a *slim, metallic target*, which performs a translatory linear movement along the symmetry axis of the active sensor surface and parallel to this surface at a constant air gap. The coil is excited in a comparable way by the sensor electronics (Fig. 3.1), the target causes coil damping and the path-dependent coil quality Q_L is in turn evaluated by the sensor electronics.

In this case, the sensor element is a flat, coreless coil, which is plane-parallel to the active sensor surface. Since the target dimensions are significantly smaller than the coil dimensions, the target has only a local influence on the coil and causes a change in the coil quality Q_L (Sect. 3.1.1.2). This influence is directly proportional to the target projection on the coil surface. If the coil has a path-dependent geometry (implementation of the path information), a monotone clear correlation between coil quality and path *s results*.

The corresponding topology change of the coil of an IWS compared to an INS can be seen by comparing Fig. 3.11 with Fig. 3.2. Instead of a classic, rotationally symmetrical wound wire coil with a pot core, we now have a planar structure on a dielectric carrier board (usually without a core). A ferritic foil on the back of the coil could be applied as a flat core. Such topologies allow a convenient, inexpensive implementation of the path information,

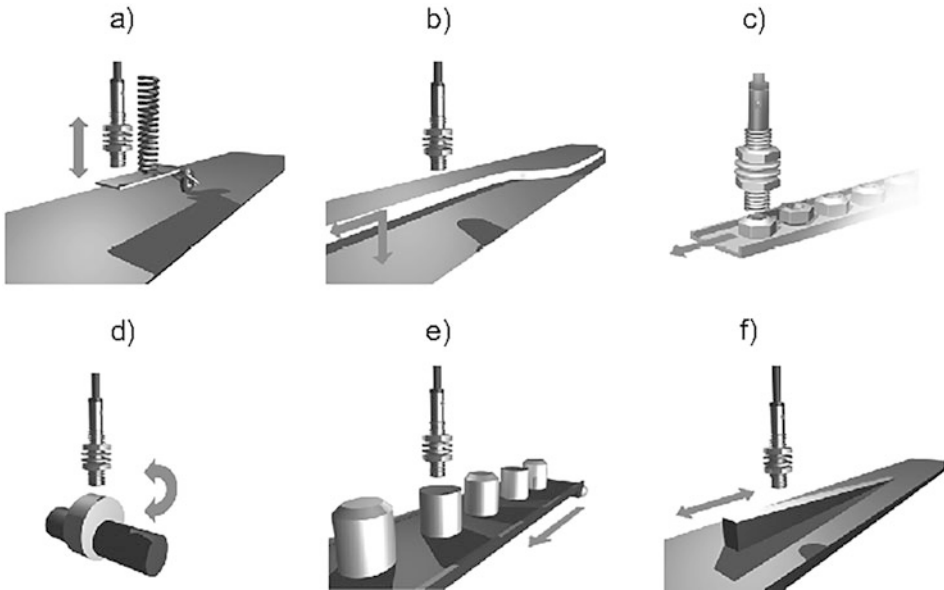


Fig. 3.10 Application examples for INS (Factory photo: Balluff GmbH)

are mechanically very flexible and can be simulated or optimized analytically and/or computationally (Fig. 3.11) in order to achieve a linearized path dependence of the Q_L function.

The block diagram of an IWS (Fig. 3.12) has many similarities to the block diagram of an INS (Fig. 3.1).

The quality of the *sensor element planar coil* is still measured indirectly with the aid of a parallel resonant circuit.

Substantial progress has been made in extending the coverage (factor 10 or more), improving the characteristics: linearity, temperature drift and the suppression of the dependency on the air gap size between sensor active area and target are achieved by using the *differential method*. The sensor element now consists of two identical path-dependent planar coils, which lie opposite one another on the same plane (Fig. 3.13). The coils are connected in a double resonance bridge with two vibrating capacitors and two resistors (Fig. 3.14) and are simultaneously influenced by the target.

For identical capacitances $C_1 = C_2$, a favourable transfer function H_0 of the bridge results:

$$H_0 = \frac{V_{\text{OUT}}}{V_{\text{IN}}} = \frac{1}{2} \left\{ \frac{Q_{L1} - Q_{L2}}{Q_{L1} + Q_{L2}} \right\} - \frac{1}{2} \left\{ \frac{R_3 - R_4}{R_3 + R_4} \right\} \quad (3.8)$$

with V_{IN} and V_{OUT} as input and output voltage respectively and R_3, R_4 as bridge resistors.

Table 3.4 Technical data of INS

Characteristics (electrical or mechanical)	Value, value range
Linearity range	Up to 50 mm
Linearity error (trend line: Regression line)	Below $\pm 3\%$ of the upper limit
Repeatability (uni- or bidirectional approximations)	Between $\pm 5 \mu\text{m}$ and $\pm 15 \mu\text{m}$
Resolution limit	Typical $\pm 0,1\%$ of the upper limit
3 dB cut-off frequencies	Up to 1 kHz
Electrical protection class	II
Protection class	IP67
Operating temperature range	$-25 \text{ }^\circ\text{C}$ ($-40 \text{ }^\circ\text{C}$) to $+75 \text{ }^\circ\text{C}$ ($+125 \text{ }^\circ\text{C}$)

Apart from the constant second term, the transfer function is proportional to the quotient: difference/sum of the coil qualities Q_{L1} or Q_{L2} . The target movement parallel to the sensor element and along its longitudinal direction causes opposite changes of Q_{L1} and Q_{L2} and thus a *monotonic* and *linear path dependence* of H_ϕ . The path detection is reduced to measure the transfer function.

Sensor Electronics and Realization of the IWS

An assembly comprising the sensor element and an FR4 carrier board with the sensor electronics is shown in Fig. 3.15. This unit can be installed in two different compact sensor housings (Fig. 3.16) and ensures a detection range of 14 mm for air gap between sensor active surface and target $d = 1 \text{ mm}$ to 2 mm .

Characteristics and Applications of the IWS

With the exception of the detection range, which is increased several times, the IWS described above are characterized by performance features that are comparable to the characteristics of INS (Table 3.1).

The IWS is especially used for monitoring the clamping path when clamping workpieces or tools in modern machine tools (Fig. 3.17).

The closing movement of the clamping device is converted into a linear movement along the axis of rotation. The function of the target is taken over by a thin metal disc, which is located on the axis of rotation and experiences a translatory movement during continuous rotation. Using the linear distance-output characteristic of the IWNs, various switching points can be set in the controller. This eliminates the need for mechanical readjustment of the clamping path monitoring. This is particularly advantageous if the clamping device is located in places that are difficult to access.

3.1.1.4 Differential Transformers with Sliding Core (LVDT)

A second distinctive class of inductive sensors uses the *influence of the coupling factor* between two or more coils of the sensing element. A simple measuring possibility results from the general equation for the coupling factor k :

Fig. 3.11 Schematic diagram of the single coil of an IWS (dimensions 30 mm × 10 mm)

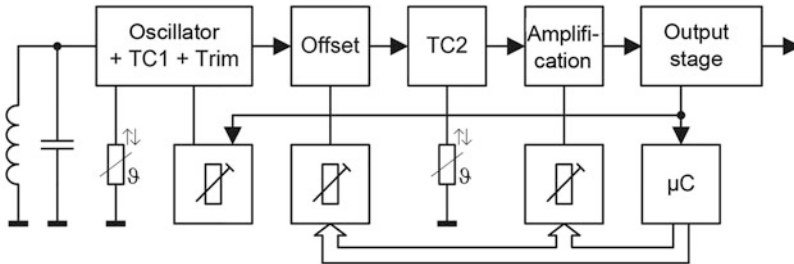
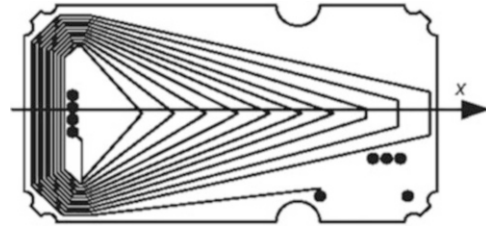


Fig. 3.12 Block diagram of an IWS

$$k = \frac{M}{\sqrt{L_1 \cdot L_2}}. \quad (3.9)$$

If the involved inductors L_1 and L_2 are slightly influenced, the change of the coupling factor can be determined in a linear way by measuring the counter-inductance M . Of particular note is the lifetime of the LVDTs. It is up to 228 years.

Sensor Element and Measuring Principle of LVDTs

The morphological realization path of an LVDT is shown in Table 3.3 with a semicolon caro line. The position sensor is now cooperative, moving within the sensor housing and having a rigid mechanical connection to the object to be detected.

In the schematic representation of the LVDT sensor element (Fig. 3.18), a primary winding, two secondary windings and a movable soft magnetic core, which is connected to the external object by a non-ferromagnetic tie rod, can be seen. In general, the wound coils have a cylindrical shape. The secondary coils are placed on both sides of the primary coil and the cylindrical core undergoes a translatory movement along the system rotation axis. The object movement in the x -direction causes a core displacement and thus a continuous, monotonous and opposite variation of the mutual inductances M_1 and M_2 between the primary winding and the respective secondary windings.

The system is supplied with an alternating input voltage:

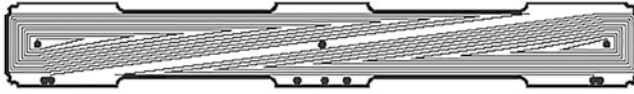


Fig. 3.13 Schematic representation of the sensor element of the improved IWS

Fig. 3.14 Double resonance bridge (coils represented by series replacement circuits)

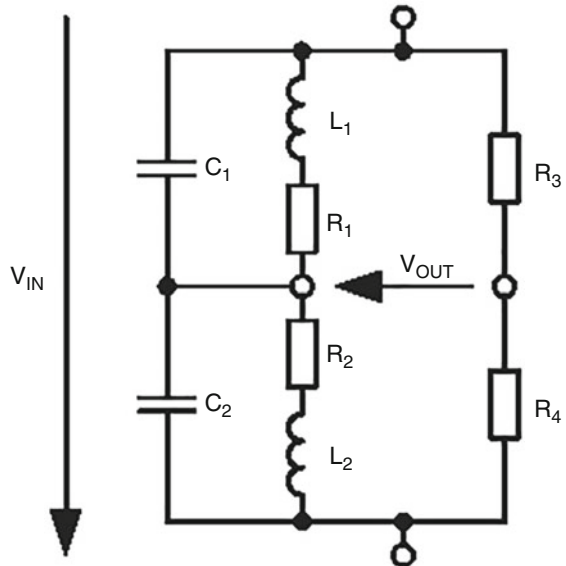
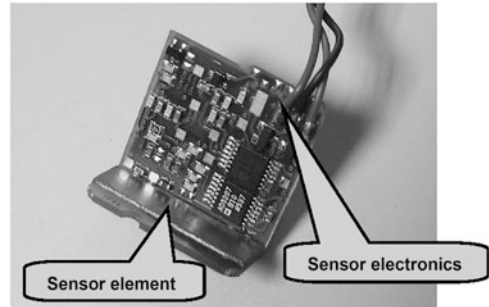


Fig. 3.15 Assembly of an IWS (the vertical electronics carrier board is mounted vertically on the sensor element); (Factory photo: Balluff GmbH)



$$u_E(t) = \hat{u}_E \cos \omega t. \quad (3.10)$$

with $\omega =$ angular frequency and $x =$ travel of the core.

The sensor element is symmetrically constructed so that in the central core position, two equal output voltages are induced into the secondary windings. If the core moves in one direction, one output voltage increases and the other output voltage decreases. To ensure good linearity of the LVDT output voltage, the secondary windings are usually connected

Fig. 3.16 Robust IWS version in metal housing with plastic cap as active surface (left) and miniaturized plastic version without mounting flange (right); (Factory photo: Balluff GmbH)

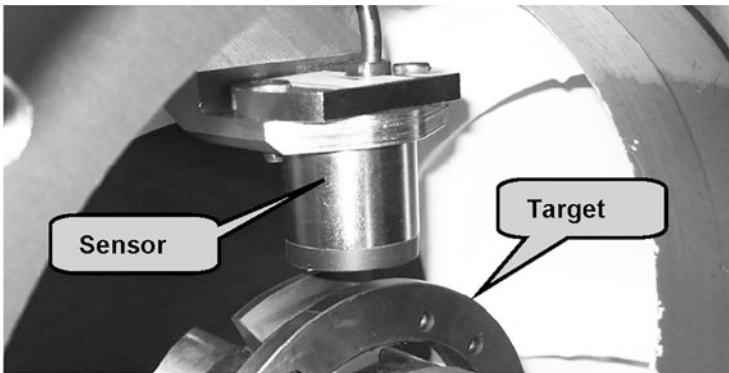
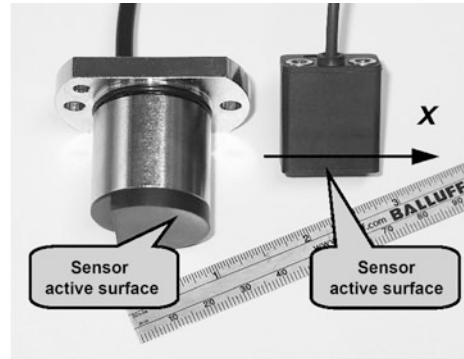


Fig. 3.17 Concrete application of an analog inductive displacement sensor IWN for monitoring the clamping displacement in a machine tool (Factory photo: Balluff GmbH)

in series (*antivalent*) and the output voltage u_A results as the difference of the induced secondary voltages.

The LVDT transfer function results for the representation in Fig. 3.18 with L_{21} or L_{22} and R_{21} or R_{22} as parameters of each winding:

$$u_A = \frac{j\omega[M_2(x) - M_1(x)]R_L}{R_1(R_2 + R_L) + j\omega[L_2R_1 + L_1(R_2 + R_L)] - \omega^2\{L_1L_2 + [M_1(x) - M_2(x)]^2\}} \cdot u_E, \quad (3.11)$$

where: $L_2 = L_{21} + L_{22}$ and $R_2 = R_{21} + R_{22}$, j is the complex variable and $M_1(x)$ and $M_2(x)$ are the mutual inductances as a function of the path x . In practice for a high-impedance output load ($R_L > 50 \text{ k}\Omega$), the Eq. (3.11) is simplified. The output voltage u_A is then directly proportional to the difference between the two mutual inductances M_1 and M_2 :

$$u_A = \frac{j\omega[M_2(x) - M_1(x)]}{R_1 + j\omega L_1} \cdot u_E. \quad (3.12)$$

The equation characterizes a high-pass filter behavior, where the -3 dB system cut-off frequency is basically determined by the parameters of the primary winding.

Functionality and Block Diagram of the LVDTs

Figure 3.19 (right-hand version) shows an improved design compared to the coaxial structure (left-hand version).

The windings are arranged overlapping. The primary winding is uniform; the secondary windings have increasing or decreasing numbers of turns per unit of travel and are arranged in opposite directions. This leads to a marked increase in the useful ratio of up to about 0.8.

Both two-piece LVDT versions and compact one-piece versions are available on the market. In the first types, an electronic extra unit (e.g. top-hat rail module) is responsible for the supply of the sensor element and the evaluation of the output voltages. The electronics generally carries out the following functions (Fig. 3.20):

- (a) Generation of the excitation voltage (typical ranges: 1 V to 24 V, 50 Hz to 20 kHz).
- (b) Supply of the primary winding with stabilized amplitude and frequency.
- (c) Synchronous detection of the sensor element output voltage(s) and, if necessary, difference formation in order to achieve a differentiation of the positive and negative core shift.
- (d) Signal conditioning, filtering and compensation.
- (e) Generation of the sensor output signal, sensor supply and protection functions.

The one-piece models with integrated electronics are more practical, more reliable and require less installation effort. Because these systems can be supplied with a DC voltage directly from the control cabinet, they are also called DCLVDT.

The widespread use of LVDT systems has motivated the semiconductor industry to produce commercial integrated circuits for LVDT systems. Figure 3.21 shows two such ICs (*Integrated Circuit*) with their schematic signal processing as a block diagram.

The AD598 module from Analog Devices performs separate rectification of the secondary voltages, calculates the sum and difference of the rectified signals and performs the proven $(A - B)/(A + B)$ procedure. Due to the asymmetry, this leads to the elimination of all common mode interference. The AD698 device uses the classic differential output voltage (u_A in Fig. 3.18) as well as the input voltage and performs quotient formation, which ensures very efficient compensation of supply fluctuations.

Implementation of the LVDTs

The sensor market offers a wide range of LVDT systems for extremely small to very large measuring ranges and with or without integrated electronics. The typical external

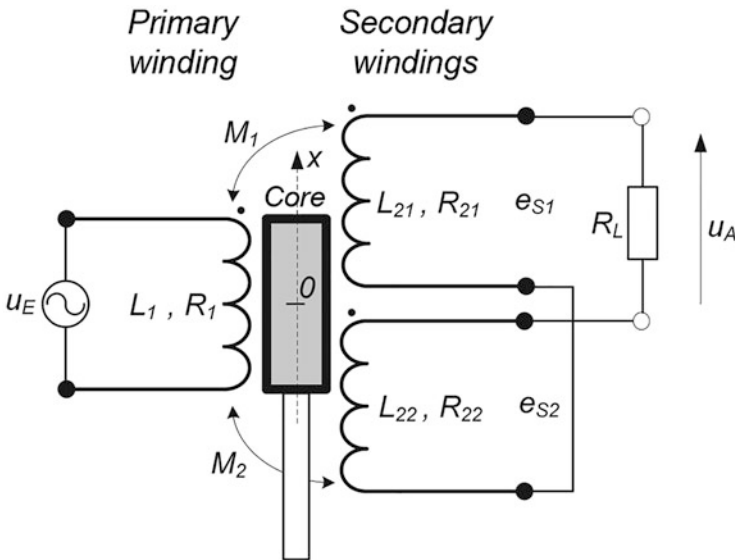


Fig. 3.18 Schematic diagram of the sensor element of an LVDT

appearance of an LVDT is the pump design (Fig. 3.22) with characteristic features: round, cylindrical metal housing (smooth or with mounting options such as thread and articulated eye) and tie rod (also with various coupling options).

Applications and Main Characteristics of LVDTs

The application range of LVDTs is very wide. The LVDTs are basically used for the detection of paths or positions. They are also suitable as *zero detectors* for *feedback positioning systems* from civil missions to submarines or aircraft. LVDTs are often used in machine tools and contour copying systems.

A second group of applications relates to the *measurement of physical quantities* that can cause the core to move (e.g. acceleration, inclination and pressure).

The LVDTs feature *excellent linearity*, *resolution* and *accuracy*. Thanks to their robustness and reliability, they are widely used in industrial applications, especially in harsh, corrosive or radioactive environments. In addition, they are easy to mount or adjust and provide galvanic isolation. Table 3.5 summarizes the main characteristics of LVDTs.

3.1.1.5 Pulsed Inductive Linear Position Sensor (Micropulse® BIW)

The BIW position sensors represent a new sensor family in the IS class, based on the influence of the coupling factor between two or more coils (Sect. 3.1.1.4). Their morphological realization path is shown with a semicolon dotted line in Table 3.3.

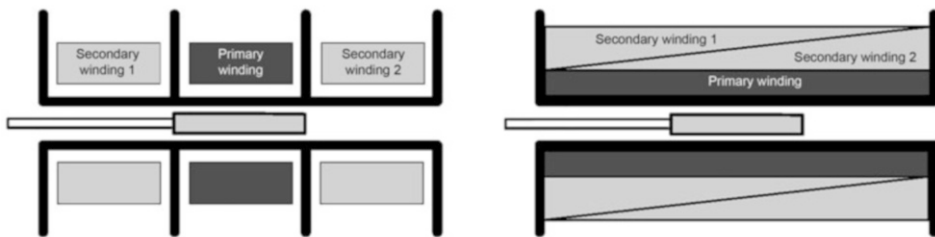


Fig. 3.19 Construction variants of the sensor element of an LVDT

Sensor Element and Measuring Principle of the BIWs

The BIW measuring principle can be used in principle, even if the position sensor is not cooperative—and thus causes a neutral non-purpose interaction with the sensor element—and also has no mechanical connection with the sensor. Nevertheless, in order to suppress parasitic couplings (e.g. air coupling) and to minimize undesired influences on the coupling factor k (e.g. mechanical fluctuations), versions with *cooperative, frequency-selective position encoders* are preferred. In addition, it is *mechanically connected* to the object to be detected and is *mechanically guided very precisely with* respect to the coils.

Striking differences to the related LVDT can be seen in the design of the windings of the sensor element or core. The sensor element consists of three planar windings printed on a vertical hard carrier board (104 in Fig. 3.23).

A rectangular winding (106) surrounds two trapezoidal windings (100 or 102) and forms the transmitting coil. The receiver coils (100) and (102) are identical, but spatially offset by 180° from each other along the direction of movement. The slim position transmitter consists of a ferritic yoke (110), which covers the carrier board on both sides. Its mechanical guide, not shown in the picture, ensures a translatory linear movement along the board axis with constant air gaps on both sides. The system has a double symmetry so that fluctuations in the transverse directions Y and/or Z are balanced out. The position sensor (110) carries a passive LC resonant circuit (112) whose resonant frequency is tuned to the frequency of the transmitter coil excitation. Under these conditions, it may be assumed that the *coupling factors* between the primary loop (106) and the secondary loops (100) and (102) have a *linear dependence* on the *path x*. Accordingly, the induced secondary voltages are clearly and directly proportional to the path x . The generation of the excitation current and the evaluation of the induced signals from the receiver coils (100) and (102) are carried out in the sensor electronics, which is identified by (108) in Fig. 3.23.

Functionality and Block Diagram of the BIWs

The block diagram of the front-end electronics is shown in Fig. 3.24. The rectangular transmitter coil is excited with rectangular current pulses (typical pulse width $t_W = 0.6 \mu\text{s}$, repetition period approx. $T = 0.02 \text{ ms}$). The resulting electromagnetic field activates the passive LC resonant circuit of the position encoder. The recharging current, which

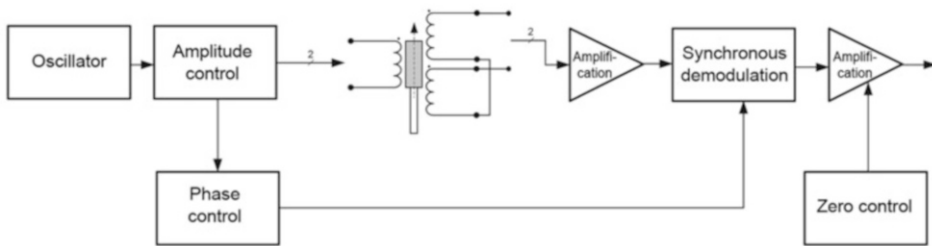


Fig. 3.20 General block diagram of the front-end evaluation electronics of an LVDT

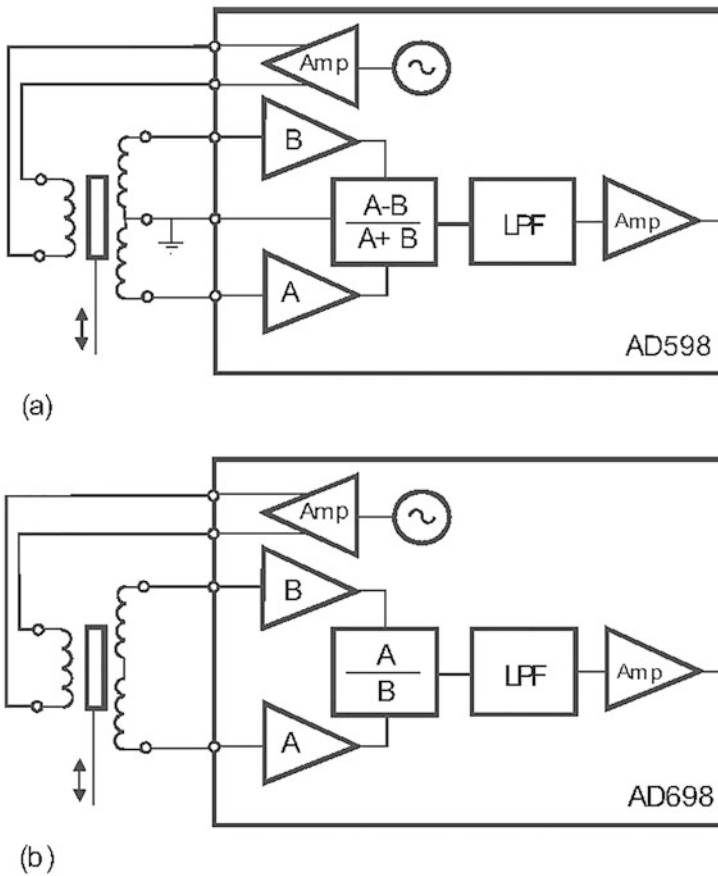


Fig. 3.21 Commercial ICs for supply and evaluation of the LVDT sensing element



Fig. 3.22 Different commercial LVDT variants (Factory photo: Inelta Sensor Systems)

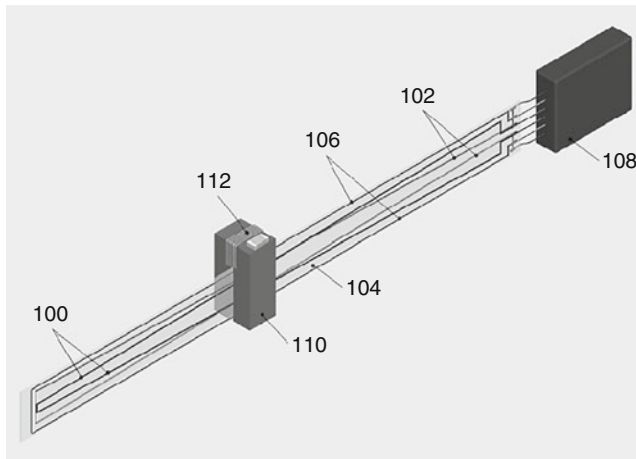
oscillates between the inductor L and the capacitor C , causes a *re-induction* in the secondary windings. This results in the two induced secondary voltages, the values of which are theoretically directly proportional to the interface between the core overlap (constant over path x) and the respective secondary loops. Two in-phase rectifiers, followed by “sample-and-hold” circuits, ensure synchronous detection of the secondary voltages and storage of their values during the pause between the supply pulses. The regulator Reg compares the sum of both secondary voltages (constant over travel) with a reference voltage and readjusts the amplitude of the supply pulses to compensate for fluctuations in the transverse directions. The back-end electronics integrated in the BIW position encoder optionally supplies a standard analog voltage or current signal (M12 industrial standard connector) and also ensures that the standard sensor requirements are met (Fig. 3.1).

Realization of the BIWs

The BIW position sensor contains the transmitter/receiver sensor element, the position sensor with the resonant circuit and the evaluation electronics, protected by an extruded aluminum profile (Fig. 3.25). The position sensor is attached to a push rod that is connected to the machine part whose position is to be determined. The oscillating circuit is briefly excited via the transmitter sensor element at a measured value rate of 32 kHz and couples a signal into the receiver sensor element at the current position. The position is immediately available at the output and is absolute. The direction of the output signal, i.e. a rising or falling characteristic curve, can be inverted via programming inputs. With the standard pitch selection, the output has 0 V when the push rod is fully extended.

Table 3.5 LVDT specifications

Characteristics (electrical or mechanical)	Value, value range
Measuring range	$\pm 1 \text{ mm}$ to $\pm 500 \text{ mm}$
Linearity error	0.05% to 1% of final value (F.S.)
Sensitivity	1 to 500 mV/V·mm
Repeatability	0.002% to 0.05% F.S.
Input voltage	1 V_{eff} to 50 V_{eff}
Input frequency	50 Hz to 25 kHz
Temperature coefficient	0.02 to 0.05% F.S./°C
Protection class	IP64 to IP 65
Operating temperature range	-250 °C to +600 °C
Air pressure	Up to 200 bar

**Fig. 3.23** Schematic representation of the sensor element and the position sensor of the BIW position sensor

Applications and Main Characteristics of the BIWs

The new patented operating principle enables position detection with high resolution and reproducibility and, above all, with an impressively *high measuring rate*. Furthermore, the sensor is characterized by

- (a) *Linear characteristic curve, potential-free measurement.*
- (b) *Insensitivity* to shocks, vibrations and interference fields.
- (c) Superior *operational reliability* and performance at no additional cost. This results in comparable prices to potentiometers.

- (d) *Contactless alternative* to classical potentiometers without friction between components of the sensing element (wear-free sensing element).
- (e) Same or similar design (including mounting brackets) and output signals ensure *compatibility* with potentiometers.

Typical applications are:

- (a) Injection moulding machines for plastics (Fig. 3.26).
- (b) Blow moulding machines, presses for rubber parts.
- (c) Pneumatically operated assembly presses.
- (d) Piston position detection in pneumatic or hydraulic cylinders.

Finally, the BIW position sensors are an ideal, wear-free *drop-in replacement* for linear contact potentiometers. Table 3.6 summarizes the main characteristics of the BIW.

3.1.1.6 Signal Processing by Phase Measurement (Sagentia)

Particularly with magnetic principles for inductive distance and displacement sensors, there are *cost disadvantages* due to

- Costs and availability of magnetic materials, especially rare earth based magnets.
- Complex and thus cost-relevant assembly and connection technology for combining magnetic materials with current-carrying components such as printed circuit boards or coils.

Technical disadvantages of magnetic principles are given by:

- Danger of magnetic foreign particles accumulating on the magnetic encoder.
- The sensor can be influenced by static external magnetic fields.

To avoid these disadvantages, the company Sagentia developed an inductive measuring principle. This offers the advantage that the signal processing is based on a *phase measurement* and not on a measurement of amplitudes, which usually requires complex measurement circuits (rectifiers, high-quality, expensive operational amplifiers).

Functionality

On a signal circuit board, there is one *sinusoidal* and one *cosinusoidal* conductor loop ($\sin(x)$, $-\cos(x)$ -transmission loop; Fig. 3.27). These are each supplied with an AC voltage ($u \cdot \sin(\omega t)$, $u \cdot \cos(\omega t)$), which is *phase-shifted* by 90° . This produces *alternating magnetic fields* $H(\omega t)$ *perpendicular* to the PCB, the strength of which is also sinusoidal or cosinusoidal over the measurement path x (Fig. 3.28).

Figures 3.27 and 3.28 show the principle of operation: According to the trigonometric addition theorem, the following relationship applies to the location-dependent sum of both magnetic fields:

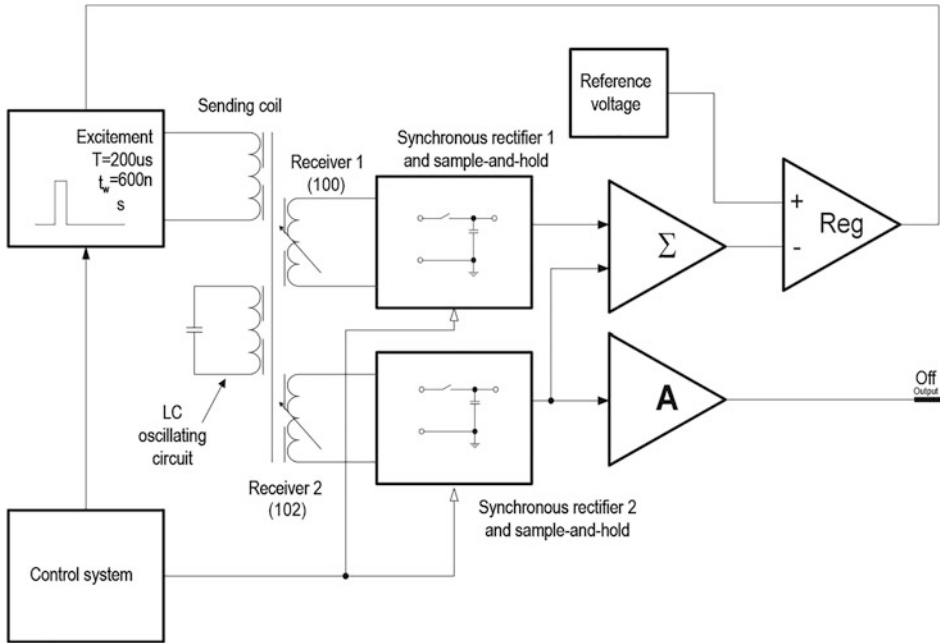


Fig. 3.24 Partial block diagram of the BIW position sensor

Fig. 3.25 BIW as a displacement transducer in a standardised extruded aluminium profile (Factory photo: Balluff GmbH)



$$H \cdot \sin(x) \cdot \cos(\omega t) + H \cdot \cos(x) \cdot \sin(\omega t) = H \cdot \sin(\omega t + x). \quad (3.13)$$

It is: H : magnetic field strength and x a path information.

This results in a signal whose *phase shift* relative to the transmitted signal is directly *proportional to the path* x .

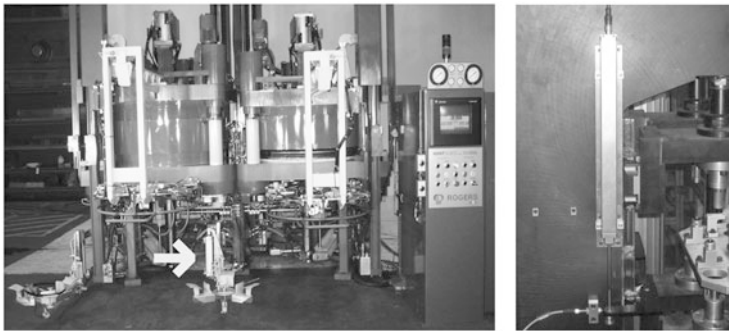


Fig. 3.26 BIW position sensor replaces worn potentiometers (Factory photo: Balluff GmbH)

The *position sensor*, which is designed as an oscillating circuit, serves as the *summation element* of the two transmitted signals. It “floats” above the signal circuit board (Fig. 3.27.). Its resonance frequency is tuned to the transmission frequency of the two fed-in signals. It is excited by them (TX) and in turn sends its magnetic field back to the circuit board (RX). The rectangular receiving coil, also integrated in the signal circuit board, receives this signal and passes it on to the evaluation electronics. Here the received signal is compared with one of the two transmitted signals. The resulting phase information is processed by the evaluation electronics into an analog voltage signal linear over the measuring path as path information.

The evaluation of the phase is carried out as a time measurement of the offset of the two signals at the zero crossing (Fig. 3.29).

Here, signals in the frequency range 1 kHz to 10 kHz can be evaluated well with a reasonable amount of circuitry in order to make a reasonable compromise between measurement resolution and update rate of the signal (dynamics of the measurement system). If, for example, a measurement signal frequency of 4 kHz and a resolution of 12 bits is selected, the result is.

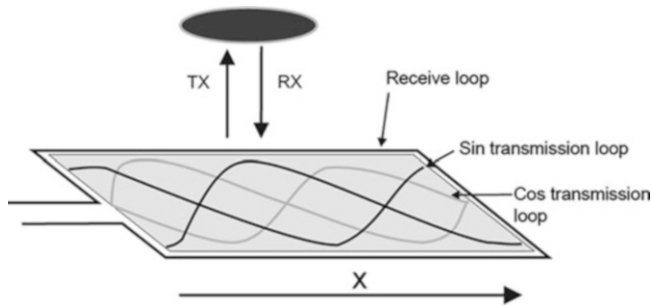
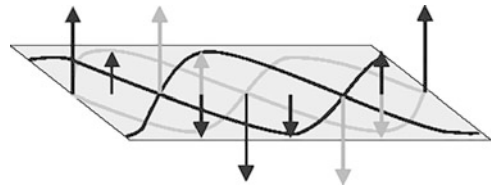
$$\begin{aligned} \text{Update time} &= 1/(4\text{kHz}) \triangleq 250\mu\text{s}, \text{ Smallest time increment} \\ &= 250\mu\text{s}/2^{12} \triangleq 61,035\text{ns}, \text{ Internal clock rate} = 1/61,035\text{ns} \triangleq 16,38\text{MHz}. \end{aligned}$$

However, the frequency range 1 kHz to 10 kHz, which is favourable for position evaluation, is not suitable for signal transmission between transmitter signal lines ($\sin(x)$, $\cos(x)$) and position encoder, since the required inductance values and resonant circuit qualities (designed as a conductor track) could no longer be accommodated in an acceptable space.

As a solution, the useful signals are transmitted amplitude-modulated on a carrier in the MHz range. This also has the advantage that several systems can be operated in *parallel* on

Table 3.6 Technical data of the BIW position sensors

Characteristics (electrical or mechanical)	Value, value range
Output signals	0 V to 10 V; -10 V to +10 V; 0 mA to 20 mA; 4 mA to 20 mA
Standard nominal lengths	75 mm to ± 750 mm (16 steps)
Linearity error	$\leq 0.02\%$ of final value (F.S.)
Resolution	5 μm
Repeatability	10 μm
Measured value rate	Typ. 32 kHz
Protection class	IP54
Operating temperature range	-20 °C to +85 °C

**Fig. 3.27** sin/cos transmission loops**Fig. 3.28** Field components

different carrier frequencies without mutual interference, which enables redundant sensors. Equation (3.13) is thus extended to:

$$\begin{aligned}
 & H \cdot \sin(x) \cdot [1 + \cos(\omega_m t)] \cdot \cos(\omega_o t) + H \cdot \cos(x) \cdot [1 + \sin(\omega_m t)] \cdot \cos(\omega_o t) \\
 &= H \cdot \sin(\omega_m t + x) \cdot \cos(\omega_o t) + H \cdot (\sin(x) + \cos(x)) \cdot \cos(\omega_o t) \\
 &= H \cdot [\sin(\omega_m t + x) + \sin(x) + \cos(x)] \cdot \cos(\omega_o t).
 \end{aligned}$$

(3.14)

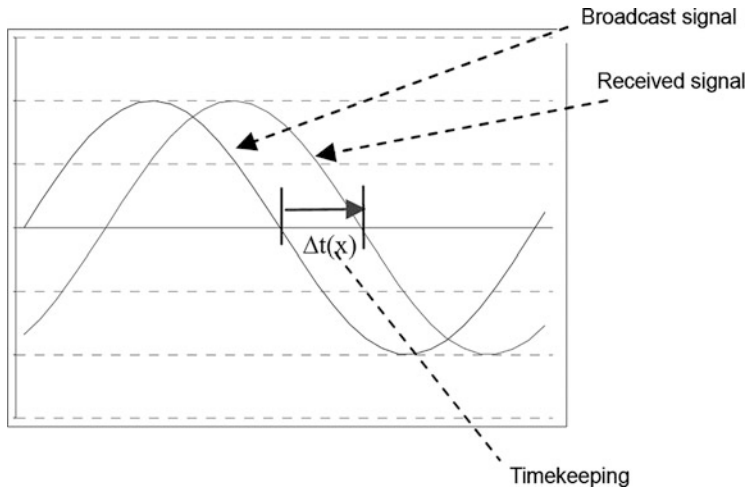


Fig. 3.29 Phase measurement

It is: H : magnetic field strength; x : path information; ω_m : modulation frequency in the kHz range; ω_o : carrier frequency. The path information x can be extracted from the *phase analysis* of the alternating component of the envelope. With this measuring principle, a further measuring length range can be covered (structure length from 15 mm to 800 mm; Fig. 3.30).

The deviation of the $\sin(x)$ and $\cos(x)$ structures from the mathematical ideal increases with shorter measurement lengths due to the design, since the necessary through-connections and line bridges cannot be arbitrarily small.

A sin- or cos-shaped field curve over the measurement path can also be generated with other non-trigonometric (e.g. rectangular) structures. Furthermore, the measuring principle is not limited to linear displacement measurement, but can also be implemented for evaluation of *rotary* movements (Fig. 3.31). The trigonometric path dependencies $\sin(x)$, $\cos(x)$ are then converted into $\sin(\varphi, r)$, $\cos(\varphi, r)$, depending on angle and radius.

The inductive sensor can be described by the following block diagram (Fig. 3.32):

The electronics are designed for a temperature range from $-40\text{ }^\circ\text{C}$ to $+125\text{ }^\circ\text{C}$. The temperature limitation for the actual measuring system “Signal PCB—Position Encoder” is essentially only dependent on the PCB material and the mounting and connection technology. Since the sensor operates without contact, a version with IP69k protection class is possible. An application example is shown below.

Actual Value Acquisition at a Hydraulic Steering Cylinder for Fork Lift Trucks

Today’s magnetostrictive sensors designed for direct integration in cylinders are generally not suitable for this application, as the continuous piston rod makes a hollow bore for sensor mounting impossible. External mounting of such sensors, in which a magnet or

magnetic ring is attached to the piston as a position sensor for a measurement through the cylinder wall, is also ruled out in this area of application. The hydraulic steering cylinders are made of steel and are therefore themselves magnetizable; measurement through the cylinder wall using magnetic methods is therefore not possible.

The inductive sensor (Fig. 3.33) is mounted on the outside of the steering cylinder (Fig. 3.34). The position sensor is attached to a metal bracket, which is connected to both ends of the piston rod. A steering movement thus leads to a linear movement of the position transmitter over the sensor housing, which is made of plastic on the top side, otherwise of aluminum. The measuring length of the sensor is 170 mm. The distance between the position sensor and the sensor housing is a few millimetres. The position of the encoder can vary in the mm range due to mounting tolerances of the entire system with regard to distance and lateral offset. It is important that this does not have a negative effect on the sensor function.

Because the inductive sensor is used within the electronic steering system of the forklift truck, it must meet safety criteria. The sensor was therefore designed to be completely *redundant*. Both the conductor loops, the position encoder circuit and the evaluation electronics are duplicated. The analog 0.5 V to 4.5 V signal with an opposing linear characteristic is available at the output. The sum of both signals is therefore constant; this allows a simple plausibility check. Mutual interference between the two measuring channels is ruled out by appropriate selection of the carrier frequencies (MHz range).

The resolution is <0.1 mm. Due to the complete encapsulation, the sensor is *very robust* with regard to *vibration stress* and *insensitive to moisture and dirt*. The sensor is steam jet resistant, resistant to all chemical substances potentially occurring in this application and is suitable for an environment with rapidly changing temperatures, as is typically the case in cold storage.

3.1.1.7 PLCD Displacement Sensors (Permanent Linear Contactless Displacement Sensor)

The PLCDs are used as very suitable non-contact displacement sensors in the range from about 20 mm to over 250 mm with control distances in the cm range. The magnetic drive is flexible and adaptable to the application. The system has been further developed so that today very compact, low-cost PLCDs realized with specific integrated circuits are available on the market. The driver for the further development of this system is clearly the automotive industry; the displacement sensor is used in large numbers in automobiles.

The PLCDs perform a non-contact displacement detection of a cooperative target (permanent magnet) and are based on the coupling factor influence due to a punctual core saturation.

Sensor Element and Measuring Principle of PLCDs

PLCDs consist of a soft magnetic core, which is surrounded by a long secondary coil along its entire length and by short primary coils at both ends (Fig. 3.35). The primary coils are driven by alternating voltage. When the position transmitter approaches the sensor, the

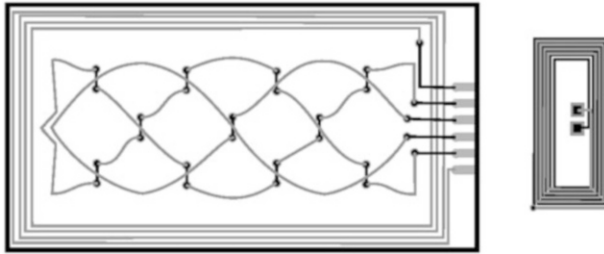


Fig. 3.30 25-mm signal PCB with position sensor

Fig. 3.31 Example of a rotative structure (Factory photo: Sagentia)

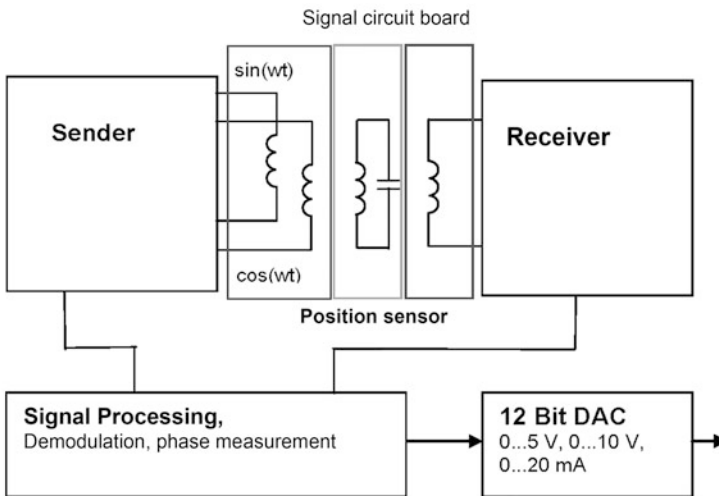
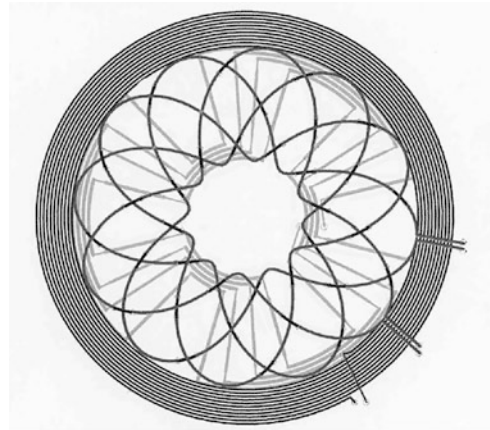


Fig. 3.32 Block diagram of total sensor



Fig. 3.33 Inductive sensor (Factory photo: Novotechnik)

aforementioned magnetic saturation of the core occurs locally. To a certain extent, the core is magnetically separated and the primary coils only act on the part of the secondary coil that lies between the primary coil and the separation point (Fig. 3.35).

From the *amplitude* and the *phase position* of the secondary voltage relative to the primary voltage, a DC voltage results through in-phase demodulation, which depends linearly on the momentary position of the permanent magnet.

Various arrangements of the magnet and thus of the magnetic field direction in relation to the sensor are possible, as well as an inverted version with a single long primary coil and two secondary coils, each with advantages and disadvantages.

The PLCD measuring principle is explained using the illustration in Fig. 3.36, where PS1 and PS2 are two primary coils excited in opposite phase with alternating current. The third coil SEKS is the secondary coil. In principle, each primary coil generates an alternating magnetic field, which continues to induce a voltage in the secondary coil.

In the case without a permanent magnet, the resulting secondary voltage is zero because the alternating magnetic field generated by PS1 is cancelled by the alternating field of PS2 due to the antiphase excitation.

If the permanent magnet is placed in the immediate vicinity, the above-mentioned local saturation occurs, which leads to a virtual core splitting or a virtual air gap LS in the core. The resulting spatial course of the AC field lines is shown in simulation Fig. 3.37; they leave the core continuously at the lateral surface. The flux leakage causes a decrease of the main flux in the core up to a value in the virtual air gap LS that has dropped to zero.



Fig. 3.34 Sensor mounted on steering cylinder; position sensor mounted on movable bracket (Factory photo: Novotechnik)

The field line density (induction) along the longitudinal axis of the PLCD decreases continuously; this course is shown in Fig. 3.38 in an idealized form. The main flux has an opposite sign to the right in the left part of the split core.

The induced voltage u_{SEK} in the total secondary coil with N turns can be calculated for an angular frequency ω of excitation using the law of induction and by superposition as the sum of the induced coils in all individual turns ($w \in 1 \dots N$):

$$u_{\text{SEK}} = \omega \sum_{w=1}^N \Phi_w, \quad (3.15)$$

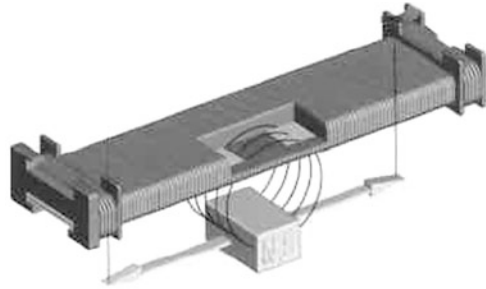
where: Φ_w is the flow through the w -turn. For the idealized, linear course (infinite air gap width) in Fig. 3.38, the sum can be replaced by area integrals and these lead to a very favorable equation:

$$u_{\text{SEK}} = \frac{\omega N}{L} \int_{-\frac{L}{2}}^{\frac{L}{2}} \Phi_w(x) \cdot dx = -\frac{\omega N}{L} \Phi_0 \cdot x \quad (3.16)$$

with: L = coil length, Φ_0 = original value of magnetic flux and x = target position.

Accordingly, the *induced voltage* u_{SEK} is *linearly dependent on the magnet position*.

Fig. 3.35 Sensor element of a PLCD with the permanent magnetic target (Factory photo: Tyco)



Functionality and Block Diagram of the PLCDs

The PLCD displacement sensors available on the market are exclusively one-piece compact designs that contain both the sensor element and the sensor electronics. The construction corresponds completely to the general sensor definition.

The electronics basically performs the following standard functions of a displacement sensor (Sect. 3.1.1.4):

- Generation of the excitation voltages.
- Supply of the primary windings with stabilized amplitude and frequency.
- Synchronous detection of the secondary voltage.
- Signal conditioning, filtering and compensation.
- Generation of the sensor output signal, sensor supply and protection functions.

In this way, both an analog output signal and a pulse width modulated output signal (PWM: Pulse Width Modulation) can be guaranteed.

The analog versions can have outputs for standardized output signals (Sect. 3.1.1.2), but also ratiometric outputs. In the second case, the upper limit of the sensor-displacement output characteristic is directly dependent on the level of the supply voltage (Fig. 3.39).

The PWM versions provide a pulse train of digital pulses with a fixed repetition frequency (0.25/0.5/1.0/2.0 kHz) and a width that varies in direct proportion to the path. This signal has a higher noise immunity; in the control unit, both the pulse width and the duty cycle can be evaluated.

The most important feature of the PLCD displacement sensor for the application is the *complete freedom of connection* between the control magnet and the sensor. The driving magnetic field is only an indicator for the target position. In the first order, the field strength is not included in the measurement signal as long as it is large enough to saturate the core. This results in a *high system tolerance* against fluctuations of the air gap between magnet and sensor. The possible magnet arrangements with corresponding applications are shown in Table 3.7.

Fig. 3.36 Schematic representation of the PLCD sensor element structure

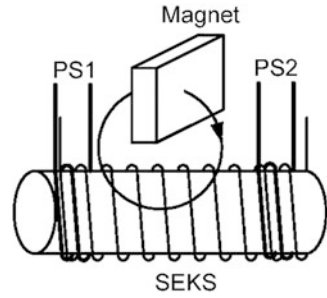


Fig. 3.37 AC field lines of the rotationally symmetrical PLCD sensor element—upper half (two-dimensional representation of simulation results of the finite element method)

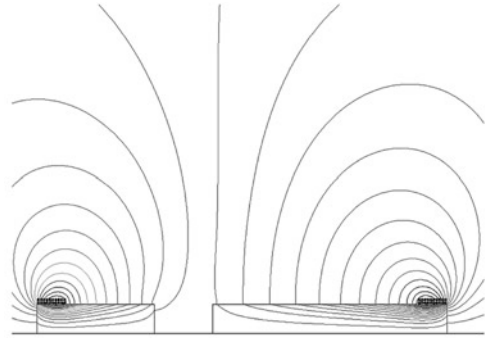
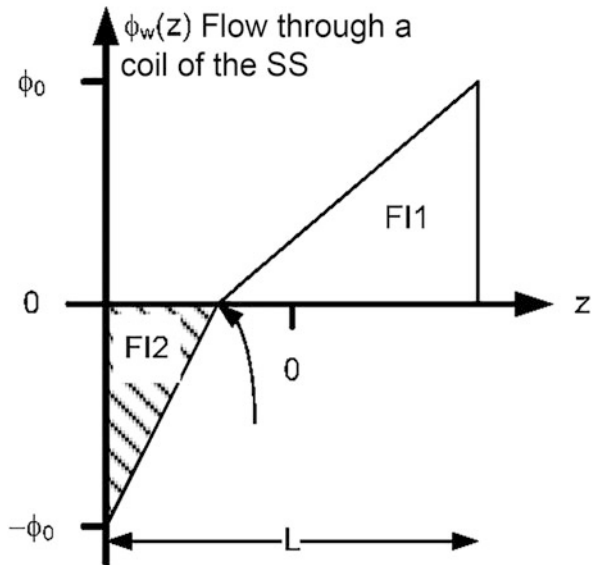


Fig. 3.38 Magnetic AC flux through one turn of the secondary coil along the longitudinal axis of the PLCD (depending on the x-target position)



Realization

The sensor market shows miniaturized PLCD systems for small to large measuring ranges and with integrated, but also with separate evaluation electronics. The typical external appearance of a PLCD is the cuboid design (Fig. 3.40), usually in a plastic housing. The electrical connections can be implemented both by cable outlet and by plug connectors.

Applications and Characteristics

The *compact, one-piece* PLCDs are characterized by *good properties* (Tables 3.7 and 3.8) and are often used in automotive engineering. They are efficient, robust, quite tolerant of fluctuations in the permanent magnet strength or the air gap between target and sensor, environmentally friendly (low power consumption) and can be used at high temperatures up to 160 °C.

Typical applications are:

- Piston monitoring in cylinders (electric, hydraulic, pneumatic),
- level monitoring,
- monitoring of valve positions,
- handling and assembly area,
- plant and process engineering,
- automotive applications, such as in:
 - master brake cylinders as brake assist,
 - clutch pedals,
 - geared or
 - automatic gears as “drive-mode-status” sensor.

3.1.1.8 Non-contact Magnetoinductive Displacement Sensors (Smartsens-BIL)

This section introduces an innovative IS: a new, slim, cuboid sensor with fully integrated evaluation electronics and with a permanent magnetic encoder whose position is converted into an absolute standard signal proportional to the displacement.

In principle, the *two-stage interaction mechanism* still applies. The position sensor acts locally on a soft magnetic foil, which acts as the core of a planar coil excited by a high-frequency current. This causes a position-dependent change in the coil inductance, which in turn is evaluated by the sensor electronics. The main difference and advantage of the BIL displacement sensors compared to other known IS are the sensor principle and design. The implementation of a patented new principle enables a highly flexible design (e.g. for linear and rotary movements) and miniaturised versions. The system has a high tolerance to the characteristics of the encoder (e.g. geometry, configuration, orientation and strength of the magnetic field). The sensor is characterized by a high sensitivity in the measuring direction with low cross-sensitivity.

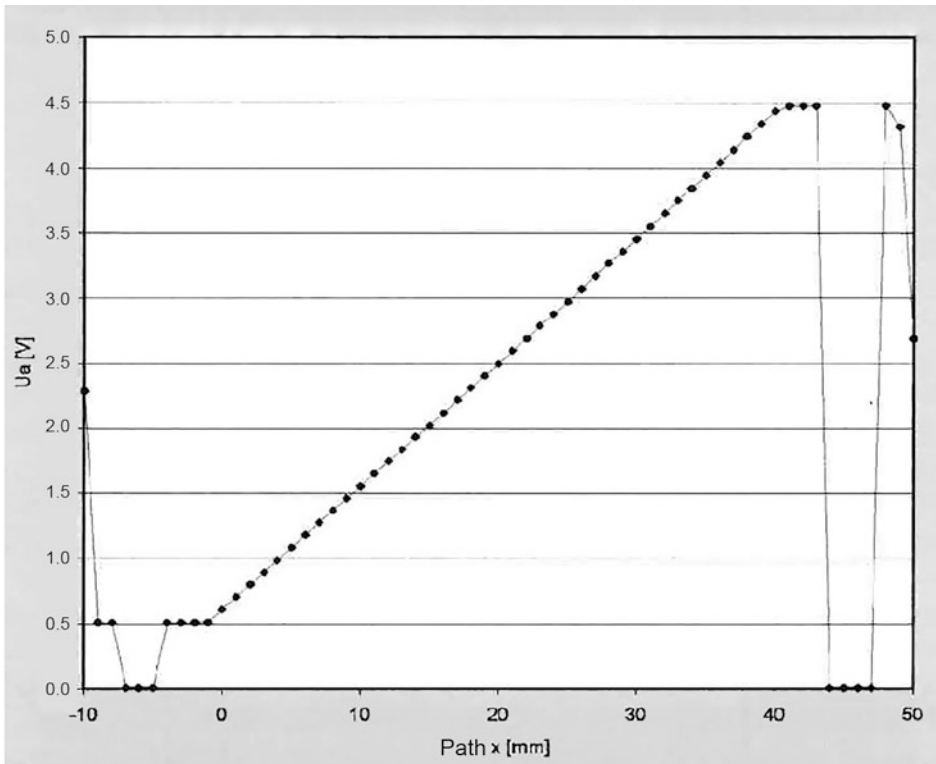


Fig. 3.39 Characteristic curve of a ratiometric PLCD sensor at 5.0 V supply. The additional end-of-range function provides 0 V alarm values outside the measuring range

Sensor Element and Measuring Principle of BIL Sensors

The sensor element again consists of two parts: A planar coil and a soft magnetic core in the form of two thin foils covering the top and bottom sides of the coil (Fig. 3.41a).

The conversion of the geometrical measured variable (displacement) into a primary electrical variable (inductance), which is evaluated by the sensor electronics and converted into an electrical sensor output signal, is carried out in two steps:

1. Effect of the magnetic encoder on the film core: *partial saturation*. The material used for the film core is an amorphous, soft magnetic material with a low coercivity ($H_C < 1$ A/m) and a very high initial permeability (about 25,000). With a suitable selection of the sensor, a saturated zone is created in the film. This zone reduces the effective foil area and thus the inductance of the sensor element compared to the case without magnet (Fig. 3.41b).

From the results of a magnetostatic field simulation for a foil (91 mm \times 10 mm \times 25 μ m) with a hard ferrite magnet (remanence 365 mT, coercive field strength 175 kA/m,

Table 3.7 Summary of solenoid arrangements and corresponding applications

Arrangement with:	Advantages	Typical applications
Parallel course of the DC field lines to the longitudinal axis	<ul style="list-style-type: none"> • Allows the largest air gap, • Has a low transverse distance dependence, • Requires an uncritical adjustment, • Is torsionally invariant when using magnetic discs 	<ul style="list-style-type: none"> • Flow measurement • Measured value acquisition • Replacement of complicated mechanical solutions • Control of lifts and conveyor systems • Direct position control • Cheaper alternative solution • Piston/valve position, positioning systems • Control through partition walls • Use in “harsh environment”
Vertical course of the DC field lines to the longitudinal axis	<ul style="list-style-type: none"> • Has the smallest expansion perpendicular to the direction of movement, • Is suitable for small magnets, which may also be mounted on a steel surface. 	<ul style="list-style-type: none"> • KWZ: Automatic level control, active chassis • Contactless, wear-free • Easy to adapt
Vertical course of the DC field lines to the longitudinal axis and closed magnet yoke	<ul style="list-style-type: none"> • Is characterized by low magnetic interference sensitivity and low stray field, • Allows the best utilization of the measuring length, • Is suitable for small magnets, which can also be mounted on a steel surface. 	<ul style="list-style-type: none"> • Replacement of linear potentiometers • Contactless, wear-free • Use in “harsh environment” • Line recorder • Operationally safe, maintenance-free • Unlimited life
Axial course of the DC field lines (ring magnet around the sensor)	<ul style="list-style-type: none"> • Is an absolutely twist-free arrangement • Has an uncritical adjustment. 	<ul style="list-style-type: none"> • Level measurement • Contactless, wear-free • High resolution • Independent of the medium

diameter 10 mm, length 10 mm, distance to the foil 3 mm), the saturated zone with a length of about 30 mm can be recognized (Fig. 3.42).

2. Effect of the film core on the spool: *path dependent change of the coil inductance*. The coil with foil core is excited by an alternating current with a frequency in the MHz range

Fig. 3.40 Series of PLCDs for general displacement measurements (Factory photo: Tyco)



Table 3.8 Technical data of PLCDs

Characteristics (electrical or mechanical)	Value, value range
Linearity range	Up to over 400 mm
Linearity error (trend line: Regression line)	Typical $\pm 1\%$ to 2% from the upper limit
Repeatability (uni- or bidirectional approximations)	Between $\pm 5 \mu\text{m}$ and $\pm 15 \mu\text{m}$
Resolution limit	Typically $\pm 0.1\%$ of the upper limit or 10 bits
Control field strength	20 mT to 30 mT
Actuation distance	Up to approx. 20 mm
3 dB cut-off frequencies	Up to 1 kHz
Temperature drift	Typical $\pm 2\%$
Protection class	IP69K
Operating temperature range	$-40 \text{ }^\circ\text{C}$ to $+150 \text{ }^\circ\text{C}$

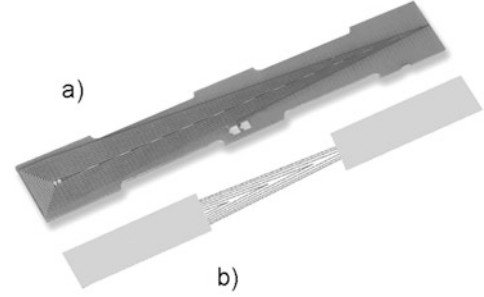
and generates a magnetic field, which is almost perpendicular to the coil plane in the air space between the foils and in the plane in the foil.

In order to convert the position of the magnet into a change in inductance and thus obtain position information, a position-dependent function must be built into the sensor element. For this purpose, the coil geometry is varied along the measurement direction, for example in the form of a triangle with several windings inside each other (Fig. 3.43).

If we consider an idealized model consisting of a coil with a single triangular winding or of two core foils completely covering the coil (no magnet, \rightarrow no local saturation \rightarrow no virtual window), the magnetic field strengths in the area A_a and in the area A_i , outside or inside the triangular winding (Fig. 3.43), neglecting the stray fields outside the foils, can be analytically calculated with the aid of Maxwell's equations. This results in the inductance of such an arrangement:

Fig. 3.41 BIL sensor element

(a) Photo of the planar coil without soft magnetic foil; (b) Simulation model (saturated zone shown as a virtual, transparent window)



$$L = \frac{\mu_0 \cdot A_i}{h \cdot (1 + A_i/A_a)}, \quad (3.17)$$

where μ_0 is the magnetic field constant and h is the core foil distance (Fig. 3.44).

The spatial partial saturation of the core foils by the magnet opens the window area (saturated zone) in the foils, which has a direct effect on the coil inductance. The wider the coil below the magnet and the saturated zone, the more the inductance L is reduced compared to the case without magnet, since wide areas of the coil contribute more to the inductance L than narrow ones. For the idealized model, this means a reduction in the effective foil area (hatched area in Fig. 3.43) and thus in the areas A_a and A_i .

Neglecting the minimum inductance in the saturated zone results in a calculation formula for the coil inductance, which depends only on the geometric dimensions of the rectangular core foil (index F), the triangular coil (index S) and the rectangular saturated zone (index Z) or the position x of the saturated zone along the measuring direction:

$$L = \frac{\mu_0}{h} \cdot \frac{(L_F - L_Z) \cdot B_F \cdot A_i - A_i^2}{(L_F - L_Z) \cdot B_F}, \quad (3.18)$$

where the inner surface A_i consists of two parts and depends on the current position x :

$$A_i = \frac{L_S \cdot B_S}{2} - \frac{L_Z \cdot B_S}{L_S} x. \quad (3.19)$$

The quadratic polynomial function $L = F(x)$, which describes the sensor element displacement dependence, can be analyzed. The derivatives of this function ultimately lead to the correct constructive sensor element design to meet the main requirements: Unconditional unambiguity (monotony of the function over the entire measuring range) and optimum linearity.

In reality, the image coils consist of several triangular windings, one inside the other, whose geometry is designed according to the above procedure. 3D simulations enabled further optimizations regarding the coil travel $L = F(x)$.

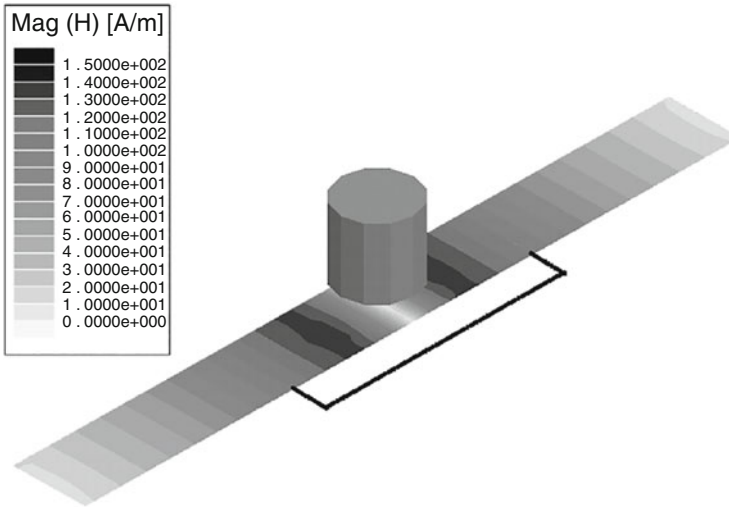
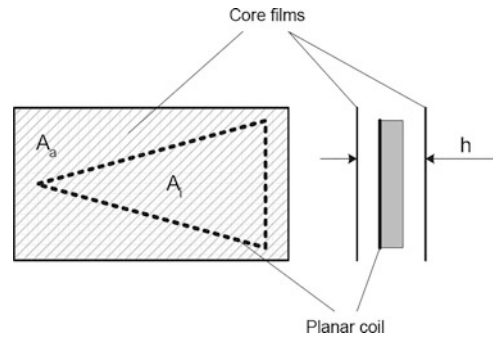


Fig. 3.42 Spectral representation of the DC magnetic field strength in the foil (three-dimensional representation of simulation results with finite elements)

Fig. 3.43 Schematic representation (top view and cross section) of the idealized model



Function and Block Diagram of the BIL Displacement Sensor

Figure 3.45 shows the block diagram of the BIL displacement sensor. The first main function of the front end electronics is to excite the planar coil of the sensing element and simultaneously measure its inductance. By using a *differential method* with two-coil push-pull oscillator, the measuring effect is *doubled*. At the same time, common-mode interference, such as changes in ambient temperature, is suppressed.

The output signal of the oscillator is rectified in the back-end section by means of a peak detector. The signal obtained in this way passes through two amplifiers with adjustment options for offset and gain and is then converted in the output stage into two standardized sensor output signals (current and voltage). Calibration of the sensor by iterative zero and endpoint adjustment can be performed manually by two miniature potentiometers or by electronic teach-in procedure with an adjustment IC.

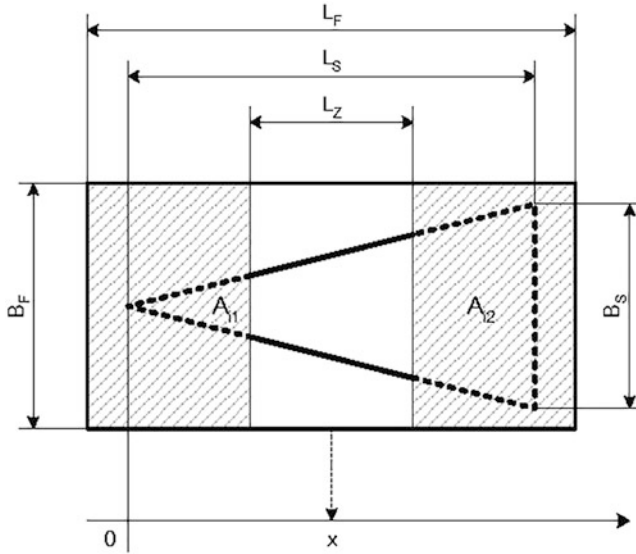


Fig. 3.44 Considering the saturated zone for the model in Fig. 3.43

The use of the two-coil push-pull oscillator requires that the sensor element is realized by two triangular, spatially opposite coils L_1 and L_2 . These are printed on both sides on the FR-4 carrier board of the sensor element (Fig. 3.46; L_1 : above and L_2 : below).

A particularly simple, proven method is used to evaluate the coil inductances. It uses the properties of a detuned resonant circuit bridge and supplies an *analog, time-continuous voltage* that depends only on the *ratio of two inductors*. For this purpose, the sensor coils and two capacitors are connected to form a bridge circuit (Fig. 3.47).

The *transfer function* of the oscillating circuit bridge is

$$\underline{A} = \frac{V_{\text{OUT}}}{V_{\text{IN}}} = \frac{Z_1 Z_4 - Z_2 Z_3}{(Z_1 + Z_2)(Z_3 + Z_4)} \quad (3.20)$$

with the impedances Z_1, Z_2, Z_3 and Z_4 . Here R_1 and R_2 are the loss resistances and L_1 and L_2 ARE the inductances of the coils, C_1 and C_2 ARE the capacitances of the capacitors. The losses of the capacitors can be neglected compared to the coil losses.

For coils of high quality and low losses, the real part A_R of the bridge transfer function is approximately only dependent on inductances or capacitances:

$$A_R \Re \left(\frac{V_{\text{OUT}}}{V_{\text{IN}}} \right) = \frac{1}{2} \left\{ \frac{L_1 - L_2}{L_1 + L_2} \right\} + \frac{1}{2} \left\{ \frac{C_1 - C_2}{C_1 + C_2} \right\}. \quad (3.21)$$

An asymmetry of the oscillating circuit bridge can be achieved by using different capacitors. If the detuning is sufficiently large, A_R is so large that the imaginary part A_I can be neglected. The transfer function thus becomes real. It does not depend on the frequency and not on the strongly temperature-dependent loss resistances of the coils.

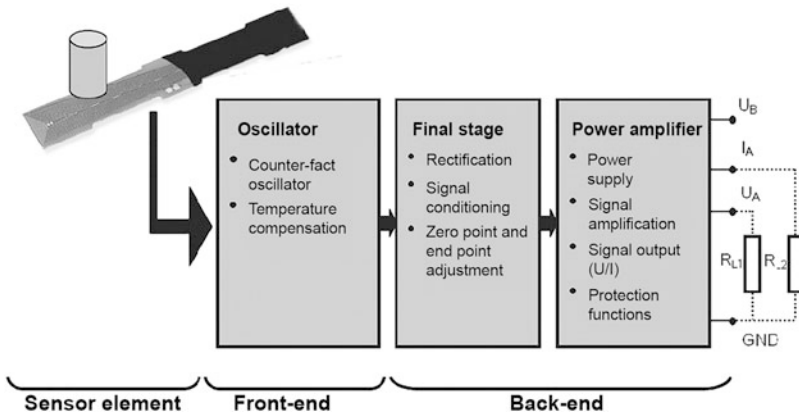


Fig. 3.45 Block diagram of the BIL displacement sensor

The direct determination of the real transfer function as the ratio of two voltages is either complex or inaccurate. A simple method for precise evaluation is to keep the input variable V_{IN} constant by means of a closed-loop control. Then the output variable V_{OUT} is proportional to the transfer function:

$$V_{OUT}(s) = V_{IN} A_R(L_1(x), L_2(x)), \quad (3.22)$$

where $L_1(x)$ and $L_2(x)$ describe the dependencies of the sensor element coil inductances L_1 and L_2 on path x .

The excitation of the oscillating circuit bridge can be done by an external oscillator. However, it is more advantageous if the oscillating circuit bridge itself is used as the oscillating circuit of an oscillator. The oscillator shown schematically in Fig. 3.48 is designed as a push-pull oscillator. This has the advantage that the output voltage is V_{OUT} referred to ground. The oscillator amplitude V_{IN} is kept constant by a control system consisting of a subtractor, a peak detector and a PI controller. The constant value of the bridge input voltage V_{IN} is set by the reference voltage V_{REF} . If the input voltage undergoes a slight change, the PI controller readjusts the control of the push-pull current sources for the excitation of the resonance bridge in such a way that the voltage V_{IN} is readjusted to the original constant value.

Realization

The patented BIL process enables a very reliable design suitable for series production: For example without wound wire coils and without brittle ferrite cores. The inner workings of a BIL sensor consist of two printed circuits: sensor element and evaluation electronics (Fig. 3.49). The assemblies are vertically butt soldered; this creates both the mechanical fixation and the electrical connections sensor element to evaluation electronics (minimum

Fig. 3.46 Front side of the BIL sensing element with coil L_1 and terminal pads (L_2 is on the back side)

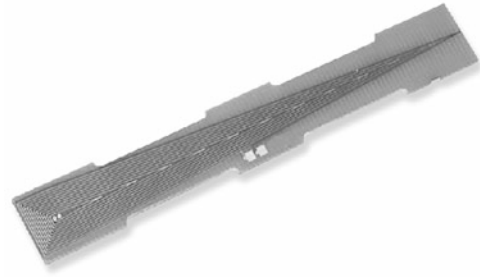


Fig. 3.47 Oscillating circuit bridge

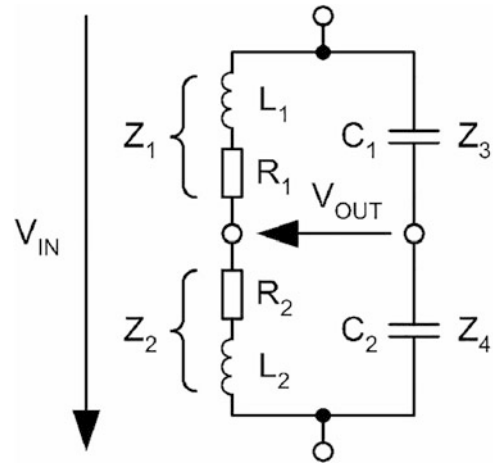
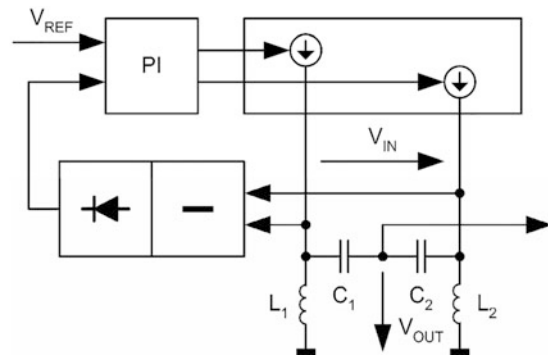


Fig. 3.48 Push-pull oscillator with oscillating circuit bridge and amplitude control



number of only three contacts required). The resulting assembly is installed in a plastic trough (sensor housing lower part). A cover with integrated fixing sockets and plug closes the housing. The components are fixed by casting. A groove in the cover forms the electrical and mechanical zero point.

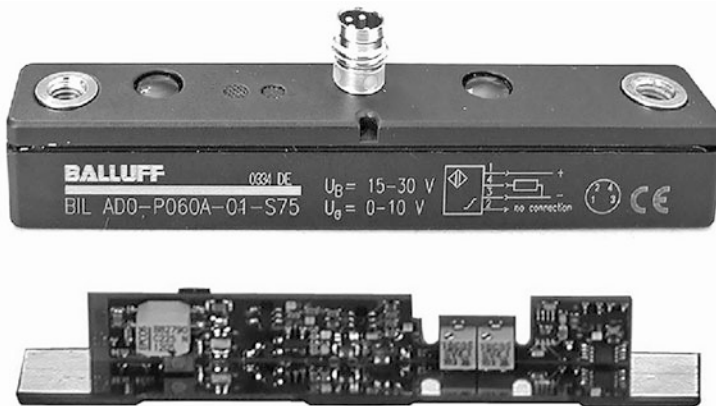


Fig. 3.49 Magneto-inductive BIL displacement sensors and subassembly and reference sensors (Factory photo: Balluff GmbH)

After the final assembly, the iterative calibration and final inspection takes place (Fig. 3.50).

Applications

The spectrum of typical applications ranges from *handling* and *robotics* to *materials handling* and *building services engineering* to *dosing* and *flow measurement tasks*.

A thorough analysis of the application requirements shows a high integration potential of this sensor. If the sensor is considered as a system consisting of the three components: position encoder, sensor (hardware) and evaluation method (software), the following four application-oriented integration levels can be identified—beginning with standard components and ending with application-specific designs of these three components (Table 3.9).

BIL Main Characteristics

The BIL displacement sensor family meets all requirements of modern robust industrial sensor technology. The most important characteristics are summarized in Table 3.10.

3.1.2 Optoelectronic Distance and Displacement Sensors

3.1.2.1 Overview

Optoelectronic distance and displacement sensors detect distances or distances with the aid of *visible* (usually red light) or *infrared light*. Light has many properties, such as straight-line propagation, high propagation speed or interference. It can be generated, shaped, guided and detected with simple means. This results in a multitude of possibilities for measuring distances and paths (Fig. 3.51).

BALLUFF

IT IS FASCINATING WHAT PRECISION
SENSOR TECHNOLOGY MAKES
POSSIBLE.



B *innovating automation*

**Outstanding precision results in
extraordinary technology.**

www.balluff.com

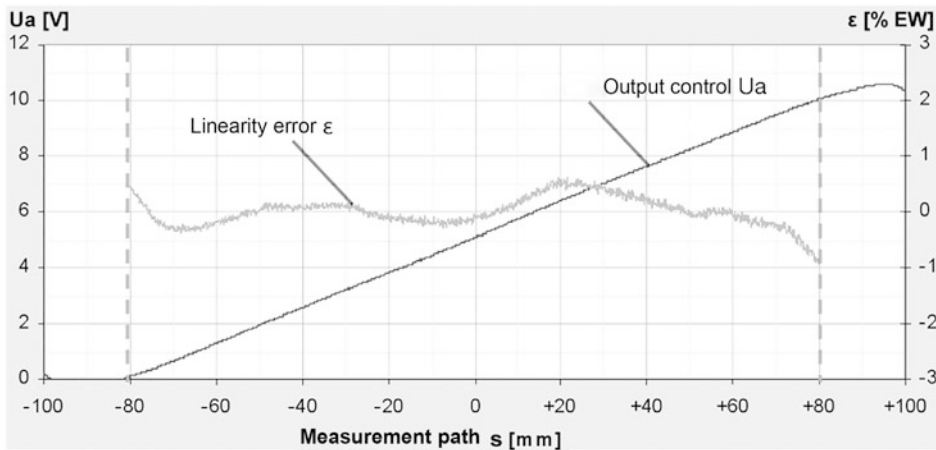


Fig. 3.50 Characteristic output curve and linearity error curve of a BIL displacement sensor

Table 3.9 Variation options for BIL sensors

Level of integration	Sensor	Evaluation	Encoder
1	Standard	Standard	Standard
2	Standard	Standard	Adapted
3	Standard	Adapted	Adapted
4	Special	Adapted	Adapted

Table 3.10 Technical data of BIL sensors

Characteristics (electrical or mechanical)	Value, value range
Linearity range	60 mm to 160 mm
Linearity error (trend line: Regression line)	$\leq \pm 0.3$ mm to ± 2.4 mm
Repeatability (uni- or bidirectional approximations)	$\leq \pm 30$ μ m to ± 500 μ m
Temperature coefficient	Typ. $+4$ μ m/K to -40 μ m/K
3 dB cut-off frequencies	1.5 kHz to 300 Hz
Measuring speed	≤ 5 m/s
Protection class	IP67
Operating temperature range	-10 $^{\circ}$ C to $+70$ $^{\circ}$ C

3.1.2.2 Optoelectronic Components

Semiconductor components are generally used to generate and detect the light (e.g. light-emitting diodes, laser diodes and photodiodes).

Light emitting diodes (LED) are used in large numbers as light sources. The diodes are operated in forward direction. Light is emitted in a certain spectral range; either red (visible) or infrared (invisible) light. In general, the diodes are encapsulated in a plastic

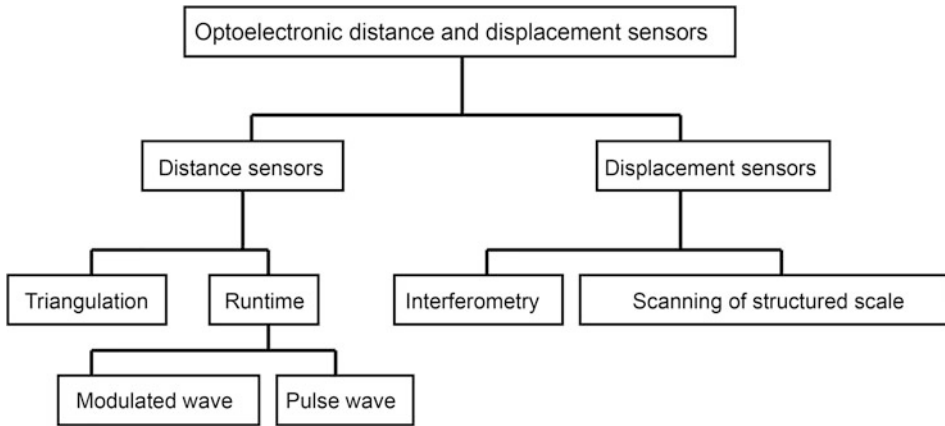


Fig. 3.51 Overview of the different measuring principles

housing. The surface of the housing acts as a lens that focuses the light into a beam cone and thus has a directional effect. Figure 3.52 shows two typical designs.

Due to the high requirements, distance sensors are increasingly equipped with semiconductor laser diodes. With the bundled light beam of the laser diode, objects can be detected with high precision. *Red laser light* has advantages over LED light in the alignment of the sensors, as the human eye is able to see the light clearly over long distances due to the focusing.

Semiconductor laser diodes have the structure shown in Fig. 3.53(a and b).

The core piece is a crystal (laser chip) with differently doped layers. The *laser crystal* is mounted on a metal block (heat sink), which dissipates the heat generated. When a forward voltage is applied, a current flows perpendicular to the layers, which generates *photons* (light quanta) as in an LED. When a current threshold is exceeded, light amplification occurs in the highly doped barrier layer through *stimulated emission*. The light wave running in this layer is partially reflected at the polished and mirrored ends. The reflected wave is amplified again in the layer, which produces the actual laser light—radiation of *one wavelength* and *one phase*. This light emerges at the front face (typically $1\ \mu\text{m} \times 3\ \mu\text{m}$) with *high luminance*. In the opposite direction, the laser light radiates onto a monitor diode, which continuously measures the luminous flux and thus allows the laser chip current to be regulated in order to keep the radiated power constant.

The laser beam exits in the main direction through the glass window attached to the housing and has an aperture angle of typically 10° to 30° with commercially available laser diodes. In order to obtain parallel or focused radiation, precise *aspherical optics* are required, which must be adjusted to the laser diode with small tolerances.

Due to the high luminance in the optically bundled beam path, the necessary measures for *personal protection must be* observed when handling laser sensors and appropriate caution is required. Laser devices are divided into different classes according to EN

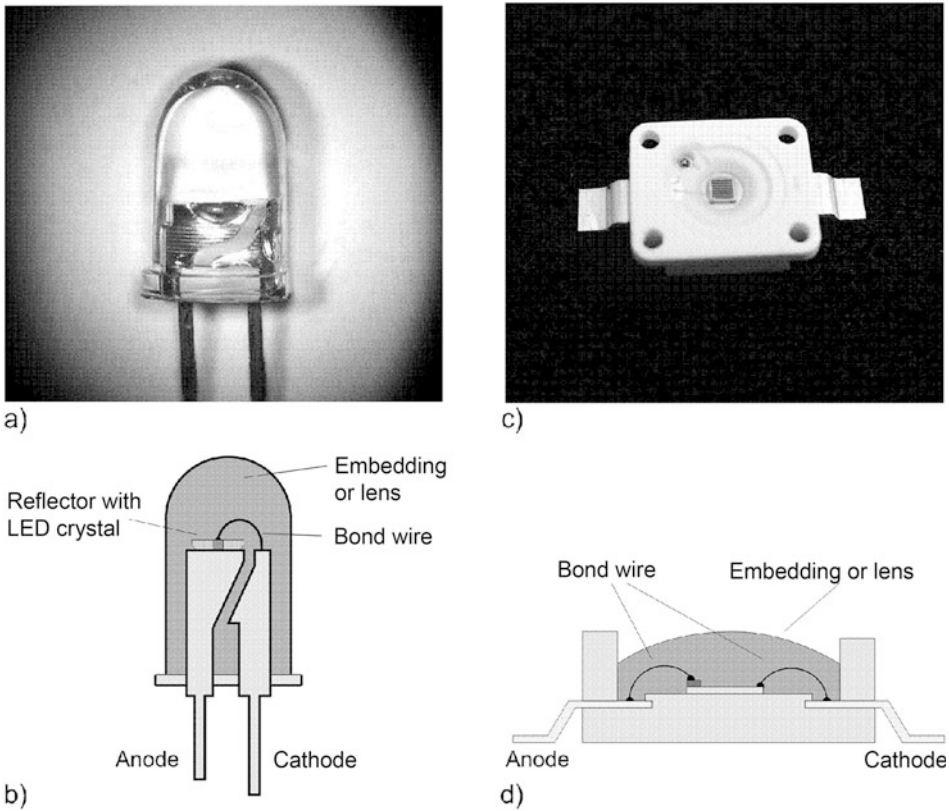


Fig. 3.52 (a) Realisation of a wired, so-called T-type LED and (b) schematic representation and (c) realisation of an SMD LED for direct mounting on the surface of the printed circuit board (SMD: Surface Mounting Device) and (d) schematic structure (Factory photos: Balluff GmbH)

60825–1 (safety of laser devices). Optical distance sensors commonly available on the market are usually devices of laser classes 1 or 2. EN 60825–1 defines the maximum permissible limit values of the emitted light power as well as the required protective measures and markings of these devices, as shown for example in Fig. 3.54. The classification of LED devices into different hazard classes is based on EN 62471 (photobiological safety of lamps and lamp systems).

Semiconductor components are used to detect the light. There is a multitude of types and designs.

In the simplest case, they are photodiodes, i.e. pn-junctions, which are polarized in the reverse direction. A photocurrent corresponding to the incidence of light is generated in these. This is then converted into a voltage proportional to the current, amplified and evaluated.

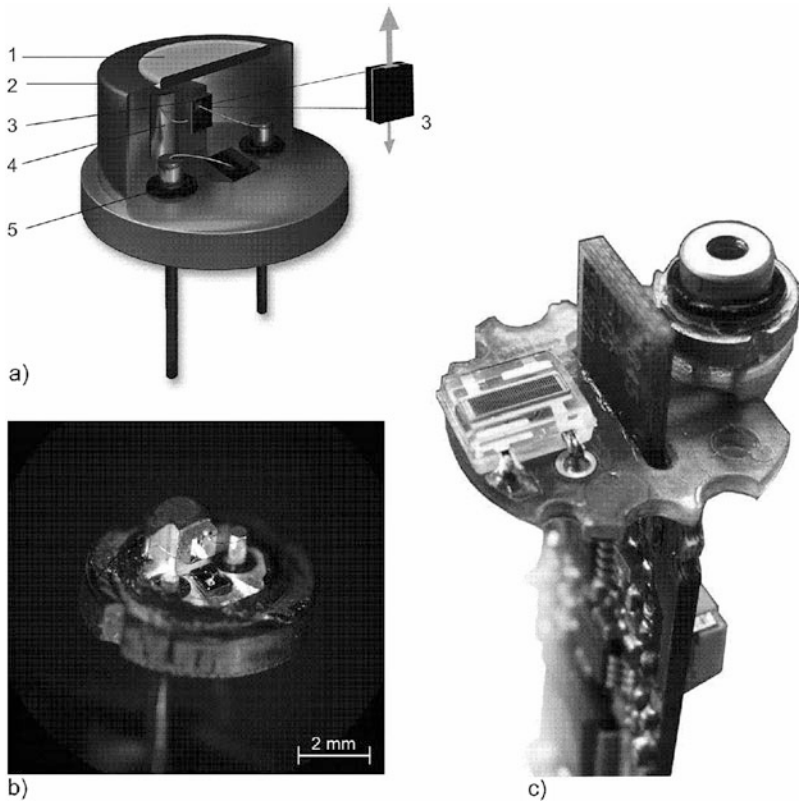


Fig. 3.53 (a) Schematic structure of a laser diode, 1: glass window, 2: housing, 3: laser chip, 4: heat sink, 5: monitor diode, (b) laser diode with removed housing and (c) laser diode and PSD element on a printed circuit board head of a distance sensor, triangulation method (Factory photos: Balluff GmbH)

Avalanche photodiodes (avalanche effect diodes, or APDs) are particularly suitable for *highly sensitive* measurements at *high modulation frequencies*, such as those required for time-of-flight measurements. A high internal amplification is achieved by accelerating the charge carriers released by the light by a high voltage in such a way that further electron-hole pairs are generated like an avalanche.

A PSD *element* (PSD: Position Sensitive Device) is a lateral effect diode with an extended light-sensitive area (Figs. 3.53c and 3.55). The light beam impinging on this surface generates a total current I in the diode, which is split into two partial currents I_1 and I_2 . The ratio of these partial currents is determined by the position x of the center of gravity of the light distribution. If the active surface has a total length L , THE FOLLOWING applies:



Fig. 3.54 Optoelectronic distance sensor of laser class 2 according to the pulse transit time method (front dimensions 35×70 mm, factory photo: Balluff GmbH)

$$\frac{I_1 - I_2}{I_1 + I_2} = \frac{L - 2x}{L}. \quad (3.23)$$

By measuring the partial currents, the position of the centre of gravity can be determined. Due to the ratio formation, the result is *independent of the incident light intensity*, i. e. of $I_1 + I_2$ and thus of the reflectivity of the object surface. In contrast, the spatial distribution of the reflection is taken into account. The lateral effect diode reacts to the center of gravity of the light and can therefore be used as a sensor element for *triangulation sensors*.

CCD lines (CCD: Charge Coupled Device) are also used in triangulation sensors and consist of a large number of photodetectors arranged in a row (Fig. 3.56). Each detector is assigned a capacity. A charge separation takes place in the photodetectors through the

influence of light. The released electrons charge the assigned capacity. The charge packets are transferred to an analog shift register by a control pulse. The charge distribution in the shift register corresponds to the intensity distribution of the light along the CCD line, which is incident on the line during the exposure time.

The control signal also discharges the detectors. The CCD line is then ready for a new exposure. Parallel to this, the shift register is used to transport all charge packets from one element to the next in a stepwise, charge-coupled manner. The charge distribution is fed to an output amplifier. A downstream microcontroller analyses the light distribution.

3.1.2.3 Optical Principles of Distance Sensors

The sensors emit visible, generally red light or invisible, infrared light (Fig. 3.57). This light is reflected by the object and detected by the sensor. The object distance d is determined from the collected information, as Fig. 3.57 shows.

The sensors available today are optimized for different applications due to their design and correspondingly selected electronics, for example, high range or small blind zone, high repeatability or high switching frequency, detection of small parts or insensitivity to dirt and ambient light.

Useful Signal

The receiving element converts the incoming light into an electrical signal S_{el} . This signal is composed of the useful signal S_{useful} reflected by the object, the signal S_{external} generated by ambient light or extraneous light (e.g. neighboring sensor using the same pulse frequency) and the internal optical cross-talk signal $S_{\text{Ü}}$ (Fig. 3.58):

$$S_{el} = S_{\text{useful}} + S_{\text{external}} + S_{\text{Ü}}. \quad (3.24)$$

Contamination Indicator

Optoelectronic distance sensors often have a so-called *contamination indicator*: either an indicator LED or a switching signal. This indicates whether the useful signal can be evaluated, i.e. whether the useful signal lies between the noise component and the overloading of the evaluation electronics. Overdriving can occur with reflective objects or excessive ambient light. Too little light can be caused by weakly reflecting objects, contamination of the optics or the light path (e.g. dust, fog).

The level of the useful signal is influenced by the transmission power and the geometrical arrangement of the beam path and on the other hand by the object distance, the surface condition of the object and the object geometry.

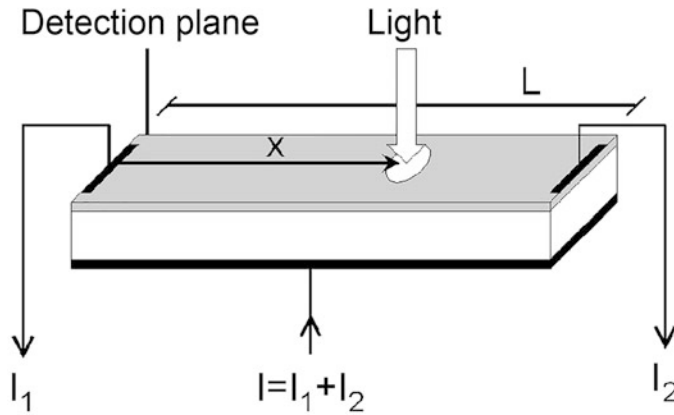


Fig. 3.55 Schematic structure and function of a PSD element (L : length of active area, x : position of the center of gravity of the incident intensity distribution)

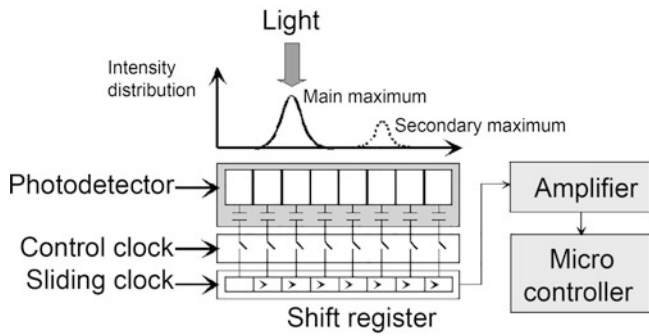


Fig. 3.56 Schematic structure and function of a CCD line

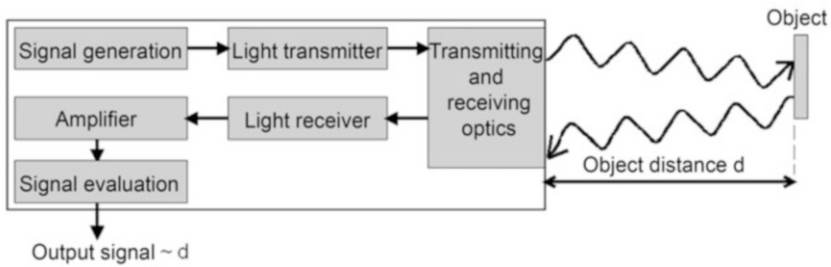


Fig. 3.57 Principle diagram of the triangulation, pulse, phase and frequency measurement method

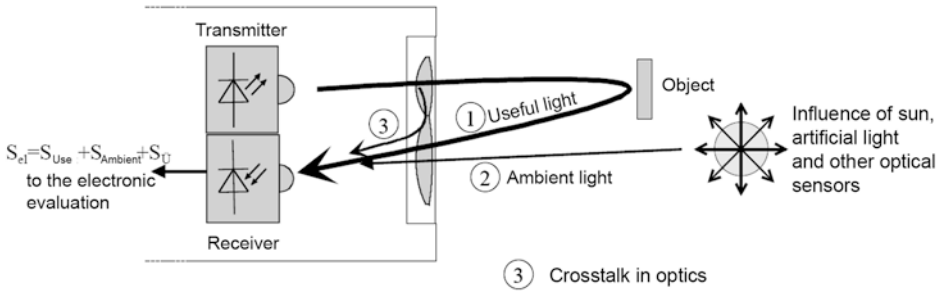


Fig. 3.58 Electrical signal S_{el} and its components

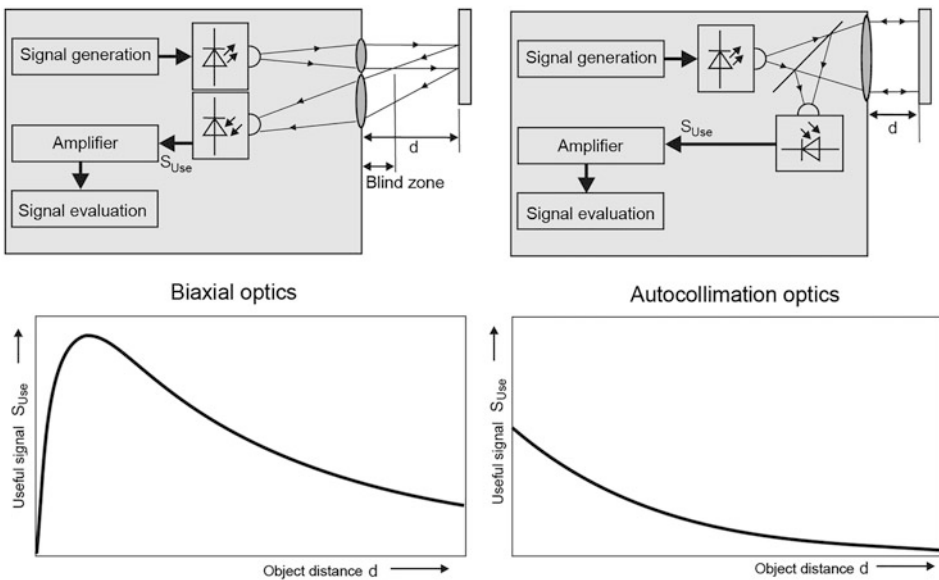
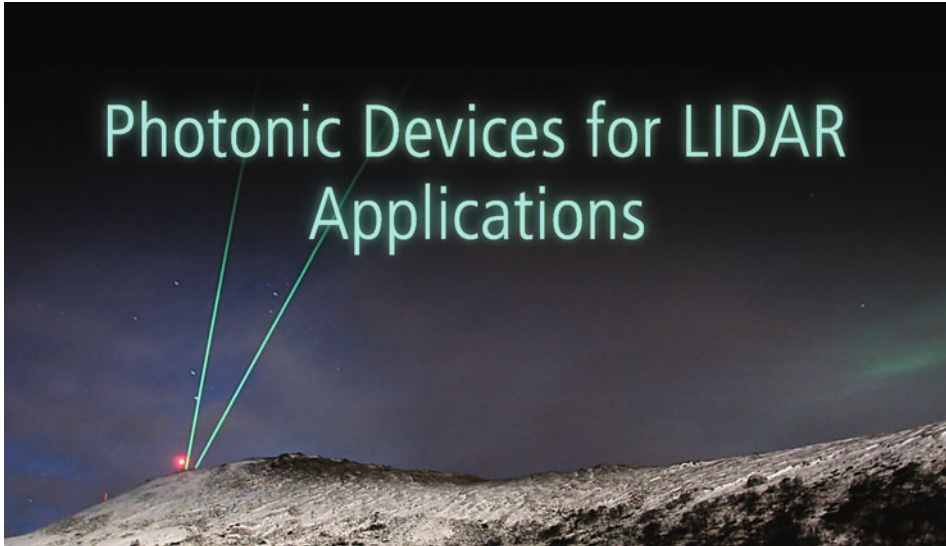


Fig. 3.59 Schematic representation of the beam path and the useful signal path at the receiver for biaxial optics and autocollimation optics

Beam Path

For the beam path, either *biaxial optics* or *autocollimation optics* are used (Fig. 3.59). In biaxial optics, the optical axes of the transmitting and receiving beam are separated from each other around the base B and usually tilted against each other in order to optimize the distance-dependent energy at the receiving element. Figure 3.59 shows the distance-dependent useful signal component reflected by the object. At close range, little energy reaches the receiver, so that a blind zone is created.

In autocollimation optics, the transmitting axis is identical to the receiving axis, which is usually realized by a beam splitter. This means that the receiver also receives light from the object at close range, i.e. no blind zone in front of the sensor. A disadvantage is that the beam splitter has a transmission of 50% and thus, compared to biaxial optics, the maximum ranges of the sensors are reduced.



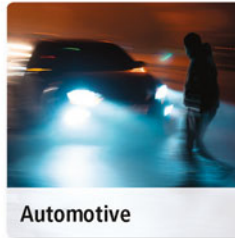
Hamamatsu Photonics, leading in cutting edge photon detectors and light sources for LIDAR and measurement applications



Consumer

Short Range (10 m)

- Near Infrared LED
- Si PIN Photodiode with COB Package
- Si PIN Photodiode Array
- CMOS Distance Linear Image Sensors



Automotive

Mid-Range (150 m)

- Pulsed Laser Diodes
- Si APD + TIA
- Si APD Array + TIA
- SiPM / MPPCs (Ceramic, Plastic and TSV package types)
- Si APD
- IR Enhanced Si APD



Industrial

Long Range (300 m plus)

- Pulsed Laser Diodes
- InGaAs APD + PIN Photodiode
- SiPM / MPPCs

HAMAMATSU
PHOTON IS OUR BUSINESS

www.hamamatsu.com

Influence of the Condition of the Object Surface

Different properties of the object influence the useful signal. These include the *size* and *curvature* of the *reflective surface*, its *reflectivity* and *reflectivity* and the *colour* of the object.

The larger the irradiated object surface, the larger the useful signal. Manufacturers usually specify a *minimum area* that should not be undercut for reliable detection (small parts detection). On the one hand, the reflection behaviour can be purely diffuse or contain a specular or shiny component, which may not reflect the light back to the sensor.

The *object color* is described by the electromagnetic reflection spectrum $R(\lambda)$. To estimate the influence of the object color on the wanted signal, the electromagnetic spectrum of the transmitter $S(\lambda)$ and the wavelength-dependent sensitivity $E(\lambda)$ of the receiver must be considered. The wanted signal is given by the overlap of all three spectra:

$$S_{\text{useful}} \propto \int S(\lambda) \cdot R(\lambda) \cdot E(\lambda) d\lambda. \quad (3.25)$$

Figure 3.60 shows the typical spectrum $S(\lambda)$ of a red LED as well as the sensitivity $E(\lambda)$ of common photodiodes and the reflectivity $R(\lambda)$ of a blue and red surface. The blue object provides the smaller useful signal because it reflects less than 5% in the red wavelength range (625 nm to 740 nm).

If infrared LEDs (wavelength $\lambda > 800$ nm) are used, the sensitivity of the photodiodes increases and higher useful signals are obtained, which is noticeable at higher ranges. However, the adjustment of the sensors is more complex due to the non-visible beam.

Grey Value Shift

The manufacturer's specifications on the output characteristics of sensors usually refer to flat object surfaces, so-called *90% grey maps*, which reflect light diffusely with a reflectance of 90% over the entire electromagnetic spectrum. If objects with other degrees of reflection are detected, the output characteristic curves shift within the tolerance range specified by the manufacturer. The characteristic curves for objects with 18% or 6% reflectivity are specified as the lower limit. The ratio between the values at 90% and 18% or 6% is referred to as the *grey-scale value shift*.

3.1.2.4 Measuring Principle: Triangulation

The triangulation method is a purely *geometric measuring method*. The light generated by the transmitter (LED or laser diode) is formed into a narrowly defined light beam by the transmitter optics and strikes the object to be detected at distance d . The light is reflected from its surface according to the surface properties. The portion reflected in the direction of the receiving optics is imaged in the detection plane, where it produces a light distribution with the center of gravity position x . If the object moves, the center of gravity moves in the

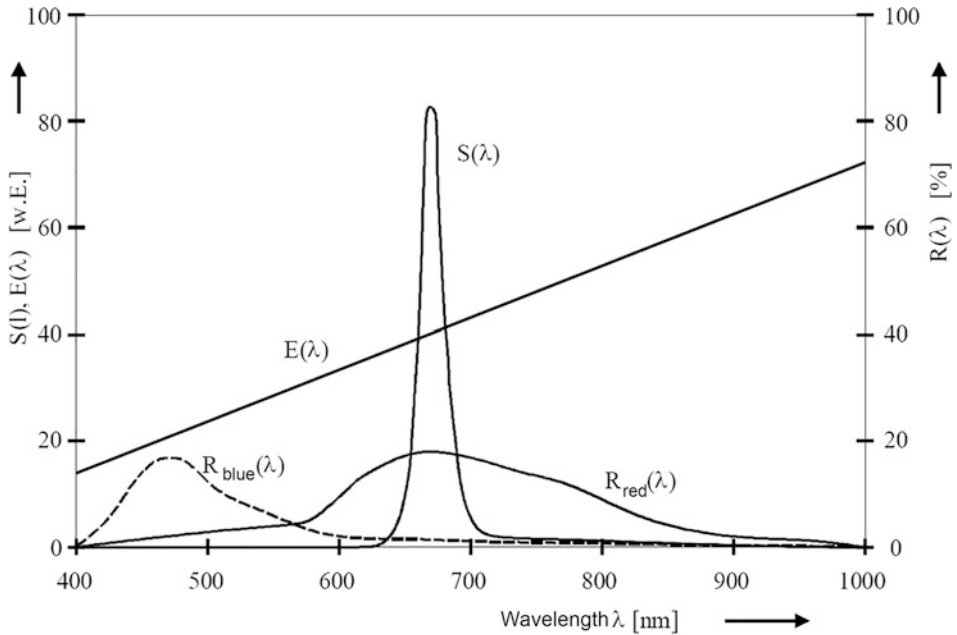


Fig. 3.60 Typical spectral distributions

detection plane. Measuring the point of incidence x in the detection plane results in the object distance d (Fig. 3.61). The following relationship applies:

$$x = \frac{B \cdot F}{d}, \quad (3.26)$$

where F is the distance between the detection plane and the optics and B is the base distance between the optical axis of the transmitting and receiving optics.

In general, position-sensitive semiconductor components, i.e. PSD elements or CCD lines, are used to determine the center of gravity x .

The output characteristic curve is linearized due to the non-linear relationship between the center of gravity x and the object distance d . The signal processing is then usually performed in a microcontroller.

In general, the *greater* the base distance B , the *higher the ranges* are possible. The maximum range is thus limited by the sensor size, among other things. Depending on the size, light type and beam shape, maximum ranges between 20 mm and 5000 mm are achieved.

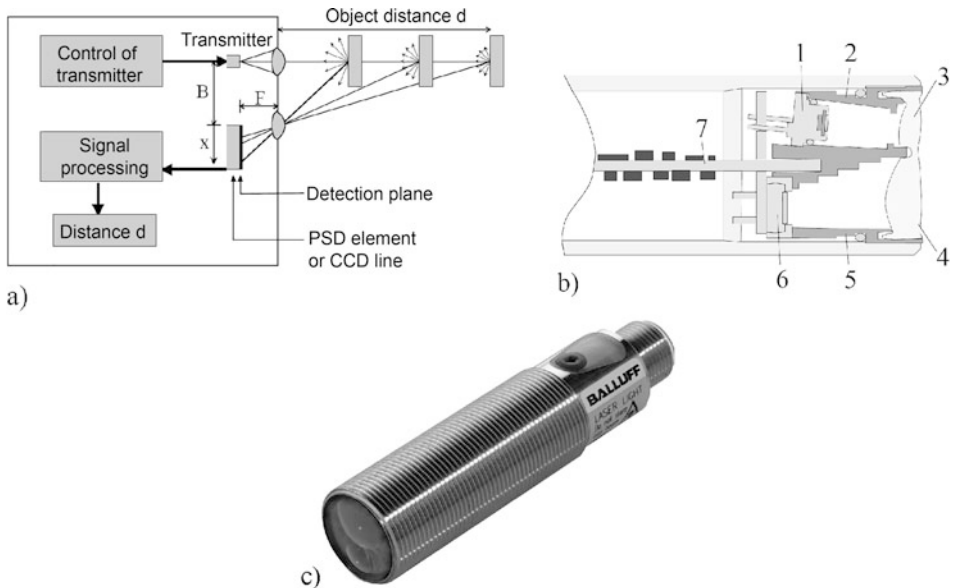


Fig. 3.61 (a) Schematic representation of the triangulation measurement principle (b) Cross section through the optical head of a cylindrical laser triangulation sensor with PSD element according to Fig. 3.61b) (1: laser diode, 2: tube for laser radiation, 3: collimator lens, 4: receiving lens, 5: tube with light traps, 6: PSD element, 7: circuit boards with electronic components) and (c) cylindrical laser sensor according to the triangulation method (length 70 mm) (Factory photos: Balluff GmbH)

3.1.2.5 Measuring Principle: Pulse Delay Method

With the pulse transit time method, the sensor sends out a light pulse and measures the time Δt that the light needs until it returns from the object. This time is a measure for the object distance d (Fig. 3.62a).

Using the speed of light c and the refractive index n of the medium passed through (usually air with $n = 1$), the object distance d is calculated with the formula

$$d = \frac{c \cdot \Delta t}{2 \cdot n}. \quad (3.27)$$

The resolution is essentially determined by the electronic measurement of time Δt . A displacement of the object by 10 mm means (in air with $n = 1$) a change of the running time of 66.7 ps.

In the course of time, various electronic methods have been developed for signal processing and time measurement. A frequently used method is shown in Fig. 3.62b. When the light pulse is emitted, the electronic clock is triggered by an internal start signal. When the light pulse arrives at the receiver, the clock is read out. Signals

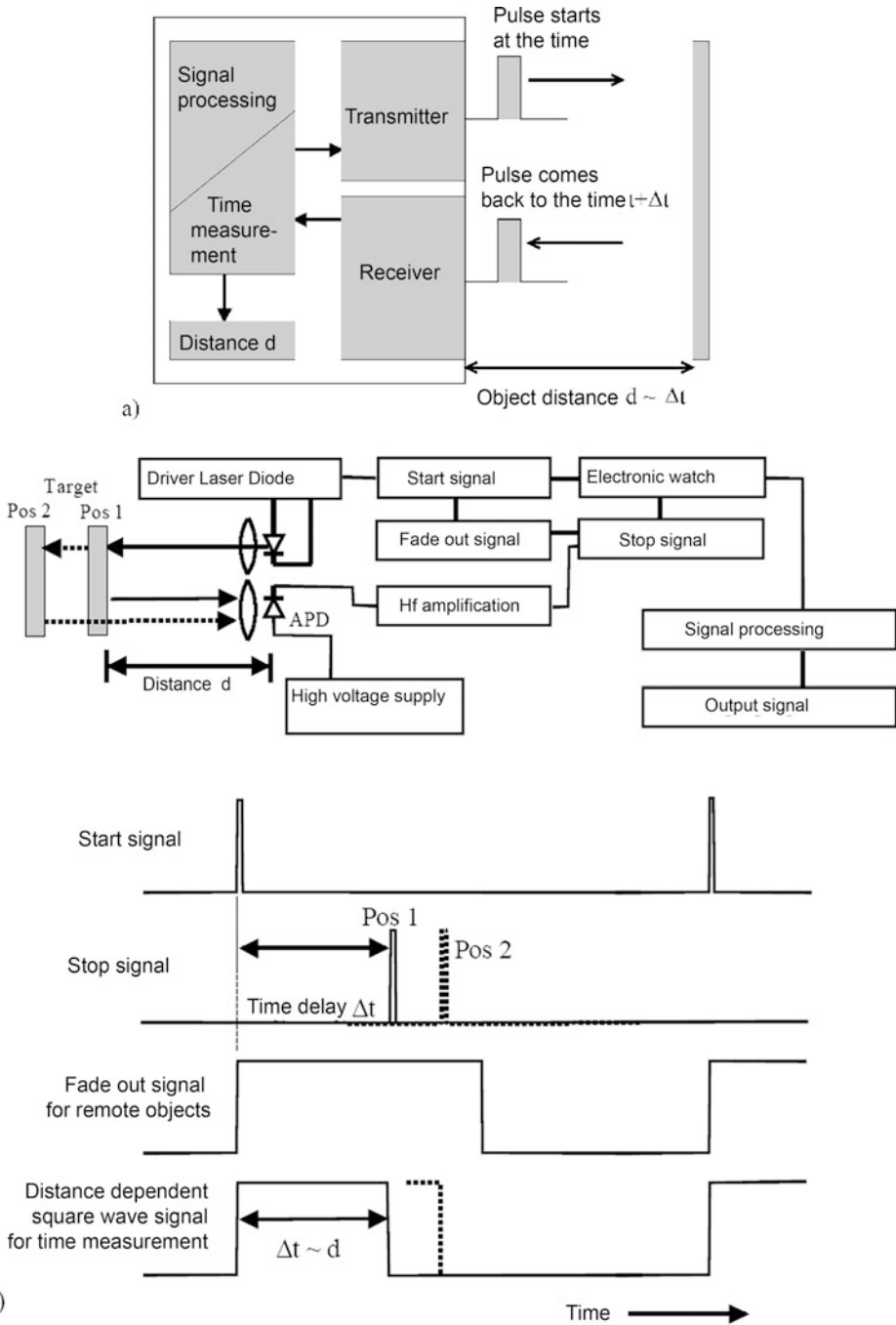


Fig. 3.62 (a) Principle of the pulse delay method and (b) functions and signals in the pulse delay method

coming back from objects further away are electronically suppressed. For pure time measurement, digital and analogue methods are used, as known from radar technology.

A certain *minimum light intensity* at the receiver is necessary for correct operation. This limits the maximum range. With the pulse transit time method, ranges of up to 6 m are achieved with diffusely reflecting surfaces and up to several 100 m with reflecting surfaces and reflectors.

3.1.2.6 Measuring Principle: Phase or Frequency Delay Method

The phase or frequency delay procedure is also a delay procedure in principle. The intensity of the light is measured with the period T_m or modulation wavelength λ_m , with

$$\lambda_m = \frac{c \cdot T_m}{n} \quad (3.28)$$

amplitude or frequency modulated. The object distance d is determined by measuring the phase difference $\Delta\varphi$ or frequency difference Δf between emitted and received light. For the *phase delay method*, the following results are obtained:

$$d = \frac{\Delta\varphi}{2 \cdot \pi} \cdot \frac{\lambda_m}{2} + i \cdot \frac{\lambda_m}{2}, i = 0, 1, 2, \dots \quad (3.29)$$

The phase shift is periodically repeated with the object distance. Clarity is only given up to the distance of half the modulation wavelength $\lambda_m/2$. The shorter the wavelength, the smaller the unambiguity range, the more accurate the measurement. For this reason, several modulation wavelengths are used (Fig. 3.63).

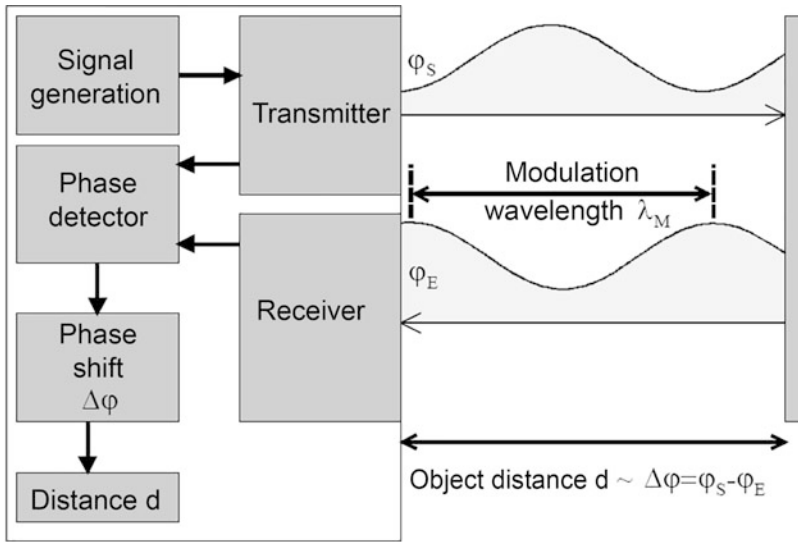
Different measuring principles have been developed for the frequency delay method, which are based on well-known procedures in radar technology.

In the simplest case, the frequency is changed with a constant frequency change $\pm df/dt$ with a frequency deviation H between the minimum and maximum value. The frequency difference $\Delta f = f_S - f_E$ is measured from the current transmit frequency f_S and receive frequency f_E . This results in the object distance d to:

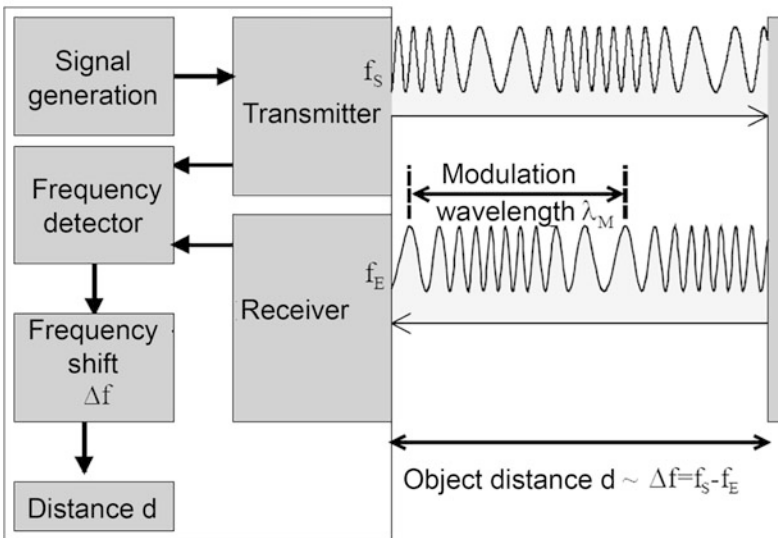
$$d = \frac{\Delta f}{2 \cdot H} \cdot \frac{\lambda_m}{2} + i \cdot \frac{\lambda_m}{2}, i = 0, 1, 2, \dots \quad (3.30)$$

Here too, ambiguity problems arise, which can be resolved by using several frequencies. With the phase or frequency measuring method, ranges of up to 10 m are achieved with diffusely reflecting surfaces.

The triangulation method and the time-of-flight method are predestined for *thickness, height and volume measurement of wood, sheet metal and other materials, contour determination of moving objects, in quality control or for sorting according to different criteria, position control of tools, position detection of parts and tolerance measurement in production*. Typical fields of application are shown in Table 3.11.



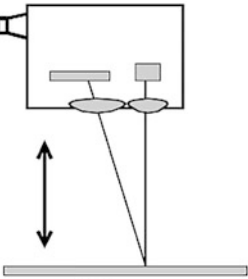
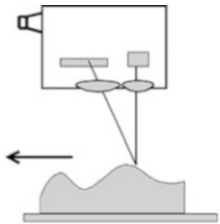
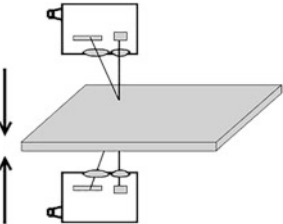
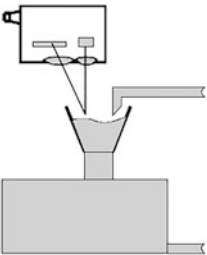
a)



b)

Fig. 3.63 Principle of the (a) phase and (b) frequency delay method

Table 3.11 Application fields of the photoelectric distance sensors (triangulation and transit time)

Application	Field of application/sector	
<i>Distance/position control</i> (distance and height measurement). For example drive-on control on overhead conveyors, web edge control, position control of tools. In particular, small parts detection over long distances. Impact measurement on rotating objects.	Mechanical engineering, storage and materials handling technology, production areas (automotive), plastics processing, iron and steel production, road and aircraft construction, shipbuilding, electrical engineering, paper processing, printing, environmental technology	 <p>Distance measurement</p>
<i>Contour determination</i> when passing objects. For example, in quality control or for sorting according to different criteria.	Quality control, production areas	 <p>Contour determination</p>
<i>Thickness or volume measurement</i> on wood, sheet metal, etc. by difference formation of the measured values of opposite sensors.	Woodworking and wood processing, pulp, paper and cardboard production, building services engineering	 <p>Thickness measurement</p>
<i>Level control and level measurement</i> , especially for controlling the filling process, container handling.	General energy industry, water supply, chemical industry, food and beverage industry, environmental technology, process monitoring	 <p>Filling level control</p>

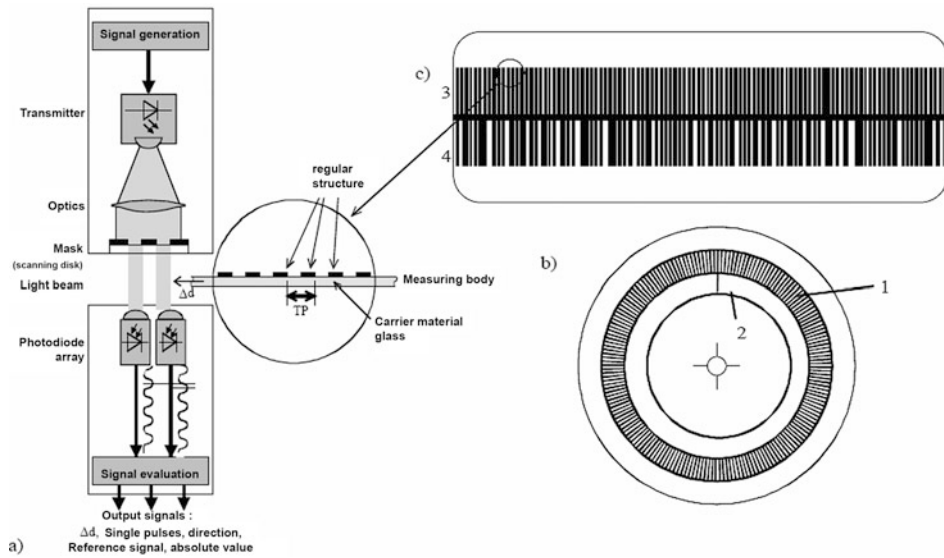


Fig. 3.64 Schematic representation of (a) scanning unit consisting of transmitter, optics, scanning plate and photodiode array, (b) rotary scale with 1: incremental division and 2: reference track and (c) scale with 3: incremental division and 4: absolute coding

3.1.2.7 Measuring Principle: Photoelectric Scanning

This is an *imaging measurement principle*. With the aid of an optical system and a mask (scanning plate), narrowly limited light beams are generated, each of which shines individually on a photodiode (Fig. 3.64a). These light beams are periodically interrupted when a measuring standard, which has a regular structure, a so-called graduation, is inserted. The photodiode current fluctuates according to the displacement Δd . The number N of interruptions gives the distance covered:

$$\Delta d = N \cdot TP, \quad (3.31)$$

where TP is the division period of the structure.

The direction of movement of the measuring standard is determined by a second photodiode, which is offset by $\frac{1}{4}TP$ or which is assigned to a second track whose structure is arranged around $\frac{1}{4}TP$ to the first track. A corresponding logic in the signal evaluation recognizes the direction of movement.

The resolution can be multiplied by using several photodiodes that are offset from the division period.

A special type of design are *rotary measuring standards* (Fig. 3.64b), which are used, for example, to measure angular positions on motors. These measuring instruments, so-called *rotary encoders*, usually do not provide the distance d as an output signal, but

the electronically processed pulse sequences of the photodiodes. These encoders can be equipped with a reference track that is used to approach the angular zero point after the encoder is switched on or to display the zero point during operation.

In addition to these incremental measuring methods, the *absolute measuring method* is also implemented. Here, the structure carries a *serial code* (Fig. 3.64c), which is irradiated with several light beams that are assigned to a corresponding photodiode array. Immediately after switching on an encoder, the current position of the measuring standard can be queried by reading the on/off states of the individual photodiodes.

The *carrier material* for the scale usually consists of *glass substrates* or *metal strips*. The scale is applied by photolithographic processes, either as a thin, non-transparent layer or as an etched, non-reflective structure. In the case of the glass scale, the light beams shine through the glass onto the photodiodes and in the case of the metal bands, they are reflected into the photodiodes. The graduation periods are usually between 4 μm and 40 μm . Accuracies down to the sub- μm range can be achieved by means of the known interpolation methods.

Depending on the application, the length of the scales varies between 50 mm and 1000 mm, in some cases even up to several 10 m.

Linear position encoders are used in *linear drives*, in *machine tools*, in *handling and automation technology* as well as in *measuring and testing equipment*. Rotary encoders are used in electric drives, especially in closed-loop *servo drives* in automation technology, robotics and handling technology as well as in tool and production technology.

3.1.2.8 Measuring Principle: Interferometric Length Measurement

Interferometric length measurement exploits the *wave character* of the laser light. For this purpose, the laser light is split into two partial beams by means of beam splitters (e.g. a partially transparent mirror) (Fig. 3.65). After passing through different optical paths s_1 and s_2 , these are superimposed again and their superimposed light intensity LI is measured with a detector.

Due to the outstanding properties of laser light (coherence), *interference phenomena* occur during superposition. This means that the light intensity periodically fluctuates when one of the two optical paths changes. The optical path of sub-bundle 1 lies within the sensor and is adjusted by a fixed mirror. The optical path of the partial beam 2 leads to the object, for example a mirror or a reflector, and back to the sensor. If the object is moved by the distance Δd , the light intensity LI on the detector changes periodically with the wavelength λ :

$$LI \propto \cos\left(\frac{2\pi}{\lambda} \cdot n \cdot 2\Delta d\right), \quad (3.32)$$

where n is the refractive index of the medium passed through. The number N of maxima passed is the distance travelled:

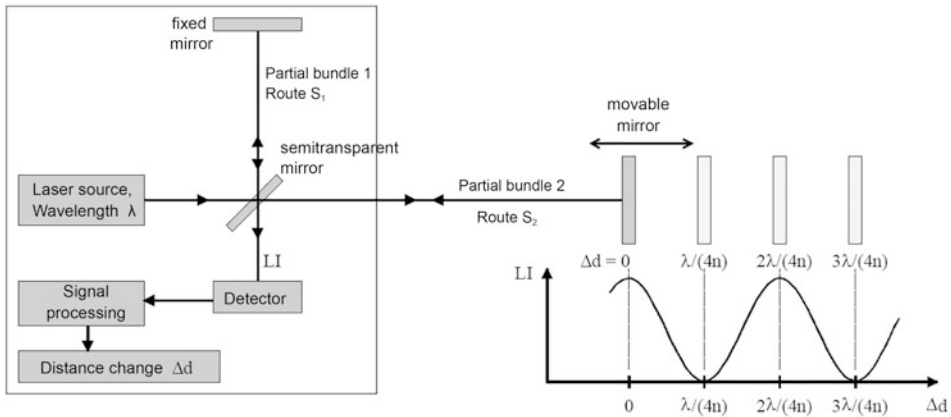


Fig. 3.65 Principle of interferometric length measurement

$$\Delta d = N \cdot \frac{\lambda}{2n}. \quad (3.33)$$

To get from a maximum of light intensity to the adjacent minimum, the mirror must be moved by $\Delta d = \lambda/(4n)$. Using a helium-neon laser with the wavelength $\lambda_{\text{He-Ne}} = 632.8 \text{ nm}$, displacements of $0.158 \text{ }\mu\text{m}$ (air, $n = 1$) can be detected with this method. Higher resolutions can be achieved by using interpolation methods. When measuring in air, the influence of air temperature, humidity and pressure on the refractive index must be taken into account.

Interferometers offer the *highest resolution* and are used wherever machines and systems are adjusted or measured in the μm area. As non-contact form and surface measurement, they are used in quality control.

3.1.3 Ultrasonic sensors for Distance Measurement and Object Detection

Ultrasound lies in the frequency range from 20 kHz to 1 GHz. The basic acoustic parameters have already been dealt with in Sect. 2.18. In the following, the functional principles, design and areas of application of ultrasonic sensors are presented.

3.1.3.1 Operating Principles and Design Touch Operation with Echo Runtime Measurement

Sensors in scanning mode are designed as *single-head systems*. Here, the ultrasonic transducer first works as a *loudspeaker* and emits a wave train that propagates at the speed of sound of the surrounding medium. After sending the ultrasonic sound, the sensor switches the transducer to *microphone mode*. The time required for the transducer to decay determines the *blind zone*, since no echo can be received during the decay process and

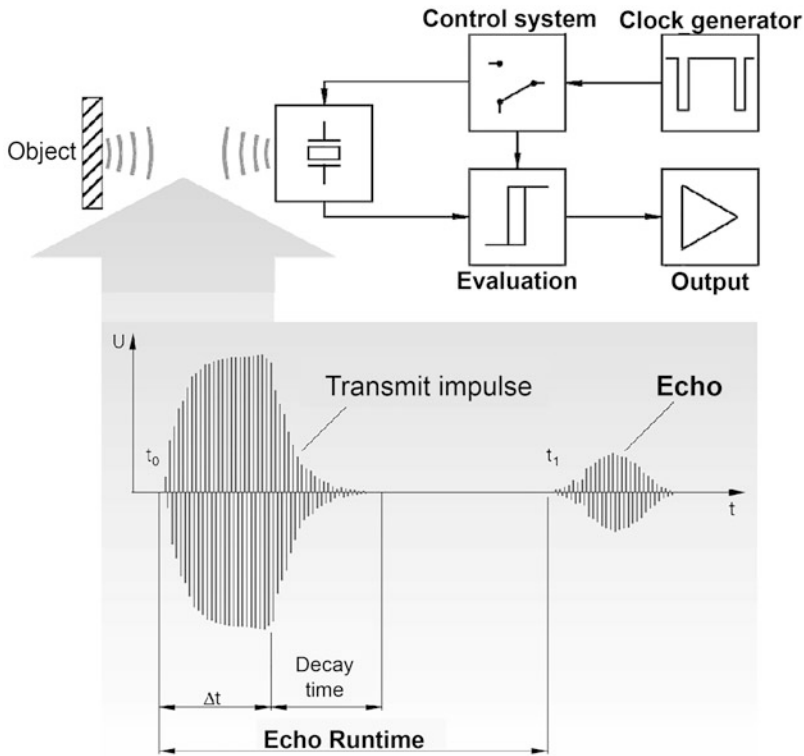


Fig. 3.66 Operating principle of an ultrasonic sensor

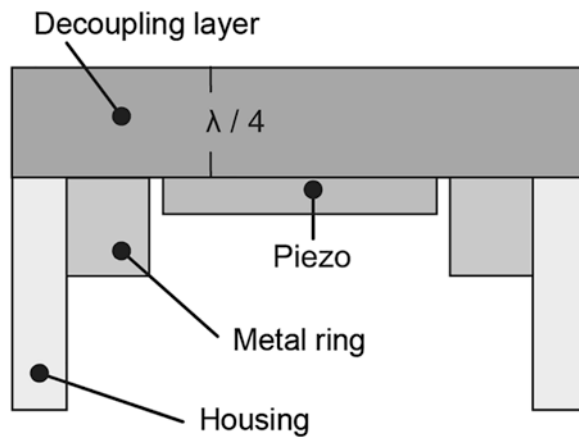
switching to microphone mode. The emitted ultrasonic wave is partially reflected by the target to be measured and returns to the sensor. In the sensor, the returning echo is detected, amplified and evaluated in microphone mode.

The determination of the object distance is thus based on a transit time measurement between the emitted ultrasound and the reception of the echo. The integrated controller calculates the distance from the *echo transit time* and the *speed of sound* (Fig. 3.66).

Barrier Operation

In barrier operation, the sensor works as a *two-head system* with a transmitter and receiver arranged opposite each other. The advantage of this system is that there is *no blind zone* in the two-head system. In contrast, the transmitter and receiver must be *aligned* with each other. However, this system is only suitable for presence monitoring, i.e. distances cannot be measured here.

Fig. 3.67 Structure of the coupling layer



3.1.3.2 Design of the Ultrasonic Transducer

The ultrasonic transducer is the heart of the sensor. It is responsible for coupling the sound in and out. This also results in the sensitivity and range of the sensor. The coupling layer is made of an epoxy resin-glass hollow sphere mixture (microballoons) in order to achieve an optimum adaptation of the sound to the medium air (acts like an electric $\lambda/4$ antenna; Fig. 3.67).

The ultrasonic transducer is periodically controlled by a power amplifier via a piezo-electric crystal. By applying an external voltage from the amplifier to the piezoelement, the latter changes its geometrical dimensions and causes the ultrasonic transducer to oscillate. This effect is reversible. The echo acting on the ultrasonic transducer as an external force causes surface charges on it, which can be measured as voltage. In transmitting mode, electrical energy is converted into mechanical energy and in receiving mode mechanical energy is converted into electrical energy. Air and piezoceramics have very different impedances. As shown in Sect. 2.18, only 0.009% of the sound intensity is radiated. With a suitable coupling layer of an epoxy resin-glass hollow sphere mixture, the impedance is adjusted and thus an optimum adaptation of the sound to the medium air is achieved. The voltage amplitude during the transmission of the sound pulse is about 80 V PP. The length of the sound pulse depends on the length of the transmission pulse and on the *decay time* of the transducer. The voltage amplitude of the received echo is in the μV range.



microsonic

Dipl.-Ing. Harry Pilz
Development

OUR HEART
SOUNDS ULTRA.

For more than 25 years, our engineers have been developing ultrasonic sensors for industrial automation technology. For example, the **nano** - the shortest ultrasonic sensor in a M12 thread still on the market -, our all-in-one **mic+** with digital display or the **esf** for splice and label detection.



microsonic.de

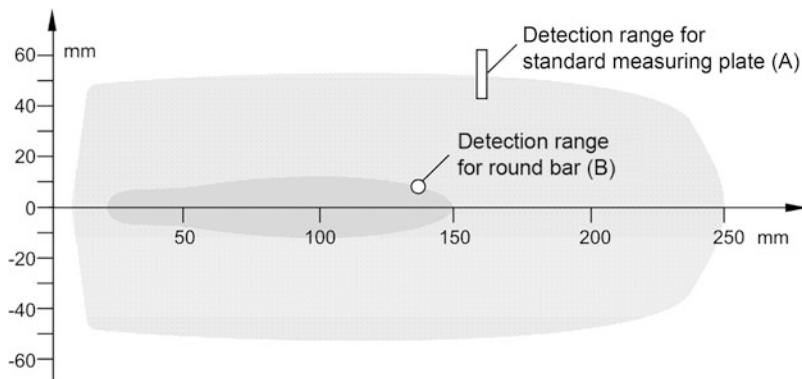


Fig. 3.68 Detection ranges of an ultrasonic sensor (Source: Balluff GmbH)

3.1.3.3 Detection Range of an Ultrasonic Sensor

The *opening angle of the sound lobes* is typically between 5° and 8° . It corresponds approximately to the 3-dB limit, i.e. the sound pressure drops to half of the original value. Even outside this angle, however, objects of corresponding size, shape and surface characteristics can still be detected. Figure 3.68 shows the detection area of a flat standard measuring plate (A) 100×100 mm perpendicular to the direction of propagation of the ultrasound and the detection area of a round rod (B) with a diameter of 25 mm. Within these ranges, the detection of the specified objects is guaranteed.

The range in which the sensor can detect objects is limited by the smallest and largest range. This is determined by the frequency used, on which the size of the blind zone also depends (Fig. 3.69). The ultrasonic sensor cannot detect an object in the blind zone. It is caused by the transmission pulse duration and decay time of the transducer, since no echo can be received while the ultrasonic pulse is being transmitted.

By using suitable focusing attachments, the ultrasound can be bundled and thus even very small objects in the sensor's close range can be detected. Figure 3.70 shows a typical design of an ultrasonic sensor with focusing attachment.

In order to obtain accurate results when measuring distances with ultrasound, the *speed of sound of the air* must be determined. For industrial sensors, it is sufficient to adjust the speed of sound to the current temperature. The dependence on the air humidity is negligible. Ultrasonic sensors work with a fixed ultrasonic frequency. This determines the maximum measuring distance. The higher the ultrasonic frequency, the greater the absorption and the smaller the maximum range. The next ultrasonic pulse can be emitted earlier, the cycle time—the time interval between the individual measurements—can be reduced as the frequency increases. Sensors with short ranges can therefore have short response times. Typical ranges for some ultrasonic frequencies are given in Table 3.12:

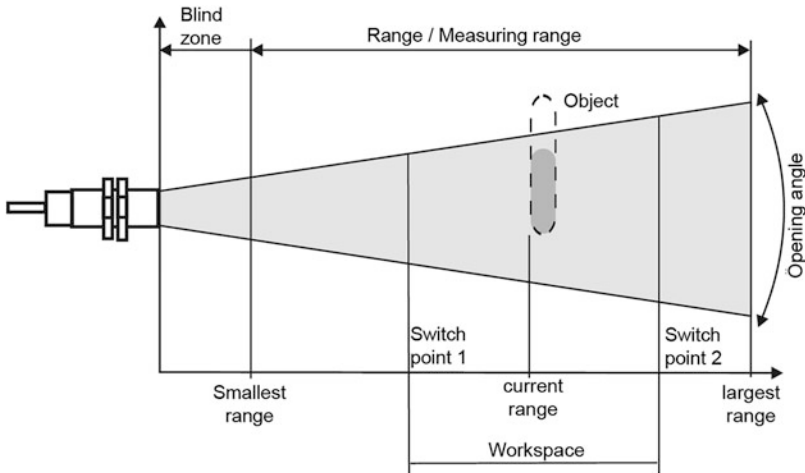


Fig. 3.69 Classification of the sound beam



Fig. 3.70 Design of an ultrasonic sensor with focusing attachment (Factory photo: Balluff GmbH)

Table 3.12 Ranges depending on the ultrasonic frequency

Ultrasonic frequency [kHz]	Maximum measurement distance [mm]	Wavelength [mm]	Cycle time [ms]
80	6000	4.1	64
130	3500	2.5	20
180	2000	1.8	16
220	1500	1.5	12
300	600	1.1	8
360	300	0.9	4

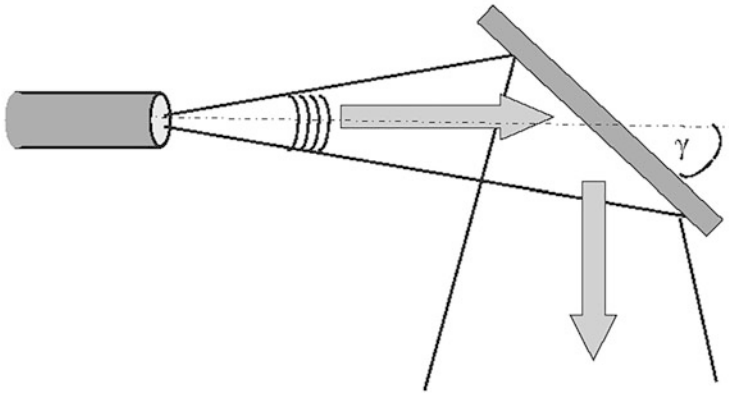


Fig. 3.71 Deflection of the ultrasound

3.1.3.4 Deflection of the Ultrasound

A flat object tilted by $\gamma = 45^\circ$ relative to the transducer axis deflects the sound beam by 90° (Fig. 3.71).

With such a sound deflection angle, the transducer can be protected and the blind zone can be placed between the transducer and the reflector.

3.1.3.5 Object and Environmental Influences

Object Influences

Almost all objects (solids, liquids, bulk materials) reflect sound and can therefore be detected. Sound-absorbing materials, such as foams, can also be detected within a limited range. In general, solid, liquid or powdery media or objects can be detected regardless of their colour, transparency or surface properties. Some examples are given below (Fig. 3.72).

At *convex (cylindrical and spherical) surfaces*, each surface element has a different angle to the axis of the club. The reflected lobe diverges as a result and the proportion of sound energy reflected to the receiver is reduced accordingly. The maximum range decreases as the cylinder (sphere) becomes smaller.

Roughness and surface structures of the object to be detected additionally determine the scanning properties of ultrasonic sensors. Surface structures that are larger than the ultrasonic wavelength, as well as coarse-grained bulk materials reflect ultrasound diffusely and may not be detected optimally by ultrasonic sensors.

In ultrasonic applications, *hard material* reflects almost all of the pulsed energy, making it very easy to detect with ultrasound.

Fig. 3.72 Deflection of the ultrasound (Factory photo: Balluff GmbH)



Soft material, on the other hand, absorbs almost all the pulsed energy. It is therefore less easily detected with ultrasound, which means a reduction in range. These materials include felt, cotton wool and foam.

Thin-walled films behave like soft materials. To be able to use ultrasound, the film thickness should therefore be at least 0.01 mm.

Liquids behave like hard material. Foam on the surface can have a negative effect on detection.

Hot tactile objects cause heat convection in the surrounding air. This can cause the sound beam to be deflected perpendicular to its axis to such an extent that the echo is weakened or cannot be received at all. Some of the sound is also reflected at the interfaces between air masses of different temperatures.

Environmental Influences

Ultrasonic sensors are designed for use in atmospheric air. *Environmental influences* such as humidity, dust and smoke *do not* affect their *measuring accuracy*. Operation in other gases, such as CO₂, is not possible because the sound is very strongly absorbed. Liquids that evaporate solvents can also impair the sensor function.

Strong air movements and turbulence lead to instability in the measurement, but can be disregarded under normal conditions in production facilities. Flow speeds of up to several m/s can be handled without any problems, so that even outdoor applications are not affected.

Precipitation such as rain or snow in normal density does not impair the function of the ultrasonic sensor and its output signal. However, the transducer surface should not be wetted.

3.1.3.6 Applications

Ultrasonic sensors can be used in almost every industry. Important and frequent fields of application are:

- Winding and unwinding control/diameter detection,
- sag control,

Fig. 3.73 Winding and unwinding control (Factory photo: Balluff GmbH)

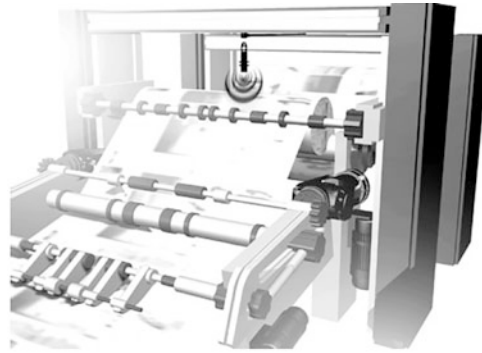
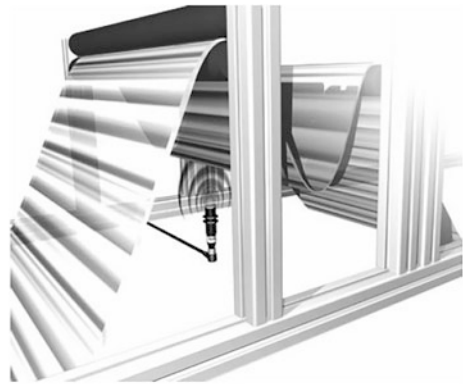


Fig. 3.74 Sag control with an analog ultrasonic sensor (Factory photo: Balluff GmbH)



- altimeter,
- position control,
- collision control,
- filling level detection and
- object recognition/count objects.

The following pictures show some possible applications for ultrasonic sensors.

Figure 3.73 shows a winding and unwinding control with an analog ultrasonic sensor, independent of color, reflections and surface condition. Ultrasonic sensors can also easily compensate for light soiling, such as paper dust.

Figure 3.74 shows a sag control with an analog ultrasonic sensor. Some analog ultrasonic sensors also have one or more switching points. Thus, switching points for min. and max. Sag or film tear monitoring can be parameterized.

Figure 3.75 shows a height measurement with an analog ultrasonic sensor independent of the color and surface condition of the object. It can also be used under harsh environmental conditions (e.g. dust, sawdust or moisture).

Fig. 3.75 Height measurement with an analog ultrasonic sensor (Factory photo: Balluff GmbH)



The filling level can also be measured with an ultrasonic sensor (Fig. 3.76). Almost all liquid, powdery and pasty media can be detected.

Figure 3.77 shows how objects can be detected and counted with ultrasonic sensors. This is done with several ultrasonic sensors arranged side by side. To prevent the sensors from influencing each other, there is the possibility of synchronisation.

3.1.4 Potentiometric Displacement and Angle Sensors

3.1.4.1 Introduction

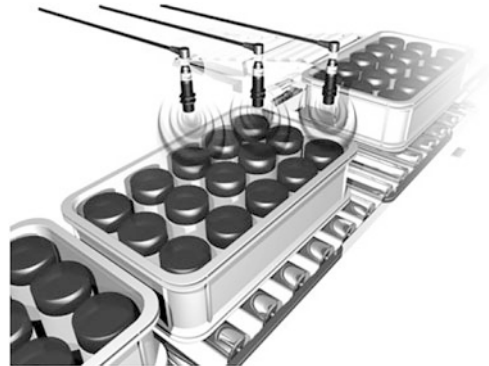
Contactless methods are becoming increasingly popular in displacement and angle measurement technology. However, potentiometer-based sensors are still without serious competition in many automotive, mobile and industrial applications due to their positive characteristics and favourable price–performance ratio. This is unlikely to change in the foreseeable future. After all, comparable measuring speeds, linearity values, resolutions, hysteresis values and temperature ranges can otherwise only be achieved with considerably greater effort.

Conductive plastic potentiometers have a firm place in numerous fields of application, both in industry and mobile machinery or in automotive engineering. Potentiometric distance and angle meters probably have a larger market share (about more than 20 million units/year) than all alternative contactless technologies combined. The argument of reliability is often used to give priority to contactless sensor technology over potentiometric solutions. It has been proven that a mature technology can achieve field data that have

Fig. 3.76 Measuring the filling level with an analog ultrasonic sensor (Factory photo: Balluff GmbH)



Fig. 3.77 Detecting and counting objects with an ultrasonic sensor (Factory photo: Balluff GmbH)



never been achieved with contactless solutions and will not be achieved in the future. In a Bosch throttle valve application, for example, there are 50 million redundant sensors in the field, operating in a temperature range from $-40\text{ }^{\circ}\text{C}$ to $+150\text{ }^{\circ}\text{C}$ and having to complete over ten million switching cycles, and this with a field return rate of 0 ppm.



Precision for your design

Precision Potentiometer

Conductive plastic, wire and hybrid potentiometer

Wide range of angle and displacement sensors
Numerous variants and options
Highest quality and durability



MEGATRON Electronics GmbH & Co. KG • Hermann-Oberth-Strasse 7 • 85640 Putzbrunn / Munich
Tel.: +49 89 46094-0 • Fax: +49 89 46094-201 • info@megatron.de • www.megatron.de

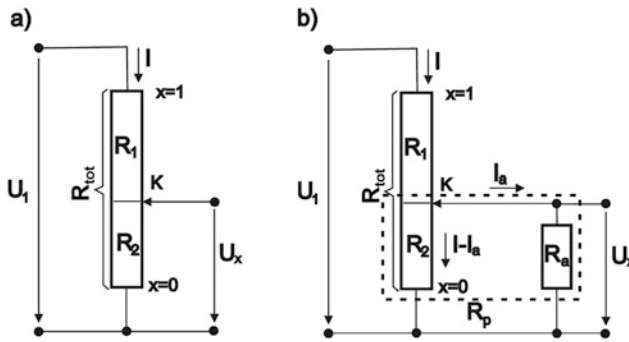


Fig. 3.78 Potentiometer circuit

3.1.4.2 Operating Principle and Characteristics of Potentiometric Sensors

Linearity Error

Figure 3.78a) shows an *unloaded voltage divider*. The output voltage is directly proportional to the displacement or angular position x . With a linear resistance function, there is a *linear relationship* between the applied voltage U_1 and the output voltage U_x . It applies:

$$U_x = U_1 \frac{R_2}{R_1 + R_2} = U_1 \frac{R_2}{R_{\text{tot}}} = U_1 \cdot x. \quad (3.34)$$

In practice, the potentiometer is loaded, i.e. a wiper current I_a flows, which influences the output signal. So that is:

$$U_x = U_1 \frac{R_2 \cdot R_a}{R_1 R_2 + R_a (R_1 + R_2)}. \quad (3.35)$$

Due to the wiper load, the characteristic curve “hangs” (Fig. 3.79), and at about 2/3 of the wiper position (grey area), there is a maximum error of:

$$F_{\text{max}} \approx 0,15 \frac{R_{\text{tot}}}{R_a} \quad (3.36)$$

For a precision potentiometer with a typical linearity of 0.03%, the load resistance R_a should be between 1000 and 5000 times greater than the terminal resistance R_{tot} in order to avoid a significant influence on the linearity. This results in load resistances of 1 M Ω to 5 M Ω for typical terminal resistances (R_{tot}) between 1 k Ω and 5 k Ω . The wiper current then remains at 1 μ A. In automotive applications, linearity values in the percentage range are

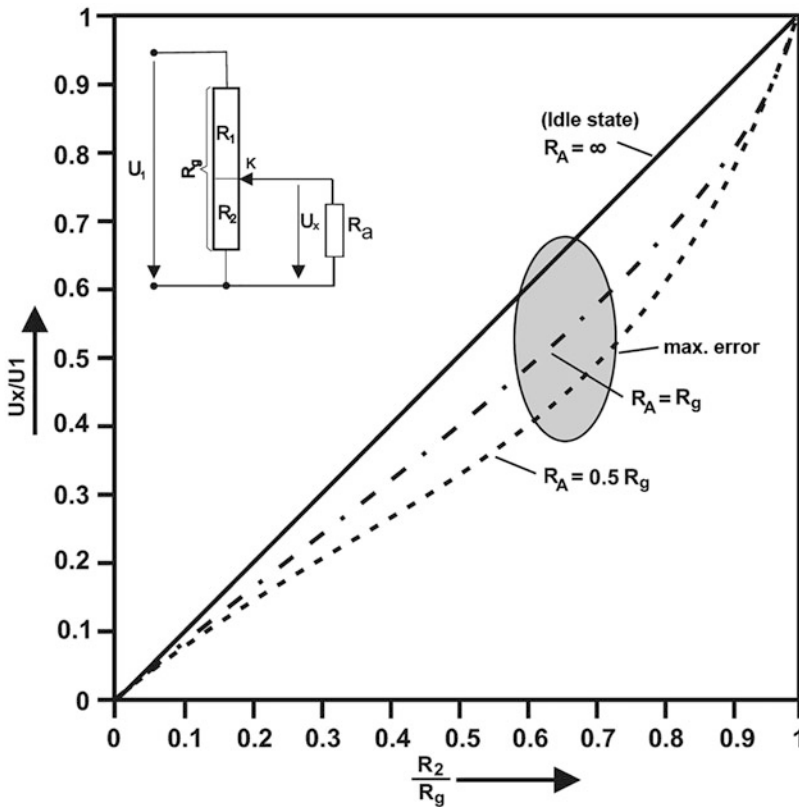


Fig. 3.79 Loaded voltage divider

usually required, so load resistors from 100 k Ω to 470 k Ω can also be used. In mass applications, it is also possible to provide this “linearity sag” for known loads.

Resolution and Hysteresis

In data sheets of many potentiometer manufacturers, you will find the statement that the resolution of a potentiometer is infinite. This statement is not correct and was derived from the wire potentiometer technology (Sect. 3.1.4.3), where the resolution is determined by the wire thickness d used (Fig. 3.80).

It cannot be assumed that this limit will not apply to conductive plastic potentiometers, since here too different parameters set limits to the resolution. In addition, in an analog system, the system noise also sets limits. A distance measuring system with a length of 500 mm, for example used on plastic injection machines, is read out today with 16 bits, which corresponds to a resolution of 0.15 mV or 8 μm at 10 V supply voltage.

The hysteresis is mainly determined by the mechanical variables such as bearing, stiffness of the slider system and the friction values between the layer and slider.

Fig. 3.80 Resolution of the wire potentiometer

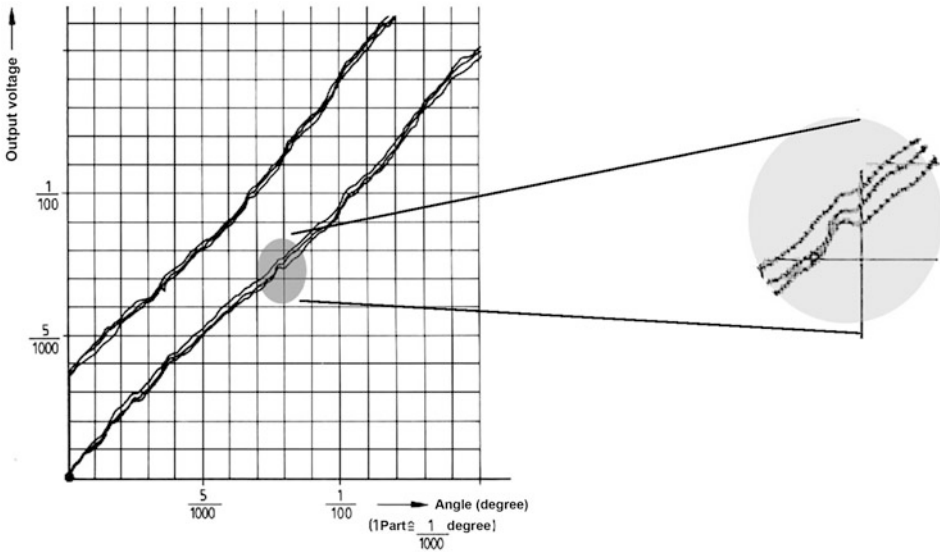
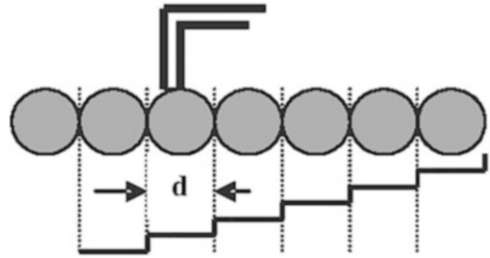


Fig. 3.81 Resolution and hysteresis of conductive plastic potentiometers

Figure 3.81 shows a section of the output characteristic curve of a rotary conductive plastic potentiometer. The measurement was repeated three times clockwise and anticlockwise. While the measurement curves almost coincide in one direction (unidirectional), they show a hysteresis of about $4/1000^\circ$ when the direction is reversed (bidirectional). This corresponds to a path of about $1 \mu\text{m}$.

In the area marked in grey, this high-resolution characteristic curve shows that the output voltage does not necessarily increase monotonically in the micro range. The gradient becomes negative in this range, i.e. the resolution limit is reached here.

Temperature Coefficient

Assuming a linear temperature coefficient, the total resistance R_{tot} will behave as follows

$$R_{\text{tot}}(T) = R_{\text{tot}}(T_0)[1 + \alpha(T - T_0)]. \quad (3.37)$$

with $\alpha < 300$ ppm/K.

The temperature response depends very much on the design of the sensor system, so that not only a linear temperature coefficient α will necessarily exist, but also higher order coefficients (β, γ) may play a role. The linear case, however, describes the behavior in a first approximation quite accurately.

If one considers the change in the dividing voltage $\Delta U_x/\Delta T$, only the difference in the coefficients is now taken into account. In first approximation applies:

$$\frac{\Delta\left(\frac{U_x}{U_1}\right)}{\Delta T} = \alpha_{\text{tot}} - \alpha_2, \quad (3.38)$$

where α_{tot} describes the temperature coefficient of the total resistance R_{tot} and α_2 the temperature coefficient of the partial resistance R_2 .

Since it can be assumed that there is no significant temperature difference between the divider resistor R_2 and R_{tot} , and that the coefficients will also be identical, since generated with the same process parameters, this difference is ideally zero. In practice, values in the order of <5 ppm/ $^{\circ}\text{K}$ can be achieved. This also applies analogously to the loaded voltage divider, since, if correctly connected, the load resistance R_a exceeds the total resistance R_{tot} by three orders of magnitude and the influence can thus be neglected.

Table 3.13 lists important technical data of potentiometric sensors.

3.1.4.3 Technology and Construction Techniques

Wire Potentiometer

Wire potentiometers are mainly used in power electronics. *So-called spiral potentiometers* are used in large quantities as sensor or adjustment potentiometers. The advantage of this technology is that up to $n = 30$ rotations can be realized with very little effort, i.e. an angular range of $n \cdot 360^{\circ}$ can be measured.

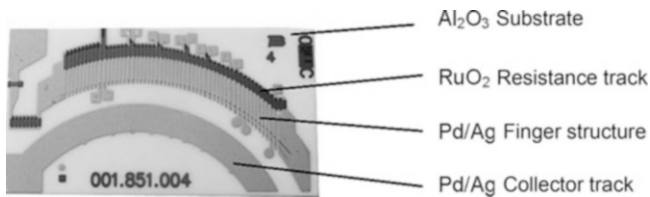
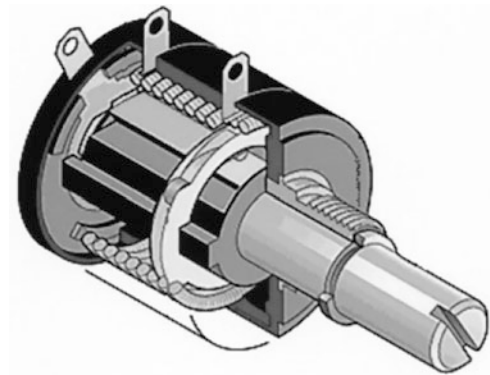
The resistance wire is wound on a linear plastic carrier or on an insulated copper wire, which is then installed in the housing as a helix (Fig. 3.82). By using a copper wire as carrier, the temperature behavior can be significantly improved as the partial resistors are subjected to the same temperature.

Cermet Layer System

Today's sensors are more and more constructed as planar systems based on thick-film cermet or polymers. The planar cermet technology is still mainly used for tank level sensors (Fuel Card; Fig. 3.83). The sensors are directly exposed to the fuel such as diesel, gasoline or ethanol. Polymer layer systems cannot be used in such an application. These sensors are usually operated in *rheostat mode* (2-wire).

Table 3.13 Technical data of potentiometric sensors

Key data	Values
Measuring range	Up to 4000 mm, or 345°
Linearity error (independent)	$\pm 0.03\%$, $\pm 1\%$ (automotive)
Resolution	$< \pm 1 \mu\text{m}$
Hysteresis	$< \pm 5 \mu\text{m}$
Temperature coefficient	Typical $\pm 5 \text{ ppm/K}$
Maximum adjustment speed	10 m/s
Lifetime	Up to 100 million cycles
Functional safety	ISO 26262 ASIL D or IEC 61508 SIL 3 possible
EMC	No measures necessary, since passive
Supply voltage	Typ. 5 V to 24 V
Power consumption	Typ. 5 mA
Output signal	Analog, voltage, ratiometric
Protection class	IP65, special housing IP 69 K
Operating temperature range	$-40 \text{ }^\circ\text{C}$ to $+150 \text{ }^\circ\text{C}$

Fig. 3.82 Spiral potentiometer
(Factory photo: Bourns Inc.)**Fig. 3.83** Fuel Card (Factory photo: Bourns Inc.)

The carrier material is a pore-free aluminium oxide ceramic (Al_2O_3 , 96%), to which the potentiometer structure is applied by screen printing. The thick-film conductor paths are realized with a palladium-silver (Pd/Ag) mixture. The mixture still contains glass

components, which ensure good adhesion to the ceramic. With the silver, a low line resistance can be achieved. The palladium increases the wear resistance of the silver and protects it against chemical and electrolytic corrosion. For higher demands, especially for so-called “*flex-fuel*” vehicles that run with a high ethanol content (85%), a tertiary mixture system (Pd/Au/Ag) is used. However, this system is currently considerably more expensive.

The resistor track is printed on a ruthenium oxide base. The resistance can be adjusted by adding Pd/Ag flakes. The resistor and conductor tracks are sintered at 850°.

Due to the rheostat mode, it is important to realize a very low contact resistance with the resistor and collector track. Therefore, the resistor track is not contacted directly, but via a finger structure (30–100 fingers) made of Pd/Ag. This allows contact resistance values of 100 mΩ to be achieved. The resolution of such a sensor is directly related to the number of fingers. The output characteristic curve is usually non-linear; rather, it is adapted to the tank geometry and can be determined via the *finger layout* and *laser trimming*.

Polymer Coating Systems

With modern epoxy and polyester resins in combination with special carbon blacks up to *carbon nanotubes* (CNT), very powerful and cost-effective sensors can be constructed today. Using the screen printing process, with working areas of 300 mm × 1200 mm, up to 200 sensors can be produced per printing process in so-called multiple-up production (Fig. 3.84).

In the meantime, only high-quality FR4 material (FR = flame-retardant) is used as substrate material. Quality 4 describes the components epoxy resin + glass fibre fabric.

FR1–3 qualities are partly still used for low-cost adjustment potentiometers. The necessary conductors can either be made as standard copper lamination or by screen printing with conductive silver. The baking temperatures for polymer layer systems are usually 220 °C.

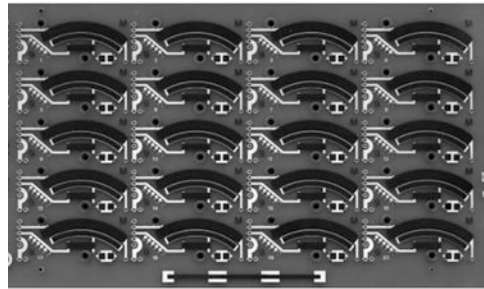
Layer Structure and Linearization

As Table 3.13 shows, linearity values of <0.05% can be achieved with potentiometers. With modern screen printing machines, one obtains layer thickness accuracy of about 2 μm. With a total layer thickness of about 20 μm, this corresponds to 10%. As can be shown theoretically, there is a linear relationship between the percentage *coating thickness variation* $\Delta d/d$ and the resulting *linearity error* F . It applies

$$F = \frac{\Delta d/d}{10}. \quad (3.39)$$

This means: In order to achieve the abovementioned linearity values, the layer thickness tolerance at 20-μm layer thickness must not be greater than 0.1 μm. Since this is not technologically feasible, the resistor path must be linearized. It applies:

Fig. 3.84 Multiple use with
 4×5 angle potentiometers
 (Factory photo: Novotechnik)



$$\frac{\Delta U_x}{\Delta U_1} \alpha \frac{\Delta R}{\Delta x} = \frac{\kappa}{b(x)d(x)}. \quad (3.40)$$

As long as the change in resistance ΔR remains constant per path increment Δx , the characteristic output curve is strictly linear and therefore the linearity error F is zero. With the condition

$$b(x)d(x) = \text{const.} \quad (3.41)$$

this can be achieved. Here the advantage of the planar structure comes into play. With the aid of a laser or a mechanical milling cutter, the condition according to Eq. (3.41) can be easily fulfilled by correcting the width $b(x)$ over the path or angle (Fig. 3.85) and thus compensating for layer thickness variations $d(x)$. If this linearization process is only carried out for an angle sensor when the resistance element is already installed in the housing, then a possible eccentric error can also be corrected at the same time.

Grinder Systems

There are basically two types of grinder geometry: the so-called *scratch grinder* (Fig. 3.85a) and the *skid grinder* (Fig. 3.85b) with a calotte or V-shaped geometry (Fig. 3.85).

For precision potentiometers, the scraper grinder has become established, which can be a punch grinder (4 fingers) or a wire grinder (4×3 wires). The advantages over a skid sander are:

- constant contact surface over the service life,
- no floating on abrasive material and
- identical contact position of all fingers over the lifetime.

In the past, simple alloys such as nickel silver (Cu,Ni,Zn) or copper beryllium (Cu,Be) were used as grinding materials, whereas today almost exclusively precious metal alloys (Au, Pt,Pd,Ag) such as Paliney 6 or Paliney 7 are used. These alloys show very good contact behavior even under difficult environmental conditions and guarantee a long service life (Table 3.13).

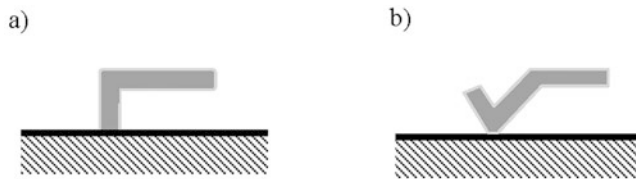
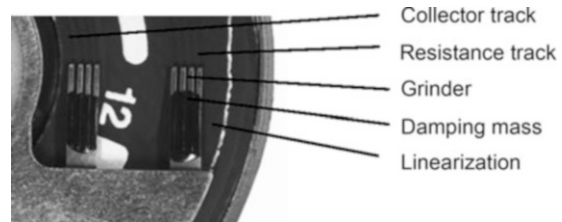


Fig. 3.85 Grinder geometries

Fig. 3.86 Damped slider system (Factory photo: Novotechnik)



For high vibration loads, the system can be damped by applying a rubber compound (Fig. 3.86), which significantly improves the contact behavior.

Foil Potentiometer

Another very interesting variant of the potentiometer is the foil potentiometer. In place of the slider, a collector foil is used, which is pressed against the resistor track by means of either a mechanical pin or a magnet and makes mechanical contact with it (Fig. 3.87).

The sensors for displacement and angle measurement consist of different foils, which are electrically insulated from each other by a *spacer*. The resistor track is applied to one of the foils or to an FR4 substrate using screen printing. On the opposite foil, the collector foil, a low-resistance collector track is printed. A mechanical pressure, usually applied by a simple pressure piece, brings the collector foil into contact with the resistance foil.

The starting point for a *magnetically* actuated foil potentiometer is a conventional foil potentiometer (Fig. 3.88). A stack consisting of several thin, highly elastic metal foils is built on the collector foil. The thickness of the individual foils is typically less than 50 μm . If a permanent magnet is guided below the foil potentiometer, the elastic *ferromagnetic foil stack* is attracted by the magnetic field and brings the flexurally elastic collector foil of the foil potentiometer into contact with the resistance path. Depending on the number of metal foils and the field strength of the actuating magnet, working distances of up to 10 mm can be achieved.

To protect the foil package against mechanical damage from the outside, these cover foils are integrated into the foil potentiometer using proven adhesive technology. Foil potentiometers can be glued to flat surfaces in the required form. In addition to linear designs, rotary systems can also be manufactured. The operating temperature range is determined by the foil material used. Standard versions made of high-quality polyester foil



Fig. 3.87 Foil potentiometer, mechanically operated (Factory photo: Metallux)

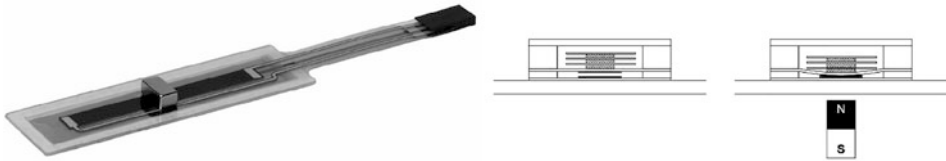


Fig. 3.88 Foil potentiometer, magnetically actuated (Factory photo: Metallux)

allow operation in a temperature range up to $+85\text{ }^{\circ}\text{C}$. If an FR4 prepreg is used as collector support, operating temperatures of $+125\text{ }^{\circ}\text{C}$ can also be achieved.

Foil potentiometers are very flat, between 0.5-mm and 2 mm thick, depending on the version. A further advantage is that the individual foils of the foil potentiometer are *hermetically sealed* together. Dirt, dust or moisture cannot penetrate and thus enable use in difficult environments. With the magnetically actuated potentiometer, so-called *transmissive operation* is also possible, i.e. there may be a non-magnetic material between the magnet and the sensor (e.g. aluminium housing wall of a cylinder).

3.1.4.4 Products and Applications

Applications in Automotive Engineering

Today's passenger cars drive with an e-gas system consisting of an accelerator pedal and an electrically driven throttle valve, both equipped with a redundant potentiometric sensor system.

The Euro 6 requirements can only be achieved with electronically controlled systems in which the driver has no direct access to relevant parameters. The driver can express a wish, for example to accelerate, but the optimum opening angle of the throttle valve is determined by the engine management electronics. As can be seen in Fig. 3.89, the potentiometer housing is part of the complete component. This is referred to as an *open sensor solution*.

In the Asian markets, two-wheeled vehicles (mopeds, motorcycles, scooters) continue to play an important role as means of transport. In India, about eight million vehicles were built in 2010 and this number will double by 2015.

Automotive applications



Fig. 3.89 Throttle valve and accelerator pedal module (Factory photo:Novotechnik)

Environmental protection is also becoming more and more important in these countries, i.e. exhaust gas and consumption values must be reduced. Virtually, no e-gas systems are yet in use. The mechanical Bowden cable connection to the throttle valve is still in place, but with the information throttle valve angle, temperature and pressure in the intake manifold an optimal ignition timing can be calculated. In these markets, only complete sensors (bolt on) are still used today (Fig. 3.90).

Typical applications for foil potentiometers in automotive engineering are adjusting systems for car and truck seats, window regulators, convertible tops and mirror systems.

Industrial Applications

The possible applications for potentiometric sensors in industrial environments are so diverse that only exemplary areas of focus are mentioned here (Table 3.14 and Fig. 3.91). It can be assumed that all industries use potentiometric sensors.

3.1.5 Magnetostrictive Displacement Sensors

Non-contact, precise and absolute measurement are the decisive advantages for the broad industrial use of *linear, magnetostrictive displacement sensors* (Table 3.15). Due to the contactless and thus wear-free mode of operation, expensive service calls and downtimes are avoided. The operating principle of these displacement sensors makes it possible to install them in hermetically sealed housings; because the current position information is transmitted via magnetic fields without contact through the housing wall to the internal sensor element. In principle, the simultaneous measurement of several positions with one measuring system is possible.

Magnetostrictive displacement sensors achieve protection classes IP67 to IP69K without complicated, costly and error-prone sealing concepts. The high insensitivity to shock and vibration extends the industrial application areas far into heavy machinery and plant construction. Measurement and position values that are immediately available as absolute values after the system is switched on are mandatory in many applications. They also



Fig. 3.90 Throttle valve sensors (TPS) for two-wheel applications (Factory photo: Novotechnik)

Table 3.14 Application areas of potentiometric sensors

Area	Applications
Mechanical and automation engineering	Actuators, level measurement, robot controls, valve positioning, use in pneumatic and hydraulic cylinders, door and gate drives, machine tools, plastic injection moulding machines, conveying and lifting vehicles, packaging machines, hydraulic presses, electro-hydraulic lifting and conveying technology, printing machines, stone moulding machines and ultrasonic welding machines
Medical technology	Operating tables, syringe pumps, respiratory equipment, mammography equipment, dialysis systems, magnetic resonance tomographs
Shipbuilding	Control of marine propulsion systems, use in joystick
Aerospace	Seating position in airplanes, landing gear positioning, handling robots on the ISS, positioning of solar panels in satellites
Mobile working machines	Forklift trucks, agricultural machinery, beach cleaning vehicles, track-laying machines



Fig. 3.91 Sensors for industrial applications (Factory photo: Novotechnik)

increase machine availability by eliminating reference runs. Table 3.15 shows the advantages.

Table 3.15 Advantages of magnetostrictive displacement sensors

Absolutely measuring	No reference run, signal immediately available.
Contactless	Wear-free.
Robust/reliable	High protection class (IP67/69 k), pressure resistant, insensitive to vibration and shock.
Multi-position capable	Up to 30 positions can be recorded with one system.
Flexible use due to many approvals	Potentially explosive areas, hygiene, automotive engineering, marine.
Capable of measuring through the “wall	Position sensor position can be detected through non-magnetic materials.
Easy to install	Passive position sensor, no cabling necessary; mechanical adaptation is roughly tolerated.

3.1.5.1 Operating Principle and Design of Magnetostrictive Displacement Sensors

Magnetostriction can be used to excite *torsional structure-borne sound waves* on tubular or wire-shaped bodies. This is a basic principle for the design of magnetostrictive displacement sensors. Figure 3.92 shows the basic design for the generation of torsional structure-borne sound waves: A copper conductor (stranded copper wire) is threaded into a tube of magnetostrictive material. The imaginary line texture in the graphic representation of the tube is intended to make it easier to describe how it works. One or more permanent magnets are placed around the circumference of the pipe. Only the field components H_p of these permanent magnets, which are bundled in the wall of the pipe and run in axial direction, are of interest for the mode of operation. If a current I is sent through the copper conductor, it surrounds itself with a circular magnetic field, as shown in Fig. 3.93. This magnetic field is also concentrated to a large extent in the wall of the tube and is represented by the arrow H_t . By superimposing the two fields, the resulting magnetic field H_{res} is obtained. Ideally, the field strengths are chosen so that H_{res} is deflected by about 45° . Since the tube is a positive magnetostrictive material, the material will stretch in the direction of the resulting magnetic field. A ring-shaped area is formed, which is slightly twisted in relation to the rest of the tube, made visible by the deformation of the imaginary line texture.

In order to trigger the desired structure-borne sound wave, the torsion is built up by a short current impulse of a few microseconds duration and immediately reduced again. This process is comparable to the triggering of a transverse wave in a tensioned rope, generated by a rapid up and down movement. As we know from experience, the time sequence of the up and down movement determines the achievable amplitude of the rope wave. It is no different with the magnetostrictively triggered wave on the tube. Here too, an exact sequence of the positive and negative edges of the triggering current pulse is decisive for the amplitude of the wave.

Figure 3.94 shows in an image sequence the build-up and breakdown of the torsion and the subsequent propagation of the wave.

Fig. 3.92 Magnetostrictive tube with internal copper conductor and permanent magnetic position sensor

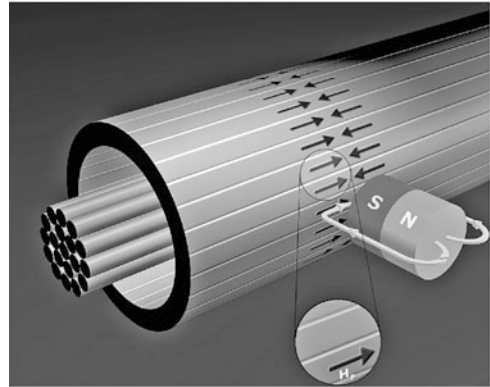
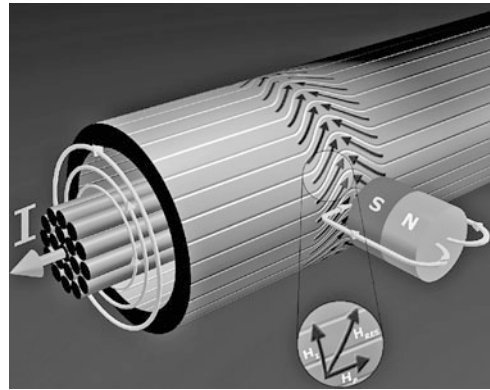


Fig. 3.93 The applied current pulse causes an elastic torsional deformation of the waveguide tube



The torsional structure-borne sound wave travels from the point of release in both directions to the pipe ends. Its propagation velocity v is calculated from the shear modulus G and the density ρ of the magnetostrictive material:

$$v = \sqrt{\frac{G}{\rho}}. \quad (3.42)$$

The propagation speed of commercially available displacement sensors is 2850 m/s. It is largely *independent* of *temperature* due to the alloy used. Torsional waves, unlike transverse waves, cannot be excited externally by shock or vibration. This has a positive effect on the *interference resistance* of magnetostrictive displacement sensors.

The basic idea for the design of a linear, non-contact magnetostrictive displacement sensor is simple. The torsional wave generated at the location of the magnet runs along the tube in both directions up to the ends. At one end of the tube, the wave is undesirable and is therefore damped away by friction. At the other end, there is a *detector* in the form of an *induction coil*, which indicates the arrival of the torsional wave (*reverse magnetostrictive*

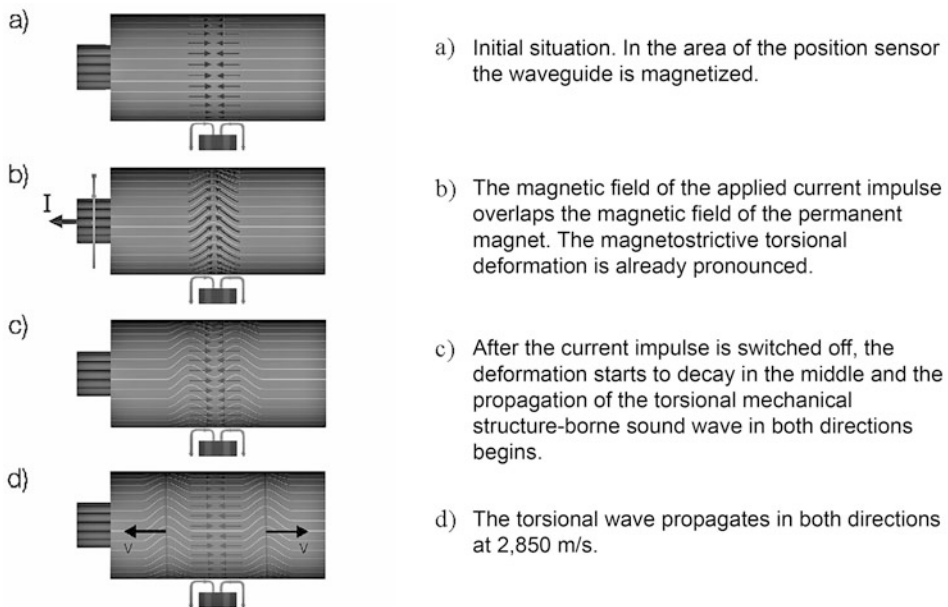


Fig. 3.94 Torsional wave propagation

effect, Sect. 2.4). The time between the triggering of the wave and its arrival at the detector coil is recorded by a high-resolution time-of-flight measurement. From the *measured time*, the distance travelled and thus the sought-after position of the magnet can be calculated directly. The time signal available in digital form is converted into the desired analog or digital output signal form by the integrated interface electronics. Figure 3.95 shows the block diagram of a magnetostrictive system with analog output.

Figure 2.14 shows the basic design of magnetostrictive displacement sensors. Central elements are the tubular or wire-shaped waveguide and one or more cylindrical permanent magnets on the circumference of the waveguide tube. These are fixed in a carrier (not shown), the so-called *position sensor*, which can be freely moved axially along the waveguide. The tubular waveguide has an outer diameter of 0.7 mm with an inner diameter of 0.5 mm. The length corresponds to the desired nominal measuring length, plus a few millimeters for the attenuation distance and the detector. A copper conductor is threaded into the waveguide tube, through which the triggering *current pulse* is sent. The copper conductor is returned outside the waveguide. If the waveguide is not a tube but a solid wire, the current is sent directly through the wire.

A measurement is triggered by a start impulse at the time t_{Start} . Depending on the electrical interface of the position sensor, the start pulse can be triggered externally (*synchronous triggering* of the measurement) or internally (*free-running cyclical generation of measured values*). In either case, the start pulse triggers a current pulse in the copper strand of several microseconds duration. The current pulse generates a torsional structure-

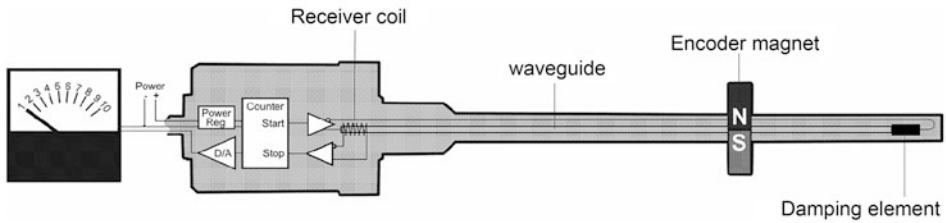


Fig. 3.95 Block diagram of a magnetostrictive displacement sensor (Factory photo: Balluff GmbH)

borne sound wave as described. If the wave arrives at the detector side, it induces a voltage pulse in the detector coil, which generates a stop pulse via a trigger circuit at time t_{Stop} . The sought-after location information can now be determined by the following simple relation Eq. (3.43) with v from Eq. (3.42).

$$s = v(t_{\text{Stop}} - t_{\text{Start}}). \quad (3.43)$$

In *multi-position measurement*, i.e. when measuring several positions simultaneously with one system, the current pulse generates a torsional structure-borne sound wave for each position sensor. The structure-borne sound waves arrive at the detector side one after the other and trigger *stop pulses* according to their number. The location information of the individual positions is again determined with Eq. (3.43).

A decisive feature of linear displacement sensors is the resolution of the location information. Industrial applications sometimes require resolution limits in the micrometer range. The required time resolution Δt for a spatial resolution Δs of $1 \mu\text{m}$ is 350 picoseconds (10^{-12} s). Thanks to modern electronic time measurement methods, magnetostrictive displacement sensors with resolution limits in the order of $1 \mu\text{m}$ can already be manufactured economically today. Table 3.16 shows some typical characteristics.

3.1.5.2 Housing Concepts and Applications

Profile Types

Figure 3.96 shows displacement sensors in *profile design*. The electronics are housed in an aluminum profile, which also provides a cylindrical channel for accommodating the measuring section. The aluminium profile is closed with two end caps and a hermetically sealed housing with protection class IP67 is obtained. The magnets of the position transmitter act on the waveguide through the wall of the aluminium profile.

Two variants of the position encoder can be distinguished: *free* or *guided* position encoders. Free position encoders are attached directly to the moving machine part to be measured and move along the profile with the part at a certain distance from the profile. It is advantageous that no great demands are made on the guiding precision. The sensors tolerate a lateral offset corresponding to the length of the position encoder as well as a

Table 3.16 Typical characteristics of magnetostrictive displacement sensors

Measuring lengths	25 mm to 8000 mm
Resolution	1 μm digital/1 μm analog
Repeatability	1 μm
Linearity deviation +/-	30 μm
Measured value rate	Up to 4 kHz
Distance system—position encoder axial/radial	Up to 15 mm
Multi-magnet technology	Up to 20 magnets
Housing	Metal/aluminium/stainless steel
Protection tube pressure resistant	1000 cash
Vibration resistance	20 g (10 Hz to 2 kHz) IEC60068–2-6
Shock resistance	150 g (6 ms) IEC60068–2-27

**Fig. 3.96** Displacement sensors in profile design (Factory photo: Balluff GmbH)

height offset of up to 20 mm. If even these generous tolerances cannot be met, guided position encoders are often used. With these, the profile housing of the position sensor also acts as a slide rail on which the position sensor runs as a carriage. An articulated rod with ball heads compensates for even highly unparallel movements.

Rod Designs

Figure 3.97 shows examples of magnetostrictive displacement sensors in *rod design*. This type of construction is most commonly used in *hydraulic drives*. Installation in the pressure range of a hydraulic cylinder requires the same pressure resistance from the displacement sensor as for the hydraulic cylinder. In practice, pressures of up to 1000 bar are achieved. The electronics are installed in a housing made of aluminium or stainless steel, the waveguide in a pressure-resistant tube of non-magnetic stainless steel, which is closed on the front side by a welded plug. The flange on the opposite side seals the high pressure area via an O-ring seal. A position sensor ring with magnets inserted in it slides on the tube with the waveguide.



Fig. 3.97 Displacement sensors in rod designs

Explosion-Proof Versions

Many applications require the use of displacement sensors in hazardous areas. Pressure-encapsulated magnetostrictive displacement transducers in various designs are available for use in Zone 0 or 1.

Redundantly Designed Sensors for Safety Applications

Magnetostrictive displacement sensors are ideally suited for applications requiring *high safety* or *availability*. They are often designed with double or even triple redundancy to ensure self-monitoring or to have a backup channel if necessary.

In order to obtain a triple redundant sensor, three measuring sections are accommodated next to each other, offset by 120° , in a common protective tube through which a position sensor is guided, as with the standard designs. The magnets of the position transmitter act on all three measuring sections. The positions are evaluated by three independent and completely separate electronics, which may, however, be accommodated in the same housing. Applications can be found in *ship propulsion systems*, *power stations* or in the *tilting technology* of trains.

Multiple Positions: One Sensor

Magnetostrictive displacement sensors, as described above, offer the possibility to detect two or more independent positions with a single displacement sensor. If, for example, a second position sensor magnet is placed over the waveguide, an initiating current pulse can simultaneously trigger two torsional structure-borne sound waves, each at the location of the position sensor magnet. These waves arrive at the detector coil with a time delay and generate two stop pulses. If the electronics are able to process both stop pulses, two independent position determinations are carried out. This technique is often used on *injection molding machines* (Fig. 3.98) to detect the position of the clamping unit and the ejector with a single displacement sensor. In the past, for example, two potentiometer displacement sensors were used for this task.

Figure 3.99 shows a displacement sensor in round profile design with two position encoders. The position of the respective measuring range is freely programmable for both

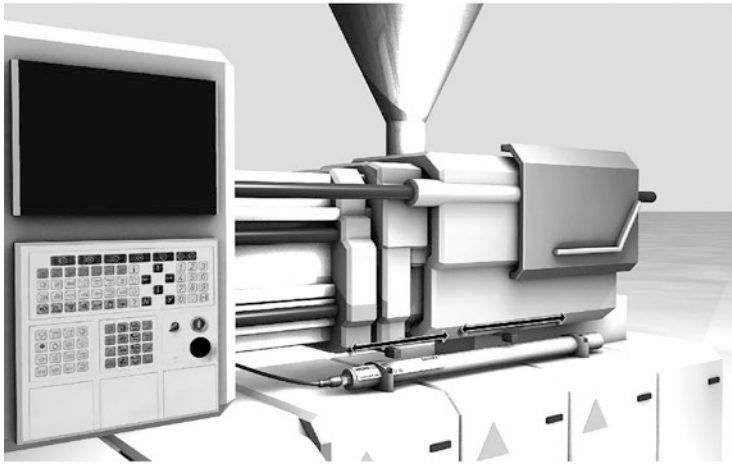


Fig. 3.98 Injection moulding machine. The positions of the clamping unit and ejector are detected by a displacement sensor with two position sensors (Factory photo: Balluff GmbH)

position encoders over the entire nominal measuring length of the displacement sensor; even overlapping of the measuring ranges is permitted.

Position-Controlled Hydraulic Drive

Magnetostrictive displacement sensors can be used to build position-controlled hydraulic drives that combine the superior force development and dynamics of hydraulic drives with the fine positioning capability of electric axes. For position control, continuous feedback of the actual position is required. In most cases, magnetostrictive displacement sensors in rod design, which are directly integrated into the cylinder, take over this task.

The protective tube of the waveguide is inserted in a slotted hole in the piston (Fig. 3.100). The piston carries the ring-shaped position transmitter with the permanent magnets.

The stainless steel protective tube of the waveguide is exposed to hydraulic pressure and is therefore pressure-resistant up to 1000 bar. The high-pressure area is sealed at the flange of the displacement sensor with an O-ring or flat seal.

An excellent control behaviour with high dynamics is achieved by synchronising the displacement measurement with the controller electronics. The controller electronics triggers the start pulse t_{Start} for the next displacement measurement and thus the new position determination of the position sensor integrated in the piston, synchronized with the read-in cycle of the current digital displacement information. Synchronization prevents flowing time differences that arise in asynchronous mode due to differences in the internal clocking of the position measuring system and the cycle time of the control electronics.

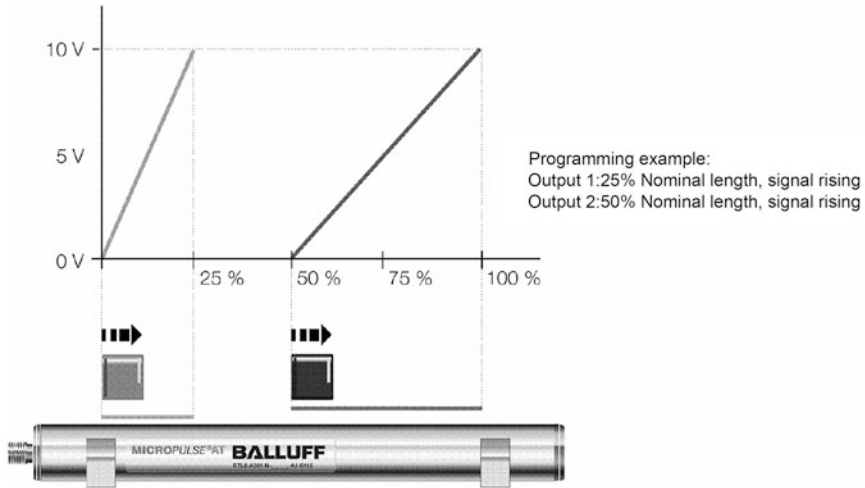


Fig. 3.99 Displacement sensor for independent detection of two positions (Factory photo: Balluff GmbH)



Fig. 3.100 Magnetostrictive displacement sensors as position feedback in hydraulic axes (Factory photo: Balluff GmbH)

Control of the Angle of Attack in Fluid Mechanics

In many fluidic machines and systems, setting the ideal operating point requires a control of the angle of attack of the machine part around which the flow passes. Examples can be found in *wind*, *thermal* and *hydroelectric power plants* as well as in *ship propulsion systems* and *large fans*. In all applications, a position-controlled linear movement must be converted into a rotary movement around the longitudinal axis of the machine part. For this reason, a controlled hydraulic drive is usually used as the drive, which provides the required—usually quite large—force and dynamics in a very small space.

In wind turbines (Fig. 3.101), controlling the *angle of attack* of the rotor blades is crucial for the achievable *energy yield* and for the *safety* of the turbine under strong wind conditions. It is important to adjust the ideal pitch angle of the rotor blades, at which the energy yield achieved is maximum at a relatively constant speed. The larger the wind turbine, the sooner the costs of these measures will be amortised. For this reason, large wind turbines today already have rotors whose blades are *individually* adjusted in their angle of attack, because the wind conditions at height differ markedly from those at ground level. As a result, the ideal angle of attack varies during each revolution and is consequently continuously optimized during the course of the revolution. Magnetostrictive displacement sensors are used as the preferred displacement sensor for position feedback in these hydraulic drives. These can be installed directly in the hydraulic cylinder as a rod design or mounted next to the hydraulic cylinder as a profile design.

Conclusion

Magnetostrictive displacement sensors have become firmly established in plant engineering and automation technology. Areas of application in which high reliability combined with precision are required are typical areas of application for magnetostrictive displacement transducers (Table 3.17).

Integrable or compact with measuring lengths from 25 mm to 8000 mm, these displacement sensors are universally applicable. The non-contact operating principle of the systems guarantees freedom from wear and an almost infinite service life. The high-precision output signal is available to the controller as an absolute signal.

The function-specific clocking of the output signal and the ratio of the housing length to the measuring length for short measuring distances (<50 mm) as well as the rigid structure of the waveguide system limit the application spectrum slightly (Table 3.18).

Energy efficiency, flexibility, safety, higher accuracy and speed are the requirements for future machine generations. To achieve this, efficient, flexible, closed-loop controlled and fast drive technology is necessary. Magnetostrictive displacement transducers for recording actual values in the drive train create excellent conditions for achieving this.

3.1.6 Displacement Sensors with Magnetically Coded Measuring Standard

3.1.6.1 Measuring Principle

This measuring system consists of two parts, a *measuring standard* and a *sensor head* (Fig. 3.102). The sensor head moves over the measuring standard without contact. The measuring standard is referred to as the measuring body in the following. It consists of a differently magnetized permanent magnetic material, for example ferrite or magnetic “plastic” on a carrier. As mechanical protection, the measuring scale can be covered by a stainless steel strip.



Fig. 3.101 Control of the pitch angle of the rotor blades of a wind turbine

Table 3.17 Typical applications of magnetostrictive displacement transducers

Markets	Application and tasks of magnetostrictive linear position measuring systems
Energy	Measurement of the displacement of large valves.
Energy	Controls the rotor blade position of wind turbines.
Energy	Measures the turbine position and the inflow of Kaplan turbines in run-of-river power plants.
Mechanical engineering	Reliably detects the closing process in injection molding machines.
Mechanical engineering	Controls the tool position in concrete block machines under extreme conditions.
Automotive production	Measures the highly dynamic joining process in rivet-free “clinching” systems.
Automotive production	Ensures a drip-free adhesive bond in the dosing cylinder.
Food production	Controls precisely, reproducibly and independently of foam the filling quantities of cans in filling lines.

Alternating magnetic poles are incorporated in the permanent magnetic material in the direction of travel. The pole width in the direction of travel is constant and ranges from 0.5 mm to over 10 mm depending on the type. In transverse direction, the expansion of the poles is usually about 50% or 100% of the dimensional element width. This is usually 5 mm to 20 mm. The entire measuring body has a thickness of about 1.3 mm. The magnetic material accounts for 1 mm of this thickness. The sensor head moves above it at a distance of almost 0 mm to 5 mm or 10 mm. It measures the magnetic field of the measuring body via two sensors. The measured signals have a sinusoidal and cosinusoidal shape above each pole. With an *analog* interface, these signals are output as analog differential signals. The controller uses these signals to calculate the position within the pole. With an incremental *digital* interface, the sensor itself calculates its *position* within a pole from these signals in

Table 3.18 Basic limits of magnetostrictive displacement sensors

Linearity deviation	If measurement results with non-linearities $< \pm 30 \mu\text{m}$ are required, other technologies are better suited (e.g. optical glass scales or magnetic tape systems).
Measured value rate	Magnetostrictive displacement transducers are clocked systems. The measuring rate depends on the length and the propagation time of the structure-borne sound wave. The typical velocity of the structure-borne sound wave is 2850 m/s. If measured value rates $> 4 \text{ kHz}$ are required, other operating principles are more appropriate.
Short measuring lengths (under 50 mm)	Due to the necessary damping zone and the non-usable area on the transducer system, short lengths result in an unfavorable ratio of the housing length to the measuring path.
Large measuring lengths (over 8000 mm)	Technically feasible. However, handling and installation are very complex and not without problems.

Fig. 3.102 Measuring element and sensor head (Factory photo: Balluff GmbH)

real time, within a few ns. A digital sensor outputs each change in position in the form of positive or negative increments via two signals. The signals are usually referred to as A/B or 90° pulses. The coding of the increments is described in Sect. 3.2.7. The sensor “answers” on its own initiative—whenever it has detected a change in position, in the form of a *signed increment*. In this way, the connected controller knows the exact position of the sensor at any time (*real-time capability*). The controller does not have to “poll” the sensor.

The following interfaces are available for incremental measuring systems:

- A digital A/B voltage signal, which is either physically transmitted as a differential signal (RS422) or single-ended signal (HTL).

This signal has two important characteristics:

1. The *resolution* (distance/increment) of the sensor head. It ranges from less than 1 μm to a few mm. The interpolator determines this resolution, which is significantly higher than the pole width. Its principle mode of operation is described in Sect. 3.2.7.
 2. The minimum temporal resolution, the edge separation, can be selected in the range from about 100 ns to 100 ms. It must be adapted to the downstream evaluation electronics, because pulses must never be output faster than the downstream control can count them. The maximum travel speed depends on the resolution and the minimum edge separation. The relationship is described in Sect. 3.2.7.
- An analog sine and cosine *voltage difference signal* whose period is the pole width of the measuring body. It is usually referred to as a sin/cos or 1Vpp signal. Digitization takes place in the subsequent evaluation electronics. The technical description can be found in Sect. 3.2.6.

Optionally, the incremental measuring system also offers a reference pulse. Then the measuring body is divided into two tracks, the *incremental* and *reference track* (see Fig. 3.102, middle system). Each track is scanned by a separate sensor. In digital systems, a reference pulse is output for the path of one increment at the point where a reference pole is located on the reference track in the measuring body. This reference pulse has a repeat accuracy better than one increment. For analog (1Vpp) systems, the width of the reference signal is approximately $\frac{1}{2}$ pole width.

When switching on the measuring system, a *reference run* must be carried out up to the reference pulse. In the worst case, this could be the entire measuring length. If the time required for this maximum travel distance is too long, several reference points can be placed at defined positions in the measuring body: With this distance-coded measuring body, the distance between two reference points is always different and therefore unique. The reference movement must then only be performed via two reference points; then the absolute position is uniquely determined. With this measure, the maximum traverse path for referencing can be significantly reduced.

Figure 3.103 shows on the left the distance between two reference points for a *distance-coded measuring body* with a 5-mm pole pitch. The distance between two reference pulses is a pole width or a multiple thereof. The position of the reference points is shown to scale on the right. The basic distance, the maximum distance between two reference points, is 80 mm in the example. The next distances are then each one pole width, in this case 5 mm, different from the previous, or pre-previous distance. The first distance corresponds to half the basic distance. The distances are thus 40 mm –45 mm –35 mm –50 mm to 5 mm –80 mm. It can be seen that there are 16 clear distances between two reference points. The maximum measuring length, in which the distances between the reference points are unambiguous, is a function of the pole width (the minimum distance of the reference pulses) and the base distance.

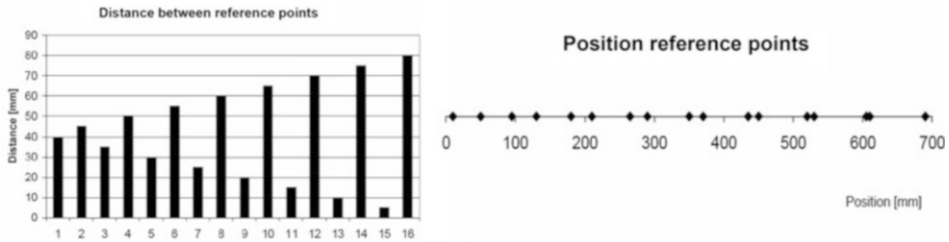


Fig. 3.103 Position of reference points for distance-coded measuring body

3.1.6.2 Design and Function of Incremental and Absolute Measuring Systems

Incremental Magnetic Measuring Systems

The sensor head evaluates the magnetic field via special sensors, which usually work according to the magnetoresistive (MR) effect. Depending on the application, sensors with special MR effects are used: AMR Anisotropic Magnet Resistive, GMR Giant Magnet Resistive, TMR Tunnel Magnet Resistive, CMR Colossal Magnet Resistive (Sect. 2.3).

Two sensor bridges, which are mechanically offset by 90° relative to the measured signal (normally one pole width), form the basis for the measurement. In the following, they will be referred to as *sin* and *cos* sensors. Usually, the structure of the sensor is adapted to the pole width. With small pole widths, the length of the sensor is even greater than one pole. In this case, averaging of the signal over several poles is possible. This increases the signal-to-noise ratio and, with suitable geometry, reduces the measured harmonics. For pole widths up to about 5 mm, the sine and cos sensors are located together in one chip. For larger pole widths, discrete sensors are used. In the evaluation electronics in the sensor head, the offset, amplitude and in some cases the phase of the sine and cos signals are adjusted to each other. This adjustment is partly done adaptively during the movement. These signals form the basis for the position measurement. Either the sin and cos signal (Sect. 3.2.6) or a digitized A/B signal (Sect. 3.2.7) is used as an interface.

A reference sensor emits a reference pulse at a defined position in the pole.

Some sensors have two further magnetically switched outputs which, in conjunction with additional permanent magnets, can take over the function of limit switches on the movement area.

There are two basic forms of execution:

- *incremental measuring system* with optional reference point and optional limit switches and an
- *absolute measuring system* with optional real-time interface.

The following metrological data are available:

- Real-time system with a conversion time of several tens of ns,
- Resolution $\leq 1 \mu\text{m}$,

- System accuracy (linearity deviation) up to $\pm 5 \mu\text{m}$,
- Hysteresis $\ll 1 \mu\text{m}$,
- Maximum travel speed over 20 m/s.

Common incremental real-time interfaces:

- 1Vpp (sin/cos),
- RS422,
- HTL (single-ended).

Usual absolute interfaces:

- SSI,
- BiSS,
- Hiperface.

Depending on the pole width, the possible air gap varies from almost 0 mm to over 5 mm (approx. 30% of the pole width). There are extremely small designs, down to $(b \cdot h \cdot l)$ 12 · 13 · 35 mm³. Even with this design, no additional amplifier electronics are required. Connectors and cable variants are available.

Absolute Magnetic Measuring Systems

In incremental systems whose magnetic field sensors are based on an MR effect, the sensor head cannot distinguish between the individual poles. Each pole is the same for the sensor head. The position within the pole is unique. For an absolute measuring system, the *pole number* n_{pol} must be recognized in addition to the position within the pole. Equation (3.15) shows the connection:

$$s_{\text{abs}} = s_{\text{pol}} + n_{\text{pol}} \cdot P. \quad (3.44)$$

Here P corresponds to the pole width. There are various methods for determining the pole number. A *parallel coding* would be conceivable, whereby each bit would be assigned its own track and sensor. However, this would result in a high number of parallel tracks and thus a very wide design of the sensor and the measuring body.

Instead, it is common practice to arrange the absolute coding in *serial form* in one track. A 4-bit PRC coding (pseudo random code or chain code) is described as an example in Fig. 3.104.

The upper horizontal column in Fig. 3.104 shows a serial arrangement for 16 values/positions for a PRC code. The serial 4-bit coding is scanned by 4 digital sensors (s_1, s_2, s_3, s_4). The 16 possible positions of the 4 sensors are symbolized below each other by the horizontal identifiers s_1, s_2, s_3, s_4 . In the middle below the code, the position of the sensors in the 16 possible positions is indicated. In the line of the respective positions, the consecutive pole number is indicated on the left and right. On the right side of the figure, the sensor values of the four sensors in the respective position are entered. The pole number

PRC Code	0	0	0	0	1	1	1	1	0	0	1	0	1	1	0	1	0	0	0
Position, Pole no.																			
0	s1	s2	s3	s4															
1		s1	s2	s3	s4														
2			s1	s2	s3	s4													
3				s1	s2	s3	s4												
4					s1	s2	s3	s4											
5						s1	s2	s3	s4										
6							s1	s2	s3	s4									
7								s1	s2	s3	s4								
8									s1	s2	s3	s4							
9										s1	s2	s3	s4						
10											s1	s2	s3	s4					
11												s1	s2	s3	s4				
12													s1	s2	s3	s4			
13														s1	s2	s3	s4		
14															s1	s2	s3	s4	
15																s1	s2	s3	s4

Pole number	Signals of the four sensors			
	s1	s2	s3	s4
0	0	0	0	0
1	0	0	0	1
2	0	0	1	1
3	0	1	1	1
4	1	1	1	1
5	1	1	1	0
6	1	1	0	0
7	1	0	0	1
8	0	0	1	0
9	0	1	0	1
10	1	0	1	1
11	0	1	1	0
12	1	1	0	1
13	1	0	1	0
14	0	1	0	0
15	1	0	0	0

Fig. 3.104 Example for 4-bit PRC code

derived from this, which the four sensors see, is entered in each case in the right table to the left of the binary sensor value representation. You can see that all possible 16 positions also appear in the serial form. Thus, it is possible to infer the pole number between 0 and 15 from the 4 sensors at each position, for example via a lookup table. If the binary values 0 and 1 are represented by N and S poles, sensors that can distinguish between N and S poles (e.g. Hall sensors) can be used to draw a clear conclusion about the number 0–15. In this way, the pole number required in Eq. (3.15) can be determined.

Figure 3.105 shows an absolute dimensional body. At the bottom is the incremental track, at the top the absolute track with PRC code. The coding of the absolute track is symbolized below with 0 and 1. As shown in the left part, it may happen that when switching on the absolute sensors (Dig0_b, Dig1_b, Dig2_b, Dig3_b) are exactly in the middle between two poles of the upper absolute track. If the poles are changed, the magnetic fields used there are exactly zero and therefore not valid. For this reason, each absolute sensor is designed as a *pair of sensors* (Digi_a, Digi_b, right in the figure). The two sensors have a distance of 1/2 pole width. One of the two sensors is always in a valid range above the measuring body. Depending on the incremental sensor (referred to as Inc. in Fig. 3.105), the sensor head can decide at which point in the pole it is located and then use the respectively valid sensors of series a or b to determine the absolute position.

Interfaces of absolute encoders transmit the absolute position in a digitized serial form (e.g. SSI, BiSS, CAN, Profibus, EnDat, Hyperface). The serial protocol always requires a finite time.

Various interfaces (SSI, BiSS) freeze the measured value at the time of the start of the request and then transmit this frozen value. After the data has been transmitted, the control system thus knows the position at the time the query was started and can set up a control loop accordingly. However, the *position data is not available in real time*.

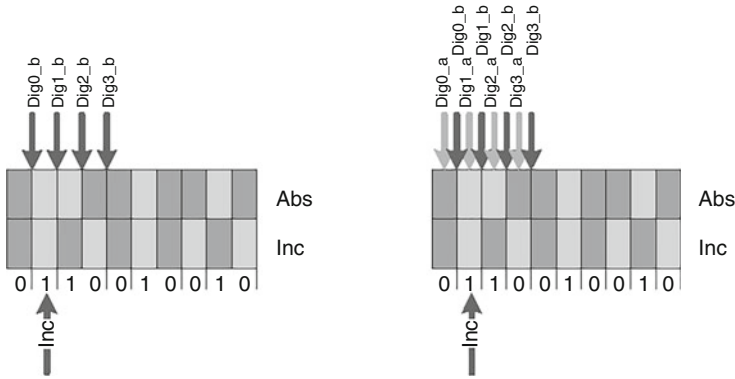


Fig. 3.105 Implementation of the absolute sensors

For real-time capability, a real-time signal (sin/cos or A/B) must be transmitted in addition to the digitized position signal. In principle, this is possible with any interface; it has been implemented so far with EnDat, Hiperface and BiSS-sin/cos BiSS-A/B SSI-sin/cos, SSI-A/B.

Often it is sufficient to read the absolute position only when switching on and then to continue working only with the real-time capable, incremental signals. In this case, no special requirements are placed on the speed of the absolute interface. After switching on, it usually only serves as a redundant system.

Various hardware standards have become established for absolute position transmission, all of which work with differential signals, some of which are given as examples: Bidirectional bus (Hiperface), a unidirectional clock and a bidirectional data line (EnDat), a unidirectional clock and data line (SSI, BiSS). Table 3.19 shows the most important properties of a selection of the absolute interfaces.

3.1.6.3 Characteristic Values

Linearity Deviation

Generally, the linearity deviation (LA) or system accuracy is defined as the *difference* between the *measuring system* and a *reference system*.

In magnetic measuring systems, the LA depends on the measuring standard and the sensor head. When comparing different measuring systems, it is important to consider more than just the numerical value; it is important to know how the numerical value is defined. Basically, there are different definitions for the linearity deviation in this area. The most important definitions are mentioned here:

The softer (cumulative) or qualitatively less demanding: Here the linearity deviation *per meter* is given. The longer the gauge length, the greater the maximum linearity deviation. For example:

Table 3.19 Properties of absolute interfaces

	SSI	BiSS	EnDat	Hiperface	CAN	Profibus
Hardware (each signal is transmitted as a differential signal)	Clock (unidir) Data (unidir)	Hardware compatible to SSI Clock (unidir) data (unidir)	Clock (unidir) Data (bidir)	Bidirectional bus	Bidirectional bus	Bidirectional bus
Topology	Stub line	Stub line with 8 slaves	Stub line	Stub line	Bus	Bus
Data transmission direction	Sensor → control	Bidirectional	Bidirectional	Bidirectional	Bidirectional	Bidirectional
Electronic data sheet possible	No	Yes	Yes	Yes	Yes	Yes
Optional real-time signal	SSI-sin/cos, SSI-A/B	BiSS-sin/cos BiSS-A/B	Yes	Yes	No	No

$$LA = \pm(k_0 + k_1 \cdot l) \quad (3.45)$$

with k_0 , k_1 : constants, l : measuring length in m.

The constant k_0 stands for the *non-linearity* of the sensor head. It is repeated in each pole and therefore has a very short-wave behaviour. k_1 represents the *measuring body*. For k_1 , there are usually different quality levels of the measuring body. The influence of the measuring body is very long-wave (in the decimetre range). As an example we assume $k_0 = 2 \mu\text{m}$, $k_1 = 10 \mu\text{m/m}$. A measuring section of 4-m length thus has an LA of about $\pm 40 \mu\text{m}$. This figure, however, does not yet tell us anything about the actual course of the LA. The LA could have the course as shown in Fig. 3.106: Within a short piece, the entire linearity stroke of $\pm 40 \mu\text{m}$ can be traversed (*microlinearity*). This definition makes no demands on microlinearity.

The stricter definition, shown in Fig. 3.107, guarantees that the maximum linearity deviation within *any one meter* is never greater than the maximum LA (k_1). To illustrate the facts of the hard definition, 5 rectangles with a length of 1 m each and an LA of $\pm 10 \mu\text{m}$ are drawn in the diagram at various points on the dark curve. In the worst case, this hard definition reaches the limits of $\pm 40 \mu\text{m}$ over its *entire length* (continuous dark curve up to about 2.7 m and then the dashed lighter curve). Typically, however, the entire LA is lower (solid curve). With this increased demand on linearity, the derivation of the linearity $\Delta LA / \Delta s$ is significantly lower than in Fig. 3.106. This difference in quality, which comes from the different definition, has a significant effect, for example, on the noise or heat generation in linear actuators or on the durability of the bearings of the actuators. Different manufacturers use different definitions of the linearity deviation.

Hysteresis

The hysteresis of (magnetic) measuring systems can be easily determined by recording the linearity deviation in forward and reverse direction (Figs. 3.108 and 3.109). The width of the curve (short arrows) corresponds to the resolution of the test sample, here $1 \mu\text{m}$. The width results from the fact that the reference system has a much higher resolution than the test sample. The linearity deviation corresponds to an *envelope curve* of the wide curve. The maximum linearity deviation is therefore determined by the difference between the maximum and minimum of the envelope curve. Similarly, the hysteresis corresponds to the difference of the envelope curves between the linearity deviation of the forward and backward direction (long arrows). In Fig. 3.109, the hysteresis is about $0.2 \mu\text{m}$.

In addition to the hysteresis, the *tracking error* (speed-dependent) is measured if necessary.

Following Error

Running or dead times in the signal path in the sensor head result in a tracking error, which can reach a much higher value than the hysteresis. The influence of the tracking error on the

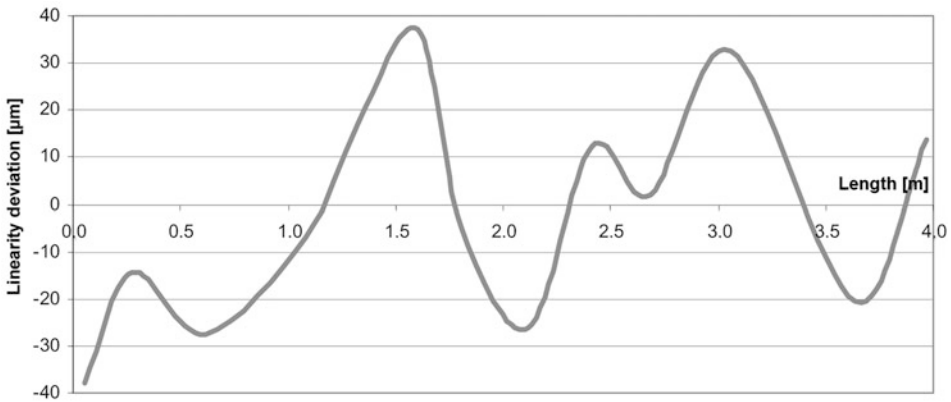


Fig. 3.106 Soft linearity definition

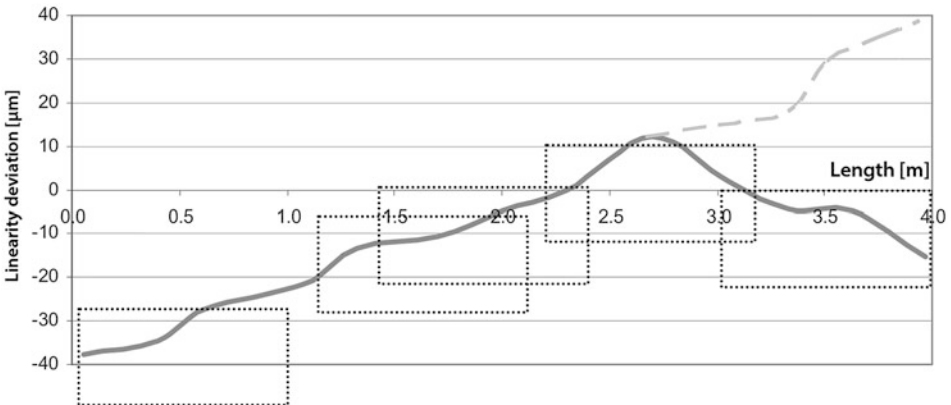


Fig. 3.107 Hard linearity definition

hysteresis measurement can be easily measured by changing the speed. If the movement is very slow, a following error has a very small influence on the measured hysteresis.

3.1.6.4 Sensor Types in Comparison

The principle of measuring the magnetic field can be used for linear movements and also for rotative movements. Magnetic sensors must first be divided into *incremental* and *absolute* measuring systems. For both systems, the *pole width* (PB)—and with it the maximum distance (the greater PB, the greater the maximum air gap) and the achievable accuracy (the greater PB, the lower the accuracy)—is decisive.

The *electrical interface* is important for the permissible cable lengths or the achievable travel speeds. In most cases, the *mechanical design* with the corresponding *travel direction* also plays a major role. Depending on the application, *limit switches* integrated in the

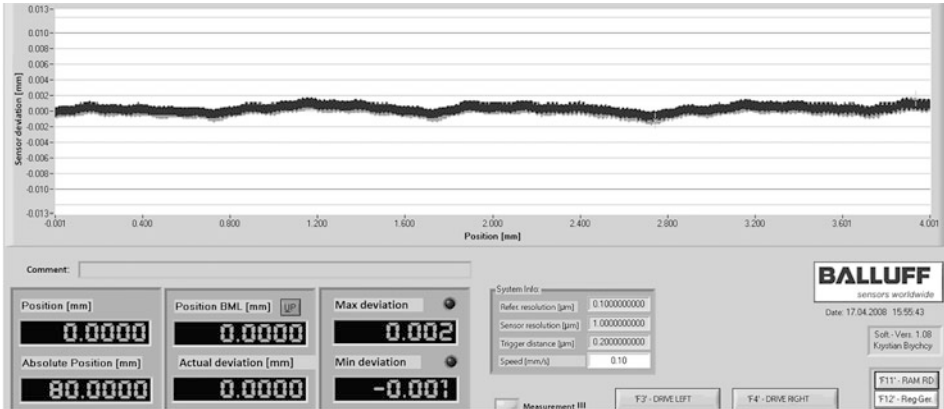
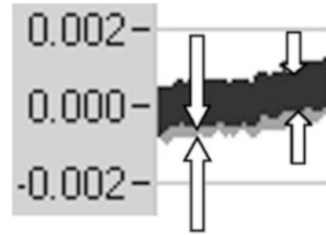


Fig. 3.108 Measuring the hysteresis

Fig. 3.109 Zoomed measuring range left hysteresis becomes visible



sensor head are helpful. Some designs are available with plug. Table 3.20 lists the properties of different sensor types.

3.1.6.5 Application Examples

Figure 3.110 shows a 3D model of a linear drive. The sensor head can be seen at the bottom right of the figure. The linear drive is controlled exclusively with the magnetic measuring system.

Figure 3.111 shows another special application of a measuring element, in which the magnetic measuring element is integrated in a rod and coated with a stainless steel layer (right figure). The measuring element is magnetized alternately on the circumference in the axial direction. The magnetic poles are made visible in the left picture by a pole pitch display card. It darkens at the poles and becomes light between the poles. With the appropriate sensor head, the position, and, derived from this, the speed and acceleration of the rod can be recorded very precisely and dynamically. A rotation around the rod axis is not important for the measurement.

Table 3.20 Selection of sensor types and their properties

	BML-S1A	BML-S1B/E	BML-S1F1	BML-S1F2
				
Pole width [mm]	1	5	1	1
Air gap [mm]	0.01 to 0.35	0.01 to 2	0.01 to 0.35	0.01 to 0.35
System accuracy [μ]	± 10	± 20	± 10	± 10
Interface	RS422/1Vpp	RS422/HTL	RS422/1Vpp	RS422/1Vpp
Limit switch available	Yes	Yes	No	No
Connector version available	Yes	No	No	No
Supply voltage [V]	5	5, 10 ... 30	5	5
Size ($l \cdot b \cdot h$) [mm ³]	60 · 11 · 14	35 · 10 · 25	35 · 12 · 13	35 · 12 · 13

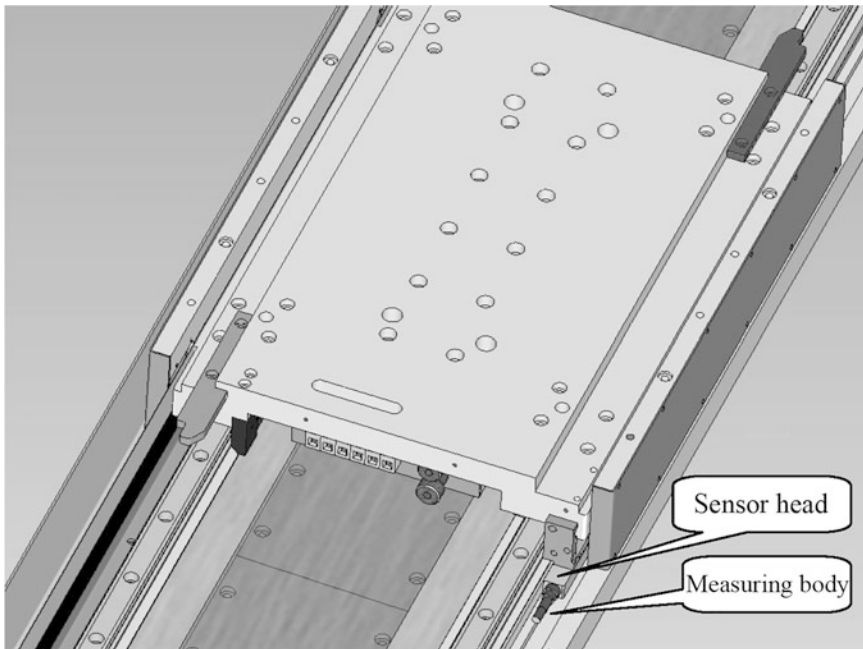


Fig. 3.110 3D model of a linear drive

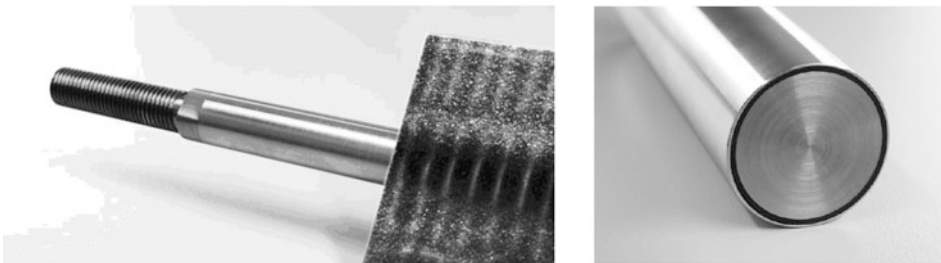


Fig. 3.111 Image of a dimensional body on a rod (Factory photo: Magnopol GmbH & Co. KG)

3.2 Sensors for Angle and Rotation

Overview

Sensors for angle and rotation convert a *mechanical angle* into an *electrical signal*. There are different terms for these sensors, such as encoder, angle encoder, angle coder, rotation sensor, rotation measuring systems, incremental/absolute encoder or rotary encoder. In this section, the term rotary encoder will be used preferably.

Fig. 3.112 Schematic representation of an encoder

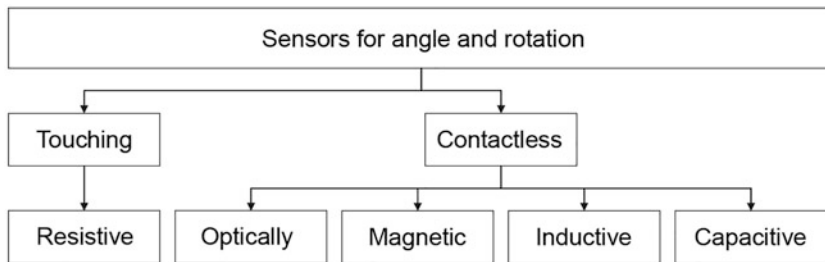
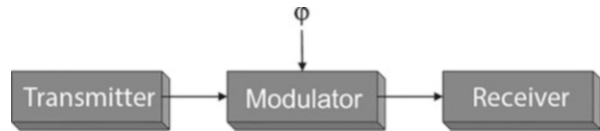


Fig. 3.113 Classification of encoders according to the operating principle

An encoder basically consists of *three elements* (Fig. 3.112; φ : angle). The *transmitter* introduces energy into the system; the *modulator* changes the introduced energy in proportion to the mechanical angle and thus serves as a measuring standard and the *receiver* converts the modulated physical quantity into an electrical signal.

As Fig. 3.113 shows, the encoders can be divided into contacting and non-contacting encoders and are further subdivided according to the physical principle.

The characteristics of the different encoder technologies are listed in Table 3.21.

Designs of Encoders

However, encoders also differ in the electrical interfaces used (Fig. 3.114). Basic encoder characteristics can be explained by means of the electrical interface.

For encoders with *incremental interface* (increment as elementary step or countable interval), the *angle information* is output *relatively*. This means that only angle changes caused by signal changes are displayed, not the absolute angle information. There are two versions: *square-wave* and *sinusoidal signals*.

In the simplest form, the incremental encoders with square-wave signals provide a single pulse line (Fig. 3.115 left). This allows only the determination of an angle change. If this one pulse line is extended by a second one, phase-shifted by 90° , a *quadrature signal* is obtained. In addition to the angle change, this allows the *direction of rotation to be* recognized by the signal edge sequence. At the same time, the resolution is doubled with the same number of pulses per line per revolution. To get the assignment to a reference point, a zero pulse information can be provided, whereby a *quasi-absolute position* can be determined (Fig. 3.115 middle). However, this reference point must be traversed once by a reference run when the encoder is switched on. The electrical specification for the pulse line for single-channel or multi-channel incremental rotary encoders can be of various types (e.g. TTL, HTL, open collector).

Table 3.21 Characteristics of the various encoders (–: below average; 0: average; +: above average and ++: very good)

Property	Resistive	Visual	Magnetic	Inductive	Capacitive
Accuracy	–	++	+	+	+
Compact structure	–	O	O	+	+
Insensitive to shock/vibration	+		+	+	+
Own consumption	+	O	–	–	+
Mounting tolerances	O	–	O	O	+
Immunity to magnetic fields	+	+	–	O	+
Temperature behaviour	–	O	O	+(resolver)	O
Sensitivity to humidity	–	–	+	+	+
Lifetime	–	–	+	+	+

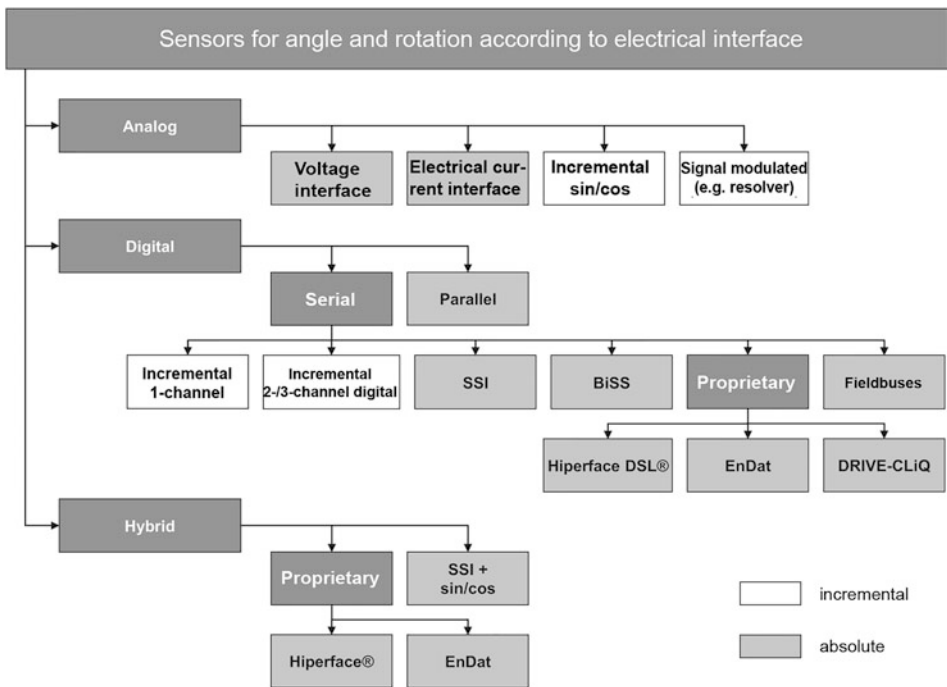


Fig. 3.114 Designs of sensors for angle and rotary motion, differentiated according to the electrical interfaces used

In incremental encoders with analog quadrature signals, *analog sine and cosine signals* are used to represent an angle (Fig. 3.115 right). These signals are simultaneously scanned, digitally converted and interpolated by the control to determine an angular position within one sine/cosine period. Interpolation in this context means the following calculation:

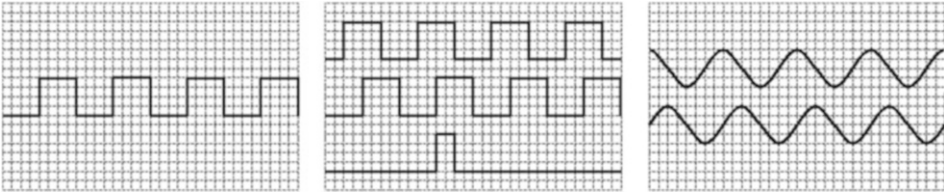


Fig. 3.115 Incremental signals—1-channel (left), 3-channel (center) and sin/cos (right)

$$a = \frac{1}{n} \arctan \left(\frac{A_{\sin}}{B_{\cos}} \right)$$

Here, A_{\sin} or B_{\cos} is the instantaneous value of the sine or cosine signal; n is the number of sine/cosine periods per revolution; a is the calculated angle. This *significantly increases* the *resolution* compared to digital incremental systems with the same code disk at the same basic resolution (i.e. number of sine/cosine periods compared to number of increments per revolution). For example, a system with 1024 sine/cosine periods and a 10-bit A/D conversion can achieve a resolution of 22 bits ($2^{22} = 4,194,304$), corresponding to 86 μgrad or 0.3 angular seconds per revolution (2 bits is obtained by squaring the sine and cosine).

In addition, for analog quadrature signals, *status* and *error monitoring* can be performed using the following formula, which applies to each angular position:

$$\sin^2 + \cos^2 = 1 \quad \text{or} \quad \sin^2 + \cos^2 = \text{const.}$$

Another advantage of analog incremental systems over digital ones is the *lower transmission requirements*, since the signal frequency of the signals is much lower. In an analog incremental system with 1024 sine/cosine periods per revolution, the signal frequency at 12,000 revolutions per minute is 204.8 kHz. With a comparable digital incremental system for 22-bit resolution (20-bit increments), this would be about 210 MHz. This is not practicable, since especially today's control units in industrial environments provide a maximum counting frequency of less than one MHz and the cables from the encoder to the evaluation unit are usually several meters up to 100 m long.

Analog incremental signals are evaluated in two ways (Fig. 3.116). On the one hand, the position within a period is finely resolved via the previously mentioned interpolation. To record the position over several periods, the signals digitized via Schmitt triggers are counted. The electrical specification for the signal line is usually such that the signals are provided differentially with 1 V_{pp}.

In contrast to incremental systems, absolute encoders indicate an absolute angular position at any time, especially when switched on. This is important if a reference run must be omitted for application reasons.

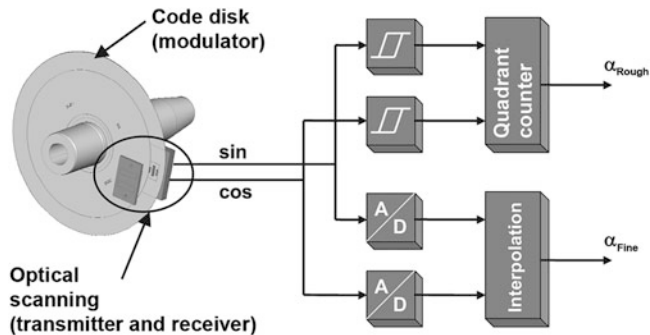


Fig. 3.116 Evaluation of analog incremental encoder signals

The internal structure of absolute and incremental encoders differs in that *additional code tracks* are scanned for incremental evaluation. In this way, the absolute information for one revolution can be determined. Possible absolute codes include binary, gray, vernier or pseudo-random codes. These are calculated in the encoder so that absolute information can be provided.

In the simplest case, the *absolute angle information* is transmitted as a *parallel code word*. However, this form is no longer very widespread today, as it is very complex and expensive to wire. Commonly used today are rather *serial interfaces*, which are either specially developed for encoders, such as SSI (synchronous serial interface), EnDat, BiSS and Hiperface DSL[®] or *field buses*, which are common in industrial plants.

Another type of encoder works with *hybrid interfaces*. Here, a *serial data channel* is combined with *incremental analog signals*. When such interfaces are used, the absolute, low-resolution information is captured after power-on and transmitted on the data channel. At the same time, the incremental position is captured by the control and offset against the absolute information. The system is thus *initialized* and work can continue with the incremental information of the analog signals. Depending on the interface, not only absolute angle information but also other data, for example for diagnosis or parameterization, can be transmitted via the data channel.

An advantage of incremental interfaces in general is that the angle information can be measured quasi in *real time*, since the sensor information is available directly at the interface and does not have to be calculated and converted into a protocol. The hybrid interfaces thus combine the advantages of incremental and absolute encoders.

Mounting and Attachment of Encoders

The design of encoders essentially depends on the sensor technology used. This is discussed in the corresponding sections of this book. However, there are some general aspects that are described here.

For encoders, it is often necessary that the basic elements: transmitter/modulator/receiver are matched to each other with sufficient *precision*. Since this can rarely be

guaranteed with the required accuracy in the end application, encoders are designed with *integral bearings*. For this purpose, the shaft carrying the modulator is coupled via ball bearings to the encoder flange, which in turn is connected to the stator of the application. With such a *self-bearing shaft encoder*, the manufacturer of the shaft encoder can precisely align the transmitter/modulator/receiver unit. In the application, care must be taken that the encoder is not rigidly connected to the shaft and stator. Small mechanical misalignments would otherwise lead to high mechanical loads, which would destroy the weakest link in the chain, usually the encoder, in a short time. *Coupling elements* are used accordingly.

Shaft couplings are used as the connecting element between the drive shaft and the encoder shaft. In dynamic drives, however, the shafts are preferably connected directly via a rigid shaft, and instead of the shaft coupling, a *stator coupling* is used as a torque support for the rotationally fixed connection of the encoder housing with the stator of the drive unit (Fig. 3.117).

Please note that the stator or shaft coupling together with the stator or shaft forms a spring-mass system capable of vibration. Their natural frequency f is:

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{J}}$$

where k is the spring constant of the coupling and J is the effective mass moment of inertia. When dimensioning the application, it must be ensured that the natural frequency is well above the maximum speed.

The *stator coupling* is advantageously designed as a spring parallelogram and compensates radial and axial movements of the drive shaft, but is at the same time torsionally rigid. Eccentricities and angular misalignments of the shafts of the drive and measuring system are also largely compensated. The installation of the encoders is simplified because, in addition to the shaft, only the stator coupling needs to be connected to the stator of the drive system by means of screws.

For rigid connection of the encoder shaft with the drive shaft to be measured, the encoder is equipped with different flanges, shaft types and shaft diameters. In addition to *solid shafts*, there are *hollow shafts in slip-on and through versions*.

Today, mechanical *wear*, especially of the ball bearing, is usually responsible for the failure of encoders, so that even *encoders without integral bearings* are justified applications. In high-performance rotary encoders without integral bearings, the measuring standard and the sensor electronics are not fixed in relation to each other at the factory. Therefore, mounting aids are required to bring the measuring standard into the correct position relative to the receiver. This positioning must be carried out reliably and accurately if the encoder is to have the required high accuracy.

However, there are also encoders without integral bearings, which, due to their design, can do without adjustment aids. This applies to systems that have a relatively low basic resolution, but which can tolerate mechanical errors of the application quite well. A classic

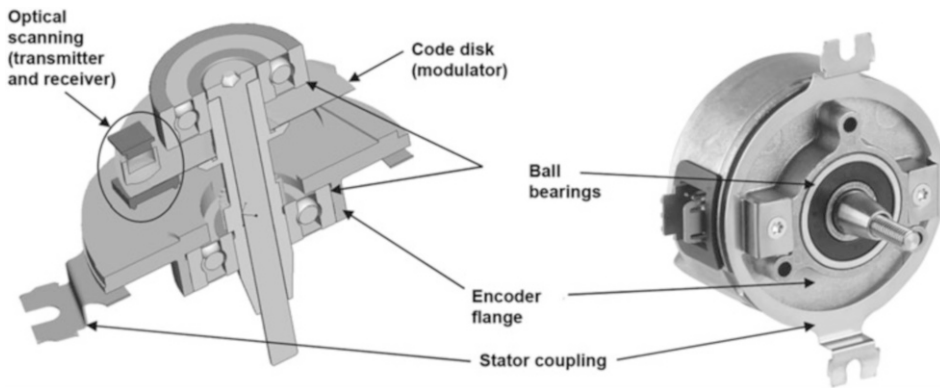


Fig. 3.117 Design of an encoder using the example of an optical system with its own bearings (Factory photo: SICK STEGMANN GMBH)

example would be resolvers (Sect. 3.2.5). In the meantime, capacitive encoders (Sect. 3.2.4) are also available, in which mechanical errors are only slightly noticeable as position errors due to the *holistic* scanning.

Resolution and Accuracy

Important *characteristics* for describing the performance of encoders are resolution and accuracy. In this section, these specifications for digital and analog incremental encoders are discussed. Since absolute rotary encoders are based on these, the corresponding errors propagate, although error compensation can be performed for absolute systems if they have a digital core.

The *resolution* describes how many measuring steps per mechanical revolution can be resolved. For rotary encoders with digital incremental interface, the resolution is defined by the number of increments. For rotary encoders with analog sine/cosine interface, the resolution results from the number of sine/cosine periods and the quality of the analog signals. This depends on how well the signals can be digitally converted, and thus primarily on their signal-to-noise ratio.

A distinction is made between several specifications in terms of *accuracy*. The *total error* is the maximum deviation of a measured value from its true value within one revolution. For encoders with analog incremental interface, a distinction is made between integral and differential non-linearity. The *integral non-linearity* indicates the measuring error over one full revolution. It describes a property that mainly refers to the accuracy of a position value and is usually caused by the mechanical structure, such as the eccentricity of the code disk. The *differential non-linearity* defines the deviation within a sine/cosine period. This is particularly relevant when deriving a speed information from the position information, since it describes inaccuracies that repeat themselves per scale period, i.e. can enter a comparatively high error frequency into the controller. The cause of this non-linearity is usually due to the imperfect calibration of signal offsets and amplitudes,

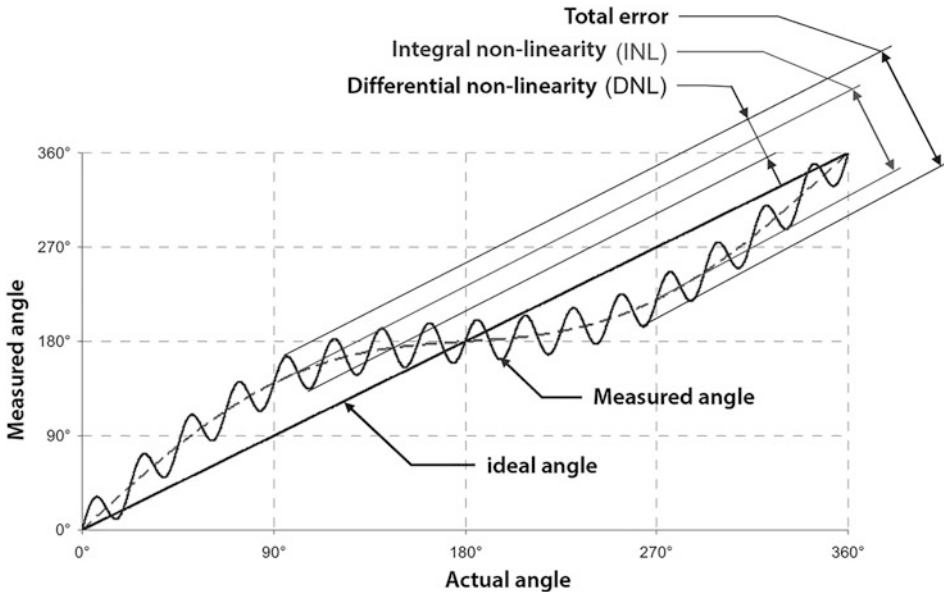


Fig. 3.118 Accuracy values of encoders with analog signal interface

the phase relationship between sine and cosine signals, or minor fluctuations in the line pattern. Figure 3.118 shows details for the accuracy specifications of rotary encoders.

It should be noted that the accuracy of encoders is not only influenced by the device itself but also by the application. Thus, *interference* on the transmission lines can lead to sporadic errors. The *input circuitry of the evaluation unit* can also introduce errors into the system, especially in the case of encoders with analog interface. If the signal paths are not optimally matched to each other, the resulting offset, amplitude or phase errors can reduce the accuracy of the entire system. The mechanical mounting of the encoder also has a great influence on the accuracy of the measurement.

Fields of Application

There are countless applications for encoders: Wherever axes rotate or rotary movements are converted into linear movements, they are used. Some examples are in servo motors as motor feedback systems, in mechanical and plant engineering, in warehousing and conveyor technology, in packaging machines, robotics and automation, in printing and paper technology, in foundry machines, the automotive industry, in elevators and in the field of renewable energies. The type of encoder used depends heavily on the application.

High-resolution optical encoders are advantageous as incremental encoders because their performance is only marginally affected by external conditions such as temperature. The reason for this is that with optical rotary encoders, which have high resolutions physically in the measuring standard, changes in offset and/or amplitude have almost no effect.

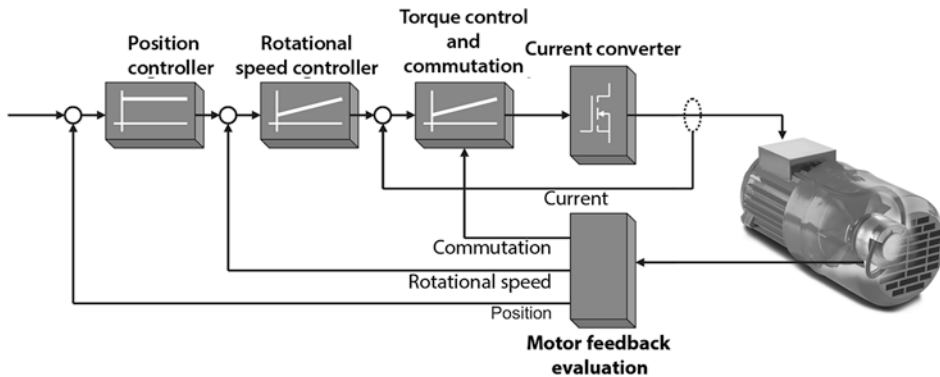


Fig. 3.119 Servo motor with cascade control and built-in motor feedback system

With absolute encoders, the fast, real-time-capable and yet simple *SSI*[®] interface is increasingly taking a back seat, as bidirectional communication capability and features such as diagnostic functions are now coming to the fore. The usual fieldbuses are used here, in recent years increasingly those based on *Industrial Ethernet*. The latter fieldbuses in particular represent such a large cost block that the technology used for the encoder core has only a minor influence on the price of the complete device. For this reason, users are increasingly opting for the more powerful optical encoders.

An important field of application for encoders is within motor control. Encoders of special design, so-called *motor feedback systems*, are installed directly in a *servo motor*. From the (real-time) position signal, all information required for classical cascade control, i.e. position, speed and commutation, is derived (Fig. 3.119).

In the control of highly dynamic servo motors, the main focus is on speed measurement. The speed control loop is often operated with cycle times of $62.5 \mu\text{s}$. This short cycle time increases the demands on speed resolution. To resolve one revolution per minute (rpm), a position resolution of at least 20 bits is required. This also makes it clear that sophisticated drive technology is the domain of optical encoders. Further requirements in this area of application are:

- high operating temperature range up to almost $150 \text{ }^\circ\text{C}$,
- high shock and vibration load of 50 g and more,
- high speeds up to over 12,000 rpm,
- Risk of contamination, for example by brake dust,
- Risk of condensation wetness.

Traditionally, resolvers were used in the past, but today, due to their dead time behaviour on the one hand and their low accuracy/resolution on the other, they are only used for servo drives in the lower performance range. With optical encoders, the limited temperature range of maximum $120 \text{ }^\circ\text{C}$ has a limiting effect. This limiting temperature is additionally influenced by the fact that the ball bearings required for high-quality optical

encoders introduce additional heat into the device at higher speeds. Sealed bearings can therefore usually not be used, which in turn increases the risk of contamination. The high vibration resistance can now be achieved by avoiding relevant resonant frequencies or by using code disks made of nickel.

Drive systems with motor feedback system require two wiring harnesses. One is used to transmit the electrical power to the motor, the other to connect the motor feedback system (power supply, data exchange). Classically, these two strings are laid as two separate cables between the inverter and the servomotor, which is a high cost factor and places high demands on the cable routing. New developments allow the combination of the power line and the line for the motor feedback system in one cable. The single cable technology based on HiperfaceDSL® is a pioneer in this respect. Here, two shielded wires are inserted into one remaining hybrid power cable (see Fig. 3.120).

The so-called “sensorless” approach is increasingly finding its way from science to industrialisation. The aim here is to remove the motor feedback system from the drive and replace the information for commutation and speed control of the motor with current and voltage measurements, coupled with sophisticated algorithms. Different methodologies are used here. Widely used is the measurement of the electrical back voltage of a rotating motor. With another method, signals are impressed into the motor’s winding and the reaction of the motor is measured. Thus, the motor is also used like a resolver. However, these “sensorless” (or better “encoderless”) approaches have only a limited range of functions compared to servo systems with a motor feedback system and do not even come close to their quality. Often an application requires that the angular position can be measured absolutely not only on one revolution, but over several. Especially when a rotary motion is mechanically converted into a linear motion, this extension is absolutely necessary. For so-called *multi-turn encoders*, the singleturn encoder is combined with an additional technology for multi-turn expansion. A common combination is realized by connecting several scales by means of one or more reduction gear stages, with the first gear stage coupling to the encoder shaft. These scales are usually optical or magnetic (Fig. 3.121).

Gear-based multi-turn encoders are *absolutely coded*, i.e. an absolute position can be formed at any time without the system having to rely on the past. In contrast to this, *absolute counting* systems form the absolute position once when the encoder is initialized and from there continue to operate in rotation-counting mode based on the singleturn information. In this approach, the singleturn encoder is combined with a battery, a reed contact, or a Wiegand wire sensor, to name just a few possible technologies.

Draw-wire encoders are a special variant for encoders. These combine incremental or multi-turn encoders with a draw-wire mechanism. A wire is wound onto a drum. The drum is connected axially to the encoder shaft and to the housing via a retaining spring, which tensions the cable. The free end of the cable is attached to a moving object, possibly also via deflection rollers. This makes it very easy to measure a linear movement. Wire-draw encoders are available for short lengths but also for applications with measuring lengths of several dozen meters.

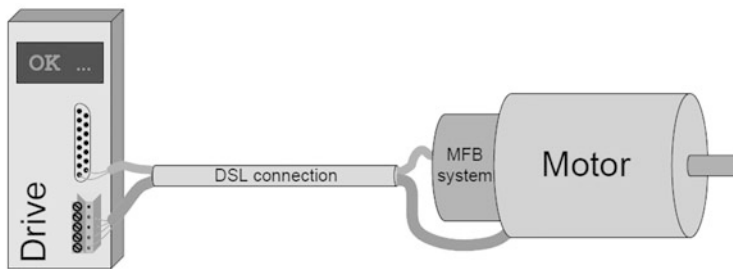


Fig. 3.120 Servo drive system with single-cable technology based on HiperfaceDSL

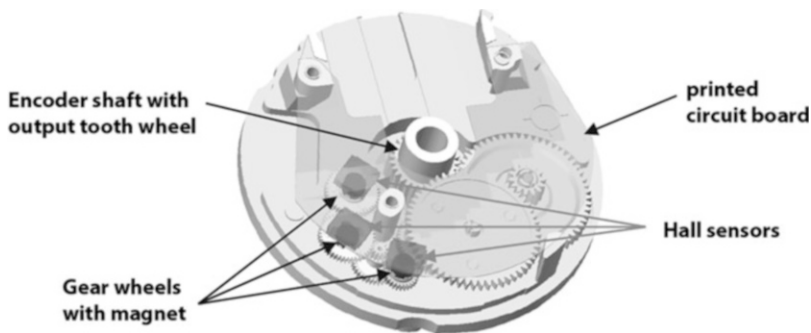


Fig. 3.121 Multi-turn encoder with magnetically levitated gear stages

3.2.1 Optical Encoders

Technically highly sophisticated encoders in terms of *high resolution* and *accuracy* use optical scanning. The majority of rotary encoders use *graticules* with imaging optics; for very high resolutions, diffraction-based systems (diffraction gratings) are also used. Figure 3.122 shows the design of an optical rotary encoder with integral bearing.

3.2.1.1 Physical Principles

Rotary Encoder with Imaging Optics

Most often, the *imaging optics* are used as the scanning principle. Here, a light source is used to project a light-dark pattern 1:1 on a code disk onto a photoelectric receiver (Fig. 3.123).

For simplified evaluation, a receiver element is usually assigned to each track signal of the code disk, which corresponds in shape and size to the track itself. Frequently, additional apertures are also used, which take up the exact size, or are even adapted so that certain error components in the signal (e.g. due to diffraction) are compensated.

Important features of this principle are:

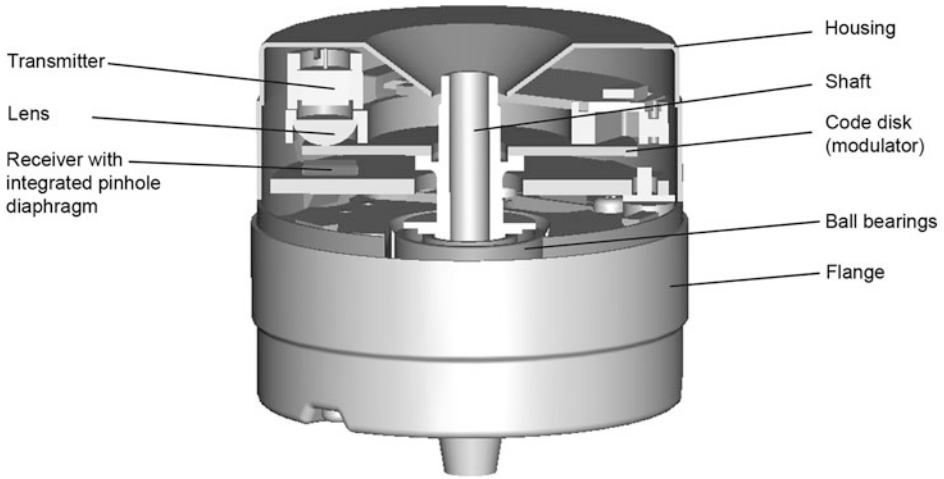


Fig. 3.122 Structure of an optical encoder with integral bearing (Factory photo: SICK STEGMANN GMBH)

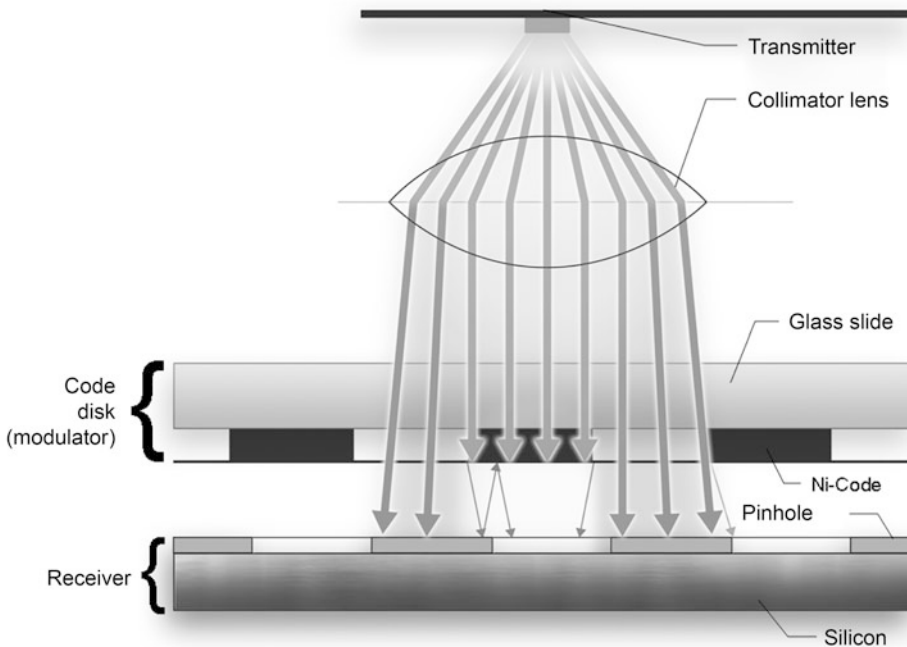


Fig. 3.123 Principle of imaging optics as a scanning principle

- Compared to other optical principles, *high mechanical tolerances* are permitted.
- Comparatively *low resolution*.

- Structure sizes usually not under 10 μm .
- The smaller the structures, the better *boundary conditions* must be met, such as
 - Small distances between code disc and receiver,
 - Low distance tolerances and
 - Parallelism of light.

Furthermore, effects such as diffraction and reflection have a strong influence on the signal quality.

Imaging can also be done on a *neutral sensor* that is not adapted to the optical structures of the code disc (e.g. a pixel-based camera sensor). This method requires *complex algorithms* and *high computing power*, since the code pattern has to be reconstructed from the image information. However, the drop in prices in the field of microprocessors and FPGA (field programmable gate array) allows this principle to be implemented in industrial environments.

Encoder Using the Moiré Effect

The structures described in the previous section can be *optically enlarged* by exploiting the *moiré effect* and thus have a *higher resolution*. If two gratings lying on top of each other with the same grating spacing g are rotated by an angle δ , *moiré stripes* are created (Fig. 3.113). The following applies to the strip spacing d

$$d = \frac{g}{2 \cdot \sin \frac{\delta}{2}} \approx \frac{g}{\delta}$$

If the grids are shifted at constant angles to each other, these stripes move crosswise to the direction of movement. With this principle, it should be noted that even slight changes in the tilting of the grids in relation to each other result in a change in the length of the period. Therefore, the evaluation is accordingly complex (Fig. 3.124).

Encoder with Diffractive Optics

The optical *diffraction principle* is used to achieve *highest resolutions* down to the order of a *few nm*. So-called diffraction gratings with grating constants in the range of a few μm are used as measuring standards. During diffraction at the diffraction grating, the light beam undergoes a deflection (diffraction orders $\neq 0$) and a position-dependent phase shift.

If a beam of coherent light is split into two partial beams and superimposed again after passing through different optical paths, *interference* is created and, if the optical arrangement is suitable, an interference pattern is created (Fig. 3.125). The interference is therefore only used here to measure the phase shift, but the *angle-dependent modulation* results from *diffraction*.

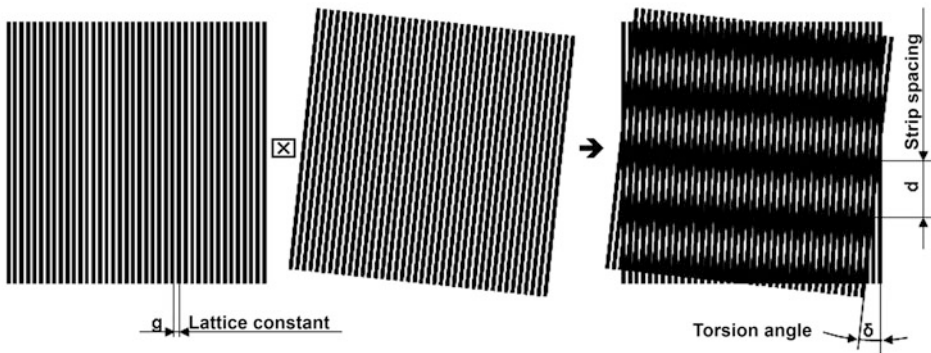


Fig. 3.124 Creation of the moiré effect

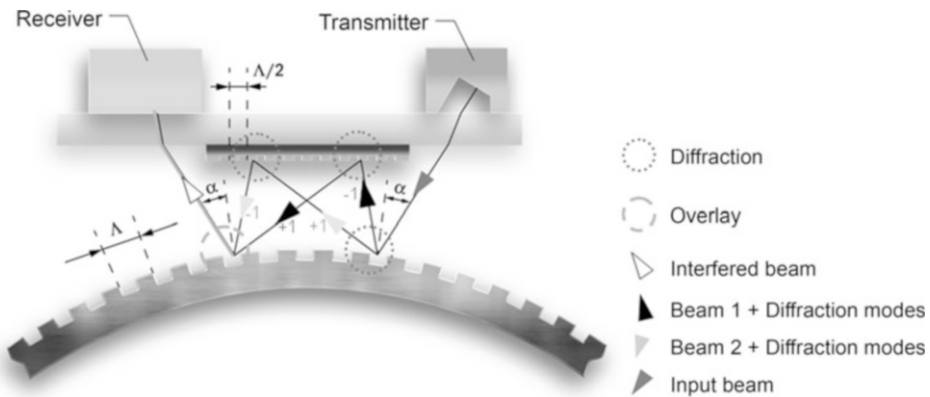


Fig. 3.125 Principle of encoders with diffractive optics

3.2.1.2 Design of Optical Encoders

Order

With optical encoders, the measuring standard, transmitter and receiver can be arranged in both radial and axial directions relative to each other. Usually, there is axial scanning relative to the code disk.

Radially scanned systems in which the code track is applied to a drum-shaped carrier are in principle only known for *reflective systems* in which the transmitter and receiver are grouped on the same side of the measuring scale. With simpler rotary encoders, the scale consists of a tape that is applied to a cylindrical base body. The accuracy is limited by the point of impact on the circumference. High-quality systems are manufactured directly on the base body (e.g. laser ablation or laser exposure).

Some optical encoders are also available *without bearings*. However, high-resolution encoders in particular require high precision, good concentricity of the shaft, adherence to

low tolerances, as well as care and experience during assembly. These systems are mainly used with low resolution encoders or with large hollow shaft diameters.

Light Source

In the field of optical encoders, *GaAs*-based *LEDs* are mostly used. These *LEDs* emit infrared (IR) light (about 880 nm). This is advantageous in several ways:

- Sensors based on Si have the highest spectral sensitivity here.
- The transmitters therefore only need to be operated at low power, which increases their service life.
- The sensor technology designed for IR is only slightly affected by daylight.

In encoders with imaging optics, a collimator lens is usually used to form a parallel light beam so that the code pattern is imaged on the receiver with as few imaging errors as possible (Fig. 3.126). Depending on the area to be illuminated, the design for parallel light requires a comparatively large overall length in the direction of the optical axis. Small structures and compact code forms, such as the *PRC* (pseudo-random code), are therefore especially advantageous for *absolute encoders* (see also Fig. 3.129). The lenses themselves are partly part of the commercially available LED.

In exceptional cases, spotlights are also used if possible. The geometries of the code disc and receiver are then matched to each other according to the beam path. A disadvantage here is the sensitivity to changes in distance between code disc and receiver.

Laser diodes are used as a further light source. This is especially the case where the coherence of the beam is a prerequisite for the principles described above, which are based on diffraction and subsequent interference.

A disadvantage is that the operating current of a laser diode is significantly higher, which is partly compensated by the pulsed operation.

Code Disk

For code discs, there are mainly three basic materials available for the encoder range:

1. Transparent plastic, metal or glass.
2. In the area of the imaging encoders, light and dark fields, i.e. transparent and non-transparent or reflecting and absorbing surfaces must be applied to the carrier.
3. In the case of glass and plastic, a chrome layer is applied to one side, a light-sensitive lacquer is applied over it, the code pattern is exposed and then the chrome layer is structured in a wet chemical process.

Plastic panes are cheaper than glass panes due to their easy processing and handling. However, they are usually only used for low line counts and light operating conditions, as their adequate thermal stability and optical homogeneity are not easy to achieve.

In the case of glass panes, on the other hand, suitable measures must be taken in the design to prevent pane breakage in the event of high shock and/or vibration.

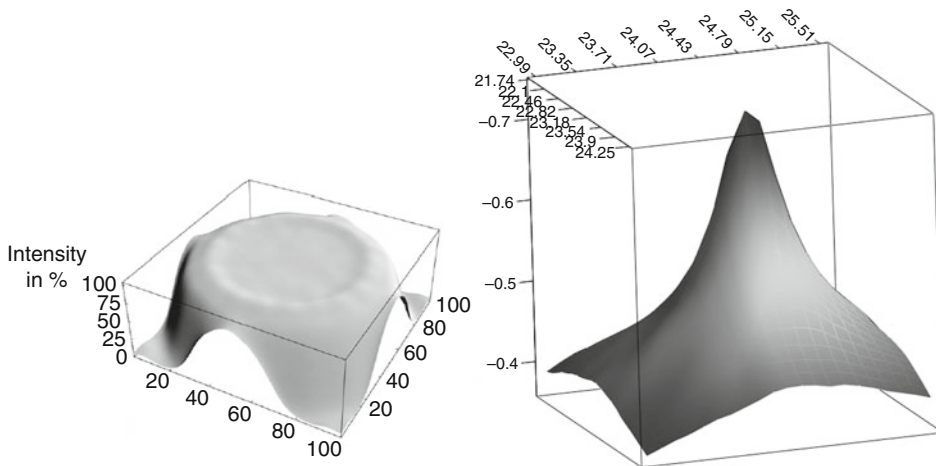


Fig. 3.126 Beam characteristics (relative light intensity) of an LED with and without a collimating lens

With metallic disks, usually galvanically manufactured from nickel (Ni), the functions of measuring standard and carrier are realised in one material, in one process. In contrast to coating, relatively thick layers (up to 100 μm) must be structured here. Until a few years ago, Ni code discs were therefore only used in the low resolution range (up to about 1024 lines/revolution). The further development and optimization of galvanic processes now makes it possible to produce very homogeneous code discs for high-resolution encoders.

Two main processes can be distinguished in galvanic structuring: the *ablative process*, in which the later slots are etched *free* (etching), and the *growing process* (electroforming), in which the bars are deposited in a Ni bath (Fig. 3.127).

In addition to its breaking strength, the possibility of self-centering metal code discs is a striking advantage over glass discs.

In the field of *diffractive optics*, *microstructured discs* made of metal or plastic are used. The structure depths and widths are approximately in the range of the wavelength of the light used. The structures are created either by micromechanical processes, such as micro-diamond milling or laser ablation, or by lithographic processes. Plastic discs can also be produced by hot stamping or micro injection moulding.

Recipient

Simple diode structures with additional *sheet metal apertures* are used for simple incremental encoders. The pinhole diaphragm can be easily and inexpensively adapted to the number of lines/structure size of the code disk (Fig. 3.128).

For incremental encoders with quadrature output, rectangular slots serve as pinholes; sinusoidal encoders use \sin^2 -like shapes. With more complex *opto-ASICs* (Application Specific Integrated Circuit), the aperture usually no longer has to be aligned and bonded as

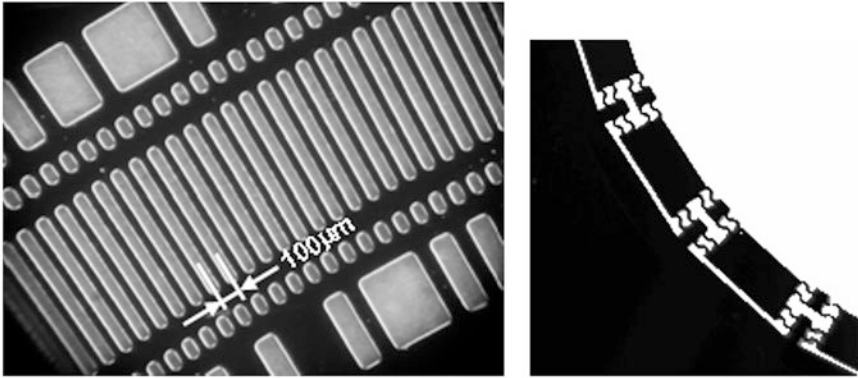


Fig. 3.127 Electroformed spring structure on the inner diameter (galvanically grown) Ni code disk for self-centering (Factory photos: SICK STEGMANN GMBH)

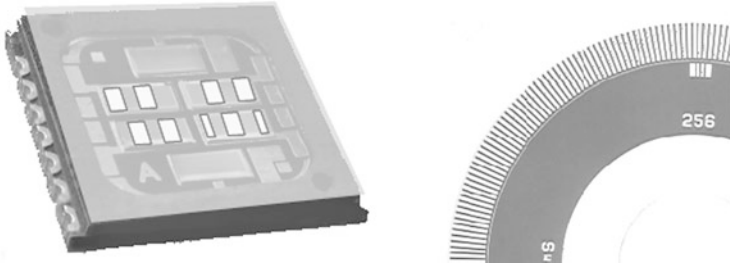


Fig. 3.128 Example of a simple opto-scanner (Source: OPTOLAB) with pinhole aperture and corresponding code disk (Source: SICK STEGMANN GMBH)

a separate process on the receiver, but is implemented directly in the metallization mask (Fig. 3.128).

The ASICs used also include signal adjustment (offset, amplitude), signal processing (AD converters, comparators, arctane calculation), as well as some monitoring and error detection functions for the reliable operation of encoders.

The acquisition of the absolute code pattern is usually realized with only one transmitter. For reliable code reading, however, the illumination must be as uniform as possible. Traditional arrangements of radial tracks of different periodicity (Fig. 3.129, right) have an unfavorable length-to-side ratio. The tangential arrangement of the code pattern as a pseudo-random code (Fig. 3.129, left) is particularly advantageous, since an almost round surface must be illuminated.

Frequently, an additional transmitter current control is integrated in the opto-ASICs, which constantly regulates the incoming light quantity via temperature, aging and contamination and thus stabilizes the operating point of the receiver circuit.

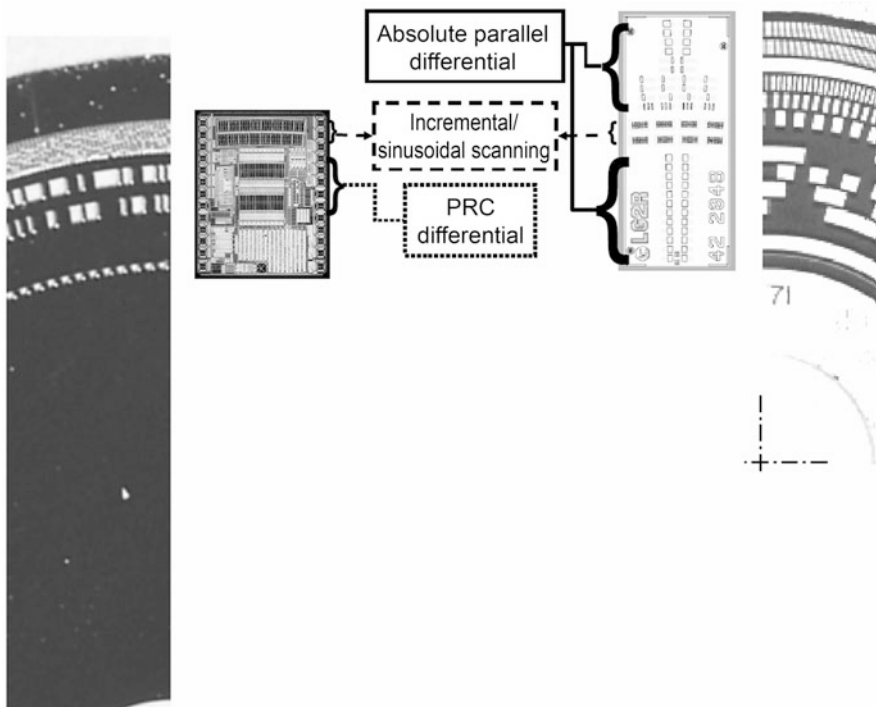


Fig. 3.129 SICK: Code disk and iC-Haus: Scanner with scanner with PRC parallel absolute code

The amount of light is either detected via a separate track or determined from the DC light component or the vector length ($\sin^2 + \cos^2$) of the incremental signals.

3.2.1.3 Special Features of Optical Encoders

Conditions of Use

Due to their high resolution and very short dead times, optical encoders are mainly used for applications of highest precision and performance (e.g. as motor feedback for speed control of servo motors).

Non-linearity

At this point, it is worthwhile to deal with the terms integral or differential non-linearity.

The *integral non-linearity* describes a property that mainly refers to the *accuracy* of a position value. It is usually caused by the mechanical structure (e.g. eccentricity of the code disk).

The *differential non-linearity* is especially relevant for the *derivation of the speed information*, because it describes inaccuracies, which are repeated per scale period. This means that a comparatively high error frequency can enter the controller. The cause of this

non-linearity is usually due to imperfect calibration of signal offsets and amplitudes or to slight variations in the line pattern. With incremental encoders, these nonlinearities are included in the specification quantity “*jitter*”.

General Mechanical Engineering and Automation Technology

In this application area, the technical requirements regarding resolution and accuracy are often met by low-cost systems (e.g. magnetic scanning). In the field of incremental encoders, high-resolution optical rotary encoders are advantageous because their performance is only marginally affected by external conditions (e.g. temperature). This is due to the fact that the high resolution of optical rotary encoders is physically contained in the measuring standard, so that changes in offset and/or amplitude have almost no effect.

Lifetime

Due to the limited durability of the ball bearing grease, the ball bearing will be damaged, resulting in encoder failure. If the expected service life is to be determined more precisely, the exact conditions of use, especially the operating temperature, and also the motion profile, speeds or bearing loads must be known. It is therefore difficult to derive general statements on service life. Often, however, several 10,000 operating hours can be realized.

Only a few years ago, the light source(s) was (were) one of the main causes of failure, especially in absolute encoders. Nowadays this can be almost completely excluded, since together with the ASICs used, usually only one light source is required and this is operated with only low power due to the highly sensitive receivers. Aging and temperature dependence of the LED as well as uniform soiling of the code disc are compensated by the previously shown transmitter control.

3.2.2 Magnetically Encoded Rotary Encoder

In Sect. 3.1.6, the function of magnetically encoded measuring systems for linear movements was described. In this chapter, a rotary measurement is considered in more detail. With this principle, small, compact rotary encoders can be very easily constructed that are also very robust against contamination. Depending on the application, different versions of the round measuring body are available.

Linear motion is a measuring behavior that does not depend on the distance between the measuring body and the sensor head. Figure 3.130 shows the principle. The sensor is precisely matched to the pole width P . The pole width does not change with the air gap, the distance between the sensor and the measuring body. It remains constant.

In contrast to linear movements, various effects occur when measuring rotary movements, especially angles of rotation, which must be taken into account during the measurement. The user must be aware of these effects and select his measurement setup or application accordingly.

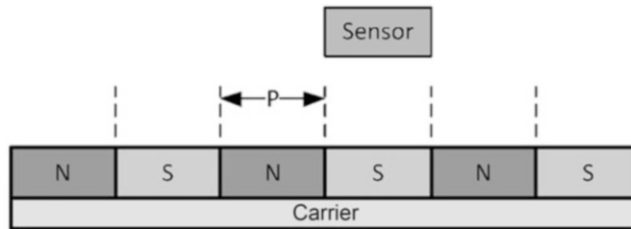


Fig. 3.130 Magnetic fields for linear dimensional bodies

Table 3.22 lists some important properties of round dimensional body concepts. In a magnetized ring, each pole has a certain arc length. This arc length is selected so that, with a defined air gap in the active area of the sensor, the pole length corresponds exactly to the pole length for which the sensor is designed. This is not possible with a linear dimensional element mounted on the circumference. Its pole width only corresponds exactly to the pole width for which the sensor is designed in a linear application or with very large radii. The correlation can be seen in Fig. 3.131: The *pole width P increases* with increasing air gap with a convex application between sensor head and measuring body.

This pole width in the active area of the sensor is referred to in the following as *effective magnetic pole width*.

When mounted on the round surface, the measuring body is deformed around the neutral fibre of the (steel) carrier material. The magnetic poles become larger or smaller depending on the convex or concave bending direction. The magnetic fields behave accordingly. An additional *linearity deviation* occurs within each pole in the sensor, which is adapted to a fixed pole width. This linearity deviation is *greater* the *larger* the *air gap* and the *smaller* the *bending radius*. However, the number of pulses per full revolution is not affected by this; for example, the sensor interprets each pole as 1000 increments. As long as the magnetic fields are correctly detected, the number of increments per revolution is still constant, as the number of poles has not changed.

If the *angle of rotation* of an axis is determined with this application, this is done via the *path* measured on the surface. The angle of rotation can be inferred from the radius. The larger the radius, the higher the achievable accuracy for the rotation angle. Since the path on the surface can be measured reproducibly with an accuracy of a few tens of times μm , the effective radius (from the axis of rotation to the neutral fiber of the substrate material) can only be produced with a reproducible accuracy of a few tens of times μm with great effort. The achievable reproducible accuracy of the radius is an important factor influencing the accuracy of the measured rotation angle.

Other conditions are achieved if the measuring body is first mounted on a curved surface and then magnetized at a precise angle. Then the *pole width* can be *adjusted* within wide limits exactly to the pole width expected by the sensor with the appropriate air gap. With the number of pole pairs on the circumference and the planned air gap, the radius and the *calculated pole width/arc angle* with which the material is magnetized are obtained. It is

Table 3.22 Comparison of magnetized ring and linear dimensional element on round surface

		Realization of the angle measurement	
Principle	Ring magnetized	Linear measuring body on round surface (convex, concave).	
Magnetization	Each pole has a certain arc length (angle).	Each pole has a constant length.	
Most important factors influencing the measuring accuracy	Difference of the Centre point during magnetisation and operation air gap	Effective radius, distance between rotational axis and neutral fibre of the scale carrier material air gap.	
Range of motion	$>360^\circ$	$<360^\circ$ with normal accuracy $>360^\circ$, if accuracy requirements can be dispensed with	
Mechanical stability	Well	Ends of the measuring body should be mechanically fixed.	

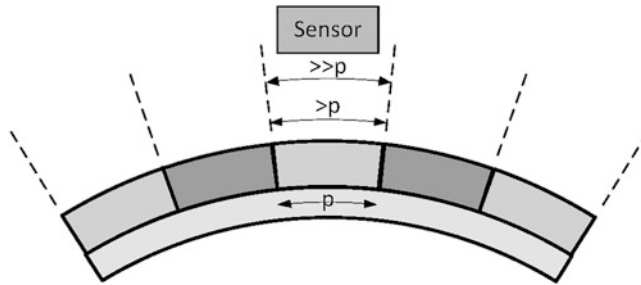


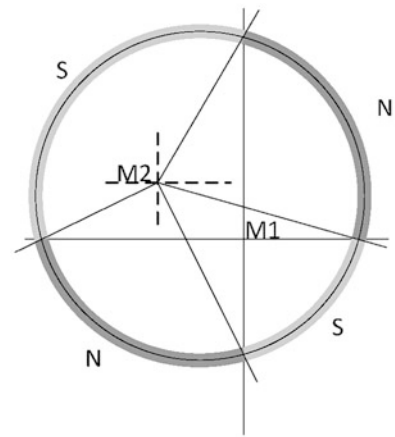
Fig. 3.131 Magnetic fields with radially bent measuring body

selected in such a way that the expected pole width is exactly adjusted at the sensor with a defined air gap. Depending on the convex or concave curvature, the physical pole width at the surface is greater or smaller than this. This dimensioning results in a minimum linearity deviation.

Another important factor influencing the accuracy of a magnetized ring is the difference between the axis around which the ring is magnetized and the axis on which it is operated. This effect occurs particularly with rings that have large tolerances, for example of the inner diameter (e.g. inexpensive ceramic rings). This effect need not be considered for rings with small tolerances (e.g. fits). Figure 3.132 shows in a very simplified and greatly exaggerated way the effect of different axes during magnetization and measurement.

A magnetic ring is magnetized around the axis M1 with two pairs of poles. Each pole has a radian measure of 90° when magnetized. If the ring rotates around the M2 axis during assembly, the measured radii are no longer 90° . The sensor interprets each pole with 1000 increments, for example. Compared to ideal conditions, the increments are too large in the area of the south pole top left, and too small in the area of the south pole bottom right. A

Fig. 3.132 Magnetic ring with different magnetization and operating axes



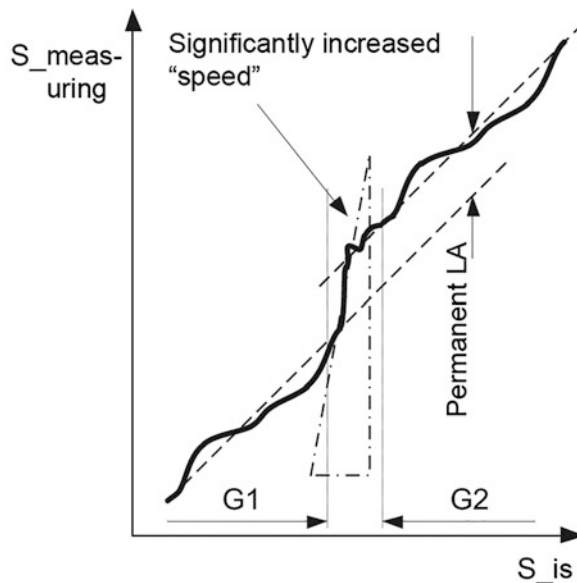
measurement error occurs again. However, the number of increments per revolution is correct.

If a linear dimensional scale is mounted over 360° on a circumference, the effects at the joint of the two ends of the scale must also be considered. In principle, the circumference of the ring must be selected so that it corresponds to a multiple of the pole pairs. Otherwise the pole would have a different width over the joint. For example, it would only be 4-mm wide instead of 5 mm. In the example, the sensor interprets each pole as a length of 5 mm and uses interpolation to make 1000 increments. In the area of the joint, increments are now output not on 5 μm but on a length of 4 μm .

Fortunately, this effect has no influence on the total number of increments per revolution, it only depends on the number of pole pairs. It does not matter if one pole has a different width. However, there is an influence on the measured *travel speed*. This is shown in Fig. 3.133. In the area of the joint, between the ends of the measuring element G1 and G2, there is a deviation (permanent LA) in the measured path in the above example. S_{is} is 4 mm, S_{measured} is 5 mm. The dashed curves correspond to the ideal relationship of the measurement (measured path over the actual physical path). The solid curve represents the measured displacement with the linear deviation. The slope of the curve corresponds to the *measured travel speed*. In the range of G1 and G2, it is constant, in the range between G1 and G2, it appears increased. This is represented by the dot-dashed triangle. (1000 increments are output here over 4 mm.)

The maximum travel speed of the measuring system is defined by the distance per increment and the minimum edge distance (minimum time per increment) (see Sect. 3.2.7). At speeds higher than this, pulses may be lost. According to the above explanations, there is a particular danger in the area of the joint. In this case, the increments of a whole pole are missing (e.g. 1000). A simple test is to mark a position on G1 and to pass the area of the joint quickly in forward direction and then slowly in reverse direction. If the measured position is then identical at the same position, the conditions in the application are sufficiently good for this effect.

Fig. 3.133 Relationships at the point of impact for a linear dimensional element mounted over 360°



A further effect for the accuracy in the area of the joint is the deformation of the magnetic material when the measuring body is cut.

If all these effects can be tolerated, there is no reason why a measurement over 360° should not be carried out with a linear measuring body glued to a ring. Then all that remains to be done is to ensure that the measuring element is well fixed, particularly in the area of the joint.

3.2.2.1 Applications

In the following, some examples of the magnetic measuring system are described. Figure 3.134 shows a linear and rotary application. In the linear movement, limit switch magnets are additionally attached.

Figure 3.135 shows a round measuring body which allows a large axial offset during measurement. The magnetic poles with a pole width of 5 mm have a lateral extension of 30 mm. The poles are made visible with a pole pitch display card. It darkens at the poles and becomes light between the poles. An application with this measuring body is in the printing industry, where the speed of an axis is measured accurately. The axis can move a few cm in axial direction without affecting the measurement.

Figure 3.136 shows a hub with the dimensional element on the circumference. This hub is used in handling applications where a large number of media must be fed through hoses into the moving area. The diameter was selected accordingly large with regard to the required accuracy (see size comparison with credit card). The measuring body is located in the area of the arrow.



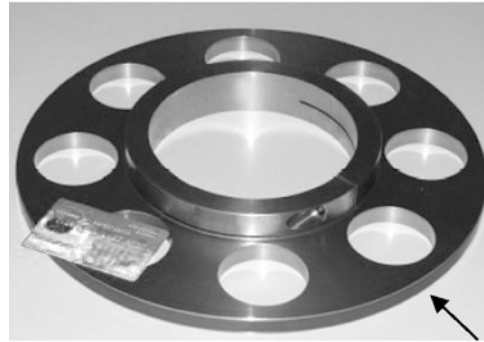
Fig. 3.134 Image of a sensor head with linear and round measuring body (Factory photo: Balluff GmbH)

Fig. 3.135 Image of a dimensional scale with a very wide pole pitch, made visible by a pole pitch display card (Factory photo: Balluff GmbH)



Figure 3.137 shows a stainless steel hose clamp in which a measuring element is located, hidden by the hose clamp. Its magnetic fields “look” through the hose clamp to the outside. The hose clamp can be very easily attached to a circumference. A sensor head can then be used to detect the *rotational movement of an axis* very easily and accurately. With the design shown, only movements $<360^\circ$ can be measured. If the mounting is on the side of the measuring body, mechanical movement ranges $>360^\circ$ can also be realized. As already mentioned, however, the point of impact of the scale body is problematic from a

Fig. 3.136 Image of a dimensional body (arrow) on a large hub (Factory photo: Balluff GmbH)



metrological point of view. For movements that are measured $>360^\circ$, a solution with a closed magnetic ring is recommended.

3.2.3 Rotation-Counting Angle Sensors

3.2.3.1 General Principle of Operation and Morphological Description of Rotation-Counting Angle Sensors

A frequently occurring measuring task in industrial systems is the measurement of several revolutions with the following additional requirements:

1. The information of the rotation must not be lost in case of a failure of the supply voltage.
2. If, in the de-energized state, the measured object is rotated even over more than one revolution, this should be detectable.

A sensor system meeting these requirements is often called a *true power-on sensor*, i.e. it always shows the true position (True) when switched on (Power On).

There are many different solutions for revolution-counting systems that do not or only partially meet these requirements. These all apply the following principle: A sensor measures only *one revolution* (or parts of it). At each transition of the measuring element, it counts this transition. This is an *incremental method of operation*. However, this information is lost when voltage is lost.

In order to solve this task with a *True Power-On sensor* (TPO sensor), there are few implementation possibilities specifically for revolution counting, which can usually be used together with a separate *singleturn sensor system* to measure angles with high resolution even within one revolution.

The following implementations are known:

1. *Mechanically* by using a gear or other mechanical means of transportation.
2. *Inductive*: Detection is combined with energy generation.
3. *Incremental* detection by battery buffering for rotation information.

Fig. 3.137 Measuring element in a hose clamp (covered by hose clamp) mounted on a shaft (Image: Balluff GmbH)



4. Measurement and storage of the rotation information with a novel *GMR system* (Table 3.23).

3.2.3.2 Gearbox-Based Revolution-Counting Methods

A very simple solution is a gear unit that maps the desired angular range $n \cdot 360$ degrees to 360 degrees. Especially for industrial applications, sometimes considerably more than 10 revolutions are required. For these applications today, *digital multi-turn* encoders based on an encoder with a multi-stage gear unit are mainly used, which usually have an angular resolution of 12 bits to 18 bits and can detect up to 16,384 revolutions (14 bits). Figure 3.138 shows such a gear unit. Depending on the requirements, different stages of the gear unit are detected in position.

If one only wanted to get by with angle detection, it is disadvantageous with gearboxes that the resolution is reduced accordingly and, in addition, hysteresis effects occur across the gearbox stages.

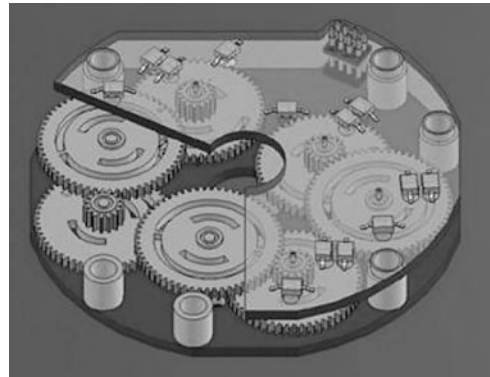
This disadvantage is eliminated simply and cost-effectively with a *multi-turn potentiometer* (Fig. 3.139). Here, the multi-turn potentiometer is created by the fact that the measuring standard in the form of a *helical resistance element* represents several revolutions. The sliding contact is usually guided by the resistance element itself and picks up a position-dependent voltage. Similar versions of the same principle convert the rotary motion via a *spindle* into a translational motion, which is then detected potentiometrically, or the position is detected magnetically via a moving magnet and a fixed Hall sensor (Sect. 2.8).

In automotive applications, such as the detection of the steering angle, significantly fewer revolutions are required. The number of revolutions is less than 8 for passenger cars

Table 3.23 Morphological modular system of TPO revolution counters (diamond line: implementation path of the GMR multi-turn angle sensor from Novotechnik)

Characteristics, function	Known designs, realizations			
1. Geometric measured variable	Angle ($> 360^\circ$ or $n \cdot 360^\circ$)			
2. Transmission principle	Mechanical: Gearbox/carrier	Magnetic field	Optical	Inductive
3. Detection (referring to rotation information)	Detection of position of stepped toothes wheel/secondary toothed wheel Conversion of angle information into path information	e.g. with reed sensors	GMR structure Acquisition: Domain shift - \rightarrow Change of resistances of a sensor structure Distribution: Resistance measurement	Generation of a voltage pulse using: a) Coil-wrapped Wiegand wire b) Coil surrounded leaf spring c) Coil-surrounded folding magnet
4. Storage of the rotation information	Position of the gear elements	Dynamic memory, buffered by battery	GMR structure	FRAM

Fig. 3.138 Multi-stage transmission for recording revolutions (Factory photo: Posital)



and trucks. Figure 3.140 shows a solution that uses the classic *vernier principle*. For example, the positions of the gear wheels m and $m + 1$ are measured with 2 AMR angle sensors (AMR: Anisotropic Magneto Resistance; Sect. 2.3). An evaluation circuit calculates the corresponding angle value.

Another subtype for a transmission solution is the *Maltese cross transmission* (Fig. 3.141). A crank mounted on the shaft to be measured generates a periodic rotational movement for a cross connected to it, interrupted by detent positions. The Maltese cross

Fig. 3.139 Multi-turn potentiometer

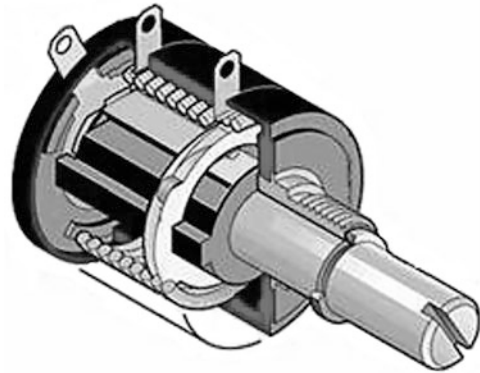
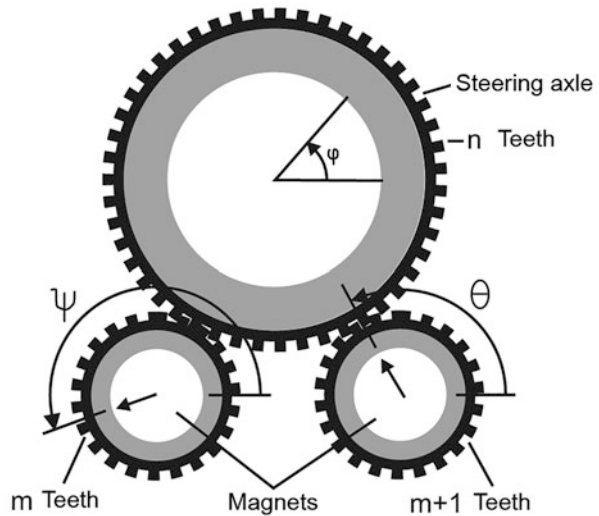


Fig. 3.140 Multi-turn angle sensor (Factory photo: Bosch)



can now be scanned using very simple means such as Hall switches (Sect. 2.8). With this variant, applications such as steering angle sensors are also possible.

3.2.3.3 Rotation-Counting Method on Inductive Basis

Revolution counters can also operate on an inductive basis (Sect. 2.5): The rotary motion of the shaft generates electrical energy as a generator to supply a counting electronics, which records the number of revolutions. However, this form of energy generation fails at a speed close to zero, because the voltage generated is not sufficient to supply semiconductor circuits. A solution approach to provide sufficient energy even at slow rotational movements is the use of *Wiegand sensors*.

The core of a Wiegand sensor is a *Wiegand wire*. This wire consists of a ferromagnetic alloy, which is given an outer magnetically hard zone (*shell*) and an inner magnetically soft

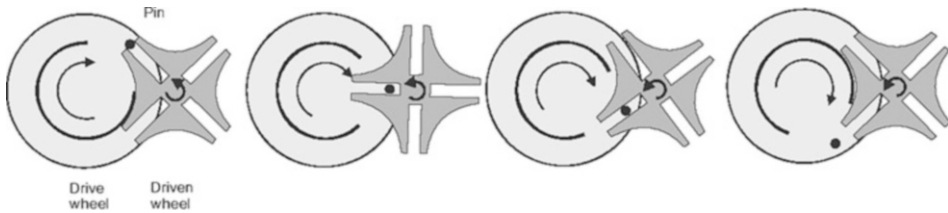


Fig. 3.141 Maltese cross gearbox (Source: Wikipedia)

zone (*core*) by a special post-treatment. With a strong external magnetic field in the direction of the wire, the shell and core can be magnetised in the same direction. The magnetization direction of the magnetically soft core is reversed if the wire is magnetized relatively weakly in the opposite direction. The magnetization direction of the shell remains unaffected. When the magnetic field strength is increased, the magnetic polarity of the shell also reverses. In a coil wound on this wire, *reversing* the direction of magnetisation of the individual areas of the Wiegand wire induces a *voltage pulse*. Voltage level and energy of this pulse are thus not influenced by the rate of change of the externally acting magnetic field.

If the magnetic field flowing through the Wiegand sensor is generated by a *permanent magnet*, which is firmly connected to the shaft of an encoder, the *rotation of* the shaft causes a *reversal of* the magnetization direction and thus of the voltage pulses. These are used to supply a counting electronics with energy for a short time. The core of the counting electronics is a non-volatile memory in which the number of rotations is stored. The counting electronics evaluates the *direction of* rotation and increments or decrements the number of revolutions stored in the memory according to the direction of rotation.

The total energy available for the counting and storage process is in the order of 100 nJ. Therefore, ideas for using the Wiegand sensor as part of the absolute position detection of encoders only became feasible with the development of corresponding *energy-efficient non-volatile memories*. Further requirements for these are a high number of write cycles and a long data retention time.

Often magnetic singleturn technologies are used in encoders together with the Wiegand sensor-based multi-turn part. If correctly designed (Fig. 3.142), a *single excitation magnet* can be used to provide both the angle information for the singleturn part and the magnetic field for the operation of the Wiegand sensor.

The *advantages of* Wiegand technology are the purely *electronic counting of* revolutions. Apart from the rotating magnet, no mechanical moving parts are required. Therefore, Wiegand sensor-based encoders are often used in areas with *high mechanical stress*. In contrast to encoders that count revolutions by means of a mechanical gear, Wiegand sensor-based encoders can hermetically encapsulate all electronic components, thus achieving very high resistance to environmental influences such as dust and moisture. Furthermore, the number of detectable revolutions is only limited by the available memory

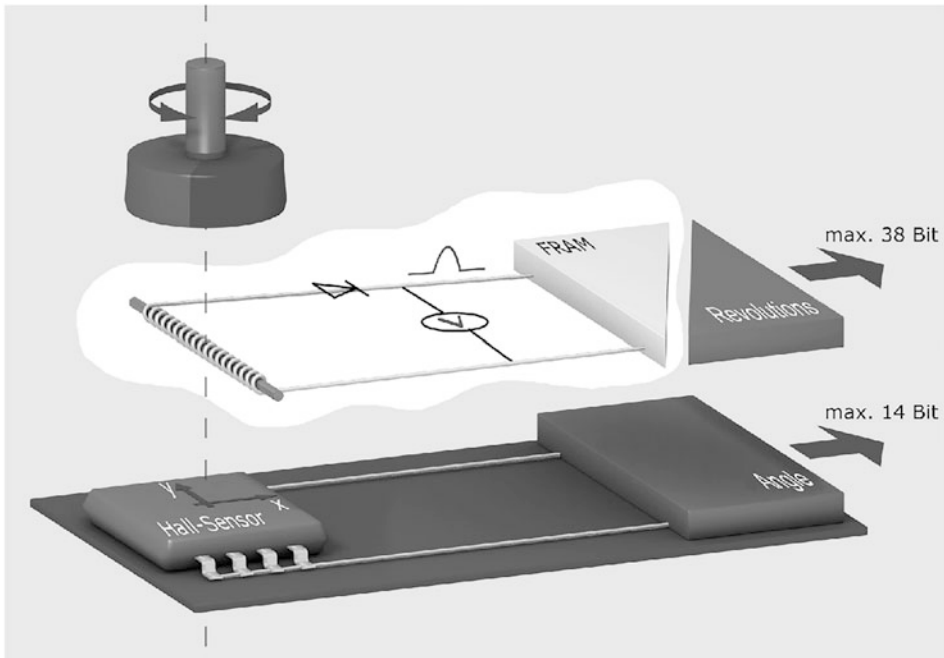


Fig. 3.142 Structure of a sensor system with Wiegand rotation detection (Factory photo: Posital)

and thus practically unlimited for practical use. Based on the inductive revolution detection or simultaneous energy generation, the following further designs are known:

- A *magnetizable leaf spring*, actuated by a rotating magnet, performs spontaneous movements.
- A *movably mounted magnet*, also actuated by a rotating magnet, performs spontaneous movements (folding magnet).

In both of these cases, *dynamic magnetic fields* are generated whose dynamics, similar to the Wiegand principle, do not directly depend on the rotational speed of the drive shaft and can be tapped inductively.

3.2.3.4 Battery Buffering of the Rotation Information

In order to operate a *true power-on multi-turn system*, it is sufficient in principle for a multi-turn sensor system to be able to detect revolutions and store them; in the best case without external supply voltage. A very simple approach is to *supply* and *buffer* a multi-turn sensor system with *battery power*. For a technically meaningful implementation of this idea, the most important requirement for this system is that it is designed to be extremely energy-efficient. Today's designs work, for example, with reed switches, a long-life lithium battery and very low-current dynamic memory components. The operation of the multi-turn sensor system is then guaranteed over an acceptable long period of time in many applications.

3.2.3.5 Novel GMR System for Detection and Storage of Rotation Information

In the following, a completely new method shall be presented, which allows to detect revolutions even in *de-energized state* and *without gear*. Combined with a corresponding single-turn sensor (e.g. Hall), a gearless, battery-free multi-turn encoder can be constructed.

The basis for this development is *magnetoelectronics*, which exploits not only the charge of the electron but also the magnetic moment (coupled to the spin). Today, such devices, which exploit magnetoresistance effects, are described under the collective term XMR (Sect. 2.3). For this purpose, a multi-turn sensor has been developed that uses the GMR effect. In addition to the rotation angle signal, this can currently count up to 16 revolutions in the currentless state without a buffer battery and without a gear unit and store them permanently. Thanks to the magnetic principle, it works contactlessly and thus practically wear-free. It provides *absolute position values* and makes the measured value available as a true “true power-on” system immediately after startup. In principle, the multi-turn sensor functions like a shift register.

Sensor Element and Measuring Principle of the GMR Revolution Counter

The GMR effect (Giant Magneto Resistance, Sect. 2.3, Fig. 2.15) is a quantum mechanical phenomenon observed in *thin film structures* of *ferromagnetic* (FM) and *non-ferromagnetic* (NM) layers: If one has such a heterogeneous structure of two magnetic layers (a soft magnetic sensor layer and as reference layer a hard magnetic layer), which are separated by a non-magnetic layer only a few atomic layers thick, the resistance of the stack can be changed by external magnetic fields.

The reference layer orientation is retained as long as the magnetic fields do not become extremely strong. The layer stack itself (Fig. 3.143) is structured in strips with a width of less than 200 nm. Due to the resulting shape anisotropy, the sensor layer can only orient itself either parallel or antiparallel to the strip. The electrical resistance changes significantly when the magnetic moments in this sensor layer *fold over*. If they are parallel to each other, the resistance decreases to the minimum value, if they are antiparallel, it reaches its maximum (Fig. 3.144). The magnetization state of such a structure can be easily determined by measuring the resistance.

In general, the sensor consists of two structures (Fig. 3.145):

1. A so-called *domain wall generator*: Due to the large geometric extension (no shape anisotropy), the magnetization in this area can easily follow the external magnetic field.
2. A certain number of *spiral arms*. These represent a very thin structure, where much higher field strengths are necessary for remagnetization (due to the high form anisotropy). The number of spiral arms determines the maximum number of revolutions to be detected.

By rotation of an external magnetic field with a suitable strength, 180° domains are created in the *domain wall generator* and injected into the spiral structure or deleted again

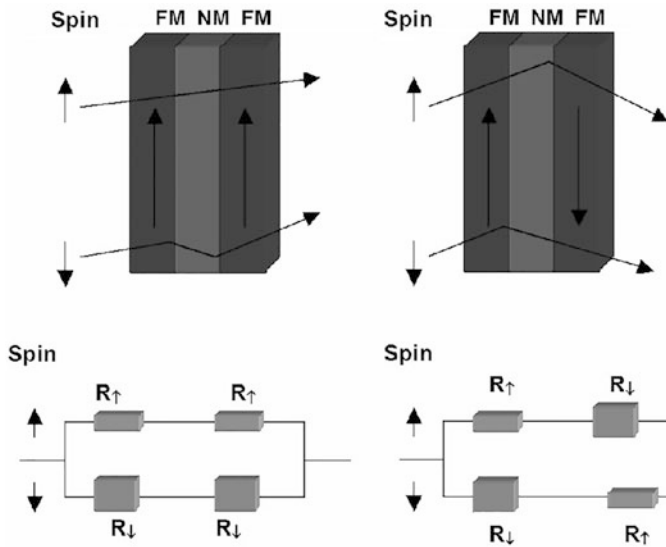


Fig. 3.143 Scattering behavior and equivalent circuit diagram with parallel and antiparallel alignment of the moving layer

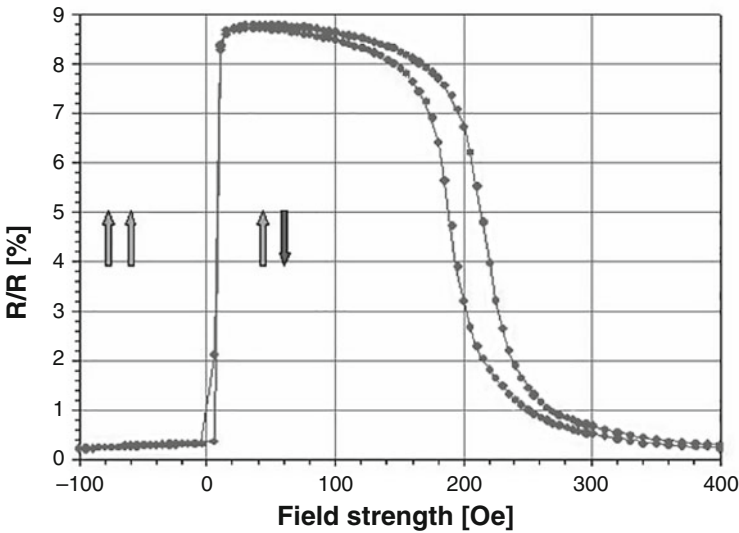


Fig. 3.144 Relative resistance of the GMR system as a function of the applied magnetic field strength

during reverse rotation. The magnetization of the sensor layer in the spiral arms is aligned either parallel or antiparallel to the reference layer (Fig. 3.146).

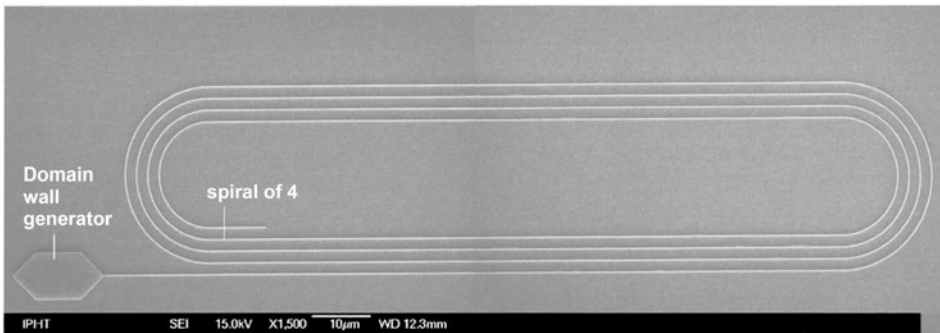


Fig. 3.145 Structure of a GMR revolution counter (Factory photo: Novotechnik)

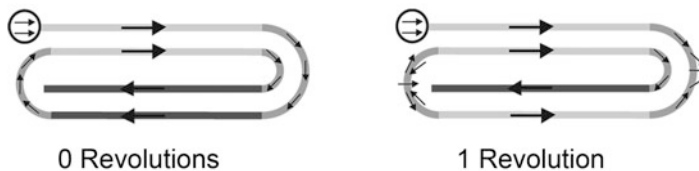


Fig. 3.146 State of the GMR revolution counter in the initial state (no domains) and after one revolution (2 injected domes)

By measuring the resistance of the complete structure or of individual straight sections of the structure, the current rotation can be deduced (Fig. 3.147).

With the described sensor element, it is therefore possible to count and store revolutions. However, it is not yet possible to determine the absolute angle over several revolutions, as the resistance or voltage changes of the GMR sensor element do not assume any clear values at the jump points. This problem can be solved with two *multi-turn elements offset by 90°*.

However, it is even more elegant to “merge” these two structures twisted by 90° into a *rhombus*. In this case, an appropriate evaluation algorithm can be used to derive an unambiguous position for more than 360° in any angular position (Fig. 3.148).

By combining a sensor based on this principle with another 360° sensor that measures the angle of rotation in the singleturn range (e.g. Hall, AMR, inductive, poten-ziometric), it is now possible to determine the angle and number of revolutions over several revolutions. Figure 3.149 shows the sensor view in the microscope. The bonding surfaces of the contacting are considerably larger than the structure.

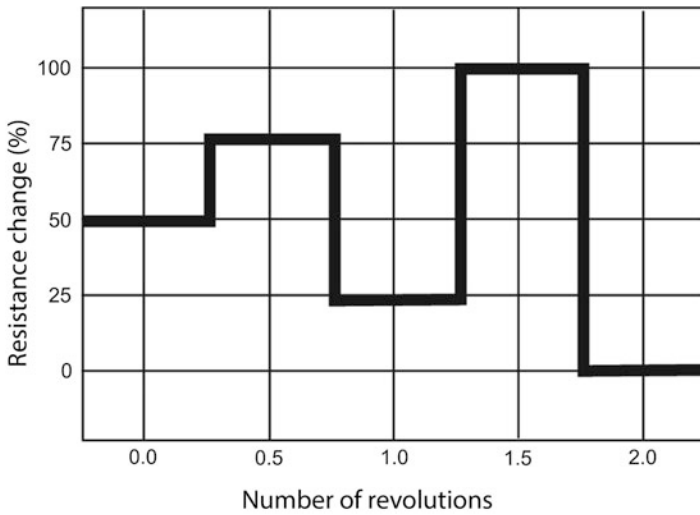


Fig. 3.147 Relative change of resistance with changes of rotation when measuring the resistance of the complete structure

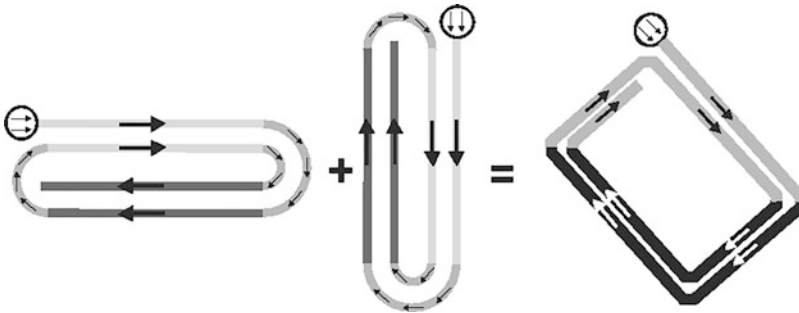


Fig. 3.148 Fusion of two spirals offset by 90° to form a diamond-shaped structure

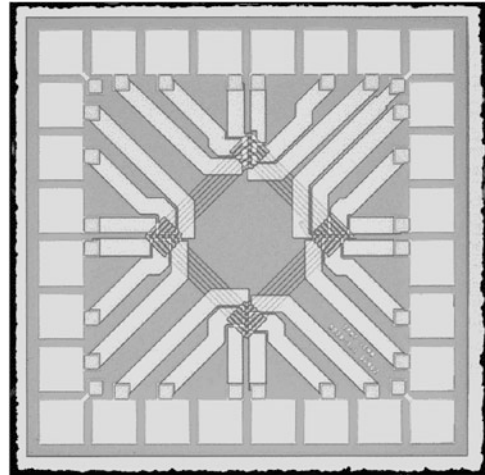
Applications and Characteristics of the GMR Revolution Counter

Compact Solution for Many Applications

Due to the magnetic principle, the revolution counter works contactlessly and therefore wear-free. It provides *absolute position values* and makes the measured value available as a real “true-power-on” system immediately after the start.

Combined with a contactless 360° angle sensor, the *current steering angle can be* directly recorded over several revolutions in automotive or mobile machines, for example. What’s more, the solution saves space and is also cost-effective. The technology is so robust that it can withstand even the toughest conditions in Formula 3.1.

Fig. 3.149 Microscope view of the sensor structure



Similar advantages are also found in other applications, for example in drives for commercial rolling doors. In order to record individual *opening positions* (memory function), the multi-turn takes over the position detection without the need for additional components for a mechanical transmission.

The sensor can be used with a wide variety of linear or rotary *actuators* to detect the position of the actuator spindle over several revolutions. In many industrial areas, this allows compact and cost-effective drives for fittings, flaps or valves to be implemented.

New possibilities are also opening up for *conveyor technology*: If the sensor shaft is combined with the winding drum of a rope length encoder, extremely compact sensor systems for position detection even for long lengths are possible.

Characteristics of the GMR Revolution Counter Using the Example of the Version in the RSM2800 Series

Figure 3.150 shows a GMR revolution counter of the RSM2800 series. The measuring range is adjustable and covers a range from 1 to 16 revolutions. With a linearity error of less than $\pm 0.03\%$ and a resolution better than 0.05° , the sensor with a good price/performance ratio is recommended for a wide range of measuring and positioning tasks.

The repeat accuracy is better than $\pm 0.2^\circ$. The RSM multi-turn angle sensor is characterized by high cut-off frequencies, i.e. it can also be used for applications with high dynamic requirements.

The angle sensor has a very compact mechanical design and is completely encapsulated, so that protection class IP67 with regard to water and dust tightness (6: dust-tight, 7: temporary immersion for 1 h and 1-m depth) can be achieved without any problems. This makes the angle sensor ideally suited for *harsh environmental conditions* in industrial or mobile applications. EMC immunity, temperature stability, cost-effectiveness, reliability, robustness and tightness are further advantages of this sensor.

Table 3.24 summarizes the most important technical data of the RSM2800 angle sensor.

Fig. 3.150 Multi-turn angle sensor RSM 2800 series
(Factory photo: Novotechnik)



Table 3.24 Technical data of the GMR angle sensor, version RSM2800 from Novotechnik

Features	Values
Measuring range	>1–16 revolutions
Linearity error	Not more than 0.03
Repeatability (uni- or bidirectional approximations)	$\leq 0.1^\circ$
Update rate at sensor output	1 kHz
Measuring speed	Up to $1000^\circ/\text{s}$
Protection class	Up to IP67
Operating temperature range	-40°C to $+85^\circ\text{C}$

3.2.4 Capacitive Encoders

Capacitive sensor technology is also suitable for detecting angles and rotary motion. Some of the advantages are: *compact design*, *long life* and *insensitivity* to mechanical tolerances as well as to dirt and magnetic fields.

A capacitor consists of two electrodes mounted in parallel, between which an electric field is generated. Based on this, there are two configurations that are relevant for the design of an encoder. On the one hand, the *electrode structures* can be *moved* or rotated relative to each other. This varies the *effective area* A between the electrodes, which leads to a change in capacitance. Secondly, by inserting a *shaped dielectric*, which moves or rotates between two stationary electrodes. This changes the effective permittivity number ϵ_r between the electrodes, which in turn leads to a change in capacity. These approaches are shown schematically in Fig. 3.151 together with the formula for the capacitance. Here it is ϵ_0 : electric field constant ($\epsilon_0 = 8.854187817 \cdot 10^{-12}$ (As)/(Vm)); d : plate spacing; A : effective area between the plates.

Figure 3.152 shows how these principles can be implemented in an encoder.

Figure 3.152 shows a 3-plate configuration on the left. Here, the *three basic elements* of the rotary encoder, i.e. transmitter, modulator and receiver, are distinct as *individual components*. The *transmitter* is a *stationary circuit board* with a circular conductive

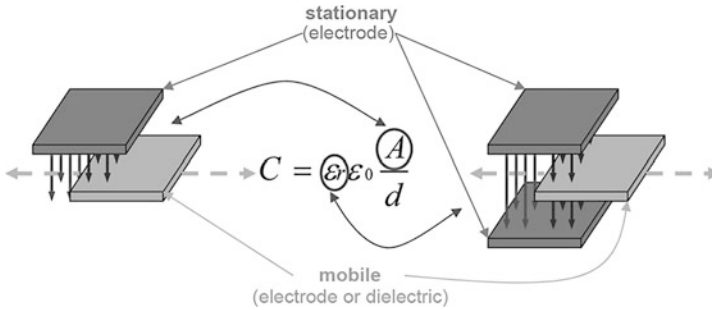


Fig. 3.151 Principles of capacitive sensors—variable electrode area (left) or variable permittivity (right)

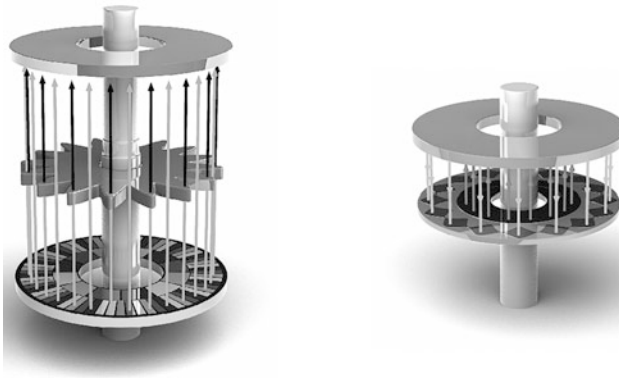


Fig. 3.152 Design of capacitive encoders—3 plates (left) and 2 plates (right) (Source: SICK STEGMANN GmbH)

plate structure. The conductive structure is segmented in the radial direction to form so-called lamellae (Fig. 3.153 left). The lamellae are excited by high-frequency signals to obtain *signal modulation* (time domain). One in four of the lamellae is electrically connected to each other to obtain a multi-electrode structure (Fig. 3.153 left). The receiver is also based on a circular, conductive, stationary plate (Fig. 3.153 right). Like the transmitter, the receiver is also manufactured as a *standard circuit board*. The third plate is the rotor (Fig. 3.153, center). This is made of *dielectric material* and *modulates (spatially)* the electric field between the transmitter and receiver, depending on the angular position of the axis on which it is mounted. This rotor is made of plastic and is sinusoidal when sinusoidal output signals from the sensor are required. The stationary plates are held at a defined distance by mechanical elements, for example a metallic spacer ring.

Four vanes each cover the area for one sine wave on the rotor (see signal processing below). The number of sinusoidal shapes—on the rotor defines the number of sine/cosine



Fig. 3.153 Electrode and dielectric structure of a capacitive 3-plate encoder—schematic (top) and with real components (bottom); (Factory photo: SICK STEGMANN GmbH)

periods per electrically conductive materialve encoder. Figures 3.152 (left) and 3.153 show a configuration with 16 periods.

The 2-plate arrangement shown on the right in Fig. 3.152 represents a *measuring capacitor with variable electrode area*. The structures for the transmitter and receiver are applied to the *stator board*. The rotor carries the sinusoidally shaped copper structure for the field modulation, which is opposite the transmitter, and a circular coupling track, which is electrically connected to the modulation structure and provides capacitive feedback to the receiver structure. This arrangement avoids, in contrast to the principle illustration in Fig. 3.151, that electronic components have to be connected on the rotor, which would then have to be connected to the stator via an inductive coupling or sliding contacts. A disadvantage of this configuration is the high amplitude sensitivity in relation to axial movements of the rotor.

Common to both arrangements is the *holistic (holistic) scanning*, i.e. the capacitive scanning is distributed over the *entire circular surface* and is therefore not limited to one spatial point, as for example in most optical scanning arrangements. This design makes the sensor very insensitive to mechanical changes.

The signal processing is identical for both arrangements (Fig. 3.154). An *excitation generator* generates four high-frequency voltages. These square-wave voltages are 90° electrically out of phase with each other. The excitation frequency is significantly higher than the maximum speed of the corresponding encoder, but still low enough to avoid typical high-frequency problems. The excitation signals are applied to the multi-electrode transmitter, whereas the receiver receives the induced electric charges of the modulated electric field. A *charge amplifier* converts the charges into a voltage, which is amplified by a downstream amplifier and converted into a differential signal. The synchronous

demodulator demodulates (time domain) the differential signal, whereby the synchronization signals are derived from the excitation generator. Two demodulators can be used to generate one sine and one cosine signal each. The subsequent low-pass filters suppress interference in the demodulation and signal noise. However, these low-pass filters have a propagation delay behavior, which is described by the group delay. This so-called *latency* is expressed by the fact that the mechanical angle is displayed with a time delay at the electrical interface. The lower the cut-off frequency of the low-pass filter, the greater the latency. Correspondingly, broadband filters are used, but this means that the previously mentioned interference is less suppressed. Depending on the application, the latency must be taken into account or possibly compensated. More recent developments aim to reduce the latency without having to forego interference and noise suppression. Figure 3.154 shows how the signals of capacitive encoders are processed.

So far, only single track systems have been described. If *several evaluation tracks* are integrated and evaluated, it is possible to implement an *absolute value encoder*. This can already be seen in Figs. 3.152 and 3.153.

If the capacitance values in the pF or fF range are low, the measuring cores must be protected against external interference. This is done by protecting the measuring core with a Faraday cage. For this purpose, a closed sheath of conductively connected elements is placed around the measuring core. Since this cage cannot be closed 100%—after all, the rotor must be mechanically connected to the axis of rotation—it must be ensured that sources of electrical interference have a sufficiently high distance to the open points (Fig. 3.155). A further protective mechanism is provided by the *bandpass behaviour* of the synchronous demodulator. Thus, only interference in a narrowly defined frequency band and sufficiently large interference amplitude can influence the measurement result.

Since capacitive encoders are not very sensitive to mechanical tolerances, they can generally be installed *without integral bearings*. This significantly reduces the mass moment of inertia and significantly increases the encoder's service life, since the susceptible component ball bearing is not present.

3.2.5 Variable Transformers, Resolvers

Variable transformers (VT) can be considered either as a family within the large class of inductive sensors (Sect. 3.1) or as a special group of electrical machines. The variable transformers are characterized by special advantages in the measurement of rotation and angle and are still today a tough competitor to the more modern systems, such as magnetically coded or optical rotary encoders or potentiometric angle encoders.

3.2.5.1 General Operating Principle of the VT

The variable transformer concept basically describes a classic transformer in which the *relative position* between the *primary* and *secondary* windings is *variable* (rotation and/or translational displacement). The relative movement of one winding to the other results in a

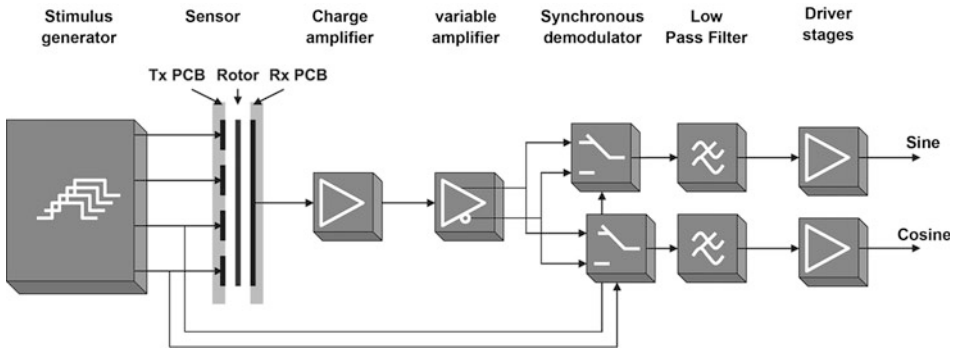


Fig. 3.154 Signal processing of capacitive encoders

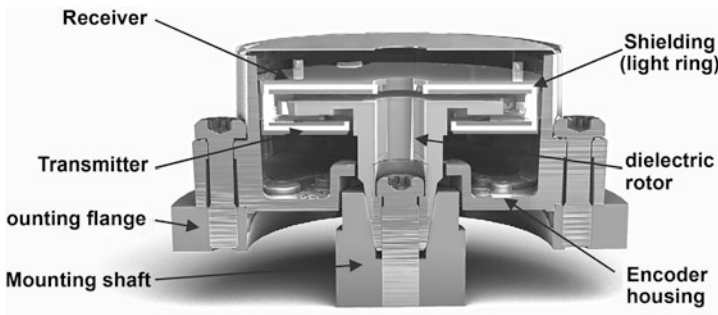


Fig. 3.155 Integrated shielding for capacitive encoders (Factory photo: SICK STEGMANN GmbH)

change in the coupling factor between the windings. If one winding is supplied with alternating current, the mutual inductance changes and thus also the induced voltage in the second winding.

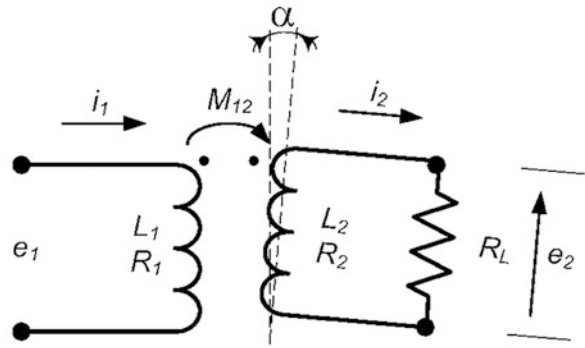
The schematic representation in Fig. 3.156 shows how the angle a between the axes of symmetry of the transformer windings can be changed.

For the mutual inductance M_{12} applies:

$$M_{12} = N_2 \frac{\Phi_2}{i_1}. \tag{3.46}$$

Here N_2 is the number of turns of the secondary winding, i_1 is the current in the primary winding and Φ_2 is the magnetic flux in the secondary winding. The rotation of the secondary winding changes the effective cross section and thus the magnetic flux Φ_2 in the following way:

Fig. 3.156 Schematic diagram of a VT with variable angle α between the windings (L_1, R_1 : primary winding; L_2, R_2 : secondary winding)



$$\Phi_2 = \mu \frac{N_1 i_1}{l} A \cos \alpha, \quad (3.47)$$

where A is the geometric cross section of the secondary winding, N_1 the number of turns of the primary winding, l the length and μ the permeability. With this equation, the result for the mutual inductance

$$M_{12} = N_2 N_1 \frac{\mu}{l} A \cos \alpha = M \cos \alpha, \quad (3.48)$$

where M is a system constant that depends on the design parameters: N_1, N_2, μ, l and A .

If the primary winding is excited with an alternating current of amplitude I_1 and angular frequency ω :

$$i_1 = I_1 \cos \omega t, \quad (3.49)$$

this results in an induced open-circuit voltage in the secondary winding of:

$$e_2 = M_{12} \frac{di_1}{dt} = j\omega M_{12} I_1 \cos \omega t = j\omega (M \cos \alpha) I_1 \cos \omega t. \quad (3.50)$$

This means: The induced voltage e_2 is proportional to $(\cos \alpha)$ in its amplitude, so that the following applies:

$$e_2 = E_2 \cos \omega t = K(\cos \alpha) \cos \omega t. \quad (3.51)$$

This dependence allows the exact determination of the angle of rotation α by measuring the amplitude E_2 of the induced voltage:

$$\alpha = \arccos \frac{E_2}{K}. \quad (3.52)$$

3.2.5.2 Significant Variants of VT

Due to the significant advantages, several physical arrangements have been developed and—very often—used under registered trademarks. The three most important representative VT designs are summarized in Table 3.25.

3.2.5.3 Resolver, a Representative Variant of VT

In the following, the most important variants of resolvers according to Table 3.25 as angle encoders (the rotor experiences a maximum of one rotation) are explained in more detail.

(a) *Sin/cos resolver (vector decomposer).*

The equivalent circuit of the sin/cos resolver is shown in Fig. 3.157.

The rotor acts as the primary winding; the stator contains two secondary windings. The output voltages form a quadrature system (Eq. (3.51)):

$$\begin{aligned} e_{S13} &= K(\sin \alpha) \cos \omega t \\ e_{S24} &= K(\cos \alpha) \cos \omega t. \end{aligned} \quad (3.53)$$

The quotient formation of the amplitudes of both output voltages enables the calculation of the function $\tan \alpha$ and thus the determination of the angle α . The quotient formation strongly suppresses or eliminates all common mode interference such as temperature, aging.

(b) *Standard resolver (electrical resolver).*

The equivalent circuit of the sin/cos resolver is shown in Fig. 3.158. Both the stator and the rotor each contain two windings, which are offset by 90° to each other.

Several operating modes of a standard turret are known (Table 3.26).

In **operating mode 1A**, the first rotor winding forms the input and is supplied with a voltage:

$$e_{R13} = E_1 \cos \omega t \quad (3.54)$$

where: ω is the angular frequency of excitation and E_1 is the amplitude of the input voltage. The second rotor winding is short-circuited.

The resolver outputs form the two stator windings (Fig. 3.158) and the output signals are two voltages offset by 90° :

Table 3.25 Significant variants of variable transformers

VT variants	Structure	Mode of operation	Designs, advantages, characteristics	Applications
Three-phase synchronous transformer (synchro)	<ul style="list-style-type: none"> • Cylindrical stator or fork rotor (H-shape) made of ferromagnetic materials • 3-phase, star-shaped secondary windings (120° spatially offset in the stator) • Rotatable single coil in the rotor 	<ul style="list-style-type: none"> • Rotor is supplied with alternating current • Output: Three induced stator voltages. They form a 3-phase geometrical system (not temporal!) i.e.: <ul style="list-style-type: none"> – frequencies and phases identical – amplitudes prop. To $\sin\alpha$, $\sin(\alpha + 120^\circ)$ and $\sin(\alpha - 120^\circ)$ 	<ul style="list-style-type: none"> • Two <i>main types</i> are known: <ol style="list-style-type: none"> 1. Rotation angle synchros (transmit rotation angle information from one shaft to a second shaft and supply power to the second shaft) 2. Standard synchros (act as angle position sensor in control system) • <i>Typical characteristics:</i> <ul style="list-style-type: none"> – Sensitivity approx. 200 mV° – Error $\leq (10/60)^\circ$ – Operating frequency: $n \cdot 100 \text{ Hz}$ 	Transmission and feedback systems in: <ul style="list-style-type: none"> • Radar systems • Cockpits • Robots • Solar cell collectors • Machine tools
Resolver	<ul style="list-style-type: none"> • Similar mechanical construction as synchro • Two-phase stator windings (spatially offset by 90° to each other) • Usually also two-phase rotor windings (same spatial offset) 	<ul style="list-style-type: none"> • Rotor and stator windings are reversible • Two windings are supplied with alternating current • The other two windings deliver induced voltages with different relationships, their phases and amplitudes 	<ul style="list-style-type: none"> • Two <i>main types</i> are known: <ol style="list-style-type: none"> 1. Sin/cos resolver (contains only one winding in the rotor) 2. Electrical resolver (contains two windings in the rotor or stator) • <i>Typical characteristics:</i> <ul style="list-style-type: none"> – Sensitivity approx. 300 mV° – Error $\leq (5/60)^\circ$ – Operating frequency: $n \cdot 100 \text{ Hz}$ to $n \cdot 10 \text{ kHz}$ 	<ul style="list-style-type: none"> • Angle measurements in feedback systems (typical, direct application) • Coordinate transformation: <ul style="list-style-type: none"> – polar \rightarrow Cartesian – artesian \rightarrow polar • Axis rotations of the Cartesian system • Phase shifting of quadrature voltages

(continued)

Table 3.25 (continued)

VT variants	Structure	Mode of operation	Designs, advantages, characteristics	Applications
Inductosyn™ [Farrand industries]	<ul style="list-style-type: none"> • Two-part system consisting of two rotating disks (stator and rotor) or of two sliding bars or belts (scale and rotor) • Each part carries a printed, meandering loop/conductor (hairpin curve) • An electrostatic shield between parts prevents parasitic capacitive coupling 	<ul style="list-style-type: none"> • Functioning as with the resolver • The fixed part is supplied with AC voltage u_e • The moving part is moved with a constant air gap to the first part (rotation or translation) • Amplitude of the induced voltage is dependent on the displacement or angle of rotation • Incremental system; the number of wavelengths passed through is also evaluated 	<ul style="list-style-type: none"> • Two <i>main types</i> are known: <ul style="list-style-type: none"> 1. <i>With a slider</i> → amplitude of the output voltage: $U_a = kU_e \cos 2\pi \frac{x}{p}$ 2. <i>With two offset sliders (P/4)</i> → amplitudes of the output voltages: $U_{a1} = kU_e \cos 2\pi \frac{x}{p}$ $U_{a2} = kU_e \sin 2\pi \frac{x}{p}$ (P is the grid dimension—typ. 2 mm -, s is the path within a geometric repeat period and k is a system constant similar to K in Eq. (3.51)) • <i>Typical characteristics:</i> <ul style="list-style-type: none"> – Repeat accuracy up to $\pm 0.5 \mu\text{m}$ – Error $\leq (5/3600)^\circ$ – Operating frequency: n·100 Hz to n·100 kHz 	Precise positioning in: <ul style="list-style-type: none"> • Computer hard disks • Machine tools • Laser beamer • Radar systems • Scanner • Antennas and radio telescopes • Aircraft • Remote control systems of the space shuttle

$$e_{S13} = \omega KE_1 (\cos \alpha) \cos \omega t$$

$$e_{S24} = \omega KE_1 (\sin \alpha) \cos \omega t \quad (3.55)$$

with α as the angle of rotation between the rotor and stator and K as a system constant, which takes into account the voltage ratio stator–rotor winding.

Fig. 3.157 Schematic diagram of a sin/cos resolver

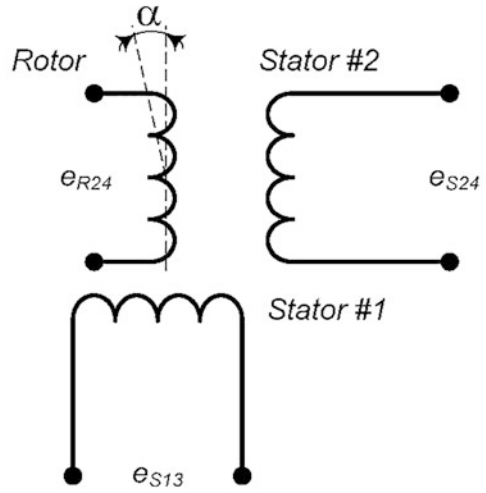
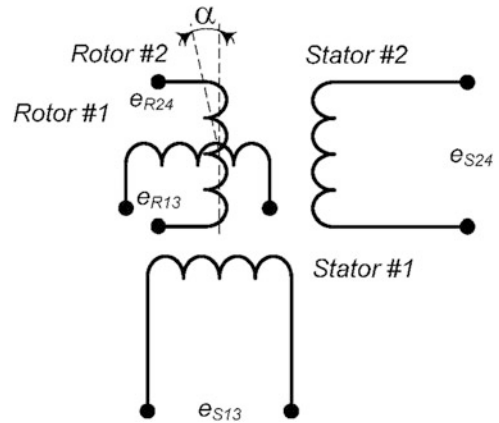


Fig. 3.158 Schematic representation of a standard resolver



The **operating mode 1B** is reversible to 1A, only the rollers are changed. The stator windings are system inputs and are supplied with the voltage:

$$e_{S13} = E_1 \cos \omega t \tag{3.56}$$

or short-circuited. In the rotor windings, the output voltages:

$$\begin{aligned} e_{R13} &= \omega K E_1 (\cos \alpha) \cos \omega t \\ e_{R24} &= \omega K E_1 (\sin \alpha) \cos \omega t \end{aligned} \tag{3.57}$$

are induced (Eq. (3.56)).

Table 3.26 Operating modes of the standard resolver

Resolver operation mode	1st stator winding	2nd stator winding	1st rotor winding	2nd rotor winding	Functionality, characteristics
1A	Both outputs		Entrance	Short circuit	Function corresponds to a classic VT.
1B	Entrance	Short circuit	Both outputs		Function corresponds to a classic VT.
2	Entrance	Entrance	Output		The stator windings are excited with two in-phase alternating voltages, which have different amplitudes.
3	Entrance	Entrance	Output		The stator windings are excited with two alternating voltages of equal amplitude, which are electrically offset by 90° to each other.

In **operating mode 2**, two synchronous AC voltages of the same frequency but with different amplitudes: E_1 or E_2 are applied to the inputs (stator connections). It applies:

$$\begin{aligned} e_{S13} &= E_1 \cos \omega t \\ e_{S24} &= E_2 \cos \omega t. \end{aligned} \quad (3.58)$$

This results in an output voltage in phase:

$$e_{R13} = E_{R13} \cos \omega t, \quad (3.59)$$

whose amplitude E_{R13} has a trigonometric dependence on the angle of rotation α :

$$E_{R13} = \omega KE \cos(\alpha - \vartheta). \quad (3.60)$$

In formula (3.61), the artificial “electrical adjustment angle” ϑ is defined by

$$\vartheta = \arctg \frac{E_2}{E_1} \quad (3.61)$$

and an amplitude E :

$$E = \frac{E_1}{\cos \vartheta} = \frac{E_2}{\sin \vartheta}. \quad (3.62)$$

In **operating mode 3**, alternating voltages of the same amplitude and frequency are applied to the inputs (stator connections), but electrically offset by 90° to each other. Then the following applies:

$$\begin{aligned} e_{S13} &= E \cos \omega t \\ e_{S24} &= E \sin \omega t. \end{aligned} \quad (3.63)$$

This results in an output voltage in phase:

$$e_{R13} = \omega KE \sin(\omega t - \alpha), \quad (3.64)$$

whose phase has a trigonometric dependence on the angle of rotation α and thus experiences a time delay τ . It then applies:

$$e_{R13} = \omega KE \sin \omega(t - \tau). \quad (3.65)$$

For the time delay τ applies:

$$\tau = \alpha/\omega. \quad (3.66)$$

In operating mode 3, the angle measurement was reduced to a *phase measurement*. The phase measurement method is a *counting method* and is usually based on the conversion of the time difference of the zero crossings of e_{S13} and e_{R12} into rectangular pulses, which form a gate signal. The counting of high-frequency pulses with a known frequency begins at the zero crossing of e_{S13} and ends at the zero crossing of e_{R12} . The number of counts gives the measured phase.

Two standard turrets can also be *cascaded* to provide *synchro-like* behavior. A classic application is a cascade consisting of a *vector decomposer* and a *standard resolver* in operation mode 2 for supplying a servo motor. The vector decomposer is supplied with alternating current. Its angle of rotation α is the setpoint and is adjusted manually. The vector decomposer supplies the two known voltages, whose amplitudes depend on α (Eq. (3.53)). These two unequal voltages supply the stator windings of the second resolver. The resulting rotor voltage (Eqs. (3.59) and (3.60)) is amplified and supplies the servomotor, whose axis in turn turns the rotor (feedback). When the servomotor reaches the setpoint, the resulting rotor voltage is zero and the motor stops.

(c) *Resolver designs for industrial applications and their characteristics.*

A large number of companies manufacture or distribute VTs. For example, LTN Servotechnik GmbH offers both encapsulated systems and completely encapsulated, mountable turret components (Fig. 3.159).

The resolvers are very *cost-effective* and extremely *robust*. They exhibit high reliability under harsh conditions such as mechanical shocks and vibrations, are relatively insensitive to extreme temperature fluctuations and to high toxic loads of chemicals and coolants. The modern brushless versions have a significantly *longer service life*; because the input voltage is transferred to the rotor winding without contact by means of a rotating

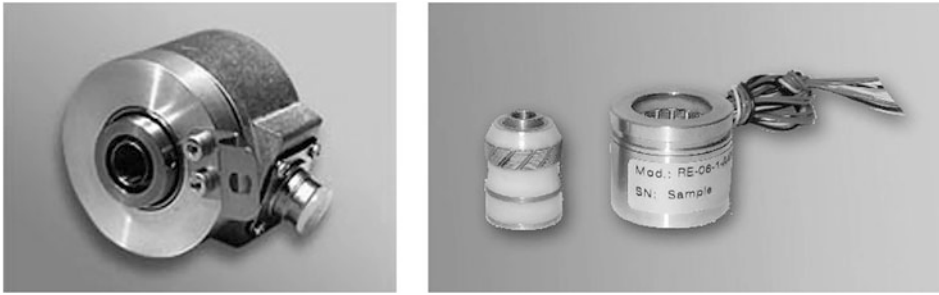


Fig. 3.159 Different resolver systems (Factory photos: LNT Servotechnik)

transformer. The resolver measures *absolute* values, in contrast to the incremental measuring devices (e.g. the linear Inductosyn). The resolver always delivers the *actual value*, even after the interruption of its supply.

The *costs* are *low*, the *designs* are extremely *miniaturized* and the electronic semiconductor industry offers a wide range of integrated circuits for further signal processing of the primary resolver output signals, if necessary directly in the resolver. The resolvers are *reversible* systems, but the manufacturer's specifications apply to a usually specified input-output assignment (unidirectional application). Table 3.27 shows the typical technical data of resolvers.

3.2.6 1Vpp or Sin/Cos Interface

This interface can only be evaluated absolutely within a period. If there are several periods, there is only one incremental evaluation for the interface.

In the measuring system with *analog sin/cos interface*, the displacement information is output via two analog values (sin and cos), each as a differential signal via four analog signals. The signals have approximately the course as shown in Fig. 3.160 as a function of the travel with a pole width of 1 mm. As an example, the offset of the sin signals is 2.5 V, that of the cos signals 2.0 V. This offset is permissible.

$$U_{\sin}(s) = \hat{U} \cdot \sin(\omega \cdot s) + U_{0\sin} \quad (3.67)$$

$$U_{-\sin}(s) = -\hat{U} \cdot \sin(\omega \cdot s) + U_{0\sin} \quad (3.68)$$

$$U_{\cos}(s) = \hat{U} \cdot \cos(\omega \cdot s) + U_{0\cos} \quad (3.69)$$

$$U_{-\cos}(s) = -\hat{U} \cdot \cos(\omega \cdot s) + U_{0\cos} \quad (3.70)$$

with $\omega = 2 \cdot \pi/P$ circular period, where P is the pole width of the measuring body.

Table 3.27 Technical data of resolvers

Shaft or hollow shaft diameter	3 mm to 165 mm
Outside diameter	20 mm to 200 mm
Length	10 mm to 100 mm
Outdoor temperature range	-55 °C to 155 °C
Speed	Up to 20,000 min ⁻¹
Absolute accuracy	To ±5'
Input frequency	Up to 12 kHz
Signal level	Up to 12 V _{eff}

$$\hat{U} \approx 0,25V$$

$$U_{0\sin} \approx 2,5V$$

$$U_{0\cos} \approx 2V.$$

In the customer control or in measuring systems with digital output, an *interpolator* takes over the digitization of the sin/cos signals. It initially forms individual signals from the differential signals:

$$\begin{aligned}
 U_{si}(s) &= U_{\sin}(s) - U_{-\sin}(s) \\
 (1) - (2) &= \left(\hat{U} \cdot \sin(\omega \cdot s) + U_{0\sin} \right) - \left(-\hat{U} \cdot \sin(\omega \cdot s) + U_{0\sin} \right) \\
 &= 2 \cdot \hat{U} \cdot \sin(\omega \cdot s)
 \end{aligned} \tag{3.71}$$

$$\begin{aligned}
 U_{co}(s) &= U_{\cos}(s) - U_{-\cos}(s) \\
 (3) - (4) &= \left(\hat{U} \cdot \cos(\omega \cdot s) + U_{0\cos} \right) - \left(-\hat{U} \cdot \cos(\omega \cdot s) + U_{0\cos} \right) \\
 &= 2 \cdot \hat{U} \cdot \cos(\omega \cdot s)
 \end{aligned} \tag{3.72}$$

Trigonometric arctan and arccot functions can be used to determine the current angle:

$$\begin{aligned}
 \alpha &= \arctan(U_{si}(s)/U_{co}(s)) \\
 &= \arctan\left(2 \cdot \hat{U} \cdot \sin(\omega \cdot s) / 2 \cdot \hat{U} \cdot \cos(\omega \cdot s)\right) \\
 &= \left(\hat{U} \cdot \cos(\omega \cdot s) + U_{0\cos} \right) - \left(-\hat{U} \cdot \cos(\omega \cdot s) + U_{0\cos} \right) \\
 &= 2 \cdot \hat{U} \cdot \cos(\omega \cdot s)
 \end{aligned}$$

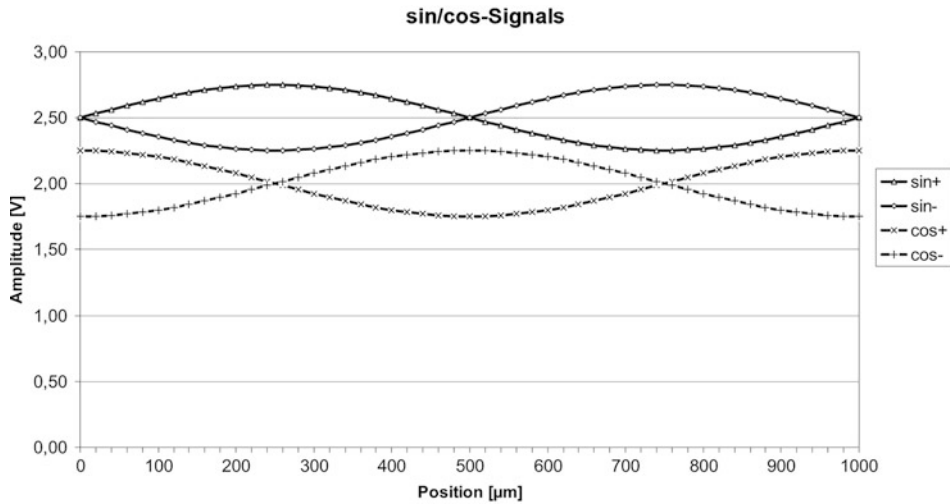


Fig. 3.160 Analog sin/cos interface

$$\begin{aligned}
 \alpha &= \arctan (U_{\text{si}}(s)/U_{\text{co}}(s)) \\
 &= \arctan (2 \cdot \hat{U} \cdot \sin (\omega \cdot s) / 2 \cdot \hat{U} \cdot \cos (\omega \cdot s)) \\
 &= \arctan (\sin (\omega \cdot s) / \cos (\omega \cdot s)).
 \end{aligned} \tag{3.73}$$

Figure 3.161 shows that the arctan function is repeated twice within one period. The distinction must be made as to whether the cos signal is >0 or <0 .

To get from the angle α Eq. (3.73) to the position s within the pole, the angle must be divided by the angular frequency ω .

$$s = \alpha / \omega. \tag{3.74}$$

The position within a pole is unique. However, each pole is measured identically.

In Eq. (3.73), it can be seen that the amplitude \hat{U} is no longer important for the determination of angles. It is only relevant insofar as \hat{U} should be sufficiently large because of the signal-to-noise ratio. Usually the angle determination is not exclusively determined by the trigonometric function, but by *tracking methods*. In this case, the singular places (division by zero) need not be taken into account. There are several already integrated chips that perform this interpolation. One manufacturer is for example the company IC-Haus (www.ichaus.de). When digitizing the sin/cos signals, more than 10 bits are possible without noise or additional hysteresis. The digitized signals are available after a conversion time of some 10 ns. With a pole width of 1 mm, a resolution of 1 μm can be achieved with 10 bits.

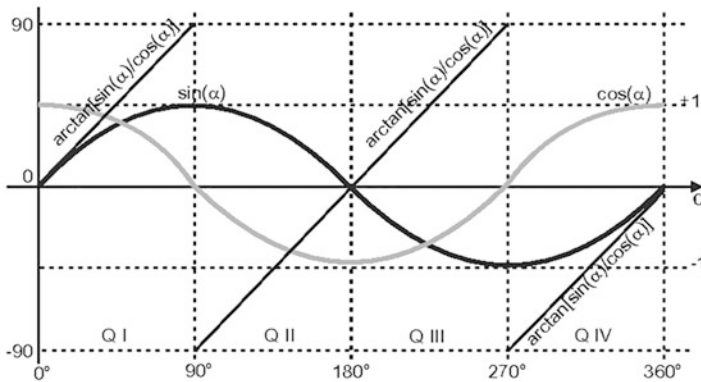


Fig. 3.161 Evaluation of the sin and cos signal

3.2.7 Incremental Encoders

In a measuring system with *digital interface*, the position signal is output via two digital signals: A and B. A common designation for this is “A/B interface” or “90° (pulse) interface”. The physical level is either a 5-V differential signal (RS422) or a single-ended signal that assumes either the level of the supply voltage or zero.

Figure 3.162 shows the logical signals of a digital A/B interface. Each edge on signal A or signal B means the transmission of one increment. The counting direction is shown in the third line. The fifth line shows the counter position, which, for example, starts at the value 40. Within one period of signal A or B, 4 increments are counted. There are two ways to decide whether the edge means a positive or a negative increment:

If signal A comes *before* signal B, each edge corresponds to one increment in *positive* direction, and if signal A comes *after* signal B, it corresponds to one increment in *negative* direction.

The counting direction of one signal edge is defined by the level of the other signal. An example is the rising edge of A in Fig. 3.162 at counter position 41 when viewed in the forward direction. Here the level of B is “low”. The edge is evaluated as a positive increment. At the rising edge of A at position 42 during the backward movement, the level of B is “high”. Accordingly, this edge is evaluated as a negative increment. For the interpretation of a movement, the path s_{ink} must still be defined, which corresponds to one increment. For magnetic measuring systems, values between about 1 μm and several mm are common here.

Another important variable is the minimum edge distance t_{\min} that the controller can still detect. Since the interface only operates incrementally, no increment must be lost in the control. The minimum edge time that a controller can still count must not be undercut when the increments are output. The maximum travel speed is defined by the distance per increment and the minimum edge distance. This value must be taken from a table of the sensor manufacturer (Fig. 3.163).

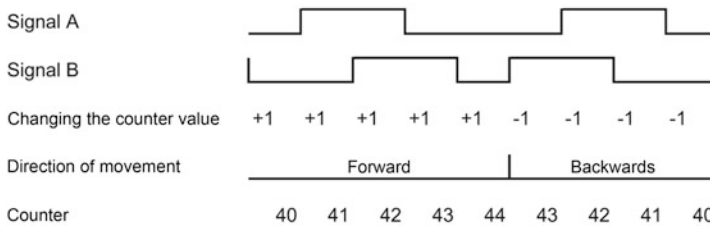


Fig. 3.162 Digital interface

min. edge separation X	mechanical resolution Y:			
	D = 1 μm	E = 2 μm	F = 5 μm	G = 10 μm
D = 0.12 μs	5 m/s	10 m/s	20 m/s	20 m/s
E = 0.29 μs	2 m/s	4 m/s	10 m/s	10 m/s
F = 0.48 μs	1 m/s	2 m/s	5.41 m/s	5.41 m/s
G = 1 μs	0.65 m/s	1.3 m/s	2.95 m/s	2.95 m/s
H = 2 μs	0.3 m/s	0.6 m/s	1.54 m/s	1.54 m/s
K = 4 μs	0.15 m/s	0.3 m/s	0.79 m/s	0.79 m/s
L = 8 μs	0.075 m/s	0.15 m/s	0.34 m/s	0.34 m/s
N = 16 μs	0.039 m/s	0.079 m/s	0.19 m/s	0.19 m/s
P = 24 μs	0.026 m/s	0.052 m/s	0.13 m/s	0.13 m/s

Fig. 3.163 Relationship between mechanical resolution, minimum edge separation and maximum speed

The maximum velocity v_{max} cannot be determined by the formula (3.75). The reason for this lies, among other things, in the component tolerances:

$$v_{max} = s_{ink} / t_{min}. \tag{3.75}$$

The digital A/B signals are usually transmitted as RS422 differential or HTL (single-ended) signals. With RS422, greater distances or higher transmission rates can be bridged than with HTL.

3.2.7.1 Summary of the Properties of Incremental Interfaces

Various interfaces exist for the incremental measuring system, the most important properties of which are summarized in Table 3.28.

3.3 Inclination

Many applications require the measurement of *inclinations*. Sensors for inclination measurement are used in many industries, for example in consumer electronics or in automotive and road construction. For this reason, the requirements for the sensors are very different. While the low price is of primary importance for consumer electronics, for example, accuracy and time stability are important for use in construction monitoring.

Table 3.28 Interfaces of incremental measuring systems

Interface	Incremental measuring system	
	Digital	Sin/cos analog
Designation	RS422/HTL	1 Vpp
Transmitted frequencies	High	Down
Harmonics	High	Low
Achievable interference immunity	Medium/low	High
Applications	Mechanical engineering	Drive technology
Flexibility in the evaluation	Low	Depending on the speed, interpolation can be performed in different ways.

The angle or inclination sensors can be divided into three groups. The first group are the *rotation sensors*, which are known from displacement measurement (Sect. 3.1.1). These sensors are attached to a rigid reference system and are mechanically connected to the rotating measuring object to be measured.

The second group of sensors does not rely on a rigid connection to a reference system, but exploits *physical effects*. The most common method is to use the *gravitational force* of the earth as a reference system. The advantage is that one is largely independent of the installation situation. In this way, very small sensors can be realized that use different measuring principles, which will be discussed in more detail in the following. However, measuring the inclination with the aid of the acceleration due to gravity is a problem in accelerated systems such as in a car, because this acceleration overlaps with the acceleration due to gravity. *Magnetic sensors* solve this problem. However, their accuracy is limited compared to other sensors.

The third group of sensors does not determine the inclination, but the deviation of an *angular momentum* from a direction. They work like a *gyroscope*. It is a gyroscope that rotates in a moving bearing. This rotation generates an angular momentum that always maintains its direction when no external moments act on it. The direction relative to the direction of the angular momentum can be measured as an angle.

Slope

Definition: *Inclination is the relative position of a direction in relation to the horizontal or vertical.*

Inclinometers, also known as *inclinometers* or *tilt sensors*, make use of the principle of “dropping a perpendicular”. This measurement is usually made in relation to the *gravitational force* or a *magnetic field*. The gravitational field can be a mechanical pendulum, a bending beam or, as in a spirit level, an electrically conductive liquid in a gas bubble.

3.3.1 Magnetoresistive Inclination Sensors

Magnetoresistive inclinometers exploit the effect that the resistance of a conductor changes as a function of a magnetic field (Sect. 2.3). Permalloy™ (81% Ni, 19% Fe) is usually used as the material for this. The resistance is greatest when the angle between the charge-carrying material and the magnetic field is 0° . The resistance is lowest when the magnetic field is perpendicular to the surface (angle of 90°). The change in the resistance value is relatively small and is in the range of 2% to 3%. It applies approximately:

$$R = R_0 + \delta R \cdot \cos(2 \cdot \alpha).$$

where R is the value of the resistance, R_0 and δR are material-dependent parameters and α is the angle between the current and the magnetic field (measured as magnetic flux density \mathbf{B}).

Figure 3.164 shows the principle structure of a magnetoresistive sensor. On the left is the material that changes its resistance value as a function of a magnetic field \mathbf{B} . On the right is the schematic structure of a sensor consisting of magnetoresistive elements. The arrow represents the sensitive axis.

The structure of a typical inclination sensor is shown in Fig. 3.165. *Two magnets* are mounted *pendulously* near two magnetoresistive elements.

The pendulously suspended magnets align themselves in the gravitational field. This corresponds to the falling of a plumb line. The output signal is proportional to the inclination. These magnetic elements form a voltage divider. At 0° , the magnets are exactly in the middle. If this system is tilted, the position of the magnets relative to the elements changes. This changes the resistance, which results in a change of the output signal (proportional to the inclination). Usually these sensors are filled with a liquid (e.g. silicon oil), which serves as damping.

However, it should be remembered that the *viscosity* of liquids is *strongly dependent on temperature*. This means that damping will also depend on temperature. The more viscous the liquid, the higher the damping. An additional control electronics is not necessary with sensors of this type. The output signal is usually proportional to the inclination. Typically, the output signal is between 40% and 60% of the input signal. With some sensors of this design, an unfavourable behaviour of the sensor signal can be observed when the sensor is tilted out of its measuring range.

Magnetoresistive sensors are suitable, for example, for monitoring traffic lights or aligning slow-moving tracking systems of solar collectors. Due to their simple design, the production costs for an entire sensor system are relatively low.

However, other systems that react to an externally present magnetic field are magnetoresistive sensors integrated into IC packages. These sensors are also used in radiant environments. In contrast to systems with highly integrated electronics, these sensors are not sensitive to radiation influences. Magnetoresistive sensors generally have *good sensitivity*. However, the sensors cannot be used if strong magnetic fields are present. The use of current-carrying compensation coils should therefore be avoided.

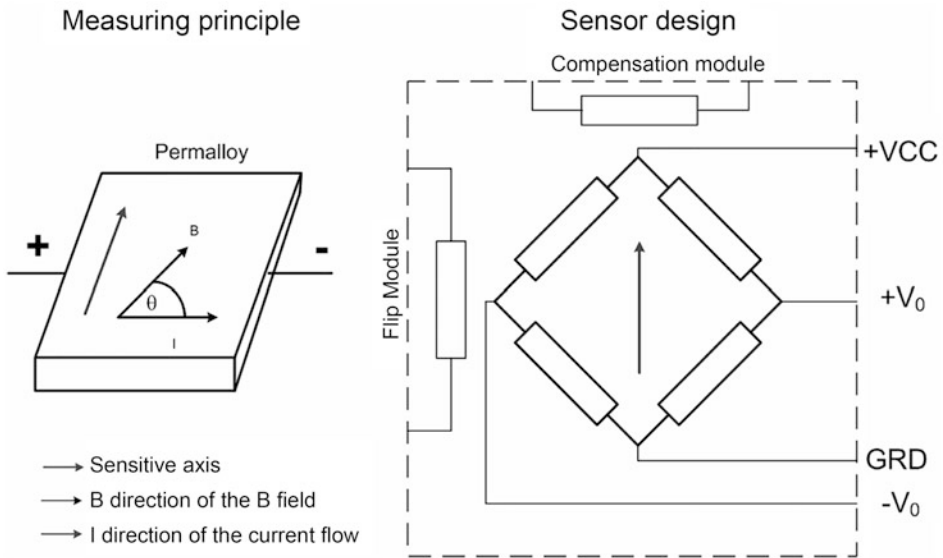


Fig. 3.164 Illustration of the magnetoresistive measuring principle

3.3.2 Compass Sensors

Compass sensors are magnetoresistive sensors that measure the *external magnetic field of the earth*. However, this is quite difficult for two reasons. On the one hand, the earth's magnetic field is relatively weak and is subject to geographical variations. On the other hand, the magnetic flux lines are inclined, i.e. they only run parallel to the earth's surface at equatorial latitudes. One area of application for these sensors is primarily in mobile applications. An exact calibration of these sensors with the measurement electronics is always necessary if the measurement environment contains magnetic materials such as iron. For these reasons, compass sensors are often combined with *MEMS inclinometers* (Sect. 3.3.5).

3.3.3 Electrolytic Sensors

Electrolytic sensors work in principle like a *spirit level*. An electrolytic liquid is contained in an air bubble (Fig. 3.166). The liquid always strives to stand "in the water", i.e. to align its surface horizontally in the resting system.

As Fig. 3.166 shows, there are three electrodes (a, b and c) in the electrolytic liquid. If the dome is tilted, the immersion depth of the electrodes is different. The electric current between electrodes a and b is greater than the current between electrodes b and c. Thus, the

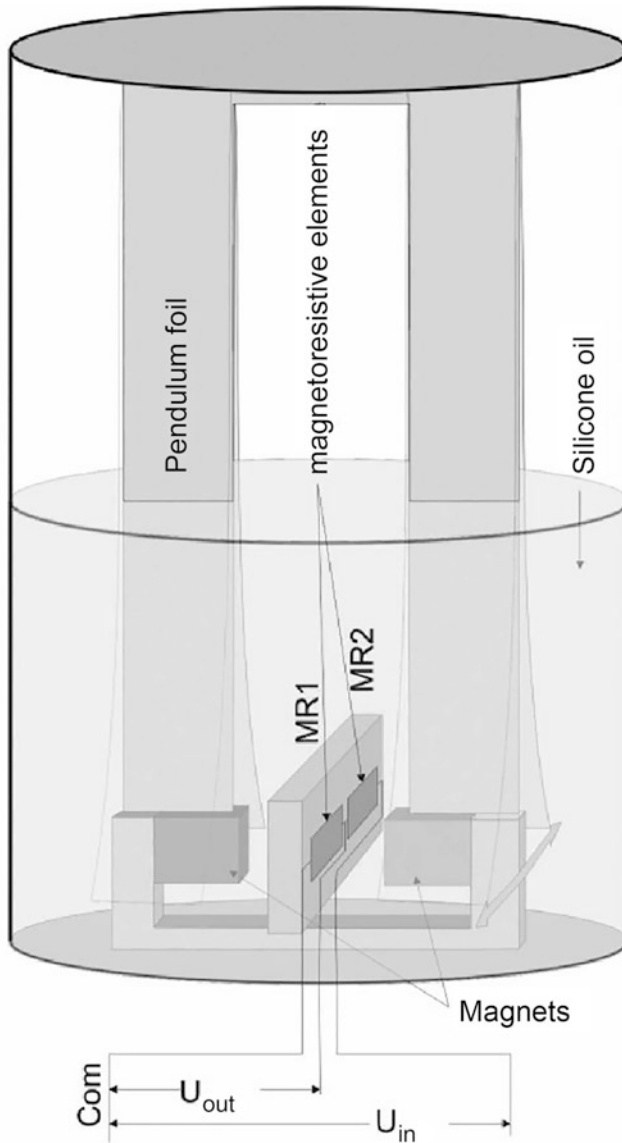
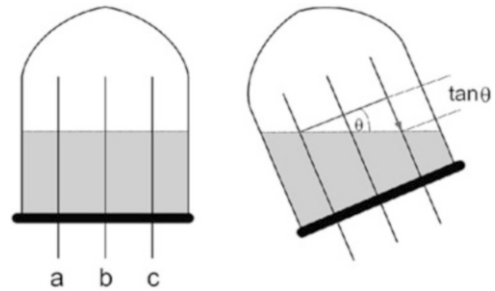


Fig. 3.165 Structure of a typical magnetoresistive inclination sensor

electrolytic sensor behaves like a *voltage divider* (potentiometer) whose *resistance* changes *proportionally to the tilt angle*.

Although the principle is simple, a voltage cannot simply be applied to this structure like to a normal potentiometer. This would result in a chemical reaction taking place in the electrolytic liquid, in which the positive ions of the liquid would migrate to the cathode to recombine with electrons. This would reduce the conductivity and eventually lead to a

Fig. 3.166 Schematic diagram of the electrolytic measuring principle



complete standstill. To prevent this, an *alternating voltage* with frequencies between 25 Hz and 4000 Hz is applied to the electrodes.

The electrolytic sensors have other special features: Due to the electrolytic liquid, their signals are *temperature-dependent*. Furthermore, the *filling* and the *size of the air bubble* play an important role. As a rule, these sensors are not suitable for systems that are subject to vibrations. In these cases, there is no clear angle of inclination between the liquid level and the horizontal, so that the measuring principle fails.

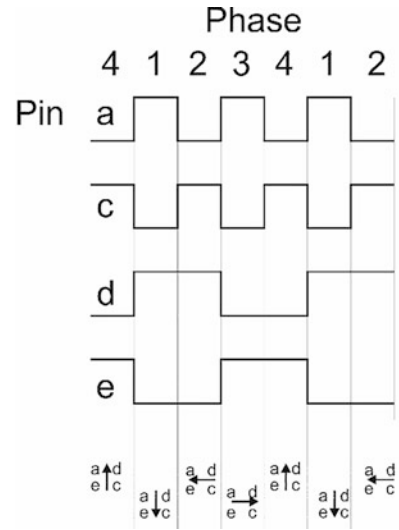
A great advantage of this sensor is that it can also be built as a *two-dimensional sensor* (Fig. 3.167).

Two solutions are conceivable: On the one hand, in the simplest case, only *one axis* can be electrically controlled at a time. During this time, the other axis is inactive. On the other hand, it is possible to apply *two excitation frequencies*, where one is selected twice as high as the other. In this case, all pins are controlled. The advantage of the former method, in which only one axis is active at a time, is the simpler electronics and lower power consumption. Furthermore, the measuring signal along the measuring axes is proportional to the inclination. This usually leads to a higher sensitivity.

3.3.4 Piezoresistive Inclination Sensors/DMS Bending Beam Sensors

Inclinometers, which operate on the basis of a *bending beam*, consist of a web to which a mass is attached. When the mass is accelerated, the web bends. The bending beam of the sensor is equipped with suitable strain gauges. This results in a *Wheatstone bridge*. The output signal is proportional to the bending of the beam and thus proportional to the strain of the strain gauge. The behaviour of such a sensor with regard to sensitivity and frequency response depends on the choice of suitable materials and the appropriate strain gages (Sects. 2.2.2 and 4.2). If silicone oil is added, the system can be damped. In this case, vibrations no longer affect the measurement behavior. However, the measurements are temperature-dependent. With *electronic low-pass filters* (e.g. due to small capacities), the signals can be damped *independent of temperature*.

Fig. 3.167 Representation of the phases and the direction of the current flow (bottom) in a two-dimensional electrolytic sensor



3.3.5 MEMS

MEMS (*Micro-Electro-Mechanical System*) sensors are systems in which electronic circuits and mechanical properties are integrated in the smallest space, usually on a substrate or chip. In general, these sensors are designed to measure small accelerations and can therefore be used to measure inclinations in relatively static cases. The measuring principle is *capacitive* (Sect. 2.6). Here the accelerated mass and the sensor body are isolated from each other. The mass and the sensor body form the capacitor plates. The charge or capacitance of this system is measured. If the distance between the mass and the sensor body decreases, the capacitance increases because the system can store more charge. On the other hand, the distance has increased, which contributes to a decrease in capacity (Fig. 3.168).

MEMS sensors offer many advantages: they have *high sensitivity* and are very *temperature stable*, i.e. only small drifts occur over a long period of time. Due to their design, these sensors can be constructed with *overload stops* so that even high vibrations of several hundred g (g: acceleration due to gravity) do not damage these systems. A disadvantage is that the electronics must be located close to the sensitive element. Parasitic currents are very disturbing for these systems.

Increasingly, however, complete accelerometers are already being offered as *single-chip solutions* for applications in consumer electronics or automotive engineering, for example. These are very inexpensive.

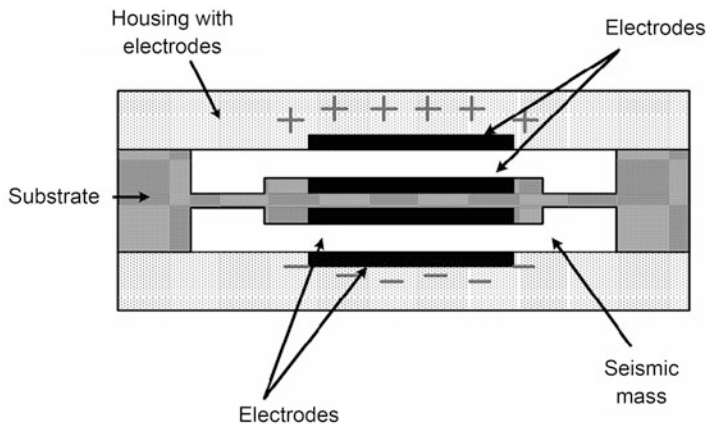


Fig. 3.168 Representation of a capacitive sensor principle of a MEMS. The seismic mass and the housing of the sensor form a capacitor. The closer the seismic mass comes to the cover of the sensor, the greater the capacitance of this capacitor

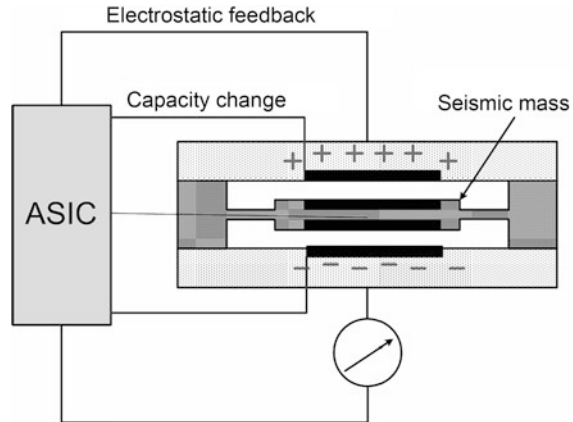
3.3.6 Servo-Inclinometer

Servo-inclinometers can be regarded as MEMS inclination sensors with particularly high sensitivity without direct current (DC) output (Fig. 3.169: ASIC is an application-specific integrated circuit). The high sensitivity is achieved because the system is *feedback*: In a mechanical, classic sensor, a permanent magnet in a coil is accelerated with its mass, so that a current flow is induced in the coil. This signal is measured and a current is generated in the opposite direction. This current builds up a magnetic field in the coils, which is opposite to the magnetic field of the magnet. Thus, the magnet is returned to its old position. The seismic mass is thus held in the zero position by a voltage proportional to the inclination. The feedback system makes it possible to measure *very small accelerations* exactly and thus to determine the inclination very accurately. This setup allows measurements above the natural frequency of the system.

Capacitive sensors based on MEMS measure accelerations *below their natural frequency*. The operating principle is however identical. The seismic mass, which is located as a capacitor between the housing cover, is kept in the neutral position by an applied charge. The charge required to achieve this is also proportional to the acceleration of the sensor element.

In order to achieve maximum sensitivity, the spring constants of the restoring mechanical elements are selected as small as possible. This allows values for a maximum acceleration of 0.2 g to be achieved. The frequency ranges apply to accelerations between 3 g and 200 g. The minimum resolution that such systems achieve in practice is $30 \text{ ng}/\sqrt{\text{Hz}}$ and a dynamic range of 115 dB. The noise of such systems is 66 times lower than that of systems with high-resolution MEMS accelerometers.

Fig. 3.169 Schematic diagram of a MEMS servo inclinometer



3.3.7 Overview and Selection of Inclination Sensors

The selection of the appropriate inclination sensor depends on the following parameters:

- *Accuracy* in resolution and repeatability,
- *Vibrations* and
- *Temperature dependence*.

Table 3.29 shows the relationships.

3.4 Sensors for Object Detection

3.4.1 Proximity Switch

Non-contact object detection is carried out cost-effectively and extremely reliably in all areas of application—from traditional industry to medical and aerospace applications—using *NS* (*proximity switches*). It represents the binary variant of the proximity sensor, which is why the basic statements can be taken from it (Sect. 3.1.1). In principle, the NS family contains *versions of* all distance sensors (e.g. inductive, capacitive, optoelectronic) with *binary outputs* (Fig. 3.170).

Because of the wide field of application and the high quantities and in order to ensure compatibility or interchangeability, the LV are specified in detail in the standard DIN EN 60947-5-2 “Low-voltage switchgear/control gear and switching elements/proximity switches”. The standard refers to Table 3.30:

- *Inductive* and *capacitive* LV, which detect the presence of metal and/or non-metallic objects
- *Ultrasonic NS*, which detect the presence of sound reflecting objects.
- *Photoelectric NS*, which detects the presence of all objects.
- *Non-mechanical magnetic NS*, which detect the presence of objects with a magnetic field.

Table 3.29 Properties and applications of the inclination sensors

Sensor	Application
Servo-inclinometer	For high-precision measurements in geophysical areas, e.g. for levelling and in building monitoring, – For alignment of precision antennas or – For measurements of rails and axles.
Mems	In systems that execute fast movements and are subject to shocks (e.g. in consumer electronics, vehicle control and robotics applications).
Electrolytic inclination sensors	High resolution (e.g. for checking construction vehicles or for low-vibration applications).
Magnetoresistive inclination sensors	As the most cost-effective solution in low-vibration environments (e.g. for the control or tracking of solar collectors and for the control of offshore installations).
Compass sensors	Use in mobile applications (e.g. vehicle navigation).

Sensor	Resolution	Non-linearity	Repeatability
Servo-inclinometer	0.00005°	<0.05°	<0.001°
MEMS	0.005°	<0.03°	<0.06°
Electrolytic inclination sensors	0.001°	<0.10°	<0.01°
Magnetoresistive inclination sensors	0.005°	<0.05°	<0.01°
Compass sensors	0.1°	<0.05°	<0.01°

**Fig. 3.170** All types and designs of proximity switches (Factory photo: Balluff GmbH)

According to EN 60947-5-2, the NS is a *position switch* that is operated without mechanical contact with the moving part and has the following characteristics:

- (a) The NS forms a *unit*.
- (b) The binary switching output of the NS contains a *semiconductor switching element*.
- (c) The NS is intended for use in circuits whose rated voltage *DC* does not exceed 300 V or *AC* 250 V - 50/60 Hz.

 **Leuze electronic**

the **sensor** people



SMART
SENSOR
BUSINESS

SMARTER **PRODUCT USABILITY**

SMARTER **APPLICATION KNOW-HOW**

SMARTER **CUSTOMER SERVICE**

www.leuze.com

The advertisement features a woman in a red shirt standing in front of a blurred industrial background. A red banner contains the text 'SMART SENSOR BUSINESS' and three bullet points: 'SMARTER PRODUCT USABILITY', 'SMARTER APPLICATION KNOW-HOW', and 'SMARTER CUSTOMER SERVICE'. Below the banner, several Leuze electronic sensors are displayed, including a large yellow sensor, a red sensor, a black sensor, and several smaller sensors in various colors (red, white, orange).

Table 3.30 provides an overview of the possibilities for implementing NS.

3.4.1.1 Block Diagram of the NS, Detection Type and Mode of Operation of the NS

The one-piece NS consists of the *sensor element* and *sensor electronics* completely integrated in the sensor. According to its main functions, the sensor electronics is divided into two *functional areas*. In the *front-end area*, the following takes place:

- the excitation/supply of the sensing element,
- the evaluation of the distance information obtained and
- converting this response into an electrical analog signal (Fig. 3.171).

In the *back-end area*, the following takes place:

- the generation of a primary switching signal from the analog signal,
- further signal processing,
- the generation of the standardized NS output signal and
- the execution of the multitude of protective functions required by the standard.

The realization of the sensing element and the front end depends entirely on the *type of detection*. In Fig. 3.171, these stages are summarized as “black-box”. Details and sources of literature in this book on these components can be found in Table 3.30.

The electronics in the back end have an almost identical structure and function regardless of the proximity switch type. The electrical analog signal U_e is fed to a *threshold value comparator*. This compares the signal with a threshold voltage U_{SCH} and generates the primary switching signal U_s , which is further processed by the logic stage. This stage provides input switching signals for both the sensor output stage and the LED driver. In the first stage, appropriate amplification and conditioning takes place so that the final bounce-free LV output signal meets the standard. The LED driver generates logic signals to supply the optical indicators:

- yellow LED (prerequisite) as NS status display,
- green LED (option) as LV supply voltage display,
- red LED (rare option) for NS error message,
- one of these colours (flashing instead of continuous) or another colour (continuous) as an indication of other functions (e.g. short-circuit message, contamination indicator, performance reserve, setting aid).

A voltage regulator stabilizes the wide permissible fluctuations of the supply voltage U_{BAT} and supplies the LV interior with constant low voltage.

The application of the LV is shown schematically in Fig. 3.171 by the voltage source U_{BAT} and the load resistor R_L , which are connected to the LV connections: $+U_B$, OUT and GND. The source symbolizes the power supply (DC/AC: regulated/unregulated) for the LV supply, the load resistance symbolizes the actual load (relay, PLC input, ohmic resistance, or complex impedance).

Table 3.30 Overview of proximity switches

Proximity sensor type and definition	Entry type	Principle of operation	Main part of the sensing element	Main stage (s) in electronics Front-end	See book section:
Inductive NS (generate electromagnetic field)	Inductive	Push button	Coil	Oscillator	3.1.1.2
Capacitive NS (generate an electric field)	Capacitive	Push button	Structured capacitor	Oscillator	2.6
Ultrasound NS (emitting ultrasonic waves)	Ultrasound	Push button	Piezoelectric ultrasonic transducer	Oscillator	3.1.3
Magnetic field NS (detect the presence of a magnetic field)	Magnetic	Push button	Hall, MR, GMR	Measuring bridge, amplifier	2.3 and 2.8
Optoelectronic NS (type D) • Light sensor (detects objects that reflect visible/invisible optical radiation)	Photoelectric	Probe; target reflects light	Transmit and receive semiconductor	Transmitter and receiver in one sensor	3.1.2
Optoelectronic NS (type R) • Reflection light barrier (detects objects that interrupt visible/invisible optical, reflected radiation)	Photoelectric	Push button; target interrupts light path between transmitter and reflector	Transmit and receive semiconductor	Transmitter and receiver in one sensor	3.1.2
Optoelectronic NS (type T) • Through beam light barrier (detects objects that interrupt visible/invisible optical direct radiation)	Photoelectric, direct light beam	Throughbeam photoelectric sensor; target interrupts light path between transmitter and receiver	Transmit and receive semiconductor	Transmitter and receiver separated into two units	3.1.2

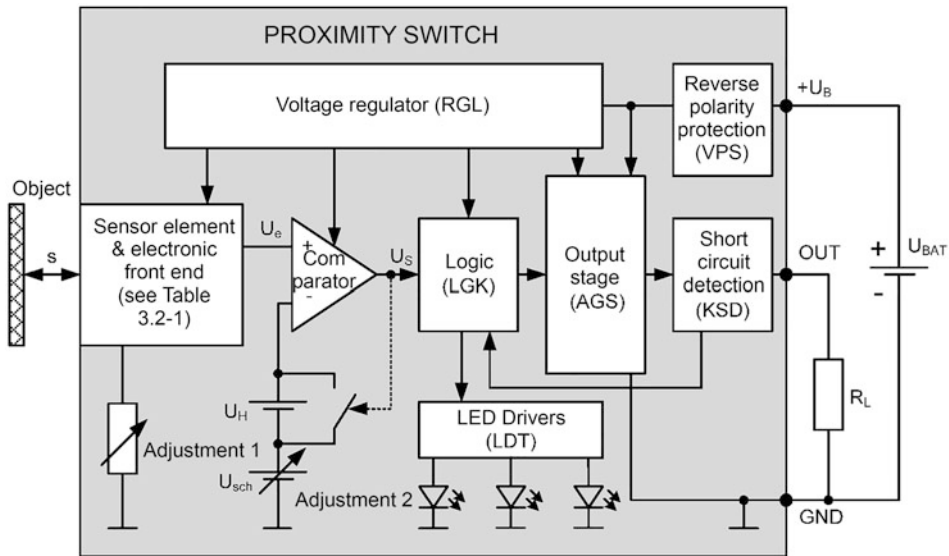


Fig. 3.171 Block diagram of a proximity switch with three connections

3.4.1.2 Main Features of the NS

If the LV is considered abstractly as a *binary converter* of *path information* into an electrical switching signal, the following features are of the greatest importance for the system description:

- (a) Switching distance and operating conditions,
- (b) Hysteresis (switching reversal span),
- (c) Switching element function and
- (d) Output type (output switching direction).

3.4.1.3 Switching Distances

The standard generally defines the *switching distance* (s) as the distance at which a standard measuring plate can be approached:

- to the active surface through which the NS field escapes, and
 - along the reference axis perpendicular to this surface (axial approximation),
- a signal change is caused at the LV output OUT.

The general term is defined in more detail by the following standard sizes (Fig. 3.172):

1. *Rated operating distance* (s_n): conventional size for determining the operating distances. It takes into account neither specimen variations nor changes due to influences such as supply voltage and temperature. The equivalent term for optoelectronic LV is the

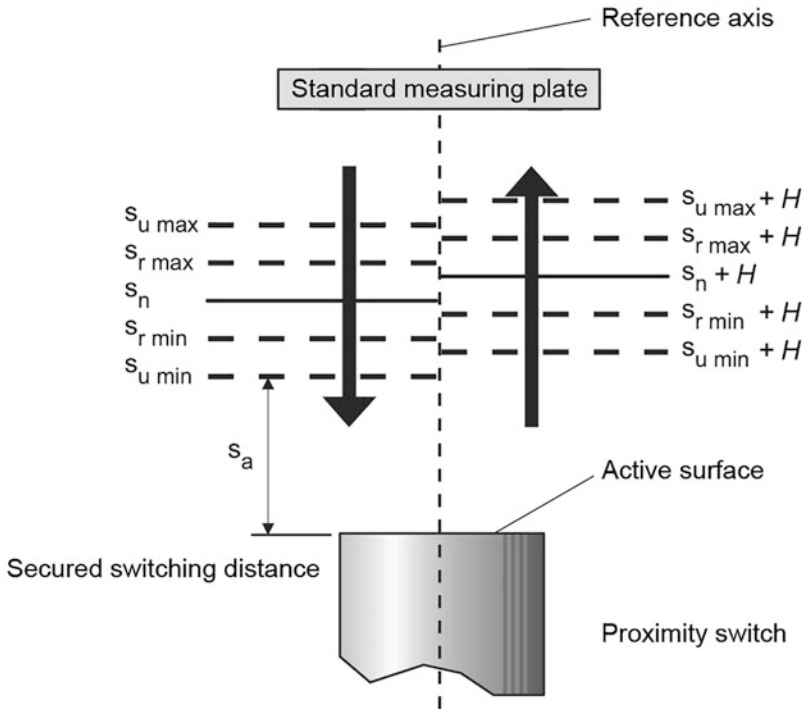


Fig. 3.172 Switching distances of inductive and capacitive NS

detection range (s_d); this characterises the range in which the switching distance can be set.

2. *Real operating distance* (s_r): the operating distance of a single LV measured under specified test conditions. Its value must satisfy the following equation:

$$0.9s_n \leq s_r \leq 1.1s_n. \quad (3.76)$$

3. *Useful operating distance* (s_u): the operating distance of a single LV, measured over the specified temperature or supply voltage range. Its value is within the range:

$$0.9s_r \leq s_u \leq 1.1s_r. \quad (3.77)$$

4. *Safe operating distance* (s_a): the distance from the active face at which operation of the LV is guaranteed, taking into account all environmental conditions and specimen dispersion. Its value satisfies the equation:

$$0 \leq s_a \leq 0.9 \cdot 0.9s_n. \quad (3.78)$$

3.4.1.4 Hysteresis

The definitions of the switching distances apply to an approximation of the standard measuring plate to the LV and along the reference axis. To ensure stable, bounce-free switching of the LV output, the LV has *increased switching distances* when the standard measuring plate is removed. This increase is called *LV hysteresis*; its value is defined in the standard by the following equation:

$$H \leq 0.2s_r. \quad (3.79)$$

Figure 3.172 shows the relationship between the switching distances for inductive and capacitive LV.

There are basically two possible implementations of hysteresis:

1. Due to the abrupt increase of the threshold voltage U_{sch} (Fig. 3.173a) In Fig. 3.171, this mechanism is symbolically represented by a switch in parallel with the additional voltage source U_H . If the measurement starts at the distance $s \geq 3 s_r$ and the switching point ($U_e = U_{sch}$) is reached as the target approaches, the comparator gate output signal U_s is switched over and the switch is opened. Accordingly, the threshold voltage jumps up with the value U_H . The movement continues until the distance $s \leq 0.3 s_r$. For the target distance, a higher distance s must be achieved to fulfill the switching condition $U_e = U_{sch} + U_H$ and to achieve the reversal of the switching output U_S (Fig. 3.173a).
2. The same behavior is achieved with a constant threshold voltage U_{sch} and with reversed, abrupt change of the comparator input signal U_e (Fig. 3.173b).

3.4.1.5 Switching Element Function

This parameter describes the relationship between the current object-sensor distance and LV output state (Table 3.29). The standard provides for three switching element functions:

N/O function: it causes the load current to flow when the object is detected and the load current not to flow when the object is not detected.

NC function: it is the inverted function of the NO function.

Changeover function: requires the combination of two switching elements and contains a normally open and a normally closed function (*complementary* switching outputs).

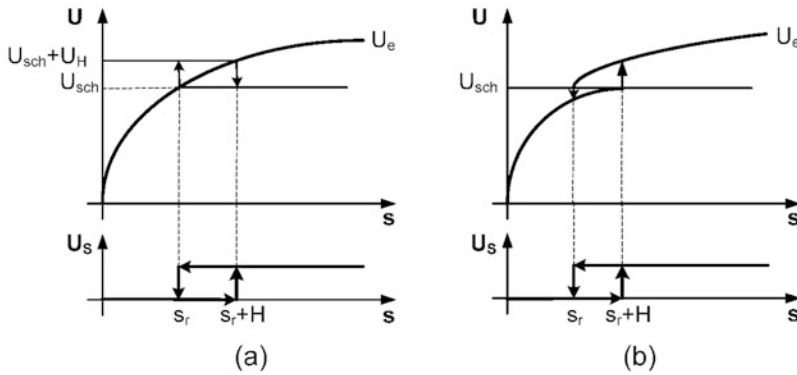


Fig. 3.173 Generation of the hysteresis by threshold jump (a) or useful signal jump (b)

3.4.1.6 Output Type

The standard basically specifies two output types for LV with DC supply voltage:

Positive switching output (PNP output): the load R_L is electrically connected to ground GND. When the NS is actuated, the load current will flow from the $+U_B$ connector through the output stage of the NS, through the NS output OUT and through R_L to GND. This output type is common in Europe.

Negative switching output (NPN output): It describes the reverse version: the load R_L is electrically connected to the $+$ pole and is switched by the NS against ground GND as long as the NS is actuated. This output type is common in Asia.

The possible combinations of the parameters switching function and output type are summarized in Table 3.31.

With NS for AC, but also DC supply voltage, there are also versions with only two connections. The NS is connected in series with the load R_L and the supply voltage source. The switching functions remain the same.

3.4.1.7 Design and Size

The NS are one-piece units and usually have the following housings:

- *Cylindrical threaded sleeve* made of metal or plastic with diameters ranging from 5 mm to 36 mm and lengths well below 100 mm.
- *Smooth cylindrical sleeve* made of metal with diameter values in the range 3 mm to 34 mm and lengths under 80 mm. Due to the permanent miniaturization of the internal electronics, minimum lengths of about 25 mm to 30 mm have been achieved (Figs. 3.174 and 3.175).
- *Rectangular housing with square cross section* made of plastic or metal.
- *Rectangular housing with rectangular cross section* made of plastic or metal.

Table 3.31 Summary of the combinations of the parameters switching function and output type

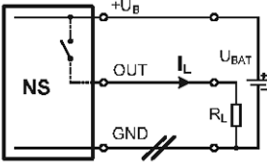


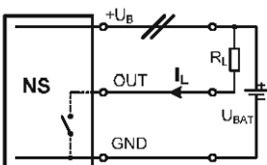
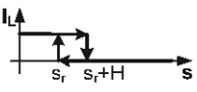
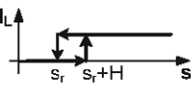
SWITCHING FUNCTION	Closer (NO)	Opener (NC)
<p>TYPE OF EXITION</p> <p>positive switching</p> 	 <p>Most used combination in Europe</p>	 <p>Second priority in Europe</p>
<p>negative switching</p> 	 <p>Combination used in Asia and America</p>	 <p>Rarely used</p>

Fig. 3.174 Miniaturised NS in metal housings: smooth or threaded with diameters of 3 mm to 5 mm (Factory photo: Balluff GmbH)

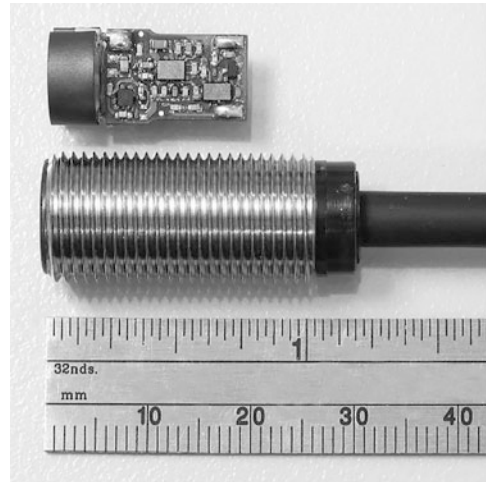


3.4.1.8 Characteristics of the NS

In addition to the main features, the following characteristics are characteristic of the industrial NS:

- Repeat accuracy (R).
- Rated voltages with the following definitions:
 - Rated operational voltage (U_e),
 - Rated insulation voltage (U_i) and
 - Voltage drop (U_d) across the output at nominal load.

Fig. 3.175 Metal threaded housing M12: extremely short (Factory photo: Balluff GmbH)



- Significant currents are among others
 - Rated operational current (I_c),
 - Residual current (I_r) of the disabled output and
 - *Idle current* (I_0) from the supply network without load (own power consumption).
- Response delay.
- Switching frequency (f).
- *Stand-by delay* (t_v) or also start-up time.
- Main protection functions:
 - the *short-circuit protection function*,
 - the *broken wire* and
 - the *reverse polarity protection function*.

The standard specifications or the usual up-to-date values of these parameters for the most commonly used metal thread designs M12 and M18 can be found in Table 3.32.

3.4.1.9 Special Designs and Their Applications

The NS usually contain a single switching output, which has a binary value. The switching output states are OFF/ON or HIGH/LOW according to the above mentioned product standard DIN 60947-5-2.

The constant increase in the complexity of automated processes has led to a continuous decentralization and relocation of some functions to the lowest sensor level. Today, the term “intelligent sensors” also applies to NS. Furthermore, some trends and realizations in the direction of increasing intelligence and complexity are presented.

3.4.1.10 NS with Several Switching Outputs

Increasing the number of switching outputs implicitly leads to an increase in application reliability or the information content of the LV. Representative designs are:

Table 3.32 Technical data of NS (@ is used to define the measurement conditions)

Characteristics (electrical or mechanical)	Standard values and value ranges	Current market values and value ranges
Rated operating distance (s_n)	<ul style="list-style-type: none"> • Inductive NS: 2 mm to 5 mm • Capacitive NS: 2 mm to 5 mm • Ultrasonic NS: 60 mm to 800 mm 	8 mm to 20 mm 1 mm to 15 mm 30 mm to 3500 mm
Detection range (s_d) of photoelectric NS	<ul style="list-style-type: none"> • Light sensor: Not specified • Retro-reflective photoelectric sensor not specified • Throughbeam photoelectric sensors: Not specified 	600 mm 7000 mm 20,000 mm
Hysteresis (H)	$\leq 0.2 s_r$	$\leq 0.1 s_r$
Repeat accuracy (R)	$\leq 0.1 s_r @ (23 \pm 5) ^\circ\text{C}$	$\leq 0.1 s_r @ (23 \pm 5) ^\circ\text{C}$
Rated operational voltage (U_e)	$\leq 300 \text{ VDC}$ or $\leq 250 \text{ VAC}$	10 to 30 VDC; 10 to 55 VDC to 250 VAC; 20 to 250 VAC/DC
Voltage drop (U_d)	$\leq 8 \text{ V}/10 \text{ V DC/AC @ 2 connections}$ $\leq 3.5 \text{ V DC @ 3/4 connections}$	$\leq 2.5 \text{ V}/6 \text{ V DC/AC @ 2 connections}$ $\leq 2 \text{ V DC @ 3/4 connections}$
Rated operational current (I_e)	50 mA DC or 200 mA _{eff} AC	
Lowest operating current (I_m)	$\leq 5 \text{ mA DC/AC @ 2 connections}$	$\leq 3 \text{ mA DC/AC @ 2 connections}$
Residual current (I_r)	$\leq 0.5 \text{ mA DC @ 3/4 connections}$	$\leq 0.1 \text{ mA DC @ 3/4 connections}$
Idle current (I_0)	$\leq 5 \text{ mA DC @ 3/4 connections}$	$\leq 5 \text{ mA DC @ 3/4 connections}$
Response delay	$\leq 1 \text{ ms}$	$\leq 1 \text{ ms}$
Switching frequency (f)	$\leq 1000 \text{ Hz (inductive NS)}$	$\leq 5000 \text{ Hz (inductive NS)}$
Standby delay (t_v)	$\leq 300 \text{ ms}$	$\leq 10 \text{ ms to } 20 \text{ ms}$
Duration of the false signal during t_v	$\leq 2 \text{ ms}$	$\leq 1 \text{ ms}$
Ambient temperature range	$-25 ^\circ\text{C}$ to $+70 ^\circ\text{C}$ (standard values) $-5 ^\circ\text{C}$ to $+55 ^\circ\text{C}$ (photoelectric NS)	$-40 ^\circ\text{C}$ to $+125 ^\circ\text{C}$
Protection class	Min. IP65, IP54 (photoelectric NS)	IP67, IP68 (x bar), IP69K

- (a) *Inductive and optoelectronic NS with antivalent switching outputs.* The implementation in the NS of both switching element functions: normally open and normally closed and the simultaneous supply of the corresponding switching signals up to signal processing is a common method for system monitoring. The received signals must *always be anti-valent*, independent of the sensor actuation. Violation of the anti-valence indicates a fault in the sensor or in the supply lines.
- (b) *Inductive NS with 3 switching outputs.* The switching outputs of the sensor have the same switching function (normally open contact), are freely programmable within a defined detection range and thus enable the division of this range into four sub-ranges.
- (c) *Inductive NS with combined switching outputs.* Basically, this is a version with two NO switching outputs DA2 and DA3 and one NC switching output DA1.
- (d) *Inductive NS with 4-bit digital interface.* The compact sensor comprises complex internal electronics with three output functions in an M12 metal threaded sleeve with a length of 75 mm:
 1. 4-bit parallel port: outputs D0 to D3. Accordingly, the very wide detection range from 1 mm to 5 mm is divided into 14 equal intervals. The target presence in an interval is output by the corresponding bit pattern. Bit patterns 0000 and 1111 are output when the target is outside the detection range—*out-of-range* function.
 2. Analog output A.
 3. Temperature output T.

This allows distance changes to be measured analogously and recorded digitally. The ambient temperature is monitored and the LV supplies a DC voltage that changes extremely linearly by -9 mV/K over a total temperature range of -10 °C to $+70$ °C.

3.4.1.11 NS with Diagnostic Information

The decentralisation of data processing is achieved by transferring and increasing intelligence at the lowest level, the sensor and actuator level. This requires additional functions for parameterization and diagnosis of the systems from a central location. Safety aspects are dealt with in Chap. 17.

3.4.1.12 Applications of the NS

The simpler the function and commissioning of the proximity sensors, the more widespread their use and application. Every year millions of proximity sensors are produced and used in industrial automation. The reason for this is the particularly good price–performance ratio of these sensors. For unit prices between 5 € and 200 €, you can find common sensors for applications with low demands up to problem solvers for applications with high sensor intelligence requirements.

The NS are *compact, precise, fast, reliable* and *wear-free*. In appropriate designs, they are also resistant to aggressive cleaning agents in extremely harsh environments and can be used in high temperature fluctuations, strong vibrations and pressures.

The detection of metallic or non-metallic objects regardless of their size, i.e. even very small and filigree, is a major task in factory automation and is mainly performed with the NS described above.

Numerous applications can be found mainly in the following areas of use:

- (a) *Machine tools*. Advances in automation engineering and sensor technology mean that the performance and flexibility of machine tools are constantly being improved. NS play an important role and perform the following classic functions in addition to standard position sensing:
 - Clamping path monitoring and recording,
 - position switching with safety switching points,
 - quality control on workpieces, completeness control,
 - level monitoring of various media (powdery, pasty or liquid) in contact with the product,
 - non-contact level detection in transparent/opaque containers,
 - detection of different diameters,
 - identification or recognition of certain features of a workpiece/tool,
 - breakage control of tools, such as drills, taps,
 - groove or thread recognition, thread control and
 - small parts detection, placement control.
- (b) *Robotics and material flow, assembly and handling*. NS monitor, control, measure and automate processes and conditions in the plastics industry, in textile machines, in wood and metal processing. Important applications are:
 - Counting objects,
 - passage, packaging and closure cap inspection,
 - part positioning and sorting,
 - query of packaging contents,
 - query the size and contents of containers,
 - query the stack height,
 - tracking of a lifting platform or forklift in high-bay warehouses,
 - query a read mark and
 - final check, cover.
- (c) *Automotive industry*
 - Speed measurement,
 - rain sensor,
 - seat occupancy detection and
 - parking sensors.
- (d) *Fluid power*

- End position detection in hydraulic cylinders for positioning of lifting units, feed components or clamping elements;
 - piston position monitoring in pneumatic actuators such as pneumatic cylinders or grippers and
 - monitoring the valve positions.
- (e) *Power generation.* Due to rising oil prices and limited resources, energy generation is more important than ever. Both conventional power plants for conventional energies as well as wind, solar and hydroelectric power plants for renewable energies require ever better LV with long service life and high robustness and reliability. Some examples are:
- Speed and direction of rotation monitoring of the rotor in wind turbines,
 - level monitoring and
 - sun position sequences of mirrors and solar modules.

3.4.2 Object Detection and Distance Measurement with Ultrasound

Ultrasounds are acoustic waves above 20 kHz, which—in contrast to electromagnetic waves—can only propagate in matter (Sect. 3.1.3 and Fig. 3.176). In ultrasonic sensors for industrial applications, the carrier medium is mainly air. If ultrasonic waves hit solid bodies, the sound is reflected. The sensor technology makes use of this principle. The sensor receives *reflected sound waves* as an *echo*, determines the *distance* from them and converts this into an output signal.

Industrial applications work with high-frequency ultrasound from approximately 80 kHz. At these high frequencies, *bundled sound lobes* are formed, which are reflected by objects to a greater or lesser extent—depending on the surface properties, shape and orientation (Sect. 3.1.3.3 and Figs. 3.68). Low-frequency ultrasound, on the other hand, spreads out spherically in all directions and is therefore unsuitable for industrial applications.

At lower frequencies, the distance and the sound beam is wider. This is necessary and desired, for example, for collision detection applications. With increasing frequency, the measurable distance becomes smaller, caused by the damping of the air. The decreasing opening of the sound beam is also noticeable, caused by the design of the transducer. Here, the dead band is also considerably smaller and ideal for *level control* in narrow environments. The following Fig. 3.177 show examples of ultrasonic beams at different frequencies.

3.4.2.1 Speed of Sound in Air

Since most ultrasonic sensors use the *echo transit time method*, it should be noted that the speed of sound can be influenced by some environmental conditions, for example

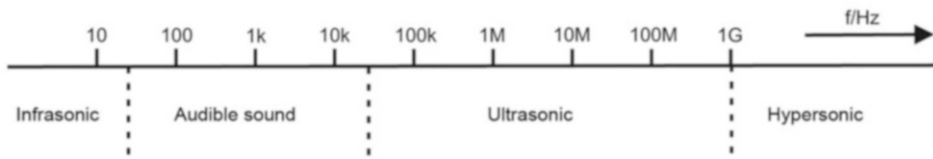


Fig. 3.176 Spectral distribution of sound waves

- the temperature,
- the air pressure or
- the relative humidity.

The following equation applies to the influence of temperature:

$$c = c_0(1 + T/273)^{1/2}$$

with

c_0 Sound velocity at $T = 0^\circ\text{C}$ (331.6 m/s)

T Temperature in $^\circ\text{C}$.

The graph shows the strong dependence of the speed of sound on temperature (Fig. 3.178).

The influence of relative humidity is rather small and does not require compensation measures for accurate measurements (Fig. 3.179).

3.4.3 Object Detection with Radar

The term *radar* stands for **R**adio **D**etection **A**nd **R**anging. With a radar, speeds, distances, shape and size can be determined. The applications are distinguished between *pulse radar* and *CW* (continuous wave) radar.

3.4.3.1 Impulse Radar

With pulse radar, oscillation packets are emitted at defined distances, which are reflected by the objects in front of them. The distance can be determined by *measuring the time of flight* and the amplitude determines the amount of reflected energy and thus the size R of the object.

$$R = \frac{c}{2} \cdot \tau$$

with τ : transit time and c : speed of light.

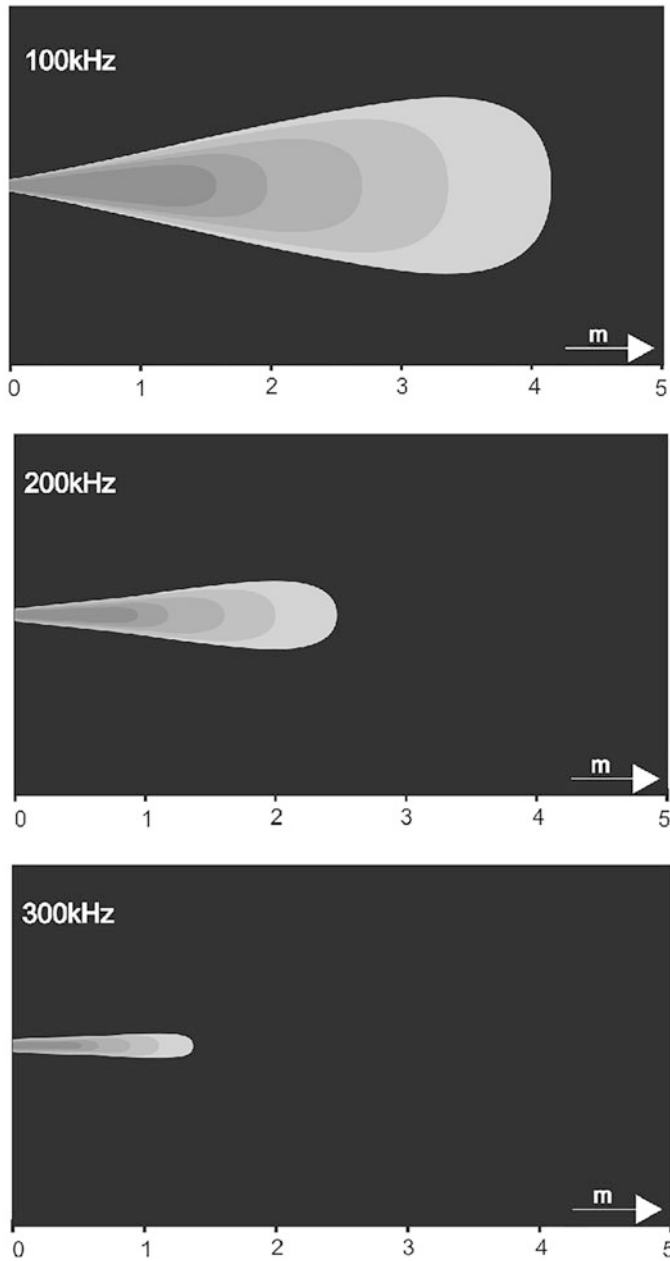


Fig. 3.177 Ultrasonic lobe at different frequencies

The resolution of the measurement is determined by the pulse width. The maximum measuring range is defined by the distance between the vibration packets.

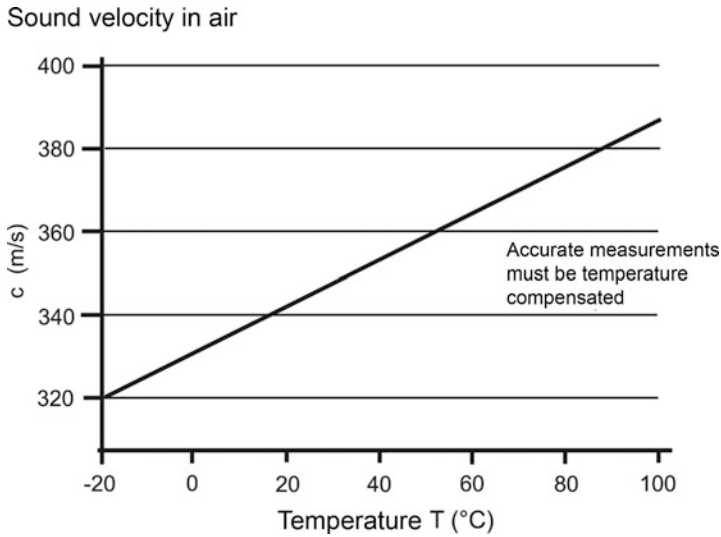


Fig. 3.178 Dependence of sound velocity on temperature

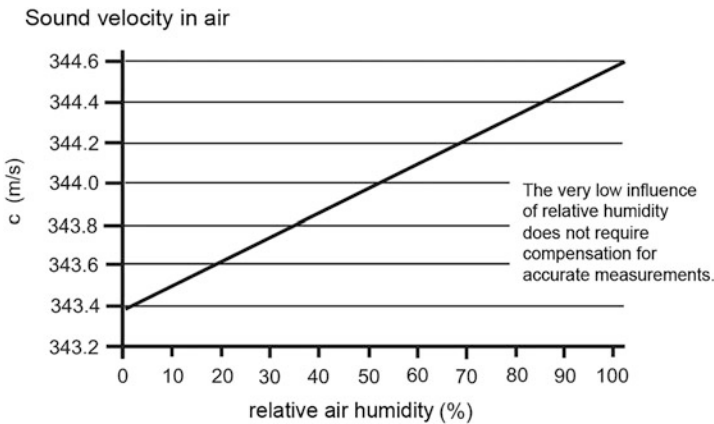


Fig. 3.179 Dependence of the speed of sound on the humidity

3.4.3.2 CW Radar

The CW radar consists of a transmitter and receiver, which operate continuously. To determine a running time, the transmitter is frequency modulated and this signal is received with a time delay. This method, called *chirp*, refers to a modulation period.

If the object moves during this measurement, the received frequency changes compared to the transmitted one. This is due to the *Doppler effect* (Sect. 2.20). This frequency shift can be used to determine not only the distance but also the speed.

3.4.3.3 Application

The determination of distance and speed by radar covers a wide range. Typical areas of application are:

Air Traffic Control:	Determination of flight altitude and speed.
Motor vehicles:	Distance alerts—in contrast to ultrasound, these are less sensitive to weather influences.
Meteorology:	Determination of height, velocity and water content of clouds.
Building services engineering:	Control of doors and safety shut-offs (in contrast to the motion detector, the temperature of the object is not important here)

Radar-based motion detectors already respond to speeds as low as 0.1 m/s.

Depending on the antenna shape, the beam angle of the measuring arrangement can be varied. By rotating the transmitter and receiver or by calculating the values of several sensors, the distance can also be determined in 2 or 3 dimensions (Sect. 3.5).

3.4.4 Pyroelectric Sensors for Motion and Presence Detection

3.4.4.1 Material Properties of Sputtered PZT Films

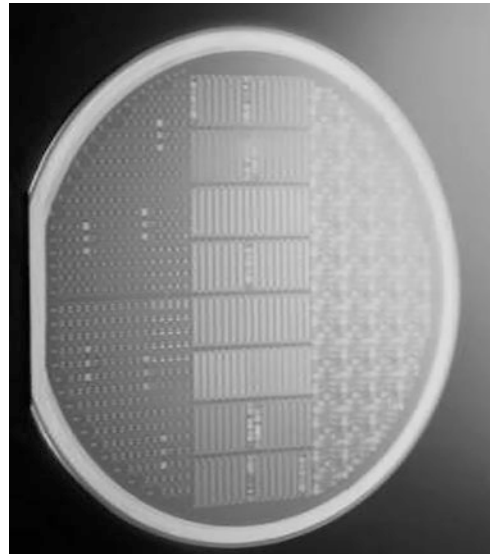
A technology for pyroelectric sensors is based on *sputtered PZT* (lead, zirconium, titanate) films, which are applied to a silicon wafer with a suitable electrode at high temperatures (>500 °C), as shown in Fig. 3.180. The PZT films are deposited in such a way that they are polarized during the sputtering process and therefore do not require any further polarization steps after discharge from the sputtering system. The PZT films form on the wafer in a perovskite structure (111) and typically have a pyroelectric coefficient of $2 \cdot 10^{-4}$ C/mK.

The sputtered pyroelectric layers have a film thickness between 0.4 µm and 1.5 µm and can therefore be structured and processed with conventional and established manufacturing processes from the semiconductor/MEMs industry. Due to the already existing infrastructure for such processes, high volumes can be produced at low cost.

The Curie temperature of the sputtered PZT films is about 550 °C. Since the films are already polarized after the sputtering process, this *polarization* is maintained even when exposed to high temperatures (up to 450 °C). In conventional pyroelectric sensors (ceramic PZT or Li₃TaO₄), the polarization of the material is performed by an external polarization step in an electric field. The polarization of the externally polarized pyroelectric sensors is strongly temperature dependent. The polarization (and thus the basis for the IR sensor) is lost when these films are exposed to temperatures of >150 °C. Therefore, conventional pyroelectric sensors cannot be processed by fully automated assembly and soldering machines (SMD technology). Temperatures of up to 400 °C occur there.

The sputtered PZT films can withstand temperatures up to 450 °C without any damage and can therefore be integrated into a fully automatic packaging process. An example of an

Fig. 3.180 6-inch silicon wafer with infrared sensor chips, based on sputtered PZT films. On the wafer, one can see single channel sensors, line sensors for infrared spectroscopy and pyroelectric arrays



SMD compatible sensor is shown in Fig. 3.181. This production-related possibility is particularly advantageous for very high quantities.

3.4.4.2 Effect on Electronics

As described above, the sputtered PZT films are characterized by a very low film thickness and a relatively high permittivity E_r of 270. Therefore, sensor chips based on sputtered PZT have a relatively high capacity. It is therefore advantageous to read out the charges resulting from heat radiation absorption on the contact surface of the sensor chip using a transimpedance converter instead of using the conventional source-follower circuit.

Compared to ceramic PZT sensors, sputtered thin-film sensors have a much *lower thermal* mass, since the film thickness is reduced by a factor of 50 to 100. This leads to a different frequency response, which must be taken into account when integrating the sensors into a system. The maximum *responsivity* (V/W) of ceramic pyroelectric sensors occurs at 0.2 Hz to 0.5 Hz and decreases steadily at higher frequencies. The maximum responsivity (V/W) for sputtered PZT films occurs at frequencies between 5 Hz and 10 Hz and can be extended up to the kHz range by a clever choice of electronic components.

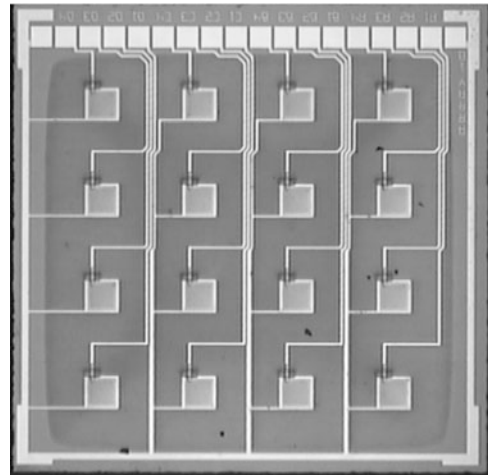
The unique combination of MEMs process technology with the sputtered and self-polarized PZT films opens up new fields of application that could not be realized with conventional ceramic sensors so far.

Figure 3.182 shows pyrotechnical sensor arrays. The arrays are structured using standard MEMs processes. The sensor generates 16 individual signals and has a chip size of 2 mm × 2 mm. As shown in Fig. 3.182, pyroelectric sensor arrays (e.g. 4 × 4) can be produced, which provide the user with 16 bits of information. This is a lot compared to the typical single-channel sensors where only 1 bit of information is available. The use of



Fig. 3.181 SMD-compatible pyroelectric sensor suitable for fully automatic assembly and soldering systems (Factory photo: Pyreos.com)

Fig. 3.182 Pyroelectric sensor arrays based on sputtered PZT films (Factory photo: Pyreos.com)



pyroelectric arrays in *motion-presence detection* makes it possible to capture more information than was previously possible with conventional sensors.

Pyroelectric sensor arrays enable the detection of the *direction of movement*, the counting/tracking of persons and the exact presence of a person in the room. This additional information can be determined very accurately by evaluating the 16 individual sensor signals. An example of such an application in people counting can be found on YouTube at <http://www.youtube.com/watch?v=sqw34TyHwVw>.

For applications of pyroelectric sensor arrays for motion detection or presence detection, the different frequency response of thin-film sensors compared to conventional sensors must be taken into account in the optical design to achieve optimal results.

However, many applications, such as *traffic counting* or *gesture recognition*, require infrared sensors that are fast and have very good sensitivity at higher frequencies. For the above application examples, the infrared signals generated (by passing cars or by a hand movement) are in the frequency range of 8 Hz to 80 Hz. This fits exactly into the optimal

frequency range of sensors based on PZT films. The following link <http://www.youtube.com/watch?v=0YpI3J2IThA> shows an example of gesture recognition, which is characterized by very high accuracy and very low power consumption, as this application is based on passive components only.

The *additional information content* that can be generated by sensor arrays can be very useful for motion detection or application detection. Data analysis of the generated signals will become increasingly important in the future.

3.4.5 Object Detection with Laser Scanner

Laser scanners are distance measurement systems based on the principle of *time-of-flight measurement* (Sect. 5.4). The technical core of the system is a time-of-flight measurement system whose beam is deflected by a system of mirrors. Thereby, distances can be measured in an aperture angle between 180° and 270° . The working radius can be between a few meters and several 10 m, depending on the model.

Figure 3.183 shows the relationships between the scan range and a measured value curve. The scanner outputs the distance measurement values in chronological order. Due to the rotational movement of the mirror system, these refer to a central point and must therefore be plotted in polar coordinates. For this reason, the bizarre corners also occur in the distance function, since the distance must always be plotted radially. A typical laser scanner is shown in Fig. 3.184.

3.4.5.1 Application

Laser scanners are mainly used in the field of *security technology*. This is done both mobile and stationary.

In autonomous transport systems, they serve the vehicle to detect *collisions* with the environment. This makes it possible to take evasive action or to stop.

The stationary application usually secures systems against *approach*, which also in these cases preferably leads to shutdown.

3.4.6 Sensors for Automatic Identification (Auto-Ident)

3.4.6.1 Overview

The following three sensor technologies enable *automatic identification* of objects:

- Barcode laser scanner,
- Auto-Ident cameras and
- RFID readers.

In the following, these different technologies are presented and the applications are explained using examples (Fig. 3.185). While bar code scanners have been used for a long

Fig. 3.183 Functional diagram of a laser scanner

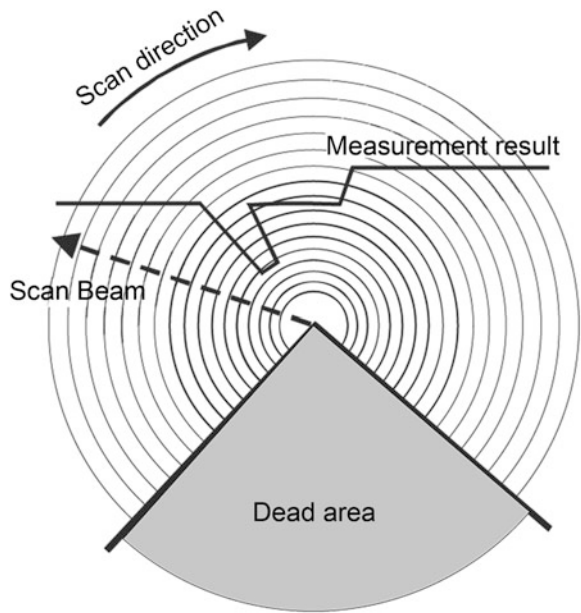


Fig. 3.184 S3000 laser scanner (Factory photo: SICK)



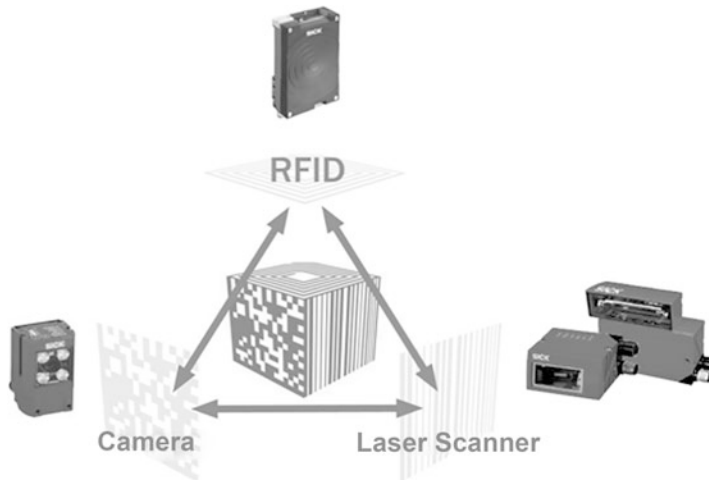


Fig. 3.185 Three technologies for automatic identification (Factory photo: Sick AG)

time, newer technologies are the compact industrial cameras with integrated evaluation and, above all, RFID readers. The different advantages of the individual technologies have their place in special applications.

3.4.6.2 Barcode Scanner

Classification of Barcode Readers

There are different types of devices for bar code identification:

- *Manual* reading pens (*pen*),
- *handheld* barcode readers (handheld barcode readers; handheld barcode scanners) and
- *fixed mount* barcode readers on the conveyor line.

In each reader, the optical received signals are converted into an electrical signal in an *optoelectronic component*. A distinction is made:

- Readers based on *CCD* or *CMOS line sensors* and
- *barcode scanner with laser scanning* (flying spot scanner).

Today's bar code scanners generally use a semiconductor laser diode in the red spectral range with a transmission power of about 1 mW to 10 mW to generate the laser beam. A distinction is made here between two construction principles, which differ in the receiving principle:

- Scanner with *omnidirectional receiver* and
- scanner based on the *autocollimation principle*.

Devices with omnidirectional receivers require less installation space in the housing and are very inexpensive and can be set up without complicated adjustment.

Although scanners based on the autocollimation principle have disadvantages in this respect, they offer the following decisive advantages over the omnidirectional receiver

- Significantly *better SNR* (signal to noise ratio); therefore longer reading distances possible and
- lower *sensitivity to ambient or extraneous light*.

Applications of Permanently Mounted Barcode Scanners

The main applications are in logistics and factory automation. Figure 3.186 shows a permanently mounted scanner on a conveyor line for automatic identification of objects transported past on a pallet (logistics automation). In factory automation, the devices are used, for example, in automobile production lines to identify car bodies in the manufacturing process. Another well-known example is barcode scanning at the supermarket checkout.

Structure and Function Modules of a Barcode Scanner

Figure 3.187 shows the essential function blocks of a scanner for identifying a bar code:

- A *semiconductor laser diode* (1) generates the laser beam, which is directed onto the polygon mirror wheel via a deflection mirror (3). The optional focusing device (1) can focus the laser beam to different distances.
- The *rotating mirror wheel* (2) deflects the beam line by line. The scanning frequency, also known as the scanning frequency, is in the range of 200 Hz to 1200 Hz for common devices.
- The *deflected laser beam* (5) emerges at an opening angle (4) and scans the bar code (7) in the reading plane (6). Due to the deflection, the laser beam becomes a “reading beam”.
- Units with *oscillating mirrors* (9) additionally deflect the laser beam perpendicular to the line deflection. This allows goods to be read even when the machine is at a standstill, for example. Common frequencies for this are in the range of 0.1 Hz to about 5 Hz.
- The light remitted by the barcode (7) is again transmitted to the photodetector via the mirror wheel and an optical filter (10). Here the *optical reception signal* is converted into an *electrical signal*. An amplifier (12) brings the low received signal into a usable or evaluable range.
- The *binarization stage* (13) converts the analog, amplified received signal into a digital one. Often, only a bivalent binary signal (8) is converted.
- This is entered into the *evaluation*. The binary signal is *run length coded* (the duration of the 1/0 phases is determined) and the result is stored in a table in the memory. This memory area is then evaluated by the decoder and compared with the possible code characters of a barcode type. The result is the *decoded character string* of the barcode



Fig. 3.186 Fixed-mounted bar code scanner for identifying moving objects (Factory photo: Sick AG)

(read result). This can now be visualized via the (optional) display in the device and output via the device interfaces.

The following formulas and relationships apply:

- (4): Opening angle $\alpha_V = 720 \text{ degrees}/n$ with $n =$ number of surfaces of the polygon mirror wheel.
- (11): The reception signal at the photodetector is proportional to $1/r^2$ for longer distances; with r : reading distance. For small reading distances, there are aperture limitations in the receiving beam path; in this case, the exact device design must be observed.

Nominal Reading Distance and Depth of Field of a Bar Code Scanner

Due to the diffraction effects of the laser light, the *beam caustic results* in the semiconductor laser, which describes the diameter of the beam as a function of the distance (Fig. 3.188). The laser beam has the smallest diameter at the focus distance. There the reading capability for high-resolution bar codes is also best. The focus distance corresponds to the frequently specified *nominal reading distance* for bar code reading.

At and near the focus distance, the light intensity of the laser beam is approximately Gaussian distributed (Fig. 3.189).

The scanning process of the barcode by the laser beam can be described mathematically as a *convolution of the intensity profile* in the beam with the light/dark areas of the barcode. The result is a *low-pass filtering* of the ideal bar code signal.

Due to beam acoustics and low-pass filtering during scanning, the readability of a bar code outside the nominal reading distance is limited to a *distance range*. This range of reading capability in front of and behind the focus distance is called the *depth of field*

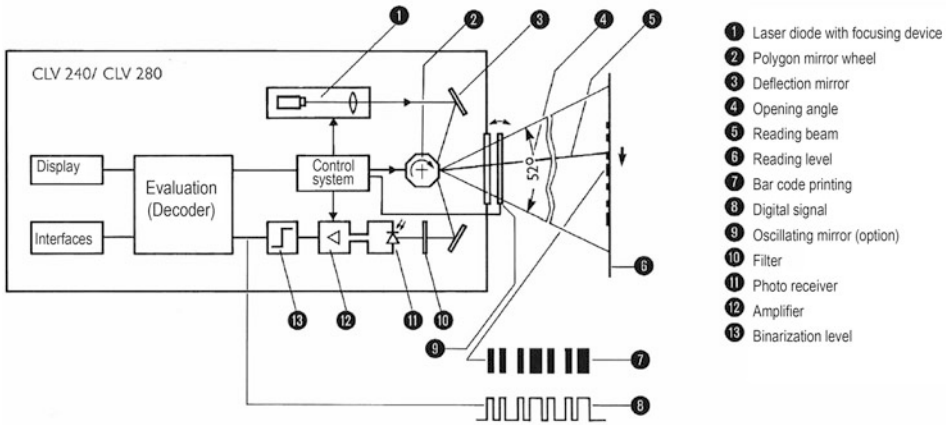


Fig. 3.187 How a bar code scanner works

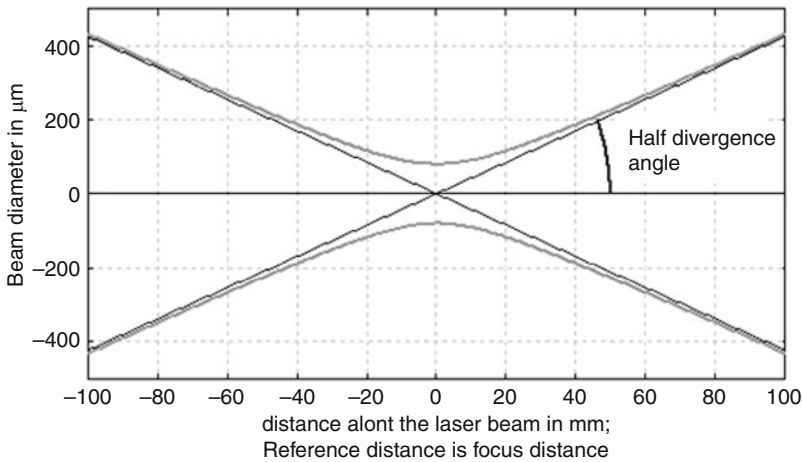


Fig. 3.188 Example of laser beam acoustics (Source: Sick AG)

(DOF). The DOF for a given focus distance is directly dependent on the *code print resolution*. For a code with high print resolution (narrow bar width), the DOF is lower.

Since the laser beam can be focused to smaller diameters at short reading distances than at large ones, the basic reading capability for high-density codes is also only available at shorter reading distances. The principle is illustrated in Fig. 3.190.

On devices with integrated focus adjustment, these reading ranges can be switched dynamically. Using the example of an industrial bar code scanner, this results in the following relationships for the readability of codes with different resolutions at different distances and with different DOF (Fig. 3.191; excerpt from an operating manual).

Fig. 3.189 Gaussian distribution of light intensity near the focus (Source: Wikipedia)

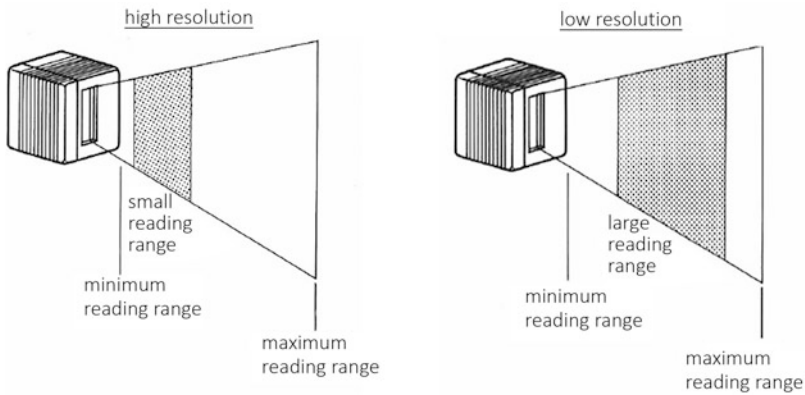
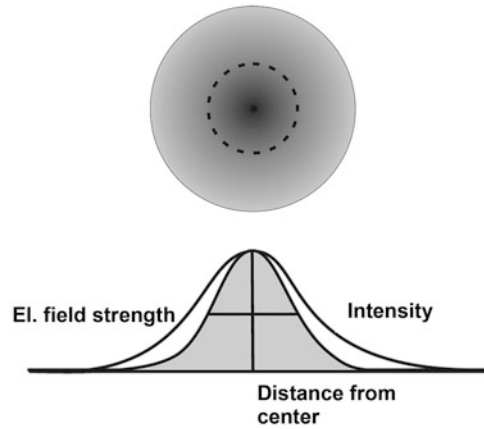
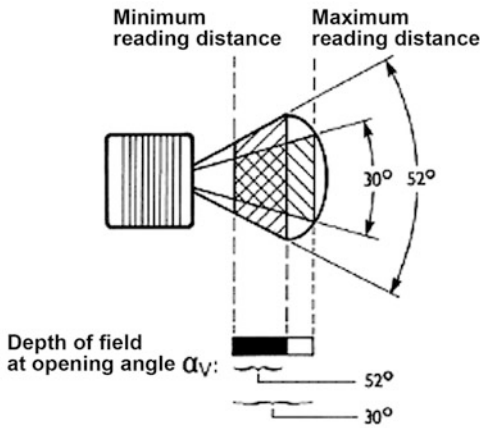
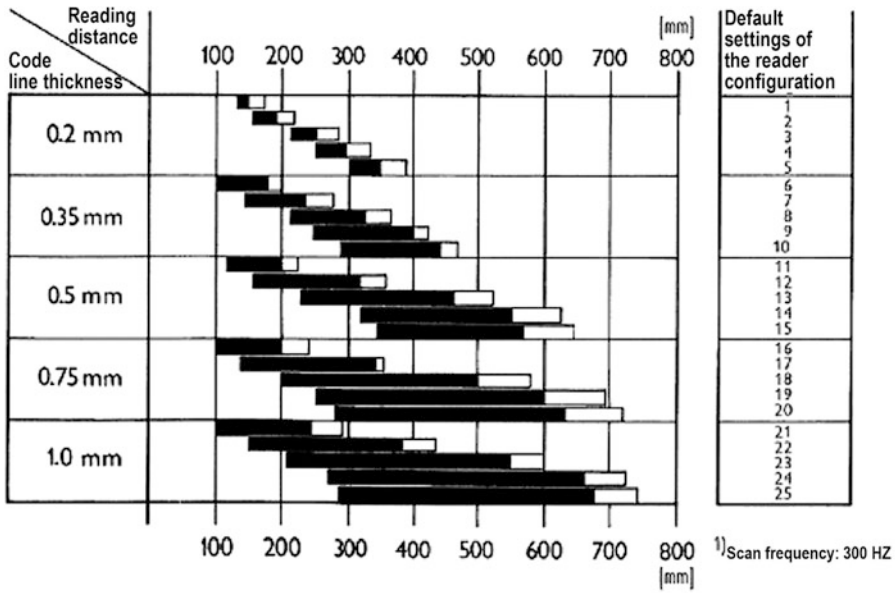


Fig. 3.190 Relationship between resolution and reading range

Important types and applications of permanently mounted bar code scanners are described below.

The reading situations are distinguished according to the alignment of the barcode on the object. In the simplest case, the bar code is arranged in “*ladder form*” (Fig. 3.192). The advantage of this arrangement according to Fig. 3.192 is that the reading result can be easily assigned to the object, the height placement of the bar code is not critical and can be driven with very short object gaps.

Figure 3.193 shows a *grid scanner* for “fence-shaped” codes rotated by 90 degrees. In contrast to the arrangement according to Fig. 3.192, the gaps between the objects must be larger to ensure a clear assignment. The barcode placement must be in a relatively narrow band.



Reading conditions:	
Test code	Code 39
Print ratio	1:3
Code Print Contrast	> 90%
Successful readings ¹⁾	> 75%
Ambient light	< 2,000 lx
Angle of rotation (Skew)	15°
Azimuth angle (Tilt)	0°
¹⁾ of 100 scans each (operating mode: percentage evaluation)	

Fig. 3.191 Overview of the reading capability of codes

The advantage of this arrangement according to Fig. 3.193 is that the reading result can be assigned to the object without any problems, the height placement of the bar code is not critical and can be driven with very short object gaps.

Figure 3.194 shows a grid scanner for “fence-shaped” codes rotated by 90 degrees. In contrast to the arrangement according to Fig. 3.193, the gaps between the objects must be larger to ensure a clear assignment. The barcode placement must be in a relatively narrow band.

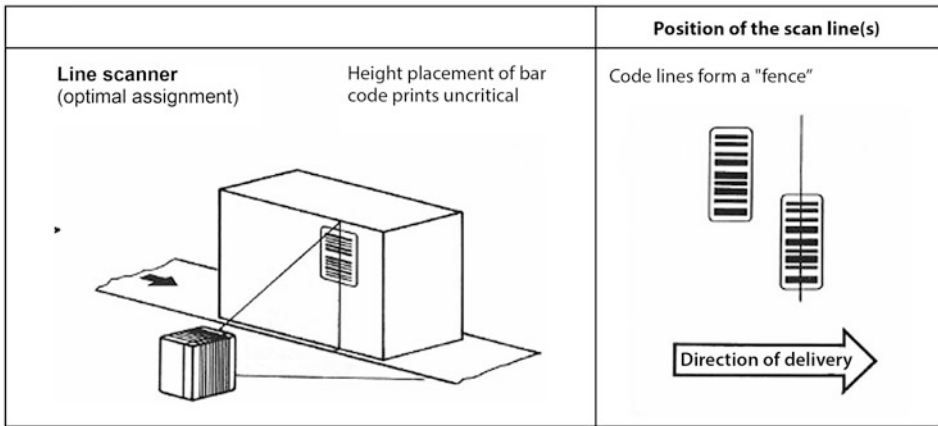


Fig. 3.192 Functionality of a line scanner

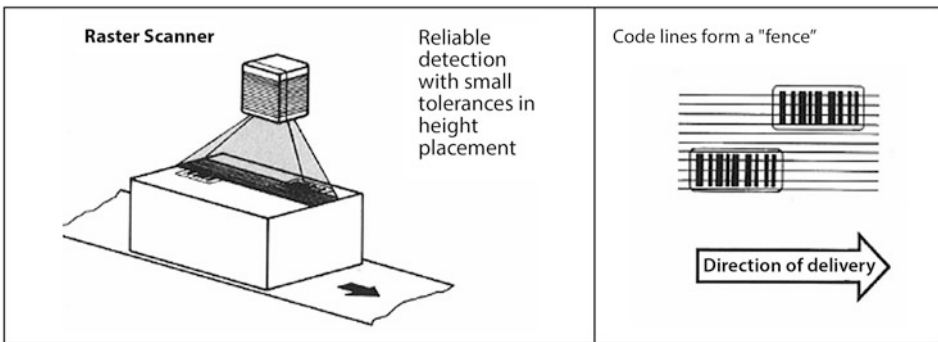


Fig. 3.193 Using a raster scanner for "fence-shaped" code lines

Devices with oscillating mirrors deflect the laser beam perpendicular to the line deflection before it exits the device. This allows, for example, goods to be read even when stationary. Common oscillation frequencies for this purpose are in the range from 0.1 Hz to about 5 Hz (Fig. 3.195).

Important Terms for Optimal Scanner Alignment

The angles *skew*, *tilt* and *pitch* are often used in connection with barcode reading. Figure 3.196 illustrates their definitions.

Optimum reading results are achieved with the following measures:

- The *skew angle* should be about 15 degrees (close to 0 degrees, there is a reflection and therefore non-readings).

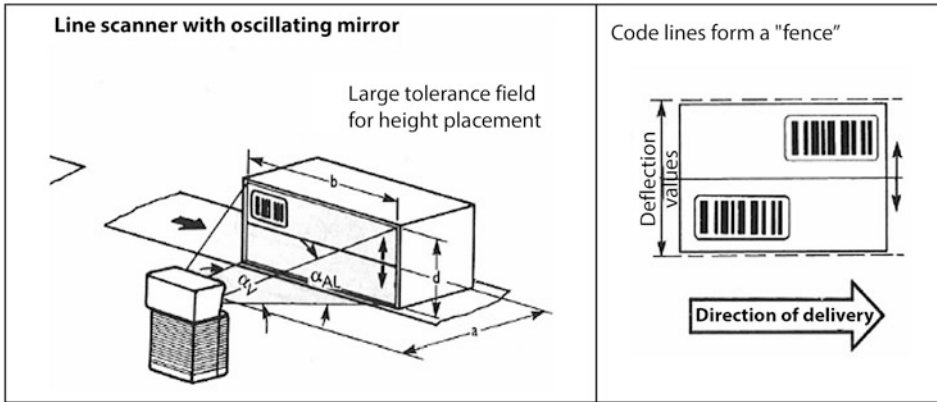


Fig. 3.194 Line scanner with oscillating mirror

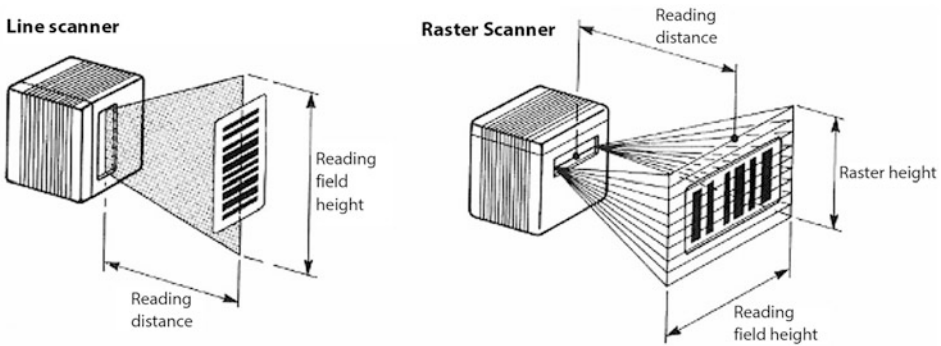


Fig. 3.195 Functional principle of line scanner and raster or oscillating mirror scanner

- For devices without code reconstruction, the *tilt angle* must be so small that in a sufficient number of scans, all code elements are overpainted. Devices with code reconstruction allow a much larger tilt angle.
- The *pitch angle* should be <45 degrees for best results.

3.4.6.3 Auto-Ident Cameras

The bar code scanners described in Sect. 3.4.6.2 are suitable for identifying conventional bar codes. If, however, so-called *matrix codes* (also 2D codes) are to be recognized or if character recognition (OCR) is required, *camera systems* are required depending on the system (Fig. 3.197).

Table 3.33 compares the advantages for the application areas of the camera systems and the laser scanner.

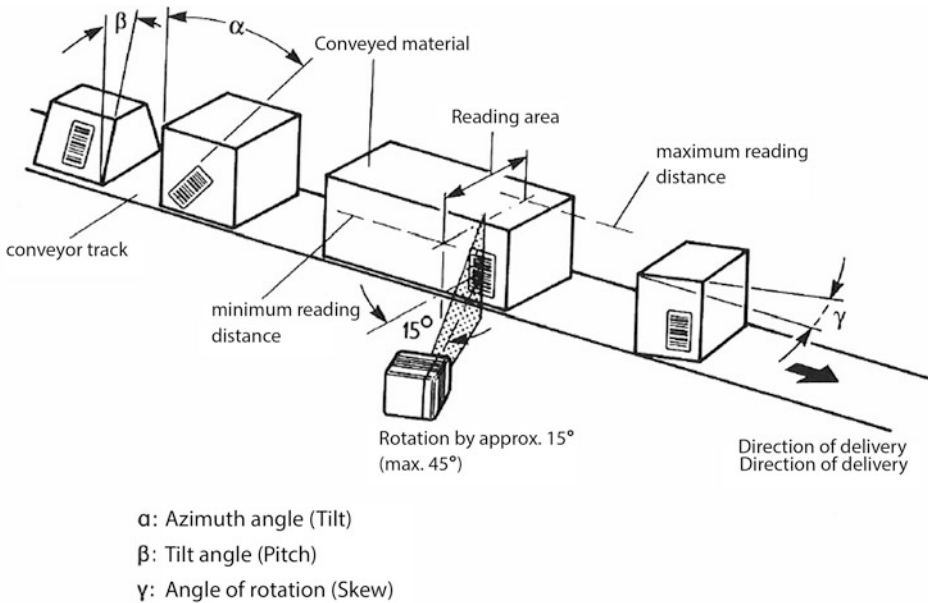


Fig. 3.196 Definition of the angles for bar code reading

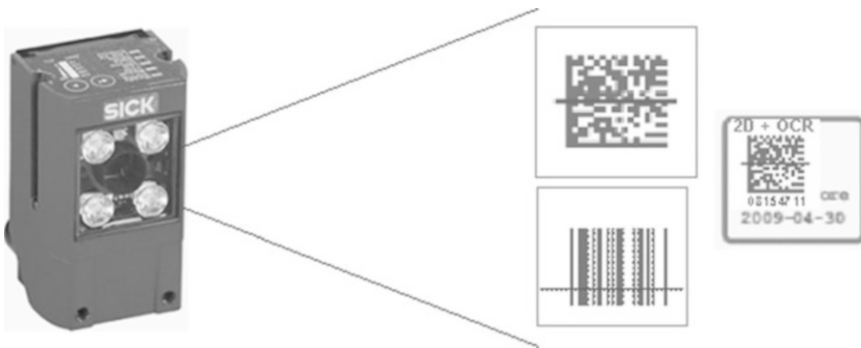


Fig. 3.197 Compact camera Lector 620 with integrated illumination for reading bar codes, matrix codes and OCR (Factory photo: Sick AG)

3.4.6.4 Construction of Auto-Ident Cameras

The block diagram in Fig. 3.198 shows the essential components of an Auto-Ident camera:

- The image sensor captures the image. The scene is illuminated by the integrated illumination flash.
- The data from the image sensor is sent to the *FPGA* (field programmable gate array) and is *pre-processed* there. This relieves the central processing unit of the computer (CPU)

Table 3.33 Comparison of the advantages as well as the application areas of camera systems and laser scanners

Camera systems	Laser Scanner
Reads 1D/2D codes and OCR	Reads 1D codes only
Reads directly marked codes (DPM: Direct part marked)	Larger reading height (larger reading field) and greater depth of field (laser beam property)
Better readability of destroyed codes or low contrast	Implicit, local, monochromatic illumination with high immunity to ambient light
Mechanical robustness, i.e. no motors or moving mechanical parts	

of computationally intensive preprocessing operations (e.g. filter operations or binarization).

- The *CPU* receives the pre-processed data from the FPGA and *stores* the image data in *RAM*. The *recognition algorithms* are executed. Analogous to the barcode scanner, a character string with the reading result is generated.
- The results data is *output* via the *interfaces*. In this case, they are available: Ethernet, serial (RS232/422) and CAN.
- Optionally, *image data* can also be output. Because of the high data rate required for this, the Ethernet interface is a good choice.

Similar to the situation with barcode scanners, the reading ranges depend on the code print resolution and the reading distance. With a *matrix camera*, *rectangular reading fields* result. Devices with integrated focus adjustment can adapt these reading areas flexibly to the conditions. Table 3.34 shows how the reading distance influences the resolution.

3.4.6.5 Important Fields of Application for Auto-Ident Cameras

The advantages of Auto-Ident cameras have already been shown in Table 3.33. This means that 1D and 2D codes can be read and the writing recognised (OCR capability). Furthermore, direct part marked (*DPM*) codes can be read. With *DPM*, the codes are not printed on a paper label, but directly on the parts to be marked. Laser marking, inkjet printing or direct dot peening is often used here. Such codes are more durable than paper labels.

In the following, some important examples for applications of Auto-Ident cameras are presented (Figs. 3.199, 3.200 and 3.201).

3.4.6.6 RFID Systems and Readers

RFID is the abbreviation for *radio frequency identification*. An RFID system consists of a *data carrier*, which is located on the object to be identified and contains an identifying code, and a *reader* to read this code. RFID reader and RFID transponder with antenna form a simple RFID system (Fig. 3.202).

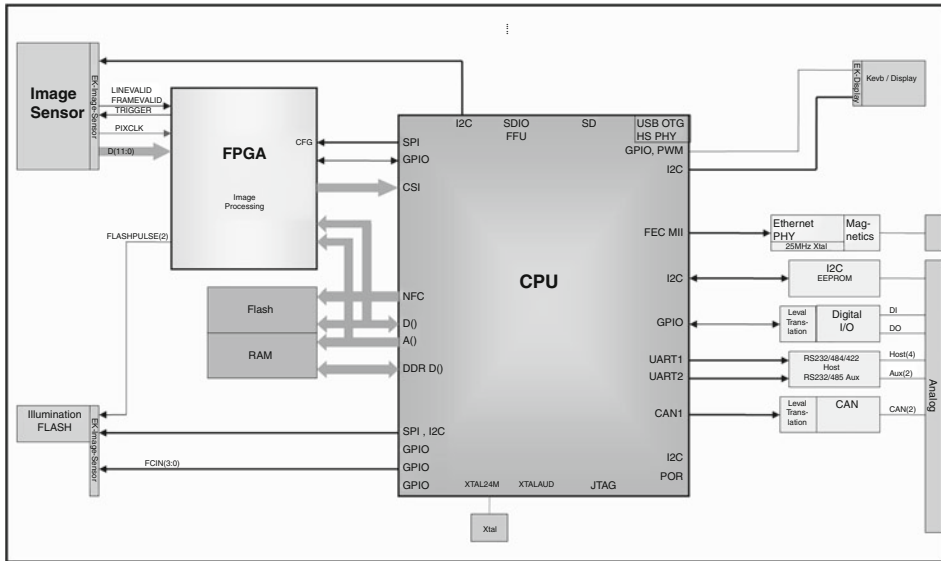


Fig. 3.198 Block diagram of an Auto-Ident camera (Factory photo: Sick AG)

Table 3.34 Example of possible read distances of the Lector 620 Auto-Ident camera (Source: Sick AG). Maximum reading distance with minimum resolution

Maximum reading distance	Minimum resolution
50 mm	0.10 mm
80 mm	0.15 mm
110 mm	0.20 mm
135 mm	0.25 mm
165 mm	0.30 mm
280 mm	0.50 mm
430 mm	0.75 mm

The information is stored in the data carrier, the so-called *transponder* (RFID transponder, often also called *RFID tag*). The reader activates the tag and reads information. In contrast to the optical methods described, the *reader* can also *write to* the tag. This makes a fundamental difference to optoelectronic identification.

Another important difference is that with RFID in the tag, the *entire product data* can be *kept on the product* and *updated* if necessary (modular, local), whereas optoelectronic identification methods usually work according to the *license plate method*: The information in the read code is used as a key to a (central) database entry, i.e. local, comprehensive information storage on the object is not possible here. Data must be changed centrally by updating the database entries.

Categorization of the Tags

One distinguishes the tags according to the type of energy supply in:

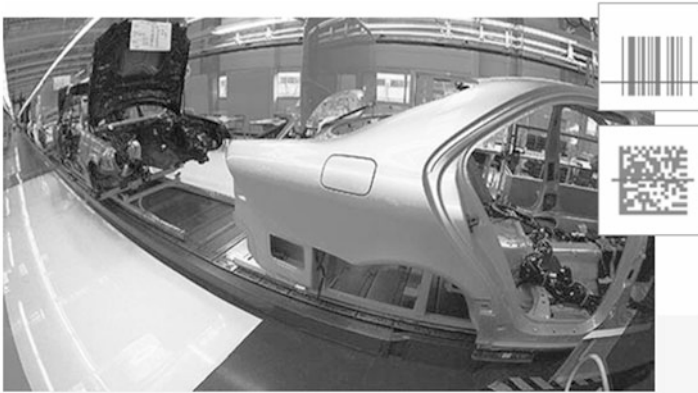


Fig. 3.199 Factory automation: Make-to-order production and traceability of components used in engine production, body production and final assembly



Fig. 3.200 Electronics manufacturing: Traceability of components, used in module production for tracking and identifying electronic cards and components

- *Passive* RFID transponders and
- *active* RFID transponders.

Passive RFID transponders *do not have their own power supply*; they are powered by the radio signals from the interrogator. Using a coil as a receiving antenna, a capacitor is charged, which enables the response to be sent in interruptions of the interrogation signal (*load modulation*). The range is limited by the activation energy.

Active RFID transponders have *their own power supply*. They enable higher ranges, but are also considerably more expensive.

Battery-operated transponders are usually in the *idle state* until they are activated (triggered) by a special activation signal. This increases the lifetime of the energy source to months to years. There are two types of separately powered RFID transponders:



Fig. 3.201 Packaging industry: Counterfeit detection of goods and especially drugs code reading and character recognition: Serial number and date code

Fig. 3.202 Interaction of the RFH 620 RF reader with the tag (Factory photo: Sick AG)



1. *Active RFID transponders* use their energy source both to supply the microchip and to generate the modulated return signal. The range can—depending on the permissible transmission power—be kilometers.
2. *Semi-active* RFID transponders are more economical because they do not have their own transmitter and only modulate their backscatter coefficient.

Typical examples of the construction of passive HF and UHF transponders and tags are shown in Figs. 3.203 and 3.204.

RFID Frequency Ranges

The following frequency ranges are distinguished for common RFID systems:

LF:	Low frequencies (30 kHz to 500 kHz).
HF:	High frequencies (3 MHz to 30 MHz; mostly 13.56 MHz).
UHF:	Very high frequencies (850 MHz to 950 MHz; also 433 MHz DOD, USA).
SHF:	Microwave frequencies (2.4 GHz to 2.5 GHz, 5.8 GHz).

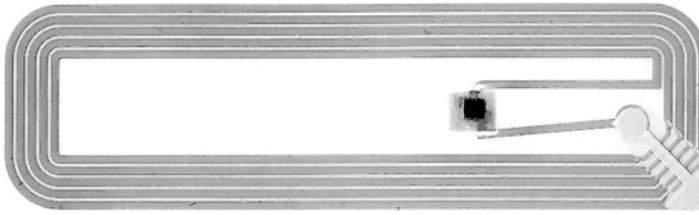


Fig. 3.203 13.56 MHz HF transponder chip with antenna (Source: Wikipedia)



Fig. 3.204 UHF transponder chip with antenna (Factory photo: Sick AG)

The most common frequencies used in industry are currently HF (13.56 MHz) and UHF (850 MHz to 950 MHz). The different frequencies are suitable for different applications:

High Frequencies (HF, Mostly 13.56 MHz)

These frequencies cover short to medium ranges and achieve a medium transmission speed. The readers are in the medium to low price range. This frequency range also includes the so-called *smart tags* (usually at 13.56 MHz).

Very High Frequencies (UHF, 850 MHz to 950 MHz (e.g. According to EPC: Electronic Product Code))

With these frequencies, a *high range* (2 m to 6 m for passive transponders ISO/IEC 18000-6C; 6 m and up to 100 m for semi-active transponders) and a high reading speed can be achieved. Short-lived passive transponders are very inexpensive. RFID is used, for example, in manual, semi-automatic and fully automated goods distribution with pallet and container identification (door seals, license plates) and for checking individual shipping and trading units (EPC tags).

The *Electronic Product Code (EPC)* is an internationally used coding system for a unique identification number, with which products and logistic units (e.g. transport pallets and outer packaging) as well as returnable transport containers can be clearly marked and identified worldwide.

The worldwide standardization according to EPC will enable a *continuous identification* in the supply chain and the “Internet of Things” in the foreseeable future.

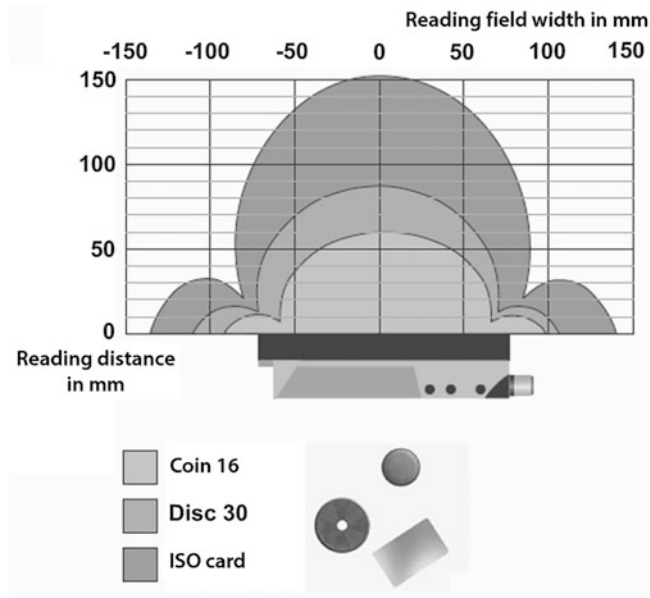


Fig. 3.205 Reading ranges of the RFH620 short-range reader for different tags (Factory photo: Sick AG)

Features and Examples of HF RFID Readers

Important for the design of a HF RFID application are the *possible reading distances* or *ranges*. These depend on the reader, but also on the type and size of the data memories (tags).

Figure 3.205 shows the reading ranges of a proximity-reader for 3 different tag types as a horizontal section through the two-dimensional “reading lobe”.

Typical, proven applications for RFID in the HF range (read distance typically from 50 mm to about 400 mm) are

- Identification of *transport boxes* in the warehouse conveyor system,
- identification on the *overhead conveyor* in the warehouse conveyor system,
- identification in the *KANBAN process* and
- location or navigation for forklift trucks.

When deciding on RFID for goods identification, the cost–benefit ratio should be considered at a very early stage. Figure 3.206 gives an overview of this relationship when using RFID. Economically worthwhile applications tend to be found in the upper left-hand corner.

Features and Applications of UHF RFID Readers

Figure 3.207 shows a compact UHF reader with integrated antenna.

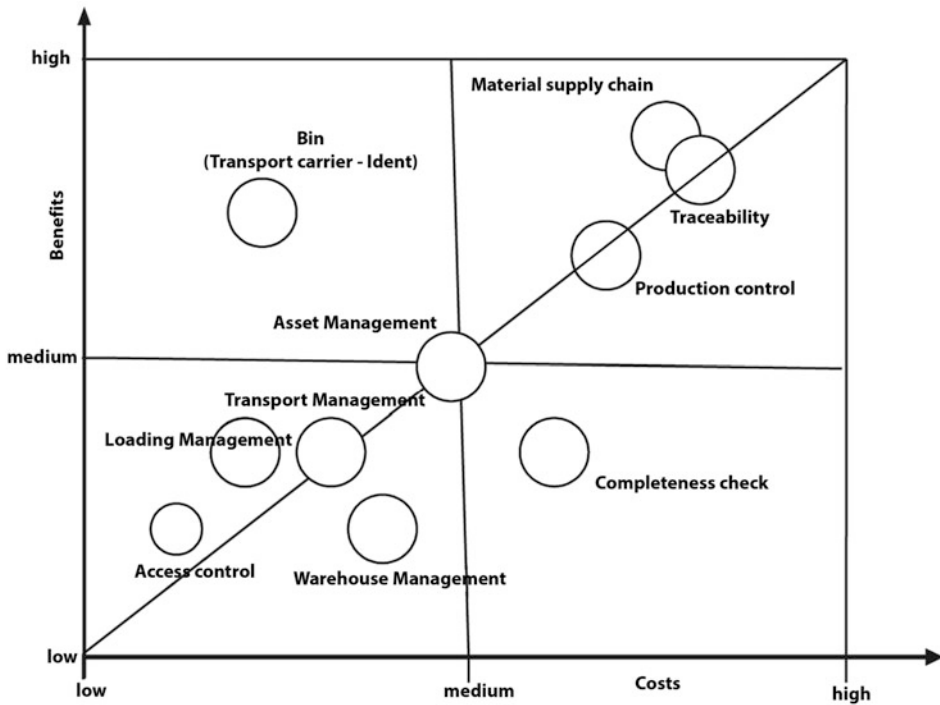
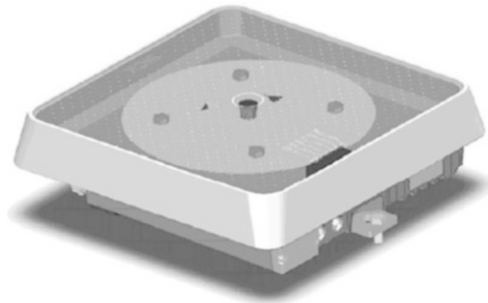


Fig. 3.206 Cost–benefit ratio of different RFID applications (Source: Sick AG)

Fig. 3.207 UHF reader with integrated antenna (Factory photo: Sick AG)



Possible reading ranges are described with the *directional characteristic* of the antenna, together with the specification of the minimum and maximum possible *reading distances*. Typical read distances for UHF readers are between 0.2 m and 6 m (for passive tags). The directional characteristic shows the reading capability as a function of the angle (Fig. 3.208). The *3-dB opening angle* (beam width) is often also given for characterization.

The directional characteristic indicates the field strength (in logarithmic dB: decibel) and thus the spatial reading capability of an antenna over the entire 360-degree solid angle. The so-called *main lobe* with the direction of the highest reading ability at 0 degrees is visible.

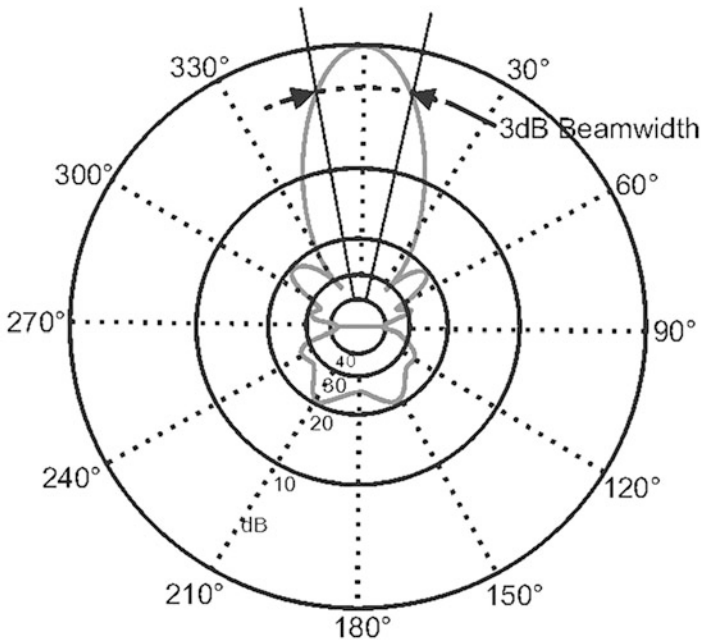


Fig. 3.208 Reading capability (field strength distribution) as a function of the reading angle (Source: Sick AG)

The “side lobes” in the diagram are usually undesirable, but often unavoidable due to the design.

UHF reader antennas can radiate with *linear polarization* or *circular polarization*. With linear polarization, a clear alignment of the tag antenna is necessary (Fig. 3.209 right).

With both linear and circular polarization, it is not possible to align the tag antenna in the direction of wave propagation, since no excitation occurs at the antenna: the electric field component in this direction is zero (Fig. 3.210).

Reflections from metallic or other electrically conductive materials in the reading range can cause *constructive* or *destructive overlapping* of the waves at different points in the reading field (Fig. 3.211). Constructive superpositions result in overreaches (Fig. 3.211: circles on the right), destructive superpositions result in “reading holes” in the regular reading field. Therefore, constellations with metal in the reading field must be avoided as far as possible while maintaining the required high reading reliability.

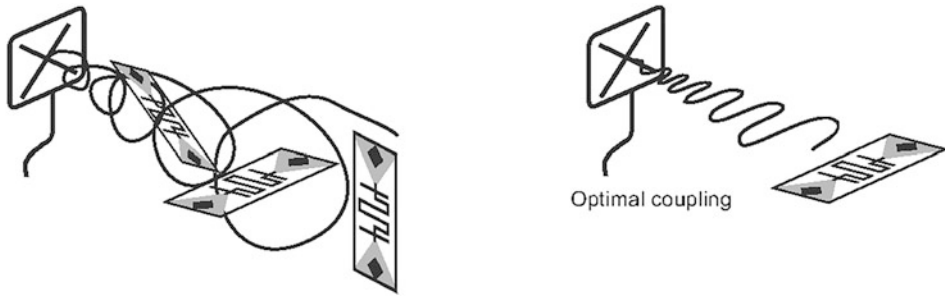


Fig. 3.209 Possible day orientations for circular (left) and linear polarized radiation

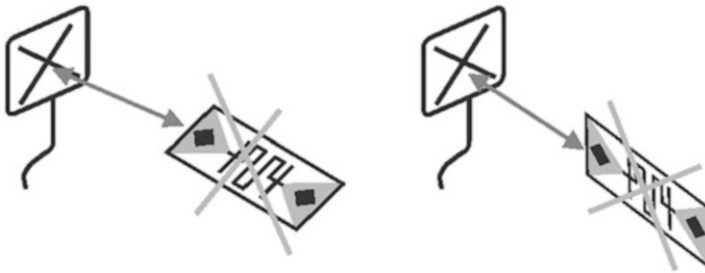
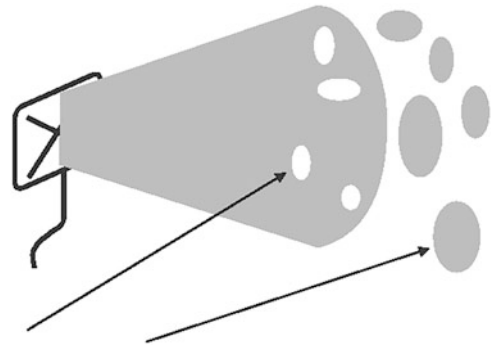


Fig. 3.210 The day orientations shown are generally not possible (no excitation, field strength = 0)

Fig. 3.211 Reading holes and areas with overreach in RFID systems (Source: Sick AG)



3.5 Three-Dimensional Measuring Methods (3D Measurement)

The 3D coordinate measurement is used to *completely capture* the geometry of workpieces or parts of workpieces in a *uniform coordinate system*. The individual features are recorded in a suitable manner with different sensors and/or styli in order to obtain a complete measurement of the entire workpiece.

Both *tactile*, i.e. touching, and *non-contact* measuring methods have become established as measuring methods. *Non-contact* measuring methods include light-optical sensors as single-point sensors and *imaging* sensors—mostly based on cameras. For imaging sensors, the usual illumination methods via coaxial bright or dark field illumination are used. This is usually done in *incident light*, but *transmitted light* is also common. To avoid disturbing shadowing effects in incident light, these illuminations can also be controlled in *segments*. Furthermore, structured illumination types (e.g. fringe projection) or oblique illumination with at least one laser line are used for so-called *laser triangulation* in order to obtain topographic information directly from an image. These methods are described in detail in the following sections.

The sensors are guided along the workpiece *step by step* or by *scanning*. The individual measured points, scans and/or images are then *combined in the correct coordinates* based on the coordinate setting and orientation of the sensor used for the respective measurement and combined to form the *overall measurement result* of the workpiece. In order to be able to determine the coordinates of the sensor during the measurement, the sensors are mostly used on coordinate measuring machines. Examples of such coordinate measuring machines are shown in Fig. 3.212.

3.5.1 Palpable 3D Measurement Methods.

3.5.1.1 Switching Sensor

Switching sensors are the simplest *tactile sensors* used in coordinate measuring technology. They operate according to a simple principle using *electrical switching contacts*. For this purpose, the stylus is inserted into a plate, which is mounted on *three points in a* mechanically clearly defined manner. The disc is pressed into the bearing point by a force applied, for example, via a spring or magnet. Furthermore, the three bearing points are designed as electrical switching contacts and are connected in series. Figure 3.213 shows a principle diagram.

When the stylus touches the measuring object (e.g. a workpiece), a force is applied. The further the sensor moves up or against the workpiece, the stronger this force increases. As a result of this force, a torque is generated around the two bearing points of the stylus plate, which are in the opposite direction to the probing direction of the sensor. As soon as this torque has risen so much that it exceeds the moment at which the stylus disc is pressed into the bearing point via the central spring, one of the switching contacts opens. This interrupts the circuit that monitors the bearing points and thus detects *probing*. With the detection of the probing, all coordinates of the axes of the coordinate measuring machine are determined in *real time* and the coordinate value of the measuring point is calculated.

For the next measurement, the sensor must be removed from the workpiece again so that the contact in the sensor closes again. Only then can the sensor be probed again and a new measuring point be recorded at a different position. Here the measuring process runs in the same way. A measured value is therefore generated with the following three steps:



Fig. 3.212 Common coordinate measuring machines with different requirements and accuracies. From left to right: Articulated arm, horizontal arm coordinate measuring machine, gantry coordinate measuring machine, gantry coordinate measuring machine (Factory photos: Faro Technologies Inc., Carl Zeiss IMT GmbH)

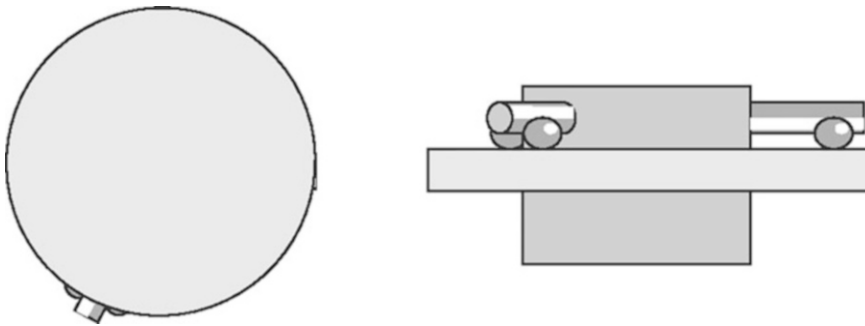


Fig. 3.213 Schematic diagram of the 3 switching contacts in a switching sensor. The three contacts are electrically monitored and connected in series (Factory photo: Carl Zeiss IMT GmbH)

1. *Probing path*—for approaching the measured object until the switching contact opens.
2. *Probing*—and reading the coordinate values of the measuring device.
3. *Scanning path*—to remove the sensor from the workpiece, closing the contacts in the sensor.

Figure 3.214 shows typical switching sensors.

If, for example, a ring gauge is measured from the outside, the following picture appears (Fig. 3.215). In a plane perpendicular to the centre axis, the measurement points do not form a circle but a triangle. This effect is called *lobing*. It results from the three-point bearing of the stylus disc together with the constant pressure force in the centre of the disc via the spring.

The effective radius for the holding torque of the disc varies depending on the direction of the maximum value in the direction of the three balls by $\cos 60^\circ = 0.5$, which corresponds to 50% of the maximum value in the middle between two balls. Assuming a symmetrical, i.e. direction-independent bending behavior of the stylus, the measuring force varies by 50% due to the lobing.

Fig. 3.214 Switching sensors in different versions. Left: TP 2, right: TP 6. The styli are screwed directly onto the sensors (Factory photo: Renishaw)

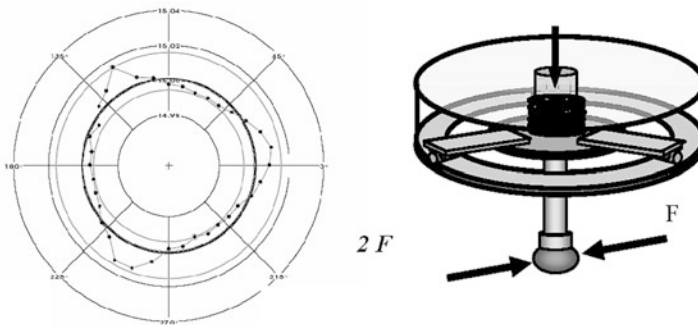


Fig. 3.215 Schematic diagram of the effect of lobing during a measurement on a ring gauge. Left: measurement result; right: associated sensor arrangement with bearing and stylus (Factory photo: Carl Zeiss IMT GmbH)

3.5.1.2 2-phase Switching Sensor

To avoid the effect of lobing, two-stage switching sensors (*2-phase touch trigger probe*) were developed. With these sensors, a very sensitive *second sensor* is added, which detects the contact of the stylus with the workpiece even before the contacts of the switching sensor open. This second, sensitive sensor can be a *strain gauge*, for example, which detects the bending of the probe system or stylus, or a *piezo crystal*, which detects the structure-borne sound when the stylus touches the workpiece. The measuring sequence is then performed in the following 4 steps:

1. *Probing path*—for approaching the measured object until the switching contact opens.
2. *Detection of a probe* from the sensitive sensor, reading the coordinate values of the measuring device.
3. *Probing* by opening the contacts of the three-point bearing of the stylus in the sensor.
4. *Scanning path*—to remove the sensor from the workpiece, close the contacts in the sensor.



Fig. 3.216 Two-phase switching sensors and a changing magazine for the sensor modules (Source: Renishaw)

Thus, the actual measurement takes place when the sensitive sensor detects contact between the stylus and the workpiece. The force on the stylus is then still comparatively low and independent of the probing direction. Opening the contacts in the sensor then serves to confirm that contact with the workpiece has occurred and that the measured value is valid. In signal evaluation, a time window can be specified for the correlation of detection and confirmation so that the *measurement* is recognized as *valid*. Typical two-stage switching sensors are shown in Fig. 3.216.

In order to adapt the stylus configurations, which are screwed to the sensors at the bottom, to the measuring task, the lower parts of these sensors can be exchanged, which contain the spring modules with the switching contacts. This can be done using the changing device with the coordinate measuring machine. Such a changing magazine is shown as an example in the right part of Fig. 3.216.

3.5.2 Optical Probing 3D Measuring Methods

3.5.2.1 Optical, Switching Single Point Sensor

In addition to tactile switching sensors, *optical switching sensors* were also developed. The mode of operation can be seen in Fig. 3.217. A spot of light is projected onto the object to be measured and then imaged via a ring lens onto a position *sensitive detector* (PSD—*Position Sensitive Detector*). Depending on the distance between the detector and the surface of the workpiece, the diameter of the circle that is imaged on the detector changes. With a *threshold value* for the signal of the PSD set to the center of the measuring range, the signal of the probe is triggered and the coordinates of the coordinate measuring machine are read out. The measuring procedure is basically the same as that used for the tactile switching sensors. Before the next measuring point is reached, it must be scanned again.

The advantage of the optical switching sensor is that it works *without contact* and can therefore also be used for *sensitive surfaces*. The sensor needs cooperative surfaces because it depends on sufficient light being reflected to obtain a good measurement signal. On edges or in corners, there may also be shadows, which can influence the measuring accuracy; so

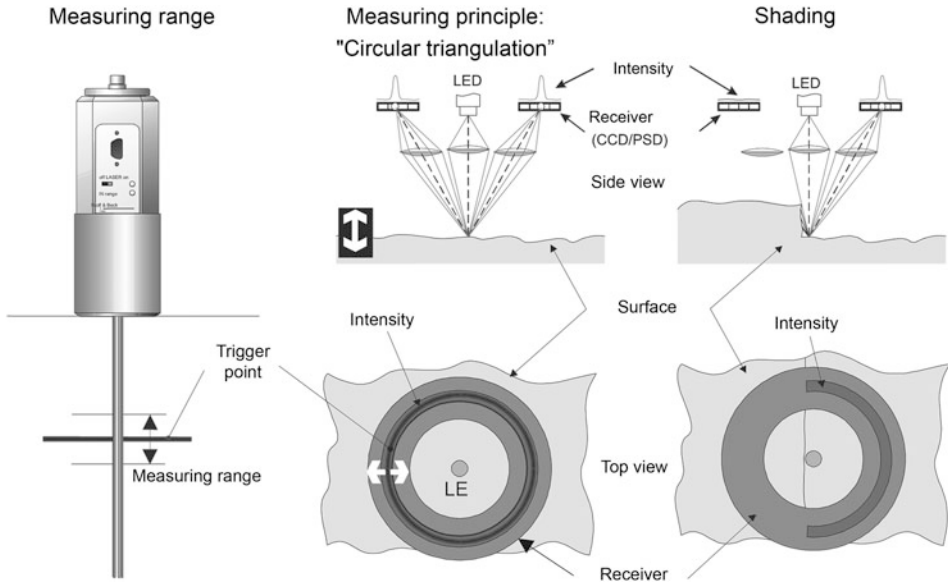


Fig. 3.217 Design and function of an optical single-point sensor on almost flat surfaces and edges (Factory photo: Carl Zeiss IMT GmbH)

when setting up the measuring plan for a workpiece, care must be taken to avoid such effects.

The single-phase and two-phase switching sensors described so far, as well as the optical switching sensors, are often used in combination with *rotary/swivel joints* in order to be able to use these sensors to probe the workpiece with different stylus directions. Before measurement, the sensors must be calibrated on a *calibration ball* so that the data measured in the different settings can be combined in a common coordinate system.

3.5.2.2 Scanning Sensors

With continuously available measurement data, scanning, i.e. *continuous scanning* of the workpiece along lines, becomes possible. The lines are measured as a sequence of individual points, whereby the point density on the line depends on the *scanning speed* of the coordinate measuring machine and the *measuring frequency*.

The sensor is guided along the contour of the workpiece, using the current measurement value of the sensor to correct the *target position* of the path of the CMM relative to the workpiece during measurement. The aim of this correction is to ensure that the sensor is always operated in the center or a target value within its measuring range, if possible. This makes the system as robust as possible against leaving the measuring range of the sensor and avoids the loss of measured values. This requirement for the control of the coordinate measuring machine is the same for optical and tactile scanning sensors.

Scanning simplifies the measuring process compared to the single-point sensors described above, since in this case the workpiece only has to be touched at the beginning of the *scanning path*. The sensor then follows the scanning path along the workpiece as specified by the control of the coordinate measuring machine. The workpiece is not scanned again until the scan is complete. If the scanning path is very curved, it may be necessary to reduce the scanning speed so that the stylus remains on the workpiece surface and does not lift off due to path acceleration and inertial forces.

The advantage of scanning compared to single point measurements is the *higher point density* and thus the additional information about the workpiece. This will be shown by an example. Let us assume that it is to be evaluated whether a cylinder fits into a bore during a fit. For this purpose, the hole is measured.

In the tactile case typically with 4 single points. Through these 4 points, a circle can then be fitted, the diameter of which is calculated. This is then compared with the diameter of the cylinder. If the diameter of the cylinder is smaller than that of the bore, the cylinder should fit. But it is possible that the cylinder still does not fit. Figure 3.218 shows the relationships.

Only a larger number of measuring points on the scanning path shows that the bore is not round. The calculated pen circle is smaller than the diameter of the cylinder and therefore the cylinder cannot fit into the bore. The larger number of points from the scanning allows a more accurate evaluation of the situation.

3.5.2.3 Other Probe and Measuring Systems

In this section, we will introduce further available buttons in sketch form to complete the overview. Optical measuring techniques such as chromatic *white light sensors* as a further development of the presented point sensors, *confocal microscopy techniques* such as *laser scanning microscopy* and *confocal scanning microscopy* are mentioned as methods. These measuring techniques allow depth-resolved images to be taken along the optical axis, so that topographic information can be obtained from them, either point by point or over a large area.

For tactile coordinate measuring technology, there is a trend towards the measurement of micromechanical and optical components. Thus, the technology used is also coming closer to the technologies used in probe and scanning probe microscopes. These are for example *AFM systems (Atomic Force Microscope)*, which use the interaction of sample and object almost without mechanical contact, *SNOM (Surface Near Field Optical Microscope)*, which use the interaction of light with the sample surface. Other sensors use measuring principles of interferometry, for example the *WhitePoint sensor* from Bosch or *optical coherence tomography OCT*, in order to achieve a high depth resolution.

In metrology, for example, *fiber probes* are used, which use a glass fiber with an attached glass bead of up to 20- μm diameter to obtain a light spot at the measurement location (Fig. 3.219). This illuminated glass bead is recorded by a camera. The coordinate measurement is then performed via the camera image. This method is also used in transmitted light, whereby the glass sphere is only used as a local probe, which provides

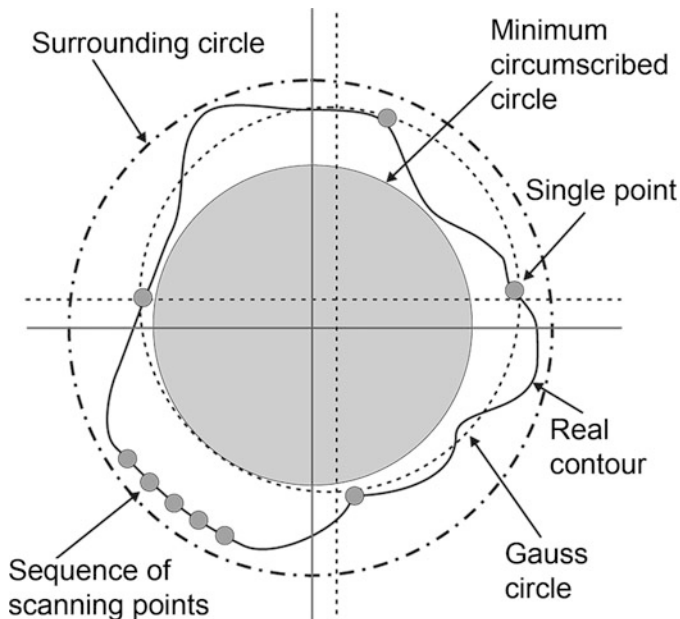


Fig. 3.218 Difference between single-point measurement and scanning for the evaluation of workpiece properties (Factory photo: Carl Zeiss IMT GmbH)

a shadow for the optical measurement in the object plane. The actual measurement is performed by evaluating video images.

Another compact probe is the SSP sensor, which is based on silicon chip technology. The entire electrical and mechanical functional structure is based on silicon technology. The probe is shown in Fig. 3.220.

On the silicon membrane and the functional structure in the central part of the sensor, structures in the form of strain gauges are applied to detect the deflection of the sensor. The contact surfaces for the electrical connections can be seen on the edge of the membrane. The problem with these high-resolution sensors is then that there are no longer any suitable traceable standards for full qualification according to DIN/EN. For this reason, the specifications in the data sheets are in the range of at least 250 nm, even if reproducibility of the measurements is in some cases significantly better.

3.5.3 3D Imaging Measuring Methods

3.5.3.1 Optical 3D Measurement (Grid and Line Projection)

The determination of spatial geometric shapes by means of *line projection* is a non-contact measuring method, which uses the possibilities of modern image processing. By means of an appropriate measuring arrangement with associated image evaluation, objects can be

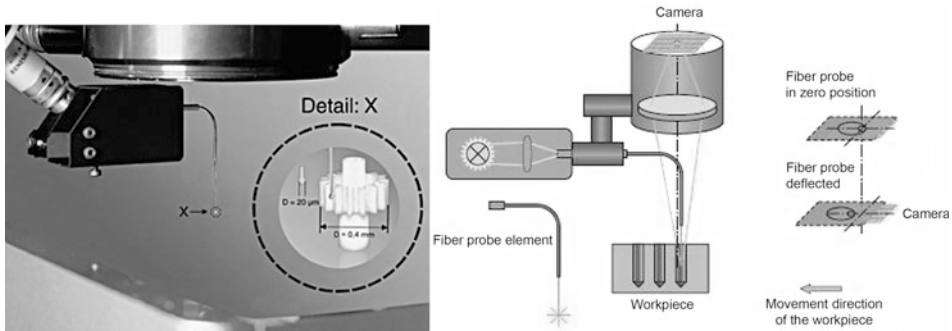


Fig. 3.219 Fiber probe for coordinate measurement on coordinate measuring machines, left: arrangement on the measuring device, right: operating principle (Factory photo: Werth Messtechnik GmbH)

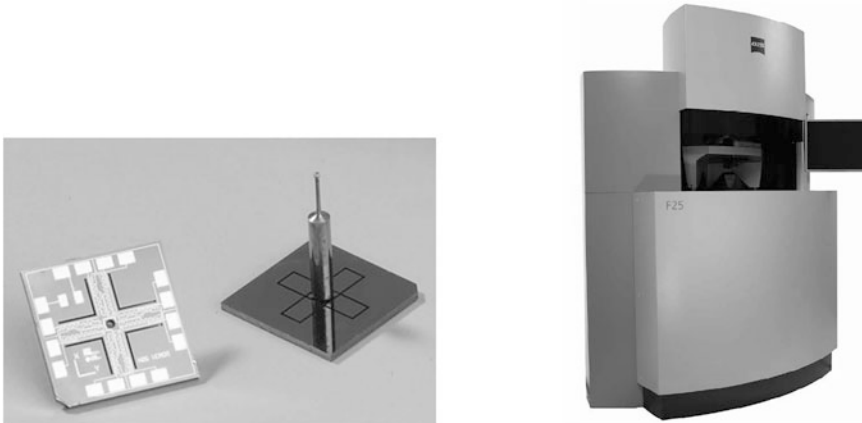


Fig. 3.220 Silicon probe with integrated stylus with stylus balls of up to 100- μm diameter, edge length chip 6.5 mm, right: F25 coordinate measuring machine (Factory photo: Carl Zeiss IMT GmbH)

recorded with their expansions in all three coordinates. This method is also called *light section*.

3.5.3.2 Measuring Principle and Measuring Arrangement

If an object is viewed vertically from above, its image can be used to determine the expansion in x - and y -direction. However, this image does not contain any information about the distance to the object to be measured. If the object is illuminated with a defined pattern at a known angle, the image is distorted by the contour of the surface. This image distortion can be detected by image processing and the third dimension of the surface can be calculated from this.

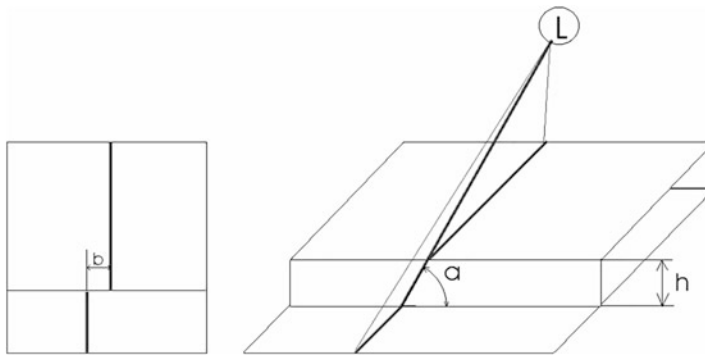


Fig. 3.221 Principle diagram of light section measurement for line projection

The mode of operation is explained in Fig. 3.221 using a simple cuboid. A laser is used to project a line onto the workpiece at an angle a to the base surface. This angle of incidence results in the projected line on the lower lying surface with an offset b compared to the surface above. The offset b is determined by image processing. The height h (expansion in the third dimension) can be calculated from the values of offset b and angle a :

$$h = b \cdot \tan(a).$$

The image of the drawn line changes with the shape of the change in height, i.e. slopes or arcs as object edge. The effort of the back calculation into the object height increases with the more complicated the profile becomes. Two examples are shown in Fig. 3.222.

The previous considerations were based on a simply designed object. With a line, the height or height profile was determined only on this line. For this reason, several lines are projected onto the workpiece, so that a height profile in x-direction is created. With this arrangement, the height profiles in y-direction are missing, which can be determined using a second line net offset by 90° . Thus, a grid is created above the object. A practical application is shown in Fig. 3.223. The here relatively coarse grid is used to identify the object. By comparing it with stored images, the resulting pattern can be assigned to a product. In the case of a metrological task, the grid would have to be much tighter.

3.5.3.3 Fringe Projection

Projected images can also be used as an alternative to line projection. In this case, alternating light and dark stripes are used, whose edges are measured in the image plane. These stripe images can also be extended by the second dimension, resulting in checks. By changing the pattern, for example with different strip widths, the measurement resolution can be changed. It is also possible to use patterns adapted to the object.

Figure 3.224 shows two objects with differently set orientations of the projection. This variation allows shadows to be compensated and the resolution to be increased.

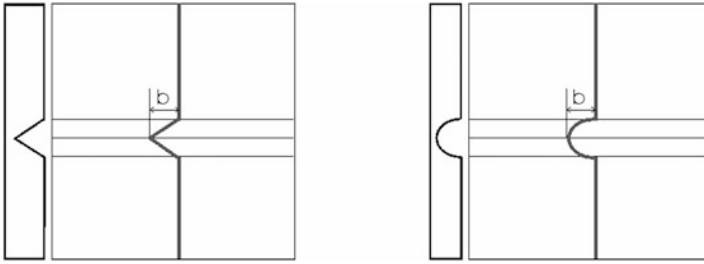


Fig. 3.222 Principle diagram of light section measurement for different profiles

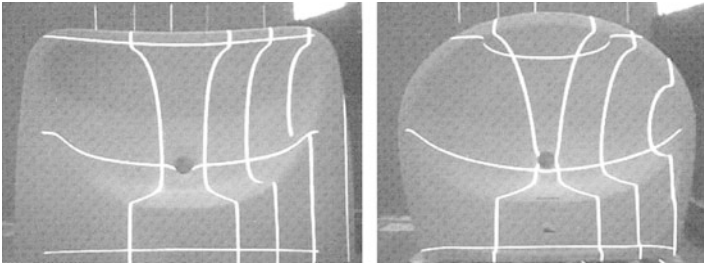


Fig. 3.223 Example of a line projection for object detection (Source: C. Schaarschmidt: “Der Laser bringt es ans Licht”, Automation&Drives, Heft 4/2007, S. 106 ff)

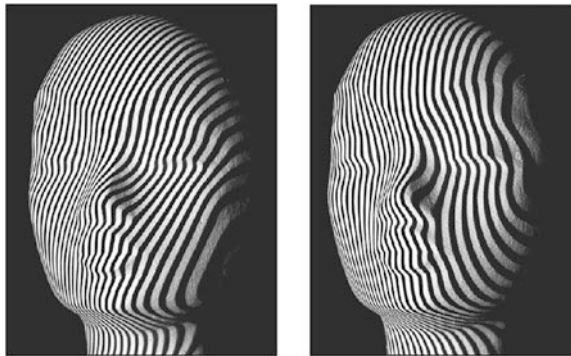


Fig. 3.224 Example of a fringe projection for 3D measurement (Source: S. Winkelbach: Das 3D-Puzzle-Problem—Effiziente Methoden zum paarweisen Zusammensetzen von dreidimensionalen Fragmenten. Fortschritte in der Robotik. Band 10, Shaker-Verlag 2006)

To increase the resolution and to support the subsequent mathematical calculations, several different images are projected for one measurement. In Fig. 3.225, the parts: Figs. 3, 4 and 6 were removed from the projection sequence of 6 images. The projector is shown on the left and the camera on the right.

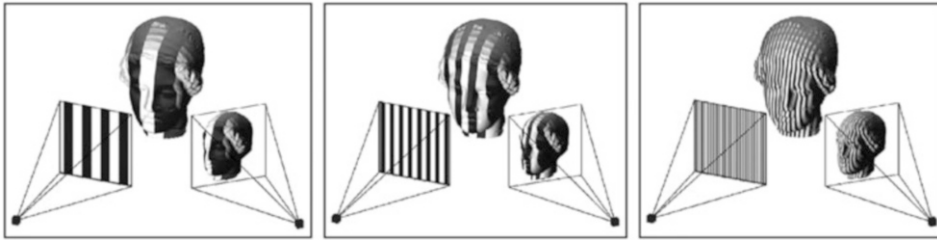


Fig. 3.225 Sequence of a measurement using fringe projection (Source: S. Winkelbach: Das 3D-Puzzle-Problem—Effiziente Methoden zum paarweisen Zusammensetzen von dreidimensionalen Fragmenten. Fortschritte in der Robotik, Band 10, Shaker-Verlag 2006)

From the entirety of the acquired data, a 3D model is calculated as a *grid* or *volume model*. A result image is shown in Fig. 3.226.

3.5.3.4 Limitations of the Procedure

In principle, the method can be used if the measured object allows line imaging, i.e. the light is not absorbed. The complexity of this procedure is determined by the image processing and the mathematical models for the recalculation of the image data to the third dimension of the object. Some difficulties may occur, which increase the mathematical effort considerably. For example, a laser line is generated by an optical system from a point source. In addition to the desired angle of attack, this beam also has a projection angle that changes over the line. The same applies to the camera and its optical aperture angle. For this reason, measuring marks are placed practically on the background of the arrangement, which allow a correction of the resulting spatial distortions.

Both distortions can be avoided for higher metrological requirements by using *telecentric optics* (with a beam path parallel to the image side), which, however, are only available for small object sizes. The method presented here is only of limited use if the object under consideration has undercuts. These then require a change of the projection location, which can be done, for example, by rotating the object.

3.5.3.5 Fields of Application

In contrast to tactile 3D measuring methods, the optical variant is a very fast method. Accuracy and resolution do not depend on the size of the object, but on the properties—especially the number of pixels—of the detecting camera. This optical measuring method can therefore be used not only for large workpieces (e.g. machine parts), but also for small objects (e.g. determining the layer thickness in the semiconductor industry). It is used in various areas of spatial measurement in industry as well as for classification and recognition of object shapes. Figure 3.227 shows a height profile on a silicon wafer created by means of a light section.

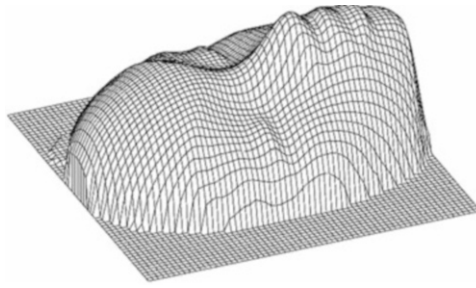


Fig. 3.226 3D model calculated from the measurement images (Source: S. Winkelbach: Das 3D-Puzzle-Problem—Effiziente Methoden zum paarweisen Zusammensetzen von dreidimensionalen Fragmenten. Fortschritte in der Robotik, Band 10, Shaker-Verlag 2006)

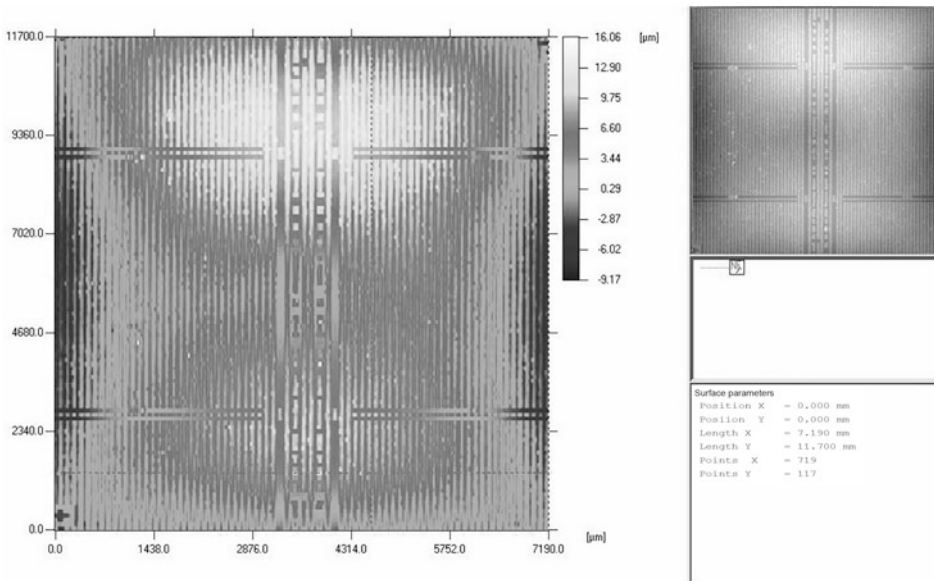


Fig. 3.227 Result of surface measurement using a light section microscope (Source: Qimonda factory photograph)

3.5.4 Overview of 3D Measuring Methods

In the following, the different sensors are arranged according to their characteristics. The overview (Fig. 3.228) shows how many different tactile and non-contact sensors and measuring principles are used. The overview clearly shows the trend in coordinate measuring technology towards *multi-sensor technology*, as the properties of the sensors determine their suitability for different measuring tasks.

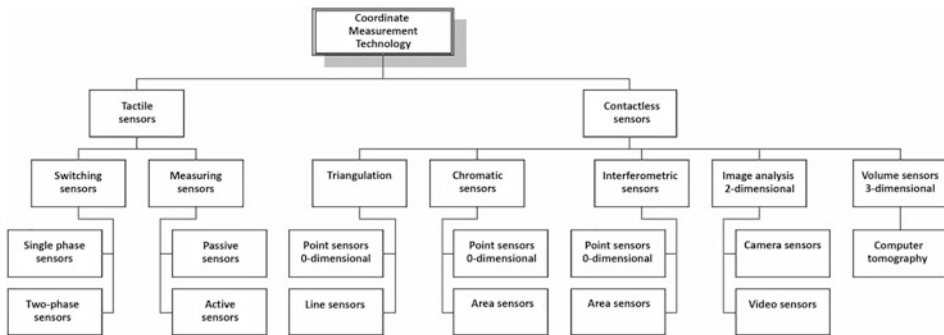


Fig. 3.228 Overview of the measuring principles used in coordinate metrology

Bibliography

1. Adam, W.; Busch M.; Nickoly B.: Sensoren für die Produktionstechnik. Springer Verlag, Berlin 1997
2. Allwood, D. A.; Xiong, G.; Faulkner, C. C.; Atkinson, D.; Petit, D.; Cowburn, R. P.: Science 309 (2005), 1688
3. Asch, G.: Les capteurs en instrumentation industrielle. Paris: Dunod, 2006
4. Balluff-Patent, EP1158266B1/2004: Wegmess-System
5. Barthélémy, A.; Fert, A. and Petroff, F.: „Giant Magnetoresistance in Magnetic Multilayers“, in Handbook of Magnetic Materials, vol. 12, K.H.J. Buschow, Editor. Elsevier: Amsterdam, 1999, pp. 1–96.
6. Baxter, L.K.: Capacitive Sensors: Design and Applications. IEEE Press, New York, 1997
7. Beach, G.S.D.; Nistor, C.; Knutson, C.; Tsoi, M. and Erskine, J. L.: „Dynamics of? eld-driven domain-wall propagation in ferromagnetic nanowires“, Nat. Mater., 4, 741 (2005)
8. Belbachir, A. N.: Smart Cameras, Springer Verlag, New York, 1. Auflage, 2009
9. Bernstein, J.: An Overview of MEMS Inertial Sensing Technology, Sensors, Febr. 2003
10. Bhatt, H.; Glover, B.: RFID Essentials, O’Reilly Media; 1. Auflage, Februar 2006
11. Brasseur, G.; Fulmek, P. L.; Smetana, W.: Virtual rotor grounding of capacitive angular position sensors. IEEE Trans. Instrum. Meas., Vol. 49, No 5, Oct. 2000, pp. 1108–1111
12. Burkhardt, T.; Feinäugle, A.; Fericean, S.; Forkl, A.: Lineare Weg- und Abstandssensoren, Die Bibliothek der Technik – Band 271, München: Verlag Moderne Industrie, 2004
13. Diegel, M.; Mattheis, R.: DE 10 2004 020 149 A angemeldet
14. Diegel, M.; Mattheis, R., Halder, E.: „Multiturn counter using movement and storage of 180° magnetic domain walls“, Sensor Letters, vol. 5, pp. 118–122, Jan. 2007
15. Diegel, M.; Mattheis, R.; Halder, E.: 360° Domain Wall Investigation for Sensor Applications, IEEE Trans. Magn. 404 (2004) 2655–2657
16. DIN EN 60947-5-2 (VDE 0660 Teil 208): Niederspannungsschaltgeräte – Teil 5-2: Steuergeräte und Schaltelemente – Näherungsschalter, November 2004
17. Dorf, C. R. (Hrsg.): Electrical Engineering Handbook, New York: IEEE Press, 2006
18. Droxler, R.: Berührungslos arbeitender Näherungsschalter, Balluff-Patent, DE 19611810 C2, 2000
19. Elwenspöck, M.; Wiegierink, R.: Mechanical Microsensors. Sprinkler-Verlag Berlin, Heidelberg New York. pp. 230–236

20. Fatikow, S.; Rembold, U.: *Microsystem Technology and Microrobotics*, pp. 224–229
21. Fericean, S.; Droxler, R.: *New Noncontacting Inductive Analog Proximity and Inductive Linear Displacement Sensors for Industrial Automation*. In: *IEEE SENSORS JOURNAL*, Vol. 7, No. 11, November 2007
22. Fericean, S.; Friedrich, M.; Fritton, M.; Reider, T.: *Moderne Wirbelstromsensoren – linear und temperaturstabil*. In: *Elektronik Jahrgang 50* (April 2001), Nr. 8
23. Ferrari, V.; Ghisla, A.; Marioli, D.; Taroni, A.: *Capacitive Angular-Position Sensor With Electrically Floating Conductive Rotor and Measurement Redundancy*. *IEEE Trans. Instrum. Meas.*, Vol. 55, No 2, Apr. 2006, pp. 514–520
24. Finkenzeller, K.: *RFID-Handbuch*, Hanser Fachbuchverlag, 4. Auflage, August 2006
25. Fraden, J.: *Handbook of Modern Sensors – Physics, Designs, and Applications*, 4. Auflage, Springer Verlag, New York, 2010
26. Gass, E.; Pali S.; Melles, A.: *Induktiver Wegaufnehmer mit einem passiven Resonanzkreis aufweisendem Messkopf*, Balluff-Patent, DE 102 19 678 C1, 2003
27. Gasulla, M.; Li, X.; Meijer, G.C.M.; Ham, L. van der; Spronck, J. W.: *A Contactless Capacitive Angular-Position Sensor*. *IEEE Sensors Journal*, Vol. 3, No 5, Oct. 2003, pp. 607–614
28. George, B.; Mohan, N. M.; Kumar, V. J.: *A Linear Variable Differential Capacitive Transducer for Sensing Planar Angles*, IMTC 2006 Conference, Sorrento, Italy, (2006), pp. 2070–2075
29. Glathe, S.; Mattheis, R.; Berkov, D.: *„Direct observation and control of the Walker Breakdown process during a field driven domain wall motion“*, *Appl. Phys. Lett.* Vol. 93, pp. 072508, 2008a
30. Glathe, S.; Mattheis, R.; Mikolajick, R.; Berkov, D.: *„Experimental study of domain wall motion in long nanostripes under the influence of a transversal field“*, *Appl. Phys. Lett.* vol. 93, pp. 162–505, 2008b
31. GS-14 geophone. Geospace Inc. 7334 N. Gessner Rd., Houston, TX 77040
32. Halder, E.; Diegel, M.; Mattheis, R.; Steenbeck, K.: DE 102 39 904 A angemeldet
33. Herold, H.: *Sensortechnik – Sensorwirkprinzipien und Sensorsysteme*. Heidelberg: Hüthig, 1993
34. Hesse, S., Schnell, G.: *Sensoren für die Prozess- und Fabrikautomation*. Vieweg Praxiswissen, 2004
35. Hoffmann, J.: *Taschenbuch der Messtechnik*, 2. Auflage, Fachbuchverlag Leipzig, München, 2000
36. Homburg, D.; Reiff, E.-Ch.: *Weg- und Winkelmessung (Absolute Messverfahren)*, PKS, Homburg, 2003
37. Horst Siedle GmbH & Co. KG, European patent application EP 1 740 909, 2005
38. Horst Siedle GmbH & Co. KG., European patent application EP 1 532 425, 2003
39. <http://www.amci.com>
40. <http://www.ltn.de>
41. <http://www.micronor-ag.ch>
42. <http://www.sick.com/>
43. Hubert, A. and Schäfer, R.: *Magnetic domains*, Springer Verlag: Berlin, 1998a, pp. 215–290
44. Hubert, A. and Schäfer, R.: *Magnetic domains*, Springer Verlag: Berlin, 1998b, pp. 35–14
45. IPHT, German patent application DE 102006039490 A1, 2006
46. Jagiella, M.; Fericean, S.: *Miniaturized Inductive Sensors for Industrial Applications*. In: *Proceedings of the first IEEE International Conference on Sensors*. Orlando/USA: 2002
47. Jagiella, M.; Fericean, S.: *Neues magnetoinduktives Sensor-Prinzip und seine Umsetzung in Sensoren für industrielle Anwendungen*. In: *Technisches Messen*, Jahrgang 72, Heft 11. 2005
48. Jagiella, M.; Fericean, S.; Dorneich, A.: *Progress and Recent Realizations of Miniaturized Inductive Proximity Sensors for Automation*. In: *IEEE SENSORS JOURNAL*, Vol. 6, No. 6, December 2006

49. Jagiella, M.; Fericean, S.; Dorneich, A.; Droxler, R.: Die magneto-induktive Variante. In TECHNICA 7/2003, pp. 36–38, Juli 2003a
50. Jagiella, M.; Fericean, S.; Dorneich, A.; Droxler, R.: Sensoren für kurze und mittlere Wege, In SENSOR report 2/2003, pp. 17–19, März 2003b
51. Jagiella, M.; Fericean, S.; Droxler, R.; Dorneich, A.: New Magneto-inductive Sensing Principle and its Implementation in Sensors for Industrial Applications. In: Proceedings of the 4th IEEE International Conference on Sensors. Wien: 2004
52. Jagiella, M.; Fericean, S.; Friedrich, Dorneich, A.: Mehrstufige Temperaturkompensation bei induktiven Sensoren. In: Elektronik Jahrgang 52 (August 2003c), Nr. 16
53. Kennel, R.: Encoders for Simultaneous Sensing of Position and Speed in Electrical Drives with Digital Control, 40th IEEE IAS 2005 Annual Meeting, Kowloon, Hong Kong, Oct. 2–6, 2005
54. Kennel, R.; Basler, S.: New Developments in Capacitive Encoders for Servo Drives, SPEEDAM 2008, Ischia, Italy, (2008), pp. 190–195
55. Kiel, E. (Hrsg.): Antriebslösungen – Mechatronik für Produktion und Logistik, Springer Verlag, Berlin, Heidelberg, New York, 2007
56. Kuttruff, H.: Physik und Technik des Ultraschalls, Hirzel Verlag 1988
57. Mattheis, R.; Diegel, M.; Hübner, U. and Halder, E.: „Multiturn Counter Using the Movement and Storage of 180 Magnetic Domain Walls“, IEEE Trans. Magn., Vol. 42, pp. 3297–3299, October 2006
58. MicroMagus – software for micromagnetic simulations, by D. V. Berkov & N. L. Gorn, <http://www.micromagus.de>
59. Müller, R. K.: Mechanische Größen elektrisch gemessen, Expert Verlag, Ehningen, 1990
60. N.N., Leitfaden Drehgeber – Begriffe und Kenngrößen, Publikation des ZVEI – Zentralverband Elektrotechnik- und Elektronikindustrie e.V., Fachverband Automation, Okt. 2008
61. Pallás-Areny, R.; Webster, J. G.: Sensors and Signal Conditioning. 2. Auflage. New York: John Wiley & Sons, 2001a
62. Pallás-Areny, R.; Webster, J.G.: Sensors and Signal Conditioning. 2. Auflage. New York: John Wiley & Sons, 2001b
63. Pfeiffer, D.; Powell, W. B.: „The Electrolytic Tilt Sensor“, Sensors, May 2000
64. Prutton, M.: Thin ferromagnetic films, Butterworth: London, 1964, p. 41
65. Reif, K. (Hrsg.): Sensoren im Kraftfahrzeug. Vieweg+Teubner Verlag, Wiesbaden 2010, 176 S., mit 221 Abb., Br. ISBN 978-3-8348-1315-2
66. Röbel, M.: Pythagoras in der Moderne, IEE, September 2006, disynet GmbH
67. Rohling, H.: „Skriptum zur Vorlesung Radartechnik und –signalverarbeitung“, Technische Universität Hamburg-Harburg; <http://www.et2.tu-harburg.de/lehre/Radarsignalverarbeitung/skriptul9.pdf>
68. Schaarschmidt, C.: „Der Laser bringt es ans Licht“, Automation&Drives, Heft 4/2007, S. 106 ff.
69. Schiessle, E.: Industriesensorik – Automation, Messtechnik, Mechatronik“, 1. Auflage, Vogel Buchverlag, Würzburg, 2010
70. Schröder, D.: Elektrische Antriebe – Regelung von Antriebssystemen, 2. Auflage, Springer Verlag, Berlin, Heidelberg, 2001
71. Shigeto, K.; Shinjo, T.; Ono, T.: Appl. Phys. Lett. 75 (1999), 2815
72. SICK Firmenseite: www.sick.de
73. Sorge, G.: Faszination Ultraschall, B.G. Teubner Verlag 2002
74. Stork, T.: Electric Compass Design using KMZ51 and KMZ52. Philips Semiconductors Systems Laboratory, Hamburg, January 2003
75. Sweeney, P. J.: RFID für Dummies (deutsch), Wiley-VCH Verlag, 1. Auflage, Juli 2006
76. Tamm, G.; Tribowski, C.: RFID, Springer Verlag, Berlin, 1. Auflage, März 2010

77. Thiaville, A. und Nakatani, Y.: „Domain wall dynamics in nanowires and nanostrips“. In: Spin Dynamics in Confined Magnetic Structures III, Eds. B. Hillerbrands, A. Thiaville, Springer Verlag, New York (2006)
78. Tränkler, H. R.; Obermeier, E. (Hrsg.): Sensortechnik. Springer Verlag, Berlin 1998
79. Tyco Electronics, Global Automotive Division: Sensors Catalog, 2010
80. VAC-Vacuumschmelze: Magnetische Sensoren, PS-000 Broschüre
81. Webster J. G. (Hrsg.): Measurement, Instrumentation and Sensors. New York: IEEE Press, 1999
82. Winkelbach, S.: Das 3D-Puzzle-Problem – Effiziente Methoden zum paarweisen Zusammensetzen von dreidimensionalen Fragmenten. Fortschritte in der Robotik, Band 10, Shaker-Verlag 2006
83. www.balluff.de



Ekbert Hering, Gert Schönfelder, and Stefan Vinzelberg

4.1 Mass

4.1.1 Definition

The mass m is a measure for the number of particles (e.g. atoms, molecules) of a body. The masses can be added up like quantities. The unit of measurement of mass is one kilogram (kg) and is defined by a calibration body. Mass is independent of both the location and the state of motion of a body (in relativistic physics, mass is dependent on speed; however, this only becomes apparent at very high speeds of the order of the speed of light).

The mass m has two characteristics:

1. *Inert mass.* The mass is a measure of the resistance of a body to a change in motion. According to *Newton's law of action*, it applies to constant masses:

$F = m \cdot a$ with F : force and a : acceleration. This means that a force F acts in such a way that a mass is accelerated with the acceleration a . If the same force is applied to two masses m_1 and m_2 , then the following applies

E. Hering (✉)

Hochschule Aalen (Rektor i. R.), Aalen, Germany

e-mail: Ekbert.Hering@hs-aalen.de

G. Schönfelder

Prignitz Mikrosystemtechnik, Wittenberge, Germany

S. Vinzelberg

Atomic Force F&E GmbH, Mannheim, Germany

e-mail: S.Vinzelberg@hs-mannheim.de

$$m_1/m_2 = a_2/a_1.$$

This means that it is possible to determine the mass of a body by measuring the accelerations.

If a mass m performs a rotational movement with the angular velocity ω , it is kept on the circular path with the radius r by a *centripetal force* F_{cp} . It then applies:

$$F_{cp} = -m \omega^2 \cdot r. \text{ A mass determination can also be made via this force.}$$

A spring-mass system carries out *mechanical oscillations*. For the natural frequency ω_0 applies:

$$\omega_0 = \sqrt{\frac{k}{m}}. \text{ The factor } k \text{ is the spring constant.}$$

The equation states that a change in mass results in a change in frequency. In this way, atomic masses can be determined (Fig. 4.2).

2. *Heavy mass*. Between two masses m_1 and m_2 , the *gravitational force* F_G acts as attracting force. It applies to the amount of the gravitational force:

$$|F_G| = G \frac{m_1 m_2}{r_{12}^2}.$$

r_{12} is the distance between the two masses. G is the gravitational constant ($G = 6.673 \cdot 10^{-11} \text{ m}^3/(\text{kg s}^2)$). Due to the force of attraction, the weight force $F_W = m \cdot g$ (g : acceleration due to gravity; $g = 9.81 \text{ m/s}^2$). Thus, the mass of a body can be determined via the weight force. This is done by force sensors (Sect. 4.2).

4.1.2 Applications

Some special applications of mass sensors are presented below. Figure 4.1 shows a mass sensor mounted on the top of a grain tank (location A in Fig. 4.1). The grains harvested by the combine harvester are thrown against a baffle plate. The impact force of the grains generates an electronic pulse. This is a measure of the mass of the grains.

At California University (California Universität), Alex Zettl's institute has developed a nano-electro-mechanical system (NEMS) that makes it possible to measure the mass of a single gold atom.

Figure 4.2a shows a double-walled carbon nanotube. It moves freely on one side. The other side is clamped to an electrode. By applying a DC voltage, a negative charge is generated at the tip of the nanotube. An alternating voltage in the radio frequency range causes the nanotube to oscillate. For the resonance frequency ω_{res} applies, as already explained in Sect. 4.1:

$$\omega_{res} = \sqrt{\frac{k}{m}}, \text{ where } k \text{ is the spring constant and } m \text{ the mass.}$$

If an atom or a molecule hits the nanotube, the resonance frequency also changes according to the above equation because of the change in mass. The mass can be

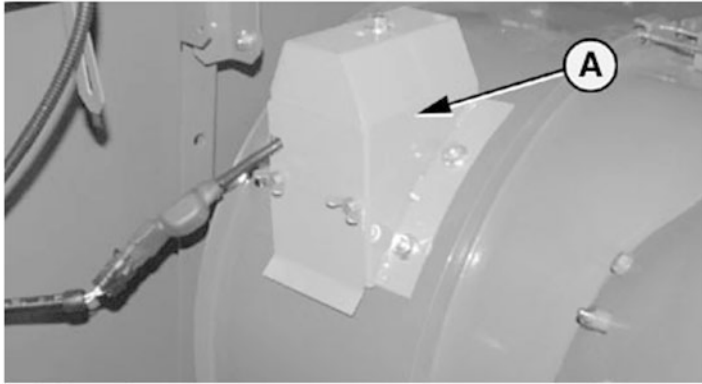


Fig. 4.1 Mass sensor on a combine harvester to determine the mass of the harvested grain (Factory photo: John Deere Harvester Works)

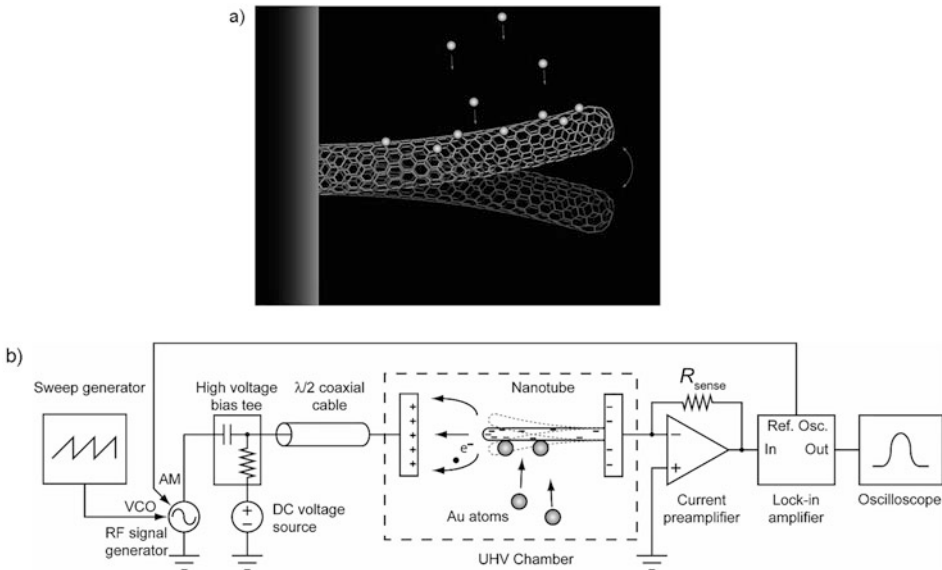
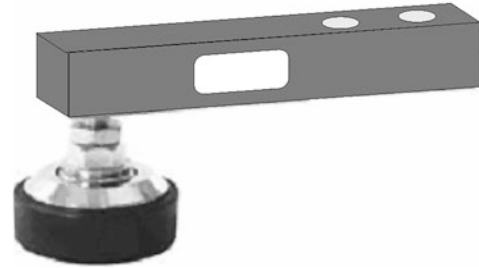


Fig. 4.2 Mass sensor for atoms and molecules through a nanotube. (a) Nanotube on which gold atoms “rain”; (b) circuit for measuring the resonance frequency (Source: Alex Zettl, California University USA)

determined from this change in frequency. Figure 4.3 shows a typical example of how the mass can be determined by measuring the weight force. Many scales are based on this principle.

The shear beam load cell shown in Fig. 4.3 is very often used as a sensor in weighing technology (e.g., as hopper scales, belt weighers, floor scales, or platform scales). They are

Fig. 4.3 Shear beam load cell
(Factory photo: Bosche)



very accurate and can be used without problems in rough industrial environments. Their nominal load is between 500 kg and 10,000 kg. Such weight sensors can also be used to check the weights of trucks and railway wagons.

4.2 Force

4.2.1 Definition

According to Newton's second axiom, the force F for bodies with constant mass is proportional to the instantaneous acceleration a . It holds:

$$F = m \cdot a.$$

The force F is thus a *vector* whose direction is parallel to the acceleration a . The unit of force is: $1 \text{ N} = 1 \text{ kgm/s}^2$. The following forces play an important role in practice:

4.2.1.1 Weight Force F_W

A body falls to the ground due to gravity (Sect. 4.1, heavy mass) with the acceleration of gravity g or has a weight force F_W , to which applies:

$$F_W = m \cdot g.$$

The acceleration of gravity g ($g = 9.81 \text{ m/s}^2$) points approximately in the direction of the centre of the earth.

4.2.1.2 Centripetal Force F_{cp}

The centripetal force keeps a body of mass m in a uniform circular motion. It is:

$$F_{cp} = -m \cdot \omega^2 \cdot r.$$

Here ω is the angular velocity of the circular motion and r the distance of the mass point from the center of the circle. The body itself feels a force (inertia force), which pushes it outward, the *centrifugal force*. This force is described by same formula, but the vector shows into opposite direction (outward).

4.2.1.3 Elastic Force or Spring Force F_{el}

Within certain deformation limits, solids show an *elastic* behaviour which is described by the spring force F_{el} :

$$F_{el} = -k \cdot s$$

where s is the change in length and k is the *spring constant*. In practice, spring scales are used as force sensors.

4.2.1.4 Frictional Force F_F

For solids, the frictional force is dependent on the *normal force* F_N . This is the force that acts perpendicularly on the base. The proportionality factor is the *coefficient of friction* μ , so the following applies

$$F_F = \mu \cdot F_N.$$

A distinction is made between rolling friction (μ_R), sliding friction (μ_G) and static friction (μ_H). The rolling friction leads to lower frictional forces than the sliding and static friction, which is why the following applies

$$\mu_R < \mu_G < \mu_H.$$

4.2.2 Effects for the Applications

It is of great importance how the sensor is installed. For *direct* force measurement, the component must be cut open to install the force sensor (Fig. 4.4a). This intervention must not impair the mechanical properties of the component (e.g., strength and rigidity).

However, direct installation has the great advantage that the *absolute force* is measured and that the force acts almost *linearly* and is *independent of the point of application*. Especially small forces can be measured correctly and accurately with direct force measurement. Figure 4.4b shows the *indirect* force measurement. It is used when it is not practical to separate the workpiece and when very large forces are to be measured. With indirect force measurement, the sensor is permanently installed in the force flow of the component. The sensor only measures a part of the real force, which is why it must be calibrated and this must be taken into account in the evaluation electronics.



Fig. 4.4 Arrangements for force measurement (a) Direct force measurement; (b) Indirect force measurement (Factory photos: Kistler)

Figure 4.5 shows the physical effects that play an important role in force sensors.

4.2.2.1 Piezoelectric Effect

The *piezoelectric effect* described in Sect. 2.1 converts the external force *directly* into an electrical signal.

As Fig. 4.6a shows, the external force displaces the crystal lattice, so that electric charges are shifted to the two outer sides. This results in a voltage that is directly proportional to the force. This direct conversion has the advantage that, in addition to the simple evaluation electronics, fast periodic force fluctuations can also be accurately recorded in real time.

The piezoelectric sensors are very robust with temperature fluctuations (from $-270\text{ }^{\circ}\text{C}$ to $+400\text{ }^{\circ}\text{C}$) and cover a wide measuring range of 9 powers of ten (10^9).

Figure 4.7 shows a *piezoelectric force washer* used to measure forces in the range between 20 N and 200 kN. The measuring rings can withstand a high overload without damage and without loss of accuracy and linearity. Other advantages of such measuring rings are that they are *self-centering* via the inner and outer diameter and therefore *do not* show any *hysteresis* in the measurements.

Piezoresistive Effect

When a force is applied to a metallic conductor, a mechanical deformation, an *elongation*, occurs. This changes the specific electrical resistance ρ of the conductor. This *resistance-elongation effect* or *piezoresistive effect* is based on a *change in the specific electrical resistance* of the conductor material on the one hand and on *structural changes on the other*. Special materials (e.g. constantan: 54Cu45Ni1Mn) and geometries of so-called *strain gauges* allow the detection of a force and a strain (Sect. 4.5) at different points and in different directions. The following relationship applies between change in resistance $\Delta R/R$ and change in length or strain ε ($\varepsilon = \Delta l/l$):

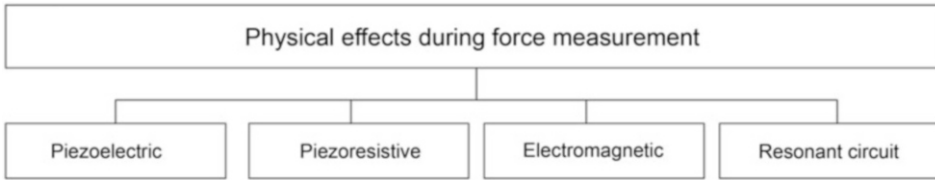


Fig. 4.5 Physical principles in force sensors

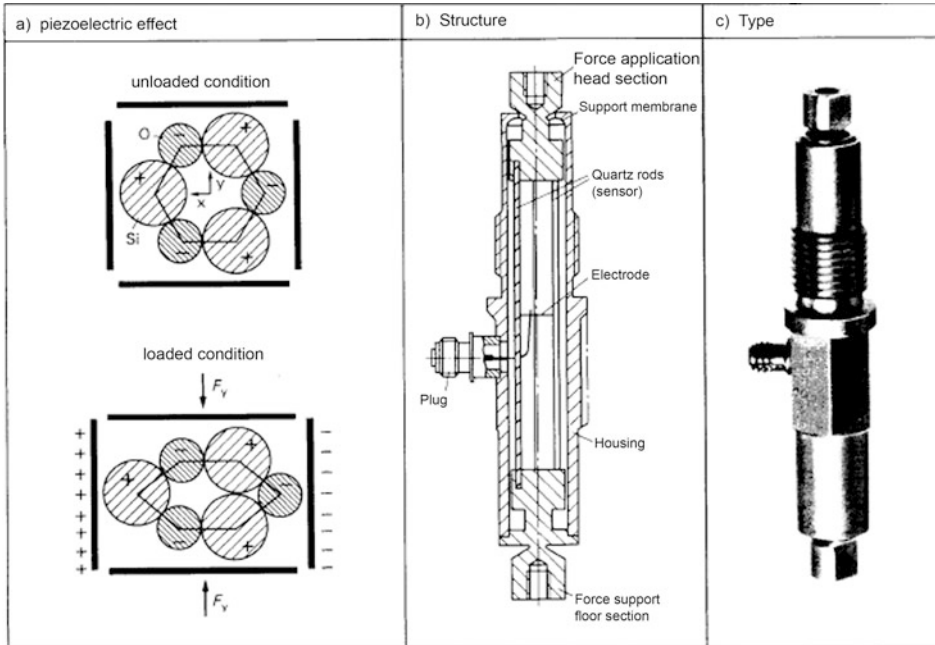


Fig. 4.6 Piezoelectric measuring principle and sensors (a) measuring principle; (b) design of a piezoelectric sensor; (c) design of a piezoelectric sensor (Factory photo: Kistler; Source: Hering, Bressler, Gutekunst: Electronic and Electric for engineers, 8 edition, Springer 2021)

$$\Delta R/R = k \cdot \varepsilon.$$

In Sect. 2.2, the k-factors of individual materials and the piezoresistive effect are explained in detail. Figure 4.8 shows the different geometries for different applications.

For a correct measurement with strain gauges, the choice of fasteners and the method of application are decisive. They ensure that the strain caused by the force is transferred to the strain gauge without loss.



Fig. 4.7 Piezoelectric force washer (Factory photo: HBM)



Fig. 4.8 Different geometries of strain gauges (Factory photo: HBM)

Electromagnetic Force Effect

If a conductor of the length ℓ through which a current of the current intensity I flows is in a magnetic field with the magnetic flux density B , then a force F acts on this conductor to which the following applies:

$$F = I \cdot \ell \cdot B \cdot \sin \varphi.$$

φ is the angle between the magnetic flux density B and the current-carrying conductor of length ℓ . This effect is the basis for the operation of an *electrodynamical* loudspeaker (Sect. 9.3).

This effect allows the measurement of very small forces. Its installation and calibration is very complex. Therefore, this effect is only suitable for special applications.

Resonant Circuit

In a spring-mass system, mechanical vibrations can occur. The following applies for the resonant frequency ω_{res} , as already explained in Sect. 4.1:

$$\omega_{\text{res}} = \sqrt{\frac{k}{m}},$$

where k is the spring constant and m the mass.

The spring force F_{el} is proportional to the deflection x in the elastic range, so that the following applies

$$F_{\text{el}} = -k \cdot x.$$

With this effect, very small forces can be measured. This principle is used in the *atomic force microscope*. The oscillation frequency of the carrier of the scanning needle changes its resonance frequency as the sample approaches. This is caused by the *Van der Waals binding forces* (Sect. 2.17).

4.2.3 Application Areas

Force measurements play an important role in many areas. Most applications are based on the piezoelectric and the piezoresistive effect. Figure 4.9 shows the most important application areas of force sensors.

4.2.3.1 Process Control

Force measurement is used in the manufacturing processes of primary forming, forming, cutting, and joining for optimal control of the manufacturing processes. For machine tools and handling equipment, force measurement is essential to guarantee the required quality and safety during production. Some examples are presented below.

Figure 4.10 shows the measurement of forces in the production process.

The pressing force plays a decisive role in the pressing of tablets (Fig. 4.11a), in the extrusion process in plastics technology (Fig. 4.11b), or in the insertion of pins or lids (Fig. 4.11c).

In machine tools or handling devices, it must often be possible to measure forces in all three spatial dimensions x , y and z . Three-dimensional force sensors are used for this purpose, as Fig. 4.12 shows.

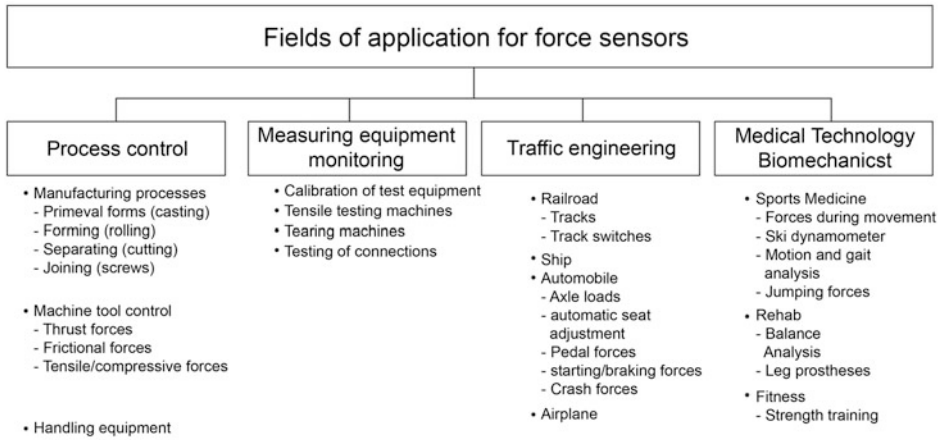


Fig. 4.9 Application areas for force sensors



Fig. 4.10 Force sensors in the production process (a) measurement of the cutting force, (b) tool monitoring, (c) control of the roll forces (Factory photos: Kistler)

Measuring Equipment Monitoring

Regular monitoring of the measuring and testing equipment is essential for production with consistently high quality. The control is usually done by means of specially calibrated *force washers* (Fig. 4.13).

These force transducers are used in particular for *highly dynamic force measurements* (e.g. in tensile or tearing machines) because they are very robust and do not show any long-term fatigue.

Traffic Engineering

Force sensors are used in rail, ship, and air traffic as well as in automobiles. Figure 4.14 shows the applications in the railway sector, for monitoring the holding forces on rails (Fig. 4.14a) and for controlling the resistance forces during switch adjustments (Fig. 4.14b).

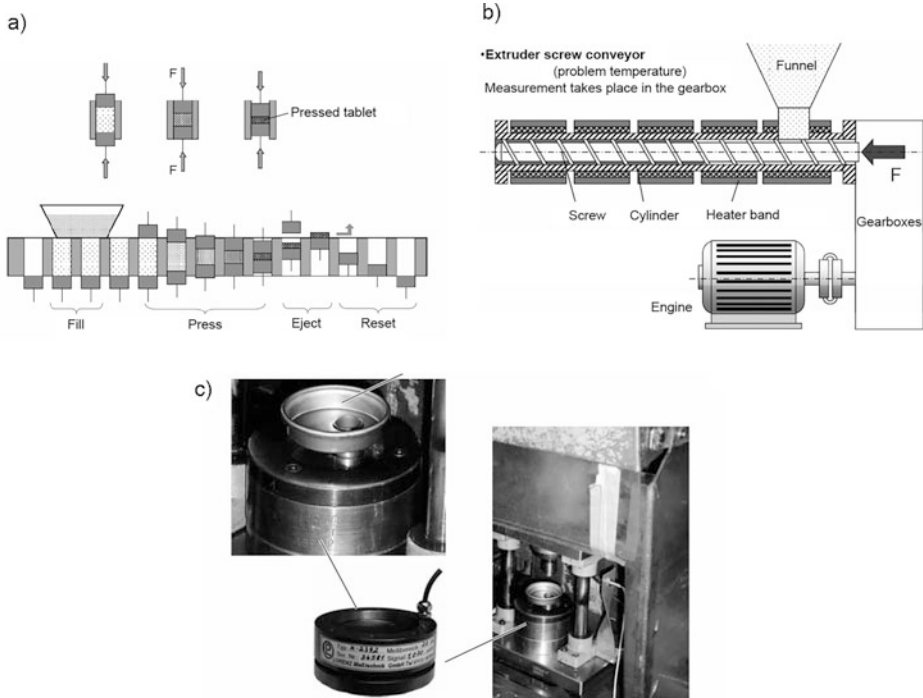
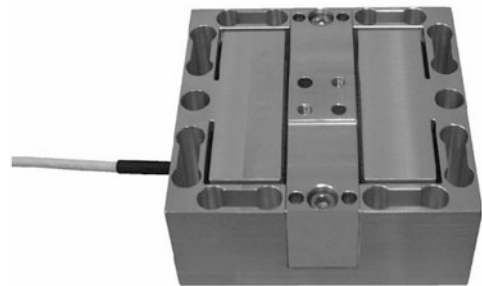


Fig. 4.11 Monitoring of the pressing forces (a) in tablet machines, (b) during the extrusion of plastics (c) during the pressing of pins and lids (Factory photos: Lorenz Messtechnik)

Fig. 4.12 Three-dimensional force sensors (Factory photo: ME-Messsysteme)



Medical Technology and Biomechanics

In medical technology, force sensors are used to measure the forces during motion sequences. This helps athletes to optimize their performance or surgeons to use force measurements in gait analyses to make the right choice of foot, hip, knee, and arm prostheses. Force sensors are also used in the rehabilitation of sick people for fitness equipment, for balance analyses, or to determine the holding forces of hands (Fig. 4.15).

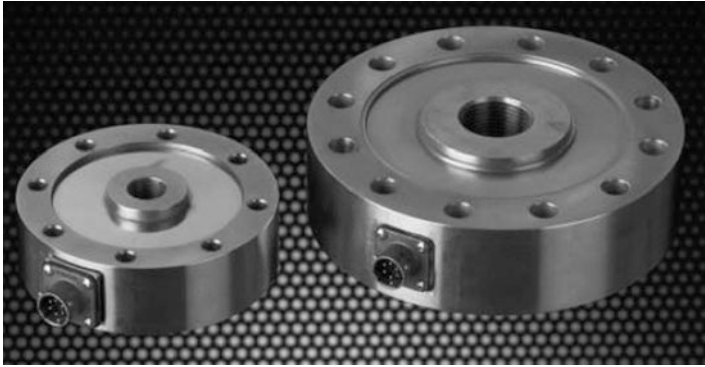


Fig. 4.13 Force sensors for the calibration of measuring and testing equipment (Factory photo: Soemer)

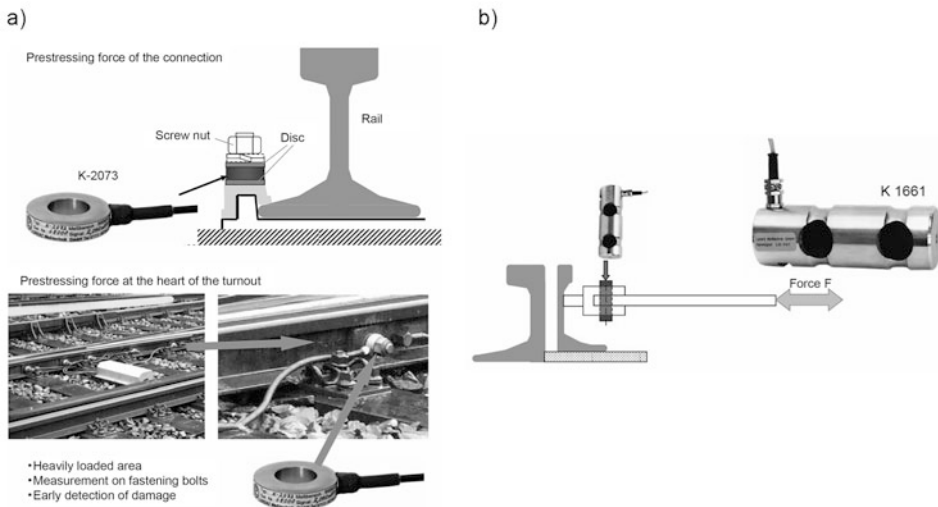


Fig. 4.14 Force sensors for monitoring (a) the holding forces of rails, (b) the forces during switch adjustments (Factory photos: Lorenz Messtechnik)

4.3 Elongation

4.3.1 Definition

The elongation ε describes the change in length of a body when subjected to an external load. This can be a force as well as a thermal influence. Here, the elongation (alternatively also compression) due to the application of force is considered in particular.



Fig. 4.15 Force sensors in medical technology: Hand force measurement (Factory photos: Lorenz Messtechnik)

$$\varepsilon = \frac{\Delta l}{l_0} \quad \text{with} \quad \Delta l = \text{change in length and} \quad l_0 = \text{output length.}$$

The elongation is unit-less and is usually given in percent or ppm (part per million). For most materials, the elongation is *linear* to the action of force in the area of elastic deformation. This is described by *Hooke's law*.

When a force is applied to a body, it not only stretches or compresses the body, but it also changes its *diameter* proportionally (Fig. 4.16). The ratio is described by the Poisson number μ .

4.3.2 Strain Measurement

If the elongation or compression acts on an electrical conductor, the resulting change in length and cross section affects a *change* in *electrical resistance* (Sect. 2.2, Fig. 2.6). This effect can be intensified by selecting suitable materials.

The use of the physical effect results in the *strain gauge* as a component, which can be constructively adapted to different tasks. The basic design of a strain gauge is shown in Fig. 4.17.

In practical applications, strain gauges are used as bridge circuits. This *compensates* for the *thermal effects* on the strain gauge and the object, and also makes use of the strain and compression that always occur together.

The sensitivity k of the strain gauge is determined by

$$k = \frac{\Delta R/R_0}{\Delta l/l_0} = \frac{\Delta R/R_0}{\varepsilon}.$$

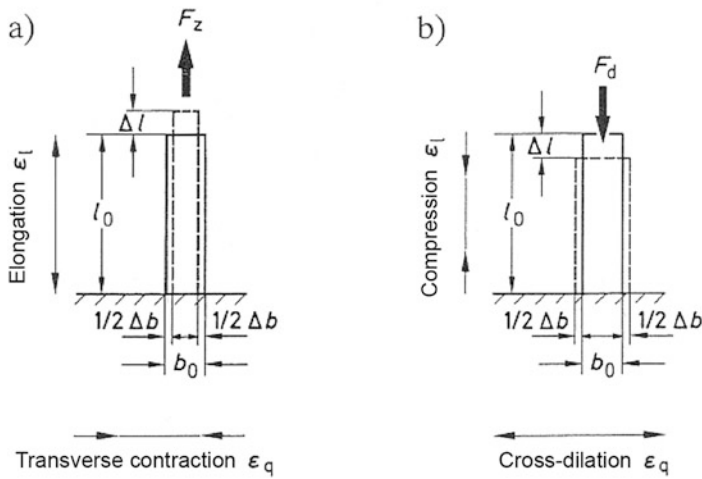


Fig. 4.16 Transverse contraction and transverse dilatation under the effect of a force (Source: Hoffmann; “Introduction to measurement technology with strain gauges”; at HBM)

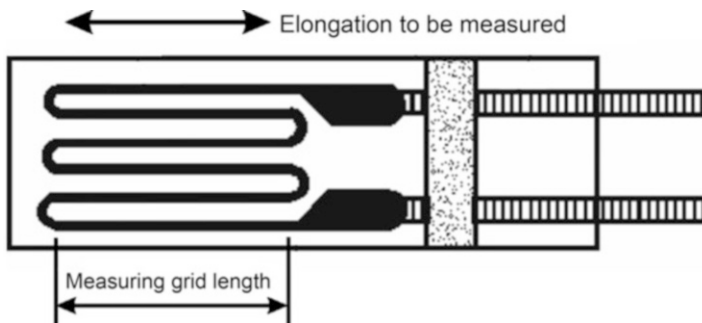


Fig. 4.17 Basic structure of a strain gauge

Using four strain gauges in a Wheatstone bridge circuit (Fig. 4.18) gives an approximate value for the output signal:

$$\frac{U_A}{U_B} = \frac{k}{4} \cdot (\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4).$$

If the strain gauges are mounted on the measuring body at an angle of 90° to each other, R1 reacts to the strain in the circumference and R2 to the compression in the length. Complete bridges are offered directly for this frequently occurring case (Fig. 4.19).

Fig. 4.18 Measurement bridge with strain gauge (Factory photo: HBM)

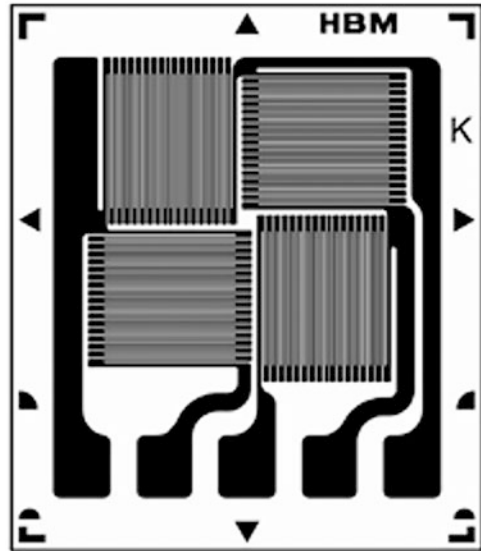
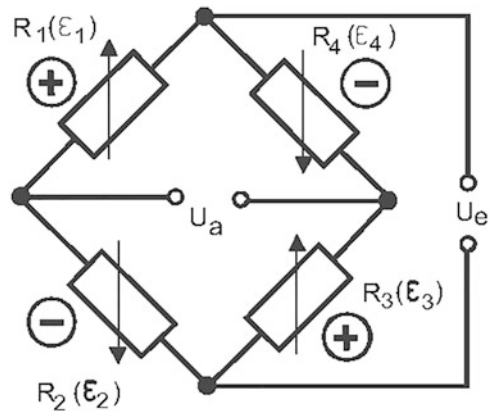


Fig. 4.19 Bridge arrangement of 4 strain gauges (Source: HBM)



4.3.2.1 Bending Beam

Another basic principle for the use of strain measurement is force measurement by means of bending beams. The effect diagram is shown in Fig. 4.20. The applied force F causes the bending beam to deflect downwards. This causes an *elongation* on the *upper side* and a *compression* on the *lower side*. These are recorded and evaluated by the strain gauges. Figure 2.9 shows a bending beam in practical use.

4.3.2.2 Fiber Bragg Grating

The fiber Bragg grating is based on an inscribed interference filter in an optical fiber. The basic mechanism is shown in Fig. 4.21.

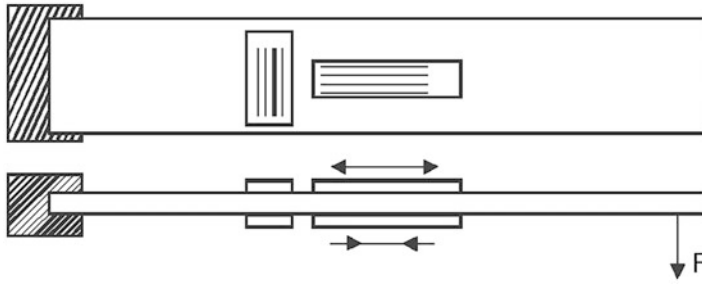


Fig. 4.20 Measuring principle of the bending beam

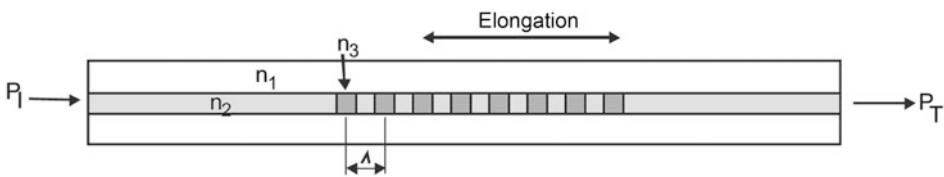


Fig. 4.21 Principle of the fiber Bragg grating

Light is fed into an optical fiber with the power P_1 at a defined wavelength λ_B . The introduced structure with the refractive indices n_2 and n_3 causes a reflection of the wave if the wavelength corresponds to the wavelength determined by the grating period Λ . This wavelength is calculated from

$$\lambda_B = \frac{n_2 + n_3}{2} \cdot 2\Lambda.$$

If the wavelength of the interference filter and light match, the output power P_T at the fiber end is low. If the fiber is stretched, the grating period Λ changes and thus shifts the filter frequency. As a result, the input power is less attenuated and thus the output signal increases.

Further examples from the field of force measurement using strain gauges are given in Sects. 4.2 and 4.5.

4.4 Pressure

4.4.1 Definition

The pressure p is a frequently used physical quantity and describes the *force* F , which acts *perpendicularly on a surface* A . If the pressure is exerted by a gas or a liquid, it acts equally strongly on all surrounding surfaces.

The symbol for the pressure is “p”. The following applies: $p = F/A$ (F: force; A: area perpendicular to the force). The unit of pressure p is Pa (Pascal). The following applies: $1 \text{ Pa} = 1 \text{ N/m}^2$. Apart from the SI unit Pascal, older and nationally coined units are also common. These include among others:

Bar	1 bar = 100,000 Pa or 1 mbar = 100 Pa = 1 hPa (hectopascal). Hectopascal is commonly used for meteorological data.
Psi	(pounds per square inch) with 1 bar = 14.5 psi. It is common in English-speaking countries.
Torr = mmHg	1 mmHg = 133.3 Pa. It is historically used in medicine (blood pressure).
mH ₂ O	10.21 m water column corresponds to 1 bar. It is mainly used for level measurements.

These units can be converted into each other using Table 4.1.

4.4.1.1 Fields of Application

Pressures are measured in many fields of application, as Table 4.2 shows in an overview.

There are different applications for certain pressure ranges, as Table 4.3 shows.

4.4.2 Measuring Principles

4.4.2.1 Measuring Principles

Pressures can be measured *statically* and *dynamically*, using different physical principles. A distinction is also made between *absolute* and *relative* pressure.

The pressure is preferably measured in gases and liquids. For solid objects, the pressure is measured by measuring the mass or force. The pressure measurement can be carried out in three ways:

- As *differential pressure*,
- as *relative value* to the air pressure and
- as an *absolute value*.

Almost all pressure measuring devices have in common that the deformation of a diaphragm is determined by the pressure difference between the front and back side. Thus, the three methods mentioned differ only in the technical design of the sensor.

4.4.2.2 Differential Pressure Measurement

The pressures to be measured are applied to both sides of the membrane. As an example, this can be the connection before and after a filter. In this case, the filter causes a pressure drop, which is a measure of how dirty the filter is.

Table 4.1 Conversion of pressure units

	Pascal	bar	psi	mmHg = Torr	mH ₂ O
Pascal	1	10 ⁻⁵	14.504*10 ⁻⁵	0.00750	0.000102
Bar	10 ⁵	1	14.504	750.0617	10.2110
Psi	6894.8	0.06895	1	51.715	0.704
mmHg = Torr	133.32	1.3332*10 ⁻³	0.01934	1	0.0136
mH ₂ O	9807	0.0980	1.422	73.796	1

Table 4.2 Application fields of the pressure sensors

Field of application/ sector	Concrete application
Automotive, commercial vehicles	Brake systems, ABS Injection technology Hydraulics on work equipment
Aerospace	Pneumatics and hydraulics for actuators (e.g. chassis and doors) air conditioning technology
Mechanical engineering	Hydraulics and pneumatics on machines Pressure on presses
Building services engineering	Air-conditioning in buildings Control of print chains in clean room technology
Medicine	Diagnostics (e.g., blood pressure measurement, respiration)
Environmental technology	Air pressure measurement Level measurement in waters

4.4.2.3 Relative Pressure Measurement

In this case, the pressure to be measured is applied to one side of the membrane. The back is open to the air pressure, which thus serves as a reference value. However, this has the consequence that the fluctuations of the air pressure are included in the measurement as errors. Therefore, relative pressure measurement is unsuitable for pressures in the range of air pressure or depends very much on the technical application. A special case of relative pressure measurement is *level measurement*. This is actually a difference measurement between the pressure of the water column including the “air column” standing on it and the “air column” alone. However, to measure the water column, the sensor must be located at the lowest point. This means that the air pressure required as a reference must be supplied to the back of the membrane via a pipe.

4.4.2.4 Absolute Pressure Measurement

In order to eliminate the inaccuracy of the air pressure as a reference, the *back of the membrane* is covered with a sealed volume of defined pressure. This is also the only way to measure the air pressure itself. Another possibility of absolute pressure measurement is the use of measuring principles that are not based on the comparison with another pressure, but

Table 4.3 Assignment of pressure ranges to application fields

Pressure ranges	Concrete application
<1 mbar absolute (<100 Pa)	High vacuum technology in the semiconductor industry space travel
0.1–10 mbar (10 to 1000 Pa) Relative/difference	Building services engineering, air-conditioning technology
1 mbar to approx. 10 bar (100 Pa to 1 MPa) Absolute/relative	Level probes (1 mbar corresponds to approx. 1 cm water column)
1 bar to 30 bar (0.1 MPa to 3 MPa)	Compressed air technology
Up to 400 bar (up to 40 MPa)	Compressed gas storage (gas cylinders typ. 200 and 300 bar) Air reservoir for starting motors
100 bar to 600 bar (10 MPa to 60 MPa)	Hydraulics
From 500 bar (up to 50 MPa)	Injection pressure at the engine (typ. 1000 bar)
Approx. 3000 bar (about 300 MPa)	Water jet cutting

the reference is based on the material structure. These include measuring methods that are based on the piezoelectric effect or exploit the pressure dependence of piezoresonators.

4.4.3 Measuring Arrangements

The classical arrangement for pressure measurement is the conversion of a *mechanical deformation*, for example of a membrane or bellows, to a display (functional principle of a pressure gauge).

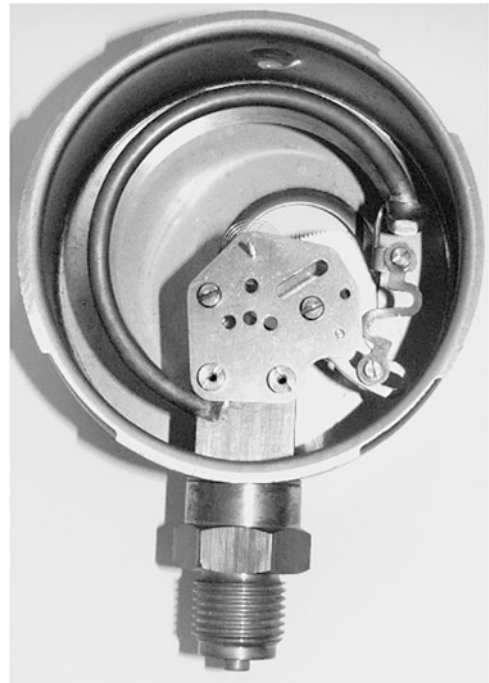
The classic among pressure gauges is the *pressure gauge* based on *Bourdon tubes*. These are still used today as they are robust and do not require any auxiliary energy. They can be used wherever no further signal processing is required. An example is shown in Figs. 4.22 and 4.23. These instruments are also available as electromechanical pressure gauges. In this case, a combined sampling of the measured value is carried out mechanically and electrically. In the simplest case, the rotary movement of the pointer is converted into an electrical signal (see magnetic angle sensors, Sect. 3.2.2).

In electronic signal processing, this mechanical deformation is detected by electrical operating principles. The most common method here is to record the *deformation* by (piezo-)resistive bridges or *capacitive distance measurement* (Sects. 3.2.4 and 2.6.1). Figure 4.24 shows the design principle of a pressure sensor with a resistive measuring bridge. The measuring range of such a cell can be adapted by varying the diaphragm diameter and thickness. Steel, ceramic, or semiconductor materials are mainly used as diaphragm materials.

Fig. 4.22 Bourdon tube pressure gauge (Factory photo: WIKA)



Fig. 4.23 Internal mechanics of a Bourdon tube pressure gauge (Factory photo: EMPEO)



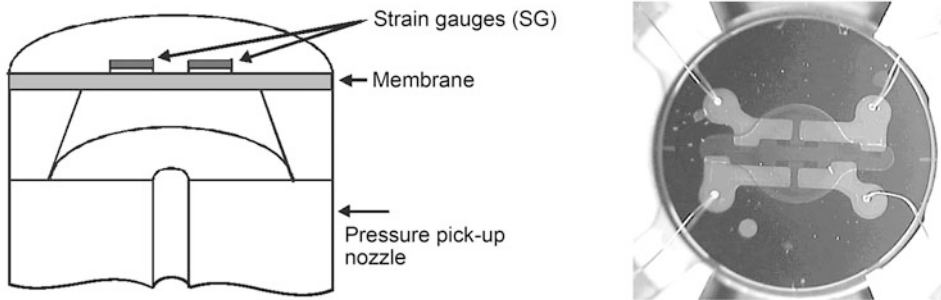


Fig. 4.24 Resistive pressure measurement via membrane deformation—schematic diagram and top view of the cell (Factory photo ADZ Nagano GmbH)

Figures 4.25 and 4.26 show sectional images of a silicon pressure sensor. The pins in the figure have a distance of 2.54 mm.

The following Figs. 4.27, 4.28 and 4.29 show examples of complete electronic *pressure transmitters*. Figure 4.27 shows a compact transmitter (sensor with signal conditioning) with a cell based on a stainless steel diaphragm (Fig. 4.24) with the required evaluation electronics. The protruding plug connector is broken off after production. Depending on the cell, it is used in the range from 1 to 2000 bar. Figure 4.29 shows a differential pressure switch for low pressure with the cell in Fig. 4.25.

Another widely used form of pressure measurement is the use of membrane deformation as a variable capacity. Here, the distance between the membrane and a reference surface is measured. The resulting capacitances are in the range of less pF and can be evaluated electronically with a resolution in the range of aF (atto Farad = 10^{-18} F).

A short overview of the possibilities of pressure measurement is shown in Table 4.4.

A fundamental disadvantage of resistive measurements is the fact that *one side of the membrane* always carries the *electronic components*. This side is therefore sensitive in contact with the medium to be measured, which causes problems in differential measurements (media contact on both sides). Furthermore, no stop for the diaphragm can be realized on this side.

With capacitive and inductive measuring methods, the mechanical construction of a diaphragm stop is possible, which provides *overload protection*, as the signal acquisition does not take place on the diaphragm. Overload protection can thus be achieved, which can be a factor of 1000 above the working pressure.

Fig. 4.25 Example of a silicon pressure sensor—top view

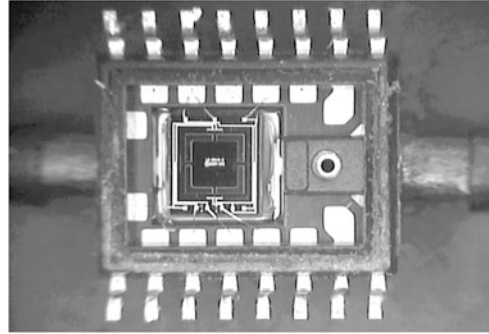


Fig. 4.26 Example of a silicon pressure sensor—sectional view

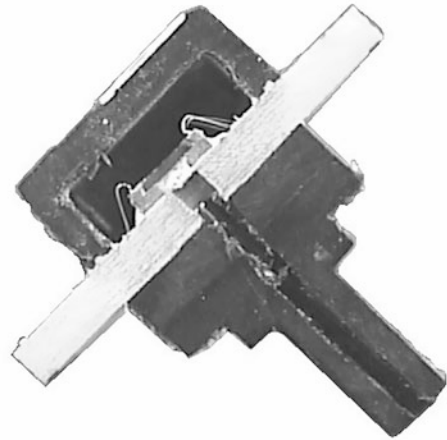


Fig. 4.27 Design of a pressure transmitter (Factory photo: Prignitz-MST.de)

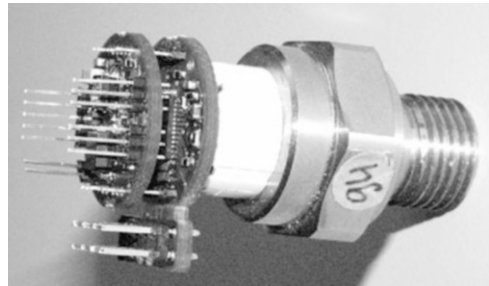


Fig. 4.28 Design of a pressure transmitter (Factory photo ADZ Nagano)

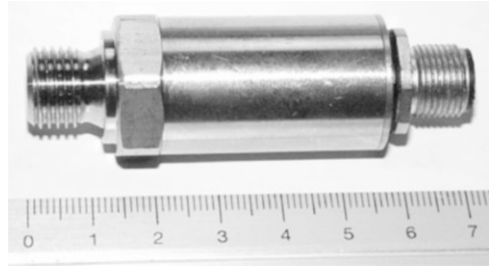
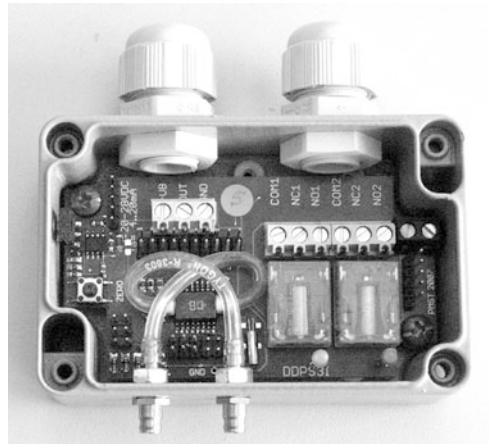


Fig. 4.29 Design of a pressure switch for low pressure (Factory photo: Prignitz-MST.de)



4.5 Torque

4.5.1 Definition

The torque M describes the force applied to rotating objects. It thus corresponds to the force during linear movements.

The unit is the Nm (Newton meter) $1 \text{ Nm} = 1 \frac{\text{kg} \cdot \text{m}^2}{\text{s}^2}$.

For drive shafts, the transmitted power P can be determined via the rotational speed n and the torque M :

$$P = 2\pi \cdot n \cdot M. \quad (4.1)$$

Table 4.4 Possible methods of pressure measurement

Measuring element	Operating principle
Diaphragm (resistive)	Mechanical deformation of the membrane. Detection of the deformation by means of resistive measuring bridges firmly coupled to the membrane. Measurement of static and dynamic pressures.
Diaphragm (capacitive)	Mechanical deformation of the membrane and detection of deformation by capacitive changes. Measurement of static and (slow) dynamic pressures.
Diaphragm (inductive)	Mechanical deformation of the membrane and detection of the deflection by inductive distance measurement.
Piezo crystal	Pressure creates a charge on the crystal surface. This charge is proportional to the pressure. Only measurement of dynamic pressures.
Quartz resonators, piezo resonators	Due to the effect of pressure on the surface, the vibrational behaviour of the elements changes. Measurement of static and dynamic pressures.
Ionization chamber	Pressure measurement by ionization of the measuring medium and the resulting conductivity. Measurement of static pressures in the high vacuum range (depending on the medium).

4.5.2 Measuring Principles

At a mass at rest, an effective torque can be achieved by measuring the force acting at a lever arm. A generally known example is the *torque wrench* for limiting or determining the tightening torque of screws. A variant of this is the design of a wrench as a bending beam (Sect. 4.3) in order to determine the applied force and thus infer the torque from the geometric design. Alternatively, the *torsional force* can also be determined in a socket wrench. Figure 4.30 shows a system based on this principle.

Another way to measure the torque is to determine the angular acceleration α . However, this requires a known moment of inertia J . It applies to the torque M :

$$M = J \cdot \alpha.$$

The transmission of torque via a shaft leads to a *twisting* of the shaft (torsion). This can be determined as strain on the surface using strain gauges. Figure 4.31 shows strain gauges specially designed for this purpose.

Fig. 4.30 Torque sensor (non-rotating) for checking screwdriving tools (Factory photo: ATP Messtechnik)

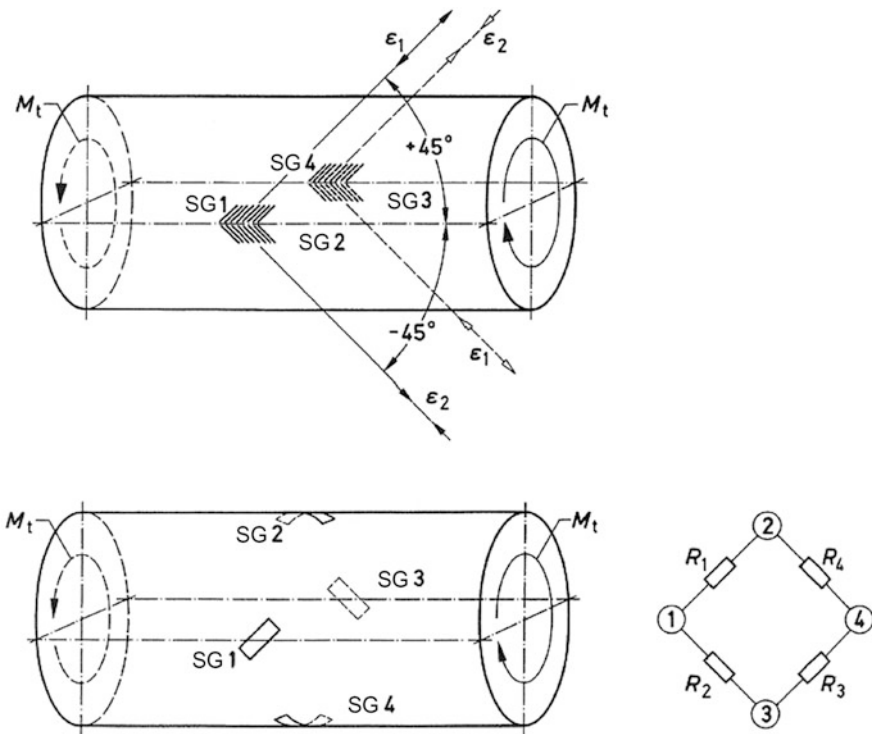


Fig. 4.31 Torsional shaft with strain gauges applied in the main directions of rotation. (a) when using special X-rosettes, (b) when using individual strain gauges (Source: Hoffmann; “Introduction to measurement technology with strain gauges”; at HBM)

4.5.3 Application Areas

Torque measurements play an important role in many areas. Figure 4.32 shows the most important areas of application of torque sensors.

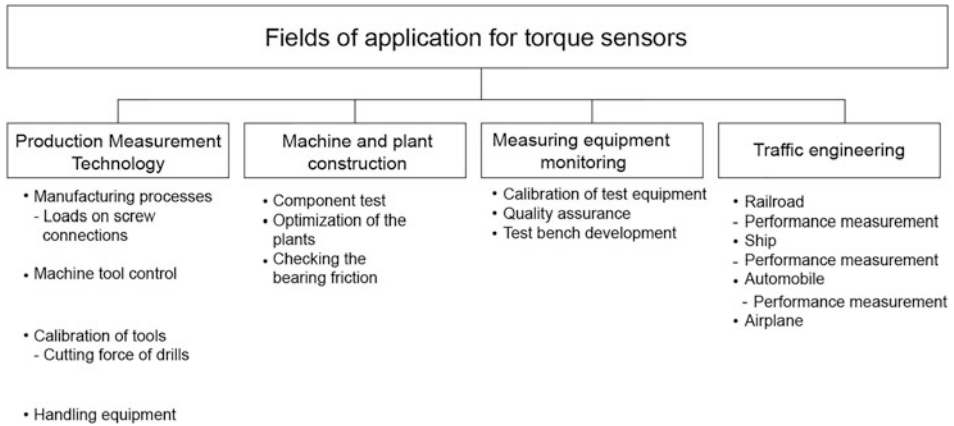


Fig. 4.32 Application fields of torque sensors

The torque can also be measured by other methods (Fig. 4.33).

Figure 4.34 shows an example of a magnetic-based application.

4.6 Hardness

4.6.1 Definition

The term *hardness* describes a material property that is phenomenologically described as “*resistance of a material to penetration by a test specimen*”. Materials can be classified according to ascending or descending hardness by means of a suitable measurement specification and a hardness definition based on this. For example, *Mohs’ hardness scale* sorts minerals in a hardness range from 1 (talc) to 10 (diamond) according to the measuring method “material A scores material B”.

The hardness values assigned to a specific material sometimes differ considerably depending on the definition. Differences in hardness of several materials on different hardness scales can also be non-linearly related. Therefore, *hardness data are primarily used to compare materials* in a certain, well-defined *class of materials* (e.g., metals, elastomers, and minerals) and are based on one and the same hardness definition that is suitable for that class.

The most common hardness definitions used today are based on the specification of a *test force F* with which a test specimen penetrates the material and a *measure of the “strength of penetration”*, for example the *depth* (Rockwell or Shore hardness) or the *projected area* (Vickers hardness) of the indentation. The *test specimens* (“*indenter*”) are usually made of carbide or diamond and have a geometry specified in the relevant standard.

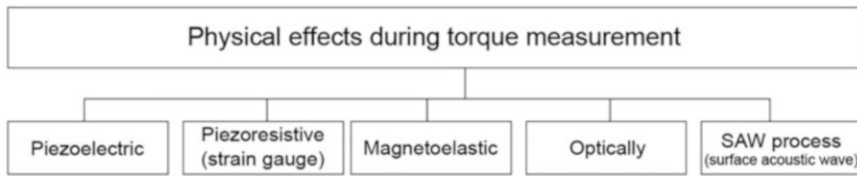


Fig. 4.33 Methods for torque measurement

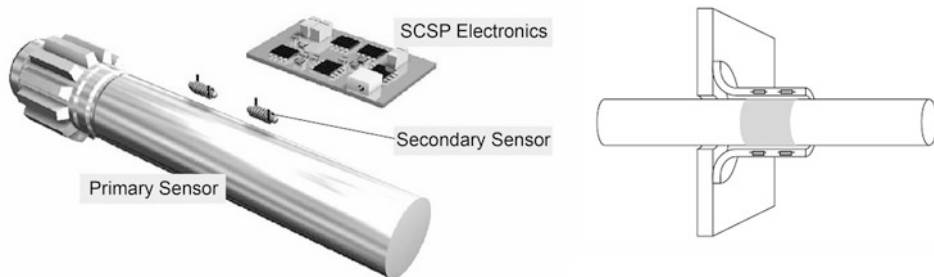


Fig. 4.34 Torque sensor based on induction (a) sensor system, (b) torque sensor for a gear shaft (Factory photo: NCT Engineering)

4.6.2 Macroscopic Hardness Determination

Depending on the field of application, the test engineer has a number of standardized methods for macroscopic hardness determination at his disposal, which were mainly developed at the beginning of the twentieth century for the control of industrially manufactured workpieces. The selection of the appropriate method depends on the respective *material*, the expected *hardness range*, the *shape* of the workpiece, and other factors. Hardness measurements according to *Brinell*, *Vickers* or *Knoop* measure the remaining indentation microscopically, while testers for *Rockwell* or *Shore hardness* contain displacement sensors for the penetration depth.

In the *Vickers test*, an *equilateral diamond pyramid* with an angle of aperture of 136° is used, while the *Knoop indenter* is a *four-sided pyramid* with diagonals of different lengths. The *Brinell* or *Rockwell* hardness test uses *carbide balls* or *conical diamond indenters*.

The *Knoop test* provides the most accurate measurements on thin films, but usually requires a polished surface of the workpiece. The *Brinell hardness* with comparatively high test force is used for rough or coarse-grained materials. So-called *Shore durometer tests* are used for *rubber* and *softer polymer materials*.

These test procedures have in common that only the *maximum test force* and the *maximum penetration depth* are measured. They therefore only provide information about the plastic deformation of the material. On the other hand, *recording hardness measurement methods* such as *Martens hardness* can also provide information about the elastic springback of the material.

4.6.3 Hardness Determination by Nanoindentation

Increasing miniaturization of components requires measurement methods where the penetration depth of the test specimen can be in the range $< 1 \mu\text{m}$. This results in practical consequences for a reasonable measurement specification of hardness in the nanometer range:

- *Geometry of the test specimen:* Three-sided pyramids are preferred because conical indenters with a very small rounding radius at the tip are difficult to produce and four-sided pyramids (e.g., Vickers indenters) often end in a roof edge rather than in a point. Both of these factors lead to incorrect determination of the indentation area, especially with low penetration depths.
- *Sample preparation:* In order to calculate the contact area from the penetration depth and the known indenter geometry, the indenter must penetrate perpendicular to the workpiece surface. The small lateral and vertical dimensions of the indentation require a very low roughness of the surface of the workpiece.
- *Combination with microscopic methods:* Samples to be examined often show inhomogeneities or topographic anomalies (e.g., trenches, scratches, grain boundaries, or inclusions). Therefore, the combination of nanoindentation with an imaging method is advantageous, in which the sample is imaged with a high spatial resolution at the location to be examined in order to determine a suitable location for the indentation. Here, atomic force microscopy is particularly noteworthy, which produces a three-dimensional map of the surface topography with a resolution in the nanometer range.
- *Measuring method for the “strength of penetration”:* As with the hardness measurement according to Martens, force F and penetration depth h are continuously measured during nanoindentation (*recording hardness test*). An exact measurement of the remaining indentation under the microscope as with Vickers hardness is not possible due to the small dimensions of the indentation. The recording hardness test, on the other hand, allows the projected contact area A_c of the indenter to be determined from the measured penetration depth h_c and the so-called indenter area function $A(h)$ in combination with a suitable model for the elastic deformation of the contact. In contrast to Vickers hardness, the recording hardness test can also be applied to materials, which deform purely elastically and leave no impression after the test load is removed (e.g., some elastomers).

The hardness H determined from nanoindentation is defined as:

$$H = F_{\max}/A_c,$$

where F_{\max} represents the maximum applied test force and the projected contact area A_c is determined from the contact depth h_c (Sect. 4.6.5). The hardness defined in this way corresponds to the average pressure on the specimen in the contact area.

Accuracy and reproducibility of nanohardness measurements are considerably influenced by experimental conditions, such as thermal drift and mechanical vibrations, due to the low penetration depth and measuring force. Nanoindent systems are therefore often equipped with an active or passive vibration isolation system and a thermo-acoustic shielding enclosure.

4.6.4 Sensors for Nanohardness Measurement

Depending on the system used, the *test force is applied* by applying a *magnetic or electric field* of known field strength or by controlled deformation of a spring. Capacitive (Sect. 2.6) or inductive (Sect. 2.5) displacement sensors are usually used to measure the displacement of the indentation.

The system in Figs. 4.35 and 4.36 represents a special feature because here the force is measured as an optically detected deflection of a calibrated spring and is kept at the setpoint value by an electronic control loop. The force resolution is below 20 nN. The nanoindenter is pressed vertically onto the sample by a piezo actuator. The penetration depth is the difference between the measured piezo movement and the spring deflection measured with the optical detector.

The entire nanoindent unit with a diameter of about 3 cm is attached to the measuring head of an atomic force microscope and integrated into its XYZ raster system so that the sample topography can be scanned with the indent tip before and after the measurement.

4.6.5 Model and Evaluation

To calculate the hardness, the projected contact area $A(h_c)$ must first be determined from the indenter surface function $A(h)$. The contact depth h_c is the difference between the maximum penetration depth and the initially unknown elastic deformation h_s (s for “sinkin”) of the surface at the contact periphery (Fig. 4.37):

$$h_c = h_{\max} - h_s.$$

During loading, the specimen is usually deformed both plastically and elastically. In contrast, the relief curve contains only information about the elastic springback of the deformed material (Fig. 4.38).

The elastic deformation is defined for a circular, flat indenter (so-called “flat punch”) is given by

Fig. 4.35 Schematic structure of the nanoindentation module

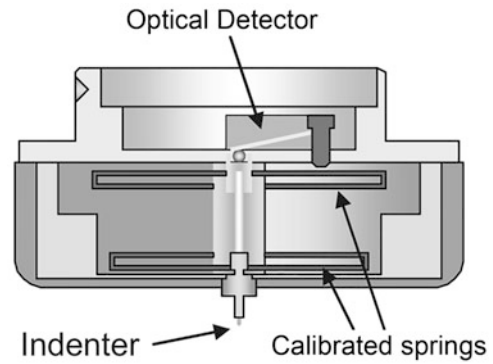
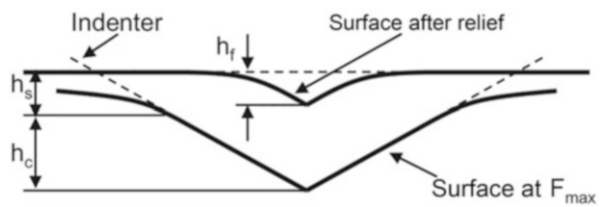


Fig. 4.36 Nanoindentation module (Factory photo: atomic force)



Fig. 4.37 Section through a nanoindentation in an elastically and plastically deformable material at maximum load and after removing the load

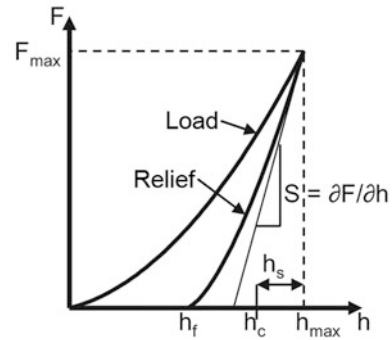


$$h_s = F_{\max}/S(h_{\max})$$

whereby the stiffness S is calculated as the slope of the unload curve:

$$S = \partial F/\partial h.$$

Fig. 4.38 Schematic representation of the measured load and unload curve with the penetration depth values h



For other indenter geometries, h_s changes by a geometry-dependent factor ε . For the most frequently used indenter geometries, the value $\varepsilon = 0.75$ is a good approximation. Thus, the contact depth is

$$h_c = h_{\max} - \varepsilon F_{\max}/S(h_{\max})$$

and finally the required hardness

$$H = F_{\max}/A(h_c).$$

The analysis assumes that no material is thrown up at the contact periphery (“pile-up”). If necessary, this assumption can be verified by imaging the impression.

The *indenter surface function* $A(h)$ is given by the geometry of the indenter. For a *Berkovich indenter* (three-sided pyramid with an angle of 142° between edge and opposite side), the relation $A(h) = 24.5 h^2$ is given. For very small penetration depths of the order of <100 nm, deviations from the ideal geometry such as the unavoidable rounding of the indenter tip must be taken into account.

4.6.6 Applications

Figure 4.39 shows a typical measurement curve on a hard material layer of a nanoindent impression only 50 nm deep with a *Berkovich indenter*. At maximum load, a waiting time of 4 seconds was inserted to wait for the plastic creep processes in this material to subside. The fit area for determining the *stiffness* $S = \partial F/\partial h$ is highlighted with a thick black line. The hardness determined from the measurement curve is 3.2 GPa.

In the atomic force microscopic image of an indent on a light metal alloy (Fig. 4.40), lateral projections can be seen, which lead to an increase in the contact area $A(h_c)$. In addition, the different heights of the projections indicate an anisotropy of the material. This example illustrates the strength of the combination of nanohardness measurement with a high-resolution image.

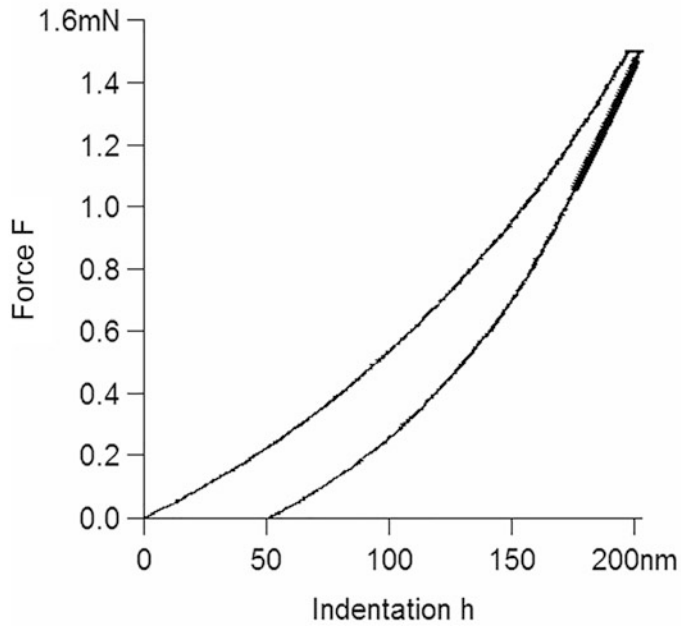


Fig. 4.39 Measurement data of a nanohardness measurement on a hard material layer

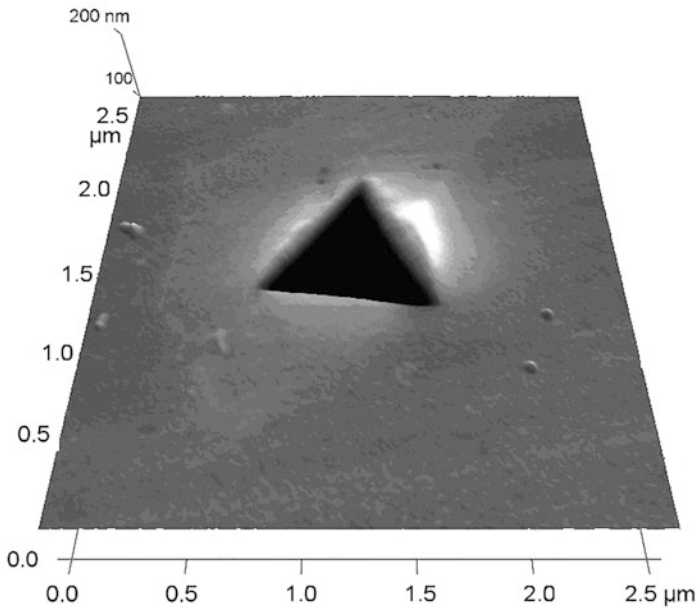


Fig. 4.40 Atomic force microscope image of the remaining impression after induction of a light metal alloy (Factory photo: atomic force)



Gert Schönfelder and Gerd Stephan

5.1 Time

5.1.1 Definition

The second is 9,192,631,770 times the period of the radiation corresponding to the transition between the two hyperfine levels of the ground state of atoms of nuclide ^{133}Cs . The time has the symbol t and the unit *second* [s]. The relative measurement uncertainty is 10^{-14} . Because of this high measurement certainty, time is the decisive reference value for many measurement methods.

5.1.2 Measuring Principles

The absolute time can only be determined by an atomic clock, which in turn is set on the basis of astrophysical processes. In order to make this reference value accessible, it is transmitted by *time signal transmitters*, which send time marks—for example second pulses—as well as absolute time information in coded form. With these signals, time and frequency references can be adjusted locally. This results in at least one reference with a high relative accuracy. For an absolute reference time, the transit times of the transmission from the time signal transmitters must be taken into account.

G. Schönfelder (✉)

Prignitz Mikrosystemtechnik, Wittenberge, Germany

G. Stephan

Quantum Hydrometrie Gesellschaft für Mess- und Systemtechnik, Berlin, Germany

In Germany, the *DCF77* works as a time signal transmitter with a carrier frequency of 77.5 kHz, which in addition to the second pulses also transmits the information about time and date every minute. For the reception of these signals, special modules are offered, which output the coded signals as a digital data stream. At the beginning of a new second, the carrier amplitude is reduced to 25%. The lowering of the 59th second is missing, so the following lowering indicates the beginning of a new minute. The length of the lowering represents a binary coding, which represents information on operating data and the current time and date. Another source for a relative time standard is information from *navigation satellites*.

In practice, an accurate *relative time reference* is generally sufficient. This is obtained by *dividing* the frequency of oscillating crystals. Quartz crystals provide a very stable oscillation, but have the disadvantage that they are thermally very sensitive. Technological and circuit measures are used to achieve accuracies in the range of 10^{-4} to 10^{-9} . The measuring arrangements are discussed in more detail in the following section on frequency.

5.2 Frequency

5.2.1 Definition

The *frequency* f is the ratio of a number of events ΔN in a defined period Δt . It applies:

$$f = \frac{\Delta N}{\Delta t} \quad [f] = \frac{1}{\text{s}} = 1[\text{Hz}].$$

The unit is [1/s] or [Hz]. As the frequency refers to the basic unit of time, the relative measurement uncertainty of 10^{-14} also applies here. It is therefore currently an equivalent measurement variable.

From a physical point of view, all periodically recurring events can be called frequency. In practice, the term frequency is only applied to *periodic oscillations*—for example, of oscillators with digital or analog output signals. For other periodic signals, which also refer to time, other terms are often used, such as

- *Baud*: *Bit per second* in a digital data stream (baud rate of serial interface).
- *sps* (samples per second): Number of samples of values per second in the measurement technology (sampling rate, sample rate).
- *rpm*: Speed in revolutions per minute (also as 1/min).
- *Bit/s*: Data transmission rate in digital technology.

5.2.2 Measuring Principles

The frequency can only be determined by *counting the periods* (number of oscillations) in a *given time*. It may be necessary to process the signal by filtering and/or triggering. The events thus discretely exposed are counted within a defined time—for example 1 second. The reference time required for this is obtained by dividing the frequency of an oscillating crystal. As will be shown subsequently in the error analysis, it is important to select the measurement time in such a way that a large number of events are recorded during the measurement.

5.2.3 Measuring Arrangements for Frequency and Time Measurement

Figure 5.1 shows the principle of a frequency measurement.

The pulses applied with the frequency f_x are detected by closing the switch for a defined period of time Δt with the counter. The result is one pulse number per time. Thus, ΔN is determined in a specified time $\Delta t (f = \Delta t / \Delta t)$.

If you exchange the reference value and count the pulses of a known frequency for an unknown time, you get a time measurement. The number of periods, i.e. the frequency during the closing time Δt of the switch is recorded (Fig. 5.2).

From these arrangements, the measurement of a frequency ratio can be easily derived, since the frequency measurement is basically a ratio determination. It has been described that the *gate time* is obtained by dividing the frequency of a reference generator. Since the counter can only display integer events, the two frequencies must be far enough apart, which is achieved by dividing the gate-controlling frequency. The resulting switching scheme is shown in Fig. 5.3.

5.2.4 Measurement Error of Time-Discrete Measurements

Since there is no time relationship between the two input signals during frequency measurement, the measurement result may fluctuate by ± 1 . The cause is the random position of the signal edges relative to each other. As Fig. 5.4 shows, depending on the position of the gate time, one pulse more or less can be detected. Since it is not known which of the two cases occurred during the actual measurement, a deviation is expected in both directions. This error, which always occurs, is called *digital residual error*. Since it always occurs with a value of ± 1 in relation to the total number of measured events, it decreases with an increasing number of measured pulses. It is represented by the ratio of this error pulse to the total number of pulses.

The total relative error F_{rel} of a time or frequency measurement is thus composed of the *error of the normal frequency* f_{rel} and the digital residual error. Since these do not correlate, they are added together quadratically. The effect of the digital residual error decreases as

Fig. 5.1 Basic principle of frequency measurement

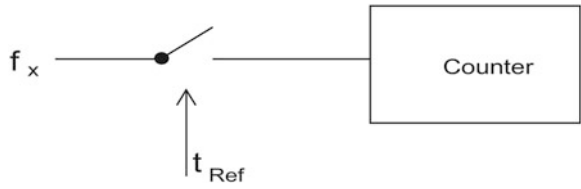


Fig. 5.2 Basic principle of time measurement

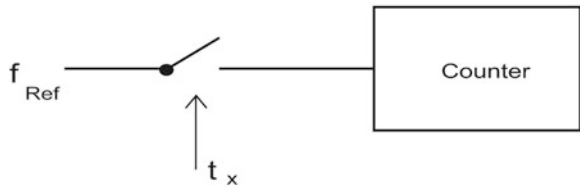


Fig. 5.3 Basic principle of frequency ratio measurement

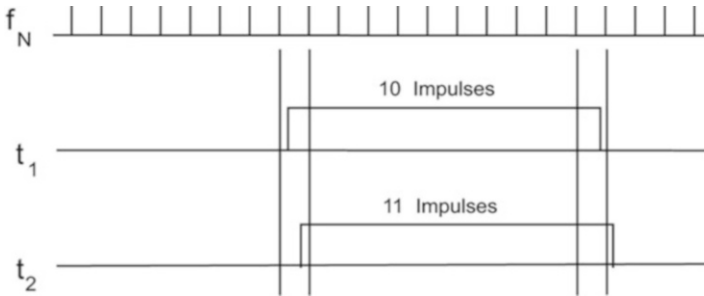
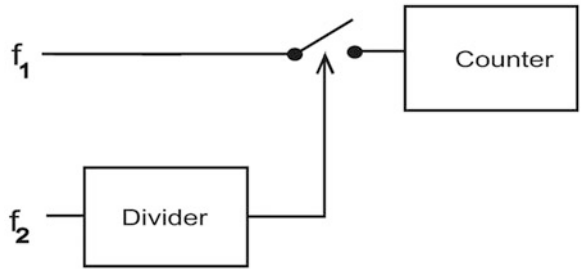


Fig. 5.4 Display of the digital residual error

the measured value increases, so that it is applied with the reciprocal of the measured value. It therefore applies:

$$F_{\text{rel}} = \sqrt{\frac{1}{\text{measured value}^2} + F_{\text{rel}}(f_{\text{ref}})^2}. \tag{5.1}$$

With this error consideration, it is also possible to determine the point at which one should switch from the frequency measurement to the period duration measurement, because the measured value for the frequency is too small and thus the error influence is too large.

Quartz crystals are used for the *frequency standards that are* common in practice. For these, the following accuracies can be assumed:

Watch quartz	10^{-4}	to 2×10^{-5}
Selected crystals		to 5×10^{-6}
Quartz and oscillator temperature stabilized		to 10^{-8}
Quartz standards	10^{-9}	to 10^{-11}

5.2.5 Measuring Arrangements

In practice, these measuring tasks are solved with meters, which are offered as devices. Since they have standard interfaces, they can be integrated into measuring arrangements at any time. They have at least the basic functions of frequency and time measurement and offer the possibility of *event counting*—which corresponds to a measurement time of infinity. The highly accurate reference frequency contained in the device should match the display range, i.e. a device with a 7-digit display should have an accuracy of at least 10^{-8} in the reference frequency.

With comfortable systems, an additional input also provides the possibility to solve further measuring tasks. These include *frequency ratio measurement* and the *pulse and phase measurement* functions described in the following sections.

Figure 5.5 shows the circuit of a simple counter control. It allows the measurement of frequency, period duration, pulse length, and counting of events. Since the input signals are rarely square-wave signals in practice, triggering, i.e. conversion into a square-wave signal, is required. The trigger threshold is set via the potentiometers. In switch position 1, the measuring arrangement corresponds to Fig. 5.1. The provided normal frequency is divided by the divider $1/N$, resulting in a gate time of 1s, for example. The measuring signal is fed to input A and the number of periods is recorded. The displayed result thus corresponds to the frequency of the measurement signal in [Hz].

Switch position 2 results in a time or period duration measurement according to Fig. 5.2. For this purpose, the normal frequency is divided little or not at all and is applied to the counting input, for example at 1 MHz. The measuring signal from input A opens the gate to the counter with the rising edge and closes it with the following rising edge. As a result, the display shows the number of μs ($1/1 \text{ MHz}$) in which the gate was open, which in this case corresponds to the period duration of the input signal. To measure pulse durations, the combination of triggering edges must be varied in the gate control (e.g. from rising to falling edge).

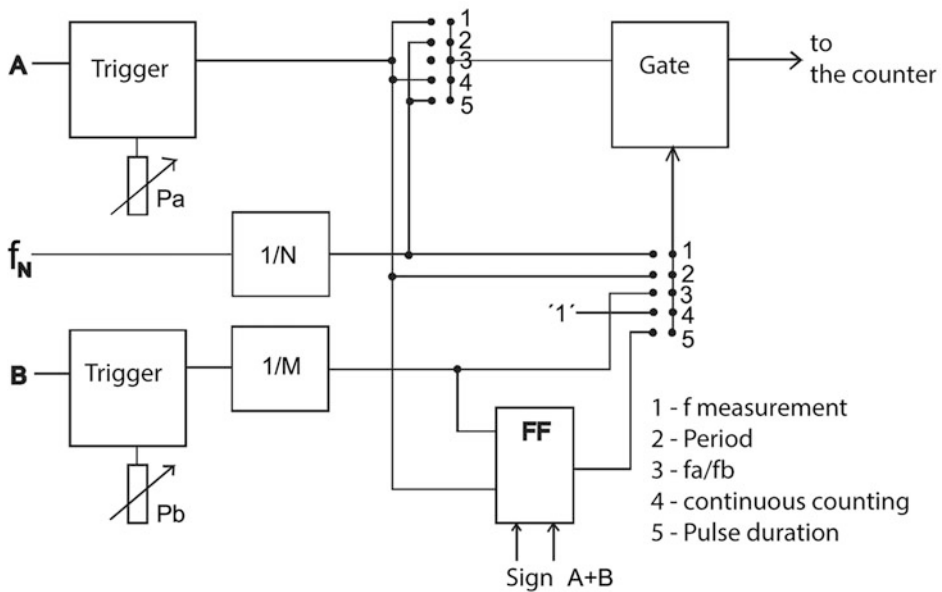


Fig. 5.5 Block diagram of a control system for digital counters

In switch position 3, the ratio of frequency A to frequency B is formed. To obtain a sufficient display value, path B is reduced by a divider $1/M$. The circuit thus corresponds to Fig. 5.3.

In switch position 4, the gate is kept permanently open, which serves to count periodic or non-periodic events.

Switch position 5 technically corresponds to position 2, but there are separate inputs for opening and closing the gate circuit. Thus, by combining the trigger polarity and the switching thresholds (levels P_a and P_b), further values such as phase, running time, edge rise time, and pulse duration can be determined for “low” and “high”.

Figure 5.6 shows the front of a counter. This is a simple version without channel B. Therefore, no frequency ratio measurements can be carried out.

An alternative to the counter is the use of *microcontrollers* for frequency and time measurement. These generally contain a *capture unit*, which has a lower resolution and does not offer the comfort of signal processing. They simulate the circuit of a counter for frequency measurement. Derivatives of the controller's quartz crystal are used as reference frequency, which means that it is not divided by decades. Their use is particularly suitable as a low-cost alternative for small resolutions and defined expected values.

Until now, only the digital measurement of the frequency has been dealt with. But there are also classical analog methods for this. A simple analog circuit results when a *monoflop* is triggered with the measurement signal. A monoflop reacts to a signal edge and generates a pulse of a defined length. This pulse must be very short compared to the maximum

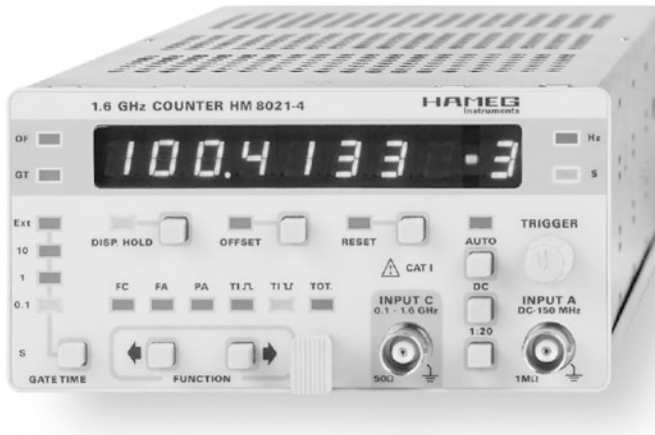


Fig. 5.6 Digital counter (Factory photo: HAMEG)

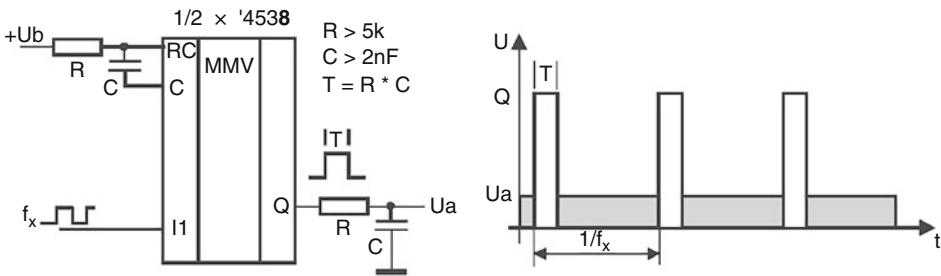


Fig. 5.7 Frequency measurement by frequency-to-voltage conversion

expected period duration and shorter than the minimum expected period duration. This results in pulse sequences at output Q, whose interval varies with the period duration. The integration of this sequence by the ohmic resistance R and the capacitance C then results in an output voltage, which is proportional to the frequency—a *frequency-to-voltage conversion* was carried out (Fig. 5.7).

The disadvantage of this arrangement is its low accuracy in the order of 10 bits. It also deteriorates due to the pulse width in the direction of high measuring frequencies. Its use is therefore suitable in the low frequency range.

An old analog method with good accuracy is the measurement of the *beat*. Here, the measured variable is subtracted from a reference frequency by means of a mixer. By tuning the reference, the result is set to zero, so that the measured variable is equal to the reference frequency. The calibrated generators required for this purpose can be easily realized with modern technology. For some years now, generators based on *DDS (direct digital synthesis) circuits have been offered*, which allow the generation of sinusoidal oscillations with

step sizes in the mHz range. Since the oscillation generation is based on a quartz reference, they achieve a *high accuracy* and *stability*. Examples of these circuits can be found at www.analog.com. One example is the AD9833, whose block diagram is shown in Fig. 5.8.

In the functionally small *DDS system* shown in Fig. 5.8, the main components are clearly visible in the horizontal central axis. They consist of the *phase accumulator*, which consists of a large *adder*. The frequency step $FREQ0$ is added to its content, resulting in the next interpolation point in the sine interpolation. The upper 12 bits of the accumulator are transmitted to the sine generator, which drives the DA converter. The output frequency results from this:

$$F_{out} = \frac{MCLK \cdot FREQ0}{2^{28}}.$$

The value of $FREQ0$ must be less than 2^{27} , since the *sampling theorem* also applies here.

5.3 Pulse Width

5.3.1 Definition

The *pulse width* P_w is the ratio of the duration of the high part of a square wave t_{pulse} to the period t_{period} (Fig. 5.9). It is the ratio of two time quantities and is therefore dimensionless. It applies:

$$P_w = \frac{t_{pulse}}{t_{period}}.$$

The result is in the range of $P_w = [0, 1]$.

In process technology, it is frequently encountered under the term *PWM (pulse width modulation)* because it offers a number of advantages in signal transmission:

- As a digital signal, it can be *easily* reconstructed at the *place of reception*;
- It does not require a reference quantity at the point of reception, since it is a ratio quantity;
- The information stored in the *pulse width can be* extracted *digitally* as time information as well as *analogously* by integrating the signal.

5.3.2 Measuring Principles

The pulse width can be evaluated with analog and digital methods.

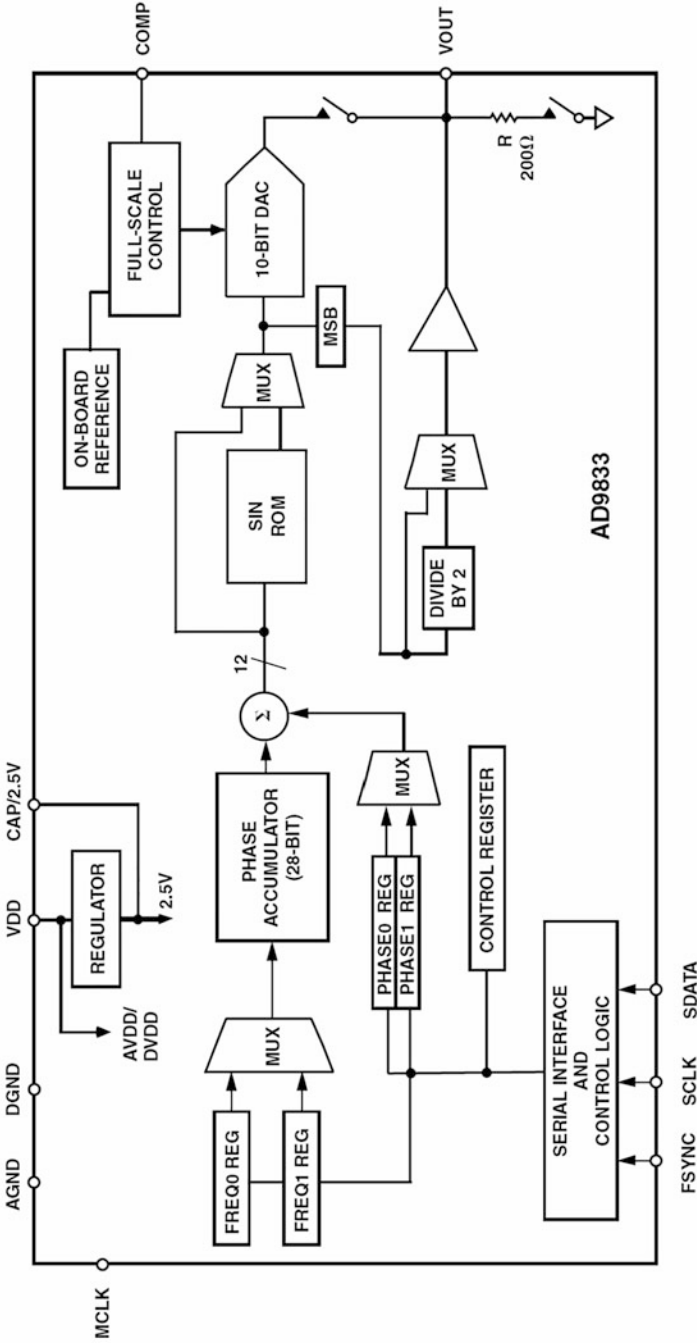


Fig. 5.8 Block diagram of the AD9833 (Factory photo: Analog Devices)

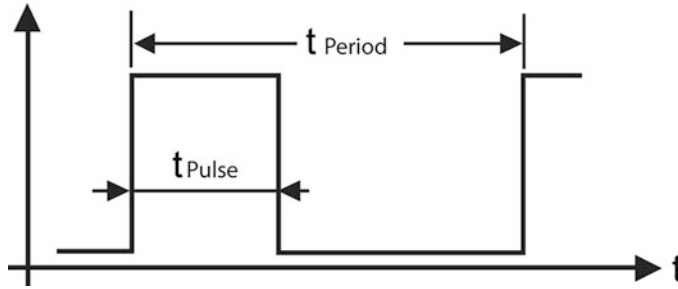


Fig. 5.9 Definition of the pulse width

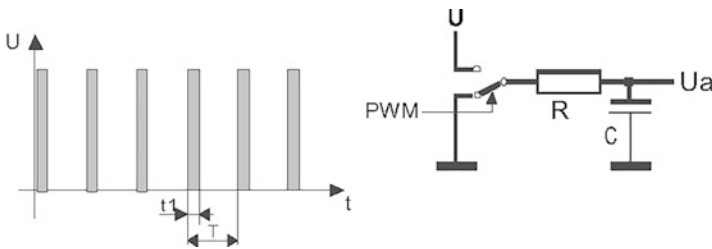


Fig. 5.10 Basic principle of pulse width measurement

5.3.3 Analogue Evaluation

The PWM signal is fed to a low-pass filter for demodulation. The result is the mean value of the pulse areas, related to the period duration (Fig. 5.10).

$$U_a = U_{\text{pulse}} \cdot \frac{t_{\text{pulse}}}{t_{\text{period}}}$$

U_a :	Output voltage after integration
U_{pulse} :	Signal amplitude of the square wave signal
t_{pulse} :	Duration of the pulse
t_{period} :	Period duration of the square wave signal

The same applies to the current as a transmission variable. For the measurement, the signal can be amplified and limited, since the amplitude carries no information.

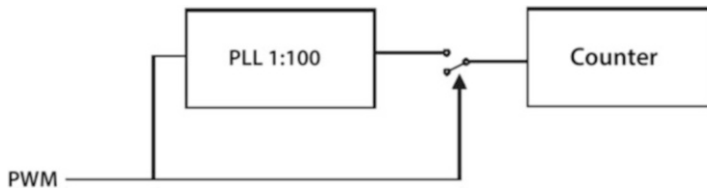


Fig. 5.11 Reference formation by a PLL stage

5.3.4 Digital Evaluation

For the digital measurement, a *pulse duration measurement* and a *period duration measurement* are carried out (Sect. 5.2). These two values give the *pulse ratio* when divided. If the period duration is sufficiently stable, its measurement can be omitted. These measurements can be performed with a frequency counter or with the capture unit of a microcontroller.

5.3.5 Measuring Arrangements

The analog version of pulse width measurement is shown in Fig. 5.10. The measurement signal is brought to a fixed level by triggering or switching between a reference voltage U and a reference voltage (ground). The amplitude of the measurement signal is therefore no longer important. By integrating the pulse train, an output voltage proportional to the pulse width is obtained, which varies between 0 V and the applied reference voltage.

During the digital evaluation of a PWM signal, two time measurements are made: that of the *pulse width* and that of the *period duration* of the measurement signal. If the double measurement and calculation is to be omitted, the measurement can be traced back to a time ratio measurement. For this purpose, the pulse itself is used to open the door on a digital counter. The pulses of the time base are obtained from the period duration of the measurement signal by means of a PLL (*phase-locked loop*). With the arrangement shown in Fig. 5.11, this would result in a display in percent.

Since in this case the reference frequency and the measuring signal are coupled in a phase-locked manner, the digital residual error does not occur here.

5.4 Phase, Running Time, and Light Running Time

5.4.1 Definition

The *phase* and the *delay time* describe the *time offset* of a signal U_2 with respect to a reference signal U_1 (Fig. 5.12).

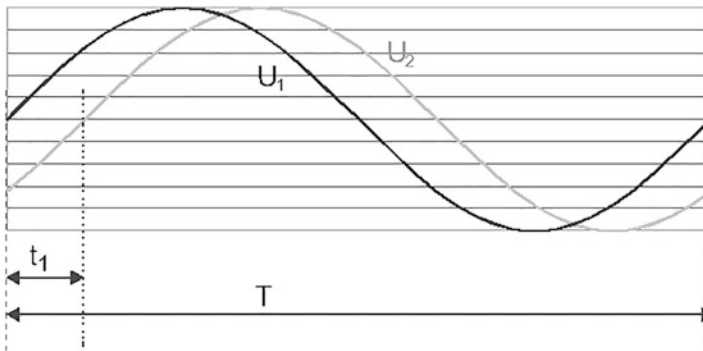


Fig. 5.12 Phase offset of periodic signals

One speaks of a *phase shift* with *periodic harmonic oscillations*. Derived from the angular frequency $\omega = 2\pi f$, the indication is in degrees or in relation to π , where $2\pi = 360^\circ$. The term *transit time* is used when the *signals are not sinusoidal* or *exceed the time of a period*. Time is used as the unit of measurement.

5.4.2 Measuring Principles

The phase can be determined by using counters with two separate inputs for start and stop. For this, the period duration must be known or also measured. For the running time, only the *signal offset* is determined. The measuring arrangements correspond to those in Sect. 5.3.

Also important is the use of *oscilloscopes* (Fig. 5.28), which allow a visual representation of the signal characteristics. With modern digital instruments, all parameters determined in this way can also be queried numerically. The measurement possibilities of oscilloscopes are discussed in detail in Sect. 5.5.

A special way is the use of circuits for time measurement with resolutions in the range up to 20 ps, as offered by ACAM (www.acam.de). Figure 5.13 shows a variant for measuring very short times. The running time of logic gates can be set in a defined manner and is in the range around 700 ps. Two chains of gates with different propagation times are used for *relative propagation time measurement*.

For the functional representation of the “Start” chain, a running time of 700 ps per gate is assumed, and for the “Stop” chain, one of 690 ps. If a 0-1 edge is applied to “Start”, it will propagate from input to input with a running time of 700 ps. If there is a delay of the measured value and this 0-1 transition is also applied to the “Stop” chain, then the “1’s” that have been applied to the flip-flops in the meantime are taken over by the delay time. At the point where the lower chain has caught up with the upper ones due to its shorter running time, no 1-values are stored, but only the 0-states. From the position of this intersection

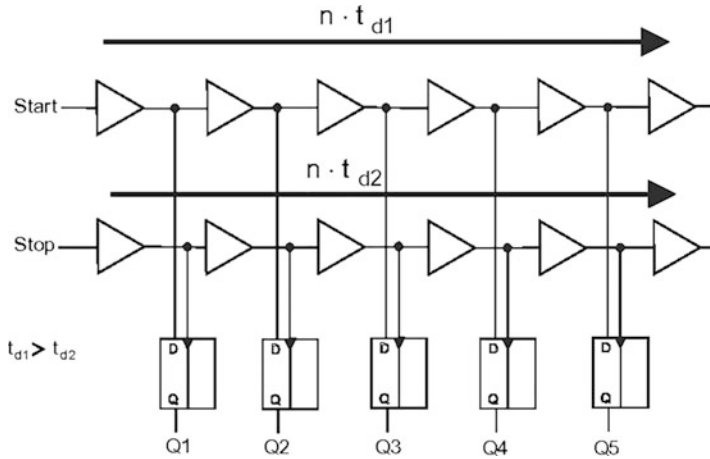


Fig. 5.13 Principle of the relative Time Delay Converter (TDC)

point and the known running times, the *time difference* between start and stop can then be deduced—it results at position runtime difference of the gates.

5.4.3 Measuring Arrangements

Another simple way of phase measurement is the generation of a PWM signal from the phase offset, which can be easily evaluated in analog form. However, this only makes sense for periodic signals (Fig. 5.14).

For the signals to be acquired, trigger pulses are derived from the positive zero crossings. The first pulse (from U_1) then switches on an output signal, which is reset by the pulse from U_2 . Since this process is repeated with the period T , an output pulse sequence is obtained whose length depends on the phase offset. If this pulse is filtered (integrated) by a low-pass filter, an output signal is obtained, which is proportional to the phase offset (Sect. 5.3).

5.4.4 Light Running Time

Due to its physical properties, the propagation time of light represents a separate field. Light propagates in a vacuum and in gaseous media at the speed of light c . It applies:

$$c = 2.99792458 \cdot 10^8 \text{ m/s (around } 300,000,000 \text{ m/s).}$$

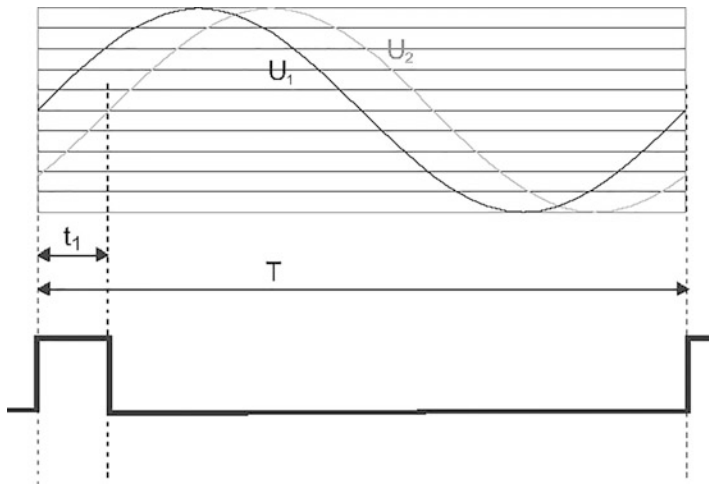


Fig. 5.14 Derivation of a PWM signal from a phase offset

This independence of propagation in these media allows the use of light for *distance measurement*. Three technically common methods will be briefly discussed here:

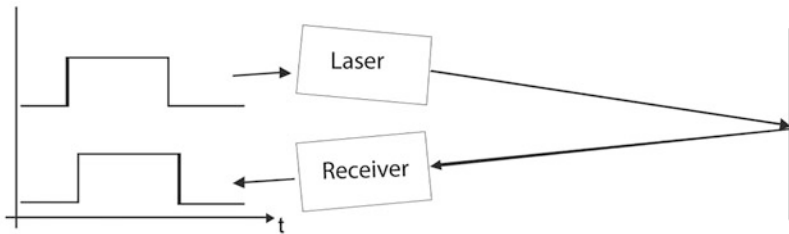
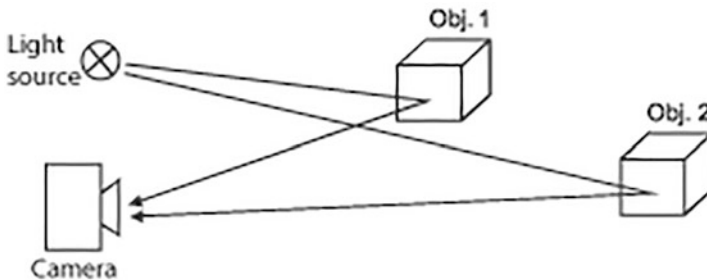
1. The *direct* measurement of running time by time measurement,
2. The transit time measurement by *pulse integration*, and
3. The transit time measurement with *modulated light* (phase measurement).

It should be remembered here that electrons in conductors also move at the speed of light. For measurements in these physical areas, the transit times within the circuits and components are already important for all arrangements.

5.4.5 Direct Time-of-Flight Measurement

In direct measurement, a *light pulse* is emitted, e.g. with a laser, whose *reflected beam* is received by a sensor (Fig. 5.15).

A few figures should show the feasibility: You can see very quickly that *small distances* result in *very small times*. As an example: A distance of 30 cm is covered in only 1 ns. With the circuit TDC-GP2 from ACAM (www.acam.de), time measurements in the range of 3.5 ns to 1.8 μ s can be performed with a step size of 65 ps. This corresponds to an object distance of 52.5 cm to 270 m at a resolution of about 1 cm. This shows that with the available technical means, a good resolution is possible.

Direct time-of-flight measurement**Fig. 5.15** Direct runtime measurement**Fig. 5.16** Light time-of-flight measurement by pulse integration**5.4.6 Transit Time Measurement Through Pulse Integration**

A method used in connection with image sensors is *distance measurement by integrating light pulses* (Figs. 5.16 and 5.17). A pulse is sent from the light source to the object and at the same time the integration, i.e. the exposure time, is started in the image sensor. The integration of the signal begins with the arrival of the reflected light. In the case of the more distant object 2, this process is delayed by the longer running time, which leads to a lower output voltage at the end of the integration time. The *output voltage* is thus *inversely proportional* to the *distance* of the object.

Figure 5.17 shows the course of the voltages on the *pixels* of the camera, whereby the basic brightness is neglected. When the exposure starts, the additional light source is switched on. But since the light has a way to travel, the reflection of object 1 arrives shortly after the start of the exposure time and is integrated in the sensor. The reflection of object 2 arrives later, so that only a smaller charge can accumulate until the end of the exposure. The reflected light beams have a “follow-up time”, which is not taken into account in the measurement.

For practical applications, a measurement consists of several partial measurements to compensate tolerances and to calculate the backlight. In a first measurement, an image is

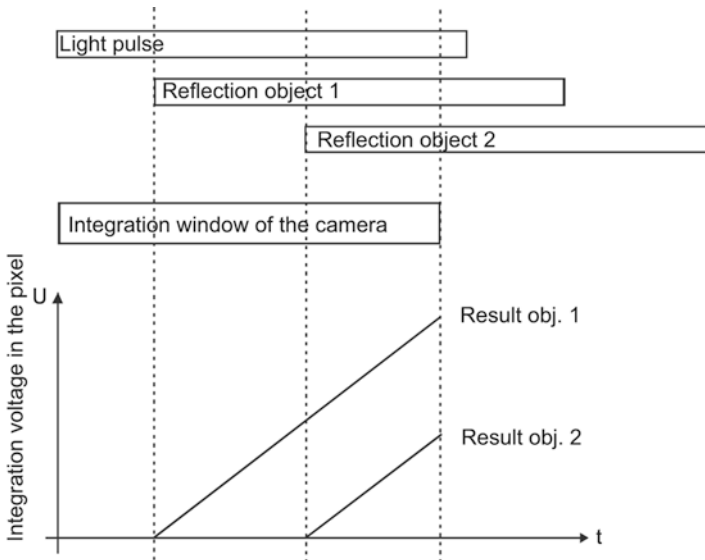


Fig. 5.17 Light time-of-flight measurement by pulse integration

Fig. 5.18 Rangefinder with time-of-flight measurement (Factory photo: Pepperl+Fuchs)



taken without the light pulse. Thereby, one gets a normal camera image. In a second exposure, which immediately follows, the light pulse is added, whereby the changes in brightness in the individual pixels are proportional to the distance of the object. Thus, in addition to the visual information, one also obtains the information about the *spatial distribution* of the objects and their movement in space. This is important in the field of security technology, for example.

Figure 5.18 shows one such sensor. It is relatively large, measuring approximately $100 \text{ mm} \times 140 \text{ mm} \times 170 \text{ mm}$.

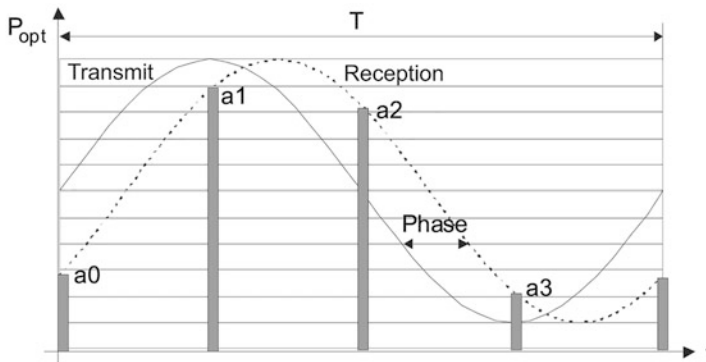


Fig. 5.19 Light travel time measurement with modulated light

5.4.7 Time of Flight Measurement with Modulated Light

Pulse integration is a *discontinuous measuring method* in which modulated light is measured *continuously*. In an arrangement as shown in Fig. 5.15, the light source is modulated with a sinusoidal signal. The receiver receives this signal offset by the propagation time, whereby the propagation time is expressed as a phase offset from the transmitted signal.

Figure 5.19 shows this process. The process “Send” corresponds to the control of the laser diode. In relation to this signal, the scanning points for the receive channel are defined—here at 0° , 90° , 180° , and 270° . At these points, the measurement of the amplitude on the receiving channel is carried out (“Receive” process). Using a *discrete Fourier transformation*, the phase φ can be determined from four sampling points within one period. It results from the amplitudes a_0 , a_1 , a_2 and a_3 :

$$\varphi = \arctan \left(\frac{a_0 - a_2}{a_1 - a_3} \right).$$

The distance d of the object is then calculated from the modulation frequency f_{mod} , and thus the period duration of the modulation, and the phase shift φ :

$$d = \frac{c \cdot \varphi}{4\pi f_{\text{mod}}} \quad \text{with the maximum distance of } d_{\text{max}} < \frac{c}{2f_{\text{mod}}}$$

with the speed of light c in the medium used.

At a typical modulation frequency of 20 MHz, this results in a maximum distance of 15 m. With a phase measurement resolution of 1° , a distance resolution of about 4 cm is then achieved.

5.5 Visual Representation of Measured Variables

For all measured variables that change over time, the aim is to record not only their characteristic values but also their *development over time*. The system must be able to record the time course of currents and voltages and to display them under given conditions. In order to see *relations* between signals in particular, the measurement must be *multi-channel*.

This section deals with the electronic representation of signals by *oscilloscopes*, also known as *oscillographs*. The area of *signal recorders*, which also allow the representation of slow temporal signal characteristics, is excluded here.

5.5.1 Measuring Principle

The technical basis is the deflectability of *electron beams* by *electromagnetic* and *electrostatic* fields (Sect. 7.5). If two such systems are arranged at right angles to each other, the beam can be deflected in the X and Y direction. The *time function* is assigned to the X-axis and the *measured value* to the Y-axis according to the mathematical representation. Since the deflection voltages at the picture tube are in the range of 100 V, appropriate amplifications are necessary. Figure 5.20 shows the block diagram of an oscilloscope.

The branch of the Y-deflection can consist of several input stages, which are sequentially displayed in a *multiplex procedure*.

The deflection in X-direction is done by means of a *sawtooth generator*. Its slew rate is adjustable and thus allows the *display period to be changed*. Since the image build-up has to be repeated constantly in order to be easily visually perceptible, the time deflection has to be adjusted so that the recorded curves are exactly on top of each other. This is achieved by *triggering* the time deflection at a defined point of the curve. The trigger signal is only effective when the generator is stopped.

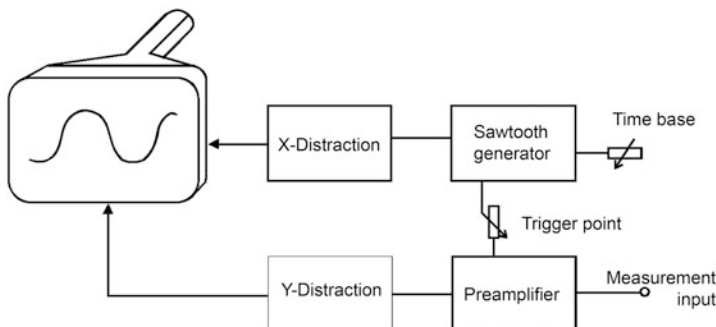


Fig. 5.20 Block diagram of an oscilloscope

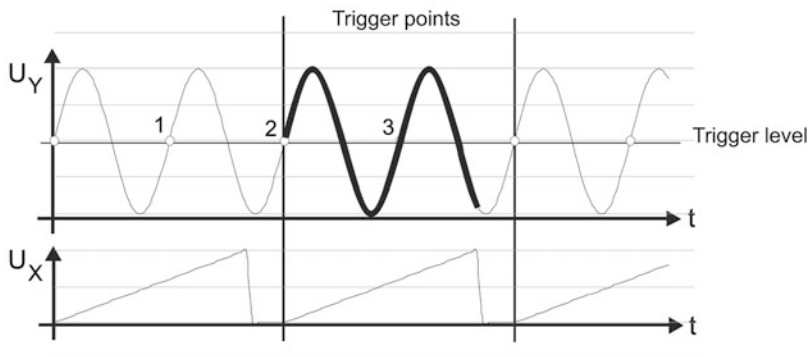


Fig. 5.21 Principle of signal triggering

Figure 5.21 shows the principle of signal triggering. A *trigger pulse* is triggered when the measuring signal intersects the *adjustable trigger threshold* in a specified direction. Since the X-deflection only reacts to this in the stop state, pulse 1 (and 3) has no effect. With trigger pulse 2, a new deflection process is started and the bold part of the input signal is drawn. The display ends with the falling edge of the deflection voltage and starts again in the example at each even trigger point.

This described operating principle is the basis of all oscillographs, regardless of their technical design. Although the continuous signal path no longer exists in modern digital devices, the result is the same.

5.5.2 Analogue Oscillographs

From today's point of view, the "old" analogue devices have a number of technical disadvantages, some of which are mentioned as examples:

- Unfavourable trigger behaviour with more complex signals;
- The representation of unique processes is almost impossible or requires special equipment;
- The determination of signal parameters is not possible or only possible with great effort
- Deteriorating image quality at high frequencies, because the beam dwell time on the luminescent layer becomes too short
- Rarely more than 2 channels are available, because the measurement quality decreases due to the beam multiplex. The exception here is image recorders for low-frequency processes, such as for ECG.

However, an analog oscilloscope has one advantage that should not be underestimated—it shows the *real signal*. Since the signal path in the oscilloscope is only limited by its cut-off frequency, everything below it in terms of frequency is

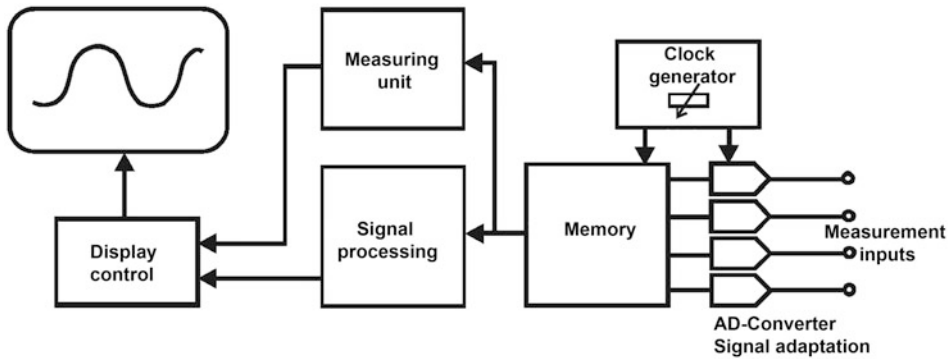


Fig. 5.22 Principle of a DSO

displayed. In a digital system, due to the time-discrete sampling of the input variable, it is only possible to make spikes or high-frequency superimpositions visible with great effort.

5.5.3 Digital Storage Oscilloscopes (DSOs)

Digital oscillographs are technically always storage oscillographs. They achieve the display by a kind of offline operation, which divides the signal path into the steps of data acquisition, processing, and display (Fig. 5.22).

The measurement inputs go through a *level adjustment process* in which offset and gain are adjusted. In the subsequent AD converter stages, the signal is converted into a digital value, which is sequentially stored in the memory. The “X-deflection” is replaced here by a clock generator, which controls the signal acquisition. Sampling is performed equidistantly in a roughly stepped grid. The data of all channels are stored in parallel in the memory, whereby data volumes in the range of megapoints (Mpt) per channel, i.e. millions of measurement points, are not unusual. This method also makes it possible to search for information in the memory at a later date and to optimize the settings accordingly.

The two units access the memory for visual display and metrological evaluation. For display, the associated areas of all channels are extracted from the memory and displayed. The data can be compressed or stretched in the time domain. The measuring unit evaluates the memory contents numerically, i.e. times between certain events or the amplitudes of signals are determined. These functions are often very extensive and can include

- Frequency	- Amplitude	- Data analysis of serial data streams
- Period duration	- Signal average	- Mathematical relations between the channels
- Phase	- Peak voltage	
- Rise times	- Minima/Maxima	- FFT (fast Fourier transformation)
- Pulse duration	- RMS (effective value)	

The display control corresponds to that of a computer. Strictly speaking, a modern DSO is a *computer* with an *analog data acquisition card*. Against this background, all additional functions such as data output to storage media, printer connection, integration into the network and output of all measurement, and image data via interfaces appear as a logical consequence.

The function blocks of a DSO (or MSO) are clearly visible on the device (Fig. 5.23). In the lower right part are the input stages, which allow the adjustment of offset and gain. Above them is the trigger setting (right) and the selection of various additional functions, such as saving image data, calling up measurement functions, and selecting operating modes. The selection of the “deflection frequency” is now only done via a button. The image display is located on the left side. This, including the measuring functions, is controlled via the 6 buttons, which are assigned depending on the image content.

This technological step has made it possible to include an oscilloscope as a normal measuring instrument in the signal and processing chain.

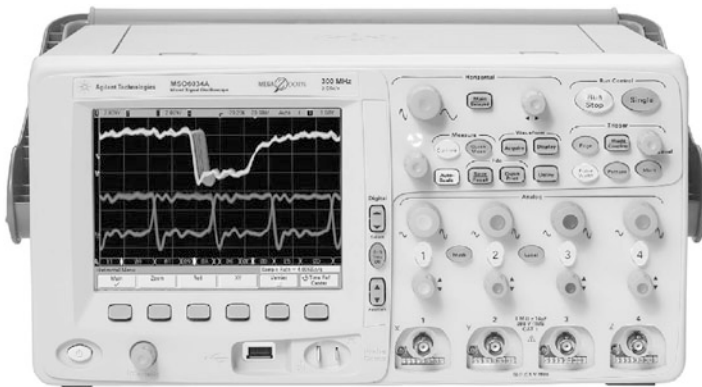


Fig. 5.23 Oscilloscope (Factory photo: Agilent Technologies)

5.5.4 Mixed Signal Oscilloscopes (MSOs)

The MSO differs from the DSO in that the memory input is directly available to the analog inputs. Digital signals, for which only the time allocation is of interest, can be recorded after triggering. The MSO thus represents a fusion of DSO and *logic analyzer*. Usually combinations of 2 or 4 analog with 8 or 16 digital inputs are used.

5.5.5 The Sampling Principle

It is technically not possible to realize a sampling rate in the range of GHz directly with an oscilloscope, because the signal runtimes on a board and the access times of memory circuits do not allow this. The access time of discrete memories is in the range of 5 ns. All higher sampling rates are achieved by integration in special circuits, time-nested parallel structures, or by the *sampling method*.

For sampling a fast signal, “wandering” *trigger points* are generated from two superimposed sawtooth functions. Sawtooth 1 runs synchronously with the vibration to be detected. A second sawtooth rises considerably slower—in Fig. 5.24 with 1/12th of sawtooth 1. The intersection points of both generators in the rising part of the flanks trigger the sampling. As they move, the scanning is performed at a different point of the measuring signal in each period. Together these result in a new function, which is slower than the original by the quotient of the rise times of the two sawtooth. In numbers, this means for the example shown in the figure that an input signal of 1 GHz is reduced to about 83 MHz, which can be easily detected and processed. If one wanted to capture the signal directly, the sampling theorem would require at least 2, practically rather 5 samples per period. The disadvantage of this measuring method is that it requires strictly periodic signals. Also, the

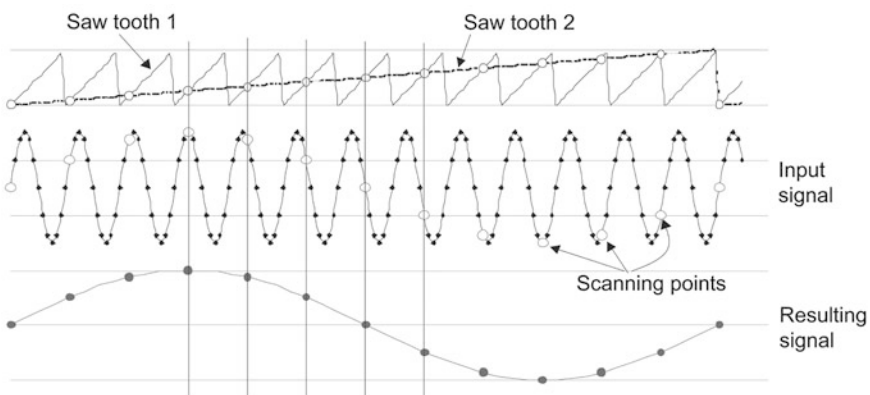


Fig. 5.24 Sampling principle for sampling fast signals

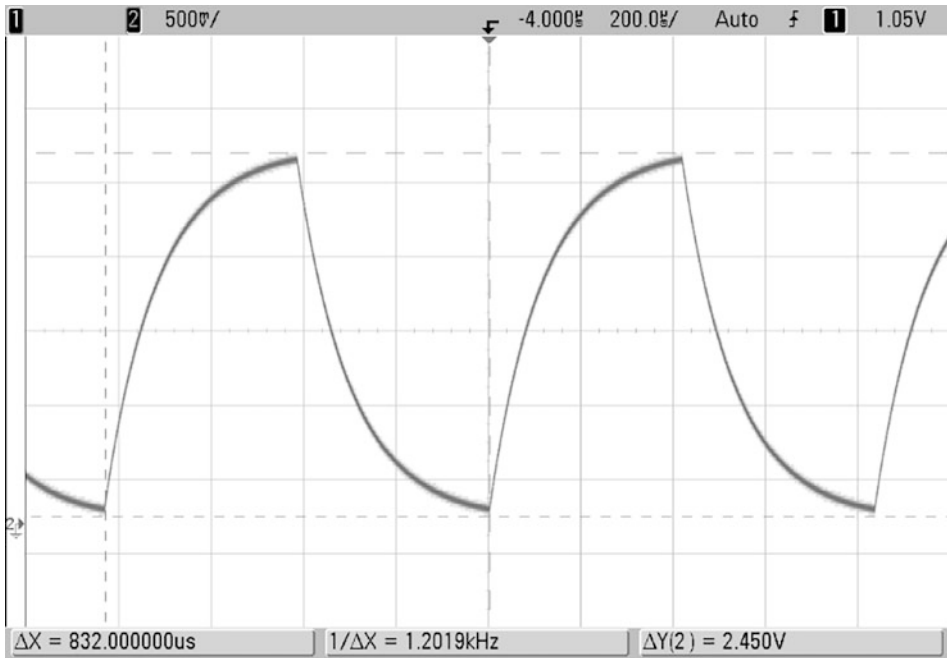


Fig. 5.25 Amplitude and period measurement

probability that interference is not detected is very high here. This method is also used, for example, to determine *rise times* of circuit outputs (Fig. 5.25).

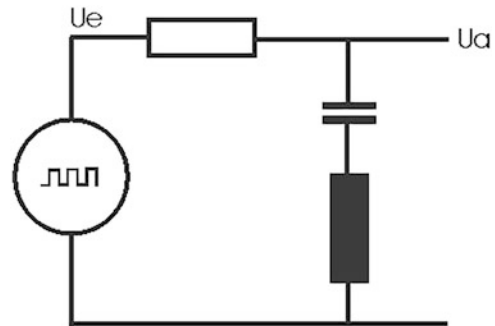
5.5.6 Measuring Arrangements

In the following part, some examples of measurements with an oscilloscope are given to provide an insight into the various possibilities. The scope of the measurement functions differs between the instruments. However, the basic functions listed here are basically available on all (digital) instruments.

5.5.7 Voltage and Period Duration Measurement

The measurement of amplitudes and period durations is the simplest measurement. This can be done by manually positioned cursor lines or automatically. Figure 5.26 shows the used measuring circuit.

The measurement results are displayed in the footer of the image. The header line shows the settings of the oscilloscope. From left to right, these are: the *input sensitivity* of the channels, the *offset* of the trigger point in the image, the *deflection speed*, the *trigger mode*

Fig. 5.26 Measuring circuit

with direction, and the *trigger threshold*. The input sensitivity and the deflection speed are related to grid units, which are stored in the image as a grid.

5.5.8 Measurement of Rise Times

Figures 5.27 and 5.28 show the measurement of rise times. This can also be done manually, but is more accurate in the automatic form, since the determination of the measurement points of 10% and 90% of the signal amplitude is done by the system. In Fig. 5.27, the automatically determined measuring points for 10% and 90% of the amplitude are entered as horizontal dashed lines. The vertical dashed lines represent the points of intersection of the input signal with the amplitude lines and thus provide the points for the measurement of the rise time. The results are output in the footer for the rising and falling edge.

Figure 5.28 shows the *zoom* option. Due to the large memory depth, it is possible to zoom into the signal in normal display (above). The section framed in black at the top is spread out over the entire width of the image at the bottom, so that very good detail resolutions are achieved. Here, the needle caused by the inductance was highlighted and measured.

Figure 5.28 shows the same process as Fig. 5.27. The temporal resolution was reduced from 200 to 100 $\mu\text{s}/\text{grid}$. The needle impulse existing at the switch-over point between charging and discharging is not visible in the picture. Only by spreading the stored data, it becomes visible and can be measured. At 80 ns for the falling edge, it is several orders of magnitude shorter than the actual rise times of the signal.

5.5.9 Measurement of Phases

There are two methods for measuring phase shifts:

1. The measurement of *time offsets* between two measuring channels and
2. The creation of *Lissajous figures*.

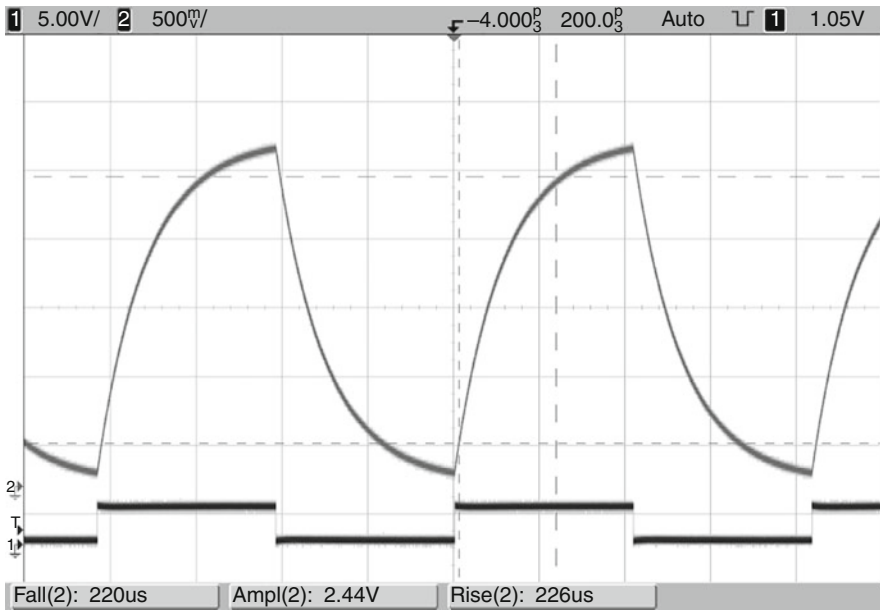


Fig. 5.27 Rise times of a signal

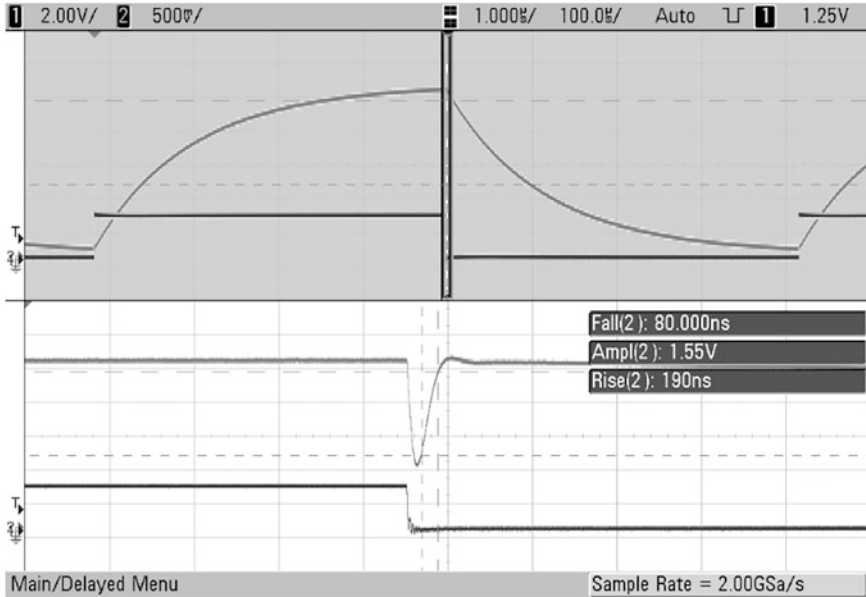


Fig. 5.28 Highlighting needle pulses

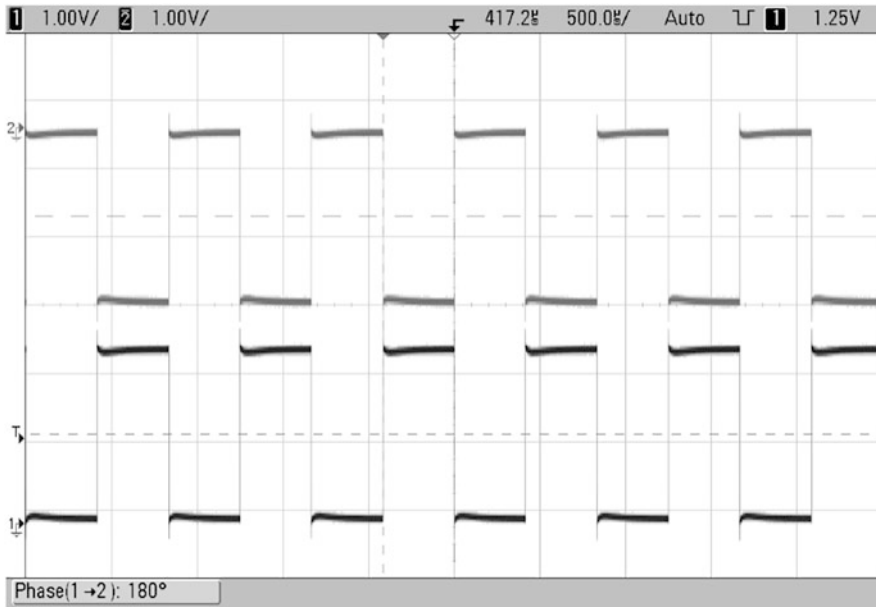


Fig. 5.29 Phase measurement

The phase measurement can also be performed by the device as shown in Fig. 5.29. The use of the manual setting is necessary if transit times are to be determined in non-periodic signals. It is always performed as a reference between two channels. Figure 5.29 shows the measurement between the rising edges of signals 1 and 2. In this case of signal inversion, these are 180° phase offset.

The other way is the creation of *Lissajous figures*—so-called *beat pictures*. They are created when the oscillograph is controlled separately in X and Y with harmonic oscillations. If the two input signals have a *fixed frequency relationship* to each other, the result is a standing waveform whose shape is determined by the phase position of the two signals. Figure 5.30 shows a selection for the frequency ratio 1:x.

Other frequency ratios are also possible, such as 2:3 or 3:4, which makes the figures increasingly complicated. If the frequencies are not exactly in line with each other, a *wandering phase shift* is obtained, which causes the figures to move. In this way, systems can be adjusted to a standing image by matching.

These images can also be generated by mirror systems with servo control and laser light, which is also done in the field of *event technology*.

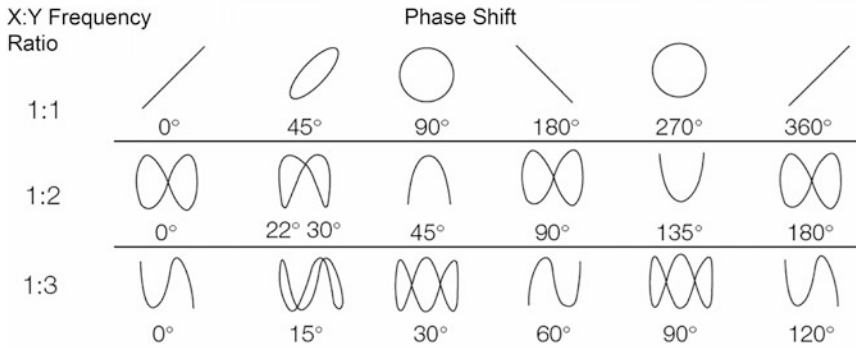


Fig. 5.30 Lissajous figures for the frequency ratio 1:x

5.5.10 Measurement of Pulse Duration and Pulse Ratios (PWM)

The following two Figs. 5.31 and 5.32 show different approaches to measurement. Both devices allow the measurement of pulse width and period duration. The duty cycle and thus the PWM value can be determined from both values together. In Fig. 5.31, the period duration and the pulse width are measured. The system determines the pulse width and the period duration of the signal (right column). The *PWM value can be* determined from these two values. The averaging (integration) over the signal results in a resulting output voltage, which can be seen in relation to the peak voltage. If you recalculate the values in the figure, you will get a deviation. The reason is that the peak-to-peak voltage U_{ss} is calculated from the minimum and maximum values and includes the disturbances present on the signal. Furthermore, an error occurs if the low level is not exactly at 0 V.

Determining the PWM value from the *time ratios*, as shown in Fig. 5.32, is the more accurate way. Here, both times are also measured and the ratio of pulse width to period is calculated.

5.6 Speed and Angle of Rotation

5.6.1 Definition

The *speed* n indicates the number of revolutions per time unit. Thus, it has the same unit as the frequency, but in practice, it is given in [revolution/minute; rpm] to distinguish it from the frequency of an oscillation. In principle, all the measuring methods specified in Sect. 5.2 apply to the speed.

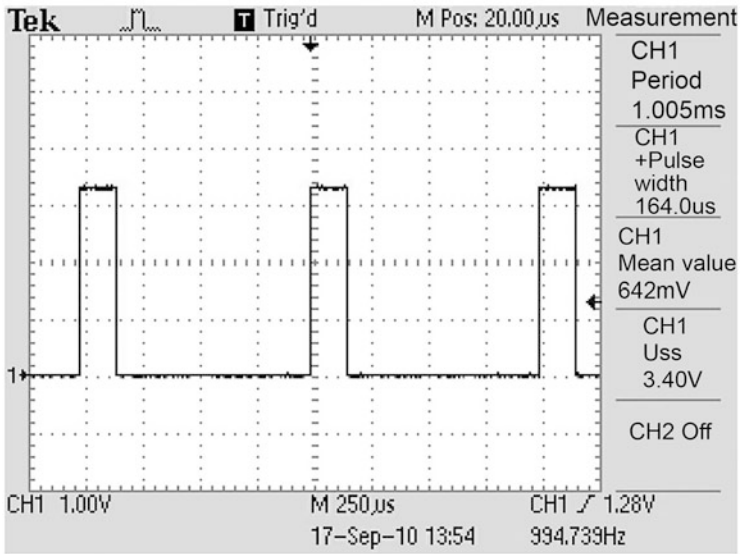


Fig. 5.31 PWM determination via the mean value

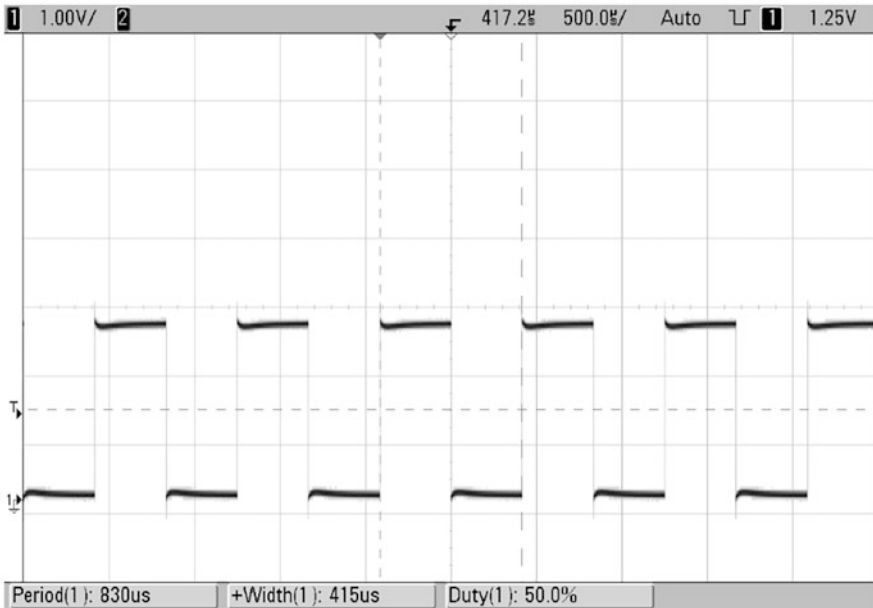


Fig. 5.32 Direct evaluation of the PWM value

5.6.2 Measuring Principles

When measuring a rotational speed, this usually takes place on a shaft or a wheel. Technically, the number of rotations must be recorded, which is done by means of markings on this shaft.

5.6.3 Measurement Through Event Recording

The marking of a shaft for speed measurement can be done in different ways depending on the application:

- Attach *markings* for *optical* scanning,
- Introduction of *magnetically detectable* materials into or onto the shaft,
- Mounting of *perforated disks* or *impellers* for optical or magnetic scanning or
- Attachment or use of *gear wheels*.

Event detection is carried out by means of light barriers or magnetic distance sensors (Sects. 3.1.5).

5.6.4 Measurement by Position Detection

If, in addition to the speed, the *direction* or the *angle of rotation* (partial rotations) is also to be recorded, a reference point for the zero position and a finer resolution of the rotation with several measuring channels is required. This can be achieved by:

- *Angular encoders*, which have a binary position code on a disk that is optically scanned. For simple applications with low resolution, this signal coding can also be done via sliding contacts.
- A relative angle and direction detection can be achieved by means of *line disks* and phase-shifted scanning.

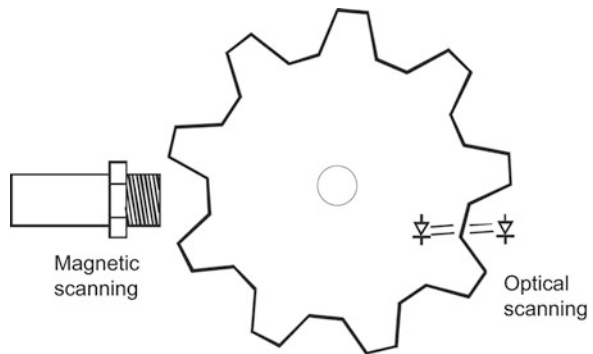
5.6.5 Speed Determination by Beating (Stroboscope)

The beat is based on the subtraction of two—nearby—frequencies, which are made optically visible. The rotating shaft is given a mark, e.g. a notch, which is illuminated by a frequency-controlled flash lamp. If the *flash period* and the *rotation are the same*, the *marking* appears *stationary* in a stable position. If the two frequencies do not match, the marking moves optically with the *difference frequency*. This method can also be used to determine angular positions, as is done when adjusting the ignition timing in a car engine. Since the result of the operation is not electrically available and therefore not electrically controllable, it has lost its importance. A still common application today is the speed control of high-quality turntables (Fig. 5.33)



Fig. 5.33 Pattern for stroboscopic measurements of the rotational speed on turntables (Factory photo: RELOOP)

Fig. 5.34 Optical and magnetic scanning of rotary movements



5.6.6 Measuring Arrangements

5.6.6.1 Magnetic and Optical Scanning

The scanning of a rotation can easily be done by means of light barriers or inductive distance sensors (Sect. 3.1.1). Distance changes during the rotation or gaps in components can be used. Figure 5.34 shows an example.

5.6.7 Digital or A/B Interface (Incremental Encoder)

If at least two phase-shifted samples are taken on a shaft, the *direction of rotation* can be determined from the *phase position*. By adding the pulses with the correct sign, the *rotational position* can also be determined, which can also be done over several revolutions. Figure 5.35 shows the arrangement and Fig. 5.36 the corresponding pulse pattern.

Fig. 5.35 Incremental sensor

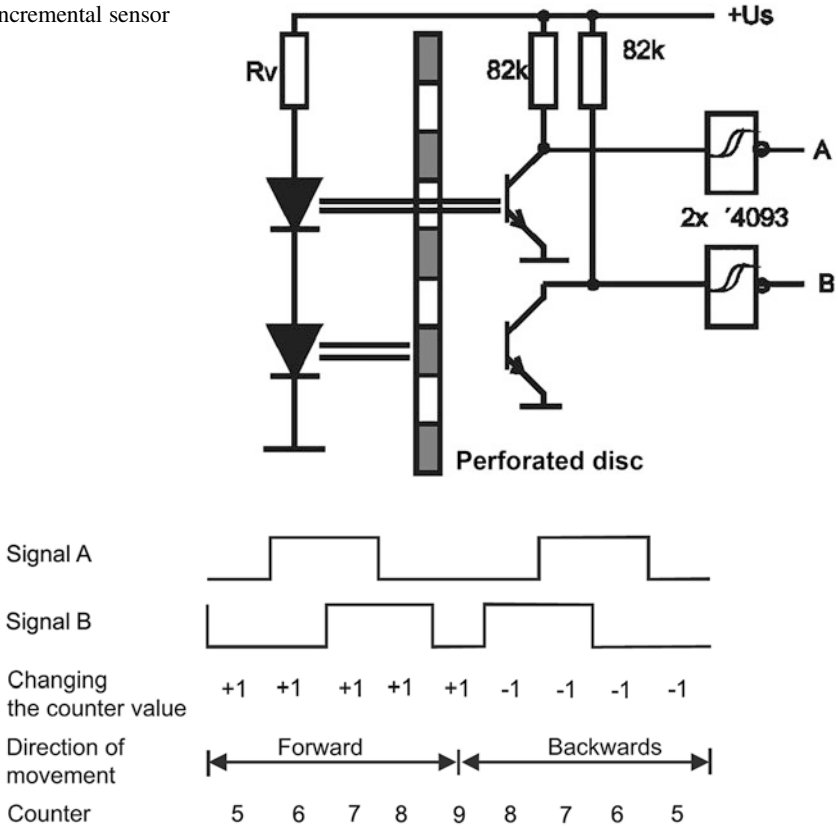


Fig. 5.36 Incremental transducer

In a measuring system with *digital interface*, the position signal is output via two digital signals A and B. A common designation for this is “A/B-” or “90° (pulse) interface”. The physical level is either a 5 V differential signal (RS422) or a *single-ended signal* that assumes either the level of the supply voltage or zero.

If signal A comes before signal B, each edge corresponds to one increment in positive direction. If signal A comes after signal B, this corresponds to one increment in negative direction.

The counting direction of one signal edge is defined by the *level* of the other signal. An example is the rising edge of A in Fig. 5.36 at counter position 41 when the direction of movement is “forward”. Here, the level of B is low. The edge is evaluated as a positive increment. On the rising edge of A at position 42 during the backward movement, the level of B is high. Accordingly, this edge is evaluated as a negative increment.

If four light barriers with a phase offset of 90° to the grating structure are used and the output signals of the optical sensors are led out in analog form, triangular to sinusoidal



Fig. 5.37 Selection of incremental sensor (Factory photo: Pepperl+Fuchs)

output signals are obtained for each channel according to the optical design. By evaluating the analog amplitudes, interpolation of the intermediate position of the grating structure is thus possible, which considerably increases the position and angle resolution. Examples of incremental transducers are shown in Fig. 5.37. Further explanations can be found in Sects. 3.2.5 and 3.2.6.

5.7 Speed

5.7.1 Definition

The *velocity* v is the ratio of the distance s travelled and the time t required for this if a uniform motion is taken as a basis.

$$v = \frac{s}{t}.$$

Special cases of speeds are the path speed for circular movements

$$v = 2\pi f \cdot r$$

with	v	Path speed	m s^{-1}
	r	Radius of the circular path	m
	f	Speed, frequency	s^{-1}
	and the angular velocity $\omega = 2\pi f = v/r$.		

5.7.2 Measuring Principle Speed

The measurement of the speed is always based on a *path-time measurement*. In systems with a variable structure, the two basic quantities of displacement and time are measured separately and calculated. With a stable measuring arrangement, it is also possible to work with a defined measuring distance, so that only time has to be measured. The time measurement can be triggered and stopped by binary signals (e.g. light barriers).

A fundamentally different approach is to use the *Doppler effect* (Sect. 2.20). It describes the change in frequency of a reflected wave by a moving object. This method can be used with sound, ultrasound, radar, or light waves, depending on the required measuring range and accuracy.

Another way is the determination of the speed by a *double distance measurement* in a defined time interval. Here, a distance determination by means of *laser time-of-flight measurement is suitable*, which allows a sufficient measuring rate with good resolution—even at long distances.

For all three methods, it must be specifically estimated where the measurement error lies. With the distance-time measurement—depending on the distance—an average speed is determined, whereas with the Doppler effect, an instantaneous speed is measured.

5.7.3 Measuring Arrangements for Speed Measurement

A description of distance and time measurement is not given here, as it is done elsewhere in the movement. A special method for determining displacement using *correlation measurement technology* is described. The technical problem is that a velocity measurement according to the Doppler effect is only accurate if the measured object moves exactly in the direction of the measuring beam. When measuring the speed of a vehicle, the use of wheel speed is also ruled out, since the circumference depends on the condition and pressure of the tires, and the speed also depends on the travel distance (e.g. inside or outside curve).

An objective statement is provided at this point by the correlation measurement technique. This is done by optical scanning of the road under the vehicle. The surface structure of the road results in a modulation in the measurement signal. At a defined distance (in the direction of travel), the second scanning takes place (Fig. 5.38). The scanning is carried out at a *fixed sampling rate*, which represents the time base for the measurement. Figure 5.39 shows a possible curve from this measurement. These two function curves are brought into coincidence by a correlation function. For this purpose, the correlation coefficient between both measurement series for the displacement $k = 0$ is formed for the section X_1 to X_n . The process is repeated for increasing k values.

The correlation coefficient is determined by the function

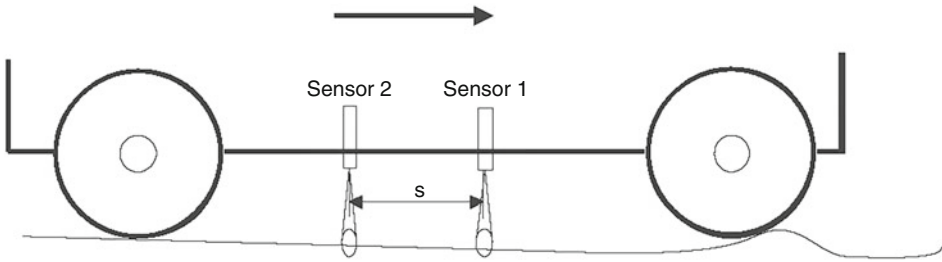


Fig. 5.38 Sensor arrangement for speed measurement

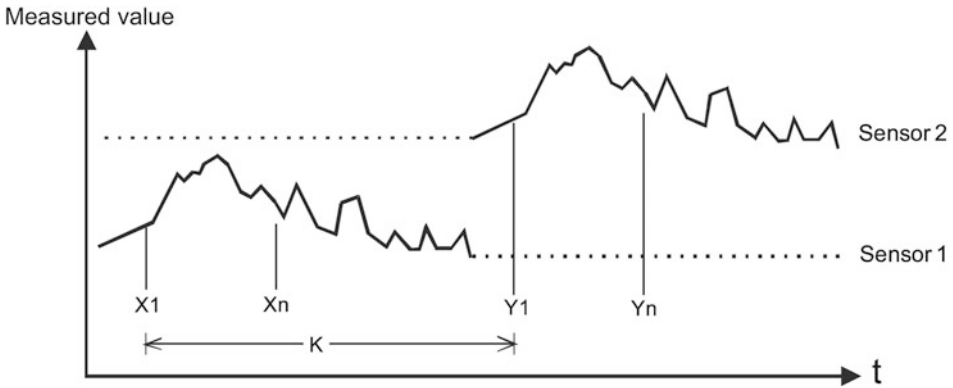


Fig. 5.39 Measured values from the arrangement in Fig. 5.38

$$r(k) = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_{i+k} - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot \sum_{i=1}^n (Y_{i+k} - \bar{Y})^2}$$

Here...

X_i	the measured value of sensor 1 of element i from the range $[X_i, X_n]$.
Y_{i+k}	the measured value of sensor 2 of element i offset by time k
\bar{X}	the mean value of the measured values of sensor 1 in the range $[X_i, X_n]$.
\bar{Y}	the mean value of the measured values of sensor 2 in the range $[Y_i, Y_n]$.

The correlation coefficient can have a value between -1 and $+1$. Where $+1$ corresponds to a *complete match* of the function progressions and -1 indicates a match of inverse functions. A value of zero describes an incoherence of both function courses.

If the results of the correlation coefficients are entered into a diagram over a changed distance k , a correlation function is obtained (Fig. 5.40).

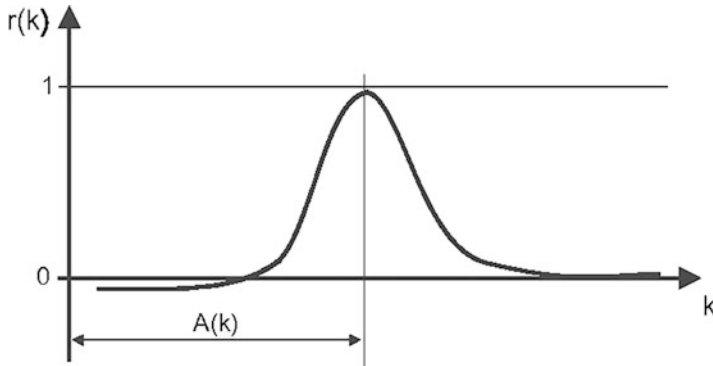


Fig. 5.40 Correlation function for the measurement in Fig. 5.38

At the scanning point $A(k)$, the function curve has a maximum, which indicates a correspondence of the function curves. This means that after k scans with the time interval t , the acquired profile of sensor 1 was recognized again at sensor 2. In other words—the vehicle has travelled the distance s in the time $k \cdot t$, which allows the vehicle speed to be calculated.

5.8 Acceleration

5.8.1 Definition

Acceleration a describes the change in velocity of an object and is described by the first derivative of velocity with respect to time.

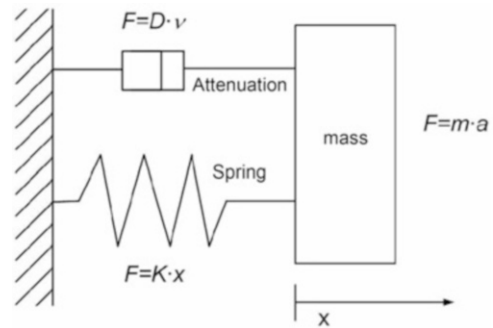
$$a(t) = \frac{dv(t)}{dt} = \frac{d^2x(t)}{dt^2}.$$

a	Acceleration	m s^{-2}
dv	Change in speed in time t	m s^{-1}
dt	Period of the speed change s	
x	distance travelled in time t	m

Acceleration can be both positive (acceleration) and negative (deceleration). In principle, all measurements of acceleration are based on Newton's law of motion:

$$\mathbf{F} = m \cdot \mathbf{a}.$$

Fig. 5.41 Physical description of an accelerometer



Here, F is the force, m the mass, and a the acceleration. An accelerometer can be imagined as a spring-mounted mass and a damping. This can be described by the differential equation

$$F_{\text{ext}} = m \frac{dx^2}{dt^2} + D \frac{dx}{dt} + Kx$$

Here, D is the damping factor, K is the spring constant, F_{ext} is the force acting from the outside, and x is the distance by which the mass was displaced in the time t (Fig. 5.41).

The same applies to the angular acceleration α : It is calculated from the derivative of angular velocity with respect to time:

$$\alpha = \frac{d\omega}{dt}.$$

5.8.2 Fields of Application

Accelerations are measured in many fields of application, as Table 5.1 shows in an overview.

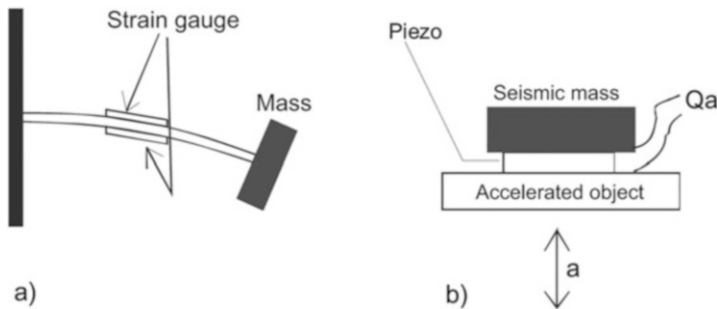
5.8.3 Measuring Principle Linear Acceleration

The simplest measuring principle is the derivation of a sequence of path-time measurements. *Accelerometers* are generally used for measurement. They are based on a spring-mounted (seismic) mass, which is deflected by the force acting during the acceleration process (Fig. 5.41).

This mechanical deflection can be detected by piezoelectric or magnetic systems. Figure 5.42 shows two basic variants of the arrangement.

Table 5.1 Fields of application of the acceleration sensors

Field of application/sector	Concrete application
Automotive	Safety systems (e.g. airbags, seat belt tensioners, immobilisers) faults in the transmission Storage of acceleration values in case of accidents maintenance request
Aerospace	Control of aircraft or movement of bodies in space
Seismology	Ground movements and earthquake measurements
Mechanical engineering	Navigation of robot arms inclination measurements Preventive maintenance of wearing parts Imbalance, loosening connections, faults in gearboxes Defects in ball-bearings Modal analysis
Construction	Load measurement of buildings, towers, bridges, and dams
Medicine	Motion Analysis Diagnostics
Consumer area	Weighing technology to eliminate the influence of gravity motion analysis and control (e.g. washing machine, video recorder) game consoles Navigation
Military sector	Control in ballistics (e.g. aircraft or projectiles)

**Fig. 5.42** Basic structure of an accelerometer

In the course of the miniaturization of these sensors using *MEMS technologies* (MEMS: micro-electro-mechanical system), the position of the mass is also detected by piezoresistive or capacitive systems (Figs. 5.43 and 5.44).

Figure 5.45 shows the frequency analysis for vibrations in a production plant. Possible damage or faults in the plants can thus be detected in advance and thus avoided.

Since the acceleration due to gravity always acts on the systems, the sensors must be installed in the specified position. In multi-axis systems, the acceleration due to gravity occurs as a disturbance variable in at least one channel. The accuracy of an accelerometer

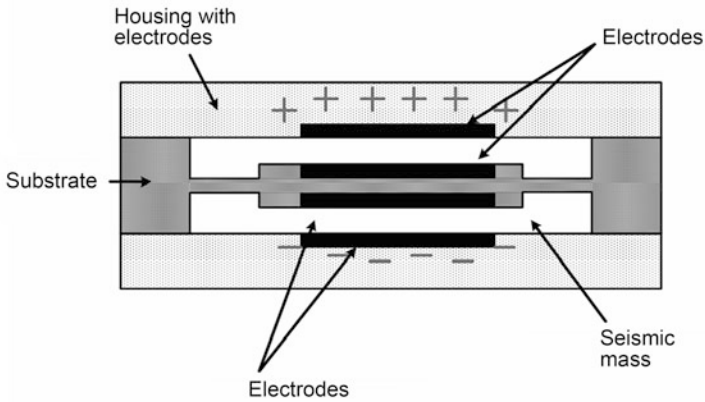
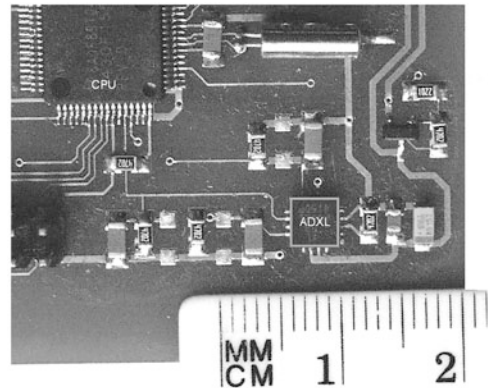


Fig. 5.43 Design of a capacitive MEMS accelerometer

Fig. 5.44 Using a capacitive MEMS accelerometer



increases with increasing seismic mass. In return, however, their cut-off frequency decreases.

If you operate an acceleration sensor with a very low cut-off frequency, you get an *inclination sensor (inclinometer)*, which determines the deviation to a given axis. This takes advantage of the fact that the acceleration due to gravity is always perpendicular to the earth's surface.

A wide range of applications for inclination sensors are mobile machines such as excavators, lifting platforms, and cranes. With these, it must be monitored that the load on the jibs does not cause the equipment to tilt. Figure 5.46 shows a tilt sensor for two axes.

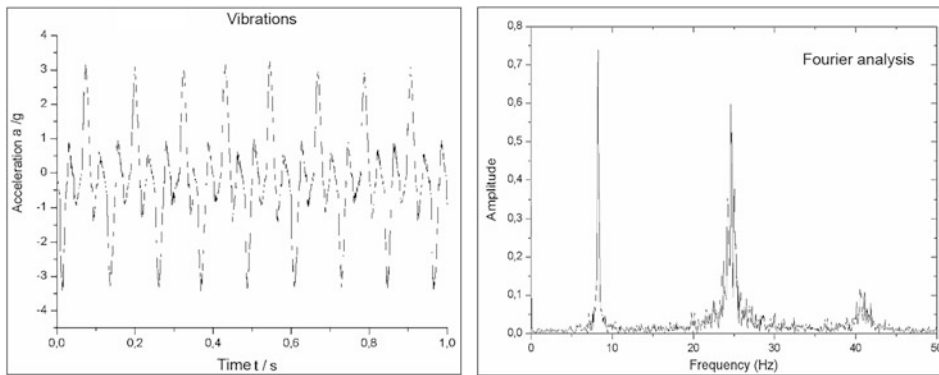
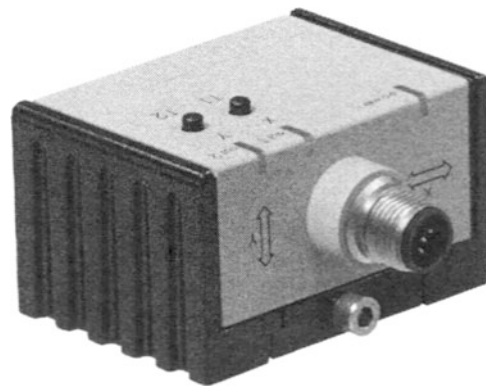


Fig. 5.45 Vibration analysis for a production plant

Fig. 5.46 2-axis tilt sensor
(Factory photo: Pepperl+Fuchs)



5.8.4 Measuring Principle Angular Acceleration

The angular acceleration is determined with an independent sensor principle, the *gyroscope*. A rotating or oscillating mass forms the core. If a torque acts on this mass from the outside, a force (*Coriolis force*) is generated, which deflects the vibrating mass laterally. This relative movement is recorded as a measured variable. The best known system on this basis is the *gyrocompass*. Microelectronic solutions such as the CMR3000 (VTI Technologies from Finland) or the MPU-6000 (InvenSense Inc., USA) have also been available for several years. Figure 5.47 shows the operating principle for the X axis.

Figure 5.48 shows the chip photo of the MPU-3000, in which the areas of micromechanics are clearly visible.

Two areas can be named as the main application—*near-field navigation* and use in *game consoles*. In the case of game consoles, the information from angular acceleration is used to control game movements in the sports sector. In combination with linear accelerations, a good reproduction of complex movements can be achieved. Today's navigation systems have two disadvantages: the *spatial resolution* is relatively *low*

Fig. 5.47 Gyroscope cell for rotation around the X-axis (Factory photo: InvenSense Inc., USA)

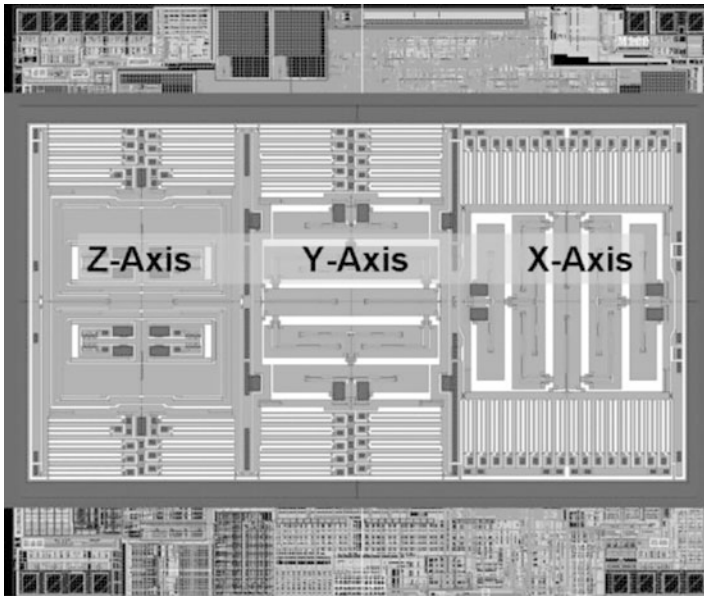
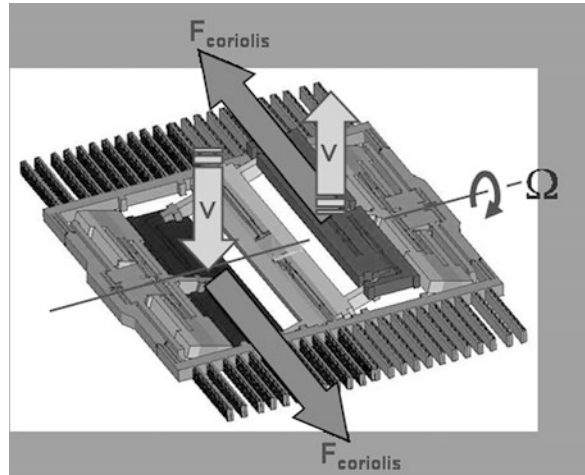


Fig. 5.48 Chip photo of the 3-axis gyroscope MPU-3000 (Factory photo: InvenSense Inc., USA)

(in the range of less than 10 m) and it requires a *line of sight* to the satellites. Therefore, it is not applicable in tunnels or buildings. By detecting rotary movements with gyroscopes, a change of direction in the range of decimeters can be traced, thus enabling navigation in closed objects.

5.8.5 Measuring Arrangement for Measuring Acceleration

Since the measurement of acceleration is traced back to a mass displaced under the influence of force, the evaluation is done at the end by determining this movement. This means that all the methods already described for determining displacement and force are used. It can thus be sampled piezoelectrically, piezoresistively, inductively, or capacitively. In practice, microsystems with 1–3 axes will be used, which are offered in sizes of a few cubic millimetres. An example is the ADXL335 (www.analog.com). This is an acceleration sensor with analog outputs for 3 axes. It can detect accelerations in the range of $\pm 3g$. Its dimensions are only 4 mm \times 4 mm \times 1.5 mm (Fig. 5.44).

5.9 Flow Rate (Mass and Volume)

5.9.1 Definition

Flow rate usually implies measurement for liquids and gases. A multitude of different measuring principles for measuring *quantity*, *mass flow* and *volume per time unit* today allows individual adaptation to the requirements of a wide range of measuring tasks. Due to the different advantages and disadvantages of the measuring principles, a variety of applications has developed.

5.9.2 Mass

The SI unit for mass is kilogram [kg] and denotes the quantity of a body or substance (Sect. 4.1). In addition to the measurement with a reference mass in the case of resting bodies, the basic Newtonian equation of forces and acceleration is also used to establish proportionality to the mass.

$$F = m \cdot a (\text{force} = \text{mass} \cdot \text{acceleration}) \rightarrow m = F/a$$

5.9.3 Volume

For solid bodies, the volume is the space required by a body with the symbol V . The SI unit is the cubic meter (m^3). For gases and liquids, the unit litre is also common. Technically, a distinction is made between a hollow volume, for example of a container, or a volume of space in the case of solid, liquid and gaseous substances.

This is very important, for example, when transporting fabrics and materials. In the case of solid materials (e.g. coal), the hollow volume required for transport is greater than the

volume of the coal to be transported. With liquid substances, it is almost the same, whereas when transporting gases, the possibility of compression is used and thus a multiple of the volume can be transported in relation to the normal pressure. The same is also used when diving with compressed air cylinders.

5.9.4 Volume Flow

Volume flow defines a medium (solid, liquid or gaseous) that moves within a known profile or cross section in a unit of time.

5.9.5 Mass Flow

Mass flow defines the movement of a medium with its mass through a cross section in a unit of time.

5.9.6 Main Groups

In practice, flowmeters are usually differentiated between *volume* and *mass flowmeters*.

In the case of gaseous substances and air, volumetric instruments provide measured values as a function of pressure and temperature with the unit operating litre/unit of time, which are then converted to the standard condition. The measurement result of mass measuring instruments provides measured values with the unit gram/litre. The flow of liquids and gases is explained in detail and illustrated with examples in Chap. 10. The following section shows how *dry flow* is measured, i.e. the mass and volume of moving bulk material or portioning in the packaging industry.

5.9.7 Measurement Methods and Application

The detection of moving solids and bodies and their evaluation according to mass and volume per unit of time also plays an important role in the chemical and especially the pharmaceutical industry. In these industries, *weighing equipment* is used to fill chemicals, granulates and powders or to prepare them for production. In the control area of tablet production, for example, this ensures that the filling process is carried out and that the completeness of the product is checked.

Figure 5.49 shows the TC8410 checkweigher from Collischan. Very fast counting machines for a wide variety of tablets and capsules are also used during production, thus ensuring that the strict regulations in these sensitive areas are implemented.



Fig. 5.49 TC8410 high-performance checkweigher for the pharmaceutical industry (Factory photo: Collischan)

5.9.8 Bulk Flow Meter

So-called bulk current meters, as shown in Fig. 5.50, are used in many areas of industry, for example, in the food industry, the building materials industry, and the pulp and paper industry. Here the materials are transported and measured on conveyor belts, steel plate belts, on troughed chain conveyors, on screw conveyors as well as in shafts and pipes in free fall.

With closed weighing systems, the flowable and homogeneous bulk material is measured during the passage. Via the adapted inlet, the fed bulk material is directed onto a radially curved surface. The centrifugal force of the bulk material is countered by the centripetal force of the same magnitude. The bulk material remains on its circular path. The throughput is measured with force sensors and is directly proportional to the centripetal force.

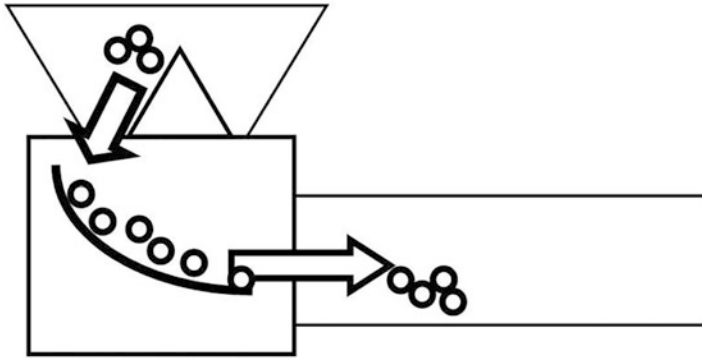


Fig. 5.50 Diagram of a bulk current meter

5.9.9 Baffle Plate Scales

With impact plate scales, the impulse of the impacting bulk material is recorded on a baffle plate. In this process, the bulk material slides down a defined slope on a tube placed at an angle or an open chute.

5.9.10 Weighfeeders

Dosing belt scales fulfill the function of weighing, transport and can be part of a control chain with measuring functions. Their fields of application are mainly in the dosing of various mixtures of the feed and building materials industry.

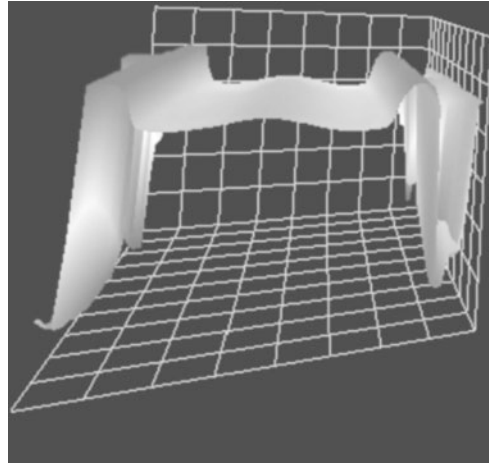
5.9.11 Belt Scales

Belt scales are installed in existing conveyor belts as weighing systems (built-in conveyor belt scales). *Material flows* are thus measured with long-term stability. An example of the application is the conveyor belt scale for potato sorting plants.

5.9.12 Differential Scales

With differential scales, the removal of the material in a container is weighed at the same time and the speed of the dosing unit is measured. From this, a processor determines the quantity of outgoing material in kg/h. It is a dust-tight system for dosing and measuring with low maintenance. In the grain and feed industry, it is used in stables for fully automatic

Fig. 5.51 3D image of a conveyor belt section (Source: EmWeA Prozessmesstechnik e. K.)



power feeding of cattle, whereby each animal is identified by its chip and double dosing can be excluded.

5.9.13 Optical Belt Weigher

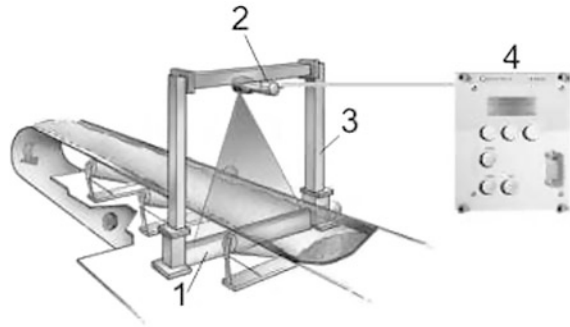
The optical belt weigher FLO-3D Sensor (EmWeA Prozessmesstechnik e. K.) uses a 3D camera to continuously take pictures of the moving bulk material on the conveyor belt. This is assigned to a previously stored profile. From this, the current volume throughput in m^3/h is determined. Figure 5.51 shows a 3D image of a conveyor belt section.

Here, too, the measurement is carried out completely contact-free. Figure 5.52 shows a sketch for a bulk flow measurement.

An easy to install frame construction (3) completes the measuring system. A shielded radiation source (1) is permanently mounted below the conveyor belt. A scintillation detector (2) is located above the conveyor belt opposite to the source. The measuring signal is determined by the ionizing radiation, which is influenced by the material on the conveyor belt. Depending on the load on the conveyor belt and the speed of the material, an evaluation unit (4) calculates the mass flow. It can also be used for shafts and pipelines, as well as in the high temperature range, at different pressure conditions and for hazardous materials of various kinds.

Other systems not mentioned here are drive-through scales, piece number scales, weighing beams, weighbridges or counting scales.

Fig. 5.52 Sketch of a bulk flow measurement (Factory photo: BERTHOLD TECHNOLOGIES GmbH & Co. KG)



Bibliography

1. Berthold Technologies GmbH & Co. KG, www.berthold.com
2. Collischan GmbH & Co. KG, www.collischan.de
3. EmWeA Prozessmesstechnik e. K., www.emwea.de
4. Henkies, A.: „Entwurf und Optimierung fremdlichttoleranter Tiefenkamerasysteme auf der Basis indirekter Lichtlaufzeitmessung“, Dissertation, Uni Duisburg, Fakultät der Ingenieurwissenschaften
5. Kutscher, N.; Mielke, B.: „3D Kameras – basierend auf Lichtlaufzeitmessung“; http://www.inf.fuberlin.de/lehre/SS05/Autonome_Fahrzeuge/3dKameras.pdf
6. Lange, R.; Seitz, P.; Biber, A.; Schwartel, R.: „Time-of-Flight range imagung with a custom solid-state image sensor“, Centre Suisse d’Electronique et de Microtechnique SA (CSEM), Uni Gießen, Institut für Nachrichtenverarbeitung



Ekbert Hering, Gert Schönfelder, Martin Liess, and Lothar Michalowski

6.1 Temperature as Physical State Variable

Thermodynamic temperature is a *physical state variable* and is one of the seven basic variables in the International System of Units (SI) to which all metrological variables can be traced. Its unit of measurement is the Kelvin (K).

The *International Temperature Scale* was introduced on January 1, 1990 (ITS-90; ITS: International Temperature Scale) and is the basis for all calibrations of temperature measuring instruments. The fixed points are determined on the basis of known relationships in the pressure–temperature diagram for substances that occur in the three states of aggregation: gaseous, liquid and solid and are characterized by the existence of a triple point (Fig. 6.1).

The defined fixed points represent *thermodynamic equilibrium states* between the three possible states of aggregation. The solidification points are always defined for a pressure of 101.325 kPa. At the triple point, the melt pressure curve is combined with the vapour

E. Hering (✉)

Hochschule Aalen (Rektor i. R.), Aalen, Germany

e-mail: Ekbert.Hering@hs-aalen.de

G. Schönfelder

Prignitz Mikrosystemtechnik, Wittenberge, Germany

M. Liess

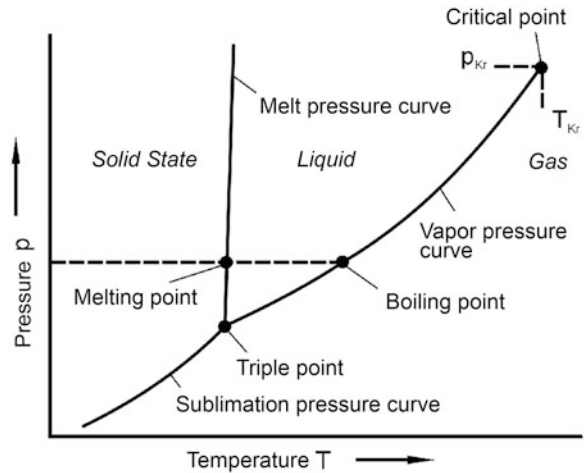
Hochschule RheinMain, Rüsselsheim, Germany

e-mail: Martin.Liess@hs-rm.de

L. Michalowski

Universität Göttingen, Göttingen, Germany

Fig. 6.1 Pressure–temperature diagram of a pure substance with triple point



pressure and sublimation curve. The gaseous, liquid and solid phases of a substance can exist simultaneously at the triple point.

To measure temperatures higher than those specified in ITS 90, the thermoelectric emf of thermocouples based on semi-precious metals such as platinum and rhodium or the superheavy metals tungsten and rhenium or their alloys is used (Sect. 6.4). Table 6.1 shows the temperature fixed points defined according to ITS-90.

The usual unit of measurement for temperature is *Celsius* (symbol t). It results from the definition of the melting and boiling points of water at $0\text{ }^{\circ}\text{C}$ and $100\text{ }^{\circ}\text{C}$, respectively. On the basis of these fixed temperature points, the relationship between temperature t in $^{\circ}\text{C}$ and T in K applies:

$$t = T - 273,15.$$

This allows a simple conversion of the Kelvin temperature into the Celsius temperature and vice versa. This equation also shows that there can be no real negative Kelvin temperatures.

In English literature, Fahrenheit ($^{\circ}\text{F}$) is still frequently used as the temperature unit. The following relationship is valid as a conversion equation:

$$y\text{ }^{\circ}\text{C} = (x\text{ }^{\circ}\text{F} - 32) * 5/9.$$

Table 6.1 Temperature fixed points according to IST-90

State of equilibrium	T90 in K	T90 in °C
Vapor pressure of helium	3 to 5	−270.15 to 268.15
Triple point of the equilibrium hydrogen	13.8033	−259.3467
Vapour pressure of the equilibrium hydrogen	17.025 to 17.045 20.26 to 20.28	−256.125 to −256.105 −252.89 to −252.87
Triple point of the neon	24.5561	−248.5939
Triple point of oxygen	54.3584	−218.7916
Triple point of argon	83.8058	−189.3442
Triple point of mercury	234.3156	−38.8344
Triple point of the water	273.16	0.01
Melting point of gallium	302.9146	29.7646
Solidification point of indium	429.7485	156.5985
Solidification point of tin	505.078	231.928
Zinc solidification point	692.677	419.527
Solidification point of aluminium	933.473	660.233
Solidification point of silver	1234.93	961.78
Solidification point of gold	1337.33	1064.18
Solidification point of copper	1357.77	1058.62

6.2 Measuring Principles and Measuring Ranges

Various physical principles can be used for temperature measurement (Fig. 6.2). The most important effects are discussed in detail in Chap. 2. The *optical systems* allow *non-contact* temperature measurement.

As Fig. 6.3 shows, the different measuring principles cover different temperature ranges. In the normal temperature range between -60 °C and 500 °C to 1000 °C , liquid thermometers, electrical resistance thermometers or thermometers based on the expansion of metals are used.

6.3 Temperature Dependence of the Electrical Resistance

6.3.1 Metals

For metals, the electrical resistance R increases with increasing temperature T (Sect. 2.11). The approximation with a third order power series is normally completely sufficient. Thus, the following applies:

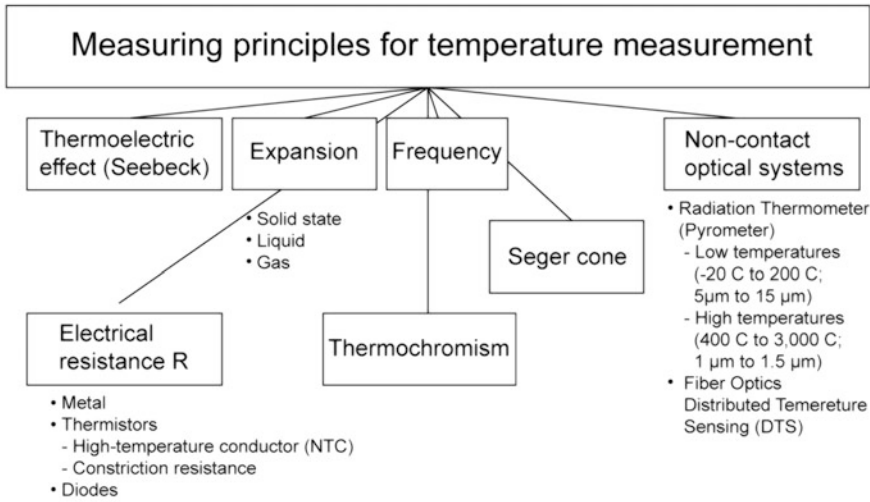


Fig. 6.2 Measuring principles for temperature measurement

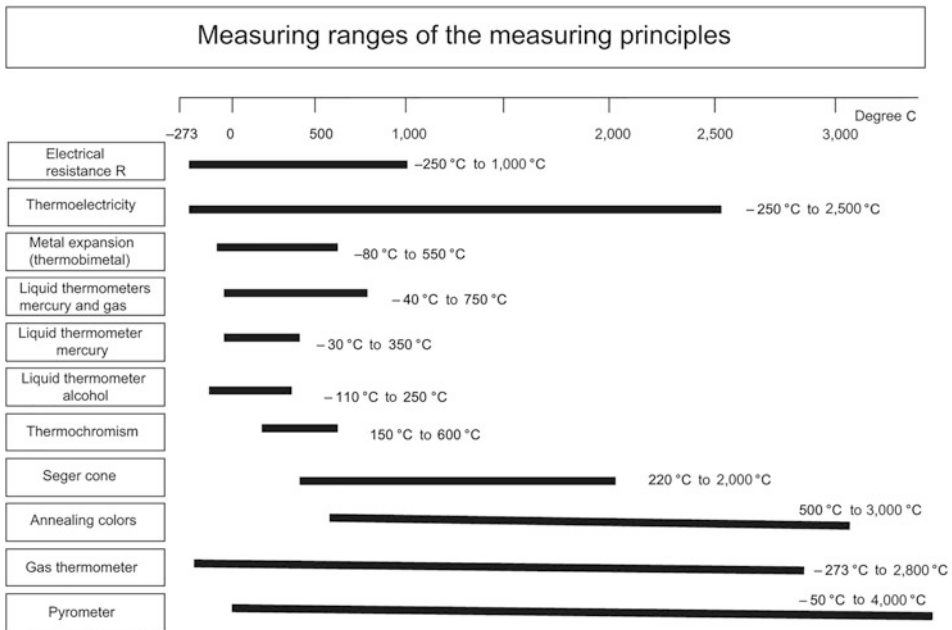


Fig. 6.3 Temperature ranges of the individual measuring principles

$$R(T) = R(T_0) [\alpha T + \beta T^2 + \delta T^3 + \dots];$$

α , β , and δ are the corresponding temperature coefficients.

In many applications in the temperature range between 0 °C and 100 °C, the *linear dependence* of the electrical resistance R on the temperature T is valid. Then the resistance R at room temperature of $t = 20$ °C is valid:

$$R(t) = R_{20}(1 - \alpha(t - 20^\circ\text{C})).$$

For the specific electrical resistance ρ ($R = \rho(l/A)$, where l is the length of the conductor and A its cross section, the following applies accordingly:

$$\rho(t) = \rho_{20}(1 - \alpha(t - 20^\circ\text{C})).$$

The *temperature coefficient* of the electrical resistance α is the ratio of the relative change in resistance to the temperature change (unit of measurement: 1/K). It applies:

$$\alpha = \frac{\Delta R}{R \Delta t} = \frac{(R_{20} - R_0)}{R_0 20}.$$

Correspondingly, ρ applies to the specific electrical resistance:

$$\alpha = \frac{\Delta \rho}{\rho \Delta t} = \frac{(\rho_{20} - \rho_0)}{\rho_0 20}.$$

Table 6.2 shows the specific electrical resistance values of various metals as a function of temperature.

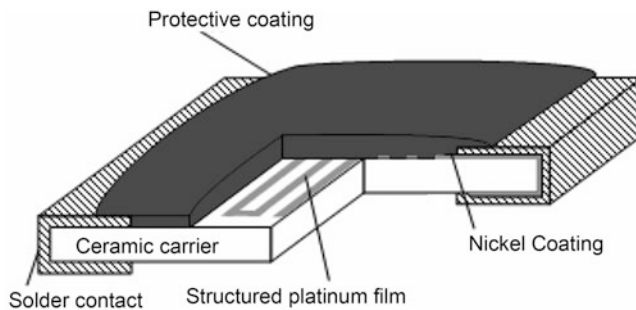
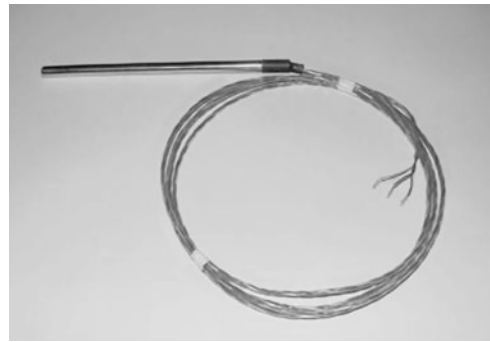
A *Pt100* and *Pt1000*, i.e. a 100 Ω or 1000 Ω platinum resistor has become the industry standard for high-precision temperature measurement. Advantages are the temperature curve defined by the material, high stability and thus high accuracy. Figure 6.4 shows a commercially available Pt100 resistor in three-wire technology (circuit according to Fig. 6.6) with a supply cable about 2-m long. The sensor length is 100 mm and the diameter is about 7 mm. The sensor is suitable for screwed connections.

Alternatively, *nickel* is also used as a base material for resistors (e.g. Ni100). Advantages are the *lower price* of the material and the *higher sensitivity* (change of resistance with temperature). Disadvantages compared to platinum are the higher non-linearity, the lower corrosion stability and the lower temperature range of the material.

Whereas in the past, resistors were manufactured from wound wire (which is still used for resistors of the highest precision), nowadays most metal resistors are manufactured in *thin-film technology* on ceramic substrates (Fig. 6.5).

Table 6.2 Temperature dependence of the specific electrical resistance ρ of various metals

Metal	Platinum (Pt)	Nickel (Ni)	Copper (Cu)	Iron (Fe)
Temperature/°C	$\rho_0/10^6 \Omega\text{cm}$	$\rho_0/10^6 \Omega\text{cm}$	$\rho_0/10^6 \Omega\text{cm}$	$\rho_0/10^6 \Omega\text{cm}$
-200	0.178	–	0.117	–
-100	0.599	–	0.557	–
0	1000	1000	1000	1000
100	1385	1663	1431	1650
200	1757	2501	1862	2464
300	2118	3611	2299	3485
400	2465	4847	2747	4716
500	2800	5398	3210	6162
600	3124	5882	3695	7839

Fig. 6.4 Pt100 temperature sensor made of wire for $-30\text{ }^\circ\text{C}$ to $250\text{ }^\circ\text{C}$ (Factory photo: Heraeus)**Fig. 6.5** Schematic structure of a metal (nickel) thin-film resistor

Especially with low ohmic temperature measurement resistances and for measurements with high precision, the intrinsic resistance of the supply lines may have to be taken into account. This is achieved with a *bridge circuit* in *three- or four-wire technology* (Fig. 6.6).

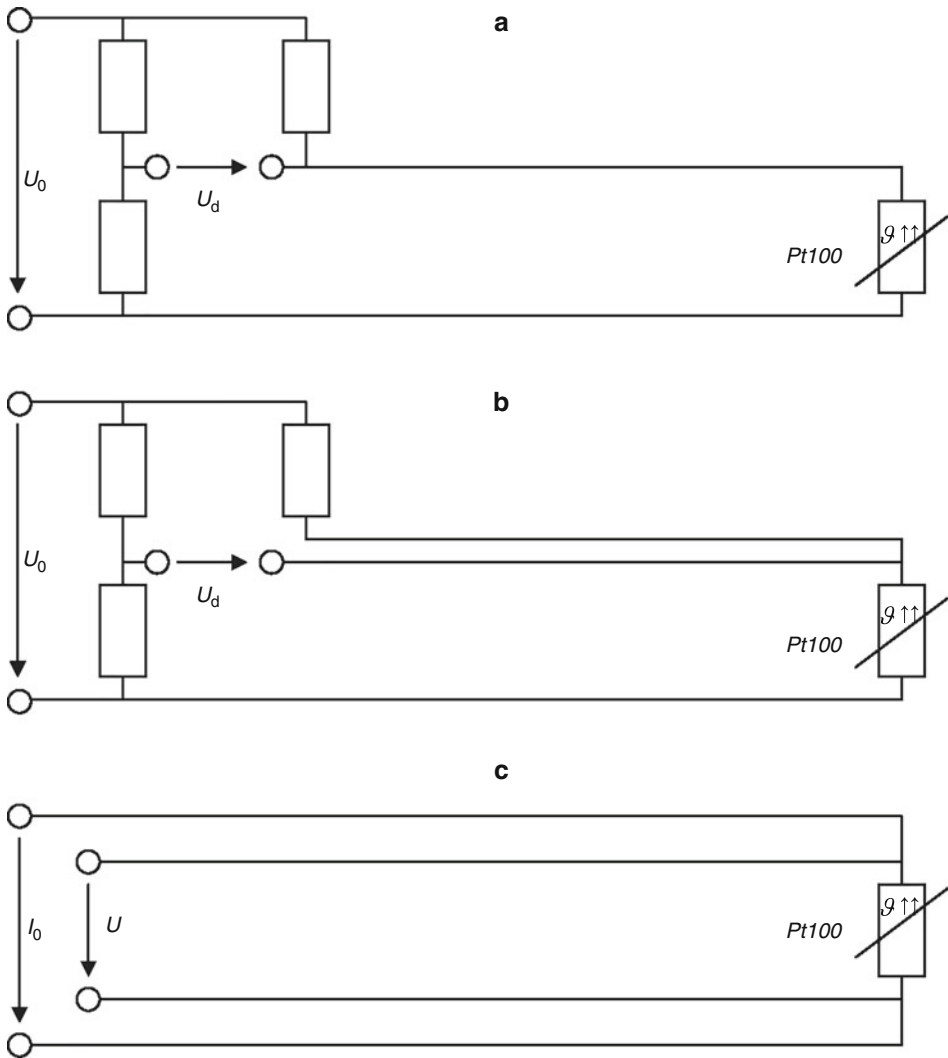


Fig. 6.6 Connection types of a Pt100 temperature measurement resistor. (a) simple bridge circuit, (b) bridge circuit with 3-wire measurement, (c) 4-wire measurement

In Fig. 6.6, U_0 is the voltage applied to the bridge, U_d is the diagonal voltage indicating the temperature. I_0 is the applied current in the V4 wire measurement and U is the output voltage indicating the temperature.

Advantages in temperature measurement with metal wire-wound or thin-film metal resistors are the high stability and low tolerance between the components.

Platinum resistors are used for temperatures between $-200\text{ }^\circ\text{C}$ and $+850\text{ }^\circ\text{C}$. The platinum temperature sensors are characterized by their nominal resistance R_0 at a

temperature of 0 °C. The following sensor types are common: Pt100 ($R_0 = 100 \Omega$), Pt200 ($R_0 = 200 \Omega$), Pt500 ($R_0 = 500 \Omega$), Pt1000 ($R_0 = 1000 \Omega$), Pt3000 ($R_0 = 3000 \Omega$), Pt6000 ($R_0 = 6000 \Omega$) and Pt9000 ($R_0 = 9000 \Omega$). The course of the resistance change is defined in DIN EN 60751. From the equations of this standard, the corresponding temperature can be determined from the measured resistance R . The following approximation methods are used:

- Temperature range between 0 °C and 100 °C (approximation: 1st order polynomial)

$$R = R_0(1 + \alpha T) \text{ with } \alpha = 3,8510^{-3}\text{K}^{-1}.$$

- Temperature range between 100 °C and 850 °C (approximation: 2nd order polynomial)

$$R = R_0(1 + \alpha T + bT^2) \text{ with } \alpha = 3,85 \cdot 10^{-3}\text{K}^{-1}, b = -5,775 \cdot 10^{-7}\text{K}^{-2}.$$

- Temperature range below 0 °C (approximation: 4th order polynomial)

$$R = R_0(1 + \alpha T + bT^2 + c(T - 100\text{K})T^3) \text{ with } \alpha = 3,85 \cdot 10^{-3}\text{K}^{-1},$$

$$b = -5,775 \cdot 10^{-7}\text{K}^{-2}; c = -4,183 \cdot 10^{-12}\text{K}^{-4}.$$

The platinum temperature sensors are divided into the following *accuracy classes* according to DIN 60751, depending on the permissible temperature error dT with respect to the real temperature T :

- Accuracy class AA: $dT = \pm (0.1 \text{ °C} + 0.0017 \cdot T)$,
- Accuracy class A: $dT = \pm (0.15 \text{ °C} + 0.002 \cdot T)$,
- Accuracy class B: $dT = \pm (0.30 \text{ °C} + 0.005 \cdot T)$,
- Accuracy class C: $dT = \pm (0.6 \text{ °C} + 0.01 \cdot T)$.

6.3.2 Metals with Defined Additives (Alloys) or Lattice Defects

The electrical resistance of metals with defined additives or even impurities has a characteristic temperature dependence. Depending on the mechanical and thermal pretreatment of the alloys, the *product* of *specific electrical resistance* ρ and *temperature coefficient* α is a *constant* for a defined temperature (*Matthiesen rule*):

$$\alpha \cdot \rho = \text{const. or}$$

$$\alpha = \frac{1}{\rho} \cdot \frac{d\rho}{dT}.$$

The temperature dependence of the specific electrical resistance of metals with a defined degree of impurity or of alloys is therefore a function of temperature and it applies:

$$d\rho/dT = f(T).$$

The integration of this equation leads to the representation

$$\rho = f(T) + \gamma.$$

Accordingly, the electrical resistance is composed of a *temperature-dependent* component $f(T)$ and a *temperature-independent* component γ . The temperature-independent component takes the influences of:

- Lattice disorder such as vacancies, interstitial atoms, dislocations,
- Phase formations,
- Solid solution formations in alloys and
- Elastic stresses, which lead to increased scattering of electrons in the conduction band.

The *temperature-dependent* proportion *decreases* with *decreasing temperature* and is determined by the strip structure of the pure metallic main component.

6.3.3 Ion Conducting Materials for High Temperatures

Temperature measurement with *ion conductors* is particularly important for the maximum temperature range of 1400 °C to 3000 °C. *High-temperature ion conduction* is made possible by the formation of migratory *oxygen ion defects* in the crystal lattice. The *thermally activated ion conductivity* σ_{ion} follows the equation

$$\sigma_{\text{ion}} = \sigma_0 \cdot e^{(E_A/kT)}$$

with σ_0 a conductivity related to the temperature T_0

E_A of the activation energy for ion conduction (typically 0.7 to 1.0 eV)

k (Boltzmann constant) = $1.3806505 \cdot 10^{-23}$ J/K and

T absolute temperature in K.

The most important oxygen ion conducting compounds are the partially stabilized *cubic zirconia* with calcium fluoride structure and the *cubic thorium oxide* with and without

Table 6.3 Ion conductor materials for maximum temperature measurements

Material	X	Activation energy in eV	Upper operating temperature in °C
$(\text{ZrO}_2)_{1-x}(\text{CaO})_x$	0.12 to 0.20	0.67	1500
$(\text{ZrO}_2)_{1-x}(\text{MgO})_x$	0.12 to 0.26	0.68	1500
$(\text{ZrO}_2)_{1-x}(\text{Y}_2\text{O}_3)_x$	0.08 to 0.35	0.70	2200
$(\text{ZrO}_2)_{1-x}(\text{Yb}_2\text{O}_3)_x$	0.08 to 0.26	0.71	2200
$(\text{ThO}_2)_{1-x}(\text{Y}_2\text{O}_3)_x$	0.00 to 0.15	1.00	3000

stabilizing oxides. Ion conduction in the high temperature range is also shown by the *uranium dioxide* UO_2 and the *cerium dioxide* CeO_2 as well as some *rare earth oxides* with oxygen defects. Important ion conductors are listed in Table 6.3.

6.3.4 Thermistors

Thermistors are differentiated according to *PTC* (positive temperature coefficient; Sect. 2.12.2) and *NTC* (negative temperature coefficient, Sect. 2.12.3). In the case of the *NTC thermistor* normally used for temperature measurement, *heat* causes excitation of charge carriers in the material, which generally consists of a polycrystalline, semiconducting or sintered metal oxide ceramic (e.g. iron oxide, zinc-titanium oxide, magnesium dichromate). The resistance is described by approximate equations, such as the *Steinhart–Hart equation*:

$$\frac{1}{T} = a + b \cdot \ln(R) + c \cdot \ln^3(R)$$

R: Resistance of the NTC thermistor,

T: Temperature,

A, B, C: Constants describing the behaviour of the NTC thermistor.

Thermistors change their resistance by several orders of magnitude even in small temperature ranges (Fig. 6.7) and are therefore well suited for the construction of simple circuits for temperature measurement.

As Fig. 6.7 shows, the nominal resistance is reached at 25 °C. The high sensitivity (dependence of the resistance on the temperature) is clearly visible.

Thermistors with a positive temperature coefficient (PTC; Sect. 2.12.2) are not normally used for temperature measurement. They serve, for example, as self-resetting fuses or are

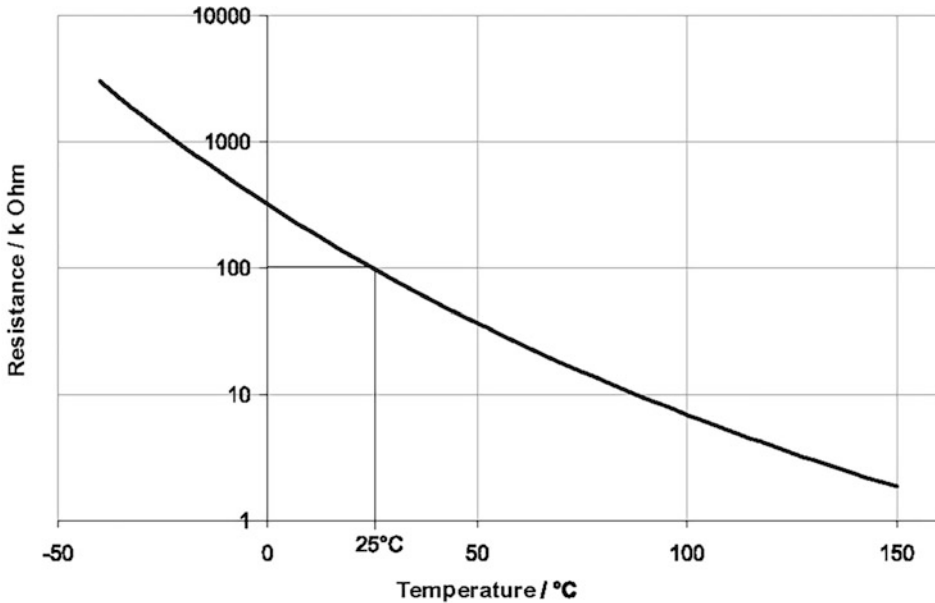


Fig. 6.7 Resistance curve of a 100-kΩ thermistor with the temperature

used in similar non-measurement applications (overcurrent protection, starting resistor for motors).

6.3.5 Constriction Resistor Temperature Sensors (Spreading Resistor)

Constriction resistance temperature sensors (Figs. 6.8 and 6.9) are manufactured in *silicon technology*. A current path is established between a contact on the upper side of the chip through a small, well-defined opening (*constriction*) in the insulating surface coating of silicon oxide in the bulk material. In the range of impurity depletion between 60 K and 600 K, *electrons* are *scattered* in the silicon by the *lattice vibrations* and thus slowed down. This effect leads to an almost linear positive resistance–temperature curve in small ranges.

Compared to thermistors, constriction resistors have a *high reproducibility* and *stability*. The total resistance is determined by the diameter of the small contact opening and the specific resistance of the semiconductor and is

$$R = \frac{\rho}{\pi \cdot d}, \quad (6.1)$$

where R : Resistance of the sensor at given temperature,

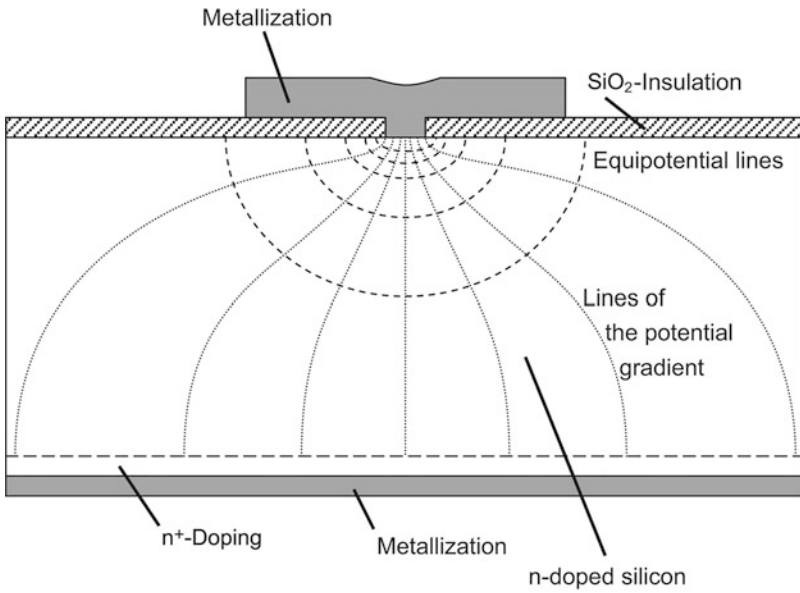


Fig. 6.8 Schematic structure and function of a spreading resistor. The resistance is determined by the contact opening and not by the size of the sensor chip

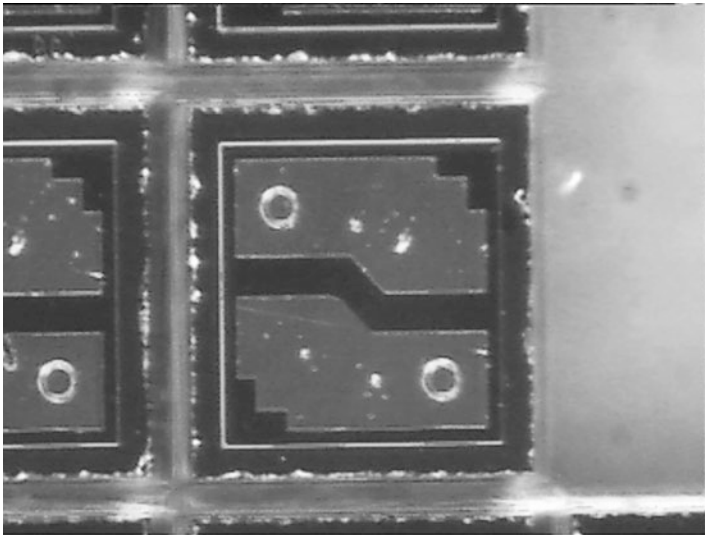


Fig. 6.9 Constriction resistances in the sawn wafer on a carrier of adhesive tape (Factory photo: NXP Semiconductors)

ρ : Specific resistance of the material at a given temperature,
 d : Diameter of the opening.

Thus, the resistor is *independent of the chip size and its tolerances*. Two elements connected in series are often used, so that the current enters the material through one constriction resistance and exits through another (Fig. 6.8).

Very uniformly and controlled *doped silicon base material* is used as base material. For this purpose, silicon wafers are exposed to neutron radiation. Natural silicon contains about 3.1% of the natural isotope $^{30}_{14}\text{Si}$, which is activated to $^{30}_{14}\text{Si}$ by neutron capture. This decays with a half-life of $T_{1/2} = 3.6$ h into $^{31}_{15}\text{P}$, emitting γ and β radiation. The remaining *phosphorus* serves as *doping*. Thus, the doping can be adjusted to an accuracy of about 1%, in contrast to 15% tolerance with conventional doping methods.

Figure 6.9 shows the constriction resistances of a sawn wafer on a carrier of adhesive tape “Blue-Tape”. Since the resistance is only determined by the entry and exit contacts to the silicon (visible on the top side of the chips), the tolerance during sawing has no influence on the basic resistance of the sensor.

6.3.6 Diodes

The *diode characteristic curve is temperature dependent*. The voltage drop U_D at a diode in forward direction at constant current I is

$$U_D = \frac{kT}{e} \ln \left(\frac{I}{I_0} + 1 \right)$$

K : Boltzmann constant ($1.38 \cdot 10^{-23}$ Ws/K)

T : Absolute temperature

e : Elementary charge ($1.602 \cdot 10^{-19}$ As)

I : Impressed current

I_0 : Saturation current (the saturation current for an ideal diode I_S describes the Shockley equation):

$$I(U) = I_S \cdot \left(e^{\frac{eU}{kT}} - 1 \right).$$

The saturation current is typically between 1 μA and 1 nA. It thus provides the basis for a linear temperature measurement. This principle can be easily integrated and miniaturized in electronic circuits with other functions. In the case of integrated temperature sensor circuits, the pn junction measuring the temperature is part of a transistor. Such a temperature sensor is called a “*PTAT*” (“Proportional To Absolute Temperature”) circuit.

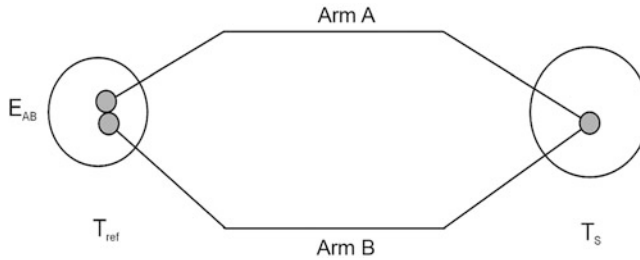


Fig. 6.10 Electromotive force (EMF) of a thermocouple with the weld T_S

6.4 Thermoelectricity (Seebeck Effect)

Two different metals A and B are connected by soldering or welding to form an electrically closed conductor circuit. If the connection points are kept at different temperatures, a *thermoelectric voltage* is measured, which is called *electromotive force (EMF)* (Fig. 6.10). This is the *Seebeck effect* (Sect. 2.10).

The following applies to the EMF E_{AB} or the thermoelectric voltage U_T according to Fig. 6.10:

$$U_T = \alpha_S (T_S - T_{\text{ref}})$$

with α_S : Seebeck coefficient,

T_S : Welded (usually warm) contact point,

T_{ref} : Reference temperature (mostly cold contact point).

Table 6.4 shows the Seebeck coefficient for selected metals, measured against platinum.

With this simple relationship between measured voltage and temperature shown above, the thermoelectric effect is very well suited for temperature measurement. Temperature measurement with the help of thermocouples is today a widely used state of the art technology. Table 6.5 lists the important properties for thermocouple pairs used in temperature measurement technology.

The thermocouples are standardized according to DIN IEC 584-1 and DIN 43710. An extract from these standards is shown in Table 6.6.

The thermocouple type J listed in the DIN standard has pure iron as its second leg. Above-average corrosion problems are to be expected when using this thermocouple, especially when the thermocouple is welded as a direct element into the protection tube bottoms and the protection tube is used in water-carrying cooling water networks of plants. Here rusting through occurs within a very short time.

With the standard types Cu-constantan/Fe-constantan, NiCr-Ni/Ni-constantan or CuNi/FeCuNi, a reduction of sensitivity can be observed. In any case, *cold embrittlement* must be

Table 6.4 Seebeck coefficient measured against platinum

Metal	$\alpha_s/\mu\text{VK}^{-1}$
Bismuth	-65
CuNi44	-35
Nickel	-15
Platinum	0
Iron	+18
Antimony	+48

taken into account in this range. The thermocouple combination gold/iron, NiCr is particularly suitable for the temperature range 220 °C to 270 °C and exhibits good stability and a largely linear characteristic curve even in continuous operation.

It is generally true for the magnitude of the thermoelectric voltage that the base metal pairs have the higher thermoelectric voltages compared to the semi-precious metals platinum-rhodium. The base metal pairings are available on the market at low cost. However, they are generally unsuitable for applications at temperatures above 1200 °C and for operation in aggressive media if they are not operated in gas-tight thermowell arrangements.

For precision temperature measurements, the non-linearities between temperature and thermoelectric voltage must be observed and, in the case of digital signal processing, electronic corrections must be made according to the linearization table. For analog indication, the instrument manufacturers use scales corrected according to the accuracy class. The standardization of the elements ensures their exchangeability in case of failure.

The *long-term stability* of thermocouples is essentially determined by the specific thermal, mechanical, abrasive, atmospheric and electrical stresses. The most important damage processes are:

- the oxidation of unprotected thermocouples;
- the alloying by diffusing *foreign atoms*, for example the siliconization of platinum-platinum/rhodium thermocouples;
- the *recrystallisation* and *segregation*;
- the *hydrogen embrittlement* of the superheavy metals (tungsten-rhenium and molybdenum-rhenium elements) and
- the *partial oxidation* of chromium in chromium/nickel-nickel elements in connection with green rot.

These damage processes must be taken into account by periodic calibrations and possible replacement of the thermocouples. The observed drift of the thermoelectric voltage increases with increasing temperature and operating time. According to IEC 584, the limit value deviations are divided into 3 tolerance classes each. Table 6.7 shows the standardized classes with the respective limit value deviations:

Table 6.5 Properties for thermocouple pairs used in temperature measurement technology

Element	Pt-PtRh10	Chromel P-Alumel	Fe Constantan	Cu-Constantan	Chromel P Constantan
Composition	100Pt 90Pt10Rh	90Ni10Mn 95Ni2Al2Mn1 Si	99.9Fe 55Cu45Ni	99.9Cu 55Cu45Ni	90Ni10Cr 55Cu45Ni
Temperature range/°C	0 to 1450	-200 to 1200	-200 to 750	-200 to 350	0 to 800
Max. temp. °C	1700	1350	1000	600	1100
Spec. resistance $10^{-6} \Omega \text{cm}$	9.8 to 18.3	69.4 to 28.6	10 to 49	1.71 to 49	70 to 49
Temp. coefficient.	0.0039	0.00035	0.005	0.0039	0.00035
0 °C to 100 °C	0.0018	0.0000125	0.00002	0.00002	0.00002
Melting temperature/°C	1773 1830	1430	1535 1190	1081 1190	1430 1190
EMF/mV related to 0 °C	100 °C: 0.643 200 °C: 1.436 400 °C: 3.251 600 °C: 5.222 800 °C: 7.330 1000 °C: 9.569 1200 °C: 11.924 1400 °C: 14.312 1600 °C: 16.674	4.10 8.13 16.39 24.90 33.31 41.31 48.85 55.81	5.28 10.78 21.82 33.16 45.48 58.16	4.28 9.29 14.86	6.3 13.3 28.5 44.3

Chemical resistance	Oxidizing and reducing atmosphere, not resistant to as-, Si- and P-containing gases, reducing gases, H ₂	Oxidizing atmosphere, not stable in reducing gases, Sulphur and water vapour atmosphere	Up to 400 °C in oxidising atmosphere, not resistant to Sulphur and Sulphur compounds	Oxidizing and reducing atmosphere up to 400 °C	Oxidation-resistant
Applications	Standard up to 1.450 °C	Electrical applications in oxidizing atmosphere	Industrial applications in process and control technology	Industrial applications in process and control technology, low temperatures	Industrial temperature measurement technology

Table 6.6 Thermocouple pairs according to DIN IEC 584-1 and DIN 43710

Metal partner	Element	KB	T_{max}° C	T_{Norm}° C	Application range/ ^p C
Iron constantan	Fe-CuNi	J	750	1200	-200 to 700
Copper constantan	Cu-CuNi	T	350	400	-250 to 300
Nickel chrome-nickel	NiCr-Ni	K	1200	1370	-250 to 1000
Nickel chromium constantan	NiCr-CuNi	E	900	1000	-250 to 700
Nicrosil-nisil	NiCrSi-NiSi	N	1200	1300	-250 to 1200
Platinum rhodium-platinum	Pt10Rh-Pt	S	1600	1540	-50 to 1400
Platinum rhodium-platinum	Pt13Rh-Pt	R	1600	1760	-50 to 1400
Platinum rhodium-platinum	Pt30Rh-Pt6Rh	B	1700	1820	0 to 1600
Iron constantan	Fe-CuNi	L	600	900	
Copper constantan	Cu-CuNi	U	900	600	

KB code letter of the thermocouple pairing

T_{max} Maximum temperature for defined limit value deviations

T_{Norm} Upper temperature limit for the validity of the standards

Table 6.7 Classification for limit value deviations of thermocouples

Thermocouple	KB	Class	Temperature range and accuracy
Fe-CuNi	J	Class 1 Class 2 Class 3	40 °C to +750 °C: $\pm 0.004 t$ or ± 1.5 °C 40 °C to +750 °C: $\pm 0.0075 t$ or ± 2.5 °C
Cu-CuNi	T	Class 1 Class 2 Class 3	40 °C to +350 °C: $\pm 0.004 t$ or ± 0.5 °C 40 °C to +350 °C: $\pm 0.0075 t$ or ± 1.0 °C 200 °C to +40 °C: $\pm 0.015 t$ or ± 1.0 °C
NiCr-Ni	K	Class 1 Class 2 Class 3	40 °C to +1000 °C: $\pm 0.004 t$ or ± 1.5 °C 40 °C to +1200 °C: $\pm 0.0075 t$ or ± 2.5 °C 200 °C to +40 °C: $\pm 0.015 t$ or ± 2.5 °C
NiCr-CuNi	E	Class 1 Class 2 Class 3	40 °C to +800 °C: $\pm 0.004 t$ or ± 1.5 °C 40 °C to +900 °C: $\pm 0.0075 t$ or ± 2.5 °C -200 °C to +40 °C: $\pm 0.015 t$ or ± 2.5 °C
NiCrSi-NiSi	N	Class 1 Class 2 Class 3	40 °C to +1.00 °C: $\pm 0.004 t$ or ± 1.5 °C 40 °C to +1200 °C: $\pm 0.0075 t$ or ± 2.5 °C 200 °C to +40 °C: $\pm 0.015 t$ or ± 2.5 °C
Pt10Rh-Pt	S	Class 1 Class 2 Class 3	0 °C to +1.00 °C: $\pm [1 + (t - 1.00)0,003]$ or + 0,5 °C 40 °C to +1.00 °C: $\pm 0.0025 t$ or ± 1.5 °C
Pt30Rh-Pt6Rh	B	Class 1 Class 2 Class 3	600 °C to 1700 °C: $\pm 0.0025 t$ or ± 1.5 °C 600 °C to 1700 °C: $\pm 0.005 t$ or ± 4.0 °C

6.4.1 Thermocouples for Very High Application Temperatures

For temperature measurements above the specification temperatures of platinum-based thermocouples ($>1500\text{ }^{\circ}\text{C}$), metallic and inorganic-non-metallic materials are available for thermocouples.

6.4.2 Metallic Thermocouples for Very High Application Temperatures

The basis for the construction of thermocouples is formed by the supermetals of the 5d elements of the periodic system of elements, especially tungsten, rhenium, osmium, iridium and homologues such as molybdenum, rhodium, ruthenium and the palladium of the 4d elements. These metals are characterized by high melting points. Technically interesting thermocouples are:

- Iridium-iridium 40 ruthenium,
- Iridium- Iridium 50 tungsten,
- Molybdenum-molybdenum 41 rhenium,
- Molybdenum 5 rhenium-molybdenum 41 rhenium,
- Tungsten 3 rhenium-tungsten 25 rhenium and
- Tungsten 5 rhenium-tungsten 26 rhenium.

For extremely high temperatures ($> 2500\text{ }^{\circ}\text{C}$), thermocouples based on tungsten/rhenium and ruthenium/osmium are recommended. Tungsten/rhenium thermocouples for application temperatures up to $2500\text{ }^{\circ}\text{C}$ in reducing and carbon-free atmospheres are used in industrial applications. Protection tubes on this basis with additives of tantalum, molybdenum or niobium increase the service life and prevent drift. At temperatures above $2000\text{ }^{\circ}\text{C}$, these thermocouples exhibit thermoelectric voltages in the range of 30 to 40 mV.

6.4.3 Inorganic-Non-metallic Thermocouples for Very High Application Temperatures

Graphite and the carbides sublimating at high temperature form the basis for the structure of inorganic-non-metallic thermocouples. Thermocouple pairs used are:

- Boron tetracarbide (B₄C) graphite,
- Silicon carbide graphite,
- Graphite tungsten,
- Graphite tantalum carbide and
- Silicon carbide-silicon infiltrated silicon carbide.

The assembly of these element pairs is carried out in graphite tubes and the operation requires reducing atmosphere. Disadvantages of these thermocouples, which can be used up to a maximum of 2500 °C, are their mechanical susceptibility and the short service life of these elements

6.5 Thermal Expansion

6.5.1 Thermal Expansion of Solid Bodies

The thermal expansion of solid bodies is closely related to the *chemical bond* and the *atomic* or molecular near-order that is formed. Experimental experience shows that the *thermal expansion of solid bodies increases with decreasing bonding energy*. Accordingly, the thermal expansion will increase with the following forms of bonding:

- Covalent chemical bond,
- Heteropolar chemical bond,
- Metallic bonding,
- Hydrogen bond and
- Van der Waals binding.

The following applies to the *change in length Δl* of a solid:

$$\Delta l = l_1 \alpha \Delta t \text{ or for temperature measurement: } t_2 = \Delta t / (l_1 \alpha) + t_1$$

l_1 : Length of the body before the temperature change

l_2 : Length of the body after the temperature change

Δl : Change in length ($l_2 - l_1$)

Δt : Temperature change ($t_2 - t_1$)

α : Coefficient of linear expansion in K^{-1} ($\alpha = \Delta t / (l_1 \Delta t)$).

Table 6.8 lists the coefficients of thermal expansion of various materials.

For isotropic or cubic solids at constant pressure, the volume expansion coefficient γ results:

$$\gamma = 3\alpha.$$

For solids with hexagonal, trigonal or tetragonal crystal structure applies:

$$\gamma = 2\alpha_x + \alpha_z.$$

For rhombic, monoclinic and triclinic crystal systems, the same applies:

$$\gamma = \alpha_x + \alpha_y + \alpha_z.$$

Table 6.8 Expansion coefficients of selected materials

Metal/ metallic bond	α 10^{-7} K^{-1}	Ceramic/ covalent bond	α 10^{-7} K^{-1}	Ceramic/ heterogeneous bonding	α 10^{-7} K^{-1}
Tungsten	44	Silica glass SiO_2	5	BeO	80
Molybdenum	45	TiN^1	24	Al_2O_3	75
Platinum	90	Si_3N_4	32	ZrO_2	100
Iron	122	BN	41	Mn-Zn ferrite	95
Gold	193	AlN	54	Ni-Zn ferrite	105
Copper	165	TiC^1	21	MgTi_2O_4	
Nickel	172	B4C	45	BaTiO_3	
Cadmium	308	SiC	45		
Lithium	580	WC	52		
Sodium	725	TaC	63		

Solid bodies expand in three dimensions. It applies:

$$\Delta V = V_1 \gamma \Delta t \text{ or for temperature measurement: } t_2 = \Delta V / (V_1 \gamma) + t_1$$

V_1 : Length of the body before the temperature change

V_2 : Length of the body after the temperature change

ΔV : Volume change ($V_2 - V_1$)

Δt : Temperature change ($t_2 - t_1$)

γ : Volume expansion coefficient in K^{-1} ($\gamma = \Delta V / (V_1 \Delta t)$)

Thermal expansion plays an important role in temperature measurement technology. In *precision metrology*, differences in *thermal expansion* in film resistances on substrates, in sandwich arrangements and joints often contribute to the *hysteresis* of measured values and to *aging* and *degradation* of the sensor systems. The cause is then usually mechanical stress, which can lead to subcritical crack growth or catastrophic failure. For high-temperature applications, it contributes significantly to *thermal shock resistance* or *thermal fatigue*. The permissible *temperature shock difference* ΔT during heating up or cooling down results:

$$\Delta T = \frac{\sigma_m \cdot (1 - \nu)}{\alpha \cdot E}$$

E_m : Modulus of elasticity

ν : Poisson's ratio

σ_m : Mechanical strength.

Thermal shock resistant materials must therefore have a small thermal expansion. Due to the high melting or sublimation temperatures, covalently bonded ceramics are particularly suitable for thermowell applications in aggressive media and at high temperatures.

Bimetallic thermometers are particularly robust and inexpensive (Fig. 6.11). *Thermobimetals* consist of metals that are firmly bonded together and have *different*

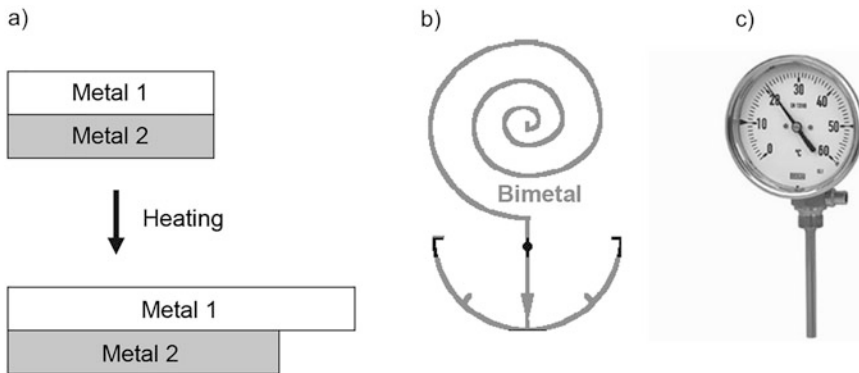


Fig. 6.11 Bimetallic thermometer (a) principle; (b) mode of operation; (c) thermometer (Factory photo: WFKA)

coefficients of thermal expansion. If the thermobimetal is heated, the metals expand differently (Fig. 6.11 a)). If the metals are firmly bonded to each other (usually by a plating process), a *curvature* is created in the direction of the metal with the lower thermal expansion. If the thermobimetal is spirally wound, the spiral expands and a pointer indicates the temperature (Fig. 6.11 b and c). The metal with the higher coefficient of thermal expansion is called *active component* (coefficient of thermal expansion $\alpha > 15 \cdot 10^{-6} \text{ K}^{-1}$), and the metal with the lower coefficient of thermal expansion is called *passive component* (coefficient of thermal expansion $\alpha < 5 \cdot 10^{-6} \text{ K}^{-1}$). Bimetal thermometers can be used between $-80 \text{ }^\circ\text{C}$ and $+550 \text{ }^\circ\text{C}$. The standard DIN 1715 applies to the thermobimetals.

Table 6.9 lists the properties of common thermobimetals.

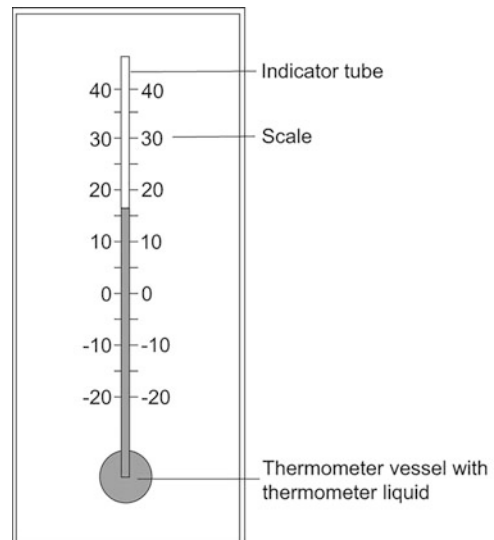
6.5.2 Thermal Expansion of Liquids

The expansion of liquids is very often used for temperature measurement. According to VDI/VDE Guideline 3511, a distinction is made between *liquid-glass thermometers* (measurement of volume change) and *liquid spring thermometers* (measurement of pressure change; deformation provides a pointer deflection).

In the liquid-glass thermometer, the liquid is contained in a small glass bulb. At a certain temperature, the liquid in the glass capillary rises and a scale shows the value (Fig. 6.12). The liquids must be liquid in the temperature range to be measured and must not attack the walls of the measuring tubes. Mercury is particularly well suited for measurements ($-40 \text{ }^\circ\text{C}$ to $+360 \text{ }^\circ\text{C}$), but today it is only used in scientific applications due to its high toxicity.

Table 6.9 Chemical and thermal properties of typical thermobimetals (Source: VAC)

Chemical composition weight-%	Therm. curvature 10^{-6} K^{-1}	Therm. deflection 10^{-6} K^{-1}	Linearity range ($^{\circ}\text{C}$)	Critical temperature ($^{\circ}\text{C}$)
25.3% Ni, 30.3% Fe, 39, 1% Mn, 5.3% Cu	39.0	20.8	-20 to +200	400
28.2% Ni, 68.6% Fe, 3.2% Mn	28.5	15.2	-20 to +200	450
30.0% Ni, 1.3% % Cr, 68.7% Fe	26.3	14.0	-20 to +200	450
31.2% Ni, 65.3% Fe, 3.5% Mn	22.0	11.7	-20 to +380	450
32.8% Ni, 63.9% Fe, 3.3% Mn	18.6	9.9	-20 to +425	450
32.8% Ni, 63.6% Fe, 3.3% Mn	18.6	9.9	-20 to +425	500

Fig. 6.12 Liquid glass thermometer (Source: Wikipedia)

The following liquids are mainly used:

- pentane (-200°C to $+35^{\circ}\text{C}$),
- ethanol (-120°C to $+60^{\circ}\text{C}$) and
- toluene (-90°C to $+110^{\circ}\text{C}$).

Further information and regulations for liquid glass thermometers are compiled in DIN 12770.

6.5.3 Thermal Expansion of Gases

The *gas thermometer* measures pressure p and volume V of a precisely defined gas quantity. The equation of state for ideal gases is: $p V = m R T$. This results for the temperature T :

$$T = (pV)/m R$$

where m : mass of the gas and R : specific gas constant.

In practice, the temperature T of a *gas thermometer* is measured by the *gas pressure* p . It applies:

$$T = \frac{(p - p_0)}{p_0 \alpha_p}.$$

It is p_0 : gas pressure at 0 °C and α_p : stress coefficient (ratio of the relative pressure change $\Delta p/p_0$ and the temperature T : $\alpha_p = (\Delta p)/(p_0 T)$) The stress coefficient of an ideal gas is equal to the volume expansion coefficient γ and is: $\gamma = 3.661 \cdot 10^{-4} \text{ K}^{-1}$.

The measurement shall be made either at constant pressure or at constant volume. The gas thermometer has a very wide application range (from -273 °C to $+2800 \text{ °C}$), as shown in Fig. 6.3. Thermometers usually filled with helium (He) or nitrogen (N_2) have an operating range of -200 °C to $+800 \text{ °C}$. Figure 6.13 shows a gas pressure thermometer. The electrical output signal is provided by a thermocouple integrated in the measuring system.

Tension thermometers are used in low temperature measurement technology. In these, the measured *vapour pressure* is a measure of temperature according to the *Clausius–Clapeyron equation*.

6.6 Temperature and Frequency

Many physical quantities are *temperature-dependent*. This fact must be taken into account for most sensors. However, there is also the possibility to use temperature dependence as a *measuring method* for temperature sensors. For this purpose, the *characteristic curve*, i.e. the temperature course of the physical quantity (e.g. pressure, electrical resistance, capacitance, inductance) is measured. The temperature can then be calculated from the measured value of the physical quantity. However, this generally involves a great deal of effort and therefore has little practical significance.

A method shown in patent DE4200578A1 can be implemented easily and cost-effectively with conventional components. The sensor described there consists of *two piezoelectric oscillating crystals*, a memory device and an electronic evaluation circuit. The first oscillating crystal shows high temperature dependence and the second oscillating

Fig. 6.13 Gas pressure thermometer (Factory photo: WIKA, Type 75 with Type K thermocouple)



crystal is temperature independent. The frequencies of the first oscillating quartz and the second oscillating quartz are counted. The temperature is determined by comparing the two counter readings.

6.7 Thermochromism

Thermochromic substances change their *colour* at certain temperatures. This can happen reversibly or irreversibly. The reason is that at certain temperatures, *phase transitions* take place in which other regions of the visible spectrum are absorbed. The reflected light therefore has a different colour. Inorganic materials are zinc oxide, rutile and mercury iodide (II) compounds. As organic materials, bixanthyliene (9.9') and bianthronyliene (10.10') are possible. The following applications are typical:

- *Thermal lacquer.* The colour of a painted surface is a measure of the temperature. This serves as a safety estimate for parts that are difficult to access in the chemical industry.
- *Food packaging.* The colour indicates whether the cold chain was interrupted or not.
- *End users.* *Mood rings* indicate the skin temperature, *magic cups* give an indication of the drinking temperature or the filling level of a container. *Spoons* for small children

indicate whether the food is at a tolerable temperature for consumption. *Thermofoils* on aquariums give an indication of the temperature in the aquarium.

6.8 Seger Cone

Seger cones are used for monitoring and control during *firing processes*. Even the most modern control and regulation technology and optical monitoring sensors (Sect. 3.1.2) have not been able to displace these robust sensors. The reason is that the Seger cones are made of the *same raw materials* as the firing masses (mostly ceramic). During the firing process, they are thus subjected to the same effects of temperature and time as the fuel. These complex interrelationships can be seen very well with this simple method. Figure 6.14 shows three Seger cones, on the left before the insert and on the right after the insert.

The standard cones are usually 5-cm high and are placed in the kiln with the material to be fired. At a certain temperature, the cones tilt around touching the bottom. The firing temperature can be read from *tables*. In Fig. 6.14, the *middle cone (guide cone)* indicates the *desired firing temperature*, while the right one tilts as an observation cone at a higher temperature and the left one at a lower temperature. In the right figure, you can see that the middle cone has melted with its tip at the bottom, but the right one has not. This means that the firing temperature (and also the whole firing process) was just right. If the middle cone had melted too much, there would have been an *overfiring*; if it had only melted, there would have been an *underfiring*.

Temperature measuring rings are also used instead of tapered cones. The *diameter of the ring* is used to determine the temperature (and the combustion process) in the kiln.

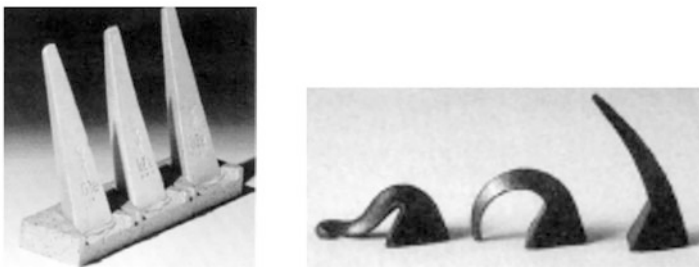


Fig. 6.14 Seger cone: left before its use and right after its use (Factory photo: Töpferhof Brockhagen)

6.9 Non-contact Optical Temperature Measurement

The following two effects are used in optical temperature measurement:

- The *thermal radiation* of the object with *radiation thermometers (pyrometers)* and
- The *backscattering of light quanta* (photons) at the *phonons* (quantified lattice vibrations). The underlying effect is the *Raman effect*.

6.9.1 Radiation Thermometer (Pyrometer)

Every body emits radiation according to its temperature. The *radiation power P_s* is calculated according to *Stefan–Boltzmann’s law*:

$$P_s = \varepsilon \sigma A(T_1^4 - T_2^4).$$

It is: ε : emissivity of the radiating surface; σ : radiation constant ($5.671 \cdot 10^{-8} \text{ W}/(\text{m}^2\text{K}^4)$)

A : Radiating surface of a body

T_1 : Temperature of the radiator

T_2 : Temperature of the environment (sensor).

The radiant power P_s contains the wavelengths of the entire spectrum. The contributions to radiant power are in turn dependent on the wavelength and are not evenly distributed over the spectrum. Figure 6.15 shows the specific spectral emission M as a function of the wavelength λ and the temperature T .

The product $\varepsilon \sigma$ is a measure for the *optical coupling* between the surface to be measured and the sensor. If the optical coupling, the radiating surface A and the temperature of the sensor T_2 are known, the temperature at the surface of the radiator can be measured by measuring the radiant power P_s . This is done with a *pyrometer*. Figure 6.16 shows a pyrometric application in medical technology.

With suitable optics, a sensor with 0.5 mm^2 sensitive membrane surface (total membrane surface 1 mm^2) receives approximately 0.1-mW radiation from a $40 \text{ }^\circ\text{C}$ warm object at $25 \text{ }^\circ\text{C}$. This results in an electrical signal of about 1.5 mV, which can be used to determine the temperature of the emitting radiator to an accuracy of $0.1 \text{ }^\circ\text{C}$ within about 1 s. This method is used in commercial *ear thermometers* (Fig. 6.16). On the top of the sensor is the silicon window transparent to infrared radiation (with additional coatings to filter the wavelengths and reduce reflection).

Pyrometric temperature measurement covers a very wide temperature range (from $-50 \text{ }^\circ\text{C}$ to $+4000 \text{ }^\circ\text{C}$).

A non-contact, pyrometric temperature measurement has the following advantages:

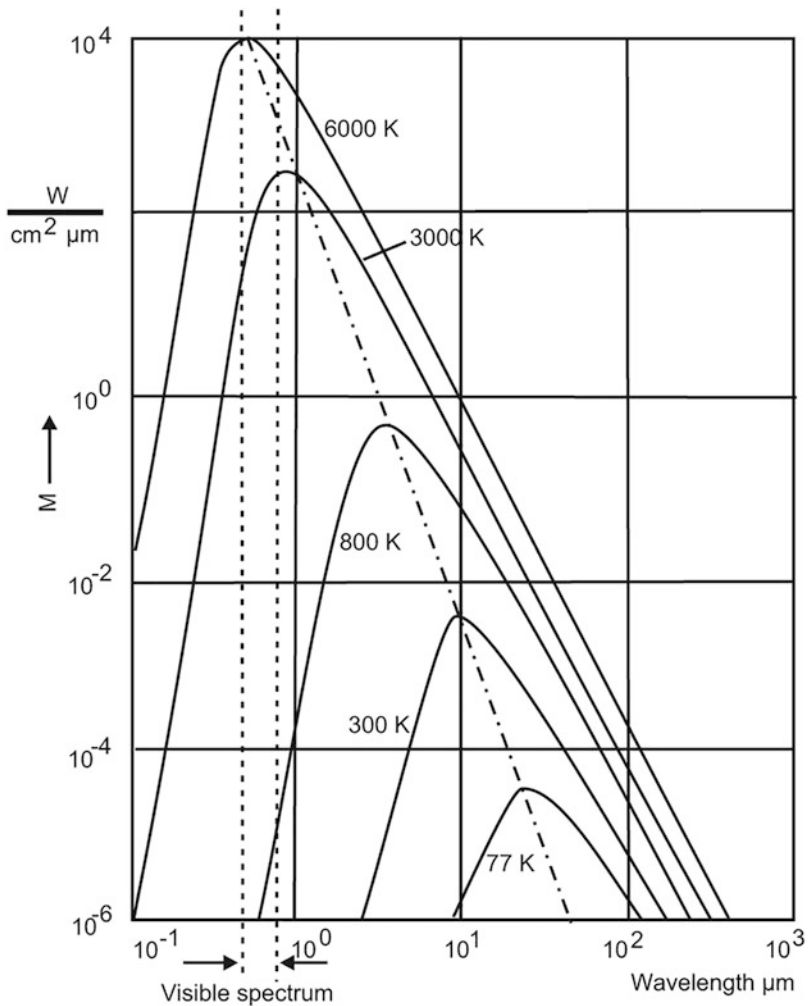


Fig. 6.15 Spectral emission M as a function of wavelength and temperature

- *non-contact measurement*,
- *no wear and tear*,
- *no mechanical damage* to sensitive measurement objects (e.g. thin foils),
- *very fast measurement* (10 μ s to 1 s),
- *very large, coherent measuring ranges* (e.g. from 200 $^{\circ}$ C to 3000 $^{\circ}$ C),
- *no problems with moving targets* and
- *measurement even in adverse environments* (e.g. aggressive media, high electric or magnetic fields).

The disadvantages are:

Fig. 6.16 Pyrometric sensor (ear thermometer) for contactless temperature measurement in medical technology (Factory photo: PerkinElmer Optoelectronics)



- the *emissivity* e of the measured object must be known and
- in the case of metals, *strong fluctuations in emissivity* (depending on the surface condition) often cause great problems in the accuracy of the measurements.

The pyrometers can be used portable or stationary. The *portable* measuring instruments are used for *diagnosis* and *inspection* in the automotive sector as well as in heating, air conditioning and ventilation technology. Figure 6.17 shows one such portable device.

With the measuring instrument in Fig. 6.17, temperatures from $-32\text{ }^{\circ}\text{C}$ to $+530\text{ }^{\circ}\text{C}$ can be measured within a measuring time of 0.3 s and a measuring accuracy of $\pm 0.1\text{ }^{\circ}\text{C}$. The built-in laser fixes the measurement object and the measurement is triggered by pressing a button.

Table 6.10 lists some important applications of mobile pyrometers.

Stationary pyrometers are mostly used for quality assurance in production lines. Not only can the measured values be recorded, but they also serve to *control* the temperature curves. Figure 6.18 shows some applications in rolling processes in the paper and glass industry, in laser cutting systems and in automatic filling machines.

Table 6.11 shows typical applications of stationary pyrometers.

Thermographic cameras can also be used. They are *spatially resolving* pyrometers.

The various detection options are described in detail in Chap. 14 (Photoelectric Sensors).

6.9.2 Fiber Optic Applications

6.9.2.1 Intrinsic Sensors, DTS (Distributed Temperature Sensing)

DTS is based on the *backscattering* of *photons* (light quanta) at *phonons* (quanta of mechanical vibrations in the solid state). Since the distribution of photons in a medium is determined by its temperature, conclusions can be drawn from the *backscattered light about the temperature* of the fiber at the point of scattering. Based on the transit time of the



Fig. 6.17 Portable pyrometer (Factory photo: Optris)

Table 6.10 Applications of mobile pyrometers

Field of application	Examples of use
Electrical machinery and equipment	Electrical contacts and connections Control of fuses Overload control Defects in cable ducts Winding short circuits in transformers
Automotive	Check the brakes, air conditioning, heating, cooling system Detection of motor faults
Heating, air conditioning and ventilation technology	Detection of leaks Checking supply and exhaust air outlets and safety valves Adjusting the indoor climate Testing of burners in gas and oil heating systems Testing of heat exchangers and heating circuits



Fig. 6.18 Applications of stationary pyrometers in manufacturing (Source: Optris)

Table 6.11 Applications of stationary pyrometers

Field of application	Examples of use
Mechanical engineering	Plastics, glass and paper industry Laser welding and cutting machines Control of the temperature profile on rollers
Electronics	Temperature control during wafer production and testing

light in the fibre, the location of the respective scattering can be determined. DTS thus allows the measurement of a *temperature distribution* in a *longer optical fiber* (up to about 20 km).

The predominant scattering mechanism in the material is elastic Rayleigh scattering, in which the energy and thus the wavelength of the scattered photon remains constant. In addition, there is *Raman scattering*, in which energy is exchanged between the photons and the material. If energy is transferred to material vibrations, this is called a *Stokes process*; in the opposite case, it is called an *anti-Stokes process*.

While the *amplitude of the Stokes band* (at longer wavelength) is almost *temperature independent*, the amplitude of the *anti-Stokes band* (at slightly shorter wavelength) depends on the *temperature* of the scattering medium. The *temperature can be calculated* from this *ratio*.

The method is used for *temperature measurements over longer distances*, for example for leakage monitoring in pipelines, fire monitoring in tunnels or temperature control in nuclear repositories. Other applications include temperature monitoring of *large hardening concrete pieces*, for example in the construction of dams. The glass fibres can simply be left in the structure after the measurement is completed. The measuring device is then disconnected.

Advantages of these methods are:

- The sensor itself (multimode optical fiber made of quartz glass) is *inexpensive* and can be left behind if required.
- The measurement is *independent of influences* (aging, mechanical stress) of the sensor.
- *Continuous measurement over long distances*.
- The sensor is a non-conductor, therefore *not affected by microwaves* or *electrical currents*.

However, the disadvantage is that these measuring systems are very complex and therefore expensive.

Since the optical fiber itself is also the temperature sensor in the DTS, it is referred to as an *intrinsic sensor* (Fig. 6.19).

6.9.2.2 Extrinsic Sensors

In extrinsic sensors, the *sensor is not identical to the fiber*. Figure 6.20 shows a measuring device for temperature measurement using fiber optic probes (left). The fibers end on a

Fig. 6.19 DTS system for temperature measurement in a quartz glass fibre (intrinsic sensor) (Factory photo: Sensornet)

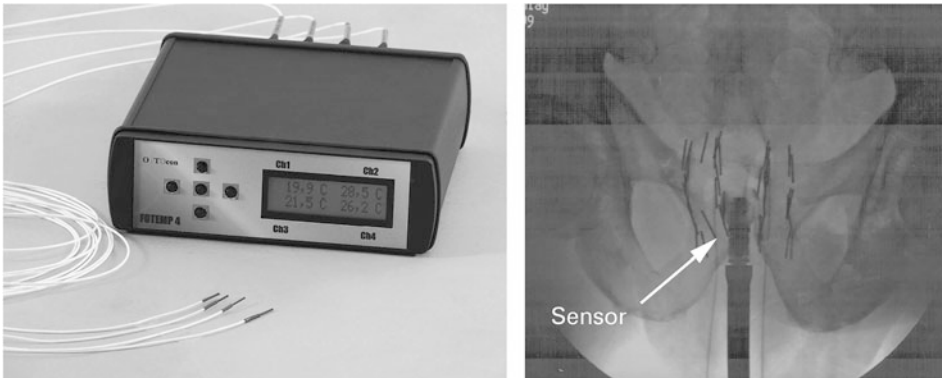


Fig. 6.20 Left: Measuring device for temperature measurement with fiber-optic probes. Right: X-ray image of a thermal cancer therapy (Factory photo: Dr. C. Renschen, Optocon company)

GaAs crystal whose optical absorption edge is temperature-dependent (extrinsic sensor). On the right side of the figure, you can see an X-ray image of a thermal cancer therapy. Fiber-optic probes of 0.5-mm diameter monitor the tissue temperature in a minimally invasive way while the tissue is heated with an alternating magnetic field. The thin lines are metal needles that switch off their conductivity above 41 °C. The thick black part is a video camera that is removed before the heating process begins. The basis of the therapy is the lower heat resistance of the cancer tissue.

Fiber-optic methods for temperature measurement are also used where no other possibilities can be used due to the special boundary conditions of the measurement. This is the case, for example, when microwaves are used for heating and the temperature must be monitored at the same time without the temperature measurement and heating

influencing each other. Instead of the complex DTS technique, temperature-sensitive *optical absorbers* or *Bragg gratings* burnt into the fiber can be used, for example.

Bibliography

1. DIN IEC 60751:2005-03 Industrielle Platin-Widerstandsthermometer und Platin-Sensoren
2. H. Worch GmbH & Co. KgaA Weinheim, 9. Aufl. (2002), ISBN 3-527-30535-1
3. Huhnke, D.; Maier, U.: Temperaturmesstechnik, atp Praxiswissen kompkt, Oldenbourg Industrieverlag Wie, 4. Auflage 2006 ifm electronic: Schulungsunterlagen Temperaturmesstechnik, 2011
4. Michalowsky, L.: „Neue Technische Keramikwerkstoffe“, Verlag für Grundstoffindustrie GmbH Leipzig, (1994), ISBN 3-342-00489-4
5. Physikalisch-Technische Bundesanstalt: Praktische Temperaturmessung, 2010
6. Physikalisch-Technische Bundesanstalt: Thermodynamische Temperaturen und die IST-90-Skala, 2010
7. VDI-Verlag: VDI/VDE-Richtlinien 3511, Blatt1-41, Technische Temperaturmessungen, Grundlagen und Übersicht über besondere Temperaturmessverfahren
8. Weber, D.; Nau, M.: Elektrische Temperaturmessungen, Verlag M. K. Juchheim, 6. Auflage 1997



Gert Schönfelder and Andreas Wilde

7.1 Voltage

7.1.1 Definition

The voltage U is a measure of the charge separation work W per charge Q . It applies:

$$U = W/Q. \quad (7.1)$$

The definition can also be extended by the dimension of time. If the work W is described as the product of the power P and the time t ($W = P t$) and the charge Q as the product of the current I and the time t ($Q = I t$), then the following applies: The voltage U is the quotient of the power P and the current I :

$$U = P/I. \quad (7.2)$$

The unit of measurement for the voltage is 1 V. One volt is between two points of a metallic conductor if an energy of 1 J (Joule) is required to transport the charge of 1 C (Coulomb) (Eq. 7.1). One volt is between two points of a metallic conductor when, with a constant current I of 1 A, a power P of 1 W is converted between the two points.

The following units are therefore valid for the unit of measurement volt:

G. Schönfelder (✉)
Prignitz Mikrosystemtechnik, Wittenberge, Germany

A. Wilde
Fraunhofer Institut für Integrierte Schaltungen Dresden, Dresden, Germany
e-mail: Andreas.Wilde@eas.iis.fraunhofer.de

$$1 \text{ V} = 1 \text{ J/C} = 1 \text{ W/A} = 1 \text{ m}^2\text{kg}/(\text{s A}).$$

The electric voltage occurs in different forms:

- as *electrostatic voltage* caused by charge separation in a limited environment (e.g. electrical charging of objects or thunderstorms);
- in *chemical processes* by migration of electrically charged particles (ions), for example in accumulators or galvanic effects (Sects. 2.16 and 2.11);
- as *contact* and *thermoelectric voltages*;
- as *noise voltage* due to the movement of charge carriers.

The voltage can be static (DC voltage), dynamic (AC voltage) or a mixture of both forms. The *Weston standard element* is considered to be the normal, which is a primary chemical element based on cadmium sulphate (electrolyte) and mercury and cadmium amalgam as electrodes. It provides a voltage of 1.01865 V, which is temperature-dependent.

A more modern alternative is the *Josephson tunnel element*, which emits a voltage of about 1 mV under microwave irradiation. By connecting large quantities of these elements in series, almost any reference voltage can be generated with a reproducibility of up to 10^{-12} .

7.1.1.1 AC Voltage

An *alternating voltage* is characterized by a constantly repeated change of polarity. The shape of the curve can be varied, such as sine, square, triangular or sawtooth oscillations. These and other waveforms can also occur in combination and with a superimposed DC voltage. The most common form is the sine wave, as it technically produces the least distortion and losses.

The sine wave is technically the basic form, since all waveforms can be traced back to a composition of sine waves of different frequencies by means of the *Fourier series*. The frequency components of any AC voltage are determined by means of Fourier transformation (*FFT—fast fourier transformation*), which as a complex representation gives the frequency spectrum and its power components. The practical side of Fourier analysis is, among other things, the dimensioning of bandwidths for the transmission of non-sinusoidal signals.

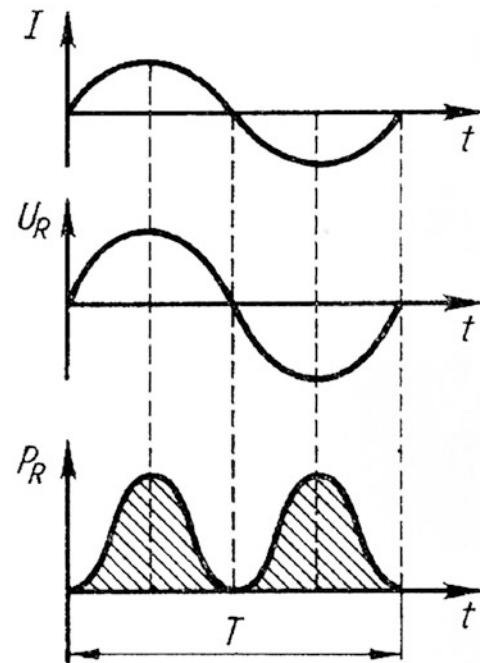
Important characteristic values of an AC voltage are the *period duration*, the *peak voltage* and the *effective value*, which are shown in Fig. 7.1.

The *period duration* is the time between two zero crossings in the same direction.

The *peak voltage* is the highest occurring voltage value U_s of the oscillation. In the case of asymmetrical oscillations, the value from positive to negative peak voltage is also indicated as U_{ss} .

The effective value is the voltage value of a DC voltage that would produce the same effective power at an ohmic resistor as the AC voltage used.

Fig. 7.1 Characteristic values of a sine wave



$$P = \frac{1}{T} \int_0^T I_m \sin \omega t \, dt \cdot U_m \sin \omega t \, dt \quad P = \frac{1}{2} \frac{U_m^2}{R} = \frac{1}{\sqrt{2}} U_m \cdot \frac{1}{\sqrt{2}} I_m \quad (7.3)$$

These are

U_m, I_m : Maximum values (peak values) of voltage and current

T : Period duration

P : Active power.

7.1.1.2 Measuring Principles

An old principle is the *electrometer*, which uses the repulsion of equally charged bodies. It is a very rough measurement, but is still used in high voltage technology. An electrometer takes no energy from the source. This can be a great advantage.

The generally applied measuring methods *take energy from* the voltage source, which is converted into another form for measurement. This loading of the source results in a measurement error, which must be taken into account accordingly. The following conversions are relevant:

7.1.1.3 Conversion into an Electromagnetic Field

These include the majority of classic *pointer instruments*, such as moving coil and plunger coil instruments. A mechanical comparison is made between the force generated by the field and a mechanical counterforce (e.g. in the form of springs).

These systems are particularly suitable for measuring *DC voltage*. The measurement of AC or DC voltage superimposed with AC voltage requires additional measures such as the generation of a phase-shifted induction or capacitance by a short-circuit winding.

7.1.1.4 Conversion to Heat

The measured variable generates a quantity of heat at a heating element, which can be evaluated as a change in temperature at the heating element. This method has the advantage that it determines the effective value of the voltage and is therefore also suitable for measuring alternating voltages. The curve shape of the voltage is not important.

7.1.1.5 Conversion into a Current

Although the methods mentioned above are also based on the conversion of voltage into current at a defined internal resistance, they draw a not insignificant amount of power from the voltage source for this process.

If the voltage to be measured is converted via a resistor into a proportional current and this current is integrated, then the applied voltage can be inferred from a time measurement during integration. The conversion of the voltage into a time achieved in this way ensures a comparison with a readily available standard. The input currents can be kept very low. (Fig. 7.2).

For a DC voltage at the input, the output voltage follows the equation

$$\Delta U_a = -\frac{U_e}{C \cdot R_e} \cdot \Delta t \quad \text{and thus results in } U_e = -\frac{\Delta U_a \cdot C \cdot R_e}{\Delta t}.$$

Since the integration reacts to the sign of the input voltage, it can be deduced that a symmetrical AC voltage would yield the value 0 as a result.

This circuit is susceptible to tolerances and drift of the components and thus of the reference points. To compensate for this, a double integration is used in practice, in which the charge ($Q = I \cdot t$) accumulated in a fixed time is removed in a second integration phase by means of a reference voltage. This results in the construction of the *two-flank converter* as shown in Fig. 7.3.

Figure 7.4 shows that the result of the measurement is only determined by the ratio of the two times and the reference voltage.

7.1.1.6 Measurement by Comparison with a Standard

In order to minimize the energy extraction from the voltage source for the measurement, the unknown voltage can be determined by comparison with a standard. Although this method is old, it is gaining importance again due to the availability of graduated standards—mostly

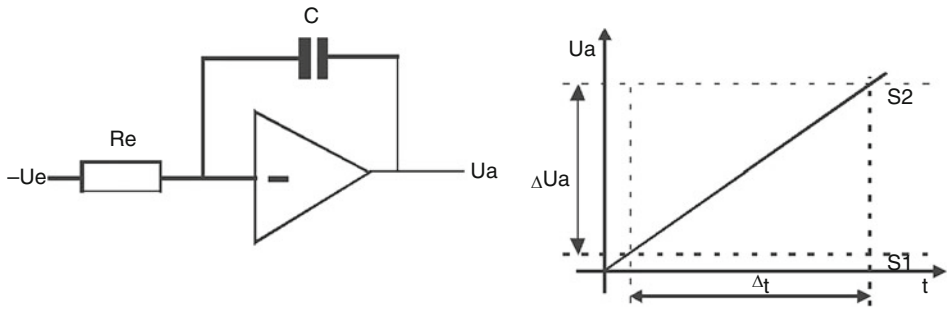


Fig. 7.2 Integration of the measuring voltage

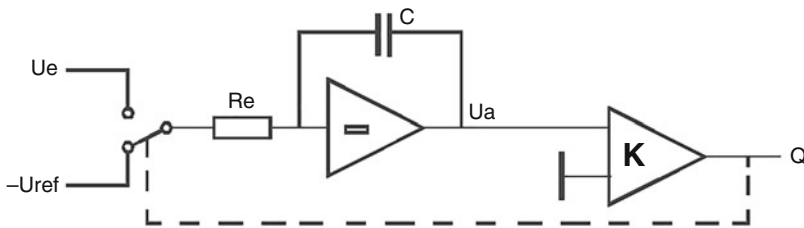


Fig. 7.3 Functional diagram of a two-edge converter (double integrator)

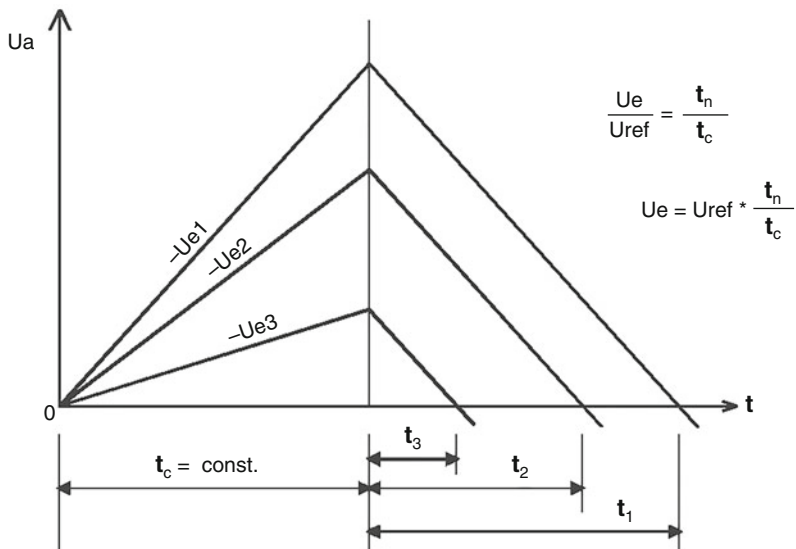


Fig. 7.4 Progression of the output voltage with the two-edge converter

on an electronic basis. Here, too, the problem is that the measured variable must remain stable during the measurement process, which excludes its use for AC voltages. Ways out of this are:

- The measured variable is “frozen” during the measuring process, which is done by scanning and holding stages. However, this results in the measurement of an *instantaneous value*, which, depending on the curve shape, can deviate considerably from the actual value.
- A measurement is carried out with the aid of *additional* modules such as rectifiers or peak value rectifiers.
- A measurement with sufficient sampling frequency is carried out so that the required characteristics can be determined mathematically from the measured voltage curve.
- Measurements are taken at a very high sampling rate and compared with randomly generated voltage values. *Statistical methods* (e.g. the Monte Carlo method) can be used to determine the characteristic values of the applied voltage signal, such as the mean value, the RMS value or the DC component.

7.1.2 Measuring Arrangements

7.1.2.1 Measurement by Energy Extraction

A voltage measurement is performed by *connecting* the measuring device in *parallel* to the voltage source. Since both have an internal resistance, the result is an electrical series connection of the actual source, its internal resistance R_{iq} and the internal resistance of the measuring device R_{im} . These internal resistances thus represent a voltage divider, which causes a deviation of the measured voltage from the actual voltage (Fig. 7.5).

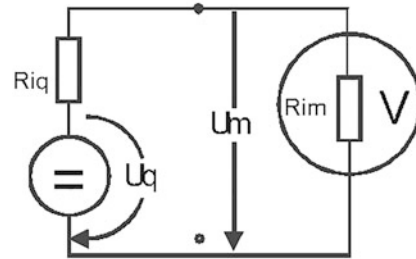
The relative measuring error of the measuring voltage U_m is thus

$$\text{Relative measuring error} = 1 - (R_{im}/(R_{im} + R_{iq})).$$

The accuracy of the measurement is thus essentially determined by the ratio of the two internal resistances R_{im} and R_{iq} . For an error-free measurement, the source resistance R_{iq} must assume the value zero or the resistance of the measuring device must assume the value infinite.

The internal resistances of pointer instruments are in the range of some 10 k Ω /V, i.e. the internal resistance was specified in relation to the measuring range. The internal resistance of electronic voltmeters is in the range of some M Ω .

Fig. 7.5 Arrangement for voltage measurement



7.1.2.2 Measurement Through Integration

Another possibility for voltage measurement is the extraction of *small amounts of energy* and their *integration*. In this way, the current load on the source can be kept low and disturbances caused by the integration process are suppressed.

Many electronic voltmeters work with this method. The most important circuitry representatives here are the two-flank converter and the sigma-delta converter.

With the *two-edge converter* (Figs. 7.3 and 7.4), a charge is built up within a fixed time with the measured variable on a capacitor. In the second phase, this charge is compensated by means of a known reference voltage and the required time is determined. The charge and thus the input voltage can be determined from the known two times and the reference voltage.

A large part of the *multimeters in use today* work on this basis, since the integration makes the measurement immune to interference and can eliminate superimposed AC voltage or noise. A measurement of pure AC voltages is therefore not possible without additional modules, since the integral is zero via a symmetrical oscillation.

The *sigma-delta converter* (also known as 1-bit ADC) also integrates the input variable. If the output voltage of the integrator exceeds a *threshold value*, a reference voltage is subtracted at the input until it falls below the threshold value again. The threshold value is interrogated in a time-discrete manner. In its distribution of states 0 and 1, the *bit stream* generated at the output represents the ratio between the measured variable and the reference voltage, which is output as a numerical value by filtering.

The circuit in Fig. 7.6 consists only of an integrator, a clocked comparator and a 1-bit DA converter, which corresponds to an analog switch between 0 V and a reference voltage. As a condition for the function, it is assumed that the input voltage is lower than the reference voltage.

The output signal of the sigma-delta converter is a pulse train whose ratio corresponds to the ratio of input voltage to reference voltage. (Fig. 7.7) This pulse sequence is filtered in the downstream unit and output as a numerical value.

It can be assumed that almost all integrated AD converter circuits with a resolution above 16 bits work according to this principle.

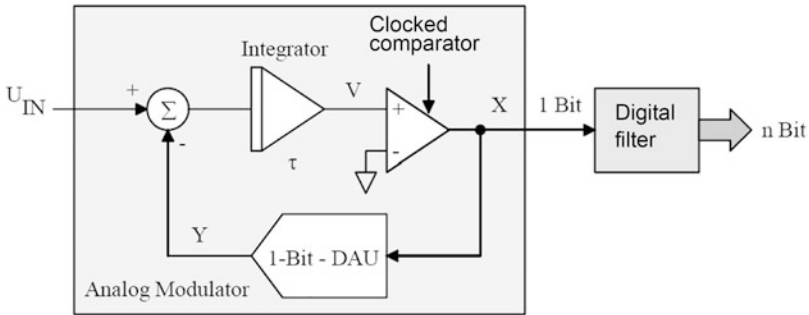


Fig. 7.6 Basic circuit of a sigma-delta converter

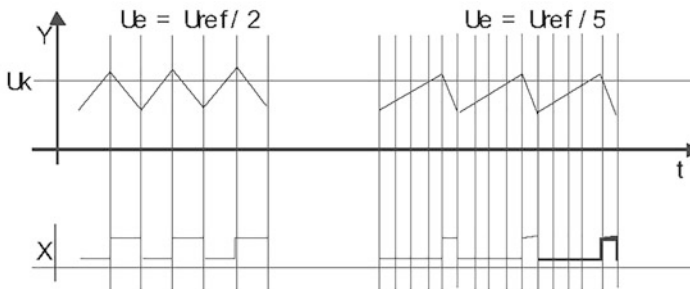


Fig. 7.7 Signal waveform at the sigma-delta converter

7.1.2.3 Measurement by Comparison

For comparison, one side is formed by the voltage to be measured, the other by a *variable reference value*. The basic scheme of all comparative measurement methods is shown in Fig. 7.8. The core is a digital-to-analog converter (DAU), which derives stepped output voltages from a reference source. These are compared with the voltage to be measured. The control system determines the method used for this.

7.1.2.4 Servo Converter

In this case, the control consists of an *up/down counter*. If the input voltage is higher than the output voltage of the DA converter, the counter value—and thus the output voltage—is increased, if it is lower, it is decreased. The analog comparison value, and thus also the digital representation, oscillates around the input value. This measuring method is suitable for static or slowly changing voltages.

7.1.2.5 The Successive Approximation

This method also works with the arrangement shown in Fig. 7.8, but uses a different method of controlling the DAU. Here binary stepped voltages are generated for comparison. After each step, it is checked whether the resulting voltage is higher or lower than the

Fig. 7.8 Basic circuit of a measurement by comparison

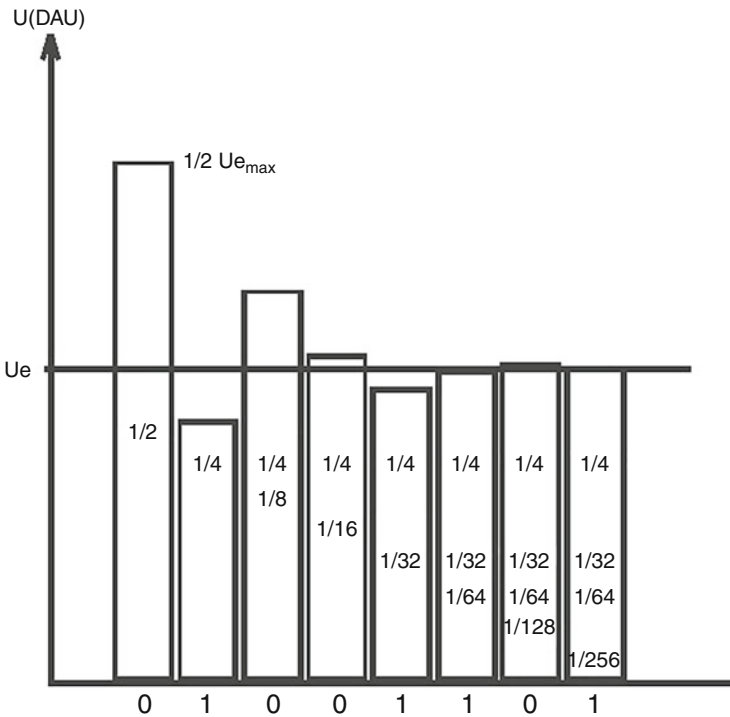
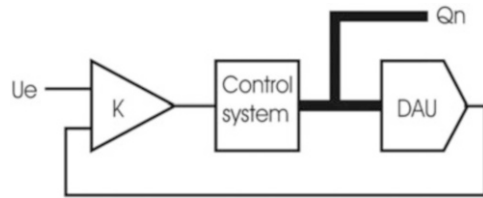
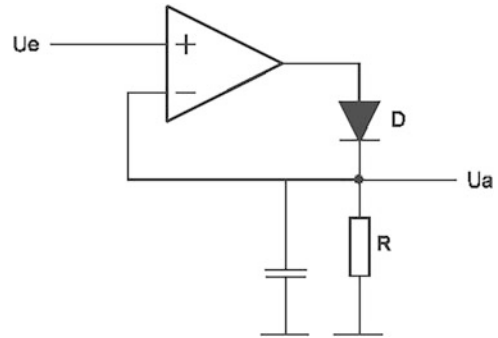


Fig. 7.9 Voltage curve on the comparator during successive approximation

input voltage and it is decided whether the last stage switched on is needed or not. The procedure is shown in Fig. 7.9.

The advantage of the comparison is that in the state of voltage equality of both sides, no current flows and thus no energy is taken from the source. This state corresponds to an internal resistance of the measuring circuit of infinite. The comparison of the two sides is carried out by high-impedance comparators or operational amplifiers—for example with SFET input.

Fig. 7.10 Peak rectifier

7.1.2.6 Measurement of Alternating Voltages

When measuring alternating voltages, the *effective value* and the *peak value* are of particular interest. The peak value can be determined by rectification. The charging capacitor is charged to the peak voltage.

The *peak voltage* can be obtained by a *peak value rectifier*. In the simplest case, this consists of a diode and a downstream charging capacitor. With this simple circuit, the flux voltage of the diode causes a significant error, so that the use of active rectifier circuits is more favourable (Fig. 7.10).

If the measured value is asymmetrical or superimposed by a DC voltage, the negative peak voltage must be measured separately and both must be offset.

In the case of a sine wave, the *effective value* of the voltage is added:

$$U_{\text{eff}} = \frac{U_{\text{top}}}{\sqrt{2}}$$

However, this value is too imprecise if the sine wave is superimposed by a DC voltage component or other disturbances. For other waveforms, different conversion factors apply (see also Eq. (7.3)).

Another way to measure AC voltages is to use *statistical methods*. These have the advantage that they are completely independent of the curve shape. Although they require a high sampling rate of the measured variable, they can be easily mastered with today's electronics. The basic idea behind this measuring arrangement, which works according to the *Monte Carlo method*, is that voltage values are generated by a random generator and compared with the input voltage. By counting the random values that are (with the correct sign) greater than the measured quantity, the effective value of the measurement signal can then be determined by simple mathematical operations.

7.2 Amperage

7.2.1 Definition

If the charge dQ moves through a cross-sectional area in the time span dt , then the current I is calculated:

$$I = dQ/dt.$$

From this equation, it follows that the charge can also be calculated by the duration of the current flow, as described in Sect. 7.1:

$$Q = \int_{t_1}^{t_2} I(t) dt. \quad (7.4)$$

This formula can be used to calculate the charge for time-dependent currents (Sect. 7.3). The charge can be clearly understood as the area under the $I(t)$ curve. At constant current I , the well-known formula $Q = I t$ (Sect. 7.1) is obtained.

The unit of measurement for the current I is the ampere (A). The current I has the value of 1 A if the current I flowing through two straight parallel conductors (with zero diameter) at a distance of 1 m from each other produces a force of $2 \cdot 10^{-7}$ N per metre of length. This definition was chosen in order to be able to measure the electrical and mechanical energy in the same units. It then applies:

$$1 \text{ V A} = 1 \text{ J} = 1 \text{ N m}.$$

7.2.1.1 Measuring Principles

The transport of electrical charge cannot be measured itself. This is always based on the effect of the current flow. The different measuring principles can also be derived from this:

7.2.1.2 Current Flow Through a Resistor

For many conductors, there is a linear relationship between the current I and the voltage U . The proportionality constant is the resistance R , so that *Ohm's law* applies:

$$U = R I \text{ or, after current } I \text{ resolved : } I = U/R.$$

According to this law, a current flow generates a voltage drop U at a resistor R that has been passed through. Thus, any current measurement can be traced back to a voltage measurement—and vice versa, as described at the previous chapter. The accuracy of the

measurement is thus largely dependent on the properties of the resistor used. Apart from the *accuracy of the value*, this refers above all to its *thermal behaviour*, since a (loss) power is always converted in the measuring resistor. This leads to the heating of the resistor and thus to a change in value.

When measuring large currents, *current dividers* are used, which split the flowing current into two paths. By designing the resistance values in the measuring branch and shunt, the ratio of the partial currents can be designed. In addition, special designs of measuring resistors in 4-wire technology have been developed for this purpose, which enable the interference to be eliminated by connection points.

7.2.1.3 Thermal Effect of Current Flow

Conductors through which current flows heat up, thereby changing their length (or volume) and often also other temperature-dependent variables, such as electrical resistance or colour. The power converted at a measuring resistor causes it to heat up. This circumstance can also be evaluated as a measurement signal.

7.2.1.4 Magnetic Field Due to Current Flow

Current-carrying, straight conductors are surrounded by a cylindrically symmetric magnetic field. The path on the closed field line with radius r is $s = 2\pi r$, so that the following applies: $I = H \cdot 2\pi r$.

This relationship can be used for current measurement. The measurements via the magnetic field are all galvanically separated from the signal to be measured and can therefore also be used for *high* voltages and currents. Two basic methods are known:

- The magnetic field is detected by *magnetic sensors* (Hall sensors, Sect. 2.8). This method is often offered as an integrated circuit. Since the current-carrying conductor can be attached directly to the sensor, it is also suitable for currents in the mA range.
- The current transformer is a *current-fed transformer*, which translates the current according to the winding ratio. For this purpose, the output of the transformer is operated with a defined load, over which a voltage proportional to the input current drops. The method is only suitable for alternating currents.
- A current-carrying conductor element of length dl exerts a force dF and the following applies: $dF = I (dl \times B)$. For the magnetic induction, $B = \mu \cdot H$ (μ : permeability). The force effect associated with the magnetic field is compared mechanically with a counterforce, such as in a moving coil instrument.

7.2.2 Measuring Arrangements

The most common way to measure current is by converting it into a voltage drop across a resistor. (Ohm's law: $I = U/R$.) Since this resistor must always be greater than zero to generate this voltage drop, it acts as an additional internal resistor in series with the source.

Fig. 7.11 Arrangement for measuring the current

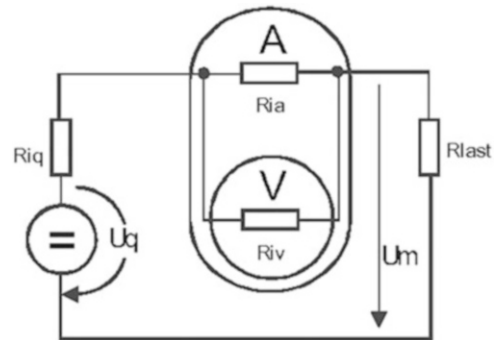


Figure 7.11 shows the basic structure of a current measurement. When a voltage drop is generated at a measuring resistor, the current measurement is performed indirectly by a voltage measurement. If one assumes with today's technology that the internal resistance of the voltage measurement (R_{iv}) is several powers above the internal resistance of the current measurement (R_{ia}), the influence of the voltage measurement can be neglected. The error in current measurement is caused by the increase of the load resistance by the internal resistance of the current measurement, which results in a lower measured current than in the case without measurement. It is thus determined by the ratio of load resistance and internal resistance of the measuring system.

Another source of error is the thermal properties of the measuring resistor R_{ia} . Here the heating of the resistor leads to a change in value. The temperature coefficient T_k describes the temperature dependence of the resistor as the ratio of the relative change of the resistance $\Delta t/R$ and the temperature change Δt : $T_k = (\Delta t/R) \Delta t$. Because of this temperature dependence, components with the lowest possible temperature coefficient (T_k) should be used. The T_k is more important than the precision of the value, since the error resulting from the value tolerance can usually be adjusted in subsequent processing.

With increasing measuring currents, the influence of the contact points of the measuring resistor increases. The resistance value of the solder or terminal point is in the range of some $m\Omega$ and is therefore almost equal in value to the actual measuring resistor, which can also be in the range of some 10 to 100 $m\Omega$. To eliminate this disturbance, special measuring resistors in 4-wire technology are offered, which have connections for the measurement in addition to the connections for the useful circuit. The latter are directly connected to the (internal) precision resistor. Since no current flows through them, their intrinsic resistance has no significance for the measurement. Figures 7.12 and 7.13 show the design and construction of such resistors—here the VPR220 (www.vishay.com).

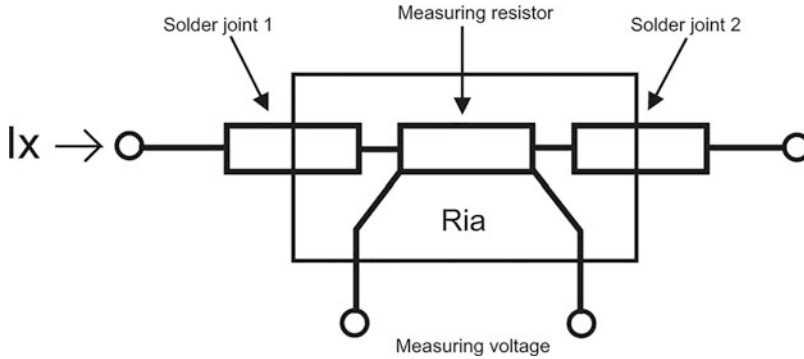
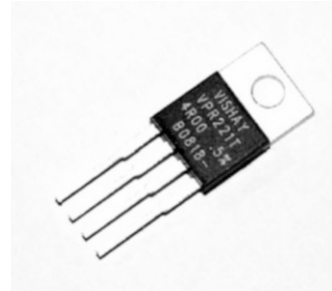


Fig. 7.12 Measurement of large currents in 4-wire technology

Fig. 7.13 Measuring resistor



7.3 Electrical Charge and Capacity

7.3.1 Definition

The relationships between electric current I and charge Q have already been explained in Sect. 7.2. It applies:

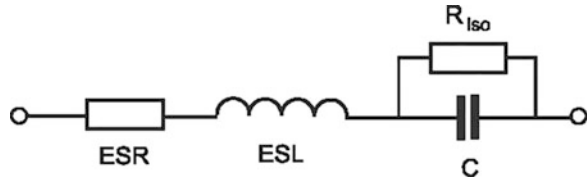
$$Q = \int_{t_1}^{t_2} I(t) dt.$$

With constant current, the following applies: $Q = I \cdot t$.

Clearly, the *electrical charge*—or quantity of electricity—is a measure of the charge carriers transported by the current with strength I .

The unit for the charge Q is the coulomb (C): $1 \text{ [C]} = 1 \text{ [A]} \cdot \text{[s]}$.

Fig. 7.14 Equivalent circuit diagram of a capacitor



The *capacity* C indicates how much charge Q can be stored per unit of voltage 1 V. The following applies: $C = Q/U$. A component that can store the electric charge is called a *capacitor*.

The unit of capacity is the farad (F):

$$1[\text{F}] = 1[\text{C}]/[\text{V}] = 1[\text{A}] [\text{s}]/[\text{V}].$$

If this relationship is rearranged, a capacitor of 3600 F charged to 1 V carries an electricity quantity of 1 Ah—a clear comparison between accumulator and capacitor.

The capacitor as a component for storing charge theoretically allows the infinitely fast exchange of this charge. *Parasitic components*, which are caused by the technical construction, hinder this fast charge exchange. Technically, the capacitor consists of two electrically conductive surfaces, which are as large as possible and which face each other with a minimum distance. The insulation resistance R_{iso} between the layers that occurs in this process causes a discharge and thus leads to a loss of charge. Due to their geometric dimensions, the conductive surfaces have an ohmic resistance ESR (Equivalent Series Resistance) and an inductance ESL (Equivalent Series Inductance L). Both limit the maximum current flowing during charging or discharging, as they form a low-pass filter together with the capacitance. The equivalent circuit diagram of a capacitor is shown in Fig. 7.14.

The ESR is intended to illustrate this. When a voltage is applied across the real capacitor, charging is performed according to the equation

$$U_a = U_e \cdot \left(1 - e^{-\frac{t}{RC}}\right) \quad (7.5)$$

changed over after the loading time

$$t = -RC \cdot \ln \left(1 - \frac{U_a}{U_e}\right) \quad (7.6)$$

with U_a : Charging voltage at the capacitor

U_e : Applied voltage

R : Value of the ESR

C: Capacitance of the capacitor.

If the voltage at the capacitor is to reach 99% of the applied voltage, the time elapses

$$t(99\%) = RC * 4.6$$

With practical values for a low ESR charging capacitor in a switching regulator with 220 μF and an ESR of 100 $\text{m}\Omega$, the time is about 100 μs .

7.3.1.1 Measuring Principle Charge

The following measuring principles are conceivable:

- The electric charge can be measured with a coulombmeter. This determines, based on Faraday's law, the amount of charge that has flowed through the galvanically deposited masses (see also the old definition of current).
- According to Eq. (7.4), the charge can be determined from the capacity C and the voltage U ($C = Q/U$). Based on the changed equation above to $Q = C \cdot U$, a change in capacity (with constant charge) inevitably leads to a change in voltage. If a further capacity of known size is connected in parallel to the charge-carrying capacity, the voltage decreases in proportion to the ratio of the capacities (Fig. 7.15). The unknown capacity can be determined from the change in voltage and the connected capacity, and from this the existing charge. Since the charge is not lost (infinitely high-resistance voltmeter), it must be the same at C_x and $C_x + C_1$.

$$Q = C_x \cdot U_x = (C_x + C_1) \cdot U_1$$

$$C_x = C_1 \cdot \frac{U_x - U_1}{U_x - U_1}$$

$$Q_x = C_1 \cdot \frac{U_x \cdot U_1}{U_x - U_1}$$

with U_x : Voltage before closing S

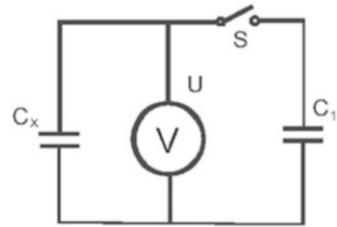
U_1 : Voltage after closing of S

C_1 : Known capacity.

This measurement requires very high-impedance measurement inputs in order not to lose any charge above the input resistance. This is possible by using *electrometer tubes* or *semiconductor circuits* with SFET input.

- The determination of large amounts of charge—e.g. of accumulators—is done by *circuits*, which determine the amount of charge moved by means of current and time measurement. By integrating these values with the correct sign, the state of charge of storage capacitors or accumulators can be calculated.

Fig. 7.15 Charge determination by charge division



7.3.1.2 Measuring Principle Capacity

For the measurement of capacitances, a variety of methods have been developed, the application of which depends on the size of the capacitance, the technical design and the accuracy.

- *Capacities can be measured* by the relationship $C = Q/U$. An uncharged capacity of known size is connected in parallel with a charged capacity of unknown size. The unknown capacity can be determined from the voltage change that occurs (Fig. 7.15). It is possible to take advantage of the fact that the definition of capacity includes time.

$$C = \frac{Q}{U} = \frac{I \cdot t}{U} \quad \text{or} \quad U = \frac{I \cdot t}{C}.$$

If a capacity is charged with a constant current to a defined voltage, the time required for this is proportional to the capacity. The same effect can be achieved by setting the charging time and then measuring the resulting voltage (Fig. 7.16).

- This method described above can also be performed with constant voltage. Here the capacity is charged to the applied voltage via a known resistor. Since the current through the resistor decreases as the capacitor voltage increases, the process follows an e-function. The time required for charging to 63.2% of the reference voltage is determined, since the logarithm in eq. (7.6) has the value -1 in this case. From eq. (7.6) (Fig. 7.17)

$$t = -RC \cdot \ln \left(1 - \frac{U_a}{U_e} \right)$$

in this particular case

$$t = -RC \cdot \ln \left(1 - \frac{0,632 \cdot U_e}{U_e} \right) = RC \quad \text{and} \quad C = \frac{t}{R}.$$

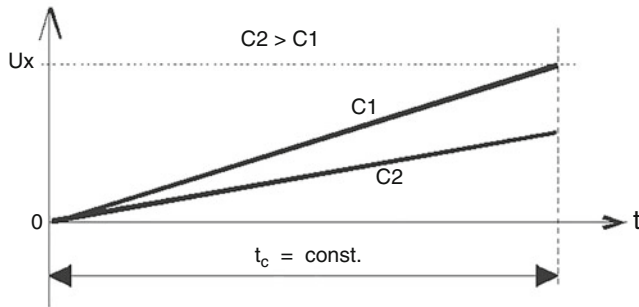


Fig. 7.16 Capacity determination by charging with constant current

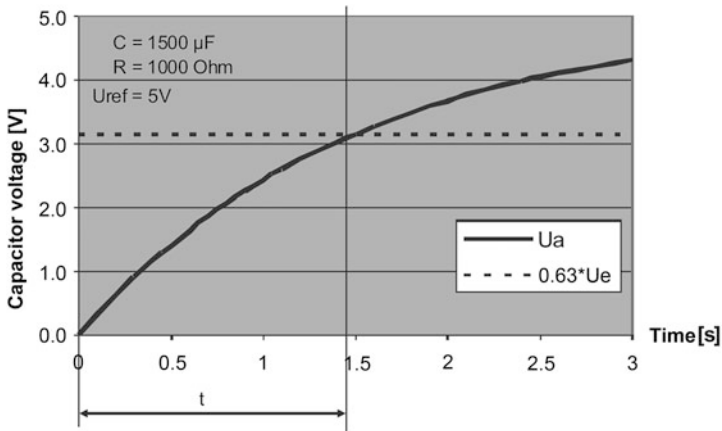


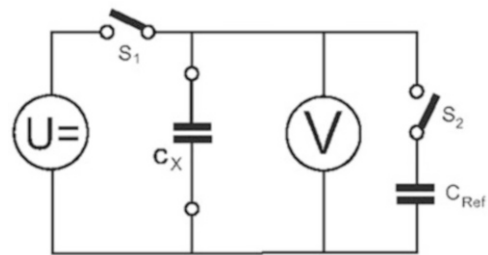
Fig. 7.17 Determination of capacity by charging with constant voltage (Eq. (7.5))

- The capacitance can be determined by use in oscillating LC or RC circuits. The *frequency* f , which is proportional to the capacitance, is then available as the evaluable variable (Fig. 7.19).
- The alternating voltage behaviour of a capacitance can be used to determine it. A capacitor has an AC resistance—the *reactance*, which depends reciprocally on the frequency and the capacitance.

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f \cdot C}$$

In contrast to the previous point, we are not working on the resonance frequency here. The frequency used for measurement should be selected according to the desired measurement range.

Fig. 7.18 Capacity measurement by charge sharing



7.3.2 Measuring Arrangements

In the following section, some examples of practical measurement of capacity and charge are given.

7.3.2.1 Capacity Measurement Through Charge Sharing

The parallel connection of known capacitors described above can be used as a discrete experimental setup, in the form of electronic circuits or as an integrated circuit. Figure 7.18 shows the basic principle.

Starting from the equation $Q = C U$, the following equations can be drawn up neglecting the losses due to the voltage measurement for the charging phase (S_1 closed) and the measuring phase (S_2 closed):

$$\text{Phase 1 :} \quad Q = C_x \cdot U_1$$

$$\text{Phase 2 :} \quad Q = (C_x + C_{\text{ref}}) \cdot U_2$$

$$\text{after conversion you get} \quad C_x = (C_{\text{ref}} \cdot U_2) / (U_1 - U_2).$$

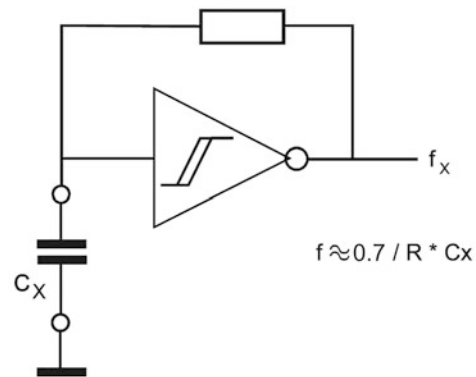
7.3.2.2 Capacity Measurement by RC Generators

A very simple method is the installation of the unknown capacitance in an oscillating circuit. The range of capacitances that can be determined by the circuit in Fig. 7.19 can extend over 7 decades. In this circuit, the *hysteresis* of the logic gate—an inverter—is used. When starting with an uncharged capacitor, the output voltage is equal to the operating voltage and the capacitor is charged. When the threshold voltage is exceeded, the output switches to 0 V and discharges the capacitor until its voltage falls below the threshold. This process is repeated cyclically.

7.3.2.3 Complex Components for Capacitance Determination

- For the measurement of very small capacitances, two examples are of particular interest—the AD774x from Analog Devices and the PS02 from ACAM.
- The AD774x is a direct *capacity digital converter* (Fig. 7.20). It takes advantage of the fact that when charging a capacitor, only a quantity of charge determined by the

Fig. 7.19 RC generator for determining the capacitance



capacitance and voltage can be moved. This amount of charge is collected on a larger capacitor at the input of an analog-to-digital converter, which converts the voltage formed there. With this principle, it is possible to determine extremely small capacitance values in the range of a few pF and a resolution of 4 aF to 10 aF (1 “atto” Farad aF corresponds to 10^{-18} F).

- With PS02, the determination of the charge and discharge time of an RC combination is carried out. It is possible to use both R and C as variables. The time measurement works in ranges smaller than 1 ps and thus also allows capacitance measurements in the pF range (Fig. 7.21).
- Charging and discharging functions are generated here as shown in Fig. 7.21. For this purpose, the capacities are charged to operating voltage by closing all switches. Discharging is performed individually by S2 and S3, whereby the time required is measured. S2 and S3 are used alternately and serve to compensate for temperature drift. The discharge times are in the range of 2–100 μ s and are measured with a resolution of 15 ps (ps: pico seconds; pico = 10^{-12}).

7.4 Electrical Conductivity and Specific Electrical Resistance

7.4.1 Definition

The *specific electrical conductivity* κ (Kappa) describes the property of a substance to conduct electricity. This property is determined by the number of freely moving electrons in a substance. Since this is strongly dependent on temperature, it is referred to a temperature of 25 °C. The following relationship exists with the resistance R of length l and the cross section A :

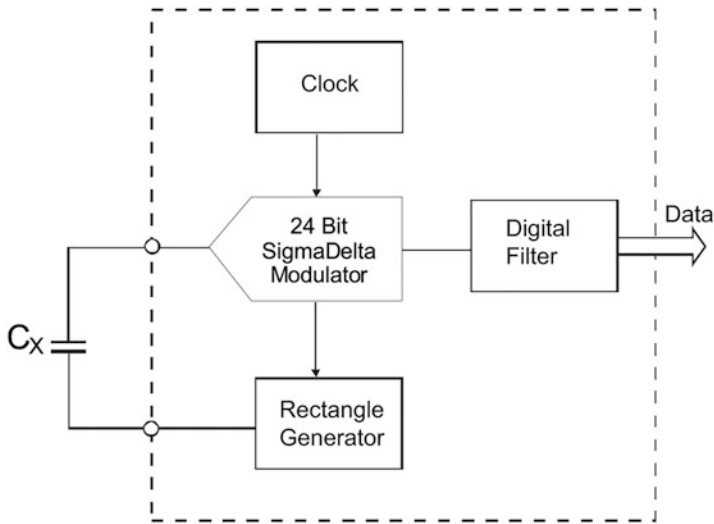


Fig. 7.20 Simplified block diagram of the AD7745 (Source: Analog Devices Inc.; www.analog.com)

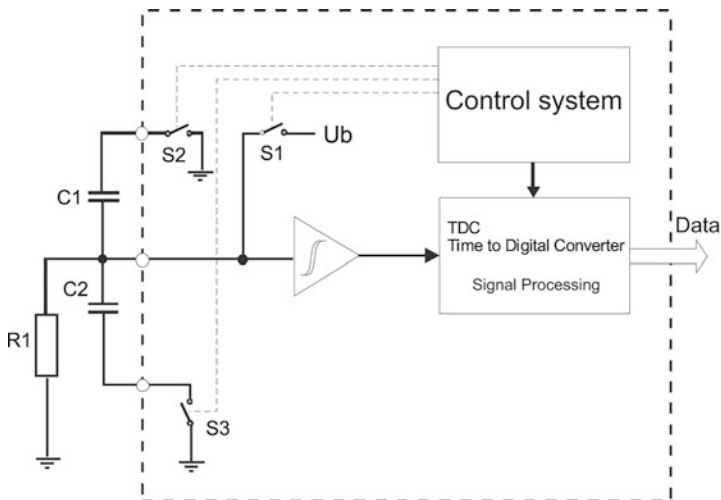


Fig. 7.21 Simplified block diagram of the PS02 (Source: ACAM-Messelektronik GmbH; <http://www.acam.de/>)

$$\kappa = l/(A \cdot R).$$

The unit of measurement of electrical conductivity κ is the Siemens per meter, which corresponds to the reciprocal value of resistance and is expressed by the SI unit

$$1 \text{ S/m} = 1/\Omega = 1 \text{ s}^3 \text{ A}^2 \text{ m}^{-3} \text{ kg}^{-1}$$

It corresponds to the *reciprocal value* of the specific electrical resistance ρ ($\rho = 1/\kappa$).

According to Ohm's law ($R = U/I$), the *electrical resistance* R describes the required voltage U , which is necessary to transport a certain current I through a conductor. The resistance value should ideally be independent of current, voltage and frequency. It is always dependent on temperature and is practically influenced by other parasitic effects. The SI unit is

$$1 \text{ } \Omega = 1 \text{ m}^2 \text{ kg s}^{-3} \text{ A}^{-2}.$$

The reciprocal value of the electrical resistance R is the electrical conductance G with the unit Siemens [S]. The following applies: $G = 1/R$.

The electrical conductivity κ or the specific electrical resistance ρ varies over 25 powers of ten. The materials are divided into the following three categories: *conductors* ($\rho =$ from 10^{-10} to 10^{-5} Ωm), *semiconductors* ($\rho =$ from 10^{-5} to 10^7 Ωm) and *insulators* ($\rho > 10^7$ Ωm).

7.4.1.1 Measuring Principles for Resistance

Resistance is generally measured by measuring the *current* and *voltage* of a target. Another way is to determine the value by compensation in a *bridge circuit*. This method is mainly used for *very small resistances*.

When measuring the AC resistance, the parasitic properties of the components come into play. The resistance is described as a *complex quantity* $X_R = R + jX$. The real part R corresponds to the *active resistance* and the imaginary part jX to the *reactance*. These values are measured in a four-pole arrangement using an AC voltage source and vector voltmeter.

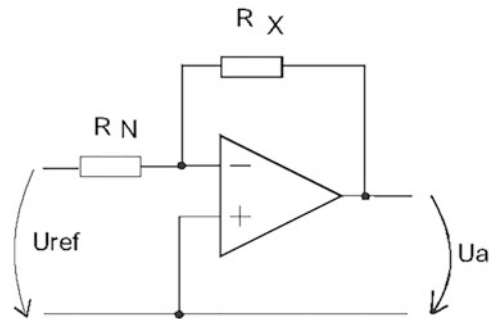
Conductivity is measured when determining the properties of solutions (e.g. of water). However, special measures must be taken to ensure that galvanic or chemical effects can be excluded from the measurement.

7.4.2 Measuring Arrangements

7.4.2.1 Determination of Current and Voltage at the Measured Object

In the simplest case, a voltage is applied and the current is measured. Since the variable quantity here is the current, the measuring function is a $1/x$ characteristic ($I = U/R$), which leads to a *non-linear* output signal. With this arrangement, the *internal resistances* of the measuring instruments must also be taken into account, as these lead to a distortion of the current and/or voltage values (see current/voltage correct measurement).

Fig. 7.22 Resistance measurement using OPV



To obtain an output signal proportional to the resistance, the target can be supplied with a constant current. This can be achieved by a current source or with an arrangement as shown in Fig. 7.22.

The disadvantage of the circuit in Fig. 7.22 is that the characteristics of the reference voltage are still included in the measurement. If one wants to switch this off, this can be done with a *ratimetric measurement* (Fig. 7.23).

The output value of the AD converter is determined by the equation

$$N = U_e \cdot N_{\max} / U_{\text{ref}},$$

where: N: Conversion result in digits

U_e : Input voltage of the converter (here difference between U_{in-H} and U_{in-L})

N_{\max} : Maximum number of conversion steps = resolution of the ADC

U_{ref} : Reference voltage that corresponds to the maximum resolution.

Since the same current flows through the measuring resistor and the reference resistor in the arrangement, the above equation can be changed to the resistors. This results in:

$$N = I \cdot R_x \cdot N_{\max} / I \cdot R_n = R_x \cdot N_{\max} / R_n.$$

This means that the resistance measurement is no longer dependent on the quality of the supply.

In order to keep the errors in the measurement of small or precise resistances small, the *4-wire measurement* is used in these cases.

The supply lines to the test object are disconnected from the test leads. This ensures that the resistance of the supply line and the connecting contacts is not included in the measurement (Fig. 7.24). No current flows through R_{Z3} and R_{Z4} on the test leads themselves, so that no voltage drop can occur on them.

For the same reason, *precision measuring resistors* are offered as 4-pole components. In addition to the signal path, they also have separate connections for the measurement. (Sect. 7.2).

Fig. 7.23 Ratiometric resistance measurement

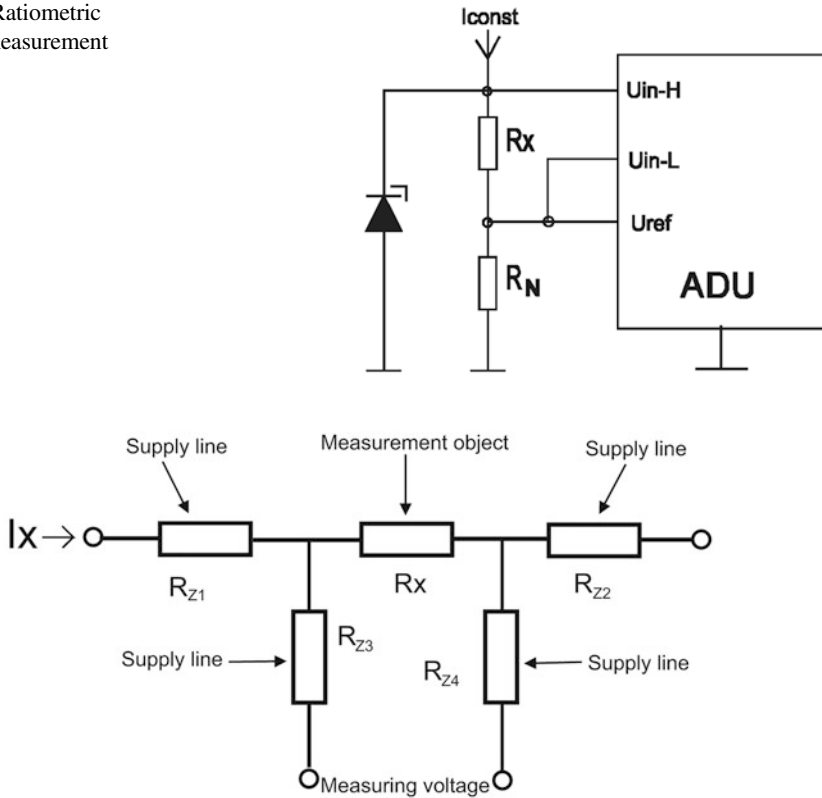


Fig. 7.24 4-wire measurement for determining small resistances

A special form of the 4-wire measurement is the *4-peak measurement*. This is mainly used in semiconductor production to determine the resistance of *surfaces*. For this purpose, 4 measuring tips are placed at equal distances on the resistance layer. A measuring current is impressed via the two outer tips and the voltage across the resistance to be measured is determined at the two inner tips.

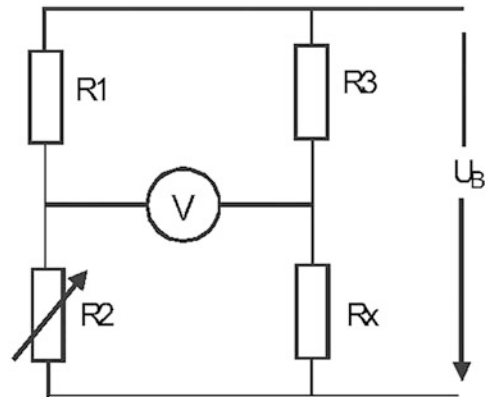
7.4.2.2 Resistance Measurement Through Compensation

Resistance measurement by compensation has the advantage that in this bridge circuit, the *cross current* is *zero* in the compensated case. This means that no error is caused by the instrument.

The following applies to the adjustment:

$$R_2 / (R_1 + R_2) = R_x / (R_x + R_3).$$

Fig. 7.25 Resistance measurement by compensation in a bridge



For R_x :

$$R_x = R_2 \cdot (R_3/R_1).$$

Since this method compares two resistance ratios, the absolute values on both sides do not have to be the same. Thus, the arrangement also allows the measurement of *very low* (and *very high*) values for R_x , which is used for example in milliohmmeters. A simple arrangement is shown in Fig. 7.25.

7.4.2.3 AC Voltage Resistance Measurement

The resistor itself has no frequency dependence. This results from its *constructional structure*, i.e. an *inductance* by the stretched or wound arrangement of the resistor track or a *capacitance* by adjacent tracks. These parasitic values move in areas that only become important in high frequency applications. They are generally not specified in the data sheets.

Figure 7.26 shows the basic structure of a *four-pole measuring arrangement*, with resistors R_1 and R_2 also being used for the RF adaptation of the signal paths. The vector voltmeter measures the magnitudes of the voltage for both inputs and the phase between them. The measurement is performed over a *frequency spectrum*, since the effect of the parasitic components changes with frequency.

7.5 Electric Field Strength

7.5.1 Definition

The *electric field strength* E is a *vectorial field quantity* that, according to its magnitude and direction in a field, indicates the force F acting on the charge Q . It is identified by the formula symbol E and the following applies: $E = F/Q$. It can also be represented by the differential change of the potential dU over a distance ds , resulting in

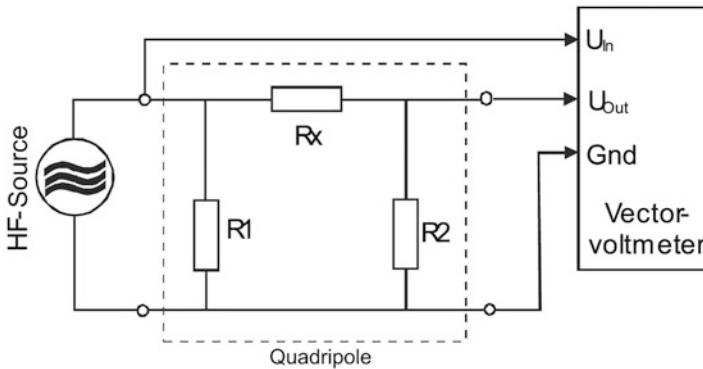


Fig. 7.26 Measurement of the AC voltage resistance using a four-pole measuring arrangement

$$E = -dU/ds.$$

The SI unit is $1 \text{ V/m} = 1 \text{ m} \cdot \text{kg} \cdot \text{s}^{-3} \cdot \text{A}^{-1}$.

There is a close relationship between electric and magnetic field, which is described by the *Lorentz transformation*.

7.5.2 Measuring Principles for the Electric Field Strength

The measurement of the electric field strength E can basically be performed by mechanically measuring the attractive forces between two differently charged electrodes. However, this has no significance as an industrial measuring method.

However, the force in the electric field also acts on *freely moving charge carriers*, which move in a conductor or in a vacuum. This influence is used and made visible, for example, in the *electron beam tube* (oscillograph tube, magic eye). The deflection of the electron beam here is proportional to the field strength between the deflection electrodes, which are constructed in the form of a plate capacitor. The field lines are perpendicular to the direction of movement of the electrons.

Cylinders (capacitors) partially pushed into each other result in field lines, which are almost parallel to the electron beam. They thus have a scattering or concentrating effect on the electron beam and thus form an electrostatic optics for *beam focusing* due to the field lines. Figure 7.27 shows the operating scheme of the deflection of an electron beam by an electric field.

In practical use, the “optics” consists of several electrostatic lenses, as can be seen in Fig. 7.28 on the five cylinders.

The course of field lines can be displayed in the *electrolytic trough*. A vessel is filled with (tap) water and the electrode assembly to be measured is immersed. The area in the trough is scanned by a measuring electrode and the voltage potentials are determined. The

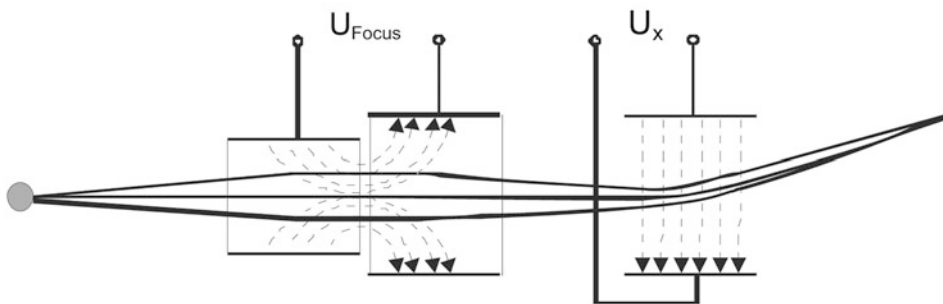


Fig. 7.27 Principle of electrostatic deflection

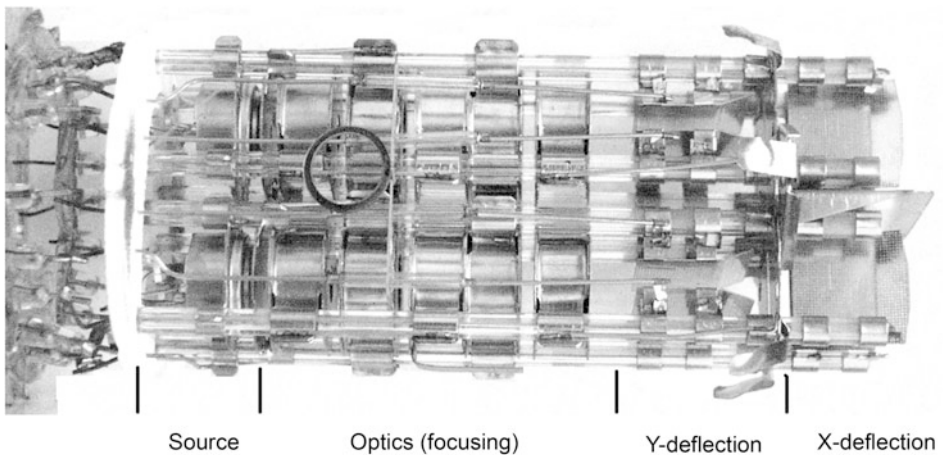


Fig. 7.28 Electron beam system of an oscillograph tube (double system B7S22)

scanning is carried out at a very high resistance or by a compensation method. The image of the field lines can easily be derived from the measured *equipotential lines*, since the field lines are perpendicular to them.

In the field of electromagnetic compatibility (EMC, Sect. 17.2), electrostatic fields are generated and measured. Special electrodes are used for this purpose, which generate the fields or absorb potentials.

7.6 Electrical Energy and Power

7.6.1 Definitions

The *power* P generally describes the amount of work dA performed in a unit of time dt . The unit of measurement is Watt [W].

Power P occurs in different physical contexts. These units can be converted into each other:

$$1 \text{ Watt} = 1 \text{ J/s} = 1 \text{ m}^2 \cdot \text{kg} \cdot \text{s}^{-3}$$

as well as

$$1 \text{ Watt} = 1 \text{ V} \cdot \text{A}.$$

The *energy* or *work* is the power generated or extracted in a certain time. Its unit is the Joule [J]. It applies:

$$1 \text{ Joule} = 1 \text{ W} \cdot \text{s}.$$

7.6.1.1 Practical Case Accumulator

For batteries, the energy content is often given in Ah. This unit describes the *charge storage capacity* of a battery. The power that can be drawn, however, can only be calculated approximately by multiplying the value by the average terminal voltage (Sect. 7.3).

7.6.2 Forms of Power

7.6.2.1 Power in DC Circuit

In a DC circuit with static conditions, the converted power P is the product of the measured current I and voltage U :

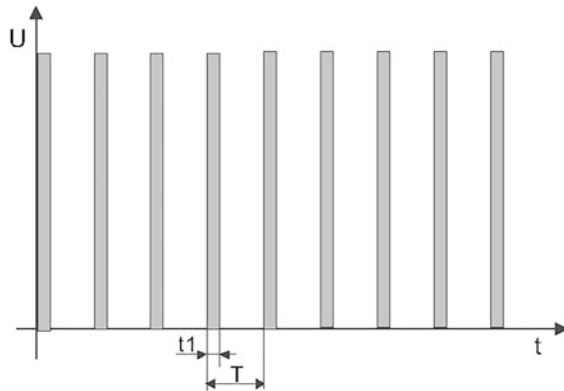
$$P = U \cdot I.$$

If current and voltage are not continuously present over time (e.g. in switched systems), an integral power over time results, which, with constant U - I quantities, gives a temporal mean value of the power. This circumstance is used, for example, when dimming incandescent lamps (direct and alternating current) by means of pulse width modulation (PWM). Figure 7.29 shows the basic principle. With a continuously applied voltage, the power at the consumer is.

$$P = U \cdot I = U \cdot (U/R) = U^2/R.$$

In the switched system, the duration of the energy conversion as a relative value must be included in this equation. The result is:

Fig. 7.29 Pulse width modulation



$$UP = \frac{U^2}{R} \cdot \frac{t_1}{T}$$

with t_1 : Switch-on time

T : Period duration

7.6.2.2 Power in AC Circuit

The DC circuit assumes constant values for current and voltage. This is not the case in the alternating current circuit.

If the current flows through an ohmic arrangement, the converted power corresponds to the area below the curve (grey area A in Fig. 7.30). As a resistor does not react to the direction of the current flow, the amount of area under the half-waves must be used, which is represented by the left-hand area A and area B. The equal distribution of the area B over the period gives the resulting area C, which corresponds to the active power (see also Sect. 7.1).

For sinusoidal signals, the resulting *real power* is:

$$P_w = U_{\text{eff}} \cdot I_{\text{eff}} = 0,5 \cdot U_{\text{max}} \cdot I_{\text{max}},$$

where U_{max} or I_{max} corresponds to the peak value of current and voltage.

If there are capacitors and inductors in the circuit, there is a *phase shift* between current and voltage. This means that the maximum current does not occur at the time of the maximum voltage (I_{ind} in Fig. 7.31). This means that only a power is produced at the load that corresponds to the product of the current and voltage values at time t . The phase shift can be between 0° and 90° .

The result is a part *real power* and a part *reactive power*, which cannot be converted at the consumer. The vectorial superposition of both quantities produces the *apparent power*.

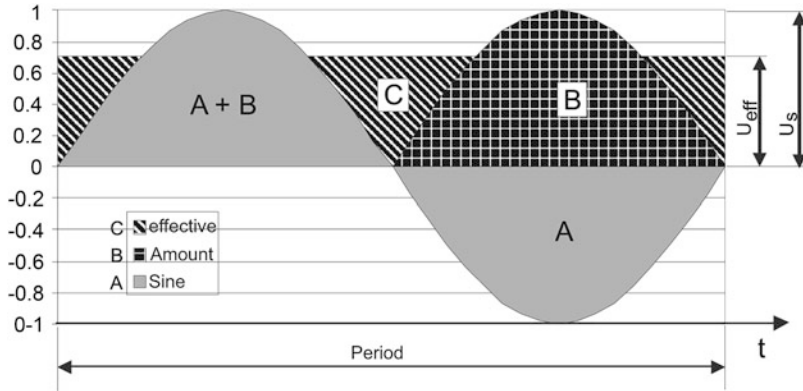


Fig. 7.30 Power at an ohmic resistor

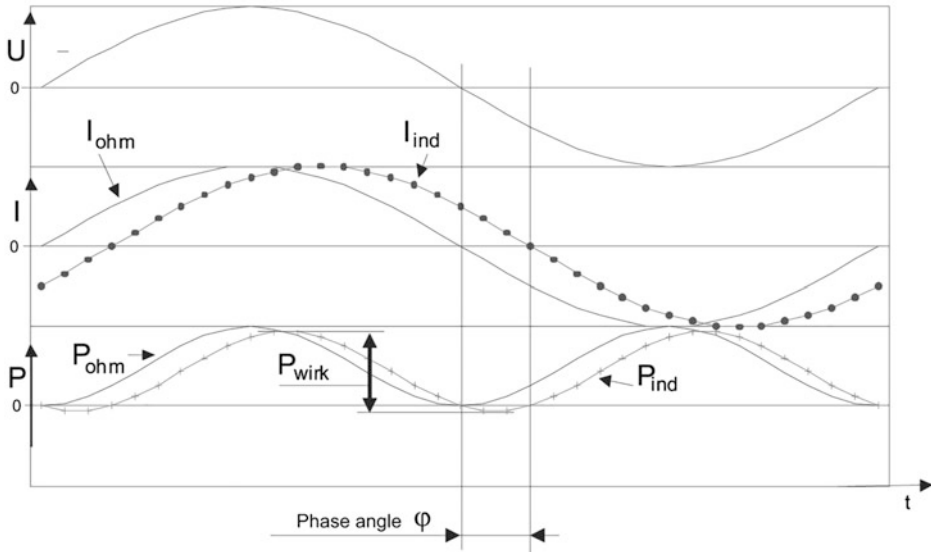


Fig. 7.31 Active power with phase shift

The proportion of active power to apparent power is described by the *power factor* $\cos \varphi$. This results in the following relationships:

$$\begin{aligned} \text{Active power} \quad P_w &= U_{\text{eff}} \cdot I_{\text{eff}} \cdot \cos \varphi, \\ \text{Reactive power} \quad P_b &= U_{\text{eff}} \cdot I_{\text{eff}} \cdot \sin \varphi \quad \text{und} \\ \text{Apparent power} \quad P_s &= U_{\text{eff}} \cdot I_{\text{eff}} = \sqrt{P_w^2 + P_b^2}. \end{aligned}$$

The mathematical and technical effort increases considerably if the signals to be measured are not sinusoidal or discontinuous due to superimpositions.

7.6.3 Measuring Principles

Due to the physical definition of power, a multiplication of current and voltage including the consideration of phase positions and time occurrence must always be carried out. As a result, almost all methods use this basic principle.

7.6.3.1 Electromechanical

The oldest method of direct power measurement is the *cross-coil instrument*. It uses the multiplying effect of superimposing magnetic fields.

In a mechanical moving coil instrument, normally the force effect between the magnetic field in a coil is generated by the measured variable and a constant permanent magnet. In the case of the cross-coil instrument, this permanent magnet is replaced by a coil through which the measuring current flows and the measuring voltage is applied to the moving coil (Fig. 7.32). The deflection of the instrument is then proportional to the product of current and voltage. The averaging is due to the mechanical inertia of the system.

Another principle of mechanical power measurement is the *energy meter*, colloquially known as the *electricity meter*. It is also called *Ferraris meter* after its inventor. In this meter, the current and voltage measuring coil and its fields act on a rotor—an aluminium disc—and generate a torque there (comparable to a synchronous motor). An eddy current brake ensures that the speed of this rotor is proportional to the power output. The revolutions of the rotor are counted, so that in the result the sum of the power consumed over time—i.e. the energy—can be read.

These mechanical meters are increasingly being replaced by electronic systems.

7.6.3.2 With Analogue Electronics

The most obvious way is to “replicate” the equation, i.e. *multiply the measured current and voltage values* and then *average them*. This can be done by *multiplication circuits* or, in the case of AC voltage, by *four-quadrant multipliers*. Depending on the dynamics of the measurement signals, this can be done directly or by logarithmizing the input signals, adding them up and then delogarithmizing them. In both cases, a downstream averaging is performed.

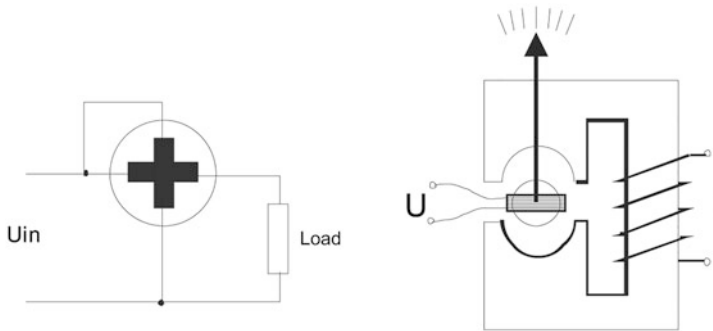


Fig. 7.32 Principle of the cross-wound instrument

The advantage is the *fast* analog signal processing. However, the dynamics of the output signal limits the possible input dynamics of the measured variables or reduces the achievable resolution of the measurement.

Another almost analogous way is the fact that the “area under the current–voltage curve” corresponds to the power. If this area corresponds in amplitude to the applied voltage and is modulated by a *pulse width modulation (PWM) signal*, which is proportional to the current, then “partial areas” in the size $U \cdot I$ are created in time sequence. The integration (averaging) of these areas then corresponds to the power (Fig. 7.33).

This circuit is much simpler, but is only suitable for signals with limited dynamic range, as this is limited by PWM formation. It is well suited for measurements in the 50-Hz range.

7.6.3.3 With Digital Electronics

The digital solutions equally implement the multiplication of current and voltage. For this purpose, current and voltage are measured with a sufficiently *high sampling rate* using an AD converter. Multiplication and filtering is then performed in the software. The measured values can also be used to determine other values, such as phase positions, real power, peak values or superimposed DC components, through targeted processing.

An important point in the digital acquisition of the input variables is the simultaneous acquisition. If a microcontroller with multiplexed inputs is used, one loses the possibility of phase measurement and can therefore no longer determine real power, for example. For this reason, there are special controllers with *DAS structure* (Data Acquisition System), which have several AD converters operated in parallel. This is the only way to ensure the physical requirement of simultaneous measurement/sampling.

With the “E” option, the MSP430FE4xxx processor family contains special processors designed for single- and three-phase energy measurement systems. The ESP430 unit is implemented in these processors, which processes these tasks at hardware level by means of multiple AD converters and multipliers with downstream signal processing. Figure 7.34 shows the block diagram of such a processing unit.

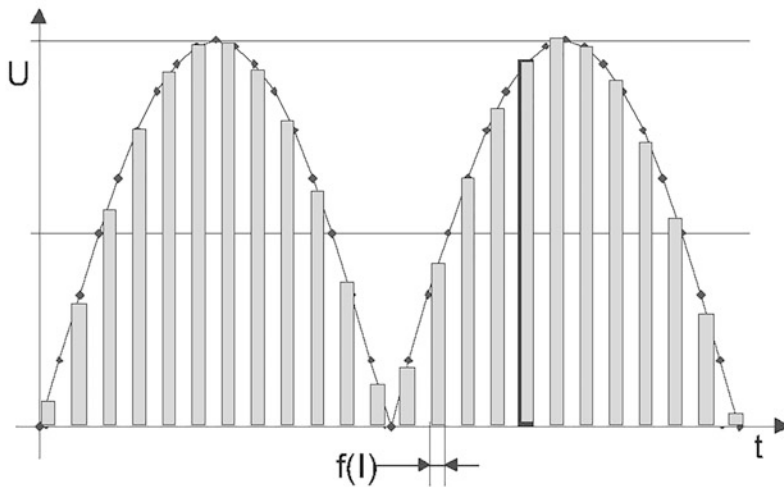


Fig. 7.33 Power measurement by current–voltage modulation

7.6.3.4 With Statistical Methods

The prerequisite for an exact determination of power in an AC circuit is the determination of the *effective value* of current and voltage. This is possible, for example, with the measurement described in Sect. 7.5 using the *Monte Carlo method*, which works independently of waveform and superpositions.

7.7 Inductance

7.7.1 Definition

The *inductance* L describes the relationship between the number of turns and the magnetic resistance of an arrangement. The unit of measurement is the Henry (H). The inductance is a derived SI unit. It is determined by:

$$L = \frac{w^2}{R_m} = \frac{w^2 \cdot \mu \cdot F}{l} \quad [L] = 1 \frac{\text{Wb}}{\text{A}(w)} = 1 \frac{\text{Vs}}{\text{A}(w)} = 1\text{H}$$

with w : Number of turns

l : Length of the magnetic conductor = path of the magnetic flux

F : Cross section of the magnetic conductor

μ : Permeability ($\mu = \mu_0 \cdot \mu_r$)

μ_0 : Magnetic field constant ($\mu_0 = 1.256 \cdot 10^{-8} \text{ V s A}^{-1} \text{ cm}^{-1}$)

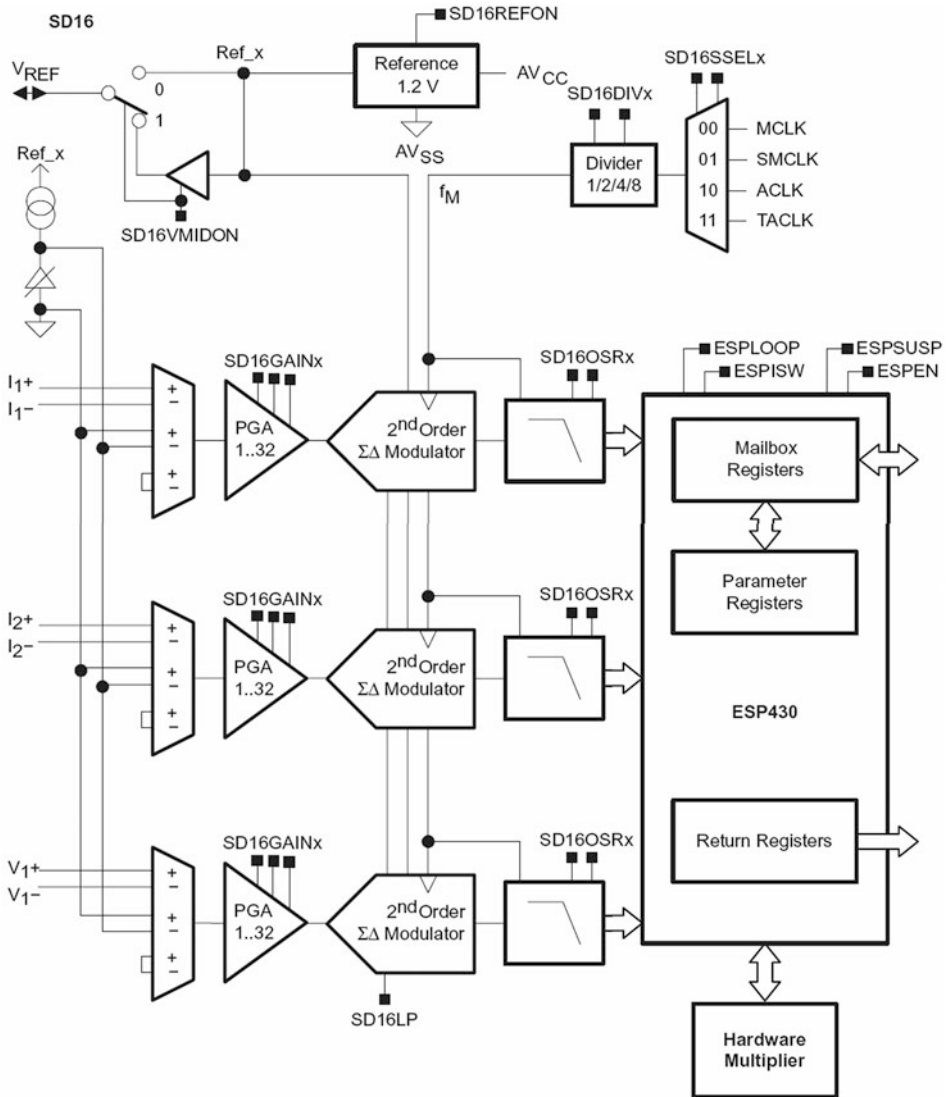
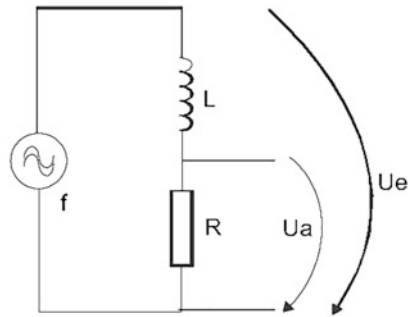


Fig. 7.34 ESP430 unit of an MSP430FE4xxx (Source: <http://www.ti.com/>)

μ_r : Permeability number (material constant, except for ferromagnetic materials usually 1).

In an alternating current circuit, the inductance acts like a resistor, since the change in voltage in the coil generates a counter-voltage caused by the magnetic field. The current flow can therefore only reach its maximum when the counter-induction has decayed. For this reason, the voltage precedes the current, which corresponds to a *positive phase angle*. The inductive resistance X_L is frequency dependent and is determined by:

Fig. 7.35 Voltage divider with inductance



$$X_L = \omega L = 2\pi f \cdot L$$

7.7.2 Measuring Principles

The inductance is measured by a *voltage divider*. The divider ratio and thus the inductive resistance can be determined from the ratio of input and output AC voltage. The inductance can then be inferred from the measuring frequency used (Fig. 7.35).

$$L = \frac{R \cdot U_a}{(1 - U_a) \cdot 2\pi f}$$

This path is usually taken with *multimeters where* no high demands are made on measurement accuracy.

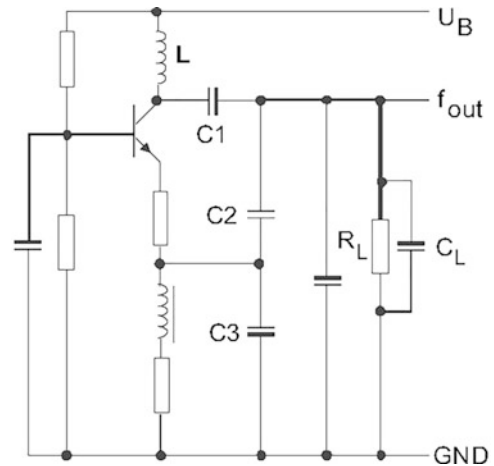
The disadvantage of this method is that on the one hand the resistance is very low at very small inductances and on the other hand the ohmic resistance (resistance of the winding), which always adheres to the component, is included in this measurement as an error.

The alternative is to build a *resonant circuit* consisting of inductance and capacitance (Fig. 7.36). The resulting frequency is only determined by the two components. The ohmic components of the arrangement influence the *quality* of the resonant circuit, but not its frequency. Thus, the inductance can be determined with good accuracy over a *wide range*. The resonant frequency of the arrangement is determined by:

$$f = \frac{1}{2\pi\sqrt{LC}} \quad \text{which results in} \quad L = \frac{1}{(2\pi f)^2 \cdot C}$$

The circuit in Fig. 7.36 shows a Colpitts oscillator, which operates with an undivided coil. The resonant frequency is determined by the components L and C1 to C3.

Fig. 7.36 LC oscillator according to Colpitts



7.8 Magnetic Field Strength

7.8.1 Definition

The *magnetic field strength* H is defined as the product of the current strength I and the number of turns N divided by the coil length l . It is valid:

$$H = \frac{I \cdot N}{l}. \text{ The unit is Henry (H) : } [H] \text{ } 1 \frac{\text{A}}{\text{m}}.$$

Here, too, there is a correlation between the magnetic field strength H and the *magnetic flux density* B . The flux density is the product of the field strength H and the permeability μ :

$$B = \mu \cdot H.$$

The unit for the flux density is the Tesla T. One Tesla is equal to the area density of the homogeneous magnetic flux of 1 Wb, which vertically penetrates the area of 1 m².

$$[B] = 1\text{T} = 1 \frac{\text{Vs}}{\text{m}^2} = 1 \frac{\text{Ws}}{\text{Am}^2} = 1 \frac{\text{Wb}}{\text{m}^2} = 1 \frac{\text{kg}}{\text{A} \cdot \text{s}^2}$$

7.8.2 Measuring Principles of Magnetic Quantities

For the determination of magnetic quantities, there is a large spectrum of sensors, which are specialized for different applications. The best known in industrial applications are the *Hall sensor* (Sect. 2.8) and the *field plate* (Sect. 2.7).

7.8.2.1 Hall Sensor

The Hall sensor is one of the most common sensors found in industrial applications. It is based on the effect discovered by Edwin Hall that a voltage can be taken from a semiconductor wafer when a current flows through it and a magnetic field is applied perpendicularly to it (Fig. 2.49 in Sect. 2.8). The resulting voltage is proportional to the product of the magnetic field strength and the flowing current.

If the current is kept constant, the magnetic field strength can be inferred. However, the reverse is also possible—the field strength is constant and the current flow is the measured variable.

7.8.2.2 GMR Sensor

The GMR effect (Figs. 2.11 and 2.17 in Sect. 2.3) occurs in thin layers of ferromagnetic and non-magnetic materials. The electrical resistance of this structure depends on the direction of magnetization of the layers among themselves, which can be influenced by external fields.

The GMR effect is mainly used for read-write heads of hard disks.

7.8.2.3 Field Plate

With the field plate (Fig. 2.48 in Sect. 2.7), it is exploited that the path of the free charge carriers can be influenced by a magnetic field. This effect appears as a change in resistance depending on the magnetic field. The field plate is increasingly being replaced by GMR sensors, which have better technical properties.

7.8.2.4 SQUID

The SQUID is a sensor working in the field of superconductivity, which reacts mainly to the smallest changes in the magnetic field. However, it is not possible to perform absolute field strength measurements. Due to its high sensitivity, it also reacts to natural and artificial interfering fields, which is why it is usually constructed as a double system to compensate for these interferences.

Typical applications are the measurement of fields from brain waves and *magnetic resonance imaging* (MRI). However, there are also applications in the fields of geology and archaeology for the exploration of deposits and objects.

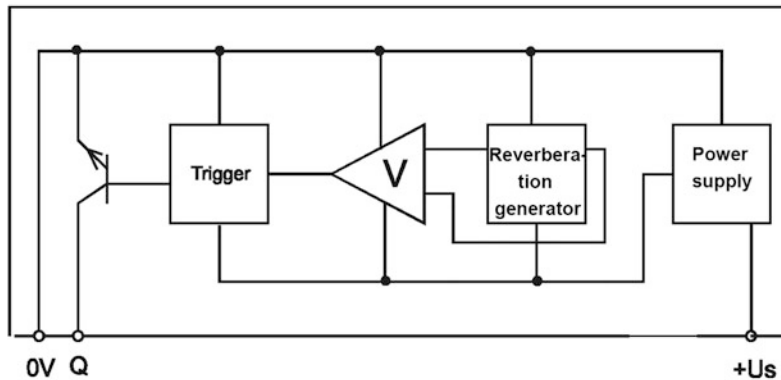


Fig. 7.37 Basic structure of a Hall switch

7.8.3 Measuring Arrangements

7.8.3.1 Hall Switch

The simplest design is *Hall switches*, which emit a bistable signal dependent on the magnetic field strength (Fig. 7.37). The switching point and switching hysteresis are specified in the product or can be set via an interface. The typical response values are in the range of a few ten mT.

The main application area of Hall switches is *position* and *attitude detection*. They generally replace mechanical limit position switches, as they have the advantage of being wear-free.

7.8.3.2 Hall Sensors

With Hall sensors (Fig. 2.53 in Sect. 2.8), the output signal is a value proportional to the magnetic field. It differs from the switch in the *analog processing* of the Hall voltage. Here, both simple amplifiers and more complex signal processing for linearization of the characteristic curve and thermal compensation can be found. The output signals are sensitive to the field direction. Typical measuring ranges here are in the range up to 100 mT.

7.8.3.3 Angle Sensors

Since the output signal depends on the field direction, it is possible to arrange several sensors in such a way that a rotating field is detected (Fig. 2.52 in Sect. 2.8). This is technically used, for example, to determine angles. However, it can also be used to detect multi-dimensional movements—like a contactless joystick.

7.8.3.4 Current Sensors

To measure a current without contact, the magnetic field of the conductor is typically used. Two main applications are the *current clamp* and the *current transformer*.

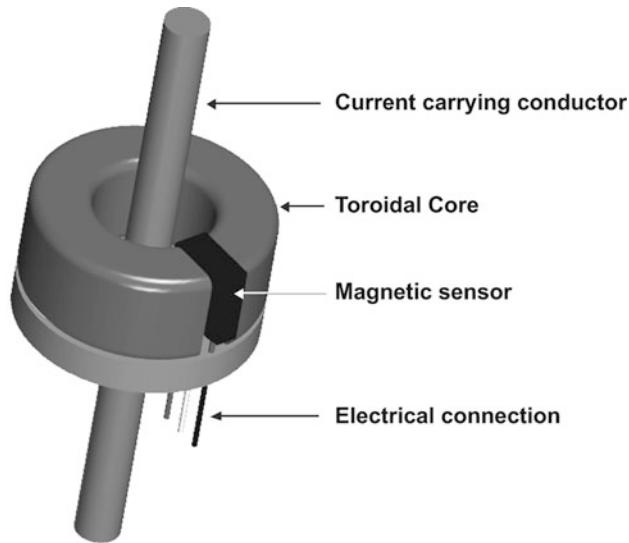


Fig. 7.38 Basic structure of a current transformer

With the *current clamp*, a core is placed around the conductor through which it flows, resulting in an inductance with 1 turn. The field concentrating in the core can be detected by a Hall sensor.

The *current transformer* uses the same principle, except that several windings can be applied to the core, thus increasing the sensitivity. An example is shown in Fig. 7.38 with a current transformer of type RAZC-2 (www.raztec.co.nz).

By using *programmable signal processing* on the Hall circuit, non-linearities and noise can be compensated. One such specialized circuit is the CUR3105 (www.micronas.com).

7.8.4 Multidimensional Measurements with the Hall Effect

7.8.4.1 Basics

A simple *Hall plate* can be used to measure the component of the magnetic field normal to the current direction and Hall voltage. The components of the magnetic field vector parallel to the current direction or Hall voltage do not produce a measuring effect. There are two ways to measure the missing components of the magnetic field vector with the Hall effect: Either the field can be distorted locally by *flux concentrators* made of ferromagnetic material in such a way that the direction of the field changes locally and thus becomes measurable for conventional Hall sensors, or special sensor geometries can be used, which are directly sensitive to the applied field (Fig. 7.39).

While comparatively sensitive sensors can be realized with flux concentrators, the special horizontal sensors can be arranged several times next to each other, for example

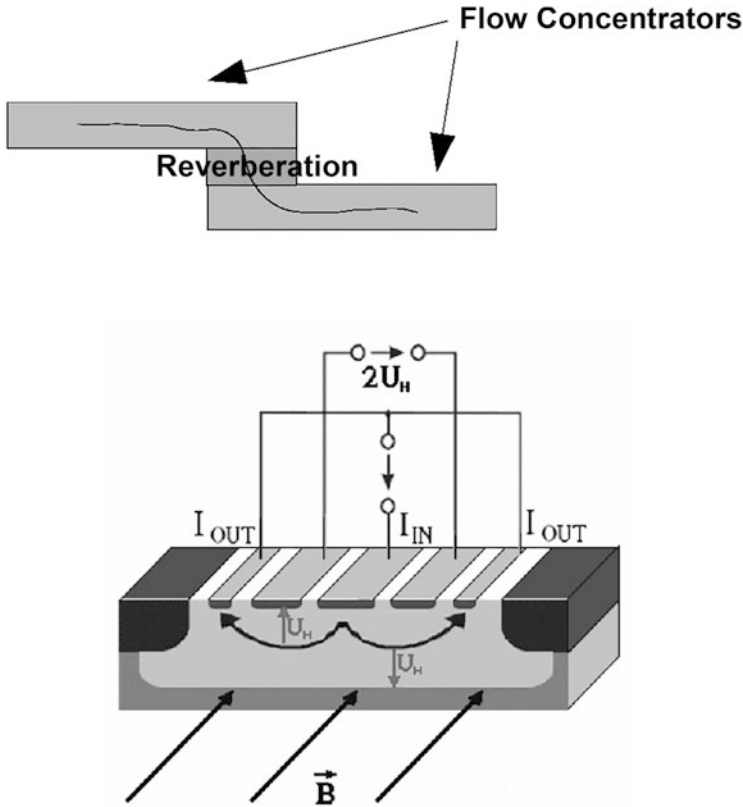


Fig. 7.39 Options for using the Hall effect in the third dimension

to detect spatial changes of the magnetic field. A symmetrical arrangement of several sensors also makes it possible to measure all three components of the magnetic field virtually at one point.

7.8.4.2 Applications

In practice, Hall sensors are often used to measure the *position* of a test magnet. This method has the advantage that it is contactless and therefore wear-free and insensitive to contamination. There are different stages for position measurements: In the simplest case, only the presence of a body is registered, for example when detecting whether a seat belt is fastened or a keypad on a mobile phone is unfolded. In this case, the information sought can be determined by a simple *threshold detection*. Often, however, a position is to be determined more precisely (e.g. with angle sensors (Sect. 3.2), rotary encoders or displacement sensors (Sect. 3.1)). In this case, there are several possibilities to determine the position of the test magnet based on the measured field. For *absolute value measurement*, the system consisting of magnet and sensor is designed in such a way that the measurable

magnetic field is clearly related to the position during movement. After linearization, the position of the magnet can then be deduced directly from the measured value for the magnetic field. The problem with this method is that, on the one hand, the magnetic materials and thus the magnetic field generated is clearly dependent on the temperature, and that considerable errors in position determination can be caused by external magnetic fields, misalignment of the sensor or sensor errors. To make the measurements more robust against such sources of error, the following methods are suitable: By simultaneously measuring the magnetic field at several locations, the spatial change of the magnetic field can be determined. When observing the changes, superimposed homogeneous fields will fall out. Since magnetic sources of interference are normally further away from the magnetic sensor than the test magnet, their interference field is often homogeneous in good approximation. This significantly reduces the disturbance influence. Another possibility is to measure different components of the magnetic field vector. By subsequently forming the ratio of the components, it is possible to infer the local direction of the field lines, which can be used for angle measurements, for example. When forming the ratio, the absolute magnitude of the magnetic field is eliminated, so that, for example, temperature variations in the magnetic material or in the sensitivity of the sensor no longer have a negative influence on the measurements.

Bibliography

1. Bretschneider, J.; Wilde, A.; Schneider, P.; Hohe, H.-P. und Köhler, U.: Entwurf multidimensionaler Positionssensorik auf Basis von HallinOne(R) Technologie 7. GI/GMM/ITG-Workshop Multi-Nature Systems: Entwicklung von Systemen mit elektronischen und nichtelektronischen Komponenten, Günzburg, 3. Februar 2009, zu beziehen über <http://www.eas.iis.fraunhofer.de>
2. Germer, H.: Fachhochschule Wilhelmshaven; Bereich Elektrische Messtechnik, Elektronik; 1999
3. Sacklowski, A.: Einheitenlexikon, Anwendung, Erl. von Gesetz und Normen; Beuth-Verlag, Berlin, Köln 1986, Praktikumsunterlagen elektronische Messtechnik, Technische Universität Chemnitz, 2004, <http://www.tu-chemnitz.de/etit/messtech/praktikum/adu.pdf>
4. Texas Instruments: ESP430CE1, ESP430CE1A, ESP430CE1B peripheral Modules user guide (slau134b. pdf), Texas Instruments, 2008



Ekbert Hering and Gert Schönfelder

8.1 Radiometry

Radiometric quantities are measured in *radiometry*. They are objectively measured physical effects and are given the index “*e*” (*energetic*). If the *eye* evaluates the radiation, *photometric* or lighting quantities are produced, which are marked with the index “*v*” (*visual*). In contrast to radiometric quantities, photometric quantities are limited to the visible range of the spectrum.

8.1.1 Radiometric Quantities

8.1.1.1 Energy Density

Light is an electromagnetic wave and transports energy. The *energy density* of an electromagnetic wave w is the sum of the energy densities of the electric field w_e and the magnetic field w_m . It therefore applies:

$$w = w_e + w_m = \frac{\epsilon_0}{2} E^2 + \frac{1}{2\mu_0} B^2.$$

Here E is the electric field strength and B is the magnetic induction. Furthermore: $E = c B$ and for the speed of light c applies:

E. Hering (✉)
Hochschule Aalen (Rektor i. R.), Aalen, Germany
e-mail: Ekbert.Hering@hs-aalen.de

G. Schönfelder
Prignitz Mikrosystemtechnik, Wittenberge, Germany

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

with the electric field constant $\epsilon_0 = 8.854 \cdot 10^{-12}$ (As)/(Vm) and the magnetic field constant $\mu_0 = 4\pi \cdot 10^{-7}$ (Vs)/(Am).

Thus, the following applies to the *energy flux density* S :

$$S = w c = \frac{1}{\mu_0} E B.$$

The energy flux density is a *vector* perpendicular to the vectors of the electric field E and the magnetic induction B . It is called the Poynting vector, so in vectorial notation

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B}.$$

For a measurement, the time average \bar{S} is decisive, since the Poynting vector \mathbf{S} varies very quickly with the frequency of the light. This temporal mean value of the energy flux density is the *irradiance* E_e . It is the quotient of the radiant power Φ_e and the irradiated area A_2 :

$$E_e = d\Phi_e/dA_2 \text{ with the unit of measurement } \text{W/m}^2.$$

8.1.1.2 Radiant Power Φ_e

The radiant power Φ_e describes the *power* of electromagnetic radiation and indicates how much *radiant energy* Q_e per time unit hits a receiver. It is:

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

The radiant power is measured in W or in J/s.

8.1.1.3 Solid Angle Ω

The radiant power Φ_e depends not only on its area A but also on the distance r between the light source and the surface under consideration. The *solid angle* Ω is defined as (Fig. 8.1):

$$\Omega = A/r^2 \Omega_0, \text{ mit } \Omega_0 = 1 \text{ sr.}$$

Its unit of measurement is the *steradian*, for which the following applies: $1 \text{ sr} = 1 \text{ m}^2/\text{m}^2$. The largest solid angle Ω_{\max} is:

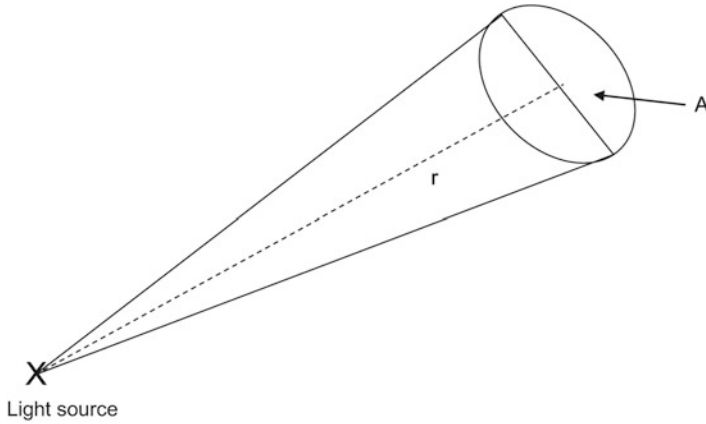


Fig. 8.1 Definition of the solid angle Ω

$$\Omega_{\max} = (4\pi r^2)/r^2 \text{ sr} = 4\pi.$$

In this case, the radiation affects the whole room. If the radiation hits a half space, the solid angle $\Omega = 2\pi$ sr.

A transmitter emits light with an area of A_1 . This light strikes a receiver with an area of A_2 (Fig. 8.2). Then the effective solid angle is valid:

$$\Omega = \frac{A_2 \cos(\varepsilon_2)}{r^2} \Omega_0.$$

$A_2 \cos(\varepsilon_2)$ is the projection of the area A_2 on the connecting line of transmitter and receiver.

8.1.1.4 Radiant Intensity I_e

The radiant power Φ_e is proportional to the solid angle Ω , so that applies:

$$\Phi_e = I_e \Omega.$$

The proportionality constant is the *radiant intensity* I_e . The radiant intensity is thus the quotient of the radiant power and the solid angle into which the radiation emerges. It applies:

$$I_e = d\Phi_e/d\Omega_1 \approx \Phi_e/\Omega \text{ in W/sr.}$$

The radiant intensity I_e depends on the material, temperature, surface and other transmitter properties. A *polar diagram* usually shows the radiant intensity I_e as a function of the radiation angle ε_1 (Fig. 8.3).

Fig. 8.2 Solid angle for a transmitter with area A_1

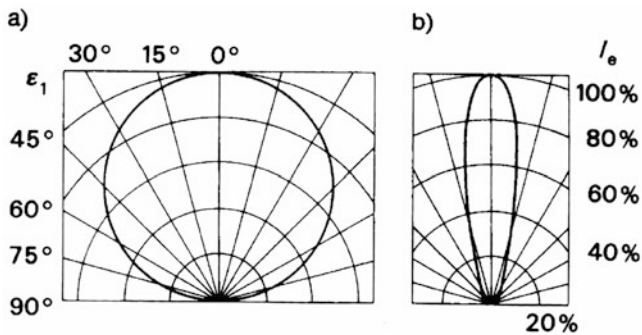
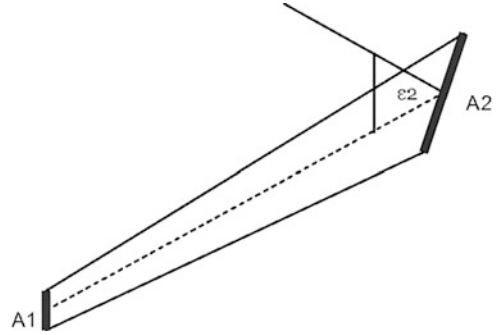


Fig. 8.3 Polar diagram of radiant intensity I_e as a function of the beam angle ϵ_1 a) Lambert radiator; b) light-emitting diode (Source: Hering, Martin, Stohrer: Physik für Ingenieure)

Figure 8.3 a shows a so-called *Lambert radiator* in which the radiance I_e in space is constant. The *Lambert's law of cosine* applies:

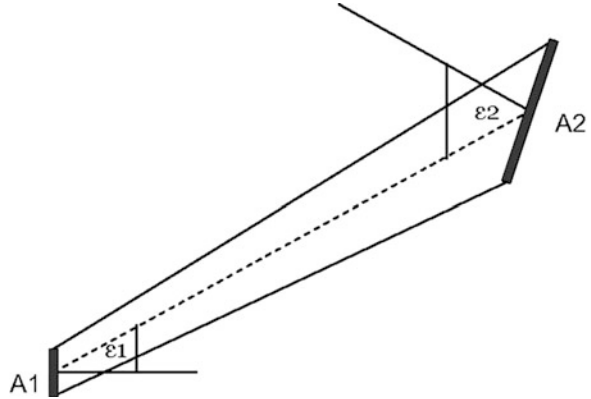
$$I_e(\epsilon_1) = I_e(0) \cos \epsilon_1.$$

All bodies with diffusely reflecting surfaces (e.g. plasterboard walls, white woodchip walls, concrete walls, paper or cardboard) are approximately Lambert radiators. LEDs show a slender lobe of light intensity forward. They are preferably used for light barriers (Sect. 2.14.3).

8.1.1.5 Radiance L_e

The radiant intensity I_e is proportional to the area of the transmitter A_1 . If the transmitter surface is viewed from the side at the angle ϵ_1 , only the projection $A_1 \cos(\epsilon_1)$ is effective, so that the following applies

Fig. 8.4 Relationship between radiant intensity I_e and radiance L_e



$$I_e(\epsilon_1) = L_e A_1 \cos(\epsilon_1).$$

Thus, the radiance L_e :

$$L_e = I_e / (A_1 \cos(\epsilon_1)).$$

Figure 8.4 shows the relationships. The unit of measurement for *radiance* L_e is $W/(m^2 \text{ sr})$. It indicates the radiant power emitted per area and solid angle.

Table 8.1 clearly summarizes the radiometric quantities.

8.1.2 Measurement of Electromagnetic Radiation

The range of electromagnetic radiation extends over a very large range of about 22 powers of ten: Maximum from the photon energies of over 10^{12} eV for *cosmic gamma radiation* to minimum to *low frequency radio waves* with photon energies below 10^{-10} eV. This section deals with the range between *X-rays* and *far infrared*.

For this wide range of electromagnetic radiation, there are different basic measuring principles depending on the wavelength, which, with the exception of thermal detectors, are in turn based on the photoelectric properties of different materials depending on the sub-wavelength range. These photoelectric sensors are discussed in detail in Chap. 14.

8.2 Photometry

The photometric quantities take into account the sense of *brightness* of the human eye. The perception of brightness not only depends on the physical radiant power in W that hits the eye, but also to a large extent on the wavelength λ . For example, a green LED appears about 16 times brighter than a red LED for the same radiant power. Figure 8.5 shows the course of

Table 8.1 Compilation of radiometric quantities. The simplified equations apply under the condition that the radiation energy is constant with respect to time, area and solid angle. If this is not the case, the simplified equations apply to the mean values (Source: Hering, Martin, Stohrer: Physik für Ingenieure)

Size	Symbol	Unit	Relationship	Declaration
Radiant energy	Q_e	W s = J	$Q_e = \int \Phi_e dt$	Energy transport through electromagnetic radiation
Radiant power	Φ_e	W = J/s	$\Phi_e = \frac{dQ_e}{dt}$	Power of electromagnetic radiation
Specific radiation	M_e	W/m ²	$M_e = \frac{d\Phi_e}{dA_1} \approx \frac{\Phi_e}{A_1}$	Ratio of radiated power and transmitter area
Radiant intensity	I_e	W/sr	$I_e = \frac{d\Phi_e}{d\Omega_1} \approx \frac{\Phi_e}{\Omega_1}$	Ratio of radiant power and solid angle into which the radiation is emitted
Radiance	L_e	$\frac{W}{sr \cdot m^2}$	$L_e = \frac{d^2\Phi_e}{d\Omega_1 dA_1 \cos \epsilon_1}$ $= \frac{dI_e}{dA_1 \cos \epsilon_1}$ $L_e \approx \frac{I_e}{A_1 \cos \epsilon_1}$	Ratio of radiant power to solid angle (i.e. radiant intensity) and projection of the transmitter surface on a plane perpendicular to the viewing direction
Irradiance	E_e	W/m ²	$E_e = \frac{d\Phi_e}{dA_2} \approx \frac{\Phi_e}{A_2}$	Ratio of radiant power and irradiated area
Irradiation	H_e	J/m ²	$H_e = \int E_e dt \approx E_e t$	Time integral of irradiance

the brightness sensitivity of the human eye for daytime vision and for night vision. These progressions were determined experimentally with a large number of test persons and defined by the CIE (Commission Internationale de l'Eclairage) and documented in tables in DIN 5031.

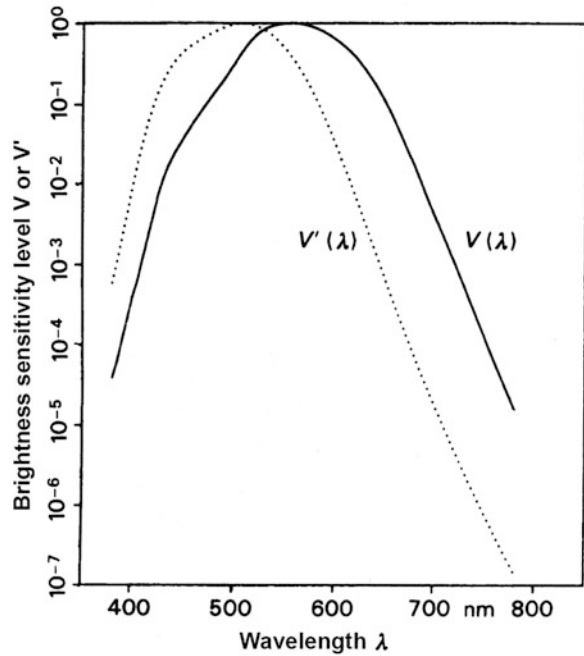
8.2.1 Photometric Quantities

The corresponding photometric quantities are explained below. The index “v” indicates that these are quantities adapted to the eye (v: visual).

8.2.1.1 Luminous Flux Φ_v

The luminous flux Φ_v is the light output emitted by a light source in all directions. It reflects the *impression of brightness* and is measured in *lumen* (lm). To determine the luminous flux, the radiometric quantity of the radiant power Φ_e is evaluated with the luminous sensitivity $V(\lambda)$. The following then applies to daylight vision

Fig. 8.5 Brightness sensitivity level of the standard observer. $V(\lambda)$: day vision (cone vision), photoptical adaptation; $V'(\lambda)$: night vision (rod vision), scotopic adaptation (source: Hering, Martin, Stohrer: Physik für Ingenieure)



$$\Phi_V = K_m \Phi_e V(\lambda)$$

K_m is the maximum value of the *photometric radiation equivalent* for daytime vision and is: $K_m = 683 \text{ lm/W}$ (lumen per Watt).

The same applies to night vision:

$$\Phi_{V'} = K_m' \Phi_e V'(\lambda).$$

In this case, $K_m' = 1699 \text{ lm/W}$. In the following, only the formulas for daytime vision are mentioned.

If the radiation covers not only one wavelength, but a large number of wavelengths, i.e. is broadband, then the luminous flux Φ_V is *obtained* by integration over the visible spectrum, so that the following applies

$$\Phi_V = K_m \int_{380 \text{ nm}}^{780 \text{ nm}} \Phi_{c,\lambda}(\lambda) V(\lambda) d\lambda.$$

From the radiometric quantities X_e , the corresponding photometric quantities X_v can be determined by this evaluation. In general, the following applies:

$$X_V = K_m \int_{380 \text{ nm}}^{780 \text{ nm}} X_{c,\lambda}(\lambda) V(\lambda) d\lambda.$$

8.2.1.2 Light Quantity Q_V

The amount of light corresponds to the radiant energy of a light source, weighted with the sensitivity curve $V(\lambda)$. It is measured in lm s .

8.2.1.3 Luminous Intensity I_V

The photometric quantities are linked to the SI basic unit *candela* (cd) as a unit of measurement for *luminous intensity* I_V . The luminous intensity indicates the light radiation of a light source in a certain direction. It applies:

$$I_V = \Phi_V / \Omega.$$

The SI unit of measurement *candela* is defined as follows:

1 candela (cd) is the *luminous intensity* in a given direction of a radiation source emitting monochromatic radiation at a frequency of 540 THz (corresponding to a wavelength of 555 nm) and whose radiant intensity in that direction is 1/683 W/sr.

This means that for the luminous intensity of 1 cd:

$$I_V = 1 \text{ cd} = K_m I_e = K_m (1/683) \text{ lm/sr}.$$

The constant K_m according to this relation is $K_m = 683 \text{ lm/W}$. This is called *photometric radiation equivalent* and concerns day vision. For night vision, the constant $K_m' = 1699 \text{ lm/W}$.

8.2.1.4 Luminance L_V

Luminance reflects the *impression of brightness* that is produced in the eye by a luminous surface. It is a measure of the luminous intensity of a light source in relation to its surface. The unit of measurement is cd/m^2 . The luminance L_V is the ratio of luminous intensity I_V to the luminous surface A_{luminous} . The following relationship applies:

$$L_V = I_V / A_{\text{luminous}}.$$

The luminance depends on the reflection properties of the illuminated materials. In lighting technology, illuminance E_V is therefore used as a planning parameter.

8.2.1.5 Illuminance E_v

Illuminance is a measure of how *intensively* a surface is illuminated. The illuminance E_v is the ratio of the luminous flux Φ_v to the size of the illuminated area A . The unit of measurement is lm/m^2 or Lux (lx). It applies:

$$E_v = \Phi_v/A.$$

8.2.1.6 Specific Light Emission M_v

Here, the emitting luminous flux $\Phi_{v, \text{emitted}}$ is set in relation to the *light-emitting* area A_{emitted} . The unit of measurement is also lm/m^2 or lx. It applies:

$$M_v = \Phi_{v, \text{emitted}}/A_{\text{emitted}}.$$

8.2.1.7 Luminous Efficacy η

It indicates how large the *luminous flux* Φ_v is in relation to the connected electrical load P :

$$\eta_e = \Phi_v/P,$$

measured in lm/W . The radiometric quantity for the *radiation yield* is

$$\eta_v = \Phi_e/P$$

measured in W/W with unit of measurement 1. Multiplied by 100, it indicates how much % of the electrical power is available as radiant power.

The luminous efficacy is a measure of the *energy efficiency* of a light source. The higher the value, the more light is generated per input power. Table 8.2 shows the values for different light sources.

Table 8.3 compares the radiometric and photometric quantities.

Table 8.4 shows the luminous flux of some light sources, Table 8.5 the luminances and Table 8.6 the illuminance of some light sources.

8.2.2 Measurement of Photometric Quantities

In the following, some measuring instruments for photometric quantities are shown. Section 2.14.3 (Applications of photoelectric effects) also describes measurement possibilities of photoelectric quantities (e.g. Fig. 2.92: Luxmeter). Figure 8.6 shows a measuring device for measuring luminous flux in lumens. With it, LEDs can be tested by setting a reference value to which one can orient oneself. This measuring device can also be

Table 8.2 Luminous efficiency of some light sources

Light source	Luminous efficacy in lm/W
Flame, candle, oil lamp	0.2
Light-emitting diode, blue	8.5
Light-emitting diode, red	47.5
Incandescent lamp, 230 V, 40 W	10.0
Incandescent lamp, 230 V, 70 W	12.0
Halogen, 230 V, 500 W	20.0
Fluorescent lamp, 230 V, 70 W	75
Fluorescent lamp with electronic ballast	95
Xenon high-pressure gas discharge lamp (video projection)	22.5
Xenon high-pressure gas discharge lamps (cinema projection)	47
Xenon arc lamp (car headlights)	90
Low-pressure sodium vapour lamp	175
High-pressure sodium vapour lamp	150

Table 8.3 Comparison of the radiometric and photometric quantities

Radiometric quantity	Character	Unit	Photometric size	Character	Unit
Radiant energy	Q_e	W s	Amount of light	Q_v	lm s
Radiant power	Φ_e	W	Luminous flux	Φ_v	lm
Radiant intensity	I_e	W/sr	Light intensity	I_v	cd = lm/sr
Radiation density	L_e	W/(m ² sr)	Luminance	L_v	cd/m ²
Irradiance	E_e	W/m ²	Illuminance	E_v	lx = lm/m ²
Irradiation	H_e	W s/m ²	Exposure	H_v	lx s
Specific radiation	M_e	W/m ²	Specific light emission	M_v	lm/m ²
Radiation yield	η_e	W/ W = 1	Luminous efficacy	η_v	lm/W

Table 8.4 Luminous flux Φ_v of some light sources

Light source	Luminous flux in lm
LED (white)	0.02 to 200
Incandescent lamp, 230 V, 60 W	730
Incandescent lamp, 230 V, 100 W	1380
Fluorescent tube, 230 V, 20 W	1200
Fluorescent tube, 230 V, 40 W	2300
Mercury vapour lamp, 230 V, 125 W	5400
Mercury vapour lamp, 230 V, 2000 W	125,000

Table 8.5 Luminance L_v of some light sources

Light source	Luminance in cd/m^2
Night sky	10^{-3}
Moon	2500
Clear sky	3000 to 5000
Fluorescent lamp	$1.3 \cdot 10^4$
Incandescent lamp matt	$5 \cdot 10^4$ to $4 \cdot 10^5$
Bulb clear	$2 \cdot 10^6$ to $2 \cdot 10^7$
Halogen incandescent lamp	$5 \cdot 10^7$
Metal halide lamp	108
Xenon lamp (short arc)	$1.5 \cdot 10^8$ to $2.7 \cdot 10^9$
Black body	$6 \cdot 10^5$
Sun ($T = 2200 \text{ K}$)	$1.5 \cdot 10^9$

Table 8.6 Illuminance E_v of some light sources

Light source	Illuminance in lx
Street lighting	1 to 15
Living room lighting	120
For reading at least	100
Job with normal entitlement	500
A workplace with high standards	1000
Colour perception limit	3
Stars, without moon, clear night	10^{-3}
Full Moon	0,25
Daylight, overcast sky	1000 to 2000
Sun, winter	6000
Sun, summer	70,000

used in the development of LED luminaires. The data can be read into a data memory and transferred to a computer with a USB port. There they are further processed accordingly.

Figure 8.7 shows a *light meter* (luxmeter). It can also be used as a *luminance meter* (cd/m^2) with an attachment. This device is used in industry, trade, at photo and film studios as well as in research and development. Due to its high precision and sensitivity, it is also very suitable for certification applications.

8.3 Application of Brightness Sensors

A very complex task for the brightness sensors is the control of lighting equipment. This multidimensional optimization includes the following questions

Fig. 8.6 LED luminous flux meter PCE-LED 1 (Factory photo: PCE Instruments)



Fig. 8.7 Luxmeter and illumination density meter Mavolux 5032 C (Factory photo: Gossen, Foto- und Lichtmesstechnik GmbH)



- the optimal lighting as a function of the work tasks taking place there,
- energy efficiency in lighting,
- the compensation of the light colour—also under the influence of daylight and time of day and
- the current usage situation of the room.

The use of LEDs makes it possible to implement new lighting concepts that allow not only dimming but also a change in the colour of the light. Figure 8.8 shows a room with very flat, wide-area light sources. These are electronically controlled by a brightness sensor (Fig. 8.9), which also detects the presence of people.



Fig. 8.8 Energy-efficient LED lighting with Slimpanel SP 595 (Factory photo: Richter lighting technologies GmbH)

Already the use of this sensor combination—brightness and presence—brings a qualitative progress and a noticeable energy saving.

The spectrum can also be defined by the targeted use of special LEDs and, thanks to electronic control, allows free selection in this available range. The spectrum of room luminaires is shown in Fig. 8.10. The CIE standard colour chart shows the colour coordinates and colour temperature of the luminaire (cross approximately in the middle).

Fig. 8.9 Combined presence and brightness sensor (Factory photo: PEMA Elektro GmbH). A—Motion detector, B—Brightness sensor, C—IR receiver for programming unit

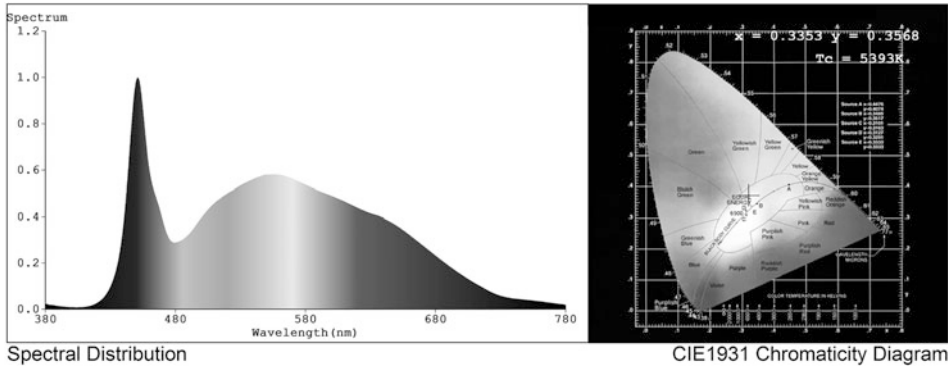
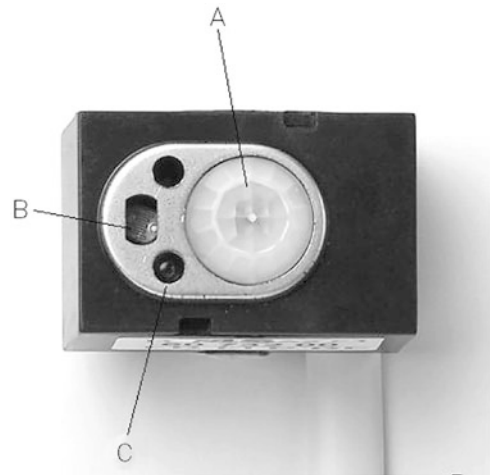


Fig. 8.10 Spectrum of the Slimpanel SP 595 (Factory photo: Richter lighting technologies GmbH)

8.4 Colour

8.4.1 Colour Perception

Despite all difficulties, it is necessary to describe colours *clearly* and *objectively*. With these figures, colours can be output from screens, printed out by printers and clearly identified as a colour in the painting field. The perception of colours by the eye is done by *receptors* located on the retina. These are the *rods* for light-dark contrast and the *cones* for *colour recognition*. There are three different cones, which have their maximum sensitivity for red, *green* and *blue*. Any colour can be created by *mixing* these *primary colours*. Table 8.7 shows the *primary valences* defined by the CIE (Commission Internationale de l'Éclairage) as early as 1931. The relative radiation power refers to the

Table 8.7 Primary valence triplets for red, green and blue with their relative radiant power

Colour	Wavelength in nm	Relative radiant power
Red <i>R</i>	700.0	72,096
Green <i>G</i>	546.1	13,791
Blue <i>B</i>	435.8	1000

primary color blue, whose power was defined as 1. It can be seen that the colour red has a radiant power that is about 72 times greater than that of blue.

A color F can be represented in a three-dimensional color space. The vectors R , G and B span a three-dimensional space in which the components red (R), green (G) and blue (B) are represented in this space:

$$F = RR + GG + BB.$$

For the representation of colors on monitors and in printers, the components have been standardized and can be displayed as hexadecimal numbers in the computer (Table 8.8).

In order to be able to define all colours uniformly, a *standard valence system* was introduced. This system has the *standard valences* X , Y and Z . In this three-dimensional system, any color F can be represented, so that the following applies

$$F = XX + YY + ZZ.$$

X , Y and Z are the *standard color values*, which are calculated as follows:

$$X = k \int \Phi_{\lambda} \bar{x}(\lambda) d\lambda,$$

$$Y = k \int \Phi_{\lambda} \bar{y}(\lambda) d\lambda \text{ and}$$

$$Z = k \int \Phi_{\lambda} \bar{z}(\lambda) d\lambda.$$

The constant k can be chosen appropriately. The *standard spectral values* $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ were determined by measurements on test persons and are specified in CIE (1931) and DIN 5033. They correspond to a *standard colorimetric observer* with a field of view of 2° . The course of the standard spectral values is shown in Fig. 8.11. The curves are drawn so that the areas under the curve are always the same size.

The standard spectral value function $\bar{y}(\lambda)$ is identical to the curve for the brightness sensitivity level $V(\lambda)$ of Fig. 8.5. Therefore, the standard color value Y is proportional to the photometric quantities.

If the *brightness* is omitted, the real color can be defined by *two specifications*. In this way, you get a *standard color table*. The *proportions* of the standard colour values are

Table 8.8 Color table for red, green and blue with hexadecimal code

Colour	R,G,B	Hexadecimal number
Black	0,0,0	000000
Blue	0,0,128	000080
Green	0,128,0	008000
Red	255,0,0	FF0000
White	255,255,255	FFFFFF
Yellow	255,255,0	FFFF00
Purple	128,0,128	800080
Olive green	128,128,0	808000
Grey	128,128,128	808080
Silver	208,208,208	C0C0C0

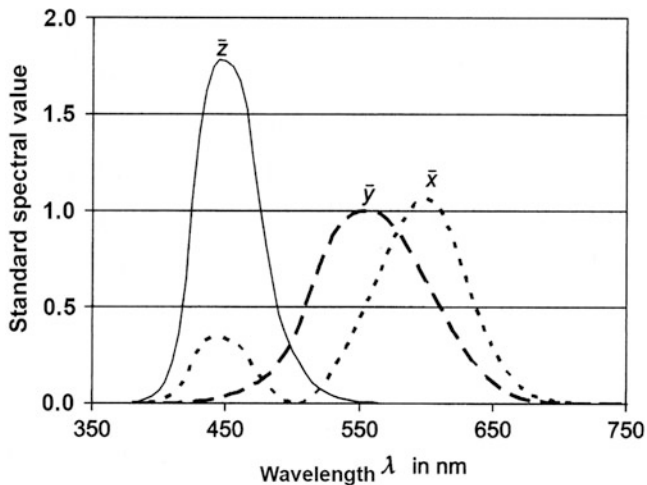


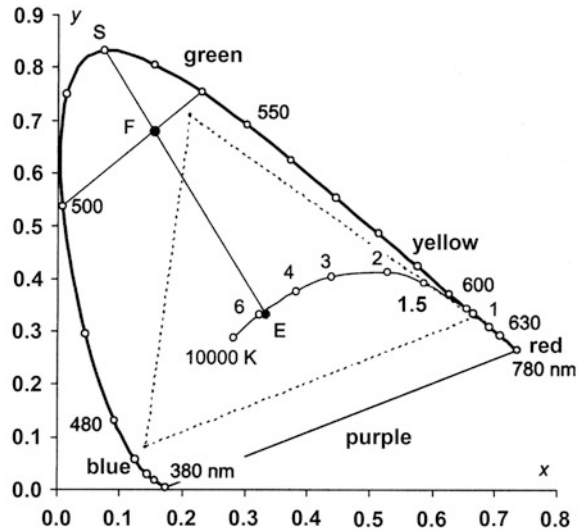
Fig. 8.11 Standard spectral value functions for the standard colorimetric observer with 2° field of view (Source: Hering, Martin, Stohrer: Physik für Ingenieure)

determined and y is plotted as a function of x . It is not necessary to give the value for z , because $x + y + z = 1$. The *standard chromaticity coordinates* are calculated:

$$\begin{aligned}
 x &= X/(X + Y + Z), \\
 y &= Y/(X + Y + Z) \text{ and} \\
 z &= Z/(X + Y + Z).
 \end{aligned}$$

In this two-dimensional representation according to Fig. 8.12, each *color* is assigned to a *point*. The standard color value components of the spectral colors form a closed curve, the *spectral color train*. The connecting line of its vertices is the *purple line*. All *real (true)* colors lie *within* the enclosed area. The curve, which passes through *color location E*,

Fig. 8.12 Standard colour chart for 2° field of view (Source: Hering, Martin, Stohrer: Physik für Ingenieure)



shows the color locations of the black body at 1000 °C. Within the triangle formed by dashed lines are the *RGB colour locations*, which can be realised with a colour television tube.

8.4.2 Colour Models

As mentioned in the previous section, a uniform standard is required for color measurement and evaluation. This standard was created with the CIE1931, which is based on the *RGB colour model*. In the course of further development, this was extended by the Lab standard in 1976. One reason for this was also the measurement requirements for color registration.

An RGB colour model always applies to a specific system, i.e. a monitor, printer or scanner. For technical reasons, a brick red on the monitor does not necessarily appear as such on the printer. The Lab system is based on human perception and is therefore device-independent. The *CIE Lab colour space* is standardised in DIN 6174.

The Lab color space is based on equal numerical distances for sensory equal distances. The color space is represented in the planes a and b. The third axis corresponds to the brightness (luminance L) with the end points 0 for black and 100 for white.

The Lab room is used for measurements on *non-self-emitting devices (linear subtractive mixing)*. The conversion of RGB colours into the Lab space is done using the standardised colour value XYZ according to CIE1931.

Brightness $L^* = 116 \cdot Y^* - 16$.

Green-red difference $a^* = 500(X^* - Y^*)$.

Blue-yellow difference $b^* = 200(Y^* - Z^*)$

$$\text{with } X^* = \sqrt[3]{\frac{X}{X_n}} \quad \text{for } \frac{X}{X_n} > 0,008856$$

$$\text{or } X^* = 7,787 \sqrt[3]{\frac{X}{X_n} + 0,138} \quad \text{for } \frac{X}{X_n} \leq 0,008856.$$

The values for Y^* and Z^* are formed analogous to X^* .

The *achromatic point* is then: $a^* = 0.0$ and $b^* = 0.0$.

The *colour distance* ΔE^*_{ab} thus results in

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}.$$

8.4.2.1 Other Color Spaces

For *self-emitting lamps (linear additive mixing)*, the *windward color space* is used, which was also defined by the CIE in 1976. The Lab and Luv color space are often used for technical image processing, because for the machine evaluation of images, the even color spacing provides better feature selection.

Another system is the *LCh color space*, which corresponds to the Lab, but the representation is not in Cartesian coordinates but in *cylindrical coordinates*.

The *HSB colour model* is also constructed in cylindrical coordinates. It is made up of the three components hue, saturation and brightness. The hue is described by an angle of 0° to 360° , the other two quantities by a value of 0% to 100%.

CMYK is a colour space for printing, which is based on subtractive colour mixing. The colours *cyan, magenta, yellow* and *black* are used. The colour value covers the range from 0% to 100%, whereby 0% represents no colour and 100% represents a full colour area.

8.4.3 Color Systems

Different technical colour systems have been developed in the course of industrial use. These are generally used for quality assurance in a specific market sector. Particularly worth mentioning here are:

<i>RAL colours</i>	Splits the color palette using a four-digit number key. Conversion to other color systems is done using tables. The system is mainly used in the field of decorative paint applications.
--------------------	--

(continued)

<i>Pantone colours</i>	This system is mainly used in the printing industry and represents the binding colour set. It is based on the use of the CMYK combination. In this system, special colours are also permitted, which lie outside and cannot be achieved with four-colour printing. There are several substandards in Pantone.
------------------------	---

In pure color sensor technology, the aim is often to trace the measured values back to one of these systems.

8.4.4 Colour Filters for Sensors

When using color sensors, the color considered for the measurement must always be selected from the spectrum. Since photosensors only react roughly to spectral ranges, they must be equipped with a *colour filter*. There are two basic technical approaches

- the *printing of* coloured filters in the form of a lacquer foil on the sensor element or
- the application of *diffraction gratings* as filter structure.

Both methods are used for sensors. The color filters are generally used for image sensors. If special colour measurement techniques are required, diffraction gratings are used, as these are more selective and show less losses.

For a definition of the spectral working range in semiconductors, there is another effect. Radiations with different wavelengths have different penetration depths in a semiconductor layer. On the basis of this property, the spectral working range of a sensor can basically be set up (Chap. 14).

The other implementation of this knowledge, which is also technically realized, are sensors that create a color image by taking information at different depths of the semiconductor structure. The advantage here is that the color information of the entire spectrum is captured at a common location. This is in contrast to the other filter systems, where the individual colors are resolved next to each other. The disadvantage of this method, however, is that, for physical reasons, no colourful images are created from a visual point of view. Therefore, it is only used under special conditions of image processing.

8.4.4.1 Bayer Pattern

The *Bayer pattern* is the most common filter for color sensors (Fig. 8.13). It consists of a grid of the three basic colors RGB with double green. This results from the great importance of the green component for human perception.

The already mentioned disadvantage is that always four pixels together result in a colored pixel, which reduces the local resolution. The sensor always outputs the pixel information in this raw form—available as raw format in high-quality cameras.

By means of modern interpolation methods, the calculation is not done in steps of double pixels, but rather this is adapted to the pixel sequence. If the color pixels of the first line are interpolated, the result is a calculation of the following form:

Fig. 8.13 Detail of a Bayer pattern

	x1	x2	x3	x4	x5
1x	R	G	R	G	R
2x	G	B	G	B	G
3x	R	G	R	G	R
4x	G	B	G	B	G
5x	R	G	R	G	R

$$C11 = \{P11, P12, P21, P22\}$$

$$C12 = \{P12, P13, P22, P23\}$$

....

The following line begins with the interpolation of

$$C21 = \{P21, P22, P31, P32\}.$$

P_{xx} are the pixel positions in Fig. 8.13 and C_{xx} are the resulting color pixels.

Although the Bayer pattern is the most common, there are other filters as well. For special applications, there are also sensors with color filters consisting of vertical or diagonal stripes, which are easier to produce. In the same way, other color combinations can be found, such as filters with complementary colors.

8.4.4.2 Color Correction

Since the individual color filters have different transmission properties and the pixels have different spectral sensitivities, a basic correction must then be made. This is done using the *color correction matrix*, which is documented by the sensor manufacturer in the data sheets.

$$\begin{bmatrix} R_{\text{cor}} \\ G_{\text{cor}} \\ B_{\text{cor}} \end{bmatrix} = \begin{bmatrix} K1 & K2 & K3 \\ K4 & K5 & K6 \\ K7 & K8 & K9 \end{bmatrix} \cdot \begin{bmatrix} R_{\text{raw}} \\ G_{\text{raw}} \\ B_{\text{raw}} \end{bmatrix}$$

The coefficients come from the manufacturer. The multiplication of a column corresponds to a gain change of the channel at the input.

The coefficients $K1$, $K5$ and $K9$ are generally positive, the others negative.

8.4.4.3 White Balance

After color correction has been made, the *white balance* of a sensor is required. Simplified, it assumes that there is a white point—or grey point—in each image, which corresponds to the colour temperature of the light source. For a camera, this is expressed in the choice between daylight and artificial light. However, the weak point can already be seen here—if

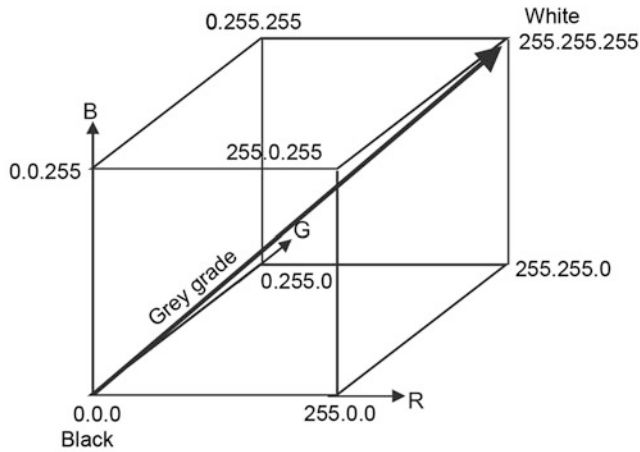


Fig. 8.14 White balance in RGB color space

a purely colored object is viewed (e.g. a green wall or lighting with colored light), this adjustment usually does not lead to success. The mathematically very complex calculations for the white balance can be illustrated in Fig. 8.14.

White balance assumes that all pixels are in the color space, with an accumulation around the gray levels. The gray levels represent all black and white gradations.

During the adjustment, a gray level of the image is calculated from the pixel data and then the content of the entire image space is rotated so that the level of the image matches the ideal level as closely as possible. During the rotation, however, it may happen that pixels leave the value limits of the space, which is why an additional adjustment of the gain may be necessary.

8.4.5 Colour Sensors

Since image sensors are discussed in detail in Chap. 14, a solution with diffraction grating is referred to below.

In Sect. 2.14.3, Fig. 2.93 shows a color sensor based on the RGB principle. Figure 8.15 shows the color sensor of type MTCSi from MAZeT, which allows to detect colors in their physical appearance. It is mainly used for the adjustment and analysis of displays and in the printing and textile industry. The course of the standard color values is shown by Z (first peak from the left), Y (second peak from the left) and X (third peak from the left). Figure 8.16 shows a color scanner based on this sensor, which determines the color values.

Fig. 8.15 Color sensor for measuring true color values (Factory photo: MAZeT GmbH)

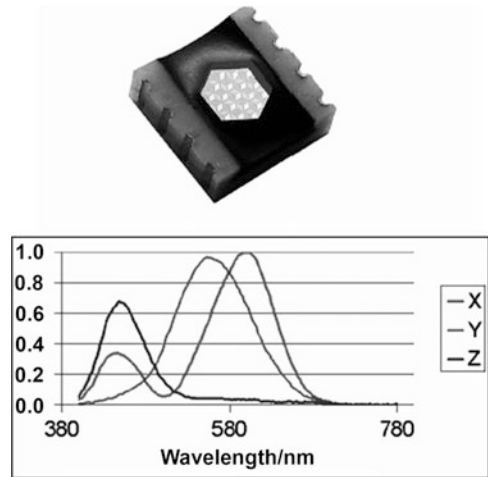


Fig. 8.16 Jencolor color scanner for measuring true color values (Factory photo: MAZeT GmbH)



Bibliography

1. Hering, E.; Martin, R.; Stohrer, M.: Physik für Ingenieure, 13. Auflage 2021, Springer Verlag <http://farbe.wisotop.de/Arbeitsfarbraum.shtml>
2. Umrechnung von Farbräumen: <http://www.frank-schibilla.de/IT/RAL-Pantone-CMYK/RAL-Pantone-CMYK.php>; <http://ngin.de/cmyk2pantone/index.php?c=50&m=10&y=50&k=25&hits=10>
3. Wagner, P.: Farbmanagement; ScanDig GmbH; <http://www.filmscanner.info/Farbmanagement.html>



Ekbert Hering

Acoustics describes the occurrence, properties and measurement of *longitudinal waves* (local compressions and dilutions of vibrating particles) in gases, liquids and solids. Of particular importance are the *propagation of sound in air* and the human *perception of sound*. A distinction is made according to amplitude and frequency response:

- *Sound*: a single sound frequency and a sinusoidal amplitude.
- *Noise*: a multitude of different frequencies and strongly fluctuating amplitude response.
- *Bang*: very many frequencies at almost constant amplitude.

9.1 Definition of Important Acoustic Quantities

Sound pressure p

The sound pressure p is a measure for the *pressure fluctuations* in a medium. For a tone (periodic fluctuation with only one frequency), the following applies:

$$p(t) = \langle p \rangle \cos(\omega t - \varphi).$$

It is $\langle p \rangle$: amplitude of the sound pressure; ω : angular frequency ($\omega = 2\pi f$); φ : phase shift. The sound pressure p is usually many orders of magnitude smaller than the static air pressure. It is a scalar quantity.

E. Hering (✉)
Hochschule Aalen (Rektor i. R.), Aalen, Germany
e-mail: Ekbert.Hering@hs-aalen.de

Sound velocity v (German: Schallschnelle)

It describes the speed with which the particles oscillate around their resting position. Like sound pressure, you can write for a sound:

$$v(t) = \langle v \rangle \cos(\omega t).$$

The sound velocity is a vector.

Sound velocity c

It indicates the speed with which the sound wave propagates in the medium. The speed of sound depends on the medium. The following relationships apply:

- *Gases*: $c = \sqrt{\frac{\kappa p}{\rho}}$ with $\hat{\epsilon}$: isentropic exponent; p : sound pressure; ρ : specific gravity.

For air, the speed of sound $c_{\text{air}} = 331.2$ m/s

- *Liquids*: $c = \sqrt{\frac{K}{\rho}}$ where K : compression modulus.

The speed of sound in water is: $c \approx 1490$ m/s.

- *Solid state*: $c = \sqrt{\frac{E}{\rho}}$ where E : modulus of elasticity.

For steel, the speed of sound $c \approx$ is 5800 m/s.

Sound power P_{sound}

This is the energy emitted per unit of time by a sound source. It applies:

$$P_{\text{sound}} = \int p(t) v(t) n \cdot dA,$$

with $n \cdot dA$: surface element dA in the direction of the surface normal vector n .

Sound Intensity I_{sound}

This describes the sound power per unit area. It applies:

$$I_{\text{sound}} = p(t) v(t).$$

Sound impedance Z_A

It is the *characteristic impedance* of the medium and indicates how the sound velocity is generated by a pressure excitation. It applies:

$$Z_A = \rho c.$$

9.2 Human Perception**9.2.1 Level**

Sound pressure waves are picked up by the human ear and perceived according to frequency and sound pressure. The hearing range is between 16 Hz and 20 kHz. The sound pressure for the *hearing threshold* is $2 \cdot 10^{-5}$ Pa and the *pain threshold* is about 20 Pa. Normal sounds are around 0.1 Pa. Because of this large range, logarithms are used (10 logarithms) and the *levels are* calculated. For example, for the sound pressure level L_P :

$$L_P = 10 \log \left(\frac{p_{\text{eff}}^2}{p_0^2} \right) = 20 \log \left(\frac{p_{\text{eff}}}{p_0} \right), \text{ with } p_{\text{eff}} : \text{Effective value of the sound pressure;}$$

with p_0 : Reference pressure.

Table 9.1 lists the corresponding sound levels. The levels are measured in decibels (dB). According to these relationships, the sound intensity level is at the hearing threshold:

$$L_{\text{hearing thres}} = 10 \log (I_0)/I_0 = 0 \text{ dB.}$$

The sound intensity of the pain threshold is 1 W/m^2 . This means:

$$L_{\text{pain thres}} = 10 \log (1/10^{-12}) = 10 \log 10^{12} = 120 \text{ dB.}$$

The following applies to sound pressure:

- Sound pressure at the hearing threshold: $p_{\text{hearing thres}} = 2 \cdot 10^{-5}$ Pa and
- Sound pressure at the pain threshold: $p_{\text{pain}} = 200$ Pa.

Between the hearing and pain threshold, the sound pressure sweeps over a range of 7 powers of 10. Figure 9.1 shows the relationship between sound pressure and sound level.

Table 9.2 lists the relative sound intensities of some sound sources.

Table 9.1 Sound levels in acoustics (Source: Hering, Martin, Stohrer: Physik für Ingenieure [4])

Sound level	Definition	Reference value	Relations
Sound pressure level	$L_p = 20 \lg \frac{p_{\text{eff}}}{p_{\text{eff},0}} \text{ dB}$	$p_{\text{eff},0} = 2 \cdot 10^{-5} \text{ Pa}$	
Sound velocity level	$L_v = 20 \lg \frac{v_{\text{eff}}}{v_{\text{eff},0}} \text{ dB}$	$v_{\text{eff},0} = 5 \cdot 10^{-8} \frac{\text{m}}{\text{s}}$	$p_{\text{eff}} = Z v_{\text{eff}}$
Sound intensity level	$L_I = 10 \lg \frac{I}{I_0} \text{ dB}$	$I_0 = 10^{-12} \frac{\text{W}}{\text{m}^2}$	$I = \frac{p_{\text{eff}}^2}{Z}$
Sound power level	$L_w = 10 \lg \frac{P}{P_0} \text{ dB}$	$P_0 = 10^{-12} \text{ W}$	$P = S \frac{p_{\text{eff}}^2}{Z}$

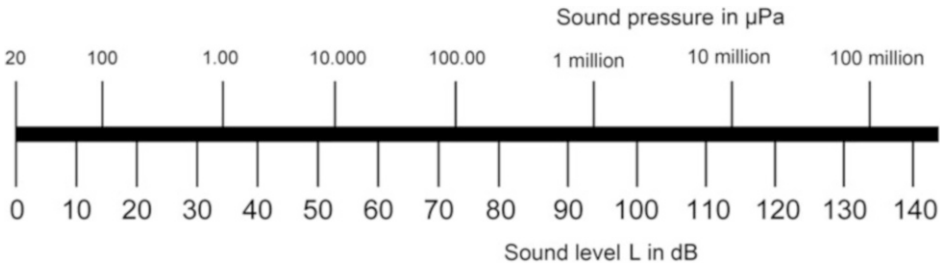


Fig. 9.1 Relationship between sound pressure and sound level

Table 9.2 Relative sound intensities and levels of some sound sources

Situations	I/I_0	dB	Description
	10^0	0	Hearing threshold
Normal breathing	10^1	10	Scarcely audible
Rustling leaves	10^2	20	
Soft whispering (5-m distance)	10^3	30	Very quiet
Library	10^4	40	
Quiet office	105	50	Quiet
Normal entertainment (1-m distance)	10^6	60	
Busy traffic	10^7	70	
Office noise from machines; factory average	10^8	80	
Heavy goods vehicle (15-m distance) waterfall	10^9	90	Permanent stress leads to hearing loss.
Old subway, concert	10^{10}	100	Damages
Pneumatic hammer	10^{11}	110	
Rock concert	10^{12}	120	
Air hammer, machine gun	10^{13}	130	Pain threshold
Take-off jet aircraft (immediate vicinity)	10^{15}	150	Deafness possible after 0.5 s; rupture of the eardrum
	10^{20}	200	

9.2.2 Volume

As mentioned above, the audible range is between 16 Hz and 20 kHz. The volume L_S is a measure of the ear's *perception of loudness*. The loudness has been determined so that at the sound frequency of 1000 Hz, the value of the loudness L_S is equal to the value of the sound pressure level L_P .

$$L_S(1.000\text{Hz}) \text{ in phon} = L_P(1.000\text{Hz}) \text{ in dB.}$$

The volume is measured in *phon*. The sound pressure reaching the ear is perceived as different loudness depending on the frequency. Figure 9.2 shows the curves of equal loudness L_S (in phon) for different frequencies f and sound pressure levels L_P .

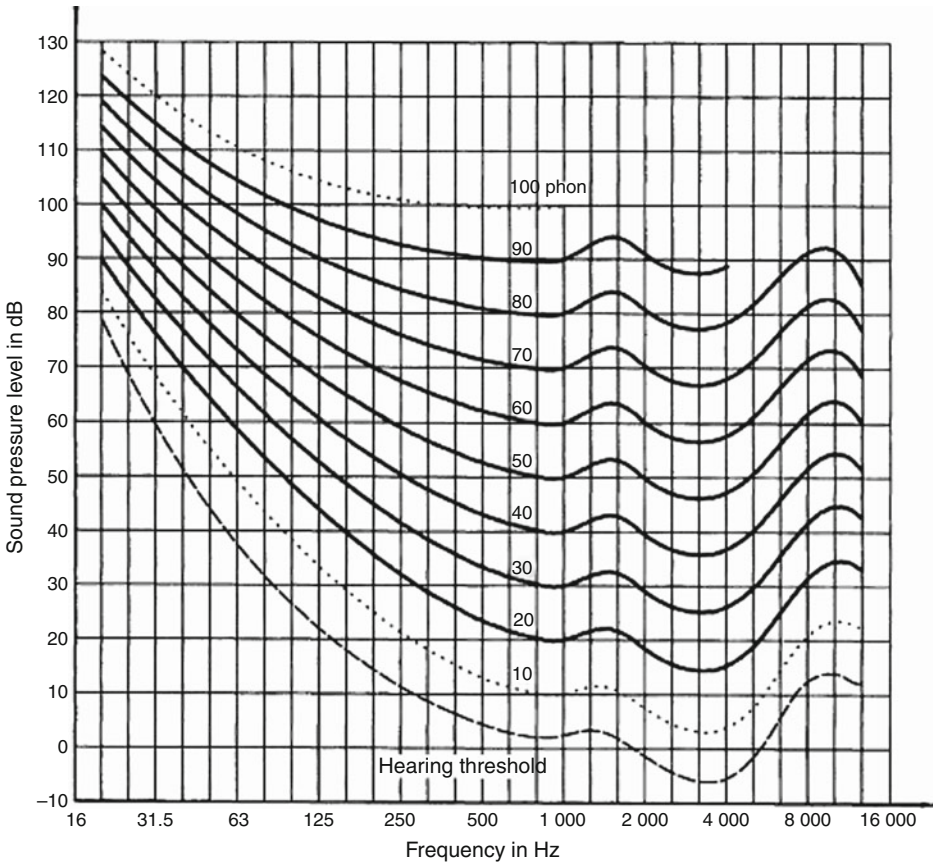
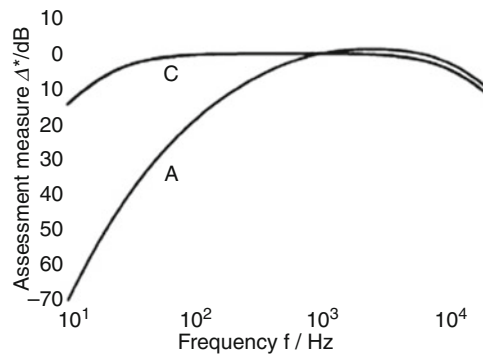


Fig. 9.2 Normal curves of equal volume levels for pure tones in the free sound field (Source: DIN ISO 226)

Fig. 9.3 Evaluation curves A and C according to DIN EN 61672-1



It can be seen from Fig. 9.2 that the greatest sensitivity of the ear is around 4 kHz. Sounds with a decreasing frequency are perceived as not being so loud. This correlation between sound pressure level and the sound perception of the human ear is shown in *evaluation curves* according to DIN EN 61672-1 (Fig. 9.3).

The A rating curve is the one most frequently used in practice. It corresponds to the sound perception for volumes below 90 phon. For low-frequency sound signals or very loud sound (above 90 phon), the weighting curve C is used. The weighted sound levels are then given as dB(A) and dB(C).

9.2.3 Loudness

The logarithmic relationships of the levels are not easy to understand in practice. Therefore, the *loudness* L in *sones* was introduced, which represents a linear relationship between volume and loudness (Fig. 9.4).

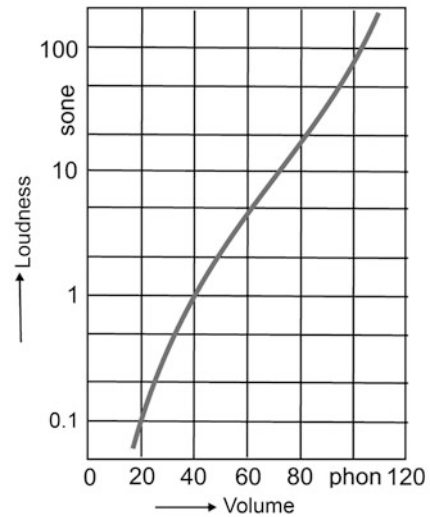
As Fig. 9.4 shows, a volume level of 40 phon corresponds to a loudness of 1 sone (DIN 1320). The dependence of the loudness L_{sone} on the loudness L_{phon} is

$$\log_{10} (L_{\text{sone}}) \approx 0,03010 (L_{\text{phon}} - 40).$$

From this equation it can be seen that an increase of the loudness level by 10 phon corresponds to a doubling of the loudness and a decrease by 10 phon corresponds to a halving of the loudness.

The above-mentioned correlation was experimentally confirmed in the range between 20 phon and 120 phon. Outside these limits, the relationships must be clarified experimentally and individually.

Fig. 9.4 Relationship between loudness in sone and loudness in phon (Source: Meyer, Neumann: *Physikalische und technische Akustik* [8])



9.3 Transducer

The sound pressure levels comprise more than 6 powers of 10. In this range, *sound transducers convert* the alternating pressure into *electrical voltages* via a mechanical oscillation system (*diaphragm*). Figure 9.5 shows the different principles of electroacoustic transducers, their technical designs and some application examples.

While *microphones* convert sound pressure into electrical voltage, *sound generators* or *loudspeakers* work in reverse. They convert electrical power into sound power.

The following converter principles according to Fig. 9.5 are implemented in the technology:

- *Electrostatic converter*: The diaphragm and a counter-cathode form a capacitor. As the diaphragm deflects, the capacitance changes and with it the electrical voltage.
- *Electrodynamic transducer*: The diaphragm moves a coil in a pot magnet. The movement induces electrical voltages in the coil.
- *Electromagnetic transducer*: The oscillation in the air gap of a magnet changes the magnetic flux in the magnet and induces an electric voltage in a winding.
- *Piezoelectric transducer*: The sound pressure generates a displacement of the charge carriers and thus a voltage proportional to the sound pressure.
- *Piezoresistive transducer*: The sound pressure changes the electrical resistance (Sect. 2.2) and thus causes a voltage change.

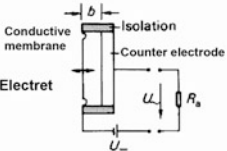
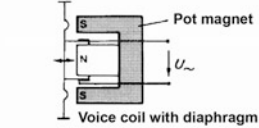
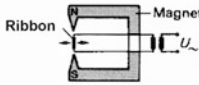
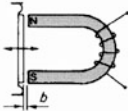
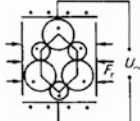
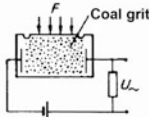
Electroacoustic transducers	technical designs	Application areas
<p>electrostatic</p> 	<p>Condenser microphone (with external polarization voltage on the microphone capsule)</p> <p>Electret microphone (permanent electrical polarization on the microphone membrane)</p> <p>Special loudspeakers</p>	<p>Sound level meter</p> <p>Studio microphone</p> <p>Clip-on microphone</p> <p>Low-frequency microphone</p> <p>Hand microphone</p> <p>Shoulder microphone</p> <p>extremely broadband headphones</p> <p>extremely broadband headphones</p>
<p>electrodynamic (voice coil)</p> 	<p>Moving coil microphone</p> <p>Speakers</p>	<p>Studio directional microphone</p> <p>Hand microphone</p> <p>Shoulder microphone</p> <p>Standard sound sources</p> <p>Acoustic irradiation system</p> <p>Headphones</p>
<p>electrodynamic (ribbon)</p> 	<p>Ribbon microphone</p>	<p>Studio microphone for highest volume levels</p> <p>Vocal Microphone</p> <p>Brass microphone</p>
<p>electromagnetic</p> 	<p>Speakers</p>	<p>Telephone receiver</p> <p>Hearing aids</p>
<p>piezoelectric</p> 	<p>Crystal microphone</p> <p>Ceramic microphone</p> <p>Piezopolymer Microphone</p>	<p>Structure-borne sound microphone</p> <p>Waterborne sound microphone</p> <p>Accelerometers</p>
<p>piezoresistive</p> 	<p>Carbon Microphone</p>	<p>Telephone</p>

Fig. 9.5 Principles of electroacoustic transducers (Source: Hering, Martin, Stohrer: Physik für Ingenieure [4])

Polar patterns of the microphones

The *directional characteristic* of a microphone indicates the preferred directions from which the sound is picked up. A distinction is made here:

- *Omnidirectional microphone*: The sound is picked up uniformly from all directions.
- *Club microphone*: The frontally incident sound is taken into account, while the side-ways incident sound is suppressed. This creates a pronounced *directional effect*.
- *Aft microphone*: Here sound pressure is received from two opposite directions.



Fig. 9.6 Types of measurement microphones (Factory photo: Brüel&Kjaer [1])

Fig. 9.7 Multi-field microphone for measuring engine noise (Factory photo: Brüel&Kjaer [1])



- *Cardioid microphone:* This microphone registers the sound pressure between the front and back of the diaphragm. The maximum sensitivity is in front of the microphone and decreases towards the sides.

Figure 9.6 shows designs of measuring microphones, especially multi-field microphones. These measure sound in unknown sound fields for each sound direction and in each sound field with high sensitivity (65 mV/Pa), in a frequency range from 5 Hz to 20 kHz and a dynamic range from 20 dB to 130 dB in a temperature range from $-20 \text{ }^\circ\text{C}$ to $+80 \text{ }^\circ\text{C}$.

Figure 9.7 shows the use of a multi-field microphone for measuring sound in automobiles.

9.4 Fields of Application

The acquisition of acoustic quantities plays an important role in technology. Some fields of application are described below:

Limit values for road noise [dB].				
Country	Key figure	Day border	Evening border	Night border
Australia	$L_{10, 18h}$	60		55
Austria	L_{Aeq}	50–55		40–45
Canada	L_{Aeq}	50		50
Denmark	$L_{Aeq, 24h}$	55	55	55
France	L_{Aeq}	60–65		55–57
Germany	L_r	50–55		40–45
Netherlands	L_{Aeq}	50	45	40
Spain	L_{Aeq}	60		50
Sweden	$L_{Aeq, 24h}$	55	55	55
Switzerland	L_r	55		45
Great Britain	L_{Aeq}	55		43

Fig. 9.8 Limit values for the noise level in dB(A) for road noise in different countries (Source: Bruel & Kjaer Sound and Vibration Measurement A/S [1])

Environmental noise

Environmental noise is unwanted sound events and vibration states. The classification differs from country to country and also depends on the respective culture. The following sound sources come into consideration:

- Car noise,
- Aircraft noise,
- Industrial noise,
- Transport noise,
- Construction noise and
- Recreational noise.

Figure 9.8 shows the limit values for the noise level in dB(A) in different countries.

Noise is often also mapped to identify particularly vulnerable areas (e.g. aircraft approach paths).

Industrial noise is essentially about making production facilities as quiet as possible and protecting the workforce from harmful noise (occupational safety). However, hazardous vibrations during production can also be investigated and eliminated accordingly. This can be done with a *modal analysis*.

Figure 9.9 shows the limit values of sound pressure levels for the noise protection zones in Germany.

Example for the application of noise protection zones						
Zone	Planning		Change		Alarm	
[dB]	Day border	Night border	Day border	Night border	Day border	Night border
Recovery area	50	40	55	45	65	60
Residential area	55	45	60	50	70	65
Mixed area	60	50	65	55	70	65
Industry area	65	55	70	60	75	70

Fig. 9.9 Limit values of levels for noise protection zones (Source: Bruel & Kjaer Sound and Vibration Measurement A/S [1])

Fig. 9.10 Dual-channel sound level meter (Source: Bruel & Kjaer Sound and Vibration Measurement A/S [1])



Acoustic material testing

The acoustic properties of materials play an increasingly important role. They are characterized in terms of absorption (absorption coefficient a), reflection (reflection coefficient r), acoustic impedance (Z), acoustic admittance (G) and transmission loss (TL). The values are specified in the relevant standards (e.g. ISO 10534-2).

Sound insulation

Sound insulation is important because the sound energy penetrates through walls and buildings or is transmitted as impact sound. For this purpose, the noise levels in two rooms are measured and compared. Figure 9.10 shows a two-channel sound level meter that can record and evaluate the sound events in the two rooms.

Room acoustics

In room acoustics, the structural conditions are examined, which are optimal for the sound events occurring in the room. Thus, room acoustics in concert halls, conference rooms, schools or studios are specially developed for radio and television. The *reverberation time*, *speech intelligibility* and *music quality* play a decisive role in this.

Sound Quality

Sound quality is the process of developing a product so that it has an optimal sound for the customer. This is done not only by physically measuring the sound parameters, but also by psychological tests at the customer's premises.

Bibliography

1. Brüel & Kjaer: Umweltlärm 2000
2. Gunther, B.; Hansen, K.H.; Veit, I.: Technische Akustik – Ausgewählte Kapitel, Grundlagen, aktuelle Probleme und Messtechnik. Expert-Verlag, 8. Auflage 2010
3. Henn, H.; Sinambari, G. R.; Fallen, M.: Ingenieurakustik: Physikalische Grundlagen und Anwendungsbeispiele, Vieweg+Teubner Verlag, Wiesbaden 2008
4. Hering, E.; Martin, R.; Stohrer, M.: Physik für Ingenieure. Springer Verlag, 13. Auflage 2021
5. Hering, E.; Modler, K.-H.: Grundwissen des Ingenieurs, Hanser Verlag, 14. Auflage 2007
6. Lerch, R.; Sessler, G.; Wolf, D.: Technische Akustik: Grundlagen und Anwendungen. Springer Verlag 2009
7. Maute, D.: Technische Akustik und Lärmschutz. Hanser Verlag 2006
8. Meyer, E.; Neumann, E.-G.: Physikalische und technische Akustik, Vieweg Verlag 1986
9. Möser, M.: Technische Akustik. Springer Verlag 2009
10. Veit, I.: Technische Akustik: Grundlagen der physikalischen, physiologischen und Elektroakustik. Vogel Verlag 2005



Climatic and Meteorological Measured Variables

10

Gert Schönfelder, Robert Krah, Gerd Stephan, and Roland Wernecke

10.1 Moisture in Gases

10.1.1 Definitions and Equations

10.1.1.1 Humidity

Humidity, or in short air humidity, describes the proportion of water vapour in the gas mixture of the earth's atmosphere or in rooms.

10.1.1.2 Water Vapour Partial Pressure

In a closed container, which is partly filled with water, water vapour forms in the air above the liquid. The water vapour exerts pressure on the walls of the vessel. This is referred to as the *water vapor partial pressure* p_d . The maximum attainable vapor concentration at a temperature T is referred to as the *water vapor saturation pressure*. The formation of water vapour is independent of the medium air. In general, the partial pressure of water vapor p_d is referred to as vapor pressure e . It applies:

G. Schönfelder (✉)
Prignitz Mikrosystemtechnik, Wittenberge, Germany

R. Krah
Krah&Grote Messtechnik, Otterfing, Germany
e-mail: info@krah-grote.com

G. Stephan
Quantum Hydrometrie Gesellschaft für Mess- und Systemtechnik, Berlin, Germany

R. Wernecke
Dr. Wernecke Feuchtemesstechnik GmbH, Potsdam, Germany
e-mail: info@dr-wernecke.de

$$p_d \equiv e = \rho_d R_i T$$

with R_i : individual gas constant ($R_i = 462 \text{ J (kg K)}^{-1}$)

ρ_d : density of water vapour.

The density of water vapour ρ_d is lower than that of dry air.

This maximum value is known as *saturation vapour pressure* or *equilibrium vapour pressure*. The concentration of water vapor at saturation depends solely on the temperature. The calculation of the saturation vapour pressure above a flat water surface E_w or above an ice surface E_e is usually done by empirical equations.

Magnus formula, named after Murray. The WMO (World Meteorological Organisation) states

$$E_w(T) = 6.1070 \exp (17.15T/T + 234.9) \quad (10.1)$$

$$E_e(T) = 6.1064 \exp (21.88T/T + 265.5) \quad (10.2)$$

with T in $^{\circ}\text{C}$ and E in hPa. Table 10.1 gives some numerical values for $E_w(T)$ and $E_e(T)$.

10.1.1.3 Absolute and Specific Humidity

The *absolute humidity* in $[\text{g/m}^3]$ indicates the *mass of water vapor* in a certain *air volume*. It is independent of the temperature. The *specific humidity* $[\text{g/kg}]$ is the *amount of moisture* in relation to dry air and is independent of temperature and air pressure.

From a meteorological point of view, an increase in absolute humidity can have various causes, such as evaporation via water surfaces (sea, lakes or river courses), the evaporation of water vapour by plants, by rainfall and the subsequent evaporation of surface water or, in general, mixing with drier or wetter air masses.

Indoors, human factors are added, such as breathing and perspiration of people, water ingress through damp clothing or moisture ingress through regular work (e.g. floor cleaning) or during renovation work (plastering or painting the walls).

10.1.1.4 Saturation Humidity

The *saturation humidity* is the *maximum amount of humidity* that the *air* can absorb at a certain *temperature*. Warm air can absorb significantly more moisture than colder air. Figure 10.1 shows that the saturation humidity increases with rising temperature. Below the curve, water vapor is distributed in the air; above it, the air is saturated with water vapor and water precipitates in liquid form.

10.1.1.5 Relative Humidity φ

Relative humidity describes the *ratio* of *absolute humidity* to *saturation humidity* at a certain temperature and air pressure. It indicates the degree to which the air is enriched with water.

Table 10.1 Saturation vapor pressure above pure water (infinitely extended, flat water surface) and ice

T in °C	100	50	40	30	20	10	0	-5	-10	-20
$E_w(T)$	1013.2	12.4	73.8	42.4	23.4	12.3	6.11	4.22	2.86	1.25
$E_e(T)$							6.11	4.02	2.60	1.03

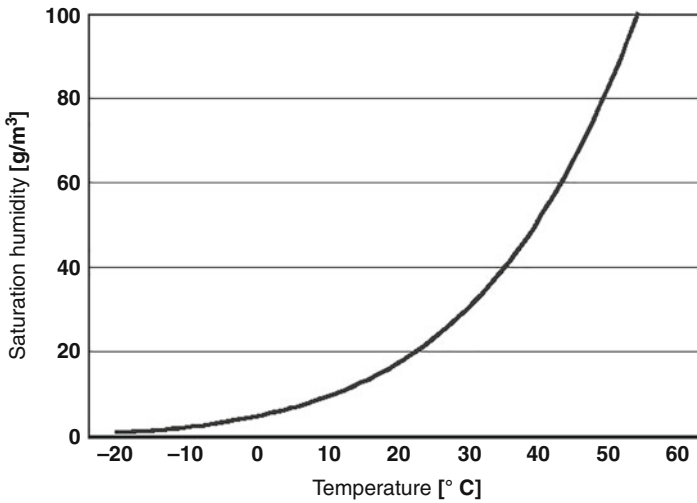


Fig. 10.1 Saturation humidity as a function of temperature

The relative humidity φ is defined as the ratio of water vapour partial pressure (p_d) to water vapour saturation pressure (p_s) at the same air pressure and temperature.

$$\varphi = p_d/p_s \cdot 100 [\%].$$

Air saturated with water vapor has a relative humidity of 100%. If the air is heated without changing the absolute humidity, the relative humidity decreases because the water vapour saturation pressure increases with temperature. When cooling a constant volume of air, the relative humidity increases to a maximum of 100%.

10.1.1.6 Dew Point

During further cooling, water vapor is emitted as *fog* or deposited on surfaces as condensation. The temperature at which water vapour is precipitated is called *dew point temperature* τ . It is the temperature at which the water vapour partial pressure e becomes the saturation vapour pressure E . It applies:

$$e(T) = E(\tau).$$

The dew point can thus be determined with the following formula:

$$\tau = 13.7 \ln (e/E(0))/1 - 0.058 \ln (e/E(0)).$$

An example: In spring, an unheated church is ventilated during the first rays of sunshine and thaw. The warm air from outside, which contains correspondingly more moisture, meets the cold surfaces of the building inside. In the still cool building, the relative humidity rises. At points where the surface temperature of the room envelope is below the dew point temperature of the room air, moisture precipitation (condensation) occurs.

10.1.1.7 Enthalpy

The enthalpy indicates the energy content of a gas mixture in J/kg.

10.1.1.8 Mollier Diagram (h-x Diagram)

The Mollier diagram (*h-x diagram* Fig. 10.2) graphically describes the relationships between temperature, absolute humidity, saturation humidity, enthalpy and relative humidity.

h: specific enthalpy (kJ/kg) is a measure for the energy in a thermodynamic system, related to the amount of substance or mass.

x: Water content (g/kg) indicates the amount of water contained in a given volume or mass of air/substance (see absolute humidity).

In Fig. 10.3, the axes and sizes used are shown again. At the respective points of intersection, the corresponding parameters for humid air can be determined.

Example: The following measured values were determined:

Relative humidity: 50%

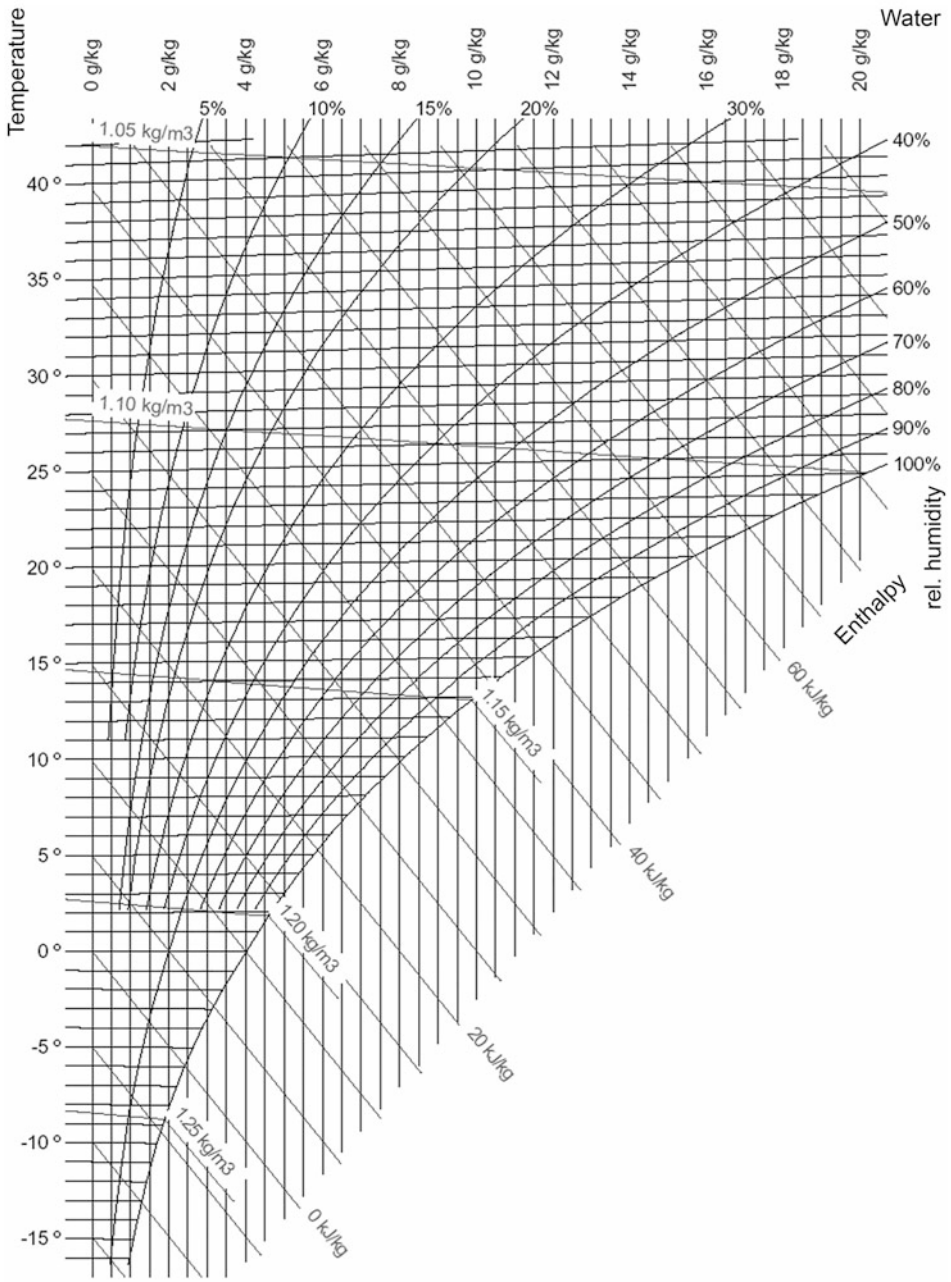
Temperature: 25 °C

Characteristics that can be derived from this:

Absolute humidity: 10,5 g/kg

Enthalpy: 52.5 kJ/kg

Density: 1,11 kg/m³



Mollier h-x diagram for humid air - pressure 0.950 bar (537.000 m / 10.000 °C / 80.000 % r.h.)

Fig. 10.2 Mollier diagram for humid air

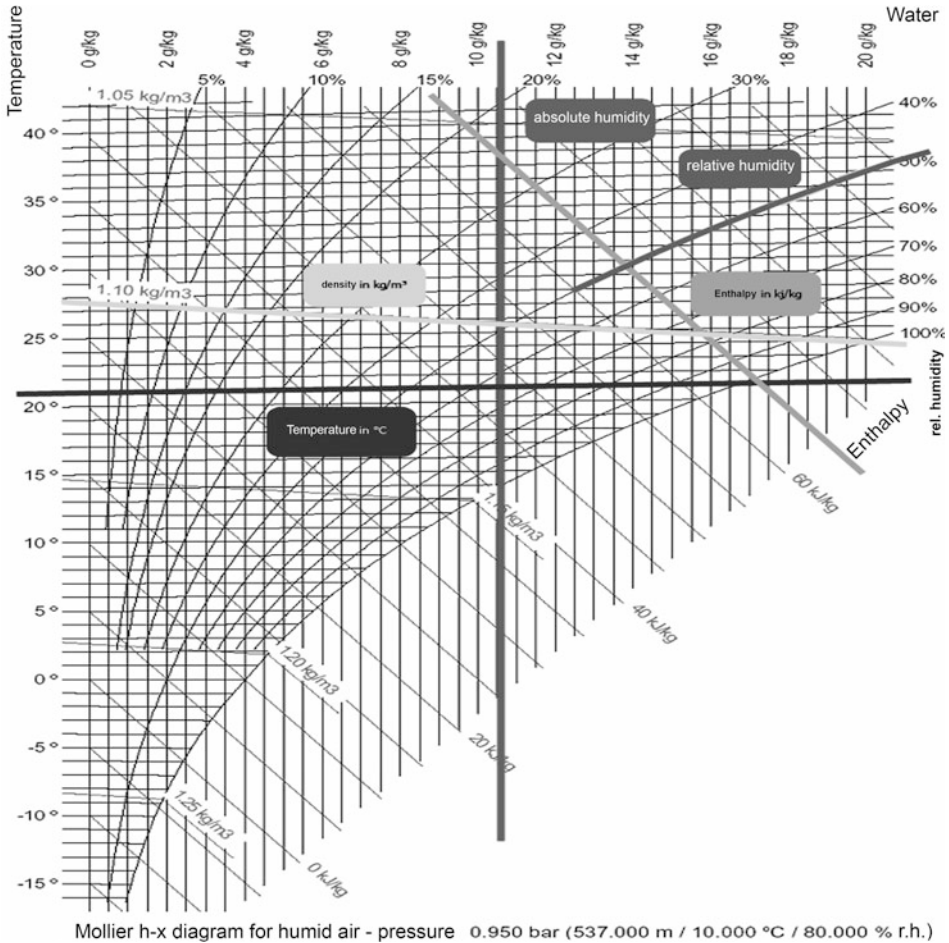


Fig. 10.3 Determination of characteristic values from the Mollier diagram

10.1.2 Humidity Measurements in Gases

10.1.2.1 Psychrometers, Design and Function

The term “psychrometer“comes from the Greek and means cool, cold. The *psychrometer* is a meteorological measuring instrument for determining the *humidity of the air*. It uses the physical effect of *evaporative cooling*. A psychrometer consists of *two thermometers*, one of which is wrapped in a damp fabric stocking (Fig. 10.4). The drier the air, the faster the water evaporates and the more heat is extracted from one thermometer. This means it cools down. The *greater* the *temperature difference* between the two thermometers, the *lower* the *relative humidity*.

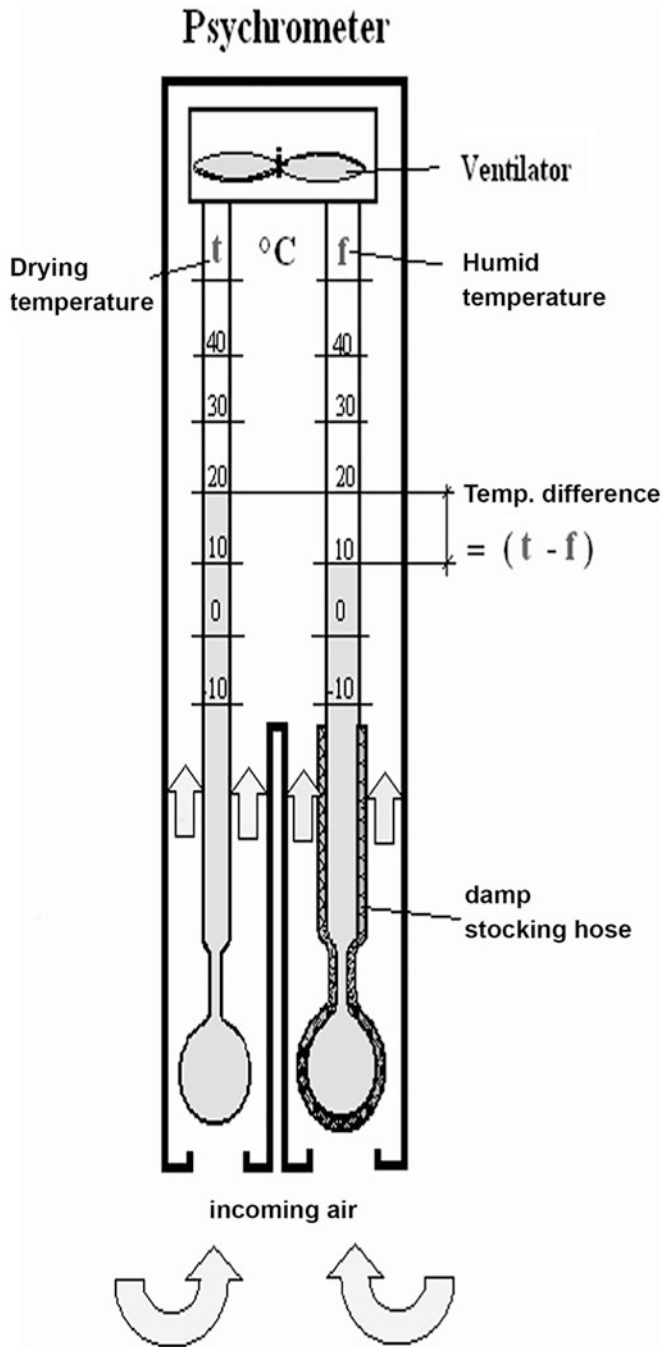


Fig. 10.4 Structure of a psychrometer

From the temperature difference, the relative humidity can be determined with the help of psychrometer tables. The *psychrometric measuring principle* is one of the most accurate and is therefore used in weather stations where accurate measurements are required or as a reference instrument. Assmann psychrometers achieve a measuring accuracy of $\pm 0.5\%$.

- *Drying temperature*: θ_L corresponds to the air temperature.
- *Humid temperature*: ϑ_L is the temperature the air parcel would have if it were adiabatically cooled to saturation at constant pressure by evaporation of water into the parcel, thereby removing the latent heat required from the parcel.

From the observed temperature difference in the psychrometer $\vartheta_L - \vartheta_f$ the vapour pressure e of the air can be calculated in good approximation according to the following psychrometer formula:

$$e = E_f - \gamma \cdot (\vartheta_L - \vartheta_f)$$

with E_f Saturation vapour pressure at the temperature of the wet surface in hPa

γ : Psychrometer constant ($\gamma = 0.67 \text{ hPaK}^{-1}$ in heights up to 500 m can be calculated with this value in a simplified way).

$\gamma_{\text{Wasser}} = 0.653 \cdot 10^4 \cdot (1 + 0.000944 t) \text{ pK}^{-1}$ with humid thermometer

$\gamma_{\text{Eis}} = 0.575 \cdot 10^4 \cdot \text{pK}^{-1}$ with frozen humid thermometer.

10.1.2.2 Design Types and Areas of Application

There are two common models on the market: the aspiration psychrometer and the *slingshot psychrometer*.

10.1.2.3 Aspiration Psychrometer

The aspiration psychrometer consists of two identical thermometers, which are protected from radiant heat by a metal tube. On one thermometer, there is a fabric stocking that is moistened with distilled water. With the help of a fan, an even air flow is led past both thermometers (Fig. 10.5).

It must be ensured that the flow velocity is greater than 2 m/s in order to avoid congestion. The relative humidity can be calculated from the temperature difference between the dry and wet bulb thermometer according to the above formula. Determination according to the psychrometric *table* (Table 10.2) is also common.

Due to their high reliability and accuracy, aspiration psychrometers are used as reference instruments for humidity measurements in meteorology, heating and air-conditioning systems and in calibration laboratories.

10.1.2.4 Slingshot Psychrometer

For mobile use, the *slingshot psychrometer* is used (Fig. 10.6). Here the necessary air flow is achieved by spinning the two thermometers. The rotating measuring part, which consists

Fig. 10.5 Aspiration psychrometer (Factory photo: Mesadan)



of two thermometers, is attached to the upper end of the handle. The thermometers have an accuracy of 0.1 K. One is kept moist by a fabric soaked with distilled water. The air flowing through it cools down on this fabric. The second thermometer serves as a dry thermometer. The temperature difference determined by the thermometer is a measure of the water vapour content of the air. The resulting vapour pressure is determined by the above formula.

10.1.2.5 Dew Point Mirror

Functionality

By cooling a mirror below the dew point temperature (Fig. 10.7), the exact humidity can be determined via the temperature at which condensation occurs (saturation vapour pressure). An optical system detects the condensation on the mirror. A Pt100 temperature sensor records the exact temperature within a range of ± 0.1 °C accuracy on the mirror (Fig. 10.17). Pt100 temperature sensors are based on the resistance change of a Pt resistor when the temperature changes. Pt100 means the resistance of 100 Ω at 0 °C. The dew point mirror is very well suited for measuring the absolute humidity in gases.

Construction of the Dew Point Mirror

A *Peltier element* is attached to the back of a small mirror. A Peltier element uses the Peltier effect, named after Jean Peltier, according to which a temperature difference occurs when different metals are connected to each other when current flows (also known as the thermoelectric effect, Sect. 2.10). By reversing the direction of the current, the heat flow can also be turned. Figure 10.8 shows the mode of operation of a Peltier element and Fig. 10.9 a dew point mirror.

Table 10.2 Psychrometric table

		Psychrometric difference [K]																					
		0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	6	7	8	9	10	11	12	13	14	15	16	18
T	2	92	83	75	67	59	52	43	36	27	20												
R	4	93	85	77	70	63	56	48	41	34	28	15											
O	6	94	87	80	73	66	60	54	47	41	35	23	11										
C	8	94	87	81	74	68	62	56	50	45	39	28	17										
K	10	94	88	82	76	71	65	60	54	49	44	34	23	14									
E	12	94	89	84	79	73	68	63	59	53	48	38	30	21	12	4							
N	14	95	90	84	79	74	69	65	60	55	51	41	33	24	16	10							
T	16	95	90	85	81	76	71	67	62	58	54	45	37	29	21	14	7						
E	18	95	90	86	82	78	73	69	65	61	57	49	42	35	28	20	13	6					
M	20	96	91	87	82	78	74	80	66	62	58	51	44	36	30	23	17	11					
P	22	96	92	87	83	79	75	72	68	64	60	53	46	40	34	27	21	16	11				
E	24	96	92	88	85	81	77	74	70	66	63	56	49	43	37	31	26	21	14	10			
R	26	96	92	89	85	81	77	74	71	67	64	57	51	45	39	34	28	23	18	13			
A	28	96	92	89	85	82	78	75	72	68	65	59	53	47	42	37	31	26	21	17	13		
T	30	96	93	89	86	82	79	76	73	70	67	61	55	50	44	39	35	30	24	20	16	12	
U	32	96	93	90	86	83	80	77	74	71	68	62	56	51	46	41	36	32	27	23	19	15	
R	34	97	93	90	87	84	81	77	74	71	69	63	58	53	48	43	38	34	30	26	22	18	10
	36	97	93	90	87	84	81	78	75	72	70	64	59	54	50	45	41	36	32	28	24	21	13
[° C]	38	97	94	90	87	84	81	79	76	73	70	65	60	56	51	46	42	38	34	30	26	23	16
	40	97	94	91	88	85	82	79	76	74	71	66	61	57	52	48	44	40	36	32	29	25	19



Fig. 10.6 Slinging psychrometer

Fig. 10.7 Functional diagram of dew point mirror

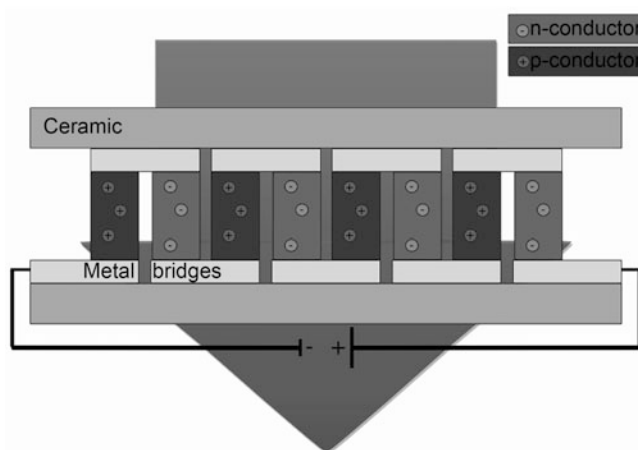
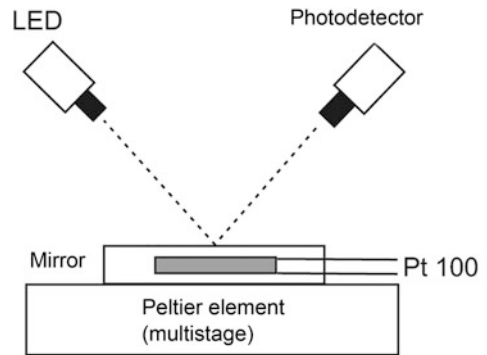


Fig. 10.8 Functional diagram of a Peltier element

Fig. 10.9 Dew point mirror
(Krah & Grote factory photo)



Hair Hygrometer

The mechanical method is based on the expansion or contraction of various (mostly organic) measuring elements. Such measuring elements are for example: hair, durometer or gut strings.

The most commonly used measuring elements are hair elements or the so-called *Durotherm*, an artificial, moisture-sensitive measuring element. The change in length of the measuring element is then transferred to the pointer via a mechanical movement.

Hair hygrometers (Fig. 10.10) require regular maintenance and care. To avoid drying out and the associated drift, hair hygrometers must be regenerated regularly. For this purpose, the hair harp is wrapped with a cloth moistened with distilled water or sprayed with distilled water so that saturation occurs. After about one hour, a measured value of about 98% r.h. is obtained. On most instruments, a single-point adjustment can be carried out via an adjusting screw.

Figure 10.11 shows a thermohygrometer as it is usually used for measuring humidity.

10.1.2.6 Capacitive Humidity Measurement

The capacitive humidity sensors work on the principle of a plate capacitor, in which a very thin humidity-sensitive polymer layer is used as the dielectric, whose permittivity number ϵ_r changes depending on the moisture absorbed (Sect. 2.6). In an oscillating circuit, this change in capacitance causes a resulting change in frequency when the humidity changes.

Figure 10.12 shows the structure of a capacitive *humidity sensor*. A porous dielectric, which is hygroscopic, is applied to a conductive substrate, which forms one electrode of the plate capacitor. The second permeable electrode is located on this. The change in capacitance is measured by a suitable measuring circuit and converted to relative humidity. Various combinations of materials have proven to be effective, for example an aluminum carrier on which aluminum oxide is applied as dielectric with a gold electrode. In aggressive environments, tantalum-coated glass plates with a hygroscopic polymer as dielectric and a porous chromium layer as counter-electrode are used.

A regular check is also necessary for capacitive humidity sensors, as they too are subject to ageing and their properties can change over time. It is recommended to check the sensors

Fig. 10.10 Hair hygrometer



Fig. 10.11 Thermohygrometer

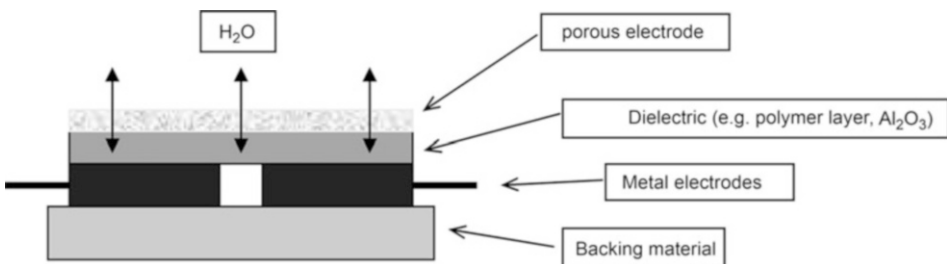
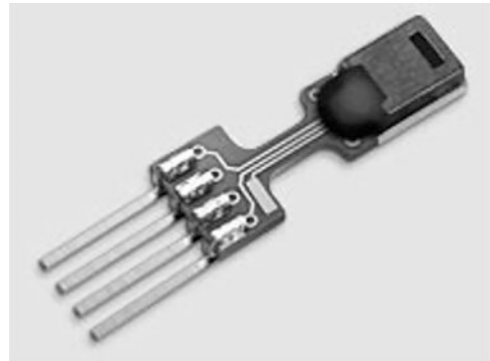


Fig. 10.12 Diagram of a capacitive humidity sensor

Table 10.3 Orientation values for checking capacitive humidity sensors

Category	Tolerance	Audit	Exchange
Inexpensive	+/- 5% r.h.	3 months	1 year
Mid-price segment	+/- 3% r.h.	1 year	5 years
High-quality sensors	+/- 2% r.h.	5 years	10 years

Fig. 10.13 Integrated capacitive humidity sensor (Factory photo: Sensirion)

after a certain period of time, depending on the type, and then readjust or replace them. Table 10.3 lists these periods.

10.1.2.7 Integrated Capacitive Humidity Sensors with Bus Output

These sensors consist of a capacitive humidity sensor, the analog part, including an oscillating circuit and subsequent signal processing using an integrated microcontroller (Fig. 10.13). They are characterized by laser calibration, long service life and high accuracy.

Table 10.4 lists the application areas of the sensors.

10.2 Moisture Analysis in Solid and Liquid Substances

In *gas humidity*, water is regarded as water vapour exclusively in its gaseous state of aggregation. The *material moisture* (or moisture in liquids) includes the liquid, solid and gaseous states of water. In production processes (e.g. drying, moistening), the interaction of the different states is used to adjust product properties in a targeted manner. Accordingly, the laws of thermodynamics (gas humidity), materials science (surface specifics, material composition, capillary properties) and chemistry (binding of water, reaction with other substances) apply to the description of material humidity. For the representation of water in solids or liquids, the following quantities must be included in the consideration:

Table 10.4 Comparison of the individual sensor types and their areas of application

Field of application	Examples	Sensor type
Paper and printing industry	Production, processing and storage (e.g. museums)	Demanding and expensive capacitive sensors
Chemistry	Monitoring (e.g. hazardous substances)	Sophisticated and expensive dew point sensors, dew point mirrors
Automotive industry		Low-cost capacitive sensors
Food industry	Production and storage (e.g. cheese dairies)	Demanding and expensive dew point capacitive sensors, dew point mirrors
Agriculture	Stables, drying, irrigation control	Capacitive sensors
Building materials industry	kilns (e.g. ceramic, bricks)	Capacitive sensors,
Components	Fuel cells, high-voltage switches	Demanding and expensive dewable capacitive sensors
Home appliances	Washing machine, dryer, humidifier	Low-cost dewable capacitive sensors
Clothing industry	Textile and leather industry: manufacturing, processing and storage	Demanding and expensive dewable capacitive sensors
Humidity generators	E.g. snow cannons, climate cabinets	Demanding and expensive capacitive sensors
Climate monitoring	Weather stations	Sophisticated and expensive dew point capacitive sensors, dew point mirrors, aspiration psychrometers
Air-conditioning in buildings and cars	Museums, offices	Demanding and expensive capacitive sensors
Medicine	Operating room, anaesthesia	Sophisticated and expensive dew point sensors, dew point mirrors
Pharmacy	Production and storage of pharmaceuticals (e.g. tablets)	Sophisticated and expensive dew point capacitive sensors, dew point mirrors, aspiration psychrometers

- Material composition,
- Stock consistency (bulk density, porosity),
- Temperature,
- Substance concentrations and
- Thermodynamic quantities (water vapour pressure).

Liquid substances can contain water in various forms:

- in small droplets or as steam (free water),
- bound to solid particles in the liquid/solid/water mixture (dispersed) or as

- chemically bound water.

Analogous to the occurrence of water in solids, temperature and pressure act as parameters in relation to the gas exchange with the environment and the formation of chemical compounds of the water with the carrier liquid. Due to the high mobility of the individual components, *gravity has a stronger effect on the arrangement of the constituents of the substance in liquids than in the case of water in solid mixtures.* In resting liquids, therefore, an inhomogeneous distribution of the water constituents can usually be observed even at lower concentrations.

10.2.1 Direct Methods for the Determination of Material Moisture

The *direct measuring methods* determine the water content of a material mixture by separating the contained water and the dry substance. The result is a direct measured value of the percentage water content independent of the material properties. The direct measuring methods can hardly be used for online measurements in the process. However, they are very important as *reference methods*, for the determination of *material-specific* characteristics and for the *traceability* of measured values.

10.2.1.1 Percentage Water Content of a Material Sample

The *gravimetric water content* WG_m is the percentage mass of water contained in a solid or liquid material. It applies:

$$WG_m = \frac{m_w}{m_{tot}} \cdot 100,$$

m_w : mass of water and m_{tot} : total mass of the mixture.

Unit of measurement: % w/w.; g/kg; ppm (part per million; $1:10^6$ particles); ppb (parts per billion; $1:10^{-9}$ particles)

Thermogravimetric measurement is the classical method for determining the gravimetric water content. Depending on the task at hand, different names have been developed for this measuring method (e.g. *kiln weigher*, dry weighing method, ATRO). The values determined are also interpreted differently.

The *volumetric water content* WG_V is the percentage by volume of water in the total material volume. The following applies:

$$WG_V = \frac{V_w}{V_{tot}} \cdot 100$$

V_w : volume of water and V_{tot} : dry volume of solids.

Unit of measurement: %vol.; ml/l.

The typical method for measuring the volumetric water content is *Karl Fischer titration*.

The conversion of the mass-related humidity into a volume-related humidity and vice versa is done via the density ρ . This expresses the ratio between mass m and volume V of a material ($\rho = m/V$).

Technical Design for the Determination of the Gravimetric Water Content

The measuring method is based on the separation of water from the dry substance of a material sample. In the thermogravimetric measuring method, the total mass (m_w) of the sample is first determined by weighing. The sample is then heated until all water evaporates and the mass is constant. The sample is then weighed again. The gravimetric proportion of water in the sample is determined from the difference between the two masses according to the equation given. The measuring instrument (Fig. 10.14) is also known as a *drying oven* or *material moisture analyzer* and consists of the following components.

- Balance (analytical balance),
- Heat source (for temperatures $T > 100$ °C) and
- Closed measuring room.

Different radiators are used as heat sources (infrared radiators, microwave generator, halogen radiators). Depending on the design of the instruments, the water content of samples up to 100%w/w or partial ranges 0 to 10%w/w can be determined.

10.2.1.2 Water Activity of a Material Sample

Especially for hygroscopic materials in the food and pharmaceutical industries, the so-called *equilibrium moisture* or the *water activity* (aw-value) of a sample is determined. These quantities are dimensionless or are given as a percentage of relative humidity (% r.h.).

Equilibrium moisture content is understood to be the state in which there is *no* longer any *exchange of moisture* between the material and the ambient air humidity. The water vapour pressures in the air and in the solid material are balanced. The equilibrium moisture content refers to a static state. It applies:

- water activity $aw(t) = \text{relative humidity}/100$ and
- Equilibrium humidity $GF(t) = \text{relative humidity}$.

This relationship presupposes that the ambient climate has balanced itself over time (t) and that the temperature is constant.

10.2.1.3 Karl Fischer Titration

Karl Fischer titration is a *chemical method* for determining the volumetric proportion of water in liquids and solutions. It is based on the separation of water and the dry substance and is mainly used to determine *low water concentrations* (trace moisture). It is a laboratory method. The titration is carried out in 2 steps. If solids are to be analyzed, the sample

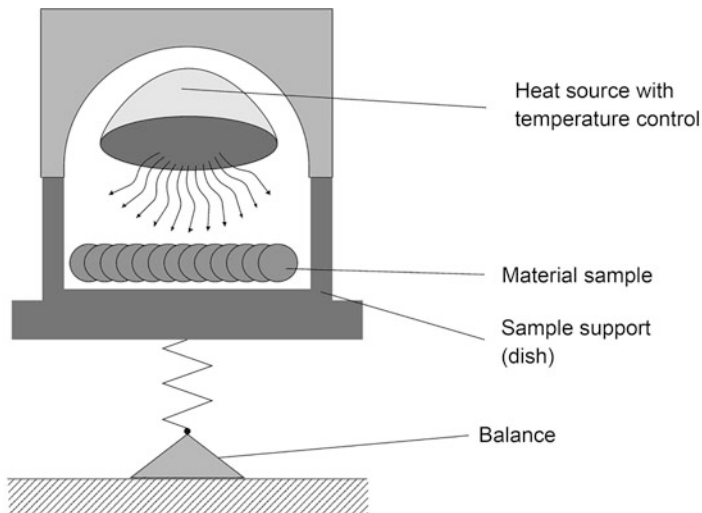
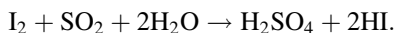


Fig. 10.14 Measuring principle of the thermogravimetric method

must first be prepared in such a way that the water is bound in a solution. Then the titration is carried out, in which the following chemical reaction takes place:



The starting materials iodine (I_2) and sulphur dioxide (SO_2) are mixed in defined quantities.

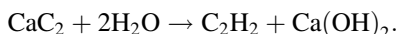
The mixture takes on a brown colour due to the iodine present. The titrant is then added to the sample to be determined using a burette. The iodine is bound by the water content of the sample and the mixture decolorizes. From the *ratio* of the added components, it can be calculated what *volumetric water content* was present in the sample. A second method is the *coulometric principle*, in which the decomposition of the water is not measured by color change but *electrically*.

Quick test kits are available as technical versions for titration. These contain the necessary chemicals and auxiliaries for rapid water determination (e.g. in the environment or in the production process).

Automatic titrators are used for analyses in the laboratory. The measurement is highly automated. However, the procedure for preparing the sample can be very complex and is reserved for the operator. Depending on the instrument version, the measuring range is between 0 and 20% vol. water content. Small material samples, which are usually additionally heated, are used for the measurement. A very precise determination in the trace moisture range is possible with this measuring method.

10.2.1.4 Calcium Carbide Method

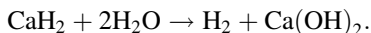
While Karl Fischer titration determines the water in a liquid solution, the calcium carbide method allows *direct determination of the water content* in powders and granulates. A defined mass of the sample is placed in a pressure-tight container and mixed with calcium carbide (CaC_2). Acetylene (C_2H_2) is formed as a reaction product.



Depending on the amount of acetylene produced, an overpressure forms in the container. This *overpressure* is measured manometrically; it is a direct measure of the amount of reacted water in the sample. Conversion tables and scales are used to convert to the gravimetric water content. This measuring method is mainly used in the construction industry to measure the moisture content in masonry, screed or concrete, for example. Mobile measuring cases allow on-site use on the construction site.

10.2.1.5 Calcium Hydride Method

In principle, the measuring principle corresponds to the calcium carbide method only with other reaction substances. The following chemical reaction takes place:



A vacuum is first generated in a measuring vessel and heat is supplied. The overpressure created during the reaction is directly related to the water content of the sample. The analytical instruments are mostly used in the laboratory for the determination of trace moisture in plastics. The typical measuring range is from 0 to 5%w/w water content.

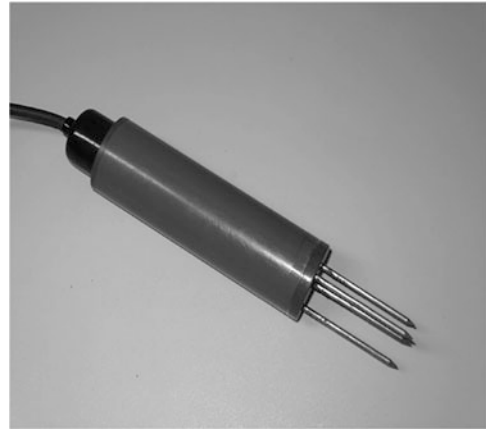
10.2.2 Indirect Measuring Methods for Determining the Moisture Content of Materials

These measuring methods are characterized by the fact that they can detect the moisture online in the process and allow direct control of the water content. Individual specific properties of water are recorded (e.g. permittivity number ϵ_r , optical absorption, thermal conductivity). From this, a signal is generated, which can be assigned to the water content of the material mixture. Indirect measuring methods require the experimental determination of a material-specific characteristic curve in order to make a statement about the water content.

10.2.2.1 Measurement of the Electrical Properties

The electrical properties of water are very often exploited to determine the moisture content of mixtures of substances. This indirect measuring method records the complex electrical resistance of the measured material. Continuous measurements in liquid and solid materials

Fig. 10.15 Puncture sensor for soils, bulk materials and granulates



are possible even with high material flows. For an accurate measurement, it is necessary that the material to be measured is as homogeneous as possible and does not contain any metallic or other well-conducting components.

Electromagnetic fields or electromagnetic waves serve as the measurement signal (Fig. 10.15). A differentiation of the measuring methods refers to

- the measurement frequency and signal power used,
- the timing (e.g. the pulse width),
- the type of evaluation of the measurement signal and
- the way the measuring signal is coupled into the material mixture.

Table 10.5 shows the measuring principles that can be used for moisture analysis using electromagnetic fields.

When measuring with *electromagnetic* fields, the *volumetric* water content is determined. The measuring volume is determined by the design of the measuring probe and the penetration depth of the measuring signal. Common to the different measuring methods is that the permittivity number of water ($\epsilon_r = 80.18$) differs significantly from that of other substances ($\epsilon_r \sim 5$ for most substances processed in the process).

10.2.2.2 Detection of the Optical Properties of Water and Water Vapour

In the entire spectral range, there are different wavelengths that are particularly selective for water vapor or water. The selection of a suitable wavelength for a specific measuring task depends on the following criteria:

- the level of selectivity for water in terms of reflection, absorption or transmission, and
- the other components occurring in the material to be measured.

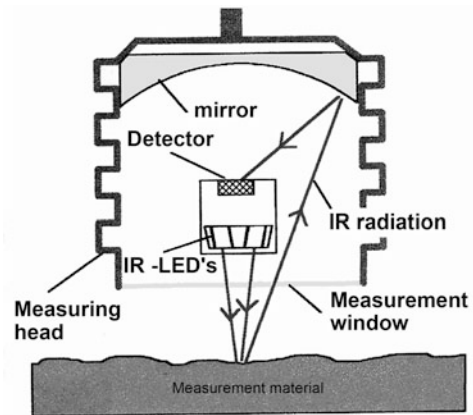
Measurement is possible in gases, liquids and solids. The *optical properties* of water or water vapour in conjunction with the material to be measured are recorded as the measuring

Table 10.5 Different principles of moisture analysis using electromagnetic fields or waves

Designation of the measuring method	Measuring signal/ measuring principle	Technical execution	Special feature/example
Low frequency	Detection of mainly ohmic resistance Effective resistance	Measuring frequency $f \leq 10$ kHz Handheld measuring instruments Fixed installations	Simple hand-held measuring devices, devices calibrated for specific materials
Radio frequency	Detection of effective resistance and reactance	Measuring frequency up to 100 kHz	Bulk materials, building materials, handheld measuring instruments
FDR Frequency Domain Reflectometry	Vibration damping	Measuring frequency ~ 100 MHz	Soil moisture, bulk materials
TDR Time Domain Reflectometry	Runtime difference of the oscillation	Measuring frequency ~ 1 GHz pulsed measurement	Soil moisture, bulk solids, handheld probe, process installation probe
Microwave	Frequency shift attenuation	Measuring frequency 2.5 MHz to 20 MHz	Permanently installed built-in devices
Radar Radio detecting and ranging	Runtime measurement		Hydrology, geology, permanent structural monitoring

effect (Fig. 10.16). The measurement is contactless and can be performed from a greater distance and through optically permeable materials (e.g. a quartz window). An advantage of the optical measuring method is that the measurements can be carried out without delay even under *extreme conditions* (e.g. high pressure, high temperature). The physical basis for the measurement of humidity is *Bouger–Lambert–Beer’s law*. This optical law describes the relationship between absorption and transmission when an optical beam passes through a material. The *absorption of light* $I(0)$ by the material is measured. The light passes through the material to be measured along the distance x and emerges with the intensity $I(x)$. The optical measurement of moisture is carried out in the entire frequency spectrum of light (ultraviolet, visible, near infrared, infrared range). Individual frequencies, which are particularly sensitive to water, are used in combination in the measuring instruments. The use of tunnel laser diodes makes it possible to use a very *narrow-band* measurement signal, which significantly increases the measurement accuracy. Solid materials (e.g. powder, bulk materials) are usually opaque. A measurement of moisture is only carried out on the *surface* of these materials. The moisture inside the material cannot be measured. In order to compensate for surface properties that distort the measured values (e.g. colour, roughness), several wavelengths are used for measurement. Even slight

Fig. 10.16 Diagram of optical humidity measurement



changes in the material composition can lead to measurement errors, which requires a recalibration of the system. An exact match between the measuring system and the material to be measured as well as keeping the material properties constant is a mandatory prerequisite for achieving low measuring errors.

10.2.2.3 Measurement of the Suction Pressure in Moist Materials (Tensiometry)

Water has the property of diffusing from solid material with a high water concentration into material with a low water concentration according to *Braun's molecular movement*. The resulting concentration gradient can be recorded as a pressure difference. The *tensiometric* measurement is based on the determination of this *water pressure difference*. Water is filled into an air-empty sealed reservoir. The sensitive surface is a fine porous ceramic (Fig. 10.17). If this sensor is inserted into the moist material to be measured (e.g. powder, soil, fiber), a negative pressure builds up on the porous ceramic (suction pressure). This negative pressure depends on the water content and the material properties of the measured material. If the sample is completely saturated with liquid water, the pressure difference is $\Delta p = 0$ Pa. If the material is very dry the negative pressure becomes too high and water exits the tensiometer via the ceramic and air enters. The tensiometer is no longer ready to measure; it must be refilled with water.

In practice, tensiometers are mainly used in agriculture, hydrology and geology for long-term measurements.

10.2.2.4 Measurement of Atomic Properties

In practice, *nuclear measurement methods* are used less and less frequently because the technical equipment and personnel costs can become very high (e.g. by radiation officers, approval procedures or radiation-proof containers). There are also considerable psychological reservations about the use of this technology.

Fig. 10.17 Tensiometer for soil moisture measurement with digital display



The various nuclear beams:

- Alpha radiation,
- Beta radiation,
- Gamma radiation and
- Neutron radiation

can be used for material moisture measurement. However, *gamma radiation* and *neutron radiation* are of technical importance. The theoretical basis of these measuring methods is the structure of the atoms. When measuring with neutron radiation, high-energy (“fast”) neutrons from a radioactive radiation source penetrate the material to be measured. The hydrogen atoms contained in the sample “slow down” these fast neutrons. According to the atomic theory, 19 collisions between neutrons and the hydrogen atoms are necessary for the neutrons to be completely slowed down and to release their energy in the form of heat. Much more collisions with other atoms are required for this. Hydrogen or the *water* contained in the sample is crucial for the attenuation of the neutron input signal. The number of braked neutrons is measured with a *scintillation detector*. A direct linear relationship can be established between the attenuation of the current and the water contained in the sample.

In the gamma ray method, radiation penetrates the material to be measured. Depending on the amount of hydrogen atoms, the measuring signal is attenuated. The remaining radiation is detected by a scintillation detector and is directly related to the water content of the material to be measured.

10.2.2.5 Nuclear Magnetic Resonance (NMR) Method

The method can be used to give a *spatial image* of the moisture distribution in a solid material. The measuring principle is based on the theory of the structure of atoms. The protons (positively charged particles) in a hydrogen atom have an angular momentum

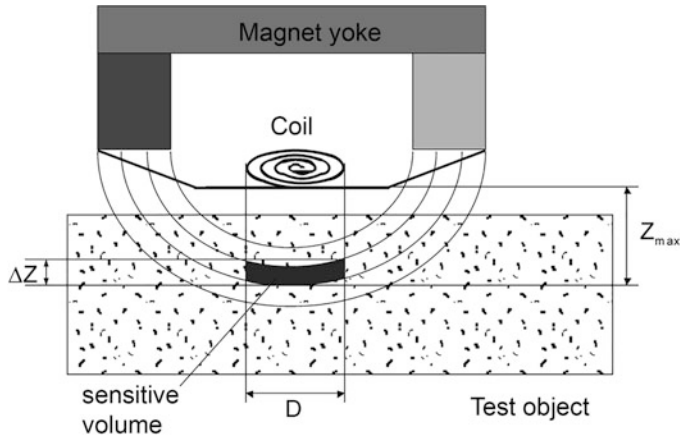


Fig. 10.18 Technical design of a moisture analyzer using the NMR method

(spin) and their own nuclear magnetic moment. If an external constant magnetic field is applied, the protons align themselves according to their magnetic polarity in this field. Due to the structure of the hydrogen nuclei, the protons perform a gyroscopic movement (*precession*) in the direction of the magnetic field. The alignment is parallel or antiparallel to the external magnetic field. If an alternating magnetic field with a certain frequency and pulse width is applied perpendicular to this field, which enters into resonance with the proton motion, the spin motion is amplified. By selective switching off (pulsing), an induction current can be induced in a coil by the rotary motion of all hydrogen nuclei and water molecules contained in the sample. This current is directly proportional to the volumetric water content of the sample. Figure 10.18 shows the arrangement. The section shown in black represents the measuring volume. The spatial distribution of the water is determined by moving the test object horizontally and vertically.

If the sample is moved past the measuring device (Fig. 10.18), a scan of the water content can be carried out and the spatial distribution of the water in the material becomes visible. The measuring method is mainly used for laboratory measurements on building materials and bulk materials. Low water contents are recorded; in principle, values up to 100%w/w can also be measured.

10.2.2.6 Measurement of Thermal Conductivity

The measurement of humidity is based on a determination of *temperature differences*. The effect is used that water and substances containing water stand out from other materials and from the environment in their *heat capacity* and in their *thermal conductivity*. Furthermore a distinction can be made by the evaporation energy. On wetted materials, the evaporation of water leads to a lowering of the surface temperature and can be measured by temperature sensors. Optical sensors, optical fibres and the known temperature sensors (e.g. semiconductor or thermocouple; Chap. 6) are used as measuring instruments. This

measuring method is used in hydrology, environmental technology, mining and civil engineering. Long-term measurements can be carried out to detect movements of water in the soil/rock or to register leaks in landfills, pipelines or dams.

10.3 Measurement of Precipitation in Outdoor Climate

Climate measurements with stationary measuring technology, ascending balloons, mobile weather stations and via satellites are carried out in a comprehensive network worldwide. Besides the daily weather report, weather observation is of great importance for:

- the transport safety (aviation, shipping, car, rail),
- the agriculture,
- the fire fighting,
- the storm warning and
- the assessment of pollution.

The measurements are carried out according to fixed scenarios and with equivalent measuring instruments. This is the only way to ensure international comparability of the data. The most important measurement data that are recorded are

- the temperature at different altitudes,
- the air humidity,
- the amount of precipitation,
- the air pressure,
- the wind (direction and speed) and
- the solar radiation/cloud cover.

10.3.1 Measurement of the Relative Air Humidity

The standard technique for recording relative humidity has been the *Assmann psychrometer* (Sect. 10.2; Fig. 10.14) since the beginning of targeted meteorological recordings. Even today, it is still regarded worldwide as the measurement standard for outdoor climate measurements. Most automatic measuring stations are now equipped with capacitive polymer sensors for recording relative humidity and temperature.

10.3.2 Precipitation Measurement

The rainfall quantity is recorded in collecting tanks with a defined opening area (Fig. 10.19) and is specified in relation to the area. The rain falling through the opening is brought into drop form via a funnel. These drops can be counted individually and calculated as

Fig. 10.19 Recording the amount of precipitation with the rain gauge



precipitation quantity per time unit. In the same way, the drops can fall onto a seesaw, at the ends of which small collecting trays with a defined volume are attached. If bowl 1 is filled, the rocker tilts, bowl 1 is emptied and the drops fall into bowl 2. The number of tilting movements is counted as a measure of the precipitation quantity fallen. The precipitation gauges are often equipped with a heating system. This also allows the measurement of snow quantities in some cases. The measured value is given in $[l/m^2]$. The specification $[mm/m^2]$ is also common. This designation describes the increase of the water level of one square meter by one millimeter: $1 l/m^2$ corresponds to $1 mm/m^2$. In the rain gauge according to Fig. 10.19, the rocker with the two collecting dishes can be seen in the middle; the collecting funnel (removed here) is placed above it.

10.3.3 Condensation Measurement

Condensation occurs when the water contained in the air (gaseous) condenses as a result of temperature drops and is deposited on surfaces as liquid water. The consequences are, for example, wetting of surfaces, formation of fog and frost or ice smoothness. Condensation has little significance in relation to the amount of liquid water released. On the contrary, wetting of surfaces can lead to the following problems:

- Slippery roads,
- Fog leads to traffic congestion,
- Wings of airplanes ice up,
- Corrosion of metal parts or
- Development of pests on plants.

Similarly, the occurrence of precipitation as an event can be important without the exact amount having to be measured (e.g. closing windows, switching on windscreen wipers). In

these cases, it is metrologically relevant to record the times (start and end) for the occurrence of dewing/wetting in order to be able to react to the dangerous situations mentioned. So-called *wetting sensors* detect the condition wet/dry on a sensor surface. The measuring principles are very different:

- the measuring surface is a *capacitor*; during wetting, the dielectric properties of the sensor change;
- the measuring surface has *electrodes* that are conductively connected during wetting;
- the measuring surface is a *translucent layer*; when wetting occurs, the reflection/refractive index on the surface changes;
- the measuring surface consists of *special paper* whose electrical resistance changes when wetted.

By attaching a heater and a temperature sensor to the sensor surface, the measuring accuracy can be increased and a distinction between the precipitation types rain and snow can be made (Fig. 10.20).

Wetting sensors are used in the following areas, among others:

- Monitoring of pests in agriculture,
- Control of windscreen wipers,
- Evaluation of road conditions and
- Meteorological measurements.

10.4 Humidity Measurement in Closed Rooms

10.4.1 Measurement of the Climate in Homes and at Workplaces

While the physical parameters dew point and absolute humidity are essential for measuring the humidity content of the ambient air, the *perception* of human beings (moist, pleasant, dry) in terms of comfort depends on the relative humidity, air temperature and air flow. The setting of air-conditioning systems is done by specifying the relative humidity, while the controlled variable is usually the dew point. An optimum humidity content of the air in the living and working environment is not only important in terms of comfort. It is much more important when the air is too dry (relative humidity <30%r.h.):

- Irritation of the respiratory tract,
- Static charges and
- Increased dust formation.

A permanently too high ambient humidity (relative humidity >70%r.h.) leads to

- Diseases of the respiratory tract,
- Growth of moulds, bacteria, pathogens,

Fig. 10.20 Sensor for measuring precipitation events



- Destruction of materials (e.g. paper, wood, textiles),
- Failure of electronic equipment (e.g. due to short circuit, leakage current) and
- Damage to mechanical assemblies (e.g. through corrosion).

Capacitive polymer sensors (Sect. 2.6) with a temperature operating range of $-10\text{ }^{\circ}\text{C}$ to $120\text{ }^{\circ}\text{C}$ and a humidity range of 20% r.h. to 90% r.h. are mainly used to control *air-conditioning systems*. If separate recording and documentation of climate values is required, *data loggers* are used. These are self-contained measuring instruments with humidity and temperature sensors, an internal clock as well as a data memory and an interface (Fig. 10.21).

The measured values are recorded at pre-programmed time intervals and time periods and stored in the data logger until they are read out. For example, the ventilation behaviour of tenants can be recorded tamper-proof in case of disputes.

10.4.2 Climate in Museums and Exhibition Rooms

Historical buildings and exhibition rooms place special demands on the measurement and control of air humidity:

- The nominal values to be maintained are specified with narrow tolerances.
- Large temperature differences can occur between the walls and the interior.
- The necessary air-conditioning (heating, ventilation) cannot be as optimal as in modern buildings.
- Individual exhibits require an extreme ambient climate (air-conditioned showcases).

In addition, historical buildings and old masonry are subject to different environmental conditions and influences today than when they were built. The rooms are often used for exhibitions and events and are equipped with heating systems. A large stream of visitors additionally pollutes the indoor climate, while the outdoor environment is subject to greater corrosion due to higher air pollution. Table 10.6 lists the guide values for the relative humidity when storing objects.

Fig. 10.21 Data logger with 6 sensors for relative humidity and temperature (Factory photo: Driesen + Kern)

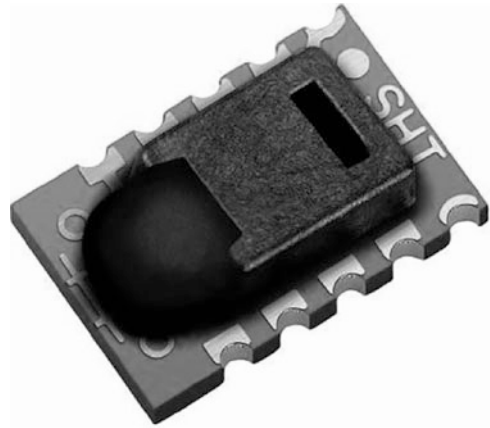


Table 10.6 Standard values for the storage of historical and art objects

Relative humidity [% r.h.]		Possible damage under unfavourable climatic conditions
Critical value	Best value	
> 25	<20	Glass, metals, minerals and photographic materials have an increased tendency to corrode
	20 to 40	Photographic material, the optimum value depends on the type of photographic material
> 55	40 to 45	Metallic exhibits can corrode
< 35	45 to 55	Irreversible changes in shape, especially of paper and textile materials; Blinding of alkaline glasses
< 40	55 to 65	Shrinkage or cracking of wood
> 55		Stones begin to discolour
> 65		Mould formation on organic materials

For works of art or historical exhibits that are to be presented, archived and stored in rooms, certain *ambient climate values* must be set (Table 10.6). The storage effect of the walls against heat and moisture in the climatic year cycle must be taken into account. Non-insulated external walls can be considerably cooler than the room temperature. Therefore, condensation can occur more frequently in places where paintings and other works of art are installed. Mould can form in places that are not visible (e.g. back of paintings or frames). It is therefore absolutely necessary to measure the humidity and also the temperature at these points or to take them into account when air-conditioning the rooms. For aesthetic or factual reasons, the installation of sensors in the immediate vicinity of the exhibits should be avoided. Miniaturised temperature sensors are very well suited for monitoring or controlling the humidity on the surface or back of art objects. The sensors are mounted directly on the outside wall. The relative humidity of the entire room, the value of which is set by the air-conditioning system, forms the reference basis in conjunction with the measured value of the wall temperature sensor. From both values, the humidity near the wall can be determined or estimated. The additional installation of a miniaturised air

Fig. 10.22 Miniaturized sensor (5 × 2 × 1)mm for relative humidity and temperature with digital signal output (Factory photo: Sensirion)



humidity sensor increases the measuring accuracy. The sensors are usually capacitive polymer sensors in combination with a temperature sensor. With dimensions of a few millimeters, these sensors can be placed almost invisibly at particularly critical locations (Fig. 10.22).

One problem with art and exhibition spaces is the *transmission of measurement data*. It is often necessary to dispense with cable routing. The sensors can be equipped with data loggers. However, climate control is not possible with data loggers. Radio modules are used to transmit measured values, which can send data over several hundred metres. Metal objects and wall reinforcements interfere with the transmission. Infrared transmitter-receiver systems can be used as a further option for wireless transmission of measured values. The ranges refer to the direct visual contact of transmitter and receiver.

In the case of incomplete air-conditioning of exhibition rooms, ventilation must be provided at certain times of the day or year by exchanging air with the outside climate (opening doors, gates, windows). As a rule, excess air humidity should be conducted from the interior to the outside. In order to determine the most favourable ventilation period, depending on the respective weather conditions, the absolute humidity (a) of the indoor and outdoor air must be determined. The relative humidity is not sufficient as the sole criterion. Only when the condition is fulfilled that

$$a_{\text{in}} > a_{\text{out}}$$

a_{in} : absolute humidity in the room

a_{out} : absolute humidity in outdoor climate

the moisture that has accumulated in the room air due to a high visitor flow or from the walls or furniture and other sources can be conducted outside. Absolute humidity is calculated from the measured values of air temperature and relative humidity. This type of climate control requires one humidity/temperature sensor each in the interior and in the exterior.

10.4.3 Climate in Electrical Installations

In rooms where high voltages and currents are converted, switched or distributed, an electrical flashover can have serious consequences for the subsequent technology or the central supply of electrical energy. As these systems are often located in the outdoor climate as unheated switch boxes, they are exposed to daily and annual *temperature and humidity fluctuations*. Depending on the location, a very high technical effort is made (e.g. switch cabinet heating, drying cartridges, ventilation systems or complete encapsulation of components) to prevent condensation on electrically conductive parts. It is possible to monitor special plant components such as insulators, cavities or housing parts; this is done as a combination of an air humidity measurement in conjunction with a temperature measurement on the special plant component. From the temperature difference between room temperature and surface temperature, taking into account the relative humidity of the room air, critical conditions can be detected at an early stage and countermeasures initiated:

- Switching on heating and cooling systems,
- Shutting down of plants in critical situations or
- Control of fans, blowers and dryers.

Furthermore, so-called wetting agents can be applied directly to insulators, switchgear or the outside of cooling water pipes. These sensors signal the condensation of water on the surface (Table 10.7).

10.4.4 Influencing the Indoor Climate

10.4.4.1 Air Humidification

Humidification of the air in rooms can be of particular importance during the heating season. Cool air (with low absolute humidity) flows from outside into the heated rooms. Here it is heated, which leads to a reduction of the relative humidity in the room. Additional humidification may be necessary to set the optimum humidity for people and surrounding objects.

10.4.4.2 Evaporator

The principle is based on the natural effect in which water vapour is released from a moistened surface into the environment. Under the condition of

$$e_w < e_{wv}$$

e_w : Vapour pressure of the ambient air

e_{wv} : Vapour pressure at the evaporation surface

Table 10.7 Parameters of wetting sensors for detecting condensation on surfaces

Parameters	Value ranges
Measuring range [%r.h.]	> 50, until formation of a condensation film
Measurement error [%r.h.]	+/-4
Temperature range [°C]	-20 to 80
Possible measuring principles	Resistance change of an electrically conductive polymer; open-stray field capacitor
Response time [s]	10 s to 200 s adjustable
Output	Analogue signal; switching signal; potential-free contact
Mounting type	Can be screwed onto pipelines; can be bonded to flat surfaces
Fields of application	Warning of moisture bridges on live components; detection of condensation on glass panes and shop windows; detection of condensation on car windows; detection of condensation on masonry surfaces and walls

the ambient air is humidified. The process is supported by the activation of fans. The evaporation surfaces are usually mats made of mixtures of cardboard and textile fabrics, which have a large surface area due to their porous or capillary structure. During evaporation, possible dirt or components (e.g. lime) of the water are not released into the room.

10.4.4.3 Steam Humidifier

A higher effectiveness of humidification is achieved by additional *heating of the water reservoir*. The water is heated up to near $T = 100$ °C. The resulting water vapour is distributed in the room by fans. Mineral residues of the water remain in the steam humidifier. The steam output and thus the humidification of the air in the room can be controlled via the temperature and the ventilation. The warm water vapour, which is released into the ambient air, heats up the room.

10.4.4.4 Nebulizer

These do not humidify the environment with steam, but with the *finest water droplets*. This means there is no energy input into the air-conditioning system through heated water. Residual components (e.g. dust or lime) are not retained in the humidification system but are released into the environment. Specially treated water is therefore required for the operation of the systems. Depending on the way these small particles are generated, atomizers are divided into:

- Ultrasonic humidifier,
- Nozzle humidifier and
- Disc or rotary atomizer.

10.4.4.5 Dehumidifying Rooms

Different principles are used to extract moisture from rooms. Depending on the size of the room and the air exchange capacity, the dehumidifiers are used as portable devices or fixed installations in air-conditioning systems.

10.4.4.6 Condensation Dehumidification

The air is cooled down to the dew point in a cooling system. The dew point can be reached by cooling or by compression of the air. Part of the water vapour condenses and is collected as liquid water. The air is then heated to the ambient temperature in a heat exchanger and fed into the room or air-conditioning system as dry air.

10.4.4.7 Adsorption Dehumidifier

Strongly *hygroscopic surfaces* (e.g. molecular sieve or silica gel) absorb the water vapour of the air flowing past and bind it to the surface. Generally, the adsorbate is located in a desiccant cartridge through which the air in the pipe system flows. When the adsorbate is used up, the system switches over to another desiccant cartridge. The adsorbate is regenerated by heating.

10.4.4.8 Absorption Dehumidifier

The humid air is passed over *salts* or *aqueous salt solutions*. The salts absorb the water vapour from the air until they reach their saturation point and dry the air flowing past. The salt can be regenerated by drying and used again for dehumidification. This type of dehumidification is used when only small amounts of air need to be dried. There is a risk that salt crystals will get into the air flowing past and deposit elsewhere in the air conditioner or in the room, which can lead to corrosion damage.

10.5 Air Pressure

Air pressure is the pressure that is created by the weight of the atmosphere on the earth's surface. Due to its constant availability, it serves as a reference value for many technical applications (relative pressure, Sect. 4.4).

10.5.1 Fields of Application

The measurement of air pressure as a special case of pressure measurement has three main applications:

- *absolute pressure value* for meteorological measurements,
- as *reference value* for *level measurements* and
- as a general *reference value* for *pressure measurements*.

It is also possible to determine the geographical altitude using the air pressure (e.g. in watches with altimeters).

10.5.2 Measuring Principles

The air pressure serves as a reference for the pressure measurement (Sect. 4.4). All measuring principles for absolute and relative pressure measurement described there are applied.

10.5.3 Definitions

The normal value of air pressure is related to sea level, where it averages 101,325 Pa. This average value fluctuates over the day, following the air temperature, which influences the density of the air. The fluctuations can be as high as 100 Pa in Central Europe. However, these small changes are strongly masked by meteorological influences, which can amount to up to 10,000 Pa.

If you move your point of view upwards from sea level, the column of air above becomes smaller, causing the pressure to decrease. As the air is compressed by its own weight, its density decreases with increasing height, so that the pressure decreases exponentially. The decreasing temperature with increasing altitude also influences its density. The *international altitude formula* exists for calculating the air pressure in the troposphere with an altitude of up to 11 km:

$$p = p_0(1 - (6.5 \cdot h)/288)^{5.256}$$

with p : Pressure at height h

p_0 : Pressure at height 0

h : Altitude above 0 in km.

In this equation, the temperature is used as the leading variable. The value 6.5 is the temperature decrease per altitude kilometre in Kelvin and the value 288 is the annual average air temperature on the ground in Kelvin.

10.6 Wind and Air Flow

10.6.1 Definition

Anemometer (derived from Anemos: wind) is the original name for devices for measuring wind. Technically, it is the name for instruments for measuring *air flow*, *volume flow* and in combination also *temperature*. In the meantime, liquids and gases can also be measured with it. In addition to climatic wind measurements, great importance is also attached to the measurement of air flow in many other areas, for example in all ventilation systems, in combustion processes in engines and in the exhaust air area (e.g. in the chimney), in model making, in aircraft construction and in many other areas. In meteorology, wind measuring instruments are often used in combination with *wind direction meters* and *temperature sensors*.

10.6.2 Methods for Wind Measurement

Seven main groups can be defined according to the physical effects and the resulting measurement methods.

1. *Mechanical* instruments set in motion by wind are the oldest and best known forms of measuring air flows. The classic *cup anemometer* is very widespread and can often be seen (Fig. 10.23). For this purpose, shells (usually 3 or 4) are connected to rods with a rotating rod in the middle. When set in rotation by the wind, the shells rotate around the central rod, whose speed is proportional to the air speed. They can be found for example at many weather stations, on high cranes and on motorways. For hand-held instruments, a propeller-like impeller anemometer is usually used.

2. An *estimate of the wind force* is common with the *Beaufort scale*. A distinction is made between 13 wind speeds (including calm = 0). They were introduced in 1805 by the English Admiral Sir Francis Beaufort for navigation under sails. The frequently amended table was also used as a basis for *land measurements*. The table according to *Koppen* has proven itself there. Starting with: Wind force 0, i.e., calm with wind speeds of 0 m/s to 0.2 m/s (corresponds to 0 km/h to 1 km/h) and the description: glassy sea, up to wind force 12, i.e., hurricane with wind speeds of >32.7 m/s (>117 km/h) and the description: air with spray and foam-filled air, the sea is white, with strongly reduced visibility, without distant vision. In 1949, another 5 hurricane stages up to wind speeds of over 200 km/h were added. Other scales and tables not listed here in detail are the Fujita scale, the Palles table and the Saffir–Simpson scale. In the past, measurements and comparisons were also made with the Wild wind speed chart.

3. Another large field of application is in the field of *differential pressure instruments*. Laminar (straight) and turbulent (swirled) flow are terms that describe the movement of gases, liquids and vapours (so-called fluids). Here the suction or *pressure effect* of an air jet on fluids is used. Giovanni Battista *Venturi* discovered that the velocity of a fluid flowing

Fig. 10.23 Cup anemometer

through a pipe is greatest where it narrows (flow velocity is inversely proportional to the flow cross section). Daniel Bernoulli discovered that the pressure drop of a fluid is proportional to the increase in velocity in a pipe through which the fluid flows (Bernoulli's law; Fig. 10.24). Thus, if the velocity of a fluid increases at a constriction (smaller pipe cross section A_2 in Fig. 10.24), the pressure drops there (height Δp ; Fig. 10.24).

The upright columns in Fig. 10.24 represent open connections to the pipe. Constrictions (e.g. Venturi flumes or sheets and Prandtl pipe) thus also deliberately allow higher velocities to be achieved in these areas (Sect. 10.7).

Venturi nozzles are mainly found in closed areas. A further development is the dynamic pressure probe deltaflowC from Systec Control (Figs. 10.25 and 10.26). It is used for measuring steam, water and compressed air quantities in numerous industrial sectors (e.g. power stations, sewage treatment plants, chemical plants and breweries) and is characterized by *high accuracy* and *low pressure losses*.

4. *Thermal anemometers*: In these instruments, a hot wire is cooled by the wind current and the difference, or *heat loss*, is measured by the change in resistance. This is then used to draw conclusions about the wind speed. Apart from impeller anemometers, thermal anemometers are the most frequently used measuring instruments to date for measuring wind or air flows in the range from 0 m/s to 5 m/s. The terms hot-wire anemometer or *thermal anemometer* are common here. Figure 10.27 shows the thermoanemometer AVM-888 of the company ATP Messtechnik GmbH from the field of ventilation and air-conditioning technology for wind speeds from 0 m/s to 20 m/s.

5. *Ultrasonic wind sensors* consist of 4 ultrasonic transmitters and receivers, with which the wind direction and wind force are measured. Even very small wind currents are

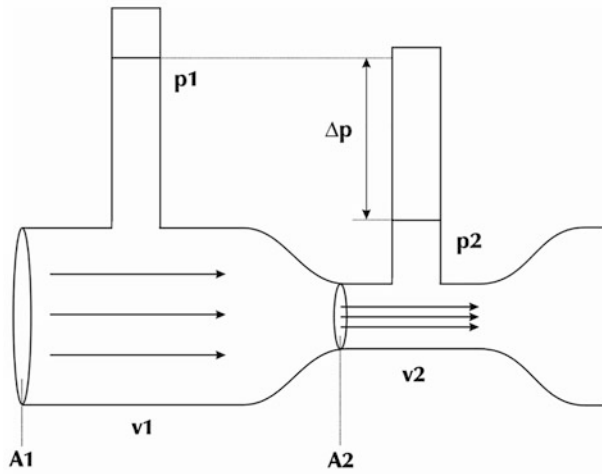


Fig. 10.24 Venturi nozzle

Fig. 10.25 DeltaflowC
differential pressure meter
(Factory photo: Systec Control)



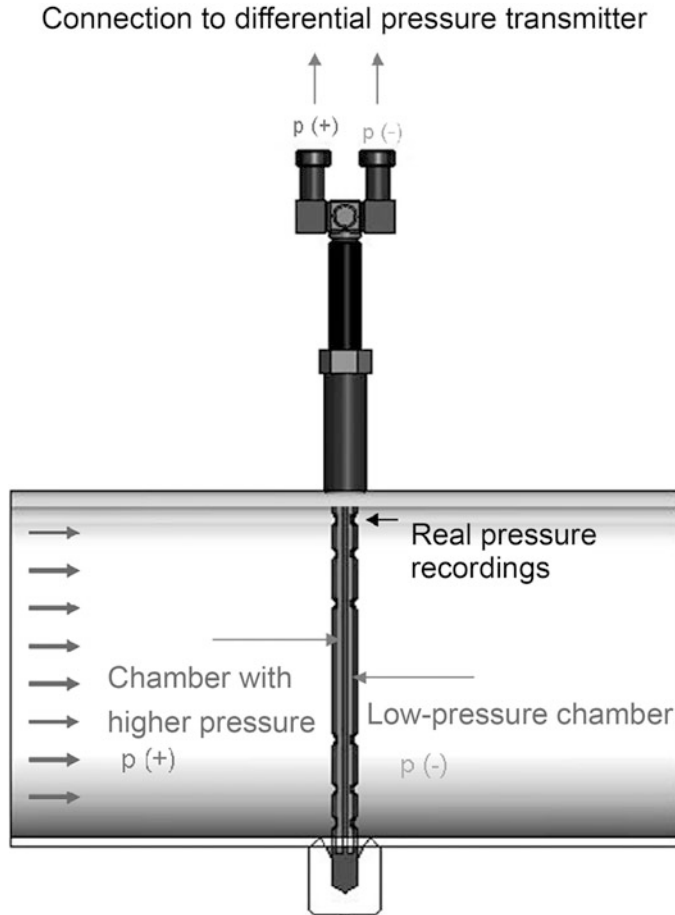


Fig. 10.26 Principle of the deltaflowC differential pressure meter (Factory photo: Systec Control)

recorded in real time and can provide information about wind strength and direction and their changes within a network. This is significant from a meteorological point of view. In addition, the smallest weather stations today also offer the real-time recording of wind gusts, temperature and humidity. *Acoustic sensors* are very reliable when spraying fields and assessing the *transport of pollutants* through the air (e.g. in the event of accidents and natural disasters). This enables authorities to react quickly to limit damage. The advantage here is the absence of rotating or revolving parts. An example of this is the weather station from Airmar Technology Corporation, which already combines all the above-mentioned functions (Fig. 10.28) and also offers software for real-time monitoring (Fig. 10.29).

6. Another application is the *glow discharge anemometer*. It works by measuring the current between two electrodes through which the fluid flows.



Fig. 10.27 AVM-888 thermal anemometer (Factory photo: ATP Messtechnik GmbH)

Fig. 10.28 Airmar system
(Factory photo: Airmar
Technology Corporation)



7. Another optical measurement method is the *laser Doppler method* (paragraph 2.20). Here, frequency differences between emitted and reflected signals are registered and the velocity in the profile is calculated by evaluation software. The 45° arrangement and crossing of the sensors (laser) produces a pattern of interference fringes with uniform heights and depths, or rather with areas of minimum and maximum light intensity. If a particle of a fluid is introduced into this pattern, it will reflect the light with a different frequency within the observed plane. This change is detected as movement within the plane and is represented as velocity. It is assumed here that the particle moves as fast as the surrounding fluid. However, since a large number of particles are measured and evaluated simultaneously, the total movement can be displayed without any doubt.

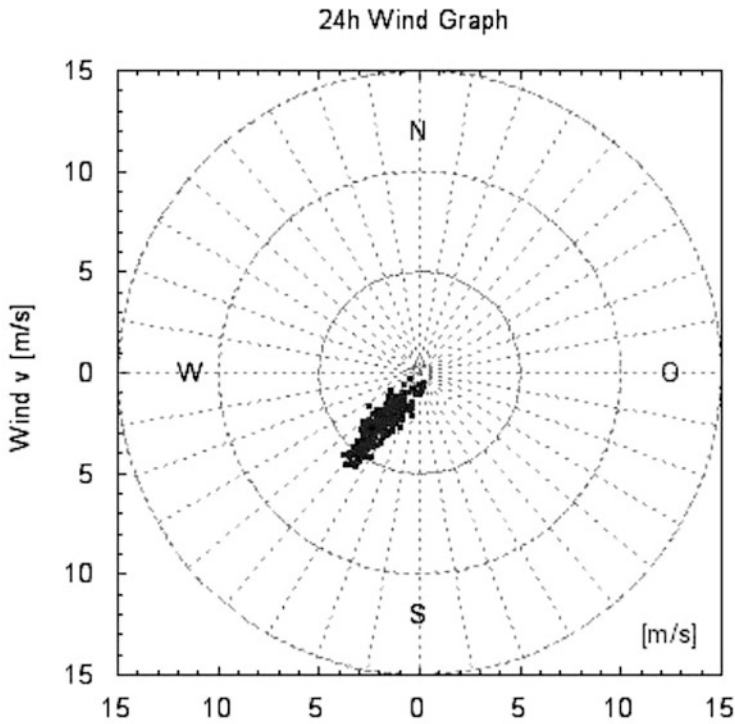


Fig. 10.29 Scheme 24 h wind graph of an LB150 (Factory photo: Airmar Technology Corporation)

10.7 Water Flow

10.7.1 Definition

In hydrology, the flow of water is described by the term *flow*. The flow rate Q is the volume of water V per time t flowing in a defined flow cross section. It applies:

$$Q = V/t.$$

The determination of flows in surface waters at predetermined river sections is an important application. For this purpose, the determined flow velocity v is multiplied by the cross section of the watercourse. Then the following applies:

$$Q = v \cdot A.$$

The result is usually given in the unit m^3/s or l/s .

In general, a distinction is made between *open* and *closed* systems. These include open channels such as rivers and canals, pressure pipes (e.g. in drinking water supply) and pipes in hydroelectric power plants. Partially filled pipes can be found, for example, in waste water and irrigation systems.

In partially filled pipe systems and open channels, the *height of the water level* is also an important part of the measurement, in addition to determining the velocity, in order to calculate the cross-sectional area through which the water flows. In addition to the range of the lath levels, the water level sensor technology is also used here (Sect. 3.1).

10.7.2 Direct and Indirect Flow Measurement

The *direct flow measurement* is carried out by means of measuring flumes (Venturi nozzle in Fig. 10.24), by volumetric measurement and by measuring weirs.

In the case of *flumes*, a cross-sectional narrowing is achieved by structural changes. The changed flow conditions lead to different water levels. The flow rate can be calculated from the difference in height.

The *volumetric measurement* is suitable for *low flow rates*. The quantity is determined by the size of the measuring vessel. The temporary diversion of the water flow into the calibrated measuring vessel is the basis for calculating the flow rate. If, for example, 5 seconds are required to fill a 10-l bucket, the flow rate is 2 l/s ($Q = V/t$).

Measuring weirs are installed in small channels (e.g. streams). This results in a narrowing of the cross section and thus an increase in the water level. According to the geometry of the weir and the overfall height, a flow equation is then created and the flow rate is determined.

The indirect flow measurement is carried out by means of an ultrasonic measurement, an electromagnetic measurement method, a hydrometric measuring blade or a measurement with tracers.

10.7.2.1 Ultrasonic Measurement

Ultrasonic flow measurement is an indirect measurement method, i.e. the flow is calculated on the basis of the *continuity equation* from a velocity measurement and a flow area assigned to the water level.

In practice, two main methods are used for measuring velocity with ultrasound:

- Ultrasonic transit time measurement and
- Ultrasonic Doppler measurement.

10.7.2.2 Term Principle

The measuring principle is based on the *direct* measurement of the *transit time* of an acoustic signal between two ultrasonic heads, the so-called *hydroacoustic transducers*. A sound wave that moves in a body of water against the direction of flow requires a longer

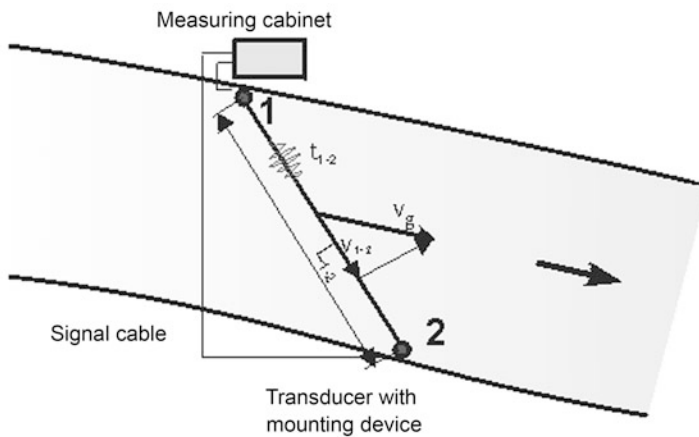


Fig. 10.30 Transit time measurement using ultrasound

transit time than a sound wave that travels with the direction of flow. The *difference in transit times* is directly *proportional* to the *flow velocity* in the measuring path and thus, with known cross-sectional and flow geometry, proportional to the *flow rate*. In the case of flow cross sections with a free water level (channel flow), the cross-sectional area through which the flow passes depends on the water level, so that the water level must always also be measured in order to determine the cross-sectional geometry (Fig. 10.30).

The sound wave generated by the transmitter (1) propagates in the medium as a spherical wave front. If the water is still, the time $t_0(1-2)$ would be required to reach the other bank. In flowing waters, the ultrasonic wave is carried along with the water, whereby its propagation speed is increased by the flow speed. This transit time, which thus appears to be shorter, is a measure of the flow velocity. The value must be corrected vectorially by the superimposed movements of the river and the sound wave.

With regard to the flow geometry, suitable assumptions shall be made whose validity shall be confirmed by calibration measurements. The *transit time measurement* is technically realized with different methods. The frequency band method and the pulse method are mentioned here. With the *frequency band method*, a defined frequency sequence is emitted into the water body and its transit time from the transmitter to the receiver is measured. In the *pulse method*, the transit time of a short-term sound pulse with a defined frequency is measured. The principle of transit time measurement with the pulse method is widely used in the field of *continuous ultrasonic flow measurement* in rivers and canals.

The alternative to cable crossing in wide waters is the use of *wireless systems*. This eliminates the need to lay cables through the body of water. It is therefore ideally suited for determining the discharge in *wide waters*. Autonomously operating measuring systems (master system and slave system) are installed on both sides of the water, each of which is coupled with a GPS receiver. The high-precision normal frequency generated from this and

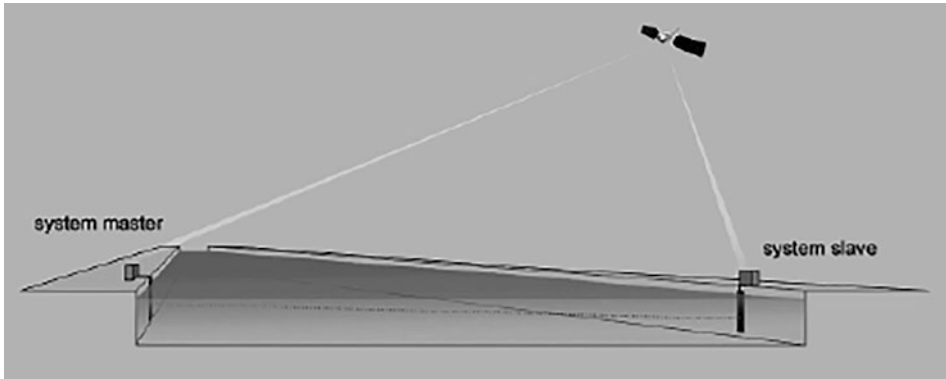


Fig. 10.31 Wireless measuring point

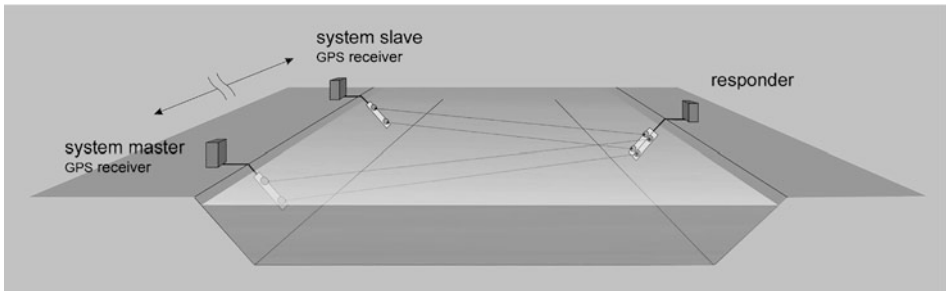


Fig. 10.32 Responder wireless measuring point

a start pulse that is exact on both sides are the prerequisites for synchronicity and thus for accurate ultrasonic measurement. Figures 10.31 and 10.32 show two applications for the use of such systems.

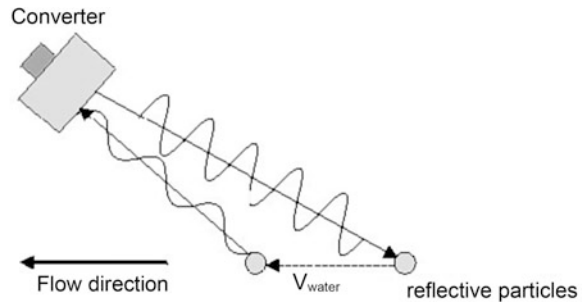
10.7.2.3 Doppler Principle

The ultrasonic Doppler principle (Sect. 2.20) exploits the sound reflecting properties of the particles in the water. It is assumed that the particles move on average in the same direction as the supporting flow. The ultrasonic transducer receives the reflected sound waves of moving particles at a different frequency than that at which they were emitted (Fig. 10.33).

The frequency shift is called *Doppler shift* (paragraph 2.20) and is proportional to the *velocity* of the reflecting particle. In addition to the frequency shift, the time it takes for a signal to return to the transmitter can be measured. From this time, the distance of the particle can be determined, which ultimately makes it possible to determine *velocity distributions* along a measurement line.

The Doppler principle is used in practice in the following areas, for example:

Fig. 10.33 Principle of Doppler measurement



- ADCP (Acoustic Doppler Current Profiler) measuring devices for *flow measurement* in rivers and canals,
- ADCP measuring instruments for *current measurements* in oceanography and
- *Ultrasonic Doppler probes* for flow measurement in sewers and small channels.

The ADCP consists of four to nine ultrasonic transducers (depending on type and manufacturer), which serve as transmitter and receiver.

10.7.2.4 Electromagnetic Measuring Method

The *magnetic-inductive measuring method* (Sect. 2.5) has been known and used in many industrial areas since 1939. Measuring instruments that record the flow rate by magnetic induction are used where electrically conductive liquids ($>5 \mu\text{S}/\text{cm}$) are measured indirectly. In the food industry, in water, sewage and sludge collection, in the hazardous area of the petrochemical industry and in many other areas (Sects. 3.1.1 and 3.1.5), a wide variety of measuring instruments of various types are found from different manufacturers.

The movement of conductive liquids through a magnetic field displaces electrical charge and generates a voltage between the beginning and the end of the magnetic field (length). The magnitude of the voltage is directly proportional to the velocity of the liquid, the path length and the strength of the magnetic field (Sect. 2.5).

Long-term measurement technology depends, among other things, on corrosion and the selection of the appropriate sensor material, which should be appropriate to the application environment. The applicable standards for stainless and acid-resistant steel, for example, and many published tests with various liquids are helpful in this regard.

10.7.2.5 Hydrometric Measuring Blade

The widely used and thus most frequently applied method of flow measurement is based on the measurement with a *hydrometric vane*, whereby the flow velocity in flowing water is documented at many points. Together with the flow cross section A and the individual velocities v in the respective measuring field, the flow $Q = A \cdot v$ is calculated. The hydrometric vane can be used both for very small watercourses and for large rivers. There, different velocities are then measured at different depths and a *velocity profile* is created from these. As at the measuring points shown in Fig. 10.34, measurements are

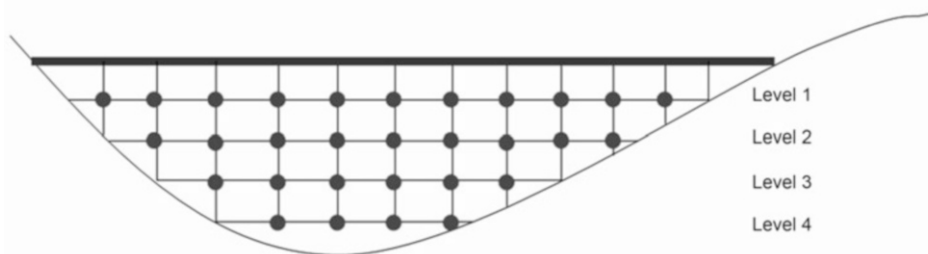


Fig. 10.34 Measuring grid for measuring the flow profile

taken in different planes corresponding to a fixed depth, each in the vertical plane. A velocity profile over the cross section can be created from the values obtained.

The measuring blade consists of a rod with a propeller-like, calibrated measuring blade at the lower end. The measuring blade generates an electrical impulse with each revolution, which is recorded by a counter over a fixed time interval. The resulting flow velocity is calculated as *pulses per unit of time*.

The mathematical representation of velocity is achieved by extrapolation using the finite element method for modelling a numerical simulation - the SIMK calibration. In this process, secondary flows are calculated as a function of the profile, water level and wall roughness, even in strongly moving channels. As a result, a *water level dependent factor* $k(h)$ for the calculation of the *mean flow velocity* is generated from a local flow velocity, from which reliable total flows are calculated.

10.7.2.6 Measurement with Markers (Tracers)

A fixed amount of water is marked by the one-time administration of an indicator (tracer) or as a continuous entry via a regulated pump control. Electrolytes, dyes and radioactive substances are used as tracers. The indicators mix completely with the water after a distance. The actual measurement is carried out after mixing by measuring the conductivity (electrolytes), the dye measurement (dye input) or by a Geiger counter (radioactive marking). In areas where a hydrometric measuring blade no longer produces reliable results, tracers are used for measurement. Fields of application are *mountain streams* and *shallow waters* under the condition that there is no level change from the beginning to the end of the measurement.

Other flow measurement methods used in industry but not listed here are Vortex, Coriolis and Delta P, as developed by Endress and Hauser.

Bibliography

1. A Guide to the measurement of humidity; NPL London 1996
2. AIRMAR Technology Corporation www.airmar.com

3. ATP Messtechnik GmbH www.atp-messtechnik.de
4. Baumgartner, A.; Liebscher, H. J.: Allgemeine Hydrologie, Gebr. Borntraeger Verlag 1996
5. Durchfluss Handbuch; Herausgeber: Quantum Hydrometrie Gesell. f. Mess- und Systemtechnik mbH; www.quantum-hydrometrie.de
6. Ergonomie des Umgebungsklimas – Grundlagen und Anwendung relevanter Internationaler Normen (ISO 11399:195); DIN EN ISO 11399; 2001–04
7. Evapotranspiration nach Haude DIN 19685
8. Gasfeuchtemessung; Kenngrößen und Formelzeichen; VDI/VDE 3514
9. Herrmann, R.: Einführung in die Hydrologie, Teubner Verlag, 2001
10. Hölting, B.; Coldewey, W. G.: Hydrogeologie: Einführung in die Allgemeine und Angewandte Hydrogeologie, Spektrum Akademischer Verlag, 2008
11. Klima am Arbeitsplatz und in der Arbeitumgebung; DIN 33403; 2001–06
12. Krah & Grothe: Messtechnik, Grundlagen der Feuchtemesstechnik
13. Kupfer, K.: Materialfeuchtemessung, expert-Verlag, Renningen-Malmsheim 1997
14. Maniak, U.: Hydrologie und Wasserwirtschaft: Eine Einführung für Ingenieure, Springer Verlag, 2010
15. Parthier, R.: Messtechnik
16. Pegelvorschrift (LAWA, BMVIT), Anlage D 1998: Richtlinie für das Messen und Ermitteln von Abflüssen und Durchflüssen
17. Scheffer; Schachtschabel: Lehrbuch der Bodenkunde, Emke Verlag, Stuttgart 1992
18. Systec Controls Mess- und Regeltechnik GmbH, www.systec-controls.de
19. Wärme- und feuchteschutztechnisches Verhalten von Gebäuden-Klimadaten, DIN EN ISO 15927, 1999–07
20. Wernecke, R.: Fachbuch Industrielle Feuchtemessung, Wiley VCH Verlag, Weinheim, 2003
21. Wittenberg, H.: Praktische Hydrologie: Grundlagen und Übungen, Vieweg+Teubner Verlag, Wiesbaden 2011



Winfried Vonau

11.1 Redox Potential

11.1.1 General

To describe the redox potential, *Nernst's equation* (Sect. 2.15.1) goes into one of its special standards, the *Peters equation* (11.1), which describes the reducing or oxidizing power of a redox system quantitatively:

$$U = U_{\Theta} + \frac{RT}{zF} \ln \frac{a_{\text{ox}}}{a_{\text{red}}} \quad (11.1)$$

a_{ox} : activity of the oxidized species

a_{red} : activity of the reduced species

Often, in addition to z electrons, m hydronium (or hydroxide) ions are also involved in the redox balance according to the following formulation.



For such a scenario, which is typical in reality, the *Peters equation* is, for example, in the case of a prevailing temperature of $\vartheta = 25\text{ }^{\circ}\text{C}$:

W. Vonau (✉)

Kurt-Schwabe-Institut für Messtechnik, Waldheim, Germany

e-mail: vonau@ksi-meinsberg.de

$$U = U_{\Theta} - \frac{59.2m}{z} pH + \frac{59.2}{z} \lg \frac{a_{\text{ox}}}{a_{\text{red}}}. \quad (11.3)$$

It is obvious and of great practical importance that the oxidation or reduction effect of redox systems can be influenced via the pH value.

In order to be able to characterize redox systems independently of the pH value, the *rH value* defined in Eq. (11.4) was introduced as the calculation variable, whereby U_{H} is understood as the value of the redox potential related to the hydrogen electrode.

$$rH = U_{\text{H}} - 2pH. \quad (11.4)$$

Provided that in Eq. (11.3) $z = m$, the combination of a redox electrode with a pH electrode also permits the determination of the *rH value*. Although this is only rarely guaranteed, in some industries this value is used. Thus sometimes a *classification of groundwaters* takes place regarding their reducing characteristics on the basis of their *rH values*. A significant correlation between water quality and the health and life expectancy of the inhabitants of individual cities in France has already been established. In places with soft spring water low in minerals, there were significantly fewer heart, circulation and cancer diseases. One of the factors used to determine water quality was the rH value. The best value determined was 22, but in most cities, it was far exceeded. In the following, the correlation between rH values and electrochemical water properties, which is common today, is given in this context, although an evaluation with regard to medical effects cannot be given here due to a lack of well-founded findings.

rH = 0 to 9 : strongly reducing properties,

rH = 9 to 17 : predominantly weakly reducing,

rH = 17 to 25 : indifferent systems,

rH = 25 to 34 : predominantly weakly oxidizing,

rH = 34 to 42 : strongly oxidizing.

However, it should be noted in this context that numerous publications ascribe a special life-prolonging effect to certain, mostly glacial waters. A typical example of this is the so-called Hunza water (found in the Hunza Valley in Pakistan) with an rH value of 21, which is ten times stronger than the best water in France in terms of antioxidant power.

In addition to the interconnection of redox electrodes with pH electrodes, combinations with other electrochemical half cells to form *potentiometric electrodes* are also possible. For many years, only the redox potential was used to assess swimming pool water with regard to its current content of active disinfectant (preferably free chlorine). Since this at best provides information about the germicidal effect of the added chlorine quantity, but not about its concentration, new regulations now require the specific determination of chlorine, for which there are several sensory possibilities. One of these is the design of an

electrochemical electrode composed of a redox and chloride-selective electrode. For the redox system of interest



whose potential U_x is determined by the ratio of *free chlorine/chloride*, the *Peters equation* for the individual potential formed on the redox electrode can be formulated as follows:

$$U_x = U_{\Theta} + \frac{RT}{2F} \ln \left(\frac{\text{Cl}_2}{\text{Cl}^-} \right)^2. \quad (11.6)$$

From the additional measurement of the chloride content by means of a chloride-selective electrode follows another equation to be considered:

$$U_{x'} = U_{\Theta'} - \frac{RT}{F} \ln \text{Cl}^-. \quad (11.7)$$

The electrode potential of the combination electrode on which this is based is thus a potential U achieved by difference formation, which according to Eq. (11.8) represents the relationship between the potential of the electrode and the chlorine content.

$$U = U_x - U_{x'} = K + \frac{RT}{2F} \ln \text{Cl}_2. \quad (11.8)$$

In particular, the poor reversibility of the redox system *chlorine/chloride* and the associated slow and poorly reproducible redox potential setting did not lead to the widespread use of this measuring method. Instead, it is currently common practice to add *iodide* to the analyte and, due to the formation of equivalent quantities of iodine in the process, to arrive at the considerably more reversible and thus better detectable redox system *iodine/iodide*.

The *redox potential* in body fluids, which is analytically understood as the sum of reducing and oxidizing substances in a solution, is sometimes correlated with the content of *free radicals*. *Antioxidants* (also known as radical scavengers) act as reducing agents. They react easily with oxidizing substances and therefore protect important other molecules in the organism from oxidation. For these reasons, the redox potential is a frequently used parameter for understanding the (bio)chemistry of reactive species. Redox potentials also provide information about the *autoxidizability* of compounds (transition metal catalysts, such as magnesium, iron, copper ions, increase autoxidation). Many biologically important molecules are oxidizable by O_2 to O_2^- (superoxide), such as glyceraldehyde, FMNH₂, FADH₂, adrenaline, noradrenaline, L-DOPA, dopamine, tetrahydrobiopterin and thiol compounds, such as cysteine. The formation and consumption of superoxide causes sometimes significant changes in the redox potential.

However, it must always be taken into account that sufficiently stable and reproducible redox potentials for the measurement only develop in so-called heavily weighted and reversible redox systems. In the case of identical activities of oxidized and reduced species, maximum weighting prevails, whereby the adjustment time and reproducibility of the potentials can be improved by adding substances known as potential mediators to weakly weighted systems. If only electrons are involved in the redox equilibrium and high exchange current densities occur in the electric field, the redox system has optimum reversibility.

11.1.2 Precious Metal Redox Electrodes

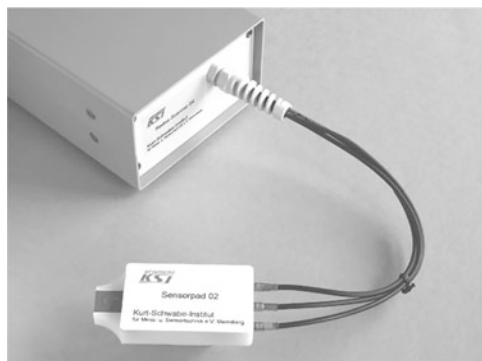
To determine the redox potential, electrodes *based on precious metals* have so far been used predominantly (mostly in compact DIN-compliant design, but sometimes also produced by *coating technology* or by *electrodeposition*). While the standard-compliant *rod-shaped redox electrodes* currently have their main area of application in *wastewater, water and sludge analysis* and will certainly continue to do so in the future, planar, miniaturized sensors with precious metal coatings are playing an increasing role in measurements in biology and medical technology. Figure 11.1 shows in this connection a redox scanner with a sensor pad as a device system for the rapid determination of the redox potential in blood/serum, whereby the disposable indicator electrode consists of a planar alumina ceramic substrate with screen-printed gold thick-film electrode.

Irrespective of the external shape of the precious metal-based indicator electrode, the usual procedure for measuring the redox potential is *potentiometry* (Sect. 2.15.2), whereby the material basis for the measuring electrode is of great importance for the measurement result. In this respect, there is a significant difference whether gold, platinum or palladium are used as the construction material.

Gold electrodes, for example, respond additionally under certain circumstances to any chlorides and cyanides present in the measuring medium, whereas platinum electrodes do not exhibit this behavior. Instead, they form hydrides just like palladium in reducing solutions. This also affects the *electrode properties*, especially on the absolute position of the potentials to be assigned to the redox electrodes in identical analyte solutions as well as on the potential setting behavior when the composition of the measuring medium changes. In addition to the inherently poor reproducibility of the potentials of precious metal-based redox electrodes (measurement error of ± 25 mV), there are several other disadvantages of this type of electrode.

The electrodes become *useless* if *catalytic poisons* such as SO_2 or other sulphur compounds reach their surface. *Proteins* also cause inactivation of the precious metal surfaces. The presence of *gaseous oxygen* or *hydrogen* in the test medium influences the half-cell potential. The precious metals themselves can act as a catalyst in certain redox media in an undesirable way (for example, they can promote the decomposition of H_2O_2). The price of precious metals is high and the extensive substitution of precious metals is an

Fig. 11.1 “Redox scanner 04” and “Sensor pad” 02 (Source: R&D results of the Kurt-Schwabe-Institutes Meinsberg)



economic necessity. Investigations on the obvious use of carbon electrodes (e.g. graphite electrodes) have shown that there is no high reproducibility of the electrode potentials there and that they are therefore no alternative to the precious metal electrodes.

It is recommended to *condition* the electrodes regularly, in the case of laboratory measurements if possible before each series of measurements, by degreasing and cleaning with detergents, rinsing with water and treating with hydrochloric acid and rinsing again with water. Any further pretreatment depends on the level of the expected redox voltage and the precious metal electrode used. While no further measures are necessary for gold electrodes, it is recommended for platinum electrodes to add an ammonia solution when measuring in oxidizing media and to treat with iron(II) sulfate or iron(III) chloride solutions when measuring in reducing media before rinsing with water again in both cases.

In connection with the operation of the instrument system according to Fig. 11.1, it is intended, with regard to the above-mentioned problem, to create defined starting conditions for each measurement, for example, in the course of conditioning the gold measuring electrode by *cyclic polarization*. This requires the realization of a *three-electrode measuring cell* (Fig. 11.2a) which is operated according to the scheme shown in Fig. 11.2b. Figure 11.2c shows the success of the corresponding procedure, namely the creation of a reproducible measuring electrode surface after each 20 polarization cycles under the test conditions shown in the diagram.

11.1.3 Redox Glass Electrodes

An alternative solution to conventional probes, where several of the above-mentioned shortcomings are not present, can in principle be redox electrodes based on glasses with very high electron conductivity. The mode of operation of these electrodes is based primarily on the presence of *iron oxides* or *titanium oxides* in the *glass compound*, whereby in the latter case, for example, titanium is present in the oxidation states +3 and +4 in a predefined ratio side by side. The iron-containing glasses in particular have the serious disadvantage that their functionality is no longer present in redox electrodes at pH values

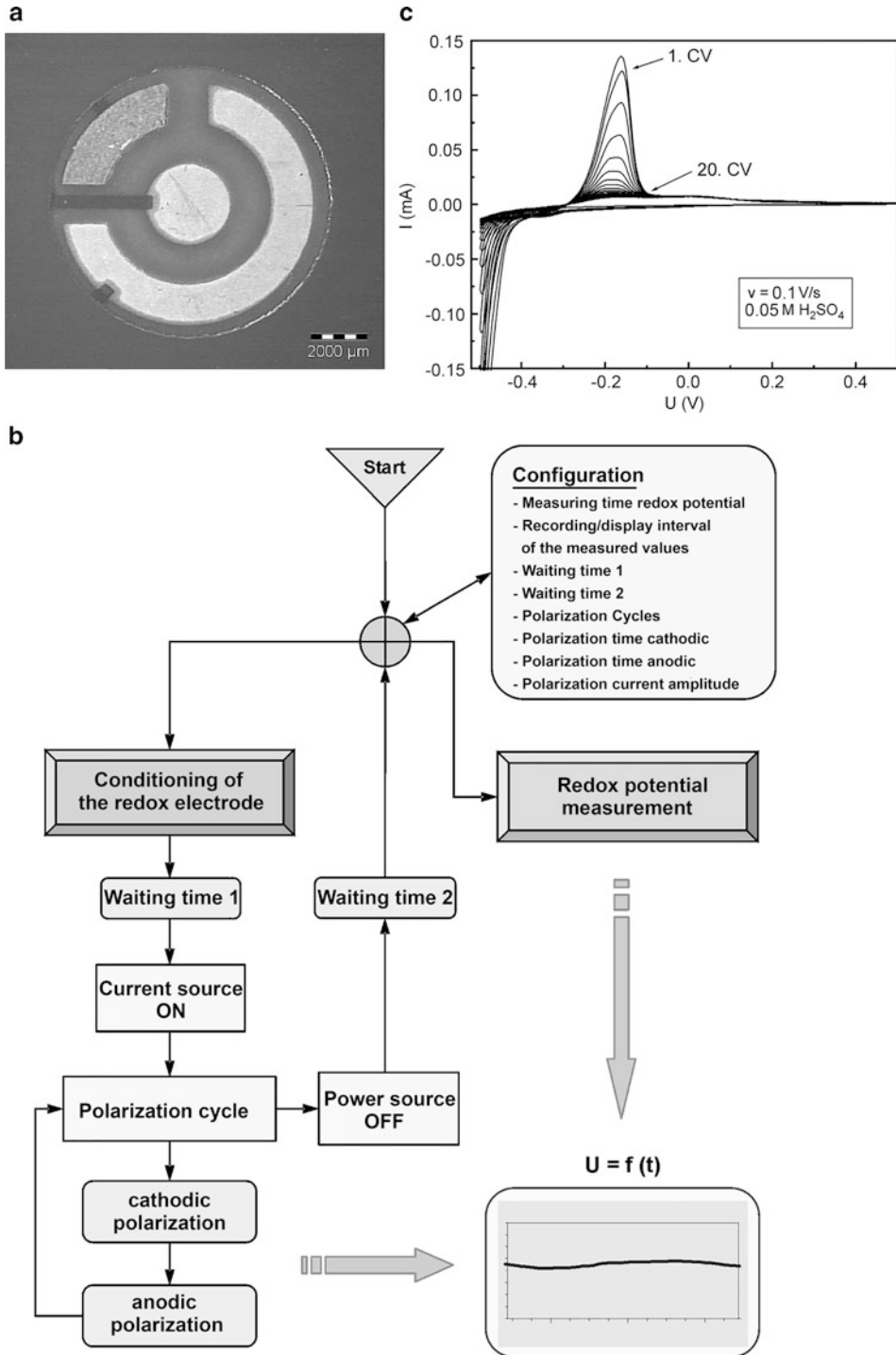


Fig. 11.2 (a) Sensor chip with a three-electrode measuring system for determining the redox potential, manufactured in thick-film technology. (b) Functional sequence for the table-top measuring

<3 and at temperatures >60 °C. The design of all redox glass electrodes described so far is currently limited to platinum wires coated with the glass mass and, in the few cases where the special glasses can be processed by glassblowing and their coefficients of thermal expansion are compatible with those of conventional electrode shaft glasses, to the geometric shapes known from conventional pH glass electrodes. An exemplary schematic representation of such a redox glass electrode is shown in Fig. 11.3. Figure 11.4 shows that under certain circumstances identical characteristic curves of redox glass and platinum electrodes can be achieved.

In contrast to conventional pH electrodes made of ion-conductive glass, the potential of redox glass electrodes is not discharged via a liquid electrolyte system (buffer + KCl or similar), but in *direct contact* via a mostly metallic electron conductor. In contrast to the pH glass electrode, where this would create an irreversible phase boundary *selective membrane/internal conduction*, this circumstance is not a disadvantage here.

The wide variability of the material composition of the redox glass membrane for the optimization of the electrode function is not given for reasons of the maximum allowed deviation of the thermal expansion coefficients of electrode and shaft glass, especially in the case of the production of classical glass electrode shapes (e.g. bulb or tip). Up to now, redox electrodes based on this approach have only played a role in niche applications. Recently, electrodes manufactured in planar technologies have also been available. Specifically, these technologies are thick film technology and laser ablation. Figure 11.5 schematically shows the structure of a thick-film redox electrode with a glass membrane. Such probes are of particular interest for measurements in the extended biological or biomedical field. If necessary, there is the option of integrating them into an ensemble with other electrochemical sensors that can be produced using screen-printing technology to form a *planar multi-sensor (lab-on-chip)*.

11.1.4 Reference Electrodes

An *electrochemical reference electrode* is additionally required to complete the ORP sensor as well as any other potentiometric sensor. This electrode is to set and maintain a constant potential independent of changes in the measured medium. For electrodes with aqueous electrolytes, the *standard hydrogen electrode (SHE)* with the temperature of the test electrode generally serves as the reference base:



Fig. 11.2 (continued) device “ORP scanner”. (c) Conditioning the surface of the thick-film Au redox electrode by cyclic polarization (CV)

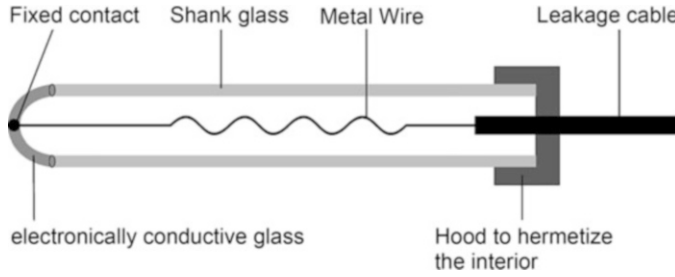


Fig. 11.3 Redox glass electrode

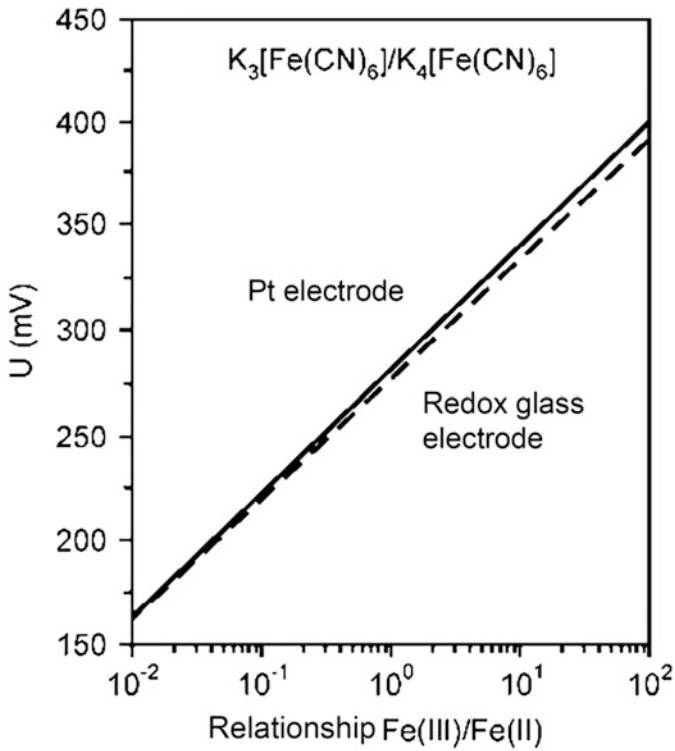


Fig. 11.4 Redox behaviour of a redox sensor made of oxide glass and a platinum electrode in $K_{3,4}[Fe(CN)_6]$ solutions vs. SSE (source: M. Miloshova et al.: New redox sensors in: meeting abstracts of the 1997 Joint International meeting of the Electrochemical Society)

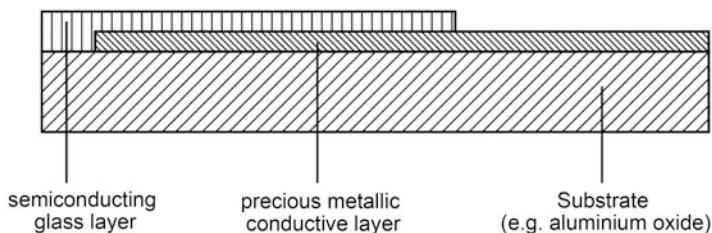
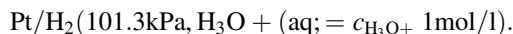
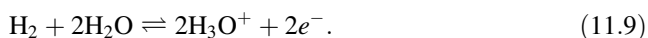


Fig. 11.5 Diagram of a planar redox glass sensor in thick-film technology



According to a suggestion by W. NERNST, single electrode potentials can be given by thinking of each electrode process combined with that of hydrogen ionization according to Eq. (11.9) and setting their standard electrode potentials to zero for all temperatures.



Most designs of hydrogen electrodes are based on the original form shown in Fig. 11.6a, which has been known for over 100 years. More recent developments have led to electrode geometries that correspond to or are even smaller than those of standard electrochemical rod-shaped sensors with a shaft diameter of 12 mm and a length of 150 mm (Fig. 11.6b). However, they only have a limited service life. Apart from the last-mentioned electrode variant, for which there could be several special additional fields of application in the future, it can be estimated that the SHE still has fundamental importance as a reference variable, but for practical measurements, it has been largely replaced by other electrodes, for example, because of its unwieldiness. First of all, the so-called *second type of electrodes* should be mentioned here, which also include the Ag/AgCl electrode mentioned in Sect. 2.15.2. In general, these are electrodes whose electrode equilibrium involves not only metal and dissolved ions but also a sparingly soluble salt of the metal in question. They are thus clearly distinguished from 1st type electrodes, which also include several gas electrodes including the SHE. The potential of electrodes of the second type depends on the *anion activity*, that is, in the case of silver chloride electrodes on that of the chloride ions.

Figure 11.7 shows the schematic structure of a *silver chloride reference electrode*, which can also be generalized to reference electrodes based on other material systems. Several such systems with practical relevance, including their potentials, compared with the SHE at 25 °C, are listed in Table 11.1. If potential measurements at other temperatures are made against electrochemical reference electrodes, their temperature coefficients must be known. To exclude so-called thermodiffusion potentials, it is recommended to keep the measuring and reference electrodes at the same temperature. For example, the following temperature function has been determined for the *saturated calomel electrode (SCE)*, which is currently one of the most frequently used electrodes in addition to the saturated

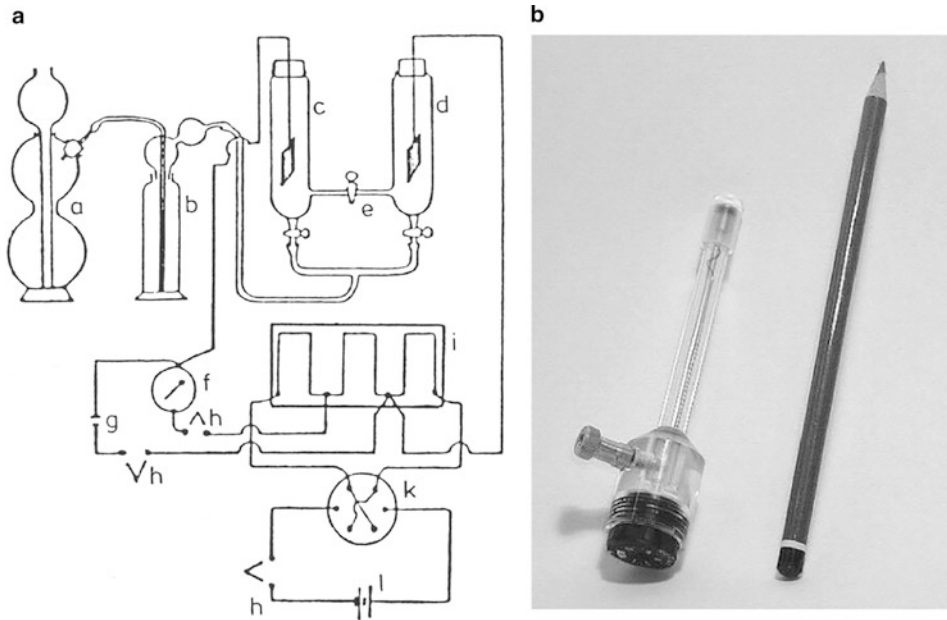


Fig. 11.6 (a) Measurement experiment with Pt/H₂ electrodes (c, d) (Source: M. Miloshova et al.: New redox sensors in: meeting abstracts of the 1997 Joint International meeting of the Electrochemical Society). (b) Hydrogen reference electrode (Source: M. Miloshova et al.: New redox sensors in: meeting abstracts of the 1997 Joint International meeting of the Electrochemical Society and US Patent 5 407 555 (1995))

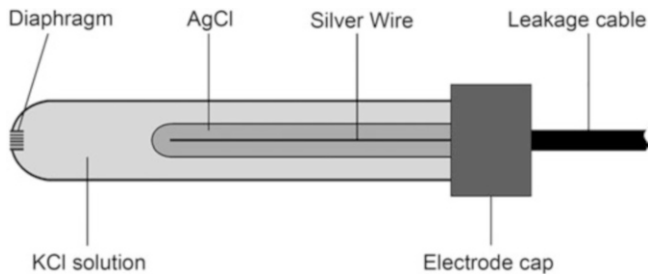


Fig. 11.7 Silver chloride reference electrode

silver chloride electrode (SSE). The standard voltage of the hydrogen electrode to which the term refers is, by definition, set to zero for all temperatures:

$$U_{\text{SCE}} = 0.2410 - 6.61 \cdot 10^{-4} (\vartheta - 25) - 1.75 \cdot 10^{-6} (\vartheta - 25)^2. \quad (11.10)$$

Table 11.1 Voltages of some 2nd type reference electrodes versus the SHE at $\vartheta = 25\text{ }^{\circ}\text{C}$

Electrode	U [V]
Hg/Hg ₂ Cl ₂ (s), KCl saturated	0.2410
Hg/Hg ₂ Cl ₂ (s), n/1 KCl (volummolar)	0.2801
Hg/Hg ₂ Cl ₂ (s), n/10 KCl (volummolar)	0.3337
Ag/AgCl(s), KCl saturated	0.197
Ag/AgCl(s), n/1 KCl (volummolar)	0.236
Ag/AgCl(s), n/10 KCl (volummolar)	0.2895
Hg/HgO(s), n/10 NaOH (volummolar)	0.165
Hg/Hg ₂ SO ₄ (s), m/1 H ₂ SO ₄ (weight molar)	0.6739
Ag/Ag ₂ SO ₄ (s), m/1 H ₂ SO ₄ (weight molar)	0.7162

Reference electrodes can be used both as part of separate electrodes and as a component of so-called *combination electrodes* in potentiometry. In principle, it should be avoided that the reference electrolyte contains the ion that is to be determined with the corresponding indicator electrode or, in the case of redox potential measurement, that influences the redox pair of interest. Alternative reference systems are then either selected or *electrolyte bridges are used*.

All reference electrodes mentioned so far contain *liquid system components* in the form of the reference electrolyte solution. This is associated with numerous *disadvantages*. Examples are the *position dependence* during storage and use, restrictions during use at *low or high pressures* and *temperatures*, the danger of *diaphragm blocking* if the measuring medium is contaminated, and *limitations* in the *miniaturization* of the electrodes. For this reason, intensive efforts have been made in recent years to arrive at so-called *all-solid-state reference electrodes*. Especially when the indicator electrode is a pure solid-state electrode, which is the case for redox electrodes or the pH-ISFET sensor (Sect. 2.16.2, Fig. 2.103 and Sect. 11.2), there is a considerable need for this. At this point, Table 11.2 lists a few examples of such solutions. So far, only the *gel-stiffened electrode types* have been able to achieve significant commercial success.

11.2 Ions Including Hydronium Ions

11.2.1 General Information

Ions can be determined by sensors in an electrochemical and non-electrochemical way. At this point a concentration on the electrochemical analysis takes place. In principle, *voltammetric* or, as its special form, *polarographic* (using mercury drop electrodes) and *potentiometric* determinations are possible. Voltammetry, which can also be carried out in special forms such as differential pulse or *square-wave voltammetry*, usually requires relatively large measurement setups. The core piece of these setups is the electrochemical measuring cell consisting of a working, counter and reference electrode (in a specific

Table 11.2 Concepts for solid reference electrodes in electrochemical sensor technology

Solids system	References
Gel or polymer stiffened reference electrolytes, for example by adding agar-agar, polyacrylamide, polyvinyl alcohol or mixtures of several gel formers, possibly with the omission of the diaphragm.	H. Galster: GIT-Fachzeitschrift Labor 24 (1980) 744 Patent DE 3 100 302 (1980), Patent DE 3 228 647 (1982)
Epoxy or polyester resin or other polymer filled with electrolyte salt with complete omission of the diaphragm.	Patent EP 0 247 535 (1987) und Patent DE 19533059 (1997)
Electrolyte-containing screen printing pastes.	W. Vonau et al.: Electrochimica Acta 29 (2004), 3745
Solidified electrolytic salt melt in diaphragm-containing containers.	Patent DE 103 05 005 (2003)
Transition metal oxide bronzes.	Patent DE 102 52 481

geometric arrangement to each other). Figure 11.8 shows the *polarogram* of a solution containing various metal ions, thus demonstrating the suitability of this method for the simultaneous qualitative and quantitative *determination of* selected ions. However, voltammetry in the narrower sense is not directly a sensory analysis method. For this reason, it will not be considered further here. Instead, the use of a series of *ion-selective electrodes (ISE) based on the principle of potentiometry* will be discussed in more detail below (Sect. 2.16.2).

11.2.2 pH Measurement

For the measurement of the pH value, one of the most frequently solved analytical tasks worldwide, there are several possibilities according to Fig. 11.9.

If, for example, one of the two hydrogen electrodes shown in Fig. 11.6a (c or d) is filled with a solution of unknown hydrogen ion activity and the other is used as SHE, it is possible to determine pH values very accurately.

This complex measuring method is still used today, particularly for testing buffer solutions. In metrological practice, the glass electrode, which is produced annually in the millions worldwide, currently dominates (Sect. 2.16.2). In most countries, standards prescribe its use for determining the pH value. Nevertheless, there are applications in which other systems shown in Fig. 11.9 are also used. Despite some progress in the development of pH optodes, the preference is still clearly for electrochemical measurement methods.

Thus, motivated by the disadvantages already described for reference electrodes with liquid system components, there are several developments towards *fixed-contact glass electrodes*. Due to the necessity of having a very thin pH-sensitive electrode glass membrane, conventional glass electrodes also have a distinct lack of mechanical instability. In the case of fixed (all-solid-state) glass electrodes, electrodes with a silver-mirrored internal

Fig. 11.8 Polarogram of a solution with different ions

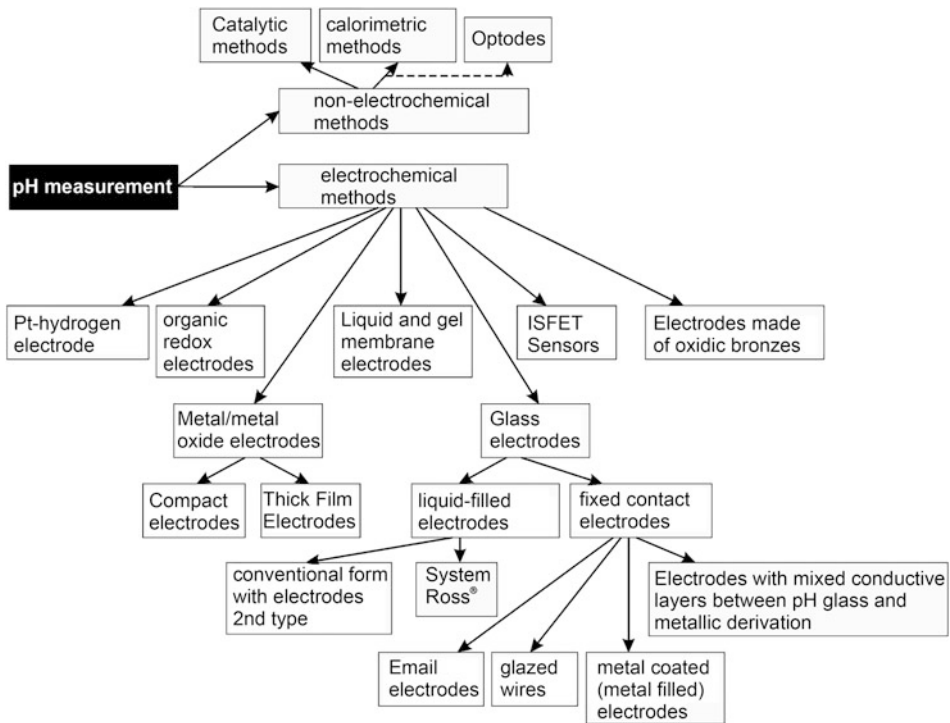
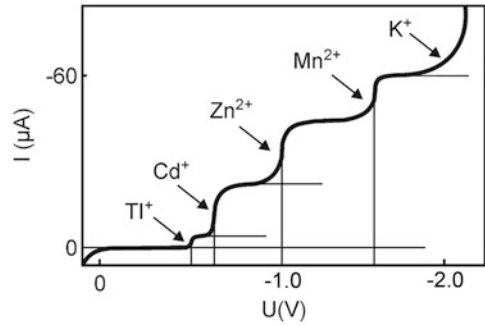


Fig. 11.9 Determination methods for the pH value

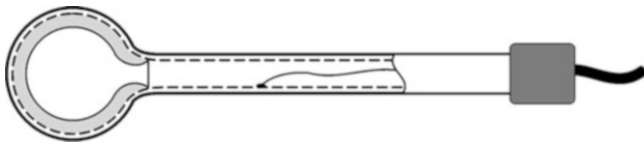


Fig. 11.10 Jena calf electrode with metal coating and supported membrane (Source: L. Kratz: Die Glaselektrode und ihre Anwendungen, Verlag Steinkopf, Frankfurt/Main, 1950)

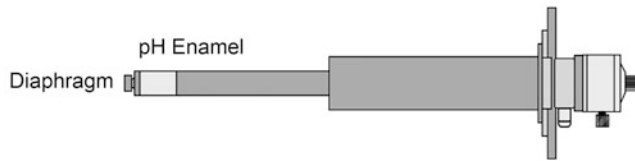


Fig. 11.11 pH combination electrode with sensitive enamel as membrane material (Source: Patent DE 213 3419 (1971))

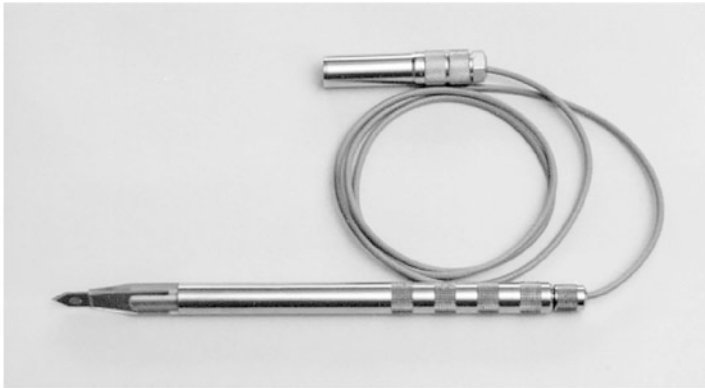


Fig. 11.12 pH insertion probe in thick-film technology on a steel substrate to check the meat quality (Source: Patent DE 199 61 210 (1999))

glass membrane surface (Fig. 11.10) are particularly well known for laboratory use. Probes based on *pH enamel* (Fig. 11.11), which have long since gained importance in process measurement technology, and glass electrodes in thick-film technology (Fig. 11.12), which can be used in perspective for controlling meat quality, have also been described.

However, there are various limitations with these probes, including the limited pH application range for enamel electrodes due to the susceptibility of special enamel to corrosion, and the minimum application temperature for thick-film electrodes due to their high internal electrode resistance. The phase boundary potential theory does not apply to the types of fixed-contact pH electrodes presented here, nor to most others. This is often used to explain the functioning of the conventional glass electrode with liquid internal electrolyte, which has reversible phase boundaries on both sides of the pH-selective membrane (Fig. 11.13 left). The situation shown in Fig. 11.13 on the right for direct metal contact with the inner surface of the pH-selective glass membrane is completely different, which often explains measurement uncertainties and potential drifts of all-solid-state glass electrodes, as shown in Figs. 11.14 and 11.15 for silver-contacted pH glass electrodes.

In more recent developments, *mixed-conducting intermediate layers* (e.g. semiconducting glass, polypyrrole or zinc oxide) are often introduced into the arrester

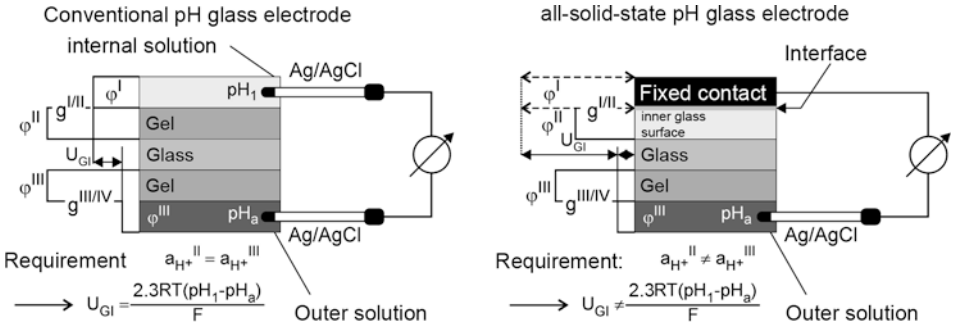


Fig. 11.13 Potential curve on pH glass electrodes with different internal discharge systems. I . . . IV Phases within a conventional and fixed-contact pH glass electrode. ϕ^i . . . Internal electrical potential of phase i. ε_{Gl} . . . Glass electrode potential. $g^{i/j}$. . . Galvanic voltage at the phase boundaries

Fig. 11.14 Electrode function of three fixed-contact pH glass electrodes (B, C, D) with silver conduction system against AG/AgCl, total KCl at $\vartheta = 25^\circ C$ (Source: W. Vonau, Kurt-SchwabeInstitut Meinsberg)

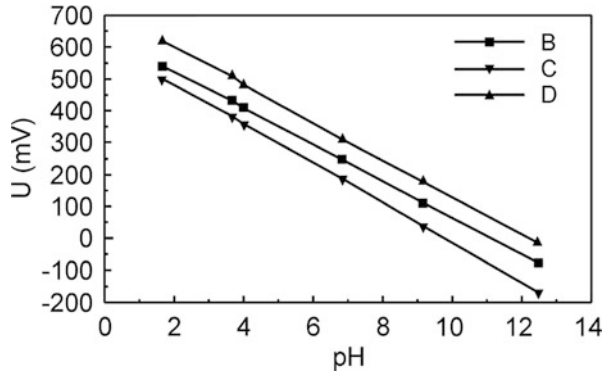
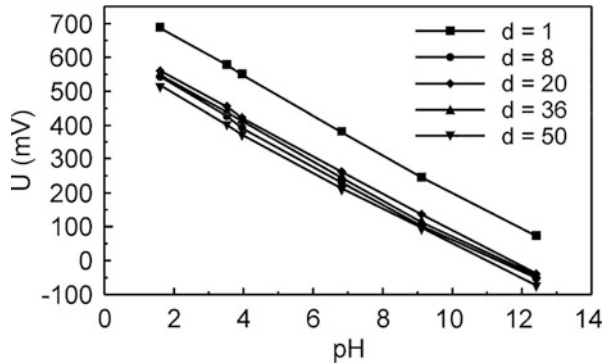


Fig. 11.15 Electrode function of a fixed-contact pH glass electrode with silver conduction system against Ag/AgCl, total KCl at $\vartheta = 25^\circ C$ on several days d (Source: W. Vonau, Kurt-SchwabeInstitut Meinsberg)



systems to eliminate this deficiency. Figure 11.16 illustrates the success of these measures for the first-mentioned solution (Fig. 11.17).

Most other possible pH sensors also play a role in various niche applications of pH measurement, for which some examples are shown in Table 11.3.

Fig. 11.16 Drift behaviour of fixed glass electrodes with different internal conduction systems (Source: W. Vonau, H. Kaden: *Glastechnische Berichte, Glass Science Technologies* 70 (1997), Nr. 5, 155)

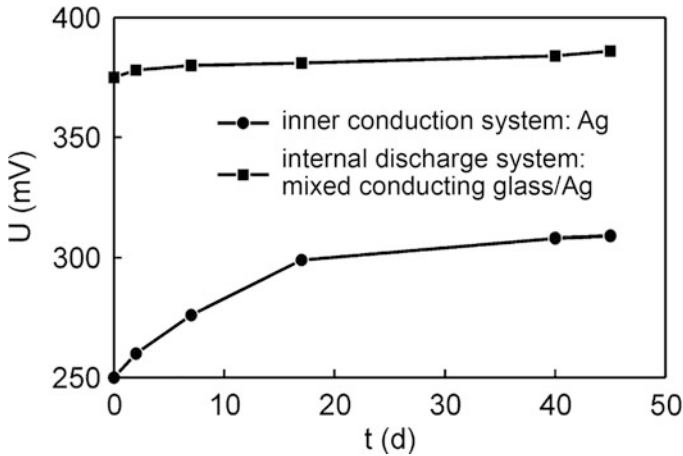
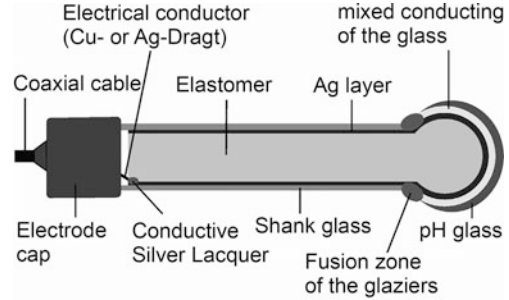


Fig. 11.17 Mechanically stable glass electrode with internal fixed contact, provided with a head of multilayer glass with layers of different conducting mechanism (Source: W. Vonau, H. Kaden: *Glastechnische Berichte, Glass Science Technologies* 70 (1997), Nr. 5, 155)

Table 11.3 Examples of applications of pH sensors beyond the glass electrode

pH sensor	Application examples
Organic redox systems (for example quinhydrone electrode)	Medical diagnostics, for example in body fluids
Metal/metal oxide electrodes [antimony and bismuth electrodes (compact and produced by electrodeposition), ruthenium oxide electrodes in thick and thin-film technology], iridium oxide electrodes in all designs	Medical diagnostics, for example, pH stomach probes Lab-on-chip systems for biological and medical investigations in low-volume analytes Food Analysis
Liquid and gel membrane electrodes	Hardly common, rare in medical technology
ISFET	Food analysis and medical diagnostics (especially in regions where glass parts must not be used) Corrosion investigations
Metal oxide bronzes	Currently not known

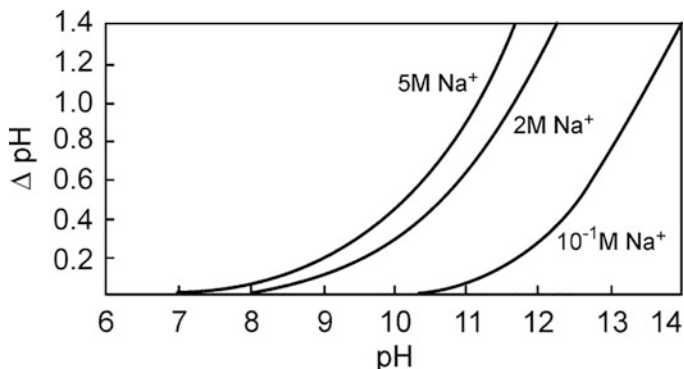
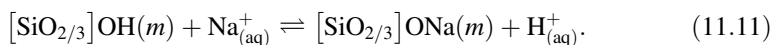


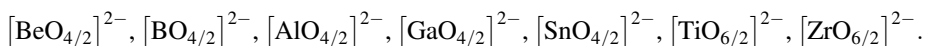
Fig. 11.18 Alkali error curves of a commercial pH glass at $\vartheta = 25^\circ\text{C}$ and different sodium concentrations in the analyte (Source: W. Vonau, Kurt-Schwabe-Institut Meinsberg)

11.2.3 Other Ions

Some monovalent and divalent cations, such as Li^+ , Na^+ , K^+ , Rb^+ , Cs^+ , NH_4^+ , Ag^+ , Zn^{2+} and Ca^{2+} can be determined with glass electrodes, whereby here the circumstance of the so-called *alkali error* (Fig. 11.18), which is disadvantageous for pH measurement with these probes, is specifically exploited. The term was coined because the pH glass electrode is often cross-sensitive to sodium ions. The extent of corresponding incorrect measurements of the hydronium ion activity can be estimated using the *Nikolsky equation* (2.15.4) introduced in Sect. 2.15.2 in the form of the selectivity coefficient K_{ij} . In the case discussed here, this corresponds to the equilibrium constant of the subsequent exchange reaction of ions between the membrane surface and the test solution. The selectivity coefficient is below -13 for commercially available pH electrodes.

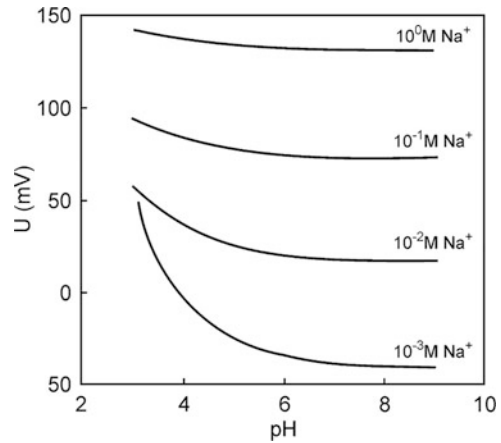


The condition for a selectivity other than the H_3O^+ selectivity of the electrode glass membrane is the presence or incorporation of a second glass former in the material, whereby its coordination must cause the occurrence of negative charges in the network. These charges are compensated for by alkali or alkaline earth ions. Typical network forming groups are:



In particular, the addition of aluminum oxide to the starting glass mixture causes a strong formation of a sodium ion selectivity of the electrode membrane. During the later

Fig. 11.19 pH influence on alkali ion sensitivity (Source: W. Vonau, Kurt-Schwabe-Institut Meinsberg)



measurement, a certain pH value must be set because pH influences on the sensor signal are still present (Fig. 11.19).

However, as the method of choice, especially for the determination of K⁺, ion-selective electrodes with liquid or gel membranes are used for this ion as well as for a number of other An- and cations, whereby the electrode active components are either organic ion exchangers or uncharged organic compounds. With components from the latter class of substances, also known as *ionophores*, the ions to be determined can form inclusion or complex compounds. *Valinomycin* (Fig. 11.20), which is embedded in a plastic membrane (polymer matrix electrode) in a concentration of approximately 0.7% and enables very precise and highly selective K⁺ detection, is probably the most important electrode-active substance. The complex formation constant for the potassium-valinomycin complex is 10⁶, whereas it is only 10 for the sodium-valinomycin complex. In recent years, there have been repeated attempts to synthesize ring molecules (preferably crown ether compounds) with cavities tailored to take up certain ions.

As *polymer matrix electrodes*, which are available both with liquid internal conduction and with fixed contact, the commercial market is currently dominated by *sensors* for the determination of *calcium*, *magnesium*, *ammonium* and *nitrate ions*. There are also combinations of these (Mg²⁺ and Ca²⁺ for water hardness determination) and with other electrodes (ammonium, nitrate and chloride) available. This allows the continuous and simultaneous measurement of both nitrogen parameters, for example in sewage treatment plants, and the chloride electrode acts as an electrode to compensate for the falsification of the measured values.

An alternative to the described potentiometric ion determination with electrodes containing polymer matrices is a patented sensor solution. The concept also known as *ISCOM* (ion-selective conductometric microsensors) concerns a cation-selective sensor that has a cation-selective coating and in which the ions to be analyzed cause a detectable change in the electrical properties of the coating. Due to an acid/base component in the cation-selective, preferably ionophoric layer, the sensor acts independently of the anions

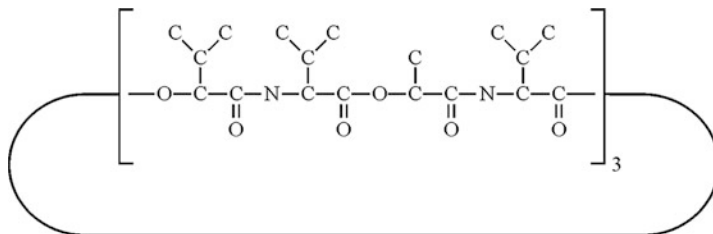


Fig. 11.20 Alinomycin as an electrode-active compound in potassium-selective electrodes

present in the solution to be analyzed. In this way, the measuring accuracy is to be improved and the detection limit reduced.

Finally, recent research on the use of *cyclodextrins* and *calixarenes* as complex-forming functional elements in ISE should be mentioned here. An example of this is the calix[4]arene crown-6 for the determination of *cesium ions*.

Several anodes and cations are accessible to sensory measurement by using ISE with single crystal and precipitation membranes. For example, the fact that the solubility product of LaF_3 at $\approx 10^{-29}$ provides the best conditions for the largely interference-free determination of fluoride activities $>10^{-7}$ mol/l using appropriate single crystals (which are doped with 0.1 mol-% EuF_2 to reduce the internal resistance of the sensor). For reasons of reversibility of all phase boundaries, it has proven to be particularly advantageous to design the constructive structure of this fixed-contact ISE as follows (Fig. 11.21):

For all other halides and a number of other anions, such as sulphides, thiocyanates, cyanides as well as cations, such as Ag^+ , Cu^+ , Cu^{2+} , Pb^{2+} or Cd^{2+} , electrodes with so-called *precipitation membranes* (poorly soluble salts) are often used. If the membranes are designed as pellets, a distinction is made between *homogeneous compacts* and *heterogeneous compacts*. In the case of the chloride-selective electrode, the homogeneous compact consists of $AgCl$ and the mixed compact of $AgCl$ and Ag_2S . The latter has better mechanical properties, lower light sensitivity and less cross-sensitivity, for example to redox systems. An extension of the lifetime of ISE with

Precipitation membranes can be achieved as shown in Fig. 11.22 if, in analogy to the filling of resins with KCl described above, silver chloride is introduced into the uncured resin starting mass for the purpose of producing solid reference electrodes. The service life is reduced, for example, in strongly flowing media, with high chloride contents in the analyte and above all at higher media temperatures.

Another possibility for the potentiometric determination of ions is the use of chalcogenide glasses as selective membranes. If the elements sulphur, selenium or tellurium are introduced into glasses instead of oxygen and (or) silicon is replaced by germanium or arsenic, one arrives at material systems with a partially pronounced glass formation tendency, known as chalcogenide glasses. The glasses are produced by melting the respective pure components together in cylindrical, evacuated quartz ampoules and quenching in cold water. To produce the electrodes, the resulting glass pieces are

Fig. 11.21 Schematic structure of a fluoride electrode based on a lanthanum fluoride single crystal (Source: T.A. Fjedley, K. Nagy: Journ. Electrochem. Soc. 127 (180) 1299)

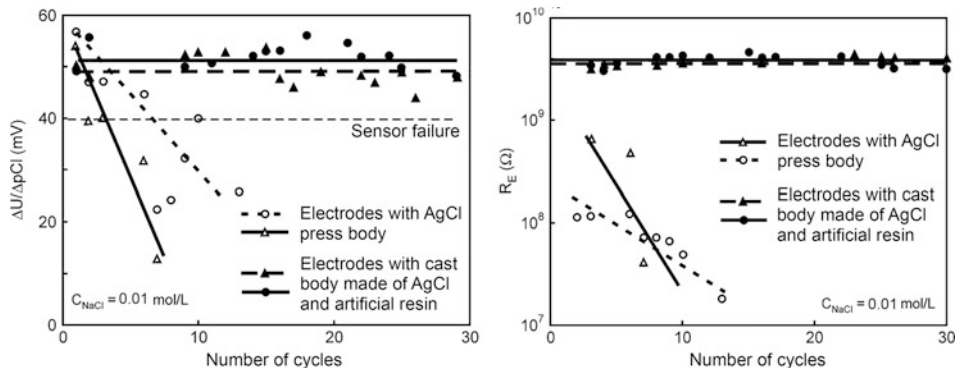
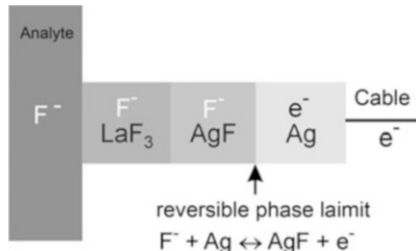


Fig. 11.22 Sensitivity (left) and resistance (right) of chloride-sensitive electrodes in use in NaCl solution at $\vartheta = 80^\circ\text{C}$ (Source: J. Zosel, Kurt-Schwabe-Institut Meinsberg)

embedded in resin and contacted. For reasons of increasing potential stability, it has proven advantageous to form an intermediate layer consisting of polypyrrole and the polymer cation exchanger Nafion[®] on the glass surface instead of the previously used bonding of arrestor wires with gold or silver conductive lacquers and to contact a platinum wire on top of it. This advantage is shown in Fig. 11.23 for a silver ion-selective chalcogenide glass electrode of the composition $\text{Ag}_{17}\text{As}_{25}\text{S}_{58}$. Chalcogenide-glass-based ISE can preferably be applied for the determination of monovalent and multivalent cations, such as Tl^+ , Cu^{2+} , Cd^{2+} , Pb^{2+} and Fe^{3+} , for example in wastewater, whereby pH dependencies and interfering ion influences must be taken into account to varying degrees. Recently, planar chalcogenide glass-based multisensor arrays have also been reported, in which the glass layers are stoichiometrically deposited onto passivated silicon substrates by means of pulsed laser deposition (PLD).

It should also be pointed out that electrodes with passivated metals of the IVth to VIth subgroup of the elements of the periodic table can be used as n-semiconducting electrode membranes, which are sometimes also called *partially selective redox electrodes*. For example, it is possible to prepare an electrode material sensitive to the Fe^{3+} determination by repeated successive anodic passivation of titanium semi-finished products with a purity of at least 99.99% in 50 % H_2SO_4 at a polarization voltage of $U_p = 4,3\text{ V}$ vs. SSE at $\vartheta = 25^\circ\text{C}$. However, the domain of partially selective redox electrodes is not so much ion

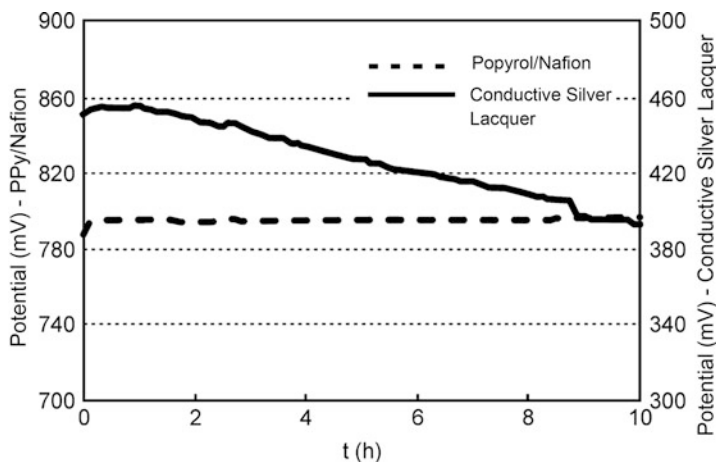


Fig. 11.23 Potential adjustment behavior of a silver ion-selective chalcogenide glass electrode with different potential derivation (Source: W. Vonau, Kurt-Schwabe-Institut Meinsberg)

analysis as the *detection of oxidizing agents* such as chlorine and hydrogen peroxide. Here too, the pH dependence of the sensor signal must be taken into account.

In summary, it can be stated that there are alternative chemosensory determination methods for a number of ions, in some cases also. Despite the greatest efforts, it has not yet been possible to provide such methods for the important anions sulfate and phosphate, which is an important research task for the future.

11.3 Gases

11.3.1 General Information

Many gases have a solubility in liquids so that their sensory determination is often also of importance as a *physically dissolved species*. For example, there is interest in the *oxygen content* in fish farms, surface, process, waste and boiler feed water or in the *carbon dioxide content* in mineral water sources, cell culture media and beverages. Both substances are of course also components of air and other gas mixtures and are therefore of chemical-analytical interest there as well. While, for example, the amperometric Clark sensors mentioned in Sect. 2.16.3 for oxygen determination and the potentiometric sensors operating according to the Severinghaus principle, which are used, among other things, for carbon dioxide determination, can be used in both states of aggregation because of their membrane covering, there are other electrochemical sensor types which can be used exclusively in the gas phase. These are also presented below.

11.3.2 Gases in a Physically Dissolved State or at Normal Temperature

The *gas solubility* in liquids depends, among other things, on temperature, (air) pressure and the presence of other dissolved species, which is particularly illustrated by Henry's law.

For the amperometric determination of oxygen, the voltam-mogram shown in Fig. 11.24, from which the well-defined plateau of the diffusion limit current in the range from -600 mV to -900 mV is clearly recognizable, is an essential basis. Eq. (11.12) shows the relationship between diffusion limit current density i_D and oxygen partial pressure p_{O_2} in the measuring medium.

$$i_D = Kp_{O_2} + i_0 \text{ with } K = zFP(T)/d \quad (11.12)$$

z : Number of electrons converted

F : FARADAY constant

T : Absolute temperature

d : Thickness of the permeable membrane

i_0 : Residual flow of the oxygen sensor at the oxygen partial pressure $p_{O_2} = 0$

$P(T)$: Permeation coefficient of the membrane.

In analogy to the amperometric oxygen electrode, other electrodes with adapted gas-permeable membranes are also available. Also of great importance in this context are, for example, *chlorine sensors* for the determination of active chlorine already discussed in the case of redox electrodes, for example in swimming pools, *H₂ sensors* for use in biogas plants or *CO₂*-, *H₂S*-, *SO₂*-, *HCN*- and *NO₂*- sensors. The amperometric measuring instruments must supply the required polarisation voltage and their circuit electronics must be matched to the sizes of the sensor electrodes. Under certain circumstances, the gas-permeable membrane is not used in the sensor construction. These are then called *open systems* or sensors with a *free working electrode*. The working electrode is then in direct contact with the medium to be measured so that undesired side reactions can occur in addition to the analytically desired electrode reaction due to other solution components. Changes in conductivity also influence the internal resistance of the sensor, so that the application area of these sensors is limited to clean solutions.

In addition, the *galvanic sensor* as shown in Fig. 11.25 plays a role in addition to the amperometric sensor, especially for *O₂* determination in the normal temperature range. If the potassium hydroxide contained in the sensor comes into contact with oxygen, a chemical reaction occurs. This results in an electric current which flows through a load resistor between a lead anode and a gold-plated cathode. The current produced is proportional to the oxygen concentration present.

The function of the electrochemical carbon dioxide sensor is based on the potentiometric measuring principle. *CO₂* permeates from the measuring medium through a thin polymer

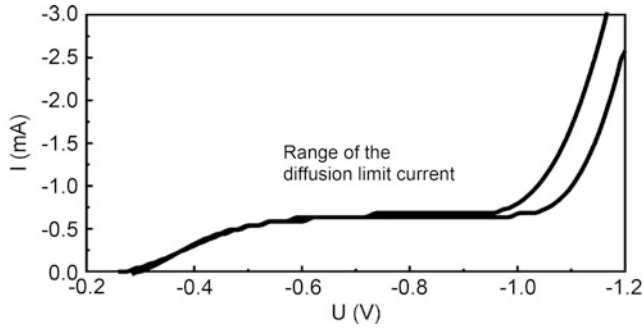
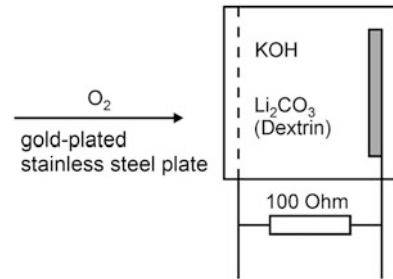
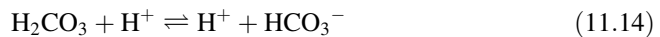


Fig. 11.24 Voltammogram of a membrane-covered oxygen sensor with microcathode (two-electrode system); measurement in distilled water at $\vartheta = 23^\circ\text{C}$, scanning rate: 0.5 mV/s

Fig. 11.25 Galvanic oxygen sensor (Anode reaction: $4\text{OH}^- + 2\text{Pb} \rightarrow \text{PbO} + 2\text{H}_2\text{O} + 4e^-$)



membrane into the carbonate-containing sensor electrolyte and causes a reproducible change in the pH value, which is measured with a pH glass electrode, for example. After calibration, the sensor voltage signal is proportional to the logarithm of the carbon dioxide concentration in the measuring medium over a wide concentration range. The following chemical equations form the basis for the chemosensory CO_2 determination (see also Sect. 2.16.2):



Other gases, such as ammonia, can also be determined according to the same principle. Figure 11.26 shows an NH_3 sensor and Fig. 11.27 as an application example of this sensor. Measurement results during fermentation with *Aspergillus niger* in different media.

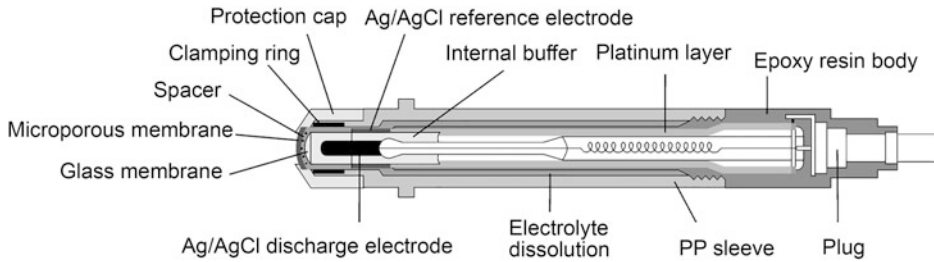


Fig. 11.26 Potentiometric NH_3 sensor

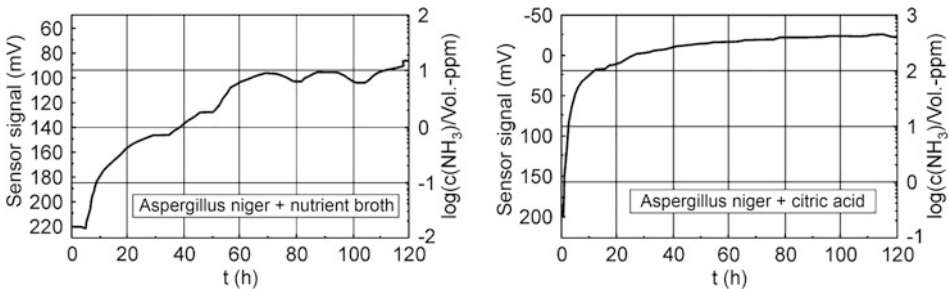


Fig. 11.27 NH_3 emission during fermentation with “*Aspergillus niger*”; left: in nutrient solution; right: in citric acid (Source: U. Enseleit, Kurt-Schwabe-Institut Meinsberg)

11.3.2.1 Solid Electrolyte Sensors

Classification and Meaning

Together with semiconductor gas sensors (MOS), resistive sensors and pellistors, solid electrolyte gas sensors are often referred to as *solid-state gas sensors* or *ceramic gas sensors*, sometimes also called *high-temperature gas sensors*.

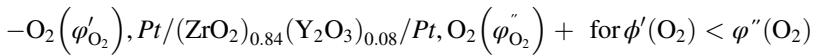
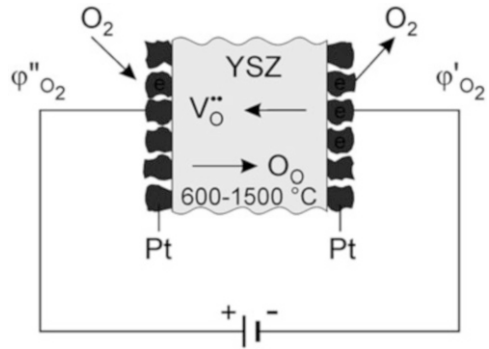
Electrochemical Cells with Solid Electrolytes

Potentiometric Oxygen Measurement

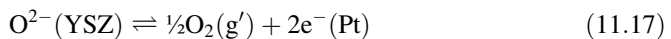
Solid electrolyte gas sensors in the potentiometric mode of operation take today a leading position among all chemical sensors, both in terms of the number produced and the effects resulting from their application. The basis of potentiometric oxygen measurement is the *currentless voltage measurement* on an oxygen concentration cell with a solid oxide ion-conducting electrolyte (Fig. 11.28). $V_{\text{O}^{\circ\circ}}$ (vacancy) refers to oxygen ion or oxide ion vacancies through which oxygen can migrate as an oxide ion (O^{2-}) at higher temperatures. In contrast to aqueous electrolytes, the charge transport is thus carried out by a *type of ion*. This phenomenon is called *unipolar ion conduction*.

The oxygen concentration cell can be described by the following cell symbol

Fig. 11.28 Oxygen concentration cell (Source: U. Guth, Gas sensors in: A. J. Bard, G. Inzelt und F. Scholz (Eds): Electrochemical Dictionary, Springer Berlin, Heidelberg (2008), 294-299)



Oxygen on the higher concentration side (the higher chemical potential) has the tendency to migrate to the side with the lower concentration. However, this is *not possible* with a gas-tight solid electrolyte. However, the electrochemical potentials can balance each other out by oxygen from the side with the higher partial pressure ($p\text{O}_2''$) taking up electrons from the electron conductor, for example, platinum (reduction, cathode), migrating as an oxidation ion (O^{2-}) via the oxygen vacancies $\text{VO}^{\circ\circ}$ through the oxidation ion-conducting solid electrolyte and leaving behind electrons on the other side with lower partial pressure ($p\text{O}_2'$) as molecular oxygen (oxidation, anode). Only a differentially small current flows until the electrochemical equilibrium is established, for example after a change in the oxygen concentration; the material conversion is almost zero. These processes can be described in formulas as follows:



As a result of the electrochemical reaction, a potential difference or an equilibrium voltage U_{eq} (or an electrochemical force $-E$) can be measured between the two electrodes with a voltmeter of large input impedance (quasi-current free), which is proportional to the logarithm of the oxygen partial pressure $p(\text{O}_2)$ according to the Nernst equation.

$$-E = U_{\text{eq}} = \frac{RT}{4F} \ln \frac{p''_{\text{O}_2}}{p'_{\text{O}_2}}. \quad (11.19)$$

With the same total pressure in both gas spaces ($p'' = p'$) and isothermal conditions ($T'' = T'$), the partial pressure can be replaced by the volume concentration ϕ_{O_2} or the mole fraction x_{O_2} and is obtained with the gas constant R and the Faraday constant F

$$\phi_{O_2} = \frac{p_{O_2}}{p} = \frac{v_{O_2}}{v} = x_{O_2} \quad (11.20)$$

$$-E = U_{eq} = \frac{RT}{4F} \ln \frac{\phi_{O_2}'' p''}{\phi_{O_2}' p'}$$

Formulation in size equations:

$$U_{eq} = \frac{8.314 \text{ mV Asmol}^{-1} 2.303 \frac{T}{K}}{4.96483 \text{ Asmol}^{-1}} \lg \frac{\phi_{O_2}''}{\phi_{O_2}'} = 0.0496 \frac{T}{K} \lg \frac{\phi_{O_2}''}{\phi_{O_2}'} \quad (11.21)$$

If air ($\phi_{O_2}'' = 20.63 \text{ vol.}\% \text{ O}_2$ at 50 % r.F.) is used as the reference gas, the following results are obtained

$$\lg \phi_{O_2} / \text{vol.}\% = \left[\frac{-U_{eq} / \text{mV}}{0.0496 T / K} \right] + 1.321 \quad (11.22)$$

$$\phi_{O_2} = 20.63 \exp \left[\frac{-46.42 U / \text{mV}}{T / K} \right] \quad (11.23)$$

Due to the logarithmic dependence of the voltage on the concentration, there is no concentration zero. Very small oxygen concentrations ($\phi_{O_2} < \text{about } 0.1 \text{ vol.}\text{-ppm}$, $p_{O_2} \approx 10^{-7} \text{ bar}$ at 500°C) can no longer be achieved freely, for example with the aid of gas mixtures. The oxygen concentration is determined by chemical equilibria such as gas dissociation equilibria, which are established at the hot platinum electrodes:



Using the example of water vapour dissociation, the calculation of the thermodynamic stress will be explained in more detail. According to the law of mass action, the equilibrium oxygen pressure at $T = \text{const}$:

$$p_{\text{O}_2}^{1/2} = K_p \frac{p_{\text{H}_2\text{O}}}{p_{\text{H}_2}}. \quad (11.26)$$

With the temperature dependence of the equilibrium constants

$$\log K_p = 2.947 - 13.008 \text{ K}/T \quad (11.27)$$

then results in the Nernst equation:

$$U_{\text{eq}}/mV = 0.0496 \frac{T}{K} \lg \left[K_p \left(\frac{p_{\text{H}_2\text{O}}}{p_{\text{H}_2}} \right) \right]^2 \quad (11.28)$$

$$U_{\text{eq}}/mV = -1290.6 + \left[0.2924 - 0.0992 \log \left(\frac{p_{\text{H}_2\text{O}}}{p_{\text{H}_2}} \right) \right] T/K. \quad (11.29)$$

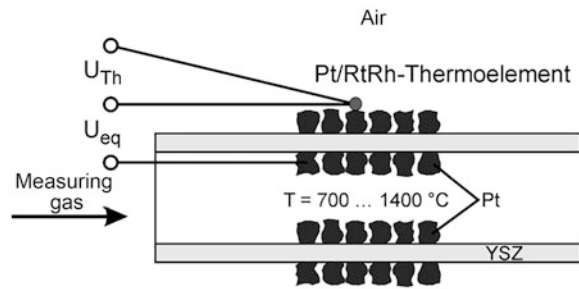
The cell voltage U_{eq} [or $-E$ (EMF)] is therefore determined by the ratio of the partial pressures. The *drier* the hydrogen, the *greater* the *absolute cell voltage*. At a constant water vapour/hydrogen ratio, the cell voltage decreases with increasing temperature.

Oxygen in Exhaust Gases

Of the principles practiced, the most common are the potentiometric sensors for “free” oxygen and equilibrium oxygen from largely adjusted gas equilibria at the electrodes. Their functionality in a wide range of temperature (400°C to 1600°C) and gas partial pressure (10 bar to 10^{-20} bar) allows their application in many high-temperature processes and thus the direct (in *situ*, *i.e. directly at the process location or in the process*) determination of process-relevant parameters in *real-time*. The main application is the *fast oxygen measurement in combustion exhaust gases* of power and heating plants. By linking information from solid electrolyte gas sensors with stoichiometric and thermodynamic relationships for gas reactions, the complex description of the state of gas phases is possible.

The sensors that can be used directly (in *situ*) in the high-temperature process and are equipped with a ceramic or metallic protective tube are called *probes* (for example, see Fig. 11.29). In the untreated exhaust gas, these often have to withstand extremely high temperatures (up to 1500 °C), mechanical loads from high exhaust gas flows and dust particles, but also chemical influences from reducing and corrosive gas components over several years without noticeable aging. If possible, the probes should be operated without calibration, that is, the measurement signal must not change over time and thus simulate a different exhaust gas composition. The great *advantage* of potentiometric oxygen measurement with solid electrolyte cells is the *high selectivity*, which is due to the defined electrochemical process on the one hand and the high operating temperature on the other. The latter ensures that any reaction products are not permanently adsorbed on the sensor surface and poison it. The *oxygen partial pressure* is measured, which has a

Fig. 11.29 Schematic structure of a tubular cell



thermodynamic effect at the measuring temperature. Gas components such as water, CO_2 do not influence the measurement signal.

Initially, the use of such sensors helped to significantly improve the quality of combustion control, and they have been used millions of times as *lambda sensors* in automotive engineering. Today, other high-temperature processes such as the optimization of burners, the production of glass or porcelain and the tempering of steels (carburization) as well as the determination of the oxygen dissolved in liquid steel are the main fields of application and further development.

The solid electrolyte most commonly used today is ZrO_2 , stabilized with Y_2O_3 , in the form of tubes and discs, which becomes noticeably electrically (ionically) conductive at a temperature $>400^\circ\text{C}$. The electron conductivity, which is smaller by orders of magnitude, and which is produced by oxygen expansion under extremely reducing conditions or by oxygen incorporation under strongly oxidizing conditions, is not important for most applications. A very simple structure with completely separated gas chambers is obtained with a tube open on both sides (Fig. 11.29). The solid electrolyte tube with platinum layers on the outside and inside is heated to 700°C to 900°C with an electric furnace. Sample gas flows through the tube while the measuring cell is surrounded by still air from the outside. This design is suitable for gas analyses in which a partial flow is taken and fed to the solid electrolyte cell, for example for determining the residual oxygen in protective gases.

For flue gas sensors, the solid electrolyte cell with a ceramic or metallic protection tube is positioned directly as a probe in the flue gas duct. If the temperature in the duct is sufficiently high ($>500^\circ\text{C}$), heating is usually not necessary. The reference air is passed through the inside of the tube, while flue gas is passed to the outer electrode by flow and diffusion. The measurement result is not distorted by introducing the reference air into the flue gas. Solid electrolyte tubes closed on one side or metal tubes with a solid electrolyte disk soldered in at the end are often used for this purpose.

Controlling high-temperature processes with the help of solid electrolyte probes has become established in many areas of industry. For example, the colour of bricks can be influenced by a specific composition of the fuel gas atmosphere. Using a voltage (EMF) temperature diagram, the thermodynamically stable ranges (calculated and measured) of metal oxides can be limited (Fig. 11.30). The voltage of an oxygen concentration cell is plotted against the temperature. The negative decadic logarithm of the oxygen partial

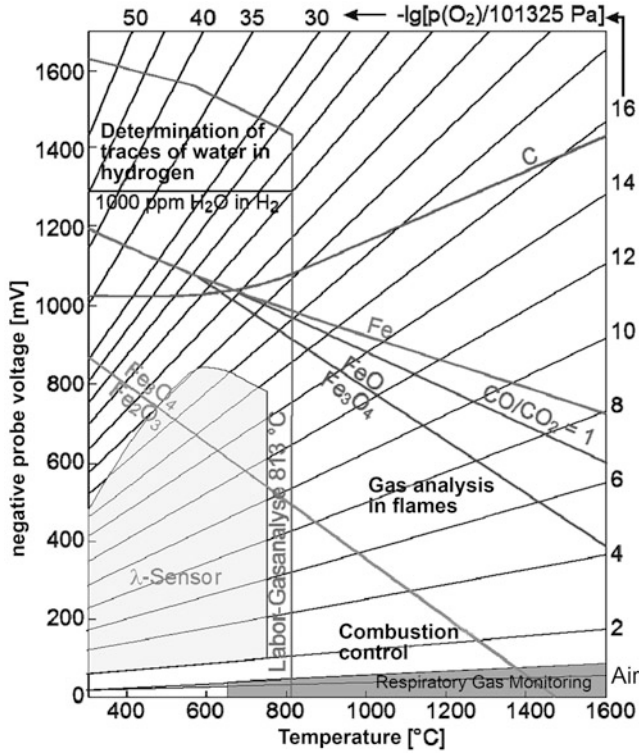


Fig. 11.30 Voltage-temperature diagram for potentiometric solid electrolyte oxygen sensors

pressure is entered as the parameter. The rising lines are calculated according to the Nernst equation if the respective oxygen partial pressure (e.g. 10^{-6} atm and air at a certain temperature: 495 mV at 1600 °C) is used.

If, for example, 400 mV is measured in a ceramic kiln with a solid electrolyte oxygen probe in a gas mixture containing CO and H₂ at 800°C, one can be sure that the iron in the ceramic is present as Fe₂O₃, which gives it a red colour. If, on the other hand, 700 mV is measured, Fe₃O₄ is thermodynamically stable, which in turn colours the ceramic grey to black. The reduction capacity is determined by the ratio of reducing to oxidizing gases. With a CO/CO₂ mixture of 1:1, FeO cannot be reduced to Fe at temperatures >800°C. The diagram is also useful to illustrate the fields of application for oxygen probes. Further fields of application, such as carburizing and nitriding of steel surfaces, can be found in the indicated literature.

The example of combustion control will be used to show how such sensors can be used to optimize technological processes. Minimizing fuel consumption and reducing CO₂ and pollutant emissions from technological processes are more important today than ever before in view of climate targets and environmental pollution. For example, more than 30 % of the emission load in urban areas is caused by small and medium-sized combustion

Lambda-Sonde im Abgasrohr (Prinzip)

- 1 Sondenkeramik, 2 Elektroden
- 3 Kontakt, 4 Gehäusekontaktingung
- 5 Abgasrohr, 6 keramische Schutzschicht (Porös)

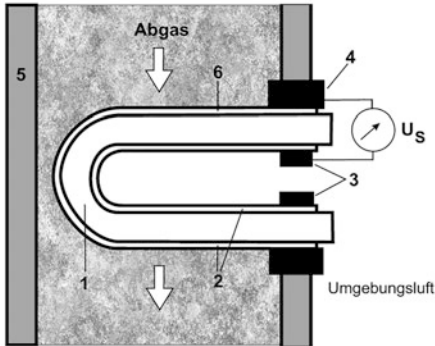


Fig. 11.31 Lambda sensor: Diagram of the sensor and component (factory photo: REGETEC)

plants during the heating season. The main emitters are combustion plants that have an insufficient combustion quality due to malfunctions or incorrect settings. Insufficient air supply leads to incomplete combustion and thus to increased emissions of unburned pollutants (soot, CO and HC). On the other hand, too high an air concentration reduces efficiency and increases NO_x emissions. The continuous control and adjustment of the optimum air volume for combustion, which mainly depends on the type of fuel and burner, is advantageously carried out by means of a fast *in situ oxygen measurement* in the exhaust gas with ceramic ZrO_2 gas sensors.

Lambda Sensor

Solid electrolyte sensors have gained great importance as lambda sensors for the catalytic exhaust gas treatment of gasoline engines. These sensors are produced in their millions as so-called thimble sensors or planar broadband sensors and are used in practically every automobile with a gasoline engine. Via the engine management system, the signal from the lambda sensor contributes significantly to optimizing combustion and reducing pollutant emissions. The measuring principle was explained in detail with the help of Fig. 11.28. Figure 11.31 shows a schematic cross-section and the sensor ready for installation.

Figure 11.32 shows the voltage curve of a lambda sensor in more detail. The continuous curve shows the voltage curve of a lambda sensor during the combustion of a hydrocarbon, for example with propene as model gas, as a function of the fuel gas concentration. The dashed vertical line marks the value $\lambda=1$. The air factor λ is defined as the ratio of the actual air quantity used in combustion to the air quantity required for stoichiometric combustion [$\lambda = m(\text{air})/m(\text{air, stoichiom.})$].

For lean mixtures, that is, with an excess of air ($\lambda > 1$), the stress is small, whereas for excess fuel gas ($\lambda < 1$) the stress is large. At the *stoichiometric point* ($\lambda = 1$) the stress

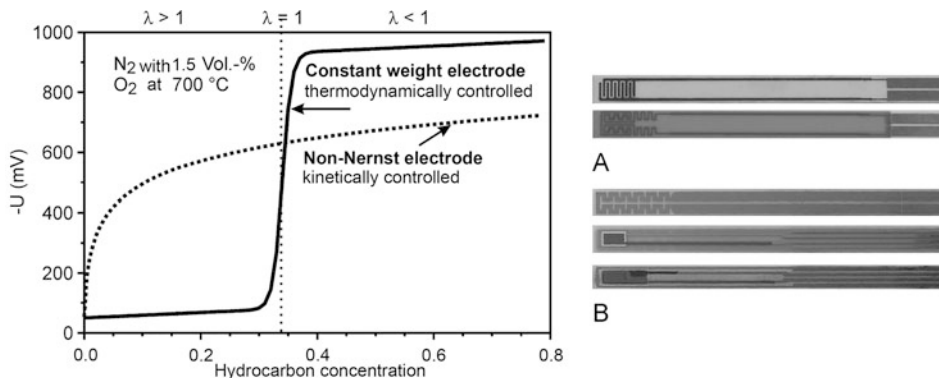


Fig. 11.32 Left: Concentration dependence of the voltage at different electrodes; right: Tongue-shaped mixed potential sensor (a) and impedimetric sensor (b), each with integrated heater

changes abruptly because the logarithmic relationship causes large stress changes due to small changes in oxygen concentrations. This voltage jump is used by the engine management system to control the three-way catalytic converter. The air-fuel ratio is changed 10 times a second by means of corresponding actuators (solenoid valves) from excess air to deficient air and thus quasi-simultaneously oxidizing (for the oxidation of hydrocarbons and CO) and reducing (for the reduction of nitrogen oxides) ratios are set on the catalyst.

The dashed curve in the picture (Fig. 11.32, left) shows the stress curve when the chemical equilibrium in a gas mixture is *not* established. Sensors that make use of this effect are called *mixed potential sensors* (Fig. 11.32, right) and play a role in the detection of unburned fuel (CO and HCs) but also nitrogen oxides (NO_x). In these sensors, a catalytically active oxygen electrode, for example, platinum/YSZ, is combined with an electrode, Au/YSZ, which is sensitive to fuel gas but inactive for electrochemical oxygen reduction. Both electrodes can be directly exposed to the sample gas or be located in separate gas chambers. As an alternative to measuring the voltage, the complex resistance, the *impedance*, which is measured over a wider frequency range, can also be used for evaluation. The most favourable method is the representation of the imaginary part via the real part of the impedance (the *Nyquist representation*), which usually leads to a semicircle (or several semicircles). Its radius, which depends on the concentration, can be easily determined. Sensors in this mode of operation are referred to as impedimetric sensors, also called "impedancemetric sensors" in the Anglo-Saxon literature (Fig. 11.32, right).

Determination of the Redox Potential of Gases

Since oxygen determination as equilibrium oxygen is also possible in reducing gases in which no free oxygen is present, the H_2/H_2O or CO/CO_2 ratio can also be determined. Through the specific utilization of special electrodes prior to stored reactions (decomposition or oxidation reactions), further gas components coupled with oxygen through reactions can be quantitatively determined, for example



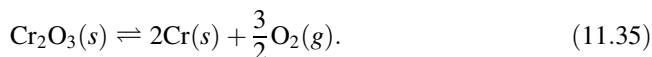
With fast oxygen control, product quality can often be decisively influenced. Examples of this are brick and porcelain production. The “red” of a fired brick, which is due to the oxidation state Fe^{3+} , is obtained with an excess of oxygen while reducing conditions lead to black bricks. The “whiteness” of the porcelain and thus its quality depends on the oxidation state of the traces of iron contained in the kaolin/clay mixture. Reducing firing in the absence of air leads to a reduction of the Fe^{3+} ions into finely distributed Fe^{2+} , which, however, appears white in the matrix of the silicate. By contrast, the oxidizing firing with excess air oxidizes the Fe^{2+} ions to Fe^{3+} , the porcelain body then shows a grey coloration. Therefore, constant product quality can only be guaranteed by adjusting an optimal firing gas/air composition. In the past, the exhaust gas was often extracted via a bypass and the CO or CO_2 concentration was measured using optical methods. Due to this method, a relatively long dead time (approx. 3 min) and a falsification of the measurement result due to the change in gas composition during cooling by condensation of water results. In this case, real-time oxygen determination directly in the firing process with ceramic oxygen measuring probe is also more favourable.

Solid electrolyte sensors have gained great importance in the *control of the surface hardening* of steels. Depending on the respective process, a distinction is made between carburizing with propane at normal pressure, gas nitriding and gas nitrocarburizing as well as the more recently established process of *low-pressure carburizing* with *ethine*. In a multi-component system in non-equilibrium, valuable information can be obtained with electrodes on which preferably one component is converted. In this way, it is possible to determine NH_3 with solid electrolyte cells, for example.

Measurement of Oxygen in Liquid Metals

An important application of potentiometric ZrO_2 solid electrolyte sensors, which cannot be solved with conventional methods, is the determination of the oxygen concentration in molten metals. It is well known that the carbon content of steels is decisive for the material properties. The carbon concentration, but also the content of Si, Mn, P, S can be precisely adjusted within small tolerances by the targeted addition of oxygen. Too much oxygen leads to a deterioration of the steel quality. In order to determine the optimum, the rapid measurement of the oxygen concentration is necessary, which is carried out alternatively with immersible ZrO_2 solid electrolyte cells. Instead of air, a fixed reference system is used

as a reference system, a metal, metal oxide mixture, for example, Cr , Cr_2O_3 , with which a certain, thermodynamically defined equilibrium oxygen partial pressure can be set according to the temperature.



In the case of pure substances, the mole fraction is $x_S = 1$ and you get

$$K_p(T) = p_{O_2}^{3/2} \quad (11.36)$$

$$p_{O_2}(T) = \text{const.} \quad (11.37)$$

The measuring cell (Fig. 11.33), whose second pole represents the liquid metal, is located at the tip of a meter-long cardboard sleeve, which burns during the measuring process. Such probes, which can be immersed in the liquid steel, only work for a few seconds until they dissolve, but during this time they supply the necessary information to control the freshness process. These probes are produced in large quantities and are now used worldwide in steel production. Oxygen is similarly determined in liquid copper.

Current Carrying Solid Electrolyte Gas Sensors

Amperometric Solid Electrolyte Gas Sensors

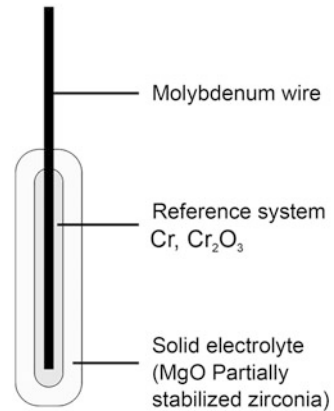
Oxygen sensors with oxide ion conducting electrolytes can also be operated amperometrically. A voltage is applied across the electrodes so high that all the oxygen reaching the cathode is reduced. Due to the diffusion-controlled transport through the porous electrode, the current flowing is proportional to the oxygen concentration. Large oxygen concentrations can be measured if a diffusion-limiting layer, for example, a porous ceramic or a capillary, is placed in front of the cathode. These layers ensure that a partial oxygen current proportional to the total concentration reaches the cathode and causes an electrical limit current I_{lim} there, which is independent of the applied polarization voltage over a wide range.

$$I_{\text{lim}} = -\frac{4FD_{O_2}A}{RTl} \quad (11.38)$$

Where A is the D_{O_2} diffusion coefficient, A is the diffusion cross-section and l is the diffusion length.

If two differently polarized electrodes are used one behind the other, amperometric O_2 and NO_x or O_2 and HC can be measured simultaneously.

Fig. 11.33 Schematic structure of a measuring probe for determining oxygen in liquid steel



Coulometric Determination of Oxygen

With potentiometric oxygen measurement, only the quotient between the measured partial pressure and the reference partial pressure can be determined. Under reducing conditions the quotient of the partial pressures between unburned and burned gas components is obtained $\frac{p_{H_2} p_{CO}}{p_{H_2O} p_{CO_2}}$.

For a complete gas analysis, it is, therefore, necessary to know further parameters such as the initial concentration or the flow velocity. The coulometric determination of oxygen, on the other hand, works, that is, the absolute mass of oxygen is obtained according to FARADAY'S law:

$$It = nzF = m/MzF \quad (11.39)$$

Where $n = m/M$ is the amount of substance z is the number of electric charges that have flowed and F is the Faraday constant. The principle is illustrated in Fig. 11.34.

The gas to be analyzed passes through a solid electrolyte electrolysis cell. This consists of a solid electrolyte tube, which is provided with a platinum layer on the inside and outside. A voltage is applied via these electrodes, at which oxygen is removed from the gas flow, for example. The charge $q = I dt$ that has flowed is proportional to the amount of oxygen transported. The current flowing is measured as a function of time. With the values for the constants

$$k = \frac{m}{It} = \frac{M(O_2)}{zF} = \frac{32 \text{ g/mol}}{496483 \text{ As/mol}} = 82.91 \cdot 10^{-6} \text{ gO}_2/\text{As} \quad (11.40)$$

is obtained for $k = 82.9 \mu\text{g O}_2/\text{As}$, that is, 82.9 μg oxygen is transported per As.

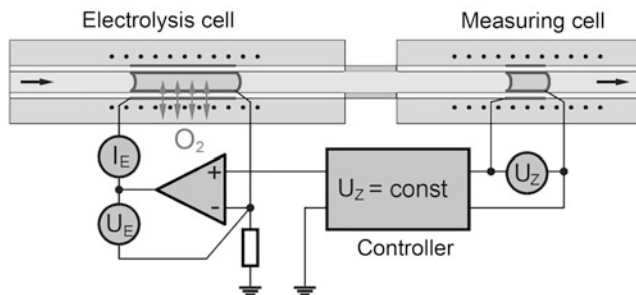


Fig. 11.34 Principle of coulometric oxygen determination

$$\Delta m = k \int_0^t (I_0 - I_t) dt \quad < \text{for } U_{\text{measuring cell}} = \text{const.} > \quad (11.41)$$

The *degree of oxygen removal* can be adjusted and controlled with the aid of the measuring cell. Usually, a control voltage of 400 mV is used, which is measured at the analysis cell at 750°C. Using Eq. (11.40), the oxygen concentration is approximately $3 \cdot 10^{-7}$ Vol.-% by volume. Up to this concentration, oxygen is electrolysed from the gas stream. In addition to measuring pure oxygen in gases, this method can be used to determine the temperature-dependent oxygen uptake or release of substances by heating them and measuring the “oxygen metabolism”. In this way, small deviations from the stoichiometry of oxidic systems δ can be determined (e.g. $2-\delta$ for $\text{SnO}_{2-\delta}$ 0.197). These deviations are decisive for the properties of catalysts, ceramic membranes and electrode materials for high-temperature fuel cells (SOFC). The method is also used to determine the *gas permeability* (permeation) of plastics and the oxygen absorption of adsorbents. It provides similar results to *thermogravimetry* (thermobalance) with the difference that the latter is used to determine unspecific mass changes. In a continuous gas flow, a combination of coulometric and potentiometric oxygen sensors can be used to titrate oxygen in or out to a specific concentration.

Gas coulometry is also suitable for producing *test gases* with specific and constant oxygen concentration.

11.3.3 Semiconductor Gas Sensors: Metal Oxide Semiconductor Sensors (MOS)

11.3.3.1 Surface Conductivity

The semiconductor gas sensors (MOS, metal oxide sensor), also known as *Tagushi* or *Figaro gas sensors*, are based on the *chemisorption of organic molecules* on the surface of an oxidic semiconductor, for example, SnO_2 , and the associated change in electron concentration. The mechanism of the change in conductivity is explained by the

displacement of chemisorbed oxygen (O_2^- and O^-) at the surface, which returns its electrons to the solid, thus increasing the electron concentration in the surface layer and thus the conductivity

$$G = \frac{1}{R} \quad (11.42)$$

increased ($G =$ conductance, $R =$ electrical resistance).

In addition to the displacement of oxygen, *oxygen consumption* due to chemical reactions with adsorbed fuel gases on the hot SnO_2 surface also plays a role. Metal oxide sensors react more or less to all reducing gases and are therefore not very selective and are best used where only a certain type of gas is expected. The correlation between gas partial pressure p_{Gas} and sensitivity S is given by the following relationship:

$$S = \frac{G_{Gas} - G_{Air}}{G_{Air}} \quad (11.43)$$

$$S = \text{const} \cdot p_{Gas}^n \quad (11.44)$$

n is the sensitivity exponent and is usually <1 .

These sensors can be operated isothermally at temperatures between 200°C and 450°C . Selectivity can be considerably increased by programmed periodic heating and cooling in the form of a triangular or sine function. Depending on the gas mixture, characteristic conductance temperature curves (CTP) are obtained which are easily reproducible.

These sensors, whose classic structure is shown schematically in Fig. 11.35, can also be manufactured as planar thick-film sensors using screen printing or dispensing technology. They are used for *semi-quantitative determinations* in *gas warning systems*, *smoke detectors*, *air quality sensors* in automobiles and *workplace monitoring*.

11.3.3.2 Volume Conductivity

Semiconducting oxides are also used for conductivity gas sensors, but these react to the respective gas phase at much higher temperatures than those used for MOS by changing the bulk conductivity.

Depending on the supply of oxygen in the gas phase, oxides change their oxygen stoichiometry (Sect. 2.17.4) at temperatures $>500^\circ\text{C}$ by removing or adding oxygen from the crystal lattice. The $SrTiO_3$ is well investigated. Oxygen ion vacancies are created by oxygen expansion and, from a chemical point of view, consist of Ti^{4+} and Ti^{3+} ions or

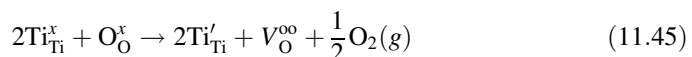
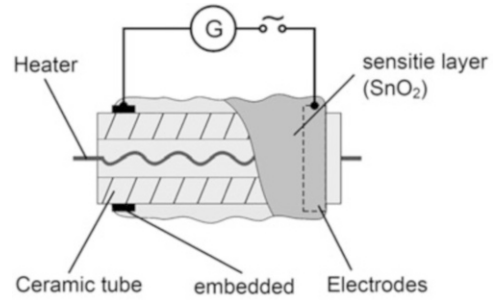


Fig. 11.35 Structure of a semiconductor gas sensor



$[Ti'_{Ti} = e']$ excess electrons, which leads to an increase in electronic conductivity σ_e . The concentration of the excess electrons and thus the conductivity is thermodynamically dependent on the oxygen pressure in the form of a root function and increases with decreasing oxygen partial pressure, for example:

$$\sigma_e = \text{const} \cdot p_{O_2}^{-1/4}. \quad (11.46)$$

Titanates can be doped with aliovalent cations to create certain electronic defects. In the so-called donor doping (doping by a higher valency ion), for example, La^{3+} is introduced into the system $Sr_{1-x}La_xTiO_3$. This means that La^{3+} ions occupy Sr^{2+} places ($La_{Sr}^o = h^o$) and V_{Sr}^{\cdot} vacancies are created. The La^{3+} ions act like electronic defects in the environment of the bivalent Sr^{2+} ions. In the reverse case of acceptor doping (doping by an ion of lower valence) $SrTi_{1-x}Fe_xO_{3-\delta}$, the Ti^{4+} ion is replaced by Fe^{3+} ion, that is, excess electrons are created $Fe'_{Ti} = e'$ and oxygen is expanded to $V_{O}^{o\cdot}$ form oxygen vacancies.

The conductivity is exponentially dependent on the temperature so that the oxygen measurement must be combined with a temperature measurement.

$$\sigma_e = \text{const} \cdot \exp\left(-\frac{E_a}{RT}\right) p_{O_2}^{-1/4}. \quad (11.47)$$

By suitable doping, mixed oxides such as $Sr_{0.95}La_{0.05}(Ti_{0.7}Fe_{0.3})_{0.95}Ga_{0.05}O_{3-\delta}$ were produced, which have only a low-temperature dependence of conductivity, that is, a small activation energy E_a of conductivity. These gas sensors, which are based on changes in volume conductivity, represent an alternative to the solid electrolyte lambda sensors currently in use.

11.3.4 Pellistors

Pellistors manufactured by conventional methods have also found widespread application. These consist of ceramic pellets impregnated with catalysts, in which a platinum filament is

embedded. The latter serves to heat and measure the temperature of the pellistor. The heat generated during catalytic combustion, for example of hydrocarbons, causes an increase in temperature and thus an increase in resistance of the platinum wire, which is measured in a bridge circuit compared to a pellistor without a catalyst layer. In this way, after calibration, a sum parameter is obtained for the concentration of all combustible gases in a gas mixture.

11.4 Electrolytic Conductivity

11.4.1 General Information

All substances that have mobile charge carriers have a finite *ohmic resistance* $R(\Omega)$. This can also be measured as *electrical conductance* G (S) or *electrical conductivity* $\sigma = \kappa$ ($S\text{ cm}^{-1}$). In aqueous electrochemistry, the symbol κ is also often used. The Eqs. (11.42, 11.48 and 11.49) provide the necessary physical relationships to characterize the conductivity of a conductor with length l and cross-sectional area A . Depending on the type of charge carriers causing the conductivity, one speaks of *electronic* (electrons), *electrolytic* or *ionic* (ions) or mixed conductivity (electrons and ions). Exclusively for solids, the *n- or p-semiconductor* is added.

$$R = \rho \frac{l}{A} \quad (11.48)$$

$$R = \sigma \frac{A}{l} \quad (11.49)$$

ρ : specific electrical resistance.

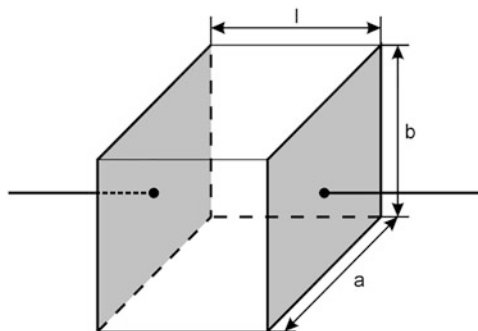
A few examples of the possible applications of conductivity sensors, which are often also referred to as conductivity measuring cells, in the electrochemical analysis are given in Sect. 11.4.5. As already discussed in Sects. 2.16.4 and 11.3.4, conductivity also serves as a *transducer principle* for electrochemical sensor technology.

The physicochemical measurement procedure is called *conductometry* (Sect. 2.16.4). If instead of direct current alternating current resistances are measured, this is called *impedimetry*. Resistances of sensor layers are almost always determined by alternating current measurements.

11.4.2 Kohlrausch Measuring Cells

A *Kohlrausch measuring cell* (schematic representation in Fig. 11.36) consists of two electrodes made of platinum, stainless steel or carbon, for example, square, facing each other at a distance l . An important parameter of conductivity sensors is their cell constant

Fig. 11.36 Scheme for a cabbage noise measurement cell



$$K = \frac{l}{A}. \quad (11.50)$$

For a hollow cube of 1 cm edge length, with two opposite surfaces ($A = a \cdot b$) designed as electrodes, this is, therefore, $K = 1 \text{ cm}^{-1}$. If a conductivity $G = 100$ is determined with a corresponding two-electrode measuring cell, the conductivity of the medium to be examined is $\sigma = 100 \mu\text{Scm}^{-1}$.

Commercially available measuring cells have cell constants in the range from 0.01 cm^{-1} to 100 cm^{-1} , which makes it possible to measure almost the entire conductivity range of aqueous solutions (Fig. 11.37).

In addition to the cell constant, a conductivity measuring cell is characterized by *polarization phenomena* at the electrode surfaces, which means all effects occurring during current flow at the interface between electrode material and analyte that make the conductivity appear smaller or the cell constant larger. These effects can be counteracted within certain limits by increasing the measuring frequency and the electrode area and by selecting the electrode material suitable for the analytical task to be solved.

11.4.3 Multi-Electrode Measuring Cells

The *four-electrode measuring cell*, for which the diagram of a possible design is shown in Fig. 11.38, represents a further development of the carbon noise measuring cell. Two separate current and voltage electrodes are used, whereby the latter act as a kind of currentless potential probe. A metrological control circuit ensures precise adjustment of the current electrodes. The decisive advantage is that polarization effects, which are particularly disturbing at higher conductivities, do not influence the measurement result. Contact resistances occurring at the electrodes due to contamination are also largely compensated. It is important that the voltage electrodes are placed at a location of low current density and that the voltage drop between them is sufficiently high. The electrode

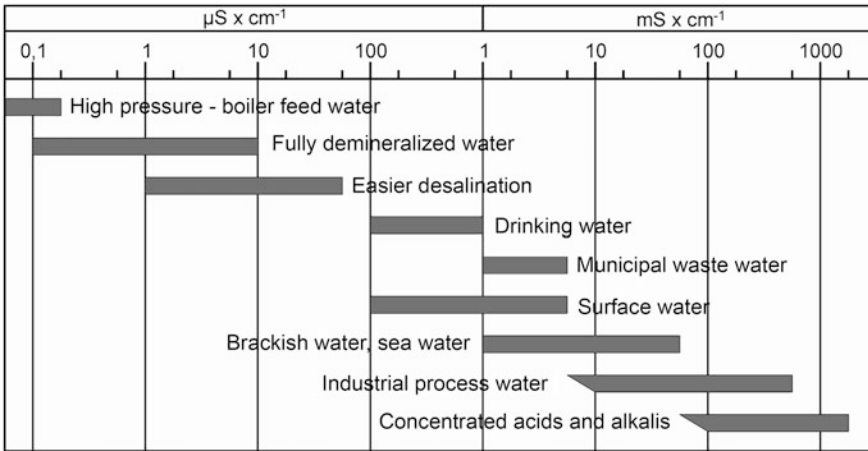
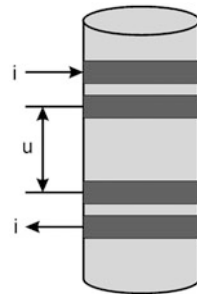


Fig. 11.37 Conductivities of aqueous solutions

Fig. 11.38 Four-electrode measuring cell



potentials of both voltage electrodes must always be identical. A cell constant can be determined using calibration solutions.

11.4.4 Electrodeless Conductivity Measuring Cells

Polarization phenomena can be eliminated by using electrodeless measuring methods. These include *inductive conductivity measurement* with a transmitter as shown in Fig. 11.39, which can operate in the range $\sigma = 1 \text{ mS cm}^{-1}$ to 1.000 mS cm^{-1} . Here, two coils replace the electrodes, one of which acts as an excitation coil.

In this coil flows an alternating current that generates a magnetic field in its environment. The measuring solution is located in the coil core, in which the magnetic field induces the current flow required for the measurement. This in turn generates a further magnetic field, which induces an alternating current with the corresponding voltage in the second coil (receiver coil), which is directly dependent on the current flowing in the

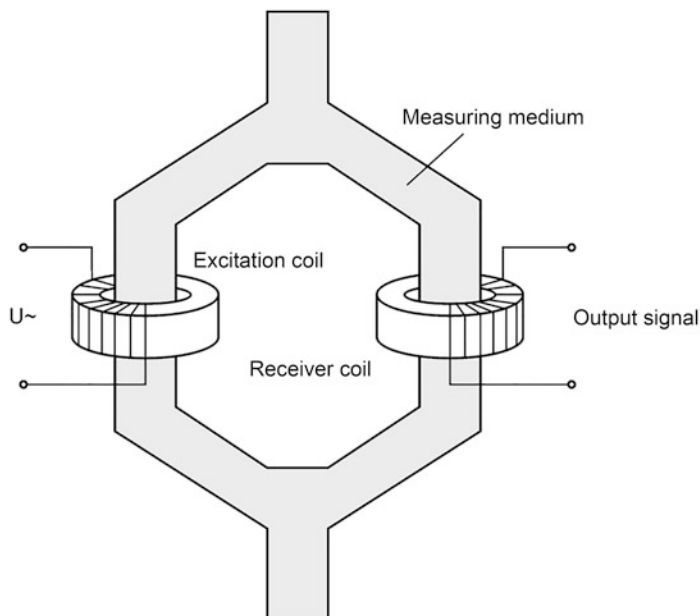


Fig. 11.39 Diagram of an inductive conductivity measuring cell

measuring solution, that is, on the conductivity. Since the magnetic field also acts via a plastic tube, no direct contact of the coils with the analyte is necessary. This *contactless measuring technique*, therefore, offers considerable advantages for measurements in aggressive media. Furthermore, high conductivities *cannot* cause *polarization effects*. However, the design of inductive conductivity measuring cells and the capacitive sensors, which also operate without electrodes, is considerably more complex than conventional systems.

11.4.5 Examples for the Application of Conductivity Sensors

Conductivity (like redox potential) is an *unspecific parameter* (sum parameter) in the spectrum of the chemosensory measurable variables. Nevertheless, the often relatively low effort required to establish appropriate sensor technology or the unavailability of specific analytical technology often justifies its use for the characterization of substance systems or the control and monitoring of process sequences. A few examples are given below.

The *purity* of many media can be detected by conductometric measurements, provided the pure medium has a very low intrinsic conductivity and the impurities are ionogenic. This is the case when monitoring the *desalination status* of water or *checking the purity* of non-aqueous liquids (including chloride determination in crude oil). Conductivity sensors

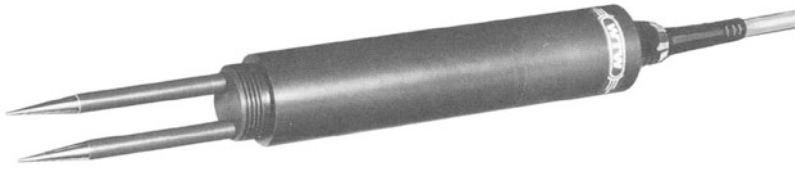


Fig. 11.40 Conductance measuring cell for meat

can also be used in the *operational control* of flushing processes, for example of pipelines in dairies and breweries and of dosing equipment. In analogy to the pH meat probe shown in Fig. 11.40, there are also stainless steel-based insertion probes for determining the conductivity of meat (Fig. 11.40), whereby the degree of progress of the glycolysis occurring after slaughtering can be determined. This makes it possible to differentiate between good, indifferent and deficient meat.

Further possible applications of conductometric sensor technology in the food sector include the differentiation of flower and honeydew honey or the detection of *adulteration* by sugar feeding honey or the *determination of ash* in sugar. Conductivity sensor technology has become indispensable, especially in the medical field. Among other things, it is common practice to determine the *blood loss* during surgery by collecting the blood that has leaked out, adding it, including cotton swabs, to a known amount of distilled water and then determining the conductivity to obtain a direct measure of the blood loss. So-called *lab-on-chip systems*, which are used extensively in medical technology, are often equipped with conductometric or impedimetric sensor structures. Figure 11.41 shows the schematic structure of a planar thin-film measuring cell, including a four-electrode conductivity sensor for the diagnosis of cystic fibrosis. Figure 11.42 shows the typical impedance spectrum recorded with this impedimetric structure of the microsensors system in samples of artificial sweat at different levels of chloride, the marker for this hereditary disease.

Such lab-on-chip systems can also be produced in thick-film technology. Finally, such a lab-on-chip system, which can be integrated into a microfluidic arrangement and is shown in Fig. 11.43, should be mentioned. In addition to several other planar electrochemical sensors, it contains a *precious metal interdigital electrode structure (IDES)* to continuously determine the degree of adhesion of biological cells in the state of cell growth and cell death by means of an impedance measurement. These states can be adjusted by adding culture media and cytotoxics.

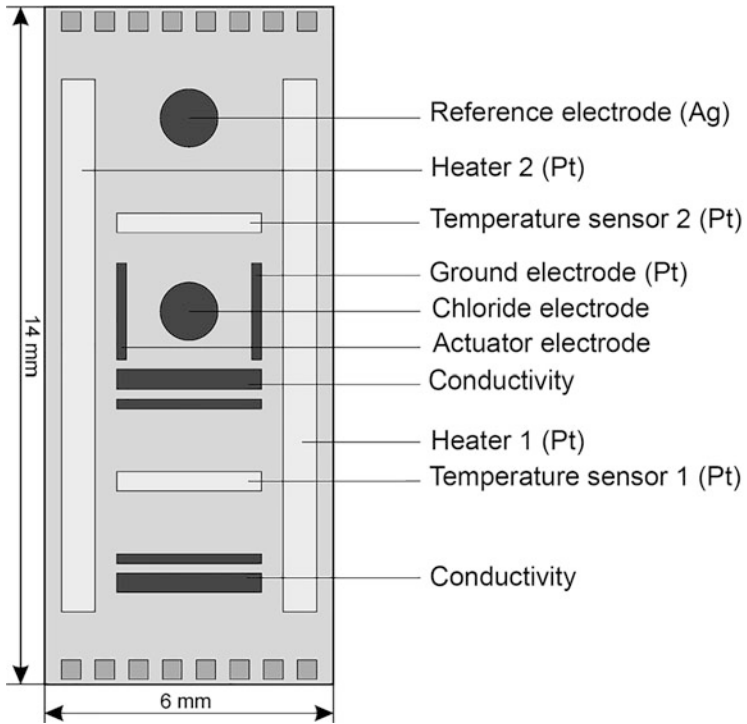


Fig. 11.41 Layout of a multi-sensor chip in thin-film technology for cystic fibrosis diagnostics (Source: S. Herrmann, W. Oelßner, W. Vonau: Entwicklung und Einsatz von Mikrosensorchips zur Schweißanalyse. 12. Heiligenstädter Kolloquium Tagungsband 2004, S. 429-436)

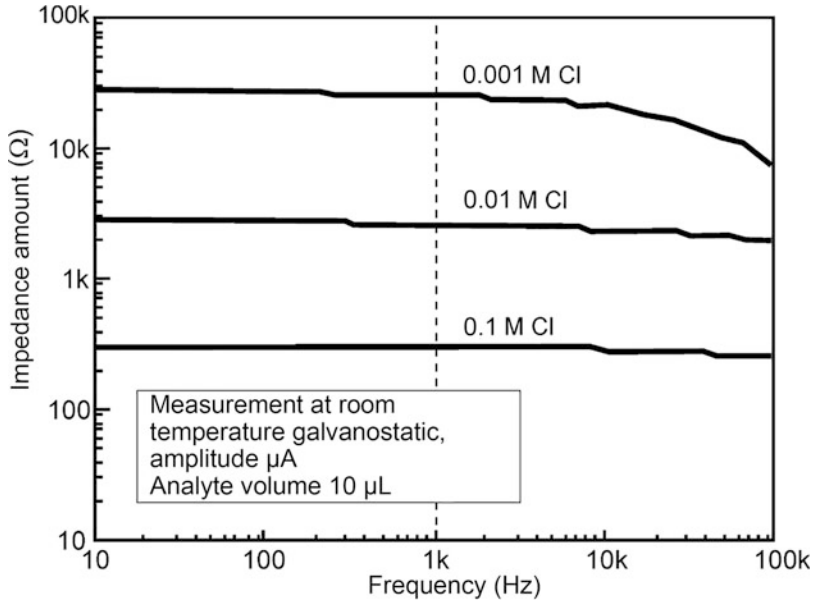
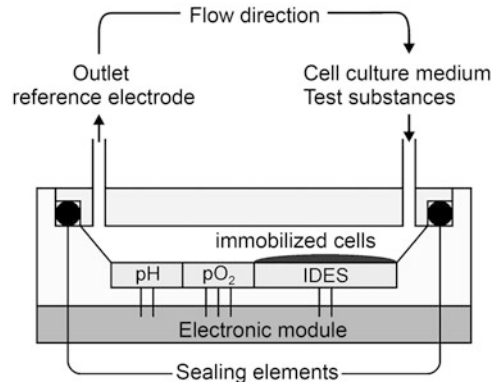


Fig. 11.42 Impedance measurements in artificial welding samples with different chloride contents (Source: W. Vonau, Kurt-Schwabe-Institut Meinsberg)

Fig. 11.43 Microphysiological array in thick-film technology for disposable lab on chip systems with electrochemical sensors (Source: W. Vonau, U. Enseleit, F. Gerlach, S. Herrmann: *Electrochimica Acta* 49 (2004), 3745)





Elfriede Simon

12.1 Biological Sensor Technology

12.1.1 Biosensor Technology

A biosensor is a *complex sensor* that uses *biological or biochemical effects* and converts them into a measurable sensor signal. The *biologically active component* (e.g. enzyme, antibody, oligonucleotide, microorganism, biological receptors) is directly connected to or integrated into a *signal converter (transducer)*. The aim is to generate an electronic signal which is *proportional to* the concentration of a specific substance or a series of substances (*analytes*). Biosensors are special cases of chemical sensors in which a *biomolecule* is used as a recognition element (*receptor*) for the analyte [1].

As a rule, all biosensors can be grouped together into a uniform functional scheme despite different functional principles (Fig. 12.1).

Through the biological element (receptor) it is possible to take advantage of the uniqueness of biological molecule specificity, since problems of selectivity and detection sensitivity arise in the field of analysis, especially at very low concentrations and under difficult environmental conditions (interfering substances).

The following processes run one after the other:

1. *Specific recognition of the analyte*, by complex formation between receptor and analyte.
2. *Conversion of the physicochemical change* (e.g. layer thickness, refractive index, adsorption or electrical charge), which has been caused by the interaction with the receptor, into a measurable electrical signal.

E. Simon (✉)
Siemens AG, München, Germany
e-mail: elfriede.simon@siemens.com

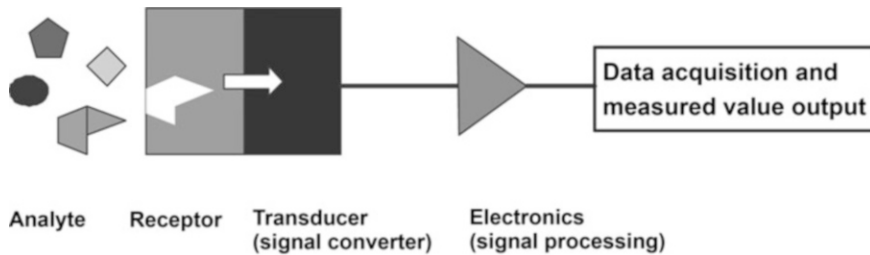


Fig. 12.1 General functional diagram of a biosensor

3. Signal processing by a processor and signal amplification

To restore the initial state of the sensor after the measurement, the binding of the analyte by the receptor must be reversible. Figure 12.2 shows the functional principle of a biosensor based on a heterogeneous system.

In a first step, the receptor molecules are *immobilised* on a surface in a spatially separated manner. After the addition of the analyte or a mixture of analytes, only certain analytes are bound on the surface of the biosensor, which can specifically interact with the receptor. After cleaning the surface, the bound analytes can be detected with or without labelling of the bound complexes, depending on the signal readout method.

The basic prerequisite for the functionality of a biosensor are various biospecific interaction processes. The central reaction of many biological processes is the so-called *key-lock* or *induced fit reaction*, which is based on the selective recognition of two high-molecular components (*ligand* and *receptor*). The interaction between ligand and receptor is based on the non-covalent binding of numerous functional groups, which leads to a stable three-dimensional complex, the so-called *affinity complex*. Since the formation of the affinity complex requires interaction of many individual reactions involving the spatial structure of the components, reactions of this type such as antibody-antigen reaction, protein-protein interaction, DNA hybridisation, DNA-protein interaction or enzyme-receptor reaction exhibit a high degree of specificity. This is used in *affinity analysis* since it is possible to detect one binding partner with the help of the other.

Table 12.1 shows the binding energies of all possible binding types. It shows that a stable affinity complex can only be formed by including as many non-covalent interactions as possible.

A sensor signal generated by the affinity reaction can be detected by various methods (Sect. 12.2): these often include spectroscopic methods such as fluorometry, colorimetry, but also ellipsometry and surface plasmon resonance. Due to the ever-increasing demands on existing analytical methods in terms of sample throughput, sample consumption and time expenditure, the main focus is on the development of methods that exhibit a high *degree of parallelization, miniaturization* and *automation*. Especially in the field of DNA

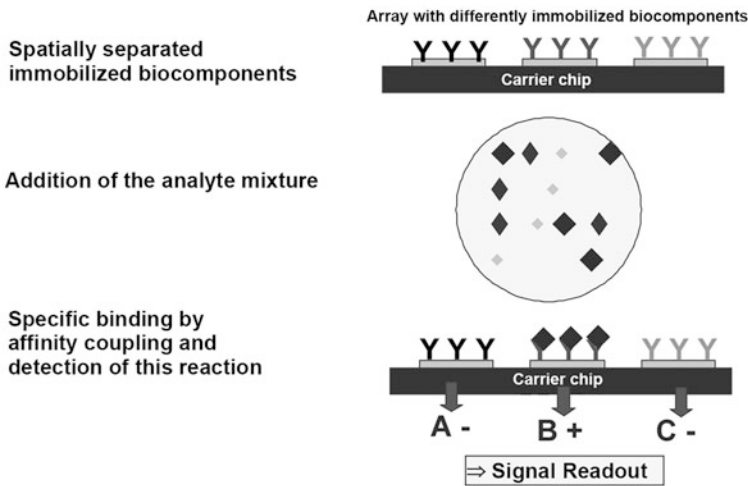


Fig. 12.2 Functional principle of a biosensor

Table 12.1 Binding energies of non-covalent binding types (according to Alberts, Bray and Lewis [2])

Binding type	Binding energies [kJ/mol]
Ionic interactions	12.5
Hydrogen bridge bonds	4
Van der Waals interactions	0.4

sequencing and diagnostics, methods are required that allow a very high degree of parallelization in the acquisition of measured values.

12.1.2 Real Biological Sensors

So-called *true* biological sensors are those sensors that *nature* itself uses as sensors and uses to collect one or more measurement parameters or events. The entire living organism is used as a sensor. With real biological sensors, specific analysis for the detection of certain substances is not possible. Hereby only the sum of all, usually negative influences, can be measured from the outside and evaluated as a *sum signal*. Biological sensors are used wherever specific analysis is not possible or an initial evaluation of hazards or effects is necessary. To evaluate the *toxicity* and *effects of substances*, for example, cells are used as living organisms and serve as representatives for a complex organism. These *cell tests* are mainly used in the pharmaceutical industry, for example in the development of active pharmaceutical ingredients. To determine the toxicity of substances, so-called *cytotoxicity tests* are carried out. For this purpose, cells from certain species (such as mice or human cells) are treated with the substance to be tested. After incubation over a certain period of time, various tests are performed to detect cell damage. This may be a microscopic test to

detect cell density and sublethal changes compared to a reference; in addition, certain enzymes such as *lactate dehydrogenase* are measured. Lactate dehydrogenase is an enzyme located in the cytoplasm of the cell. If damage to the cell membrane occurs, the enzyme can be released more intensively and leads to an accumulation of the enzyme in the surrounding medium. In addition to this, *viability tests* are carried out. The basic dye neutral red is easily absorbed into the lysosomes of intact cells. Damage to the cell membrane and lysosomal membrane results in reduced dye retention during washing. The colour can be determined by photometric measurements.

To ensure *drinking water quality*, a large number of analytes must be detected. As this is a sum parameter, *small organisms* are much more suitable, as they can be used to evaluate the effect of different substances on the entire organism without having to know the individual substances in detail. For this reason, biological sensors are still used today to evaluate drinking water quality. These include, for example, the *fish egg test*, which examines the determination of the non-acutely toxic effect of wastewater on the development of fish eggs via dilution stages. 60 healthy, fertilised eggs of the *zebrafish* are selected for the test and examined. The fertilised fish eggs cannot develop properly in toxic effluents. These eggs are then examined for *lethal malformations* such as coagulation of the eggs, absence of somites, no detectable heartbeat, no detachment of the tail from the yolk, and *non-lethal malformations* such as lack of eye contact, lack of pigmentation and deformities.

As a continuous biotest method, the *Daphnia test* is used for the timely monitoring of running waters or raw waters. The average speed, the speed distribution, the range of swimming strokes and the swimming strokes per time unit are observed. Additionally, the swimming height, turns, circular movements and size of the daphnia (growth) are determined. For evaluation, the data is recorded via a video system and the image evaluation from different measuring chambers is used. There are up to 10 *Daphnia* in one measuring chamber. This examination is mainly used for monitoring for *pesticides*, *neurotoxins* and *warfare agents*. *Mussels* are also used in continuous biological testing procedures. They show sensitivity in the range of a few mg/l to some µg/l.

For the *static luminescent bacteria test* (a short-term test of 30 min), certain marine bacteria (including *Vibrio fischeri*) are used, which emit a natural glow (*bioluminescence*). This glow, which occurs due to enzymatic, energy metabolism dependent processes, can be inhibited by pollutants (EC₂₀ approx. 0.6 mg/l dichlorophenol). In the luminescent bacteria test itself, the inhibition of cell proliferation is determined by *Vibrio fischeri* after 7 h contact time by means of turbidity measurements at 436 nm (DIN 38412-L37). In contrast to the short-term test, *chronic toxicity* is recorded here.

In the *fish test*, the swimming behaviour of fish is monitored visually or with image analysis over 48 h.

Another biological test is the *algae test*. For this purpose, unicellular green algae are exposed to light over an incubation period of 15 min (photosynthesis) and the fading dark red afterglow (680–720 nm; delayed fluorescence) is determined. Photosynthesis toxins

alter the decay kinetics, so that, for example, a few $\mu\text{g/l}$ of herbicides cause a significant change in the kinetics.

More recent developments are aimed at linking cells directly with transducers, in so-called *cell-based sensors*. This enables the various processes of *cell metabolism* to be detected directly with the aid of a biochip. A detailed description and applications of cell-based sensor technology is given in Sect. 12.7.

12.2 Functional Principles of Biosensors

The interaction of the *analyte* with the receptor is translated into an electrical signal by means of the signal *converter* (*transducer*). There are a number of different types of transducers that can be used for biosensor applications. Usually, the biosensors themselves are classified according to the type of transducer used. The transducers are divided into two groups, depending on whether they can be traced back to a physical measuring method or a chemical measuring method (Fig. 12.3). Within these groups, there are very different types of transducers, which are based on different detection methods. These can be *calorimetric* or *microgravimetric* transducers on the one hand, but also *optical* or *electrochemical* transducers on the other hand. Within the chemical measurement methods, there is an additional group that combines electrical and optical measurements—the *electro-optical* detection of biological interactions. Figure 12.3 shows the classification of biosensors according to the type of transducer.

Within these four or five groups of biosensors, there are different application examples and designs. Table 12.2 summarizes the most important transducers with the most important sensor examples and the corresponding measured variable.

The biosensors are further subdivided according to whether a *marker* is used for the signal generation or not. Sensors that *do not* use a *marker* have a *direct biosensor format*. Here, changes of a measured variable are observed that are caused by the interaction of the analyte with the receptor itself, such as an increase in mass. Sensors that use a *marker* have an *indirect biosensor format*. Here, it is not the complex formation (analyte receptor) itself that causes the change in the measured variable, but the marker. The bound amount of the marker is correlated with the concentration of the analyte. In the following, some transducers or biosensors are briefly explained.

12.2.1 Calorimetric Sensors

If a *change in enthalpy* occurs during the reaction of the analyte with the receptor, this can be measured directly as a change in the heat of reaction. The resulting increase in temperature depends on the amount of substance in the reactants. A prerequisite for this is that the temperature increase is sufficient. Usually, these effects are very small, so that these measurements must take place in very well insulated systems. Figure 12.4

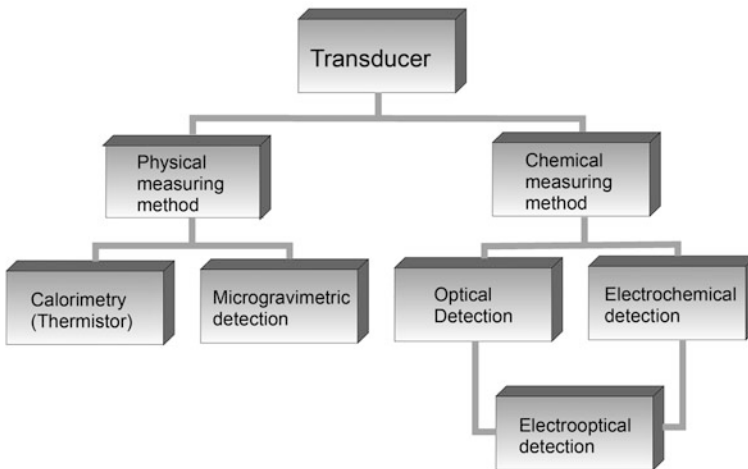


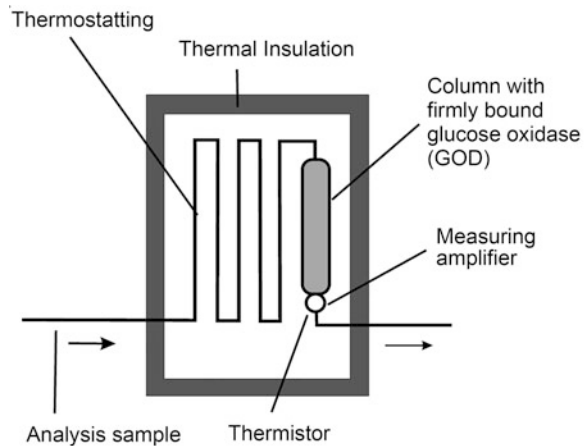
Fig. 12.3 Classification of biosensors according to the type of transducer

Table 12.2 Classification of biosensors according to the type of transducer used

Transducer type/ detection method	Application examples/sensors	Measured variable
Calorimetric/ thermal	Thermistor	Temperature change
Microgravimetric/ piezoelectric	QCM (Quartz Crystal Microbalance), microcantilever	Resonant frequency
	SAW (Surface Acoustic Wave)	Velocity of propagation of surface acoustic waves
Visual	TIRF (totally internal reflection fluorescence), optical waveguide	Fluorescence
	Luminescence/colorimetry	Chemiluminescence/UV/VIS
	Nephelometry	Light scattering
	Reflectometry	Optical coating thickness
	SPR (surface plasmon resonance)	Refractive Index
Electrochemical	Amperometry	Electricity
	Coulometry	Charge
	Impedance	Resistance
	Potentiometry	Voltage
	Conductometry	Conductivity

schematically shows the structure of a calorimetric biosensor. The more glucose is converted by the glucose oxidase, the more heat is generated. Since this heat is very small, not only the isolation is very important, but also that the reaction partners must be

Fig. 12.4 Calorimetric biosensor for determining glucose



brought to exactly the same temperature before the measurement. The temperature is measured with a miniaturized semiconductor thermistor.

12.2.2 Microgravimetric Sensors

Microgravimetric transducers, for example, consist of a *piezoelectric system*. During the complex formation of the analyte with the receptor bound to the surface of a piezoelectric crystal, an increase in mass occurs at the surface, which in turn changes the resonance frequency of the piezoelectric crystals. This effect can be used to detect and quantify complex formation. *Quartz thickness oscillators* consist of a piezoelectric plate with circular gold electrodes on both sides. Here the oscillations run through the material with a direction of propagation perpendicular to the sensitive surface, the deflection is parallel to the surface. Since quartz platelets are usually used as a substrate, the sensors are often referred to as *oscillating quartzes*.

Because of their mass sensitivity, the term quartz crystal microbalance (*QCM* or *QMB*) is also commonly used (Fig. 12.5). The following applies to the resonance frequency of the oscillations

$$f = \frac{n}{2d} v = \frac{n}{2d} \sqrt{\frac{c}{\rho}} = n \frac{N}{d},$$

where d is the thickness of the quartz, ρ its density and n the n th harmonic. c is the coefficient of elasticity in the direction of propagation of the acoustic wave and v is the propagation velocity.

The development of *microcantilevers* led to significant miniaturization of these sensor systems and the possibility to run several tests in parallel. Figure 12.6 shows the schematic

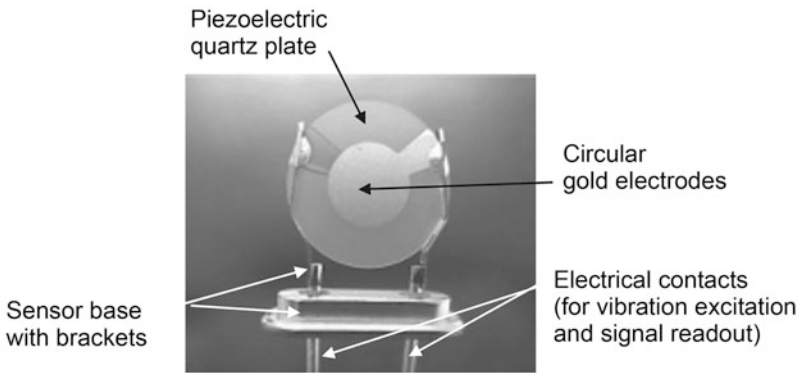


Fig. 12.5 Structure of a QMB

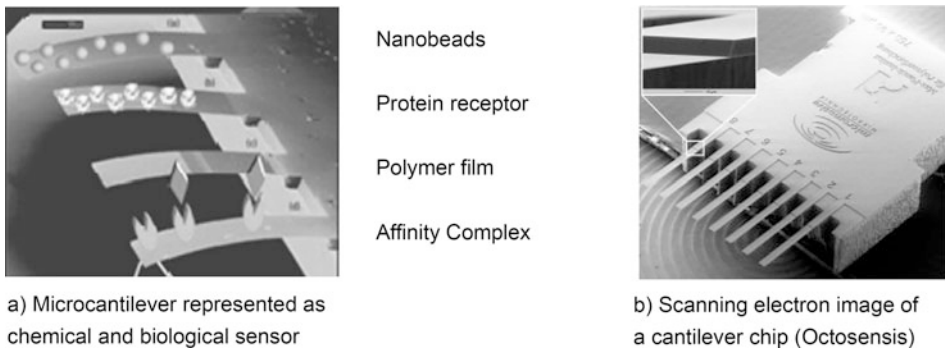


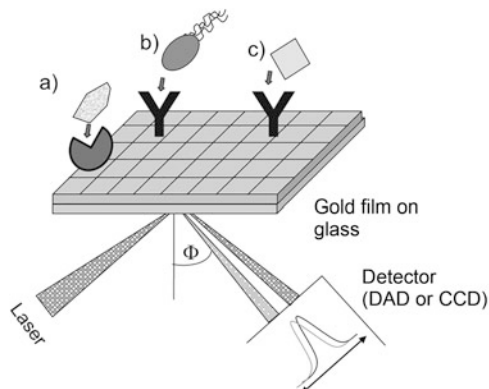
Fig. 12.6 (a) Microcantilever as a biosensor (Source: M. Sepaniak, *Analytical Chemistry*, 570-575 (2002)) and (b) ein Cantilever-Chip (Source: Octosensis)

of a biosensor based on a microcantilever and a commercial product. In Fig. 12.6a various functionalizations of the surface of the microcantilever are shown schematically.

12.2.3 Optical Sensors

All optical biosensors are based on the effect that the complex formation changes the optical properties of the system. Some systems use the generation of evanescent fields. *Evanescent field sensors* are based on the principle of *total reflection of light* at the phase interface between the optically denser light guide and optically thinner medium. The electromagnetic field of the light wave partially enters the optically thinner medium and then decays exponentially. This evanescent field can be used directly to detect complex formation when the complex is immobilized very close to the phase boundary, such as in *surface plasmon resonance (SPR)*. In SPR, polarized, monochromatic light is irradiated

Fig. 12.7 Schematic structure of a sensor based on surface plasmon resonance (SPR) (a) enzyme/substrate, (b) antibody/protein, (c) antibody/bacteria



from below onto a thin layer of precious metal (e.g. gold or silver) where total reflection occurs (Fig. 12.7). The precious metal layer is located on a glass prism. The light can now interact with the free electrons of the precious metal, which form a so-called *electron plasmon*. If this resonance occurs, an energy gap occurs in the reflected light, which has a very specific angle θ . The resonance and thus the angle θ is directly proportional to the degree of loading of the surface.

With the SPR, marker-free test systems can be set up. If there is an interaction between the receptor and a potential ligand, the degree of loading on the gold surface increases. The resulting change in angle θ is detected with a diode array detector (DAD) or a CCD camera.

In addition to evanescent excitation, the analyte-receptor complexes can also be determined by changing *light scattering* or *layer thickness*. Both methods can be used without marking the affinity complex. The disadvantage of marker-free biosensors is usually the very expensive expenditure in terms of equipment and the sophisticated optical system since the problems of unspecific binding and matrix effects are more frequent and have to be solved.

The evanescent field can also be used to excite directly (e.g. with *TIRF* (*totally internal reflection fluorescence*)) fluorescent dyes that are bound to the complex and are located directly at the phase interface. Figure 12.8 shows an optical waveguide that has bound a complex at the surface, which is marked with a fluorescent dye (left side). The fluorescent dye is excited via the waveguide by laser light (Fig. 12.8 right side).

The fluorescence signals are read out per position of the biosensor array using optical fibers (POF) directly coupled to the array positions of the waveguide.

There are also optical biosensor systems in which the bound complexes are also marked with a fluorescent dye. In this case, the excitation of the fluorescence takes place directly on the chip (without waveguide), since the affinity complexes are immobilised on simple glass or plastic chips, for example. The signals (over the entire biosensor) are read out with a CCD (charged coupled device) camera. This principle is used, for example, in the *DNA microarray system* marketed by the company Affimetrix (Sect. 12.6). The *chemiluminescence biosensor* works similarly. The complexes are labelled with an enzyme (*peroxidase*)

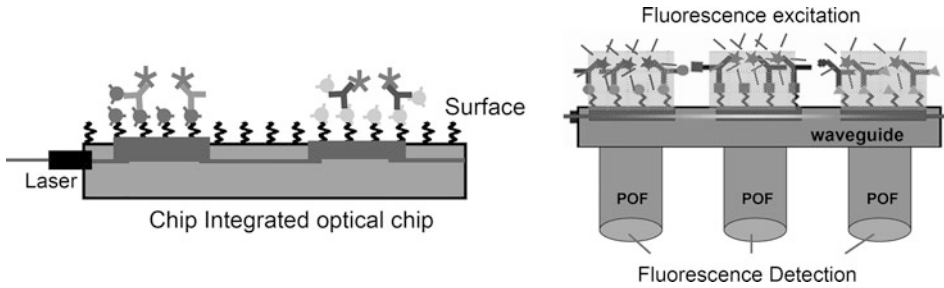


Fig. 12.8 Fluorescence-based detection of the binding of an affinity complex carrying fluorophore on the surface of a plastic optical fiber (POF) according to J. Tschmelak [3]

and then *luminol* is added. Luminol is directly oxidised by the enzyme, resulting in short light emission at 425 nm, which can be detected by a CCD camera. The advantage here is that chemiluminescence does not require an excitation light source; the energy is provided by the high-energy educt luminol itself. Figure 12.9 shows the principle of chemiluminescence detection in the formation of an antigen-antibody complex which has been labelled with the enzyme peroxidase using a secondary antibody (a) and an image of a chemiluminescent biosensor array for the detection of *E. coli* (b).

12.2.4 Electrochemical Sensors

In electrochemical transducers, the formation of the *analyte-receptor complex* causes a change in electrical parameters such as charge, current or voltage.

Amperometric sensors are often referred to as *enzyme electrodes*, as these are very widely used and are already available on the market as biosensors for the detection of glucose. These sensors use a selective enzyme reaction to produce or consume an electroactive species. A constant DC voltage is applied between measuring and a reference electrode, where the diffusion limit currents are measured. The externally applied potential causes the working electrode to be polarized, but practically no current flows. Depolarization only occurs in the presence of the electroactive species. The measured current is proportional to the concentration of the electroactive species.

The Nernst equation applies to the working electrode:

$$\frac{C_O}{C_R} = \exp \left[\frac{(E - E^0)nF}{RT} \right] \text{ Nernst Equation}$$

C_O : concentration (activity) of oxidised species; C_R : concentration (activity) of reduced species; F : Faraday constant: 96,486 C/mol; n : number of electrons; R : gas constant; E measured potential between working and reference electrode

When the voltage E^0 is constantly applied, it is valid for the *Faraday current*:

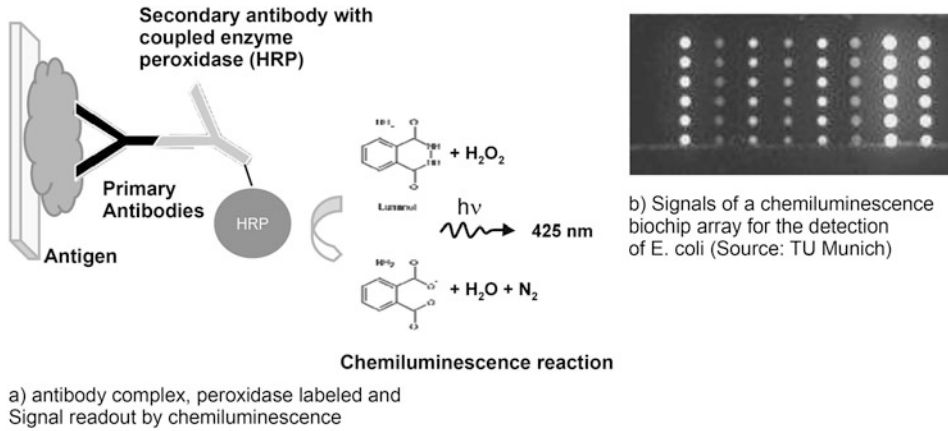


Fig. 12.9 Schematic structure of a biosensor based on chemiluminescence and biochip array

$$i_f = nFAD_0 \left(\frac{d[C_0]}{dx} \right)_{x=0}$$

A : Electrode area, i_f : Faraday current; dC/dx : slope of the concentration profile.

The current i_f is proportional to the C_0 gradient and the diffusion coefficient D_0 .

Besides the amperometric method using enzyme electrodes, there are amperometric biosensors based on the principle of *redox recycling*. Here, the complex is marked with the help of an enzyme. This enzyme serves to release an electrochemical substance which is stable in both the oxidized and reduced state and is therefore repeatedly available for the electrochemical reaction, that is, can be “recycled”. The prerequisite for successful redox recycling is therefore a *redox-active species* (e.g. p-aminophenol, ferrocene, hexocyanoferrate) and an electrode gap that is smaller than the diffusion path length of the redox species.

Figure 12.10 shows examples of a biochip microarray from FHG ISIT (eBiochip Systems GmbH, biochip array with 16 target positions, for example, for measuring penicillin in milk) and a biochip microarray from Siemens based on CMOS technology (128–512 target positions, for example, for use as a DNA or protein biochip). Both biochips have an interdigital electrode structure made of gold with an electrode gap of less than or equal to 1 μm .

A potential difference specific to the redox species is applied to the interdigital electrodes. For the example shown in Fig. 12.11 with para-aminophenol, a potential difference of 300–350 mV between anode and cathode is required. When the potential is applied, p-aminophenol is oxidized to p-iminoquinone at the anode, which in turn is reduced to p-aminophenol at the cathode. This redox process is repeated as long as the potential is applied. As can be seen in the graph on the right, the anode current and cathode current increase over time in the presence of p-aminophenol (Fig. 12.11).

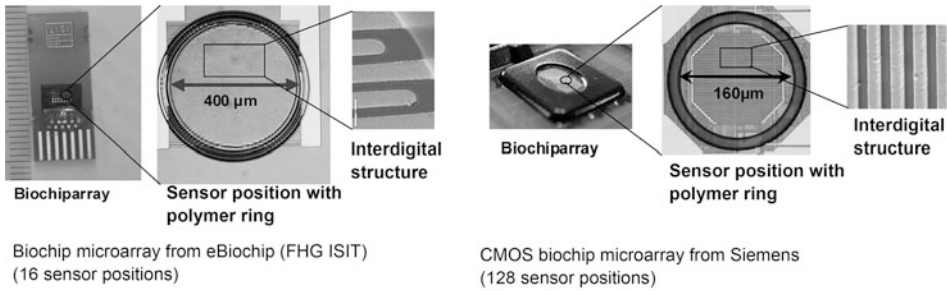


Fig. 12.10 Examples of amperometric redox recycling. Biosensors with an interdigital electrode structure made of gold

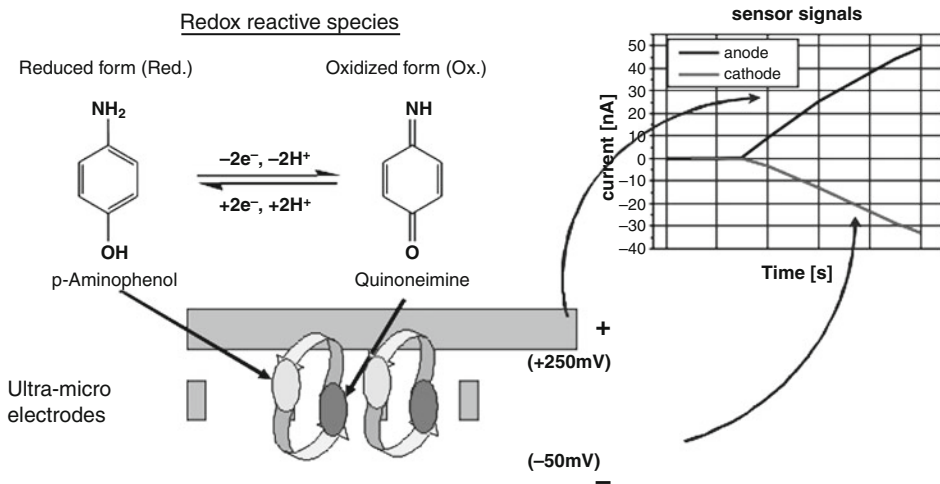


Fig. 12.11 Principle of the redox recycling process

Potentiometric biosensors are used for ionic reaction products (H^+ , CO_3^{2-} , NH_4^+). The quantitative determination of these ions is based on their electrochemical potential at the measuring electrode. Technically relevant, however, are those sensors in which the biological components can be applied directly to a silicon semiconductor. If the pH value is changed by the formation of the analyte-receptor complex, a so-called *ISFET* (ion-selective field-effect transistor) is suitable here. Here, H^+ takes over the function of the drive voltage at the gate terminal of the transistor, whereby the current flow between the source and drain increases with increasing H^+ concentration. This sensor thus supplies a sensor signal proportional to the *proton concentration* (Fig. 12.12).

IS-FET sensors on the basis of semiconductors in CMOS technology can be manufactured today very cost-effectively and reliably. In addition, evaluation electronics

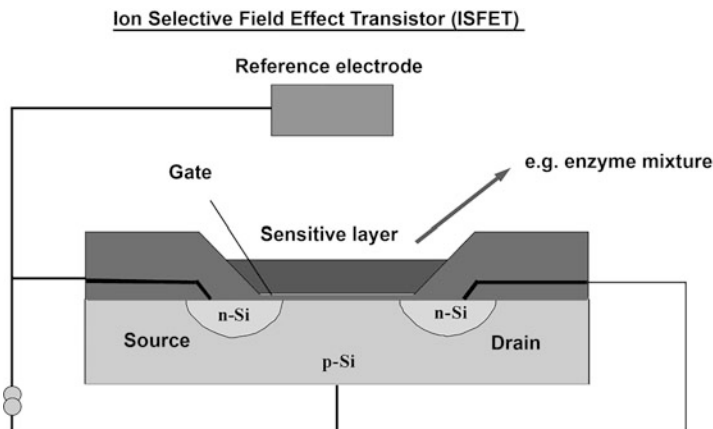


Fig. 12.12 Diagram of an ISFET (ion-selective field effect transistor)

and various sensors (e.g. microarrays) can be integrated into one chip and combined to form a compact miniaturized measuring device.

12.2.5 Immobilization Methods

In the field of biosensor technology, the *immobilization* of the receptors on the surface of the transducer is an essential element that strongly influences the analysis. This applies to applications in the field of enzyme, immuno and especially DNA sensors. There are many different methods to immobilize receptor molecules on the surface. This depends on the one hand on the molecule itself to be immobilized and on the other hand on the type of surface. A distinction is made between two basic immobilization methods.

1. Covalent immobilization

Receptors (antibodies, proteins, oligonucleotides, haptens) can be *covalently* bound to the transducer surface. This leads to an *irreversible immobilization* with a defined orientation of the receptor on the surface. Stable surfaces can thus be produced. However, this is coupled with chemical modification of the receptor. The covalent coupling is achieved by modification of the surface such as glass, Al_2O_3 -, Si_3N_4 -, Au-, Pd- or silicate substrates by silanisation, cross-linker chemistry (modified dextrans, NHS-ester coupling) or polymers.

2. Non-covalent immobilization

The non-covalent binding of receptors to surfaces is very diverse. Therefore, only the most important methods are summarized here.

- *Ligand receptor interactions* (e.g. protein A/antibody binding, biotin/avidin binding).
- *Adsorptive immobilization* (hydrophobic bonding on silanized glass surfaces, activated polystyrene surface).
- *Thiol binding* to gold surfaces (formation of stable “self-assembled” monolayers with hydrophobic or hydrophilic groups, with thiol-derivatized oligonucleotides, by coupling terminally functionalized thiols (e.g. with amino groups) for covalent binding of the receptor (e.g. the Fc fragment in antibodies)).
- *Langmuir blodgett films* (hydrophobic monolayers).
- Immobilization of proteins using *oligonucleotides*.

12.3 Physical and Chemical Sensors in Medicine

Today, essential information about the health or disease state of a person can and is obtained by examining the *blood*, but also other *body fluids* such as urine, saliva and cerebrospinal *fluid*. For this purpose, very different parameters are examined, which are necessary for the function of the body and can be correlated with the patient’s state of health or with diseases. Very different detection principles are used for these procedures and analyses, which are realized with very different sensory measurement methods. These methods range from classical physical sensors such as temperature sensors to chemical *pH sensor elements* to *enzymatic* and *immunological* sensor methods and the first genetic investigations using *DNA sensors* (Sect. 12.6) or *cell-based sensors* (Sect. 12.7). Physical and chemical sensors can be classified analogously to biosensors on the basis of the transducers used (Fig. 12.2). Since transducers used for biosensors can ultimately be traced back to purely physical or chemical effects, a comparable picture emerges for the group of physical-chemical sensors used in medicine. What distinguishes biosensors on the one hand and physical and chemical sensors on the other hand is their design with regard to the generation of the sensor signal. While a biosensor uses a biological receptor and biological or biochemical effects (e.g. antigen-antibody interaction, enzyme-substrate reactions, DANN hybridization) to detect an analyte, *chemical sensors use* chemical receptors and a chemical effect such as acid-base reactions or redox reactions, but also physical-chemical effects such as interactions between receptor and analyte (e.g. *gas adsorption*). The IUPAC (International Union of Pure and Applied Chemistry) issued an official definition of chemical sensors in 1991:

A chemical sensor is an arrangement that converts chemical information (e.g. concentration of a single analyte) into an analytically useful signal. The chemical information can result from a chemical reaction of the analyte or the physical property of the system under investigation.

Physical sensors measure a physical quantity (e.g. pressure, temperature). The signal is determined by a physical process (e.g. expansion, column position, capillary forces, mass change). Figure 12.13 shows an overview of the physical-chemical sensors.

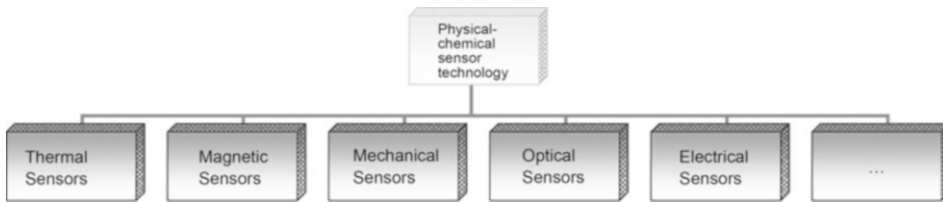


Fig. 12.13 Overview of physical-chemical sensors

12.3.1 Physical-Chemical Blood Analyses

Blood Gases

Blood gas analysis is a method for measuring the gas distribution (partial pressure) of O₂ (oxygen), CO₂ (carbon dioxide) as well as the pH value and the acid and base balance in the blood. Blood gas analysis goes back to the need to monitor and control ventilation parameters and was developed in its basic form in the 1960s. Over the years, other values were added, such as the measurement of haemoglobin, bicarbonate, glucose, lactate, or electrolytes. In the meantime, blood gas analysis is used to monitor many patients with *respiratory disorders* and *oxygen deficiency* (such as chronic obstructive pulmonary disease or cystic fibrosis). In intensive care units, blood gas analysis is usually performed at the bedside, that is, in the immediate vicinity of the patient. Preference is given to arterial whole blood from an artery or *arterialized capillary blood*, for example from the hyperemic earlobe. Especially in the neonatal intensive care unit, it is crucial that very small sample volumes (50 µl) are sufficient for the analysis. Venous blood is suitable for the assessment of respiration-specific values only with restrictions, among other things because of its low oxygen content.

As a rule, in today's devices, pH measurement (a glass electrode), pO₂ measurement (Clark electrode) and pCO₂ measurement (Severinghaus electrode) are carried out independently of each other; standard bicarbonate (HCO₃⁻), excess base and O₂ saturation are calculated. Table 12.3 lists the normal values of the blood gas parameters.

Electrochemical pH value measurement

A common method for measuring the pH value is *electrochemical measurement* using a pH meter. There is a potential difference U between two solutions that have different pH values (different H₃O⁺ ion concentrations). This potential difference can be calculated using the Nernst equation.

$$U = 2.303 \cdot \frac{RT}{F} \cdot (\text{pH}_2 - \text{pH}_1)$$

$R = 8.314 \text{ J/mol K}$; $T = \text{absolute temperature}$; $F = 96,485 \text{ C/mol}$.

Table 12.3 Normal blood gas values

Parameters	Standard
pH	7.36–7.44
pO ₂ (oxygen partial pressure)	75–100 mmHg
pCO ₂ (carbon dioxide partial pressure)	35–45 mmHg
HCO ₃ ⁻ st. (standard bicarbonate)	22–26 mmol/l
BE (excess base)	(–2) to (+2) mmol/l
O ₂ saturation	94–98%

The voltage U can be measured in a closed circuit. If one of the two pH values is known, the second value can be calculated from it. Figure 12.14 shows an electrochemical pH meter. This consists of two silver electrodes coated with AgCl and placed in a reference solution consisting of an AgCl saturated potassium chloride solution. The reference solution has a constant pH value. The only difference between the two electrodes is the type of contact with the solution to be measured.

The reference electrode has a diaphragm which allows a small exchange of ions, but the solutions do not mix, that is, the solutions remain neutral. The measuring electrode has a glass membrane through which H⁺ ions can diffuse. This creates a potential at the electrodes that obeys Nernst's law over a wide pH range.

pO₂ value measurement (Clark electrode)

The Clark electrode [4] usually consists of a platinum cathode and a silver anode, which are connected via an electrolyte solution. There are also other arrangements like gold against silver or gold against lead. The metal electrodes are separated from the sample to be measured by an oxygen-permeable membrane, usually made of Teflon (there are also membrane-free arrangements in which the sample also serves as electrolyte).

In the case of the Pt/Ag combination, a polarization voltage of -0.8 V is applied to the platinum cathode against the silver anode; other combinations, such as with lead, do not require polarization.

If the membrane of the measuring chamber is immersed in the sample (e.g. in water or an arterial blood sample) to determine the oxygen content, O₂ diffuses according to its partial pressure through the membrane into the measuring chamber and is reduced there at the cathode, producing hydroxyl ions (OH⁻).

Silver, for example, is oxidized at the anode and, in the presence of chloride, is deposited on the electrode as insoluble AgCl (the deposits of the oxidized anode metal must be removed regularly to maintain the unimpeded current capability of the measuring arrangement). The following reactions take place at the cathode and the anode:

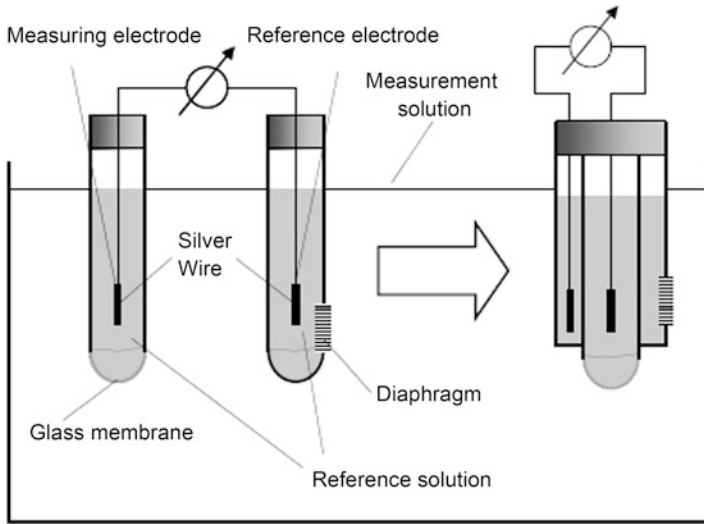
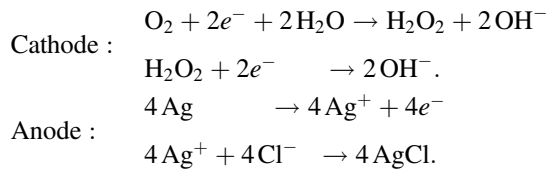


Fig. 12.14 Structure of an electrochemical pH meter (Source: UNI Aachen Verfahrenstechnik)



The measured current I is directly proportional to the partial pressure of oxygen $p(\text{O}_2)$.

$$p(\text{O}_2)\tilde{I}.$$

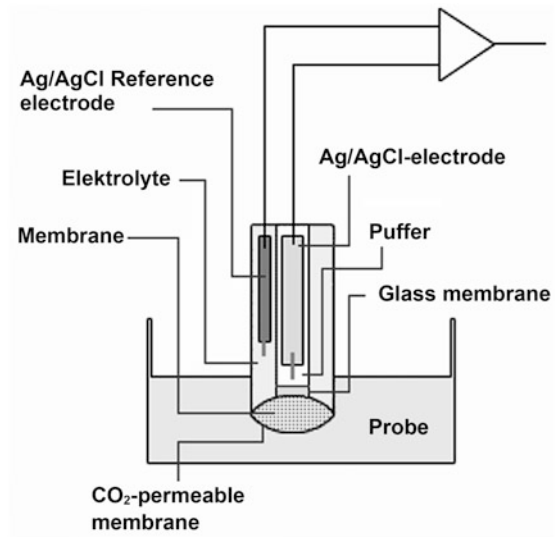
When evaluating the measuring current, it must be taken into account that both the diffusion rate of oxygen through the membrane and the oxygen solubility are temperature-dependent.

pCO₂ value measurement (Severinghaus electrode)

The Severinghaus electrode [5] is an electrochemical sensor for the quantitative determination of the concentration of carbon dioxide in a solution (or also in gases). It is a pH electrode preceded by a measuring chamber with a buffer of potassium hydrogen carbonate KHCO_3 (Fig. 12.15). Via a gas-permeable membrane, carbon dioxide from the sample penetrates the buffer solution according to the partial pressure of CO_2 there and shifts its dissociation equilibrium:

After a short time, the pH value is established in the buffer that depends on the concentration of CO_2 outside the membrane. The following reaction equation applies:

Fig. 12.15 Structure of a Severinghaus electrode



Structure of a blood gas measuring device

There is currently a whole range of different blood gas analysis systems in the field of medical laboratory diagnostics on the market. Most of them can also be used to record additional measured values: Electrolyte measurements Na^+ , K^+ , Ca^{++} , Li^+ and Cl^- , furthermore, for example, haemoglobin, bicarbonate, glucose and lactate. Figure 12.16 shows the pH, pO_2 and pCO_2 sensors of devices from Eschweiler and Bayer. These devices are designed in such a way that the individual components can be exchanged as small sensor elements.

12.3.2 Clinical-Chemical Blood Analyses

Due to the large number of parameters to be measured, there are many different measurands and detection methods in clinical chemistry. Table 12.4 lists a small selection of clinical chemistry blood values.

As an example, the detection of haemoglobin with the haemoglobin cyanind method is described in more detail here. Haemoglobin is a chromoprotein contained in the erythrocytes with a molar mass of 64.46 Da. It consists of globin (94%) and the iron (Fe) containing prosthetic group haem (6%). The haemoglobin molecule is composed of four peptide chains each containing one haem. Normally there are a total of four different



Rapidlab blood gas device (Bayer)

pCO₂ sensor (Severinghaus cell)**Fig. 12.16** Commercial blood gas measuring systems**Table 12.4** Selection of some clinical-chemical blood values

Bilirubin	Urea	Potassium
Cholesterol	Hemoglobin	Triglycerides
Creatinine	HDL cholesterol	Uric acid
Lactate	Urea	Rheumatoid factor
Amphetamines	Barbiturates	Ethanol
Methandone	Opiates	Salicylates
Dogoxin	Gentamicin	Phenobarbital

peptide chains (α , β , γ , δ). In the components of the haemoglobin, two chains are identical in each case.

Because of its ease of use, good precision and good colour consistency, the *haemoglobin cyanide method* is superior to all others and is therefore the most frequently used method. The following biochemical processes take place:

Haemoglobin (Fe^{+2}) is oxidized by potassium ferricyanide (potassium hexacyanoferrate) to haemoglobin (Fe^{+3} , methaemoglobin) and then converted with potassium cyanide (KCN) to haemoglobin cyanide ($\text{Hb-Fe}^{+3}\text{-CN}$). Hemoglobin cyanide is a very stable derivative of hemoglobin, whose color remains unchanged even after 24 h. The absorbance is measured against a transformation solution (potassium hexacyanoferrate III/KCN) at a wavelength of 546 nm.

12.4 Enzymatic Methods: Enzyme Sensors

Enzymes are proteins that can catalyze a chemical reaction. Enzymes play a major role in the *metabolism* of all living organisms: they catalyze and control the majority of biochemical reactions—from digestion to the copying of genetic information (DNA polymerase). They are divided into six enzyme classes based on the reaction they catalyze:

1. *Oxidoreductases*, catalysis of redox reactions.
2. *Transferases*, transfer of functional groups from one substrate to another.
3. *Hydrolases*, splitting of bonds with the use of water.
4. *Lyases/synthases*, catalysis of cleavage or synthesis of more complex products from simple substrates (without cleavage of ATP).
5. *Isomerases*, acceleration of the conversion of chemical isomers.
6. *Ligases or synthetases*, catalysis of the formation of substances that are chemically more complex than the substrates used (difference to lyases: only effective enzymatically under ATP cleavage).

Some enzymes are able to catalyse several reactions. If this is the case, they are assigned to several enzyme classes.

The structure of the enzymes is very different. Many enzymes consist of one protein chain (*monomers*), other enzymes form *oligomers* of several protein chains, still, others combine with other enzymes to so-called *multicomplexes* and cooperate with each other. In addition, there are individual protein chains that have several enzyme activities. Figure 12.17 shows the structure of the enzyme triose phosphate isomerase (TIM) of glycolysis.

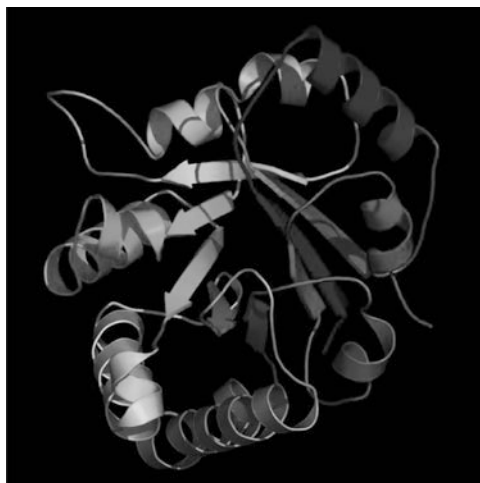
As biocatalysts, enzymes can accelerate chemical reactions by reducing the activation energy (ΔG). The *active centre* of the enzyme is responsible for this catalytic effectiveness. There the substrate is bound and converted. This active centre can consist of folded polypeptide chains or reactive “non-protein” components (e.g. cofactors).

On the one hand, a certain spatial structure of the active centre can cause only a structurally suitable substrate to be bound. On the other hand, certain (non-covalent) interactions (H-bridges, electrostatic interactions, hydrophobic effects) between enzyme and substrate are necessary for binding. Only this leads to the formation of the *enzyme-substrate complex*. Thus, a certain substrate matches the corresponding enzyme like a key matches the corresponding lock (*key-lock principle*). In addition to this principle, the (non-rigid) *induced fit model* also exists. The active centre of the enzyme can be reshaped through interaction with the substrate and thus contribute to the specific binding of the substrate.

Sometimes, small structural differences in the spatial structure or the charge distribution are often sufficient, so that the substrate is no longer recognized or bound. However, enzymes can also have a *broad substrate specificity* (e.g. alcohol dehydrogenases also degrade other alcohols besides ethanol, hexokinases can also convert other hexoses besides glucose).

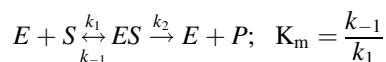
In general, the binding of the enzyme must be strong enough to bind the often small concentrations of the substrate. However, the binding must not be too strong, as the reaction only ends with the formation of the product. It is important that the transition state, that is, the *substrate-enzyme complex*, is stabilized. An important factor here is the *reaction rate*, which depends on the temperature, the salt concentration, the pH value but also on the concentration of the enzyme, the substrate and the products as well as the actual

Fig. 12.17 “Ribbon diagram” of the enzyme triose phosphate isomerase (a stylized representation of the protein structure obtained by X-ray structure analysis according to Wikipedia)



enzyme activity itself. The units of enzyme activity are *Unit* (U) and *Katal* (kat). 1 U is defined as the amount of enzyme that converts 1 μmol of substrate per minute under the given conditions (1 U = 1 $\mu\text{mol}/\text{min}$). 1 Katal is defined as the conversion of 1 mol substrate per second (1 kat = 1 mol/s). The measured enzyme activity is proportional to the reaction rate and therefore strongly depends on the reaction conditions. For example, an increase in temperature of 5–10 $^{\circ}\text{C}$ leads to a doubling of the reaction rate and thus to an increase in enzyme activity.

A model for the kinetic description of simple enzyme reactions is the *Michaelis-Menten-Theory* (*MM-Theory*). It provides a relationship between the *reaction rate* v of an enzyme reaction and the *enzyme* and *substrate concentration* $[E_0]$ and $[S]$. It is based on the fact that an enzyme forms an enzyme-substrate complex with a substrate molecule and that this complex breaks down either into enzyme and product or into its initial components. What happens faster depends on the respective rate constants k .



E (enzyme concentration), S (substrate concentration), P (product concentration), K_m (Michaelis constant).

The *Michaelis constant* is a characteristic quantity for each enzyme and each substrate it converts. It has the dimension of a concentration (mol/l) and is to be understood as the substrate concentration at which the half-maximum rate of conversion is reached. The lower the K_m value, the higher the affinity of an enzyme to its substrate.

The model states that with increasing substrate concentration the reaction rate also increases. Initially, this happens linearly and then flattens out until a further increase in

substrate concentration no longer has any effect on the speed of the enzyme, since it is already working at maximum speed v_{\max} . The MM equation is then as follows:

$$v = \frac{k_{\text{cat}}[E_0] \cdot [S_0]}{K_m + [S_0]}$$

k_{cat} (change number). The *turnover number* is a measure of the maximum reaction rate at substrate saturation (v_{\max}), also called *molecular activity*, “*turnover number*” or k_{cat} called $k_{\text{cat}} = V_{\max}/[E_0]$.

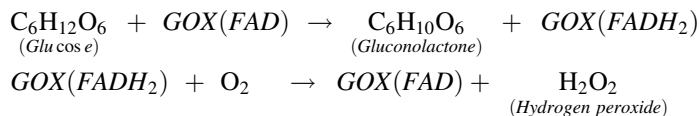
The enzyme hexokinase (it binds glucose with high affinity) has a very low K_m value of 0.01 mM. Due to the low K_m value, the hexokinase works with respect to glucose both at a blood glucose concentration of 4 mmol/l in and at 8–10 mmol/l glucose in the saturation range. The hexokinase thus has a K_m value for glucose which is far below the lowest blood glucose concentration. This ensures that the muscle or the brain is able to transfer glucose from the blood into glycolysis as required, regardless of the metabolic state.

Medical diagnostics uses enzymes to diagnose diseases. In particular, these examination methods make use of the enzymes’ high substrate specificity. The principle of *enzymatic examination methods* in diagnostics consists of the following:

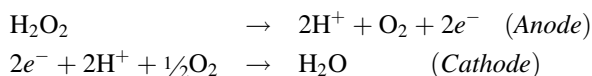
1. *The substrate specificity of an enzyme is used to measure the concentration of an analyte.* For this purpose, an enzyme suitable for the analyte is added to the sample to be analyzed (e.g. serum). The reaction product of the specific analyte is then detected by a classical measuring method (enzymatic measurement).
2. *The activity of the enzyme (enzyme kinetics) is determined.* The enzyme-specific substrate is added to the serum sample, for example. The concentration of the product resulting from this enzyme-substrate reaction is then measured within a certain time window. The activity of the enzyme can then be deduced. On the basis of such tests, conclusions can be drawn about the damage to organs. For example, liver damage can be indicated by an increase in the enzymes GOT (glutamate oxalacetate transaminase), GPT (glutamate pyruvate transamina) and γ -GT (gamma-glutamyl transferase).

12.4.1 Enzyme-Based Analyte Detection

A classic example of *enzymatic detection* (see point 1 of the previous section) is the determination of glucose in the blood. Here, the concentration of the analyte glucose is determined by means of the specific enzyme glucose oxidase, which has a very high substrate specificity for glucose. This process takes place in the following steps:

Step 1: Enzymatic reaction

GOX: glucose oxidase, FAD: flavine adenine dinucleotide (co-factor)

Step 2: Electrochemical detection of hydrogen peroxide

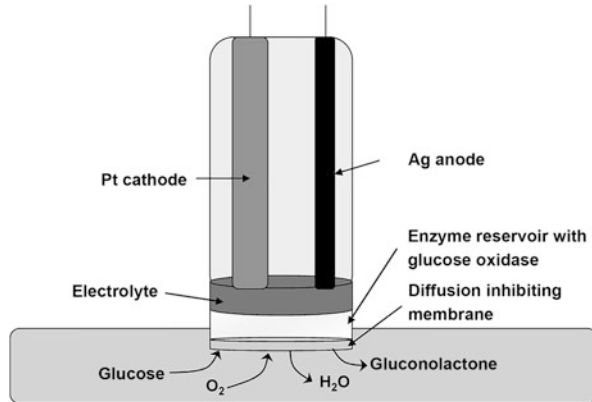
In this detection, glucose is determined *indirectly* via the conversion to hydrogen peroxide by the enzyme glucose oxidase. The detection of glucose is ultimately performed by a classical electrochemical determination of the hydrogen peroxide formed.

The glucose detection according to L. Clark (1962) was one of the first applications in biosensor technology. The enzyme is fixed directly to a platinum electrode in a membrane sandwich arrangement (Fig. 12.18). The resulting hydrogen peroxide is then converted directly by the platinum electrode (see Step 2). The anode potential required for oxidation is in the range of +600 mV against an Ag/AgCl reference electrode. The current flowing during this electrochemical reaction is proportional to the glucose concentration. By means of a calibration of the glucose sensor, a quantitative determination of an unknown glucose concentration is possible. In addition to hydrogen peroxide, the change in pH (ISFET, pH electrode), the oxygen content (oxygen electrode) or the formation of the reaction heat (calorimetry) can also be determined in the glucose determination. For this reason, there is a whole range of different measuring methods available today.

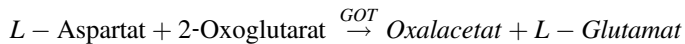
12.4.2 Determination of Enzyme Activity

An example for the detection of enzyme activity is the determination of *GOT* (glutamate oxalacetate transaminase), also known as aspartate aminotransferase (*ASAT*). Liver enzymes occur in elevated concentrations when the liver is damaged. Depending on which of the enzymes are elevated, it is possible to draw conclusions about the type of disease. The level of the enzyme increase in serum corresponds to the extent of the damage to the liver cell. The detection of *GOT* is a composite two-step enzymatic test with the measurement of the kinetic formation of NAD^+ at the wavelengths 334 nm, 340 nm or 366 nm and proceeds in the following steps:

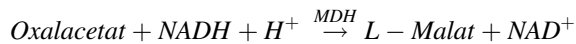
Fig. 12.18 Sensor for glucose detection



Step 1: Enzymatic conversion of the L-aspartate to oxaloacetate using GOT



Step 2: Reduction of the oxaloacetate formed with the aid of the enzyme MHD (malate dehydrogenase) with the oxidation of NADH to malate



Step 3: Determination of the increase of NAD⁺ at 37 °C

The rate of conversion from NADH to NAD⁺ will be measured at a wavelength of 340 nm, which is directly proportional to GOT activity.

Table 12.5 shows some examples of enzyme/substrate systems that are used primarily in medical laboratory diagnostics.

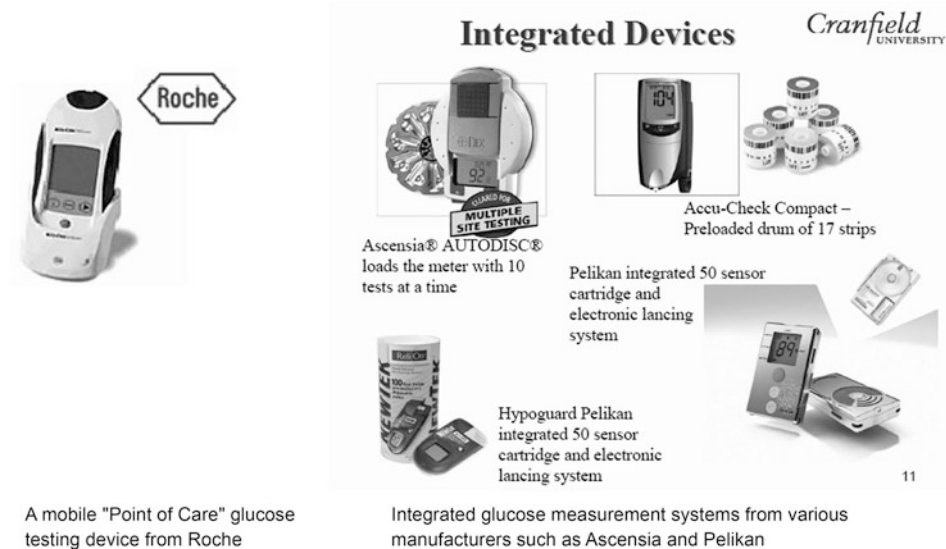
12.4.3 Application Fields of Enzymatic Tests

These enzymatic tests are mainly used in clinical diagnostics. *Glucose detection* is one of the few tests that has also been very successfully implemented in a mobile device as a biosensor and has established itself on the global market for glucose test systems in the *point of the care* sector. The glucose sensor market currently accounts for 87% of the biosensor market [6].

There are a number of different mobile devices for measuring glucose in the blood. *Roche Diagnostics* is the market leader in this field and, together with *Lifescan*, has a 70% market share. Figure 12.19 shows some products.

Table 12.5 Some enzyme/substrate systems

Enzymes	Substrate/analyte
Glucose oxidase	Glucose
Urease	Urea
Creatine Kinase	Creatinine

**Fig. 12.19** Commercial blood glucose meters

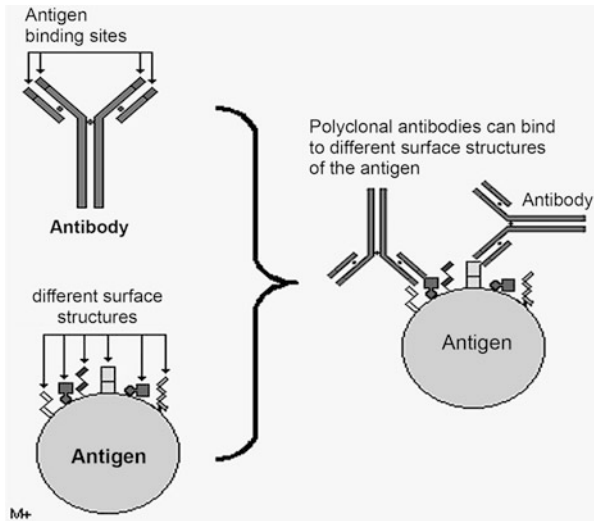
Enzymes also serve as labelling systems for other diagnostic tests and biosensor systems such as *enzyme immunoassays* (Sect. 12.5) or DNA tests (Sect. 12.6). Enzymes play an important role in medicine. Certain substances can inhibit or enhance the effect of enzymes.

12.5 Immunological Methods: Immunosensors

Compared to enzyme sensors, *immunosensors are* systems that use an *antibody* as a *biomolecular recognition element* or receptor. Immunosensors play a special role within the group of biosensors since it is possible to produce an antibody for almost any desired substance.

Antibodies are *glycoproteins* formed by the B cells (B lymphocytes) of the immune system and represent the immune system's response to contact with a foreign substance (antigen = antibody generating). A B cell carries a *specific receptor* on its surface which binds the antigen on the first contact and triggers the so-called *primary immune response*. The B cells proliferate and differentiate into plasma and memory cells, which causes larger amounts of antibodies (*immunoglobulins*) to be released during further immunisation with

Fig. 12.20 Scheme of binding of polyclonal antibodies to different surface structures of an antigen



the same antigen (*secondary immune response*). Since an antigen can be bound to B-cell receptors via different binding substructures (epitopes), different types of B-cells trigger immune reactions. Each type stimulates the production of its own antibody clone, which is why the resulting antibodies are called *polyclonal antibodies* against the respective antigen. Figure 12.20 shows the scheme of binding of *polyclonal antibodies*.

Through the cell-technical multiplication of B-lymphocytes, it is possible to produce monospecific, so-called *monoclonal antibodies*. The B cell is fused with a permanently growing tumor cell (myeloma cell) by fusion. The result is a so-called *hybridoma cell*, which combines the properties of both cell types: antibody production and permanent growth [7]. In addition to the classical methods of antibody production, there are also *recombinant* antibodies produced by *genetic engineering*. Here, the amino acid sequences and the sequence lengths of the variable sequences of the light and heavy chains are *varied* and *multiplied* using genetic engineering methods, thus artificially imitating the *evolutionary mutagenesis* in the natural production of antibodies. The resulting antibody libraries are selected by targeted screening for desired target structures and the corresponding highly specific antibodies are selected.

In addition, artificially produced *templates* are produced, which can imitate the properties of antibodies but are more stable against external influences. Mostly synthetic polymers are used as *antibody templates*. The most important representatives of this group of polymers are the so-called *molecular imprinted polymers*.

More recent developments for the targeted production of specific antibody-like receptors are the use of short, *single-stranded DNA* or *RNA oligonucleotides* (25–70 bases), which can bind a specific molecule via their 3D structure, so-called *aptamers*. They bind to proteins, low molecular weight substances and also to virus particles. Aptamers have dissociation constants in the pico- to nanomolar range. This means that

they bind to their target molecules as strongly as antibodies. This high specificity is achieved by folding the 3D structure of the oligonucleotide precisely around the binding partner. The most important interactions besides the fit are electrostatic interactions, hydrogen bonds and base stacking.

Antibodies are *immunoglobulins (Ig)* and occur in five different classes (*IgA*, *IgD*, *IgG* and *IgM*). Although they are structurally similar, they perform functionally different tasks in the process of controlling the immune response. For immunological detection methods mainly antibodies of the *IgG* class are used. They have a molecular weight of about 150 kDa and represent the largest proportion of total immunoglobulins with 80% or 8–16 mg *IgG* per ml serum. The basic structure of the glycoprotein is formed by two polypeptide chains, the so-called *heavy chains*, of about 450–550 amino acids (about 50 kDa), which are covalently linked to the light chains of about 220 amino acids (about 25 kDa) via disulfide bridges. The heavy chains carry the carbohydrate side chains responsible for the name *glycoprotein*. Figure 12.21 shows the Y-like schematic structure and structural model of a class G immunoglobulin (*IgG*).

The specific binding between antigen and antibody is based on the so-called *key-lock* or *induced fit reaction*, which also plays a central role in enzymatic reactions. The interaction between ligand and receptor is based on the non-covalent binding of numerous functional groups. This leads to a stable three-dimensional complex, the so-called *affinity complex*. Since the formation of the affinity complex requires the interaction of many individual reactions involving the spatial structure of the components, reactions of this type have a *high specificity*. The special case of affinity analysis, in which the affinity of the antibody to a certain antigen is used, is called *immunoanalysis*. *Immunosensors* or *immunoassays* with different structures are used.

The antibody is bound via *antigenic determinants*, so-called *epitopes* consisting of 5–8 amino acids and requires *steric complementarity* of both components. Spatially speaking, there are two possibilities for the formation of an antibody binding site within the antigen protein:

1. The 5–8 amino acids involved are all covalently bound to each other and belong to a polypeptide strand. This case is called a *continuous* or *linear epitope*.
2. The spatial shape of the epitope and thus also the amino acids involved are determined by the spatial folding, that is, the quaternary structure of the protein. This is then a *discontinuous* or *conformational structure* of the epitope.

When binding a protein (antigen), which has a size of about 20 kDa, the interaction with an antibody requires only 5–10% of the surface of the antigen. Depending on the steric nature of the antibody, an antigen can bind 5–10 antibodies.

The specificity in the formation of an *antibody-antigen affinity complex* is due to the fact that the totality of numerous non-covalent bonds that are formed in three dimensions between the antigen and the F_{ab} region of the antibody determines the binding energy (Table 12.1 in Sect. 12.1.1) [9].

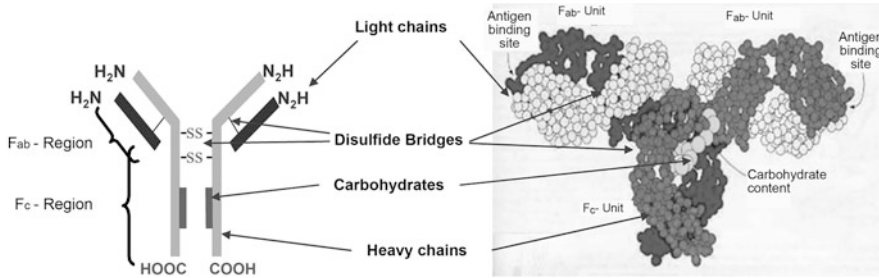
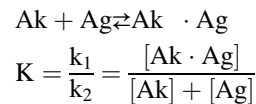


Fig. 12.21 Schematic structure and structural model of an IgG antibody (Source: Harris, Larsson and Hasel [8])

In the affine interaction, *binding energy* (ΔG_0) of about 20–90 kJ/mol is released per antibody bond, which is the sum of all occurring non-covalent bonds.

The antigen(Ag)-antibody(Ak) bond thus depends on the physical and chemical properties of the *reactants* and follows the *law of mass action*:



K: Affinity constant; k_1 : association constant; k_2 : dissociation constant; $[\text{AkZAg}]$: concentration of affinity complex formed in mol/l; $[\text{Ag}]$: antigen concentration mol/l; $[\text{Ak}]$: antibody concentration mol/l.

According to this equation, the size of K gives information about the stability of the affinity complex. Kinetic investigations have shown that k_1 assumes values in the order of 10^7 to 10^8 l/mol for affinity reactions and k_2 ranges between 10^{-5} l/mol (high affinity) and 10^3 l/mol (low affinity), which corresponds to an affinity constant of 10^4 l/mol to 10^{13} l/mol [10].

With immunological analysis methods very different analytes (antigens) can be detected. These are proteins (including cell receptors such as tumour markers, enzymes, allergens, hormones, but also the antibodies themselves), viruses, cells (e.g. tumour cells, bacteria, microorganisms) or haptens (smaller molecules such as toxins, drugs, antibiotics, xenobiotics such as pesticides, explosives). Depending on the analyte, a different method, the *immunoassay format* or *immunoassay design*, is also used.

The test formats can be divided into *homogeneous* and *heterogeneous immunoassays*. In the *homogeneous* immunoassays, *all components are in solution*; the assay is performed without washing steps. They are mostly used as *immunological rapid tests* (e.g. as test strips). In *heterogeneous immunoassays*, one component (antibody or also antigen) is bound to a *solid phase* while the *analyte is in solution*. The solid bound and free phase must be separated from each other. These different methods can be combined with different readout techniques.

12.5.1 Direct Immunosensors

Direct, label-free methods are known for reading the antigen-antibody reaction. *Direct immunosensors* do not require chemical modification of the biological components. The antigen-antibody interaction is observed in *real-time*. *Piezoelectric immunosensors* belong to this group, in which mass sensitive oscillating crystals serve as transducers that provide information about the amount of bound analyte by means of frequency change. *Potentiometric immunosensors* evaluate the potential shift on an electrode surface caused by the antigen-antibody reaction. By means of optical immunosensors (ellipsometry, surface plasmon resonance), the amount of bound analyte is deduced from the increase in layer thickness through the formation of the affinity complex (Sect. 12.2).

12.5.2 Indirect Immunosensors

Indirect immunosensors require a *label*, that is, a chemical modification of one of the immunocomponents, to detect a measurable signal. Depending on the label a distinction is made between *radioimmunoassay (RIA)*, *fluorescence immunoassay (FIA)* and *enzyme immunoassay (EIA)*. In the radioimmunoassay, this marker is a radioactive isotope. By measuring the isotope decay the concentration of bound labelled substance is determined. More often, however, an optical signal is generated. In fluorescence immunoassays, for example, one component carries a fluorescent marker. The intensity of the fluorescence leads to the sensor signal.

Enzyme immunoassays are based on the existence of an enzyme-labelled reaction partner. Common enzymes are alkaline phosphatase, β -galactosidase, urease, glucose oxidase and peroxidase which, after the addition of a suitable substrate, catalyse a cleavage reaction to an optically or electrochemically active molecule. The readout of the sensor signal is then either optical (e.g. photometric) or electrical. Due to the difficult handling of the very sensitive RIA (fast isotope decay, generation of radioactive waste), FIA and especially EIA are mainly used.

Immunoassays are further classified according to their immunoassay format into non-competitive, competitive (Fig. 12.22) and displacement immunoassays (Fig. 12.23).

Non-competitive (sandwich) immunoassay

In the *sandwich immunoassay*, a *capture antibody* is bound to the solid phase and its reaction with the antigen (analyte) in solution is read with a labelled secondary antibody used in excess. A prerequisite for this set-up is a *sufficient size of the antigen*, as it must have at least two epitopes (Fig. 12.22a). The signal measured by the enzymatic substrate cleavage is *directly proportional* to the amount of bound antigen.

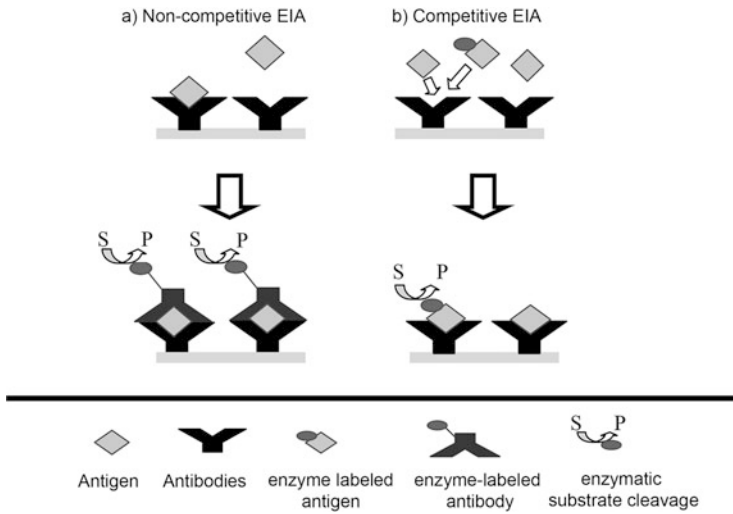


Fig. 12.22 (a) Schematic diagram of a non-competitive sandwich immunoassay and (b) a competitive enzyme-labelled immunoassay

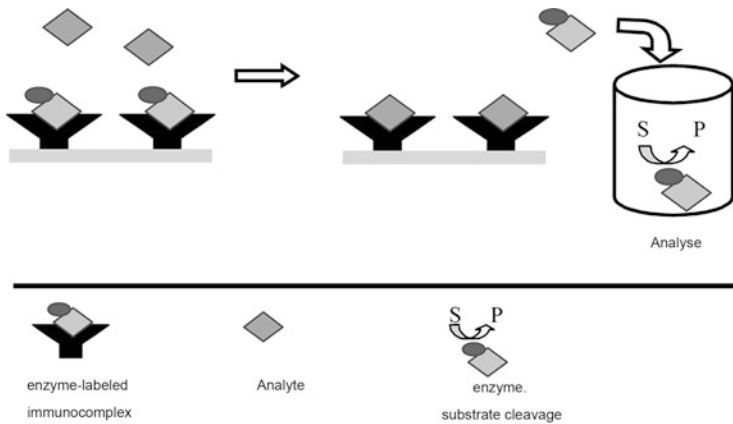


Fig. 12.23 Schematic diagram of a displacement immunoassay

Competitive immunoassay

Competitive immunoassays are preferred when the antigen does not have two binding substructures (epitopes). In this case, the analyte is mixed with an *enzyme-labelled antigen*. Under competitive conditions, a balance is established between enzyme-labelled antigen and unlabelled antigen present in the solution, both of which bind to the capture antibody (Fig. 12.22b). Since the amount of bound enzyme is greater the less analyte was present, the measured signal, in this case, is *inversely proportional* to the amount of analyte.

Displacement immunoassay

The function of the displacement immunoassay is based on the displacement of an already bound labeled immunocomponent by the analyte to be determined. Since normally the dissociation rate of an affinity complex is lower than the association rate, a low-affinity antigen-antibody binding is replaced by a higher affinity interaction. By analyzing the *cleavage products*, the amount of analyte present can be determined directly (Fig. 12.23).

12.5.3 Application Fields of Immunosensors

Immunosensors represent the further development of immunological analysis methods, which are already used today, for example, as *microtiter plate tests* in clinical diagnostics, for *trace analysis of toxic substances* in the pharmaceutical and food industries, and in *environmental analysis*. Compared to electrochemical biosensor systems for low-molecular metabolic products (e.g. glucose, lactate), which have already established themselves as *point-of-care systems*, immunosensors are only used in small numbers in clinical and commercial applications. The reason for this is that the fields of application of an immunosensor in the routine operation of a medical laboratory have not yet been clarified. In addition, technological problems such as the *comparability of the antibodies* on the respective transducer surface, the *orientation and specific properties of the antibodies* and *immobilization* are still unsolved. Prerequisites for the successful establishment of immunosensors in clinical diagnostics as well as in other application areas are that the measurements are performed with a *high degree of precision and accuracy*, that the device *can be fully automated* and that *fast analysis times* can be realized. The greatest potential of such immunosensors lies in their application as mobile analytical devices. These can be the point of care applications for the *home care sector*, in order to carry out a first health check at home and only need to visit the doctor when necessary.

Besides, immunosensors could be used as a mobile analysis device for clinical analysis in the field of emergency medicine: A rapid on-site diagnosis of heart attacks or other life-threatening diseases. In addition, many applications are now seen in the field of environmental analysis. Rapid on-site analysis to determine the contamination of water with toxins or contamination with bacteria could be further new fields of application for such immunosensors. Figure 12.24 shows an example of an immunosensor developed for use in clinical diagnostics. It is used to quantitatively detect a *heart marker cTnI* on site.

12.6 DNA-Based Sensors

DNA (deoxyribonucleic acid, DNA) is the macromolecular building block of genes. It carries the *genetic information of every living being* and some viruses. In humans, the DNA contains over 3 billion base pairs and thus a large amount of important information. This information is encoded in a *code* consisting of only four building blocks, the nucleic acids

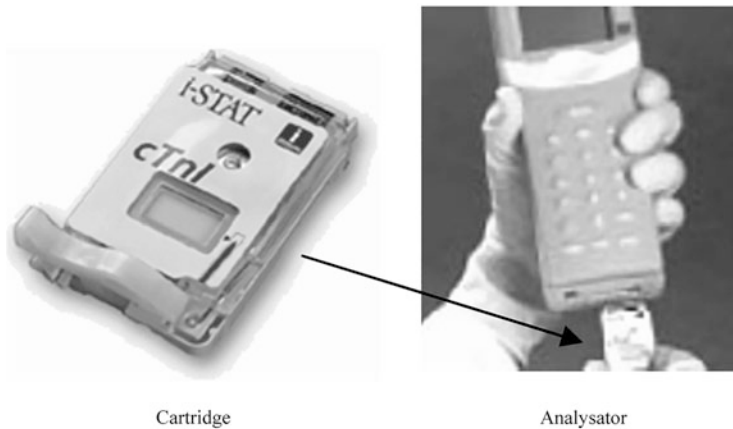


Fig. 12.24 A quantitative amperometric immunosensor for the detection of cTnI (a cardiac marker for heart attack) from i-STAT

adenine, thymine, cytosine and guanine. However, deciphering the genetic code of various organisms is only one part of DNA analysis. *Gene expression analysis*, which examines the regulation of gene activity, and the analysis of *genetic mutations*, in particular *point mutations* (single nucleotide polymorphism, SNP), are also of primary importance.

A *gene* is a section of DNA that contains the *basic information* for the production of *biologically active RNA* (ribonucleic acid). In this production process (*transcription*), a negative copy is made in the form of the RNA. The best-known RNA is mRNA, which translates a protein during translation. A gene consists of two different regions:

1. a DNA section that makes the *single-stranded* RNA copy, and
2. all other DNA segments involved in the *regulation* of this process.

The DNA itself is a *double helix molecule* (Fig. 12.25b). From a chemical point of view, it is a nucleic acid, a long chain molecule (polymer) made up of individual pieces, so-called *nucleotides*. Each nucleotide consists of a phosphate residue, a sugar and one of four organic bases, adenine and thymine, and cytosine and guanine. These are connected to each other via hydrogen bonds (Fig. 12.25a). Within the protein-coding genes, the sequence of the bases determines the sequence of the amino acids of the respective protein: In the genetic code, three bases each stand for a certain amino acid.

12.6.1 Hybridization Diagnostics

The majority of DNA analyses are based on hybridisation diagnostics, that is, the detection of the *specific interaction of two DNA strands* with regard to their complementarity. *DNA microarray technology* is usually used to implement the hybridization technique. A variety

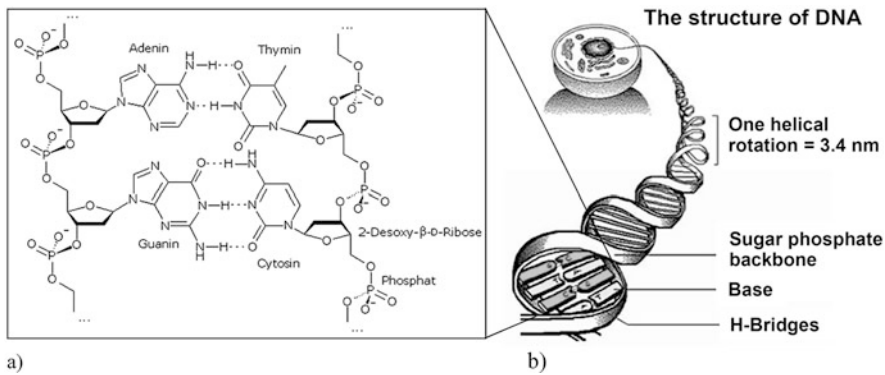


Fig. 12.25 (a) Hydrogen bond of the four nucleotides of the DNA and (b) the double helix molecule of the DNA

of different DNA probes can be immobilized and arranged on these DNA chips. The need for parallelisation is understandable, since, for example, oligomers (DNA probes) with a length of only eight nucleotides are required for sequence analysis by hybridisation. For sequencing, however, approximately 65,000 different probe oligonucleotides are needed, which must be immobilized on a DNA chip and measured in parallel. In addition, it is necessary to miniaturize the system, since the PCR (polymerase chain reaction) preceding the DNA test means that the DNA products to be analyzed are only available in very small volumes of $<100 \mu\text{l}$. Figure 12.26 shows the detection of a DNA sequence using a *fluorescence microarray*. First, the DNA probes are immobilized on the surface of the chip. Then the sample is added, which is either a PCR product or a cDNA already coupled with a biotin marker. On the chip, the respective (single-stranded) DNA molecules bind specifically to the fixed oligonucleotides on the surface according to the hybridization rules. After hybridization, a second marker (streptavidin coupled to a fluorophore or an enzyme) is added, which binds to the already bound biotin molecules. With suitable exposure or addition of a specific enzyme substrate, the characteristic pattern of coloured illuminated dots or electrical measurement signals can be evaluated.

In the field of *DNA chip technology*, the production of microarrays, parallel dosing of DNA and immobilisation of DNA probes on the surface of the chip are of central importance. DNA chips are usually small platelets made of a carrier material such as glass, plastic or a transducer surface (Au , Si_3N_4) on which many different DNA oligonucleotides with known sequences have to be fixed in a dot matrix arrangement. The manufacturing processes of the DNA chips use techniques from semiconductor production.

For the immobilization of a large number of DNA probes the following techniques are mainly used:

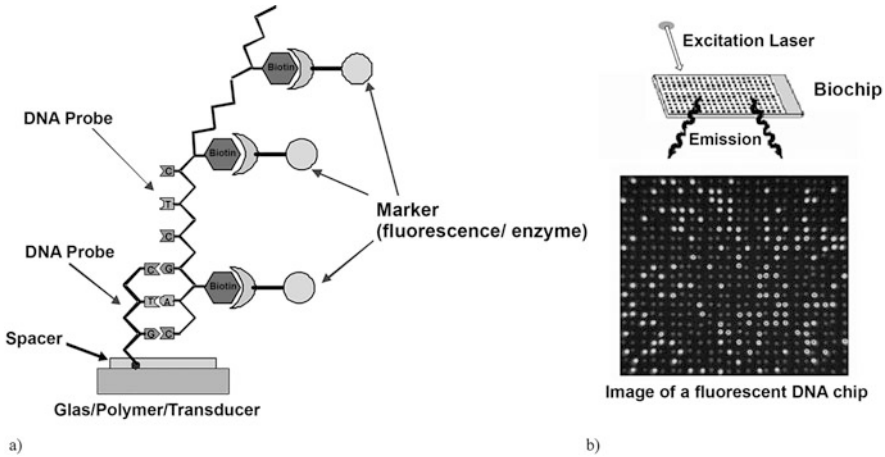


Fig. 12.26 (a) Schematic representation of the hybridization of a DNA sample on a DNA chip and (b) the result of a DNA chip measurement after excitation in the case of fluorescence-labeled DNA samples

1. Photolithographic processes

Single-stranded DNA sequences are built up at the exact positions by light-controlled coupling reactions. Each position contains about 10 million molecules of the respective oligonucleotide.

2. Spotting techniques

In order to dose the reagents required for oligonucleotide synthesis onto the smallest areas or to immobilize already prepared single-stranded DNA probes, *nanospotters* are also used, which function similarly to inkjet printers. They can place very small volumes from 1 to 1.2 nl specifically on surfaces.

Today, a number of 6.5 million assays can already be accommodated on a chip area of 1.3 cm² (Affimetrix). There is a whole range of different pre-coated chip surfaces that can be used for the fixation of oligonucleotides. This is, for example, the coupling via aldehyde, epoxy, streptavidin-modified surfaces, NHS or amino surfaces.

12.6.2 Application and Use of DNA Sensors

In 1994, the company Affimetrix launched the *HIV Gene Chip*, the first commercial DNA chip on the market. Today there are special arrays for genomic DNA, plasmids, PCR products and long oligonucleotides for a wide range of applications offered by various

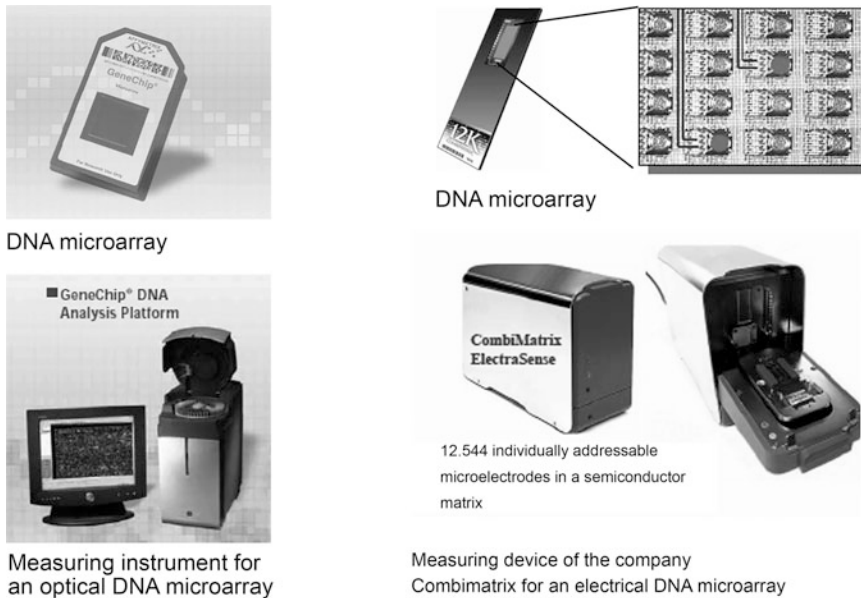


Fig. 12.27 Commercial DNA microarrays (Factory photos: Affimetrix and Combimatrix)

manufacturers. Figure 12.27 shows two commercial DNA chips (an optical affimetrix chip and an electrochemical combimatrix chip).

The current world market for DNA microarrays is US\$800 million. Affimetrix is the market leader.

To study the *gene expression* of, for example, normal cells and tumor cells, the mRNA or cDNA is isolated from the cells and labeled with fluorescent dyes (red and green). The sequence motifs of the genes are immobilised on the DNA chip as a single strand and can hybridise with the mRNA/c-DNA molecules with the complementary base sequence. The position, intensity and wavelength of the resulting mixed colour are detected with a high-resolution laser camera and provide information on the differences in gene expression between the two samples or cells. It is thus possible to elucidate the interaction of the genes in the active metabolic pathways at different cell stages. This is another reason why such DNA chips are a valuable tool in basic molecular biological research.

Another important application is *microbial and virus diagnostics*. The DNA chips carry the characteristic DNA sequences of pathogenic germs. Therefore, specific detection of microorganisms from food, (waste) water or tissue samples and blood can be carried out. The great advantage of the genetic detection of microorganisms is the speed of the examination. While a normal detection of bacteria or viruses can take several days, the analysis can be carried out within a few hours using a DNA chip.

A very large and important field of application of DNA analysis is *tumor diagnostics*. A chip has already been developed that represents 18,000 different variants of tumor genes.

This enables very similar forms of B-cell lymphoma to be distinguished on the basis of their gene activity pattern. It was found that one form of the disease, which accounts for around two-fifths of cases, responds to cytostatic chemotherapy, whereas the other variant does not.

The so-called *single nucleotide polymorphisms (SNPs)* are small variations in individual DNA base pairs that are responsible for the development of many diseases. They occur in entire genomes (depending on the region) at intervals of 100–2000 base pairs. DNA chips are particularly suitable for the rapid diagnosis of such *individual disease variants*. For this purpose, chips are used which represent all combinations of a certain *oligonucleotide*. Small differences between two individuals can be detected by means of the *different hybridization patterns*. The correlation of the individual polymorphism patterns with the varying efficacy and tolerance of drugs then provides information for individualized therapy.

12.7 Cell-Based Sensor Technology

As already discussed in Sect. 12.1 on biological sensory analysis, cell tests are mainly used in pharmaceutical research and drug development to determine the effect, toxicity and influence on the metabolism in the body of certain substances. The biosensors described in previous sections can usually only be used to obtain specific structural information about substances. The presence and effect (e.g. enzyme activity) of certain previously known substances, but do not so much determine functional effects. The effect of substances and the influence on the metabolism of cells are of great importance, particularly in the field of *pharmaceutical screening*, since drug effects are often very complex and can only be reduced to simple receptor-ligand interactions in selected cases. For this reason, it is often necessary to include whole cells or even tissue parts in the test. In addition, there are currently approaches to also investigate the quality of water using a so-called *effect sensor* such as *cell sensor technology*. Instead of the costly fish or daphne tests for rapid determination of water quality and water toxicity, great potential is seen in the use of whole cells. As with drugs, the effect of hazardous substances present in water could also be determined with a cell sensor. As described in Sect. 12.2, there are various classical cell tests in the pharmaceutical sector that investigate the effect of substances. More recent developments are therefore aimed at integrating whole cells (or tissue) as receptor surfaces in combination with different transducers in a *bio(cell) sensor system*. The living cells are cultivated directly on the surface of a carrier, where they form a cell lawn that is as homogeneous as possible. Ideally, a semiconductor material is used as a carrier in which microelectrodes, transistors (e.g. ISFETs), small sensor elements or complete semiconductor circuits can be integrated relatively easily. Two different fields of application are currently seen for such types of cell chips. These are on the one hand the *metabolic cell chip*, which studies parameters of cell metabolism, and on the other hand, the *neurochip*,

which can stimulate nerve cells as well as to measure the derivation of potentials and thus can determine the neurotoxicity of substances.

12.7.1 Metabolic Cell Chip

There are various approaches to examining the metabolism of cells, which can also be integrated into a silicon-based CMOS sensor. The change or the influence on the metabolism can be investigated, for example, via cell respiration, which is reflected in the *change in oxygen consumption*. Cell respiration as well as the glucose metabolism of the cell can also be studied by measuring the *change in acidification performance* (i.e. the change in pH). In addition, the *morphological change of cells* or the detachment of cells from the cell chip provides an indication of metabolic processes or the condition of a cell. There are various approaches to investigate these *metabolic parameters* in cells directly and online. However, in most systems, only one of the parameters described above is measured. Currently, there is a silicon chip that has integrated these metabolic functions into one system. This system from the company Bionas consists of a *metabolic silicon chip* that has integrated an interdigital structure to measure cell adhesion, a Clark-like electrode to measure oxygen consumption and an ISFET to determine the *acidification performance of the cells*. The cell chip is integrated into a module that is directly suitable for growing the cells. With this system, all three parameters can be examined online simultaneously. Figure 12.28 shows the cell chip, cells cultivated on the cell chip, and the entire chip module for measuring metabolic parameters.

12.7.2 Neuro-chip

So-called *neurochips* are used to investigate neurotoxicity in pharmaceutical screening. These neurochips contain, for example, field-effect transistors (FET) that can measure potential changes in the cell. If a receptor interacts with a suitable analyte (ligand), an ion channel is opened via a complex signal cascade. The *ion current causes* a change in potential, which must be detected [11]. Since many nerve cells also require an electrical stimulus to switch an ion channel, the cell can also be contacted by a fine electrode. Alternatively, the stimulation can be provided by microelectrodes located on the silicon substrate. This measuring arrangement is also known as a *neuro-transistor* (use of neurons as cell elements). Figure 12.29 shows a neuro-chip developed at ETH Zurich for the investigation of heart and nerve cells.

There are a number of other approaches and measuring arrangements to investigate the functionality of electrogenic cells, for example with a *translucent cell chip*. The cell is located directly above a hole in the chip through which laser light is irradiated. This allows molecules in the cell to be excited to fluoresce, which are activated within the signal cascade between receptor and ion channel. Such an approach has already been realized for

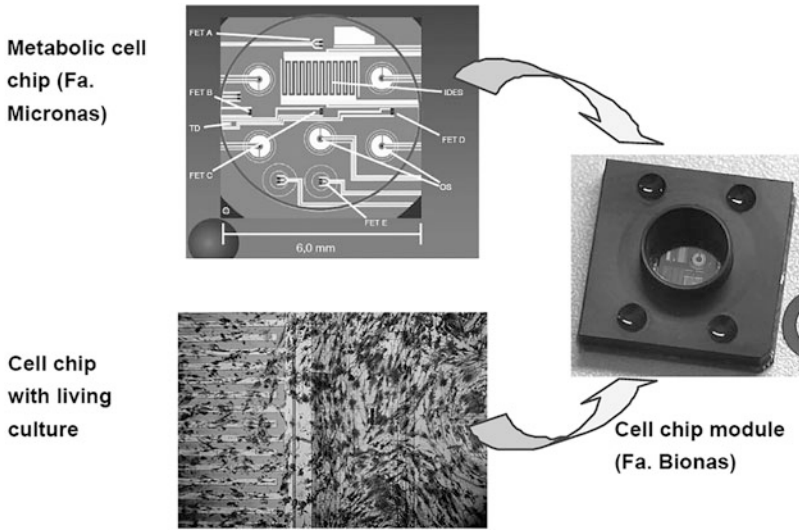
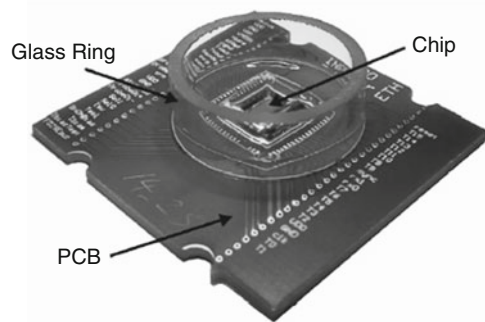


Fig. 12.28 Structure of a cell-based sensor for measuring the cell metabolism

Fig. 12.29 Neuro-chip for measuring the activity of heart or nerve cells according to Heer [12]



a biosensor for the detection of L-lymphocyte activators and active ingredients that interact with G-protein coupled receptors [13].

The application of cell-based sensors is mainly seen in the *screening of new active pharmaceutical ingredients*. In addition, these novel systems offer the possibility to screen new substances and chemicals for their toxicity and metabolic effects. In the future, this could be a new testing method for the *approval of new substances*. Completely new applications are seen in the field of *water quality* and *water safety*. Cell-based sensors are currently being developed that will in future be able to detect very quickly when harmful concentrations of substances that have been introduced into or accumulated in the water cycle by the environment are reached (e.g. the accumulation of pharmaceutical residues in water) or when toxic substances enter the water (active or passive).

Bibliography

1. Sethi, R. S.: Transducer Aspects of Biosensors, *Biosens. Bioelectron* (1994), 243–264
2. Alberts, B.; Bray, D.; Lewis, J.: *Molekularbiologie der Zelle*. 1990, 2. Auflage (VCH Weinheim)
3. Tschmelak, J. et al.: *Biosensors and Bioelectronics* 20 (2005), 1499–1508
4. Clark, L. C.; Wolf, R.; Granger, D.; Taylor, Z. (1953): Continuous recording of blood oxygen tensions by polarography. *J Appl Physiol.* 6, 189–193
5. Severinghaus, J. W.; Astrup, P. B. (1986): History of blood gas analysis. IV. Leland Clark's oxygen electrode. *J Clin Monit.* 2, 125–139
6. Newmann, J. D. et al. (2005): Home blood glucose biosensors: a commercial perspective. *Biosensors and Bioelectronics* 20, 2435–2453
7. Köhler, G.; Milstein, C.: Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, 1975. 256: pp. 495–497. Roitt, I. M., J. Brostoff, and D.K. Male, *Kurzes Lehrbuch der Immunologie*. 1995. 3. Aufl., Thieme Verlag, Stuttgart, New York
8. Harris, L. J.; Larsson, S. B.; Hasel, K. W.: Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry*, 1997. 36: p. 1581
9. van Oss, C. J.: *Antigen-Antibody-Reactions. Structure of antigens* (van Regenmortel, M. H. V.), 1992 (CRC Press, Boca Raton): p. 99–125
10. Steward, W. M.; Steensgaard, J.: *Antibody Affinity: Thermodynamic Aspects and Biological Significance*. CRC Press, Florida, 1983. Niessner, R. and et al., Development of a high sensitive enzyme-immunoassay for the determination of triazine herbicides. *Fresenius J. Anal. Chem.*, 1997. 358: pp. 614–622
11. Vassanelli, S.; Fromherz, P.: *Journal of Neuroscience*, 1999, 19(16):6767–6773
12. Heer, F. et al.: CMOS microelectrode array for the monitoring of electrogenic cells, *Biosensors and Bioelectronics*, 20 (2), 358–366 (2004)
13. Zahn, M.; Renken, J., Seeger, S. (1999): Fluorimetric multiparameter cell assay at the single cell level fabricated by optical tweezers. *FEBS Lett* 443:337–340



Hartmut Bärwolff

13.1 Introduction and Physical Quantities

Humans are surrounded by many types of radiation. These include solar radiation, that is, light, heat radiation from a radiator, radiation from a microwave oven or cosmic radiation from the depths of the universe. Radiation is thus part of our environment; it can be produced naturally or by technical devices (Fig. 13.1). In general, radiation is a form of energy that propagates in space as a *particle flow* or as an *electromagnetic wave*.

Energy is the most important criterion for distinguishing between ionising and non-ionising radiation. *Ionising radiation* is a term for any particle or electromagnetic radiation that can *ionise atoms* or *molecules*, that is, able to remove electrons from the electron shell. These are usually ionisation energies of more than 5 eV (corresponds to the energy of an electron when passing through a voltage difference of 5 V). Ionising radiation can also emanate from radioactive materials. However, the term radioactive radiation should be avoided, because it is not the actual radiation that is radioactive, but the substance that emits the radiation.

- If it is *electromagnetic radiation*, the ionization energy of 5 eV corresponds to wavelengths of less than about 200 nm. Therefore, only gamma radiation, X-rays and short-wave ultraviolet radiation can release electrons from the atomic shell or break atomic bonds.
- *Charged particles* (e.g. protons (p), electrons (e) or mesons) are also counted as ionising radiation from 5 eV upwards.

H. Bärwolff (✉)
Hochschule Köln, Köln, Germany
e-mail: baerwolf@gm.fh-koeln.de

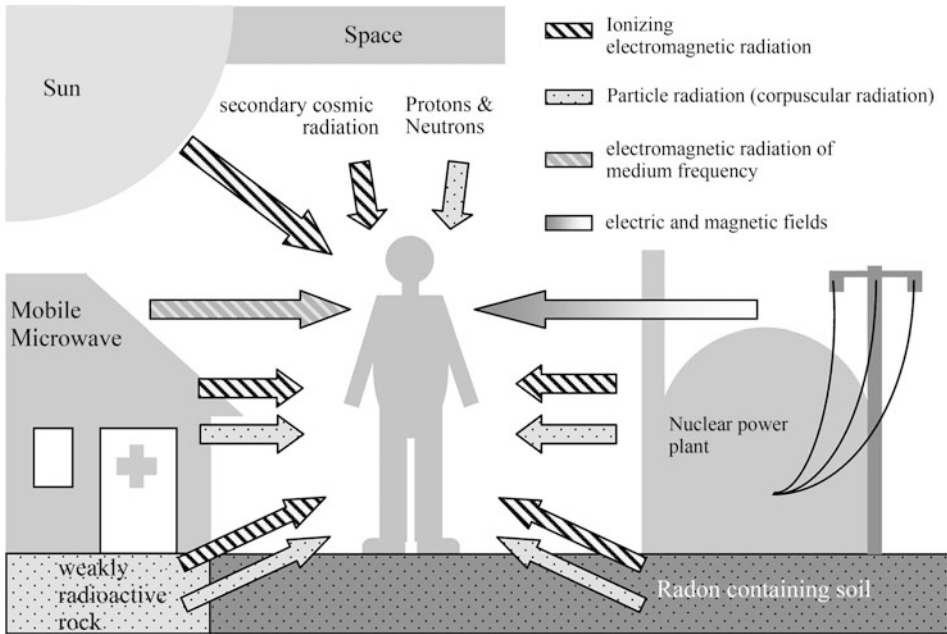


Fig. 13.1 The different types of radiation

- *Neutral particles* such as neutrons are also counted as ionising radiation. Here, ionisation is achieved by indirect nuclear reactions or scattering processes.

Figure 13.2 illustrates the conditions. Wilhelm Conrad Röntgen discovered the X-rays in 1895, which were later renamed X-rays. It should be noted that there is a *dualism* in the sense of quantum theory between the *wave* and *particle aspect* of radiation. Thus a wave can be seen as a particle and vice versa. Wave and particle radiation can hit the earth as *cosmic rays* or be generated by technical equipment such as X-ray tubes or accelerators. In this respect, cosmic gamma radiation is also listed under cosmic particle radiation.

Alpha (α -), beta (β -) and gamma (γ -) radiation is produced during *radioactive decay*. Certain atomic nuclei, the radionuclides, decay under the emission of radioactive radiation until they reach a stable final state (Fig. 13.3). *Alpha radiation* consists of *double positively charged helium nuclei*. This radiation can neither penetrate the skin of a human being nor paper. *Gamma radiation* is *high-energy electromagnetic radiation*. It has a *high penetration capacity* and can be shielded by materials such as lead or concrete. *Beta radiation* consists of *electrons* or *positrons*. It has a *medium penetration capacity*, that is, in the air some centimetres to metres, in human soft tissue up to the centimetre range.

As shown in Fig. 13.4, warning signs must be used to warn of radioactive radiation.

The effects of this radiation on humans are undisputed and depend on many factors. Only limited attention is paid to this. This also applies to numerous applications. The

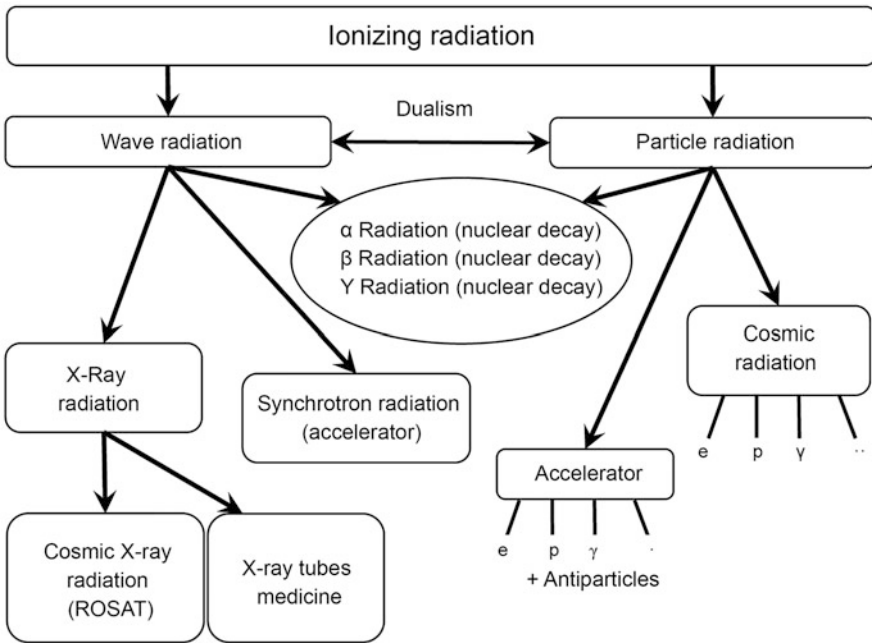


Fig. 13.2 Classification of ionizing radiation

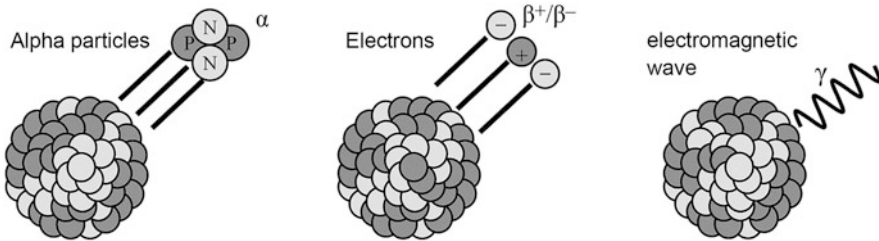


Fig. 13.3 The origin of radioactive radiation (*N* neutrons, *P* protons)

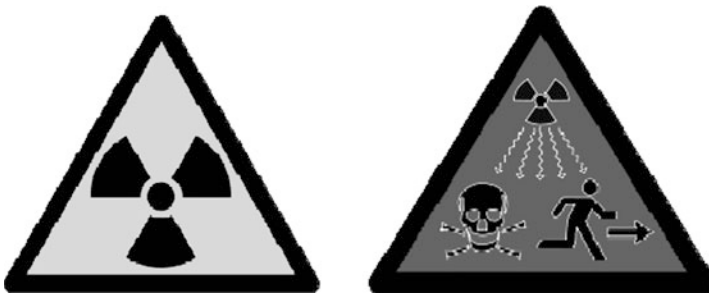


Fig. 13.4 Warning signs against radioactive materials or ionising radiation

measurement and evaluation of ionising radiation will be described below. The following four basic parameters serve this purpose:

- Radiation energy,
- Activity,
- Absorbed dose and
- Equivalent dose.

Radiant energy

The radiation energy E is an important quantity for the description of radiation. X-rays and gamma rays transport energy through photons without rest mass. The energy E results to:

$$E = hf = hc/\lambda \quad (13.1)$$

h : Planck's quantum of action ($h = 6.626 \cdot 10^{-34}$ Js), f : frequency, c : speed of light ($c = 2.9979 \cdot 10^8$ m/s), λ : wavelength.

When passing through matter, this radiation does not emit its energy uniformly along the path, but rather at points. This is called *indirect ionizing*.

The particle beams, on the other hand, have a rest mass m . The kinetic energy E_{kin} is proportional to the mass of the particles m and the square of the particle velocity v .

$$E = mv^2/2. \quad (13.2)$$

This radiation is called *directly ionizing* and is described by the Bethe-Bloch formula (Eq. 13.10 in Sect. 13.2). If *differently charged* particles are in a magnetic field, they can be distinguished by their *deflection*. In general, the energy of radiation is expressed in electron volts (eV). The energy of radiation can be determined very accurately with *detectors* (Sect. 13.2). This is the task of *energy spectroscopy*.

Activity

The unit of activity A is the Becquerel. It describes the number of decays per second (dN/dt):

$$A = dN/dt. \quad (13.3)$$

The radioactive decay law applies:

$$N = N_0 e^{-(\lambda t)}. \quad (13.4)$$

A certain number of nuclei N_0 decays exponentially; λ is the decay constant and t the time. The *half-life* $T_{1/2}$ is the time after which half of the initially present atomic nuclei have decayed. The activity can be measured with a detector and a counting device.

Absorbed dose

The activity of a radiator does not yet provide any information about how *dangerous* the radiation is and what *biological damage is to be* expected. The absorbed dose D is the *amount of energy* dE per unit mass dm *absorbed* by an irradiated object over a period of time. It depends on the intensity of the irradiation, the absorption of the irradiated material, the type of radiation, the radiation energy and geometric quantities. It is measured in Gray (Gy) ($1 \text{ Gy} = 1 \text{ J/kg}$). It applies:

$$D = dE/dm. \quad (13.5)$$

In principle, the measurement can be carried out via the small temperature increase due to radiation absorption in the volume and is correspondingly difficult (Sect. 13.2).

Equivalent dose

The absorbed dose D , measured in Gy, describes the purely physical energy deposition. No statement can be made about the biological effect of the radiation. The *relative biological effectiveness* depends on the type of radiation, the energy, the time of exposure and other influencing variables. Therefore, a quality factor w_R (*radiation weighting factor*) has been determined depending on the type of radiation. The radiation weighting factor for some types of radiation can be seen in Table 13.1. It also depends partly on the energy of the particles or the radiation.

The *equivalent dose* H (measured in Sievert; Sv; $1 \text{ Sv} = 1 \text{ J/kg}$) is a measure of the strength of the *biological effect* of a given radiation dose. It is obtained by multiplying the absorbed dose D by the radiation weighting factor w_R , which describes in a simplified way the relative biological effectiveness of the radiation in question. It applies:

$$H = w_R D. \quad (13.6)$$

The radiation weighting factor w_R is equal to 1 for beta and gamma radiation; the equivalent dose H is therefore numerically equal to the absorbed dose D . For other types of radiation, factors of up to 20 apply (e.g. α particles).

Table 13.1 Radiation weighting factors w_R for different types of radiation

Type of radiation	Radiation weighting factor
α -Radiation	20
β -, γ Radiation	1
X-rays	1
Fast neutrons	10
Thermal neutrons	3
Protons	5–10
Heavy recoil cores	20

Ion dose

Furthermore, the ion dose I is of interest. It is a measure of the *strength of the ionization*, expressed by the released charge Q per mass m (unit of measurement J/kg) of the irradiated substance. It can be determined using ionisation chambers (Sect. 13.2). The following applies:

$$I = dQ/dm. \quad (13.7)$$

All dose variables can also be recorded per time unit. Then the corresponding *output quantities are obtained*.

13.2 Interaction of Ionising Radiation with Matter

In the previous section, the basic physical measurands were described. In addition, depending on the physical problem, the temporal occurrence of radiation, the pulse, the location, the range, the half-life, the emission angle or even other quantities can be of interest. For the detection of radiation, the *phenomenology of the interaction of ionizing radiation with matter* is an important topic. Figure 13.5 gives an overview of this.

The radiation is roughly divided into *charged particle radiation*, *neutral particle radiation* and *electromagnetic radiation*. Ionising electromagnetic radiation (gamma or X-ray quanta) does not ionise continuously along its path like particle radiation, for example. The interaction of these quanta with matter essentially takes place through one of the following three processes: *Photoelectric effect*, *Compton effect* and *pairing effect*. These are described below.

Photo effect

If a high-energy quantum hits an atom, it can release an electron from its inner shells. As a *secondary process*, a characteristic *X-ray* or an *Auger electron* can be produced. Figure 13.6 illustrates this process. This effect can also occur at the nucleus as a *nuclear photoelectric effect*, and within the band model at the simple semiconductor or the pn junction. It is used

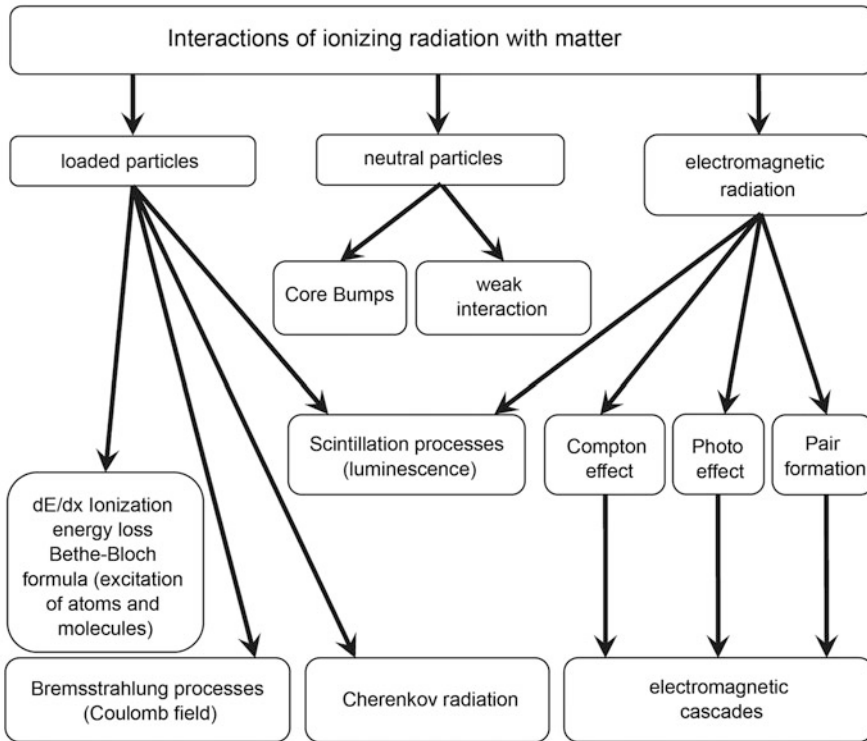


Fig. 13.5 Interaction of ionising radiation with matter

in many ways for *particle detection*. The electron generated by the high-energy radiation can be read out and amplified electronically.

If the quantum is not ionizing, we speak of the *photoelectric effect* (Fig. 13.7). This also plays a role in the detection of ionizing radiation when the ionizing radiation is converted into visible light by a luminescence effect and the light is converted into electrons by the photoelectric effect in a second process.

Compton effect

The Compton effect describes the interaction between a photon with a free or quasi-free electron. If an X-ray quantum hits a material whose electrons are only bound to the atomic lattice with low binding energy, *impact processes take place* between the photons and the electrons. The energy of the quantum $E = h f$ (f : frequency) is partially transferred to the electron as kinetic energy E_{kin} . The outgoing quantum has the energy $E' = h f'$. The following equation and Fig. 13.8 clarify the situation:

Fig. 13.6 Photoelectric effect and ionization of an atom

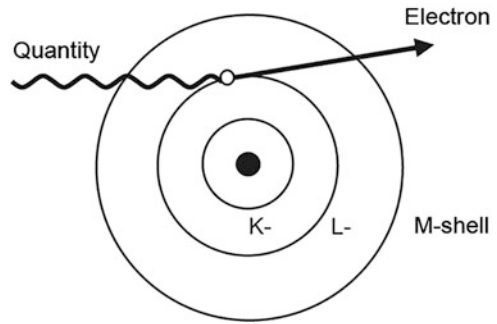


Fig. 13.7 Photoelectric effect

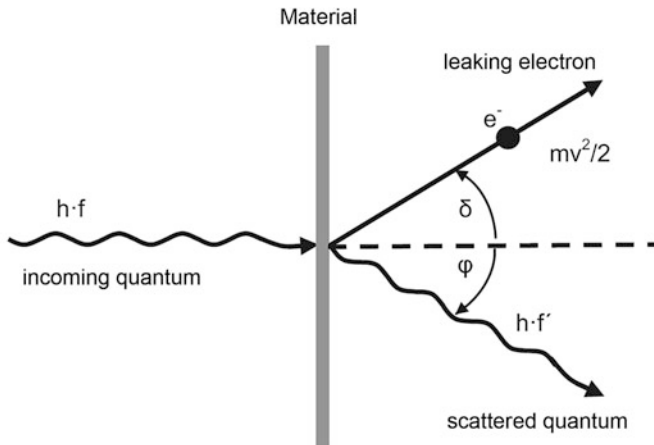
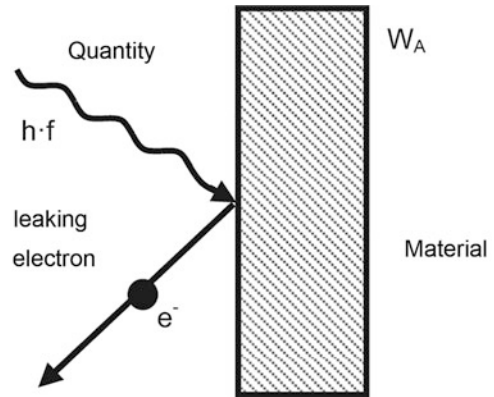


Fig. 13.8 Compton effect

$$E = hf = hf' + mv^2/2. \quad (13.8)$$

Therefore, behind the scattering object, besides the unscattered photons, *scattered photons* can be detected, which have a *deviating wavelength* and thus energy. The wavelength difference $\Delta\lambda$ is (m_0 : mass of the electron):

$$\Delta\lambda = \frac{h}{m_0 c} (1 - \cos \varphi). \quad (13.9)$$

The wavelength difference $\Delta\lambda$ depends only on the angle φ between the primary radiation and the scattered radiation. The Compton effect is an important effect of the detection of *photons*.

Pair formation effect

Although the pairing effect is less interesting for particle detection, it plays an important role in the interaction phenomenology. A high-energy quantum can be split into an electron and a positron in the Coulomb field of a heavy atom. The process is shown in Fig. 13.9.

Which of the three effects described above predominate depends on the energy of the photons and the mass numbers of the atomic nuclei involved. With charged particles, the *interaction takes place via ionization and excitation*. For example, an electron produces ionized atoms and free electrons along a path. A large number of impacts take place. The energy loss of a particle when passing through matter depends mainly on the charge, the speed and the properties of the matter. This is described by the Bethe-Bloch equation (Eq. 13.10 in simplified form):

$$\frac{\Delta E}{\Delta x} \approx \frac{z^2 Z}{v^2 A} \ln(\alpha E). \quad (13.10)$$

Here is: ΔE : energy release on the route Δx , z : particle charge; Z : nuclear charge number of the material, v : velocity of the particle, α : material constant, E : total particle energy.

The range of alpha radiation or beta radiation is quite different. For example, alpha particles with a few MeV energy are stopped after only a few cm in air. The range of beta radiation of the same energy could be a few meters. Equation (13.10) must be extended for very high energies. Additional radiation mechanisms such as brake radiation or Cherenkov radiation are gaining in importance (Fig. 13.5).

The detection of *neutral particles* is difficult and depends strongly on the particle type. Neutrons will only be briefly discussed here. Usually, the *secondary charged products* of the interaction of a neutron with nuclei or other particles will be detected. *Different processes* can take place at nuclei (e.g. elastic or inelastic scattering or absorption with particle emission).

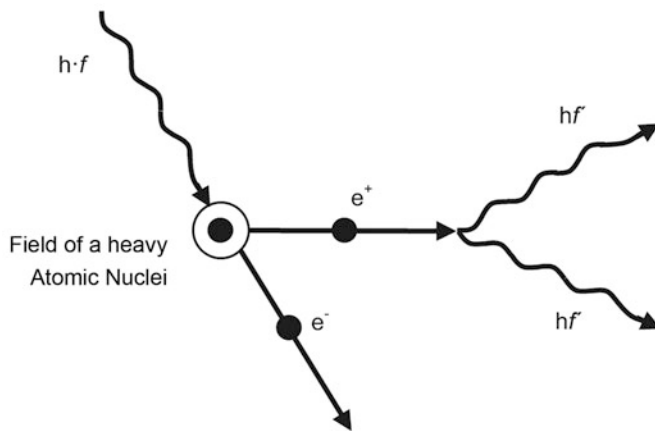


Fig. 13.9 Pair formation effect

Figure 13.10 shows the various physical processes involved in the interaction of ionizing radiation with matter. The length of the process chains is a measure for the ranges.

Radiation exposure of humans

Ionising radiation occurs in lower doses than *natural radiation exposure*. This consists, among other things, of cosmic radiation and radiation from radioactive substances that occur naturally in the earth and the atmosphere, such as radionuclides. In principle, humans come into contact with radiation through the ingestion of food, inhalation of the air we breathe, exposure in the form of cosmic radiation or contamination with radiation. The latter is normally not the case. However, the other two possibilities mentioned are given in real life. For example, certain building materials or even foodstuffs emit ionising radiation. In the medical application of diagnostic procedures such as X-rays or computer tomography, one does indeed come into contact with increased doses of ionising radiation. Radiotherapy to combat cancer uses even higher doses of radiation.

13.3 Classification of Sensors

The sensors or detectors for the detection of ionising radiation can be classified according to their different physical principles of action Table 13.2 shows a classification according to their *detection size*. In the simplest case, the *photographic* or *chemical effect* of the high-energy radiation can be exploited, as with simple *dosimeters*. In this respect, *light centers* can also be used. In the vast majority of cases, *electrical charges* are used for detection, which are generated in *gas meters* by direct absorption. In a *semiconductor meter*, the radiation can also be directly absorbed or detected via the detour of conversion into visible light.

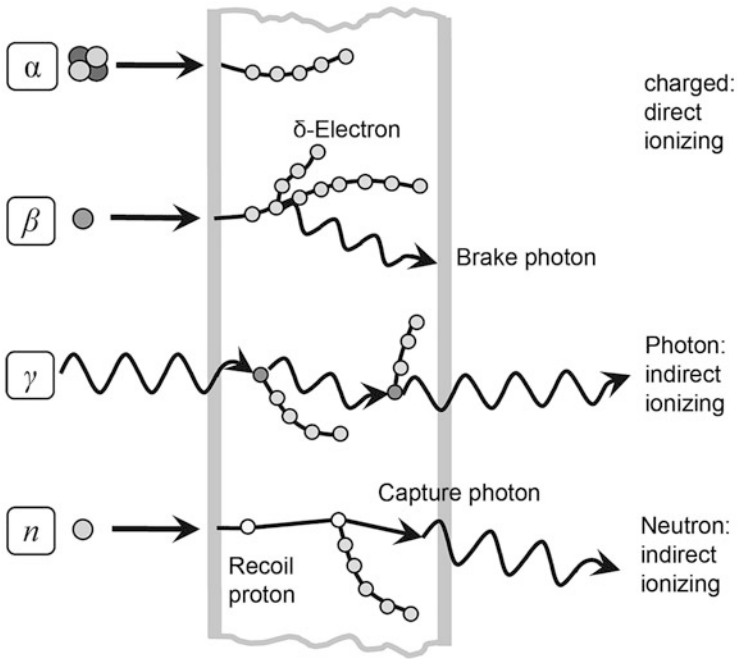


Fig. 13.10 Interaction of ionising radiation with matter

Table 13.2 Measuring principles for the detection and dosimetry of ionizing radiation

Detection size	Detector
Electric charge	Ionization chamber
	Proportional counter tube
	Photomultiplier
	Semiconductor detector
Light	Scintillation detector (optical)
Light centres	Thermoluminescence dosimeter
	Phosphate glasses
Chemical reactions	Iron sulphate dosimeter
Photographic effect	Film dosimeter

Figure 13.11 provides an overview of the many possibilities, especially with regard to modern semiconductor components. The solid-state detectors (FK detectors) can essentially be divided into PIN diodes, avalanche diodes (APDs) or CCD elements (charge-coupled devices). These components, like the photomultipliers, can also be arranged in arrays, which allows the location of the detector to be determined. Solid-state ionization chambers (FK-IK) and lithium drifted Si or Ge detectors will not be discussed in detail here. As already mentioned at the beginning, the different methods and detectors have to be

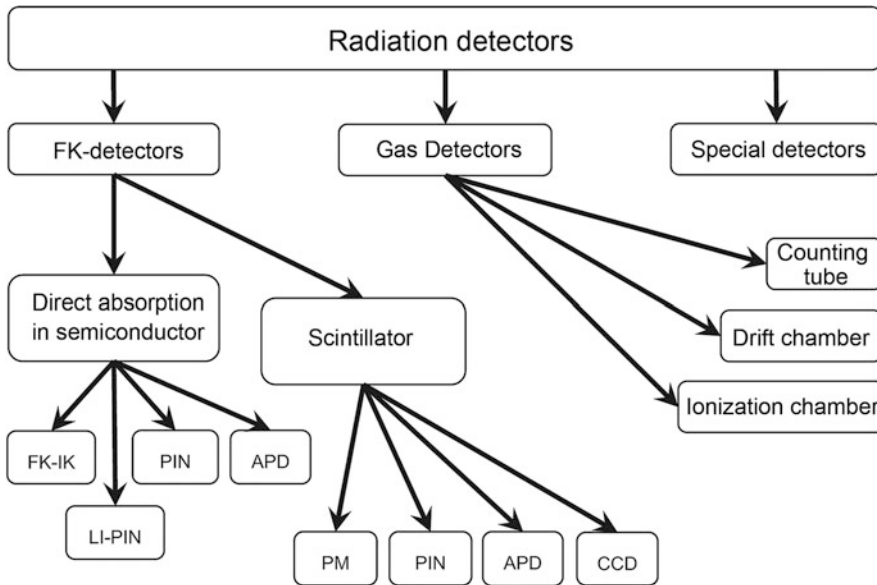


Fig. 13.11 Classification of radiation detectors

used to perform energy spectroscopy or, as in the case of radiation protection, dosimetry measurements.

In the large particle accelerators, on the other hand, *location information* is very often important in addition to energy in order to reconstruct the tracks of the particles. The *semiconductor detectors* can be used in conjunction with a *scintillator* or as directly absorbing sensors. The *gas detectors* can be roughly divided into ionization chambers, counter tubes and drift chambers. In addition, there are the fog, bubble or streamer chambers, which are not included in the illustration, as they are no longer of great practical importance today. They will be discussed briefly in the following section. There are also many other special detectors.

In the simplest case, ionising radiation can be detected via *photoemulsions*. An emulsion layer changes due to the radiation effect. The coloration depends on the type, quantity and energy of the radiation. These structures represent a simple possibility of the documentary radiation-proof. The evaluation is cumbersome and the procedure is inaccurate. It can be used in different variations for the surveillance of persons. Figure 13.12 shows various film dosimeters that are used in nuclear medicine at the Gummersbach District Hospital (Kreiskrankenhauses Gummersbach (KKH GM)).

Calorimetric methods based on the temperature increase of water are also used for calibration and adjustment. These are used, for example, in *absolute dosimetry* at the Physikalisch-Technische Bundesanstalt (PTB) in the form of *dose standards*. This also applies in a similar form to the iron sulphate detector, which makes use of chemical reactions.

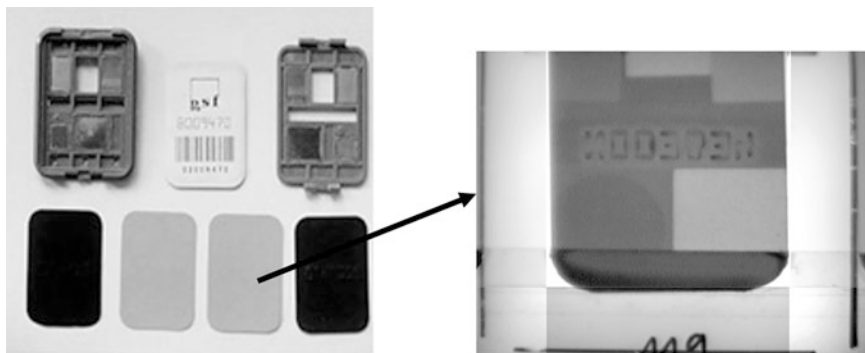


Fig. 13.12 Various film dosimeters (Plant photo: Nuclear Medicine KKH GM)

The basic electronic arrangement of a detector as a block diagram is shown in Fig. 13.13. Ionizing radiation can be converted into *light* via the *scintillation effect* or absorbed directly in a photodiode or APD. Subsequently, the electronic signal is amplified, whereby the signal-to-noise ratio is particularly important. The amplifier can, for example, be a charge amplifier (LEV) or a trans-impedance amplifier. The detector and amplifier are often integrated and cooled, which helps to improve the resolution. Subsequently, a filter is often used for pulse shaping, for example, to generate a unipolar Gaussian pulse. This can be evaluated in analog or digital form, further methods may follow.

The application spectrum of radiation detectors for ionizing radiation is very broad. In the field of *medicine*, these detectors are used in *radiation diagnostics*, for example in *computed tomography* or *positron emission tomography* (PET), and in radiation therapy for the precise application of radiation and quality control.

X-ray detectors are used for security checks at airports or in other security-critical areas. In materials science, ionizing radiation is used for structural analysis or non-destructive detection of critical weak points (e.g. in turbines). The decoding of the human genome in biology or the production of ultra-small chips in the context of LIGA technology is carried out using X-rays or particle radiation. The applications range to art or archaeology. For example, mummies or paintings can be examined with X-rays to clarify their origin. The age of bones can be determined using the C-14 method. The modern environmental analysis uses ionizing radiation.

13.4 Gas-Filled Radiation Sensors

The most important gas-filled detectors in practical use are *ionization chambers* and *counter tubes*. The *bubble chamber* is another gas-filled particle detector that makes traces of charged elementary particles visible. The bubble chamber is similar in design to the cloud chamber, which can be divided into expansion or continuous cloud chamber. Large

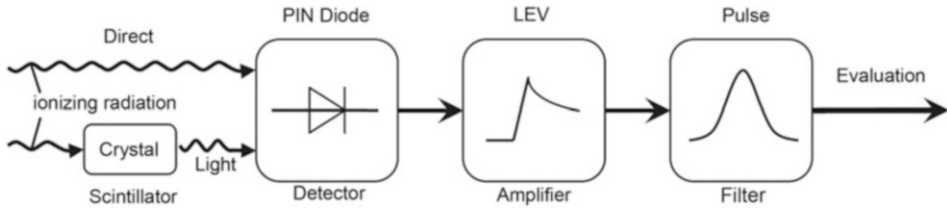


Fig. 13.13 Basic structure of a radiation detector

volumes can be detected; due to the lack of trigger possibilities, the importance of these detectors is low.

A *bubble chamber* is often filled with liquid hydrogen into which the particles to be investigated enter. Shortly before entry, the pressure inside the chamber is reduced considerably so that the temperature of the hydrogen is above the boiling point. The incoming particles now ionize hydrogen atoms, which serve as nuclei for gas bubbles. These can then be photographed and evaluated. The bubble chamber is usually located in a magnetic field, so that charged particles follow a curved path. From this, the ratio of mass to charge, the momentum and, in the case of decaying particles, the lifetime can be determined. The bubble chamber has hardly any practical significance anymore and even further developments such as the streamer chamber have been replaced by electronic detectors today.

Table 13.3 shows a comparison of different detector systems for the energy of 1 MeV. The values represent guide values that may vary in individual cases.

One of the simplest radiation detectors is the so-called *ionization chamber*. The ionizing radiation produces a specific primary ionization in the sensitive volume. The generated electrons and ions can be collected in a constant electric field. Figure 13.14 shows the characteristic curve of a gas discharge or counter tube.

At low voltages, as in the ionization chamber, a few electron-ion pairs are generated: the detector operates without amplification. If the voltage is further increased, the number of charge carriers increases in proportion to the incident radiation. In the plateau region, the behaviour weakens before the path changes to the gas discharge region. The ionization chamber has no internal amplification like the counter tube. Therefore, electronics with high amplification are required.

Various ionization chambers are shown in Fig. 13.15. The principle is shown in Fig. 13.16. There are different designs and sizes for the individual beam types and dose ranges. The delivered current is proportional to the dose rate. With an ionization chamber, the *ion dose*, the *ion dose rate* and other dose sizes can be determined very well. This chamber is particularly suitable for *alpha spectroscopy* since these particles deposit all their energy in the chamber volume. For example, parallel plate ionization chambers for automatic exposure systems with specific dose rate ranges or ionization flat chambers for X-ray diagnostics can be used well.

Table 13.3 Comparison of different detector systems for the energy of 1 MeV

Key figure	Semiconductor (e.g. silicon)	Gas meter tube (proportional range)	Scintillator (e.g. NaJ)
Energy consumption w for: one particle-hole pair one ion pair a secondary electron at the photocathode	About 3.8 eV $T = 77 \text{ K}$	About 35 eV	About 1000 eV
Number of events Per MeV, $Z = 1 \text{ MeV}/w$	330,000	30,000	1000
Fano factor F	0.13	0.2	1
Statistical fluctuation \sqrt{ZF}	205	81	32
Relative fluctuation $\sqrt{F/Z}$	0.6%	2.5%	3%
Half-width	1.5 keV	6 keV	70 keV
Internal reinforcement Gas Reinforcement SEV Reinforcement	≥ 1	$>10^3$	$>10^6$
Signal strength: Number of elementary charges	$330 \cdot 10^3$	$30 \cdot 10^6$	10^9

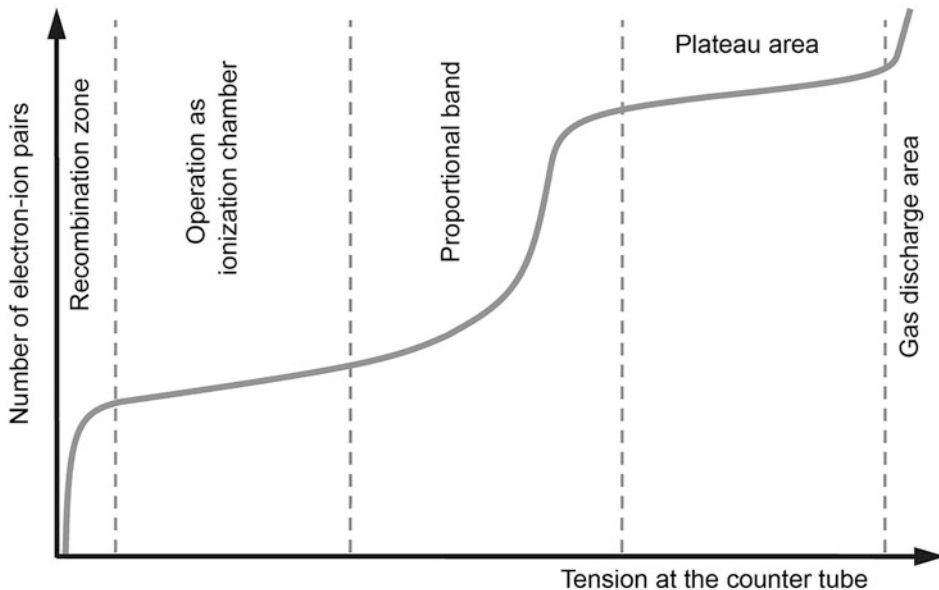
**Fig. 13.14** Characteristic curve of a counter tube

Fig. 13.15 Different ionization chambers (Plant photo: Nuklearmedizin KKH GM)

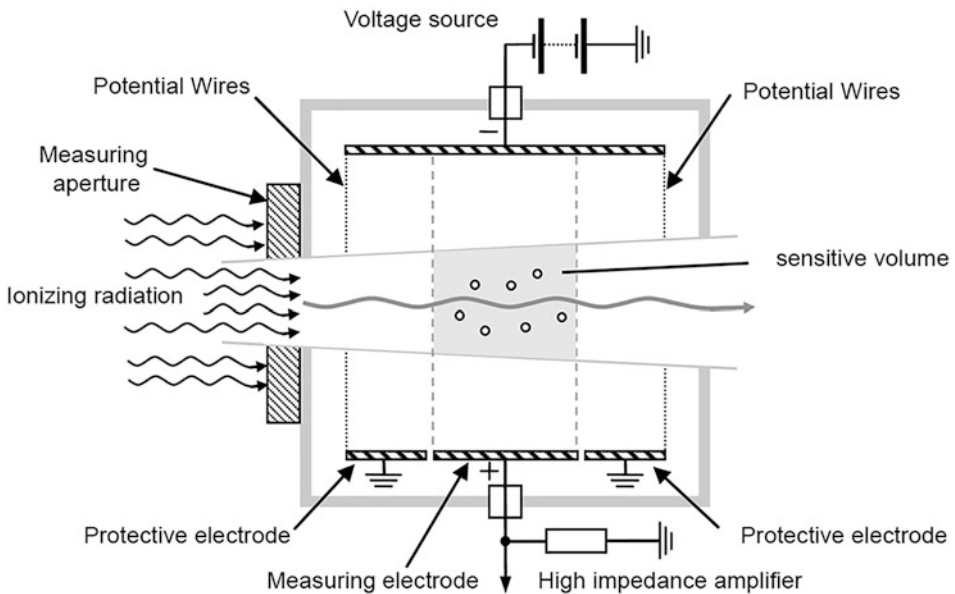
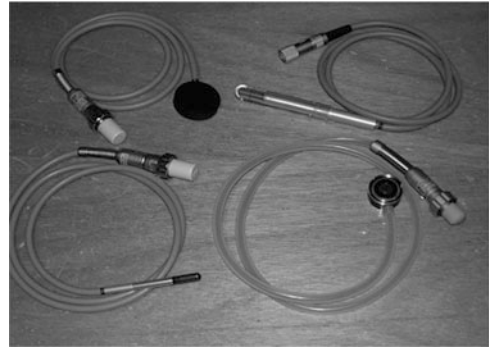


Fig. 13.16 Principle of the ionization chamber

The ionization chamber is a simple, robust detector and is particularly suitable for the following measuring tasks or applications:

- Reactor measurement technology,
- Radiation protection and dosimetry,
- Nuclear physics spectroscopy and
- Measurement and control technology.

The principle of an ionization chamber is shown in Fig. 13.16.

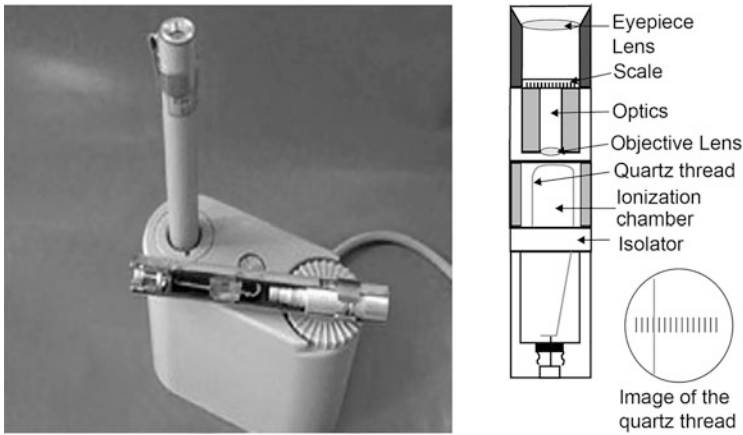


Fig. 13.17 Design and view of a rod dosimeter (Plant photo: Nuklearmedizin KKH GM)

As already explained, the absorbed dose is difficult to measure, so that the *equivalent dose* is usually measured with appropriately calibrated gas detectors. A very simple device is the *rod dosimeter*, which is charged with a voltage. Due to the primary ionisation deposited in the gas, the voltage decreases, which can be read off a quartz thread on a calibrated scale. The whole thing can then be calibrated in units of the equivalent dose in Sievert. The following Fig. 13.17 shows the principle structure and a device.

The *meter tube* is another important gas detector. The main component is a metal tube that forms the cathode. In the middle is a wire, the anode. Counter tubes, which are used for the detection of alpha radiation, have a radiolucent window (e.g. made of Mylar). This counter tube can also be used to detect alpha radiation because its range is small. If only beta or gamma radiation is to be determined, the window can be omitted because it causes design problems. Inside the tube, there is a noble gas with low pressure. Between the electrodes, there is a DC voltage of several hundred volts, which separates the ionized atoms and the electrons. The electrons are accelerated in the direction of the anode and generate further electrons by impact ionization like an avalanche. The resulting current is tapped at a load resistor. The resistor must be high-impedance so that the current is limited and the discharge is extinguished. Figure 13.18 shows the principle.

The counter tube can be operated in *various operating ranges* such as the starting, proportional or Geiger-Müller range (Fig. 13.14). To measure the dose rate and the activity of a radiation source, the *pulses* are counted. In the proportional range, the measured *current* is *proportional to the energy* of the incident radiation. The ionization chambers described or the proportional counter tubes for measuring the absorbed dose or the dose rate of a radiation source operate in this range. The *pulse height* is *proportional to the primary ionization*. The proportional counter tube can also be used for *energy determination* and *particle differentiation*. If the voltage in a counter tube is further increased, each particle triggers a saturation current. This is the actual operating range, the so-called

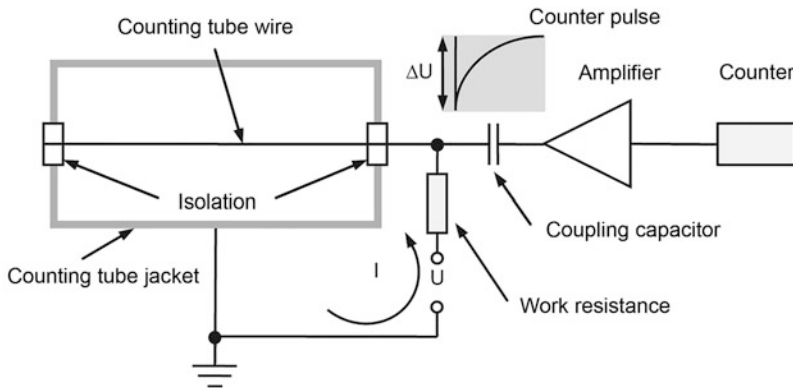


Fig. 13.18 Schematic diagram of a counter tube with detection electronics

Geiger-Müller range. Above a certain voltage, each incident particle triggers a cascade of secondary particles that “saturates” the counter tube; each particle generates the same current in the counter tube regardless of its energy. This area is the actual counting or plateau area and is used to *count* the particles. If the voltage is increased further, the counter tube is destroyed.

After triggering a gas discharge, the counter tube is not sensitive to further pulses for a short time, the *dead time*. The positively charged ions shield the anode field. The effect depends on the counter tube and can be influenced by additives. Counter tubes are offered by many manufacturers for *activity*, *dose rate* and *contamination measurement*. Counter tubes can be used to determine no or hardly any gamma radiation because the effectiveness is too low. This is a limitation.

In the large accelerator experiments, *proportional* and *drift chambers* based on the counter tube are used. They represent further developments, have very many wires and can determine the *particle location* in larger sensitive volumes.

13.5 Radiation Sensors According to the Excitation Principle

The radiation sensors based on the excitation principle or excitation detectors are an important detector class. The principle is that a physical system is excited by ionizing radiation and electromagnetic luminescence radiation is emitted during the transition to the ground state. These detectors can be roughly divided into *scintillation detectors*, *thermoluminescence detectors* and *Cherenkov detectors*.

To build a scintillation detector, you need an *optical detector* and a *scintillator*. The scintillator is a substance that is excited when ionizing radiation passes through it and emits the excitation energy again in the form of light. The *scintillation effect* is used to measure the *energy* and *intensity of ionizing radiation*. The scintillation detector is compared with the gas

meter and the semiconductor meter in Table 13.3. It lies between the two detectors. Average energy input of about 35 eV is required to generate one electron at the photomultiplier. The achievable resolutions are about 6 keV (related to 1 MeV primary energy).

The energy deposited in the scintillator can be detected with the photomultiplier, a photodiode or a charge-coupled device (CCD). Apart from the photomultiplier, these are mainly semiconductor components, which are described in the following section. For the construction of an *excitation detector*, a scintillator is configured, that is, in the simplest case, it is housed in a light-tight housing with reflector. Figure 13.19 shows the principle. Semiconductor detectors such as PIN diode or APD are discussed in the following section. A *photomultiplier* is a vacuum photo device based on the photoelectric effect, that is, the photons are converted into electrons and a charge multiplication is achieved by the dynodes indicated in the picture. The photomultiplier is a long known device and is one of the *most sensitive detectors*. Its disadvantages are the high noise rate, the magnetic field dependence and the size. In addition, a high voltage is required. The optical connection between the entrance window of the photomultiplier and the scintillator is particularly important: as few photons as possible should be lost here.

There are a large number of scintillators that can be classified into *inorganic ones* such as sodium iodide or *organic ones* such as anthracene or doped plastics. With the scintillation effect, neutral particles can also be detected indirectly, via subsequent reactions and the formation of subsequent charged particles. The scintillators can also be used in liquid or gaseous form. The time resolution is generally very good and can be in the nanosecond range (10^{-9} s). There is a large number of dosimeters for sensitive dose rate measurement that can be set up with scintillators. The response to gamma radiation is good, so the effect can be used for *gamma spectroscopy*.

The main advantages are:

- Good light output,
- The short decay time of the luminescence effect,
- Good coupling to the photomultiplier possible and
- Versatile in use.

The *thermoluminescence detectors* (TLD) are based on the thermoluminescence effect. This is the effect that the ionising radiation acting on a substance can be stored in the form of long-lived states of the crystal electrons. This energy can be released again by heating. This measuring method is only suitable for *relative dosimetry* and will not be discussed in detail here, but is also important for calibration and verification. Figure 13.20 shows a TLD detector.

The *Cherenkov effect* plays a role in particle detection, especially in high-energy physics. Cherenkov radiation is produced when *charged particles* in matter move at a *higher speed* than the *phase velocity* of electromagnetic waves in this medium. Short-term *polarization effects* produce a conical wavefront which can be determined as *Cherenkov light* with a photomultiplier or a semiconductor detector. There are other detectors, such as

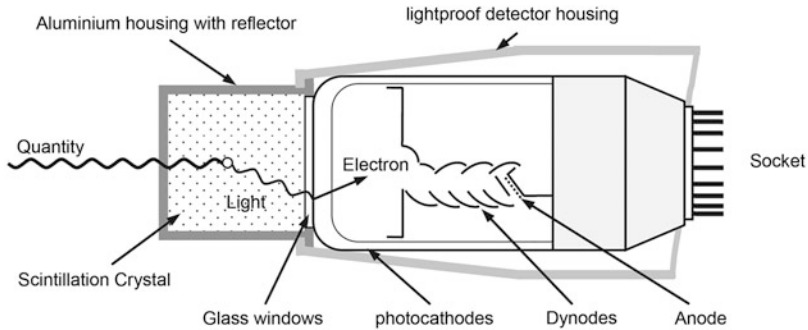


Fig. 13.19 Principle circuit diagram of a scintillation counter



Fig. 13.20 Thermoluminescence detector (Plant photo: Nuklearmedizin KKH GM)

the *transition radiation detector*, which is important in high-energy physics at high energies.

13.6 Semiconductor Sensors

Semiconductor sensors or semiconductor detectors exploit the electrical properties of semiconductors to detect ionizing radiation. Of great practical importance are the *standard pn junction*, the *PIN diode* and the *avalanche photodiode*. The semiconductor properties can be described within the band model. The radiation can produce free electron-hole pairs in the semiconductor, which can be read out in an electric field. An electron is lifted from the valence band into the conduction band: what remains in the valence band is a *defect electron*. The corresponding current signal is amplified and can continue to be available for a wide variety of tasks. The semiconductor detectors are often made of silicon, GaAs or germanium, but also, depending on the application, of other materials such as cadmium telluride. The bandgap determines how much energy is needed to generate an electron-hole pair.

The effect can take place in a standard semiconductor or at a barrier layer, as Fig. 13.21 also shows. Usually, monocrystalline silicon is used. However, amorphous silicon without

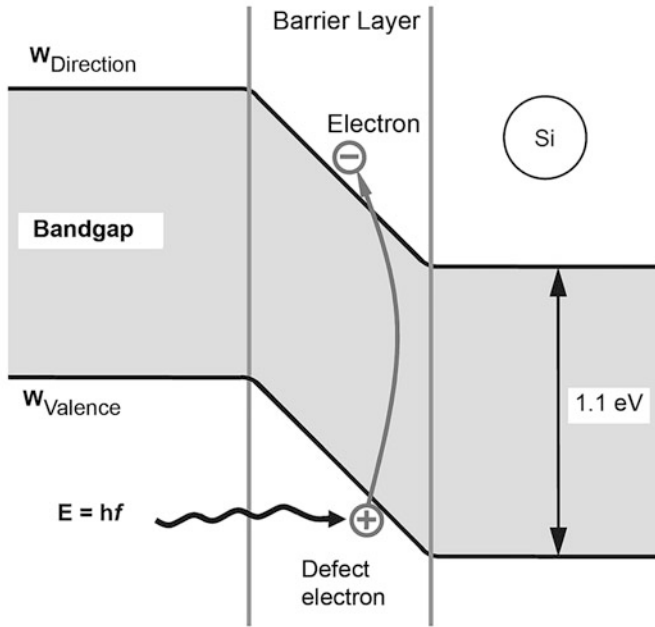


Fig. 13.21 Band model of a pn junction

a crystal structure can also be used. These detectors are deposited on a substrate by deposition from the gas phase. The maximum sensitivity is shifted towards the UV range, which is favourable for scintillation detectors. In addition, these detectors can be used in a *large radiation background*, since the sensitivity for radiation damage is lower.

In semiconductors, the conductivity can be influenced by doping within wide limits. This is specifically exploited. In a *solid-state ionization chamber*, which is modeled after the gas chamber, energies as low as 3 eV are sufficient to create an electron-hole pair in the semiconductor. Here, the photoelectric effect is exploited on a simple semiconductor without pn junction. The energy resolution in semiconductors far exceeds that of gas or excitation detectors, which is why they are often used for *spectroscopy*. With appropriate segmentation, they are *site-selective* and are used in research. Charged particles also generate charge carriers along their track, while photons interact at one point, that is, absorbed. Silicon photodiodes for radiation detection are usually PIN diodes operated in the reverse direction. The principle course of a current-voltage characteristic curve is shown in Fig. 13.22. A characteristic curve with or without external radiation is shown. Operation in the third or fourth quadrant is possible. In the vast majority of cases, operation in the third quadrant is desired. The pn-junction is biased in the reverse direction, the detector, therefore, has a very small junction capacity.

The PIN diode is one of the most important semiconductor detectors. For example, it consists of weakly n-type or intrinsically conductive silicon base material, which is usually

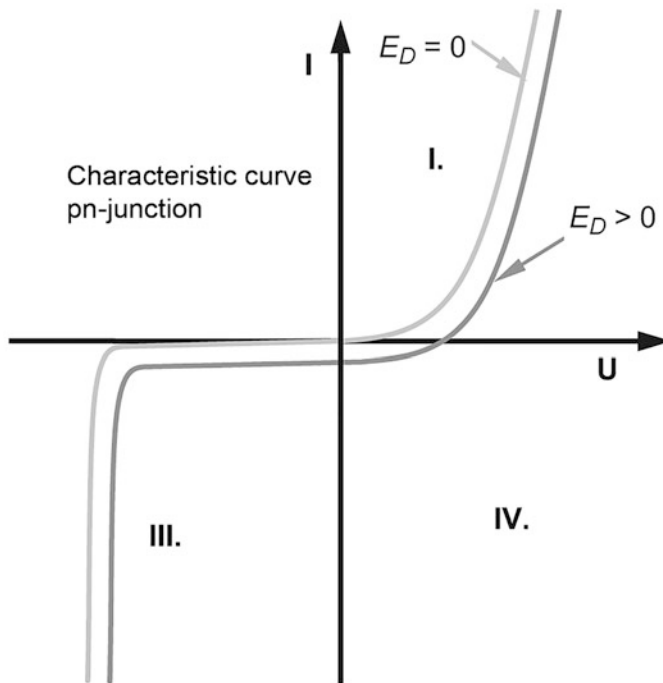


Fig. 13.22 Current-voltage characteristic of a photodiode

provided with a highly doped p-type implantation on one side and correspondingly n-type implantation on the other side. An intrinsically conductive layer is an undoped semiconductor layer in which only a few charge carriers are present, which can be generated by thermal processes, for example. For contacting, both areas are provided with metallizations (e.g. of aluminum). The basic structure can be seen in Fig. 13.23. The reverse voltage generates the space charge zone so that the diode is depleted over its entire depth. This minimizes the capacitance and results in a large radiation-sensitive volume. The *capacitance* should be *as small as possible*, which is recommended for noise and dynamic reasons.

The capacitance C of PIN diodes can be approximately calculated with the plate capacitor model. In addition to the permittivity number ϵ_r , it depends on the active area of the detector A and the depth of the space charge zone d . It applies:

$$C = \epsilon_0 \epsilon_r A / d. \quad (13.11)$$

The capacitance C should be very small and is in the *pF* range or above, depending on the detector. The dependence on the reverse voltage is in the form of a root function. In reality, one will have to find a compromise between the detector volume, the magnitude of the reverse voltage, the junction capacitance and other parameters.

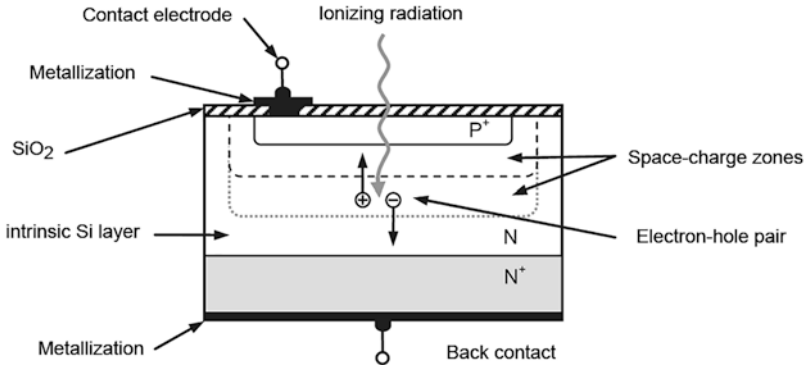


Fig. 13.23 Principle structure of a Si-PIN diode

An important criterion of a PIN diode is the *dark current*. It is generated by thermal and surface effects and increases exponentially with temperature. It should be relatively small. Since it depends on the size of the *sensitive volume*, it should not be larger than required for the response capability. PIN diodes also generate *noise*. They essentially exhibit shot noise and thermal noise. Like the dark current, this can be reduced by cooling. PIN diodes can be effectively cooled in combination with Peltier elements. Since the coupled electronics determine the overall performance of the measuring circuit, the first stage of the electronics, which is particularly important, is often also cooled. Today, the *optimization* of the entire measurement chain can be effectively supported by the use of electronic simulators such as PSPICE or SABER.

Figure 13.24 shows a PIN diode for the radiation sensor in the TO package. These diodes can be used to detect ionizing radiation in conjunction with scintillators. They have a wavelength-optimized higher sensitivity in the blue spectral range. They can also be manufactured with greater thicknesses up to about 0.5 mm. They can then be covered light-tight with a beryllium window and detect the radiation by direct absorption in the silicon. The diodes can be manufactured in *different designs*, up to arrays or other structures.

An important criterion is the *achievable resolution* of a semiconductor detector, such as a PIN diode. Due to the relatively small detector volumes, PIN diodes are only suitable for dose measurements to a very limited extent, but they have a *good energy resolution* and a *good temporal response*. The *energy resolution* of a spectrometer is defined as the *half-value width of a measured spectrum* and depends very much on the downstream electronics, cooling, pulse shape and temporal resolution. The values that can be achieved range from about 120 eV, which is the lower practical best value, to 1 keV for larger, simple uncooled systems. This applies to radiation energies below 10 keV. Very often the Fe55 preparation is used, which represents a radioactive gamma emitter at 5.9 keV.

Figure 13.25 shows such an energy distribution with a resolution of 150 eV, which is a very good value. The resolution is the full width of the spectrum at half height (*FWHM: full*

Fig. 13.24 PIN diode for the radiation sensor (Factory photo: silicone sensor)

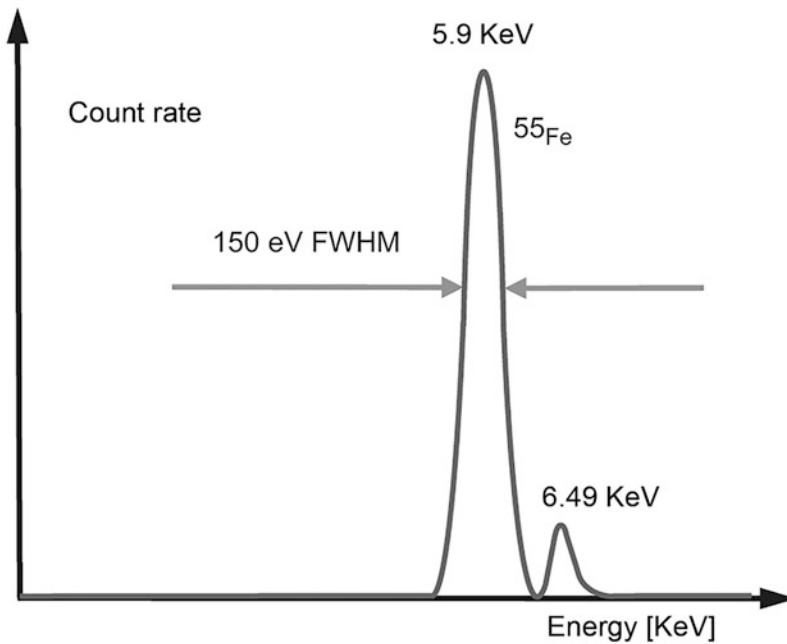


Fig. 13.25 Energy resolution of a semiconductor detector

width at half maximum). The spectrum also shows a second line at 6.49 keV, which can be additionally detected with the appropriate good resolution.

Furthermore, the *lithium enriched* silicon and germanium detectors must be mentioned here. As already described above, large detector volumes in PIN diodes can only be

produced with some disadvantages. By incorporating lithium ions into silicon, a large depletion zone with almost constant field strength can be achieved. Usually, for higher resolution requirements, these detectors are operated with liquid nitrogen. This is somewhat complicated in handling. Therefore the importance of these detectors has decreased in the last years.

Semiconductor detectors can be used with different sensitivities to detect the different types of radiation. For the detection of gamma radiation, especially at high energies, and thus reduced absorption probability, detectors as thick as possible with the disadvantages already described must be used, made of materials as heavy as possible. If light and thinner semiconductors are used, Compton scattering predominates, which worsens the energy resolution because part of the radiation is not absorbed.

Beta radiation or *electrons penetrate* deep into a detector. Along their path, they produce a uniform density of *electron-hole pairs* according to the Bethe-Bloch formula (Eq. 13.10). They can be detected relatively well. *Neutrons* can be detected by *recoil effects* at the nucleus. Semiconductor detectors are less suitable for detecting these particles. *Alpha particles* will only penetrate the detector in the *micrometer range*. A suitable construction must be used to ensure that they can be detected at all. High-energy other charged particles such as *pions* or *protons* also generate electron-hole pairs along their path and can be detected.

Avalanche photodiodes (Avalanche Photo Diode: APD) are also used to detect *high-energy radiation* down to single quanta (Fig. 13.26). They are also suitable for measurements with high time resolution. The formation mechanism of the electron-hole pairs is the same as for the PIN diode. The range of applications of APDs with respect to radiation detection is comparable to that of the PIN diode.

However, higher field strengths accelerate the electrons in the conduction band so much that numerous other electron processes are created. The *avalanche breakthrough* is associated with a *strong carrier multiplication* and is one of several breakthrough scenarios in semiconductor devices. A *breakthrough* of a pn junction is understood to be a *strong increase in reverse current* from a certain reverse voltage. The main distinction is made between *thermal breakdown*, *zener breakdown* and *avalanche breakdown*. The avalanche effect is reversible if the permissible total power loss of the component is not exceeded. In the case of an avalanche breakthrough, the current increases very strongly with the voltage compared to the Zener breakthrough. The *Zener breakdown* is used in electronics in the form of the Zener diode for *voltage stabilization*. When the temperature rises, the avalanche breakdown, unlike the Zener breakdown, only starts at a higher voltage. The *avalanche breakdown* has a *positive temperature coefficient*, whereas the *Zener breakdown* has a *negative temperature coefficient*. Both effects partially merge into each other.

Charge carriers which are lifted from the valence band into the conduction band by the influence of radiation can knock valence electrons out of the atoms otherwise not involved in the band model by impact ionization and additionally transport them into the conduction band. Due to the very high field strength, they can ionize again. The number of free charge carriers in the conduction band increases avalanche-like with an exponential characteristic.

Fig. 13.26 APD for radiation sensors (Factory photo: silicone sensor)



In the case of a PIN diode, this internal amplification effect is achieved as follows: The doping profile of an APD is distinguished from that of a PIN diode by the addition of a *further p-layer* (Fig. 13.27). This results in a strong *local increase in field strength* at the pn junction. This increase can initiate the process of impact ionization described above and the associated avalanche process.

Figure 13.28 shows the principle characteristic curve of an APD. At very low voltages, the APD acts like a PIN diode with a gain of about 1. If the voltage is increased further, but below the breakdown voltage, a reverse voltage and temperature-dependent *gain* occurs. It can be used to build highly sensitive and proportional photodetectors. The amplification here is usually between 10 and 1000. This range is called the *start-up* or *proportional range*. Similar to the PIN diode, the downstream electronics can be a charge amplifier or a trans-impedance amplifier. The characteristic curve is similar to that of a counter tube.

If APDs are operated *above the breakdown voltage* in the so-called *Geiger mode*, they achieve a gain of up to 10^6 . A single-photon triggers an avalanche. This is called *quenching behavior* or *quench mode*. The subsequent mostly digital electronics guarantee that the diode is not destroyed and then reset. This can also be easily achieved by the so-called analogue passive quenching by operating the APD via a series resistor. At the moment the current rises like an avalanche, a higher voltage drops at the series-connected series resistor so that the effective voltage at the APD becomes lower and the avalanche is extinguished.

Due to the statistical avalanche process, APDs have a *higher equivalent noise power* than PIN photodiodes. Nevertheless, low-noise photo-receivers can be built if they are carefully selected and well constructed. A decisive factor is cooling. APDs can be used to build detectors that can detect *individual photons*. The range of applications is very wide. Essentially, the radiation detection behavior described for PIN diodes also applies to APDs. *Ionizing radiation* can also be detected in APDs directly via *absorption at the crystal lattice* or indirectly via *scintillators* with silicon radiation receivers. For this purpose, very large absorption volumes and extremely small dark currents are required in some cases, while at the same time the space-charge zone is greatly extended.

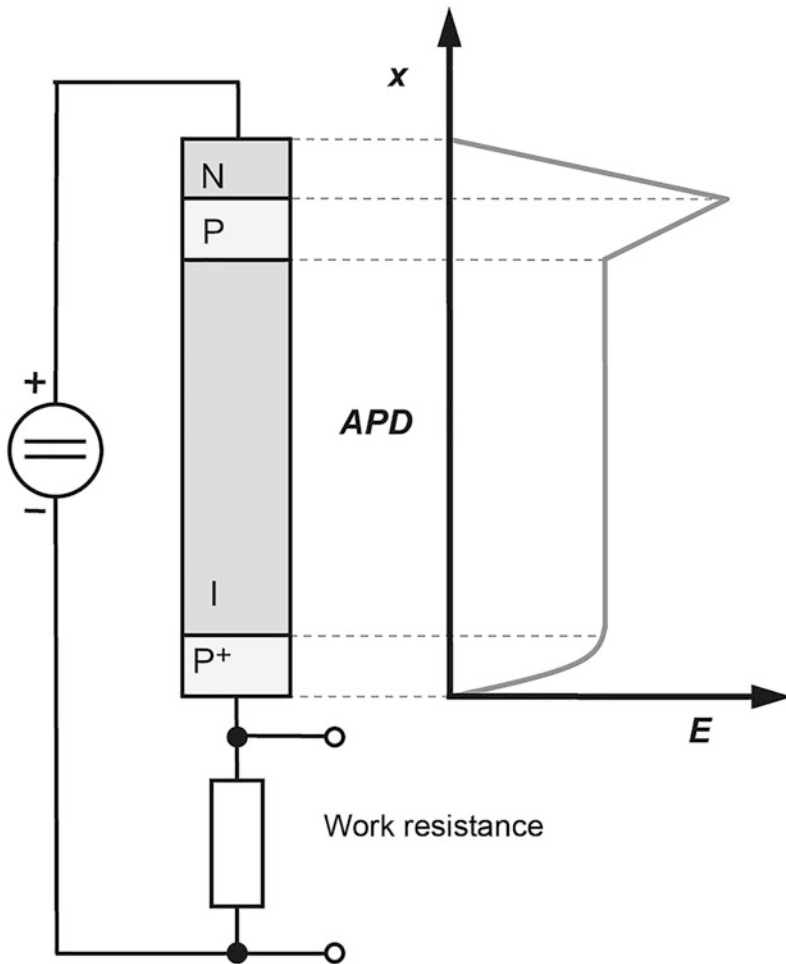


Fig. 13.27 Principle structure and field strength curve of an APD

Charge-coupled devices (CCD: Charge Coupled Device) can also be used to detect photons. The basis for these detectors is the *MOS capacitor element*. A silicon substrate can be separated from a gate electrode by an insulating layer. Photons can be generated and detected by light incidence via the photophysical effect described. The ionizing radiation must therefore be converted into light via a scintillator, as in the case of the excitation detectors. The *advantage of CCD structures* lies in their *pixel structure*. However, they are *very slow structures* because the individual pixels are not addressable and have to be read out via a protocol and an analog pipeline. This disadvantage is avoided by *CMOS sensors*, which have been used more and more recently. *Semiconductor radiation sensor technology* is developing rapidly, which is strongly promoted by the large accelerator experiments at

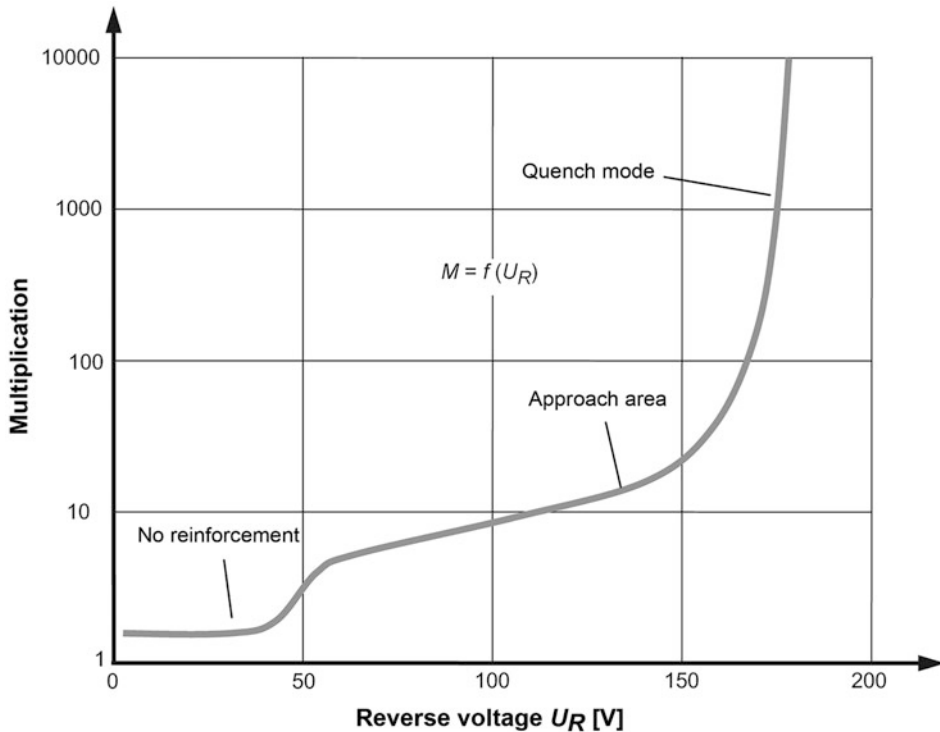


Fig. 13.28 Characteristic curve of an APD

the LHC (Large Hadron Collider) at the European particle physics centre CERN or by national centres such as DESY in Hamburg. Here, precise location information will be required, especially in the interaction area of the collision zones. This can be achieved by *strip* and *pixel detectors*, which are ultimately based on the basic structures described above, such as PIN, APD, CCD or CMOS. Recently, especially a fusion of the relatively old vacuum technologies such as the photomultiplier with semiconductor technology can be observed. This has led to sensors with completely new properties, which will not be discussed in detail here.

In the individual sections, several applications of ionizing radiation have already been mentioned in addition to the important dose measurement. Here some more examples will follow. The *intraoperative gamma probe* developed by the company Silicon Sensor allows a more precise removal of tumours such as breast carcinoma or malignant melanoma. The affected patient is injected with a radionuclide that has an energy of about 150 keV and a half-life of several hours. The probe is used to measure the radiation activity enriched in the tumour. The surgeon is thus able to use a computer and a display unit to see how close he is to the tumour or whether any tissue still needs to be removed. This allows the tumour to be removed more gently and accurately. In addition, it is often possible to find and remove the



Fig. 13.29 Intraoperative gamma finder (Factory photo: silicone sensor)

very small sentinel lymph node, as in breast cancer. Only when this has been successful can the patient be considered cured. Figure 13.29 shows an intraoperative gamma finder. The “inner workings” consist of a PIN diode and subsequent electronics consisting of amplifiers and filters. In the field of medicine, semiconductor detectors are particularly used in radiation therapy and diagnostics.

Figure 13.30 shows two monitor probes that can be used as *spectroscopy* and *counting sensors* or as *monitor sensors* for *integral intensity measurement*. They detect X-rays and the types of radiation from radioactive decay as well as synchrotron radiation in different energy ranges. The sensors are equipped with electronics, usually a transimpedance or charge amplifier.

These sensors have the following characteristics:

Monitor sensor for integral intensity measurement:

- Measurement in current mode,
- Dose rate measurement,
- Directly connectable to a voltmeter or A/D converter and
- With AD converter card configurable to the PC measuring system.

Spectroscopy and counting sensor:

- Measurement in spectral and counting mode with analog and digital output,
- Dose measurement and dose rate measurement at the digital output,
- Directly connectable to meter and pulse density meter,
- With counter card configurable to the PC measuring system and
- With MCA card configurable to a complete PC spectrometer

These and other devices or detectors can be used for the following important measurement tasks in research or industry: Nuclide detection and detection, α , β and γ nuclear radiation spectroscopy, X-ray diffraction, layer thickness measurements, coincidence

Fig. 13.30 Monitor sensors
(Factory photo: Silicon-Sensor)



measurements, contamination measurements, energy dispersive dose and dose rate measurements, dosimetry, absorption measurements, density measurements, Thickness measurements, mass per unit area determination, layer thickness measurements, dust density measurements, nuclear medicine, PET radiochemistry, gamma probes, borehole probes, level measurement technology, density scanners, material tomography, case and container inspection equipment, X-ray and synchrotron measurements. Thus, a large spectrum of applications is shown at the same time.

Bibliography

1. Bundesamt für Strahlenschutz (BfS): <http://www.bfs.de/bfs>
2. Grupen, Claus: Grundkurs Strahlenschutz: Praxiswissen für den Umgang mit radioaktiven Stoffen [Gebundene Ausgabe], 4., überarbeitete und ergänzte Auflage, Springer Verlag, Berlin 2008
3. Kleinknecht, Konrad: Detektoren für Teilchenstrahlung [Broschiert], 4., überarbeitete Auflage, Vieweg+Teubner Verlag, Wiesbaden 2005
4. Physikalisch-Technische Bundesanstalt (PTB): <http://www.ptb.de/>
5. Stolz, Werner: Radioaktivität. Grundlagen – Messung – Anwendungen [Taschenbuch], 5., durchges. Auflage, Vieweg+Teubner Verlag, Wiesbaden 2005
6. Tietze, Ulrich; Schenk, Christoph; Gamm, Eberhard: Halbleiter-Schaltungstechnik [Gebundene Ausgabe], 13., neu bearb. Auflage, Springer Verlag, Berlin 2009



Gert Schönfelder and Martin Liess

14.1 Radiation

Radiation is on the one hand *material radiation*, for example, alpha, beta, proton and neutron radiation, and on the other hand *electromagnetic radiation*. Matter radiation is generally detected indirectly, that is, it is converted into light, for example, using a scintillator, and then detected as electromagnetic radiation. There is also proof of the ionizing effect with the aid of a Geiger-Müller counter.

The range of electromagnetic radiation extends from its outermost edges with photon energies of over 10^{12} eV for cosmic gamma radiation to low-frequency radio waves with photon energies of less than 10^{-10} eV. This section deals with the range between X-rays and the far-infrared.

For this wide range of different electromagnetic radiation, there are different basic measuring principles depending on the wavelength, which, with the exception of thermal detectors, are in turn based on the *photoelectric properties* of different materials depending on the sub-wavelength range (Fig. 14.1).

Detectors for optical radiation are divided into *photon detectors* and *thermal detectors*. While thermal detectors measure the energy of electromagnetic radiation non-specifically, photon detectors use photons of a certain energy range to trigger an optical or photoelectric process in the material of the detector, which detects the light. In the case of *scintillators* as primary detectors, this process leads to the conversion of high-energy, poorly detectable

G. Schönfelder (✉)
Prignitz Mikrosystemtechnik, Wittenberge, Germany
M. Liess
Hochschule RheinMain, Rüsselsheim, Germany
e-mail: Martin.Liess@hs-rm.de

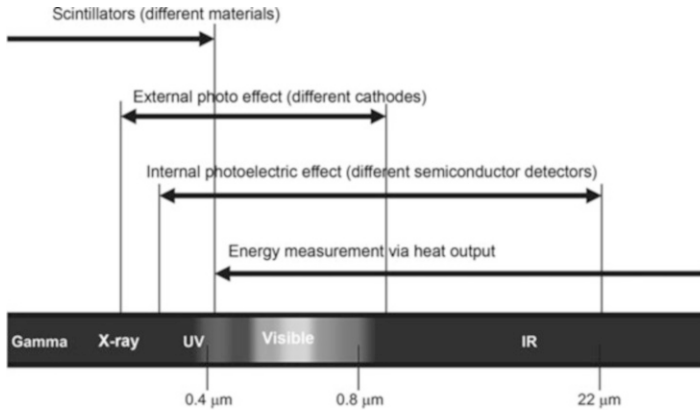


Fig. 14.1 Measuring principles for different areas of optical radiation

photons into light quanta in the visible or UV range that can be easily detected (e.g. with photodiodes or photomultipliers).

14.2 Scintillators

Scintillators convert radiation, especially electromagnetic X-rays and gamma rays, into longer-wave radiation, usually UV or visible light. This is then measured in a further step—usually with an avalanche photodiode or a photomultiplier.

The materials used are inorganic crystals, plastics and sometimes (liquefied) noble gases and nitrogen. Plastics generate UV radiation, which is quickly absorbed again in the plastic itself. They are therefore provided with fluorescent dyes in the visible spectral range so that the UV radiation is converted into visible light which can then be detected with a photodetector.

The decay time of light emission is approximately between 1 ns and 1 ps, depending on the material. Organic scintillators generally react faster, but age due to the radiation. Gases react very fast, do not age, but have (unliquefied) a low light output. Desired properties like

- Simple manageability,
- Longevity,
- Favourable emission spectrum for detection,
- The highest possible quantum yield and
- The short decay time of light emission

are often not all compatible. Therefore, the choice of scintillator depends strongly on the application.

Fig. 14.2 Photodiode rows on both sides of a circuit board, each covered with scintillator of different thickness (Factory photo: PerkinElmer Optoelectronics)



Typical applications are:

- X-ray sensitive lines in baggage screening systems (Fig. 14.2),
- area sensors for material testing and the medical sector (Fig. 14.3) and
- particle detectors in research.

X-rays of varying hardness hit the thin scintillator. The softer part is mostly converted into light, which is detected by the photodiodes on the upper side. The harder X-rays are converted into the light on the thicker, lower scintillator and detected. This creates a *two-channel X-ray detector line* that is used in baggage screening systems at airports.

14.3 Outer Photoelectric Effect

If the photon energy hf (h : Planck's quantum of action, f : frequency of light) is sufficient to overcome the exit work W of a conductor, electrons are released from the conductor, which have a kinetic energy E_{kin} of maximum $E_{\text{kin}} = hf - W$. This observation, called the *external photoelectric effect* (photoelectric effect), was explained in 1905 by Albert Einstein with the help of quantum theory. It forms the basis for numerous optical sensors.

14.3.1 Photomultiplier

Also in German, the English expression "Fotomultiplier" is generally used instead of "Fotomultiplikator". By means of the external photoelectric effect, light quanta in photocathode release electrons. These electrons are accelerated in an electric field and when they hit an electrode, a so-called *dynode*, they produce further secondary electrons, which in turn are accelerated and multiplied on further dynodes. Per dynode, a multiplication by a factor of 3–10 occurs, so that a photomultiplier with 10 dynodes generates about 10 million

Fig. 14.3 Digital X-ray imaging device, based on an array of polysilicium photodiodes and a scintillator made of gadolinium oxosulfite (technical applications) or cesium iodite (high sensitivity and resolution in the medical field) (Factory photo: PerkinElmer Optoelectronics)



electrons per detected photon. The multiplied photocurrent generated in this way can either be evaluated linearly as a measure of the received light or in the case of very low illumination, the received photons can be counted by means of the electrical pulses.

Photocathodes are used either *transmissively* (light incident on the front side, secondary electrons leave the cathode on the backside) in the so-called “*head-on photomultiplier*” or *reflectively* (light incident on the front side, secondary electrons leave the cathode also on the front side) in the so-called “*side-on photomultiplier*”. The quantum efficiency is somewhat higher for reflective photocathodes and is between 30% and 35%, in contrast to transmissive photocathodes with a yield of approximately 20–25%.

Depending on the desired spectral sensitivity, cathode materials of cesium iodide, cesium telluride, and alkali metal alloys are used. The windows are made of magnesium fluoride for UV radiation up to 115 nm, quartz or UV-transparent glass. Depending on the combination of entrance window and cathode, spectral sensitivities in different ranges between 115 nm and 900 nm are realized.

Since the path of electrons is deflected by a magnetic field, the sensitivity of a photomultiplier in the magnetic field is reversibly reduced.

14.3.2 Channel Photomultiplier

In the *channel photomultiplier*, a weakly conductive bent channel is used to multiply the electrons instead of individual dynodes (Fig. 14.4 above). The conductivity of the channel is achieved by reducing the lead glass tube with the help of hydrogen on the surface before the channel photomultiplier is evacuated and sealed in production. During operation, a small current causes a voltage drop along the channel. Electrons generated at the transmissive photocathode are accelerated along the channel and generate secondary electrons each time they collide with the channel walls. Channel photomultipliers are characterized by a *lower noise level* (dynode noise is eliminated) and a much lower dependence of the gain on an external magnetic field.



Fig. 14.4 Top: Channel photomultiplier tube. The curved channel, entrance window (right) and (transmissive) photocathode are visible. Below: Channel photomultiplier modules with protective housing in various sizes (Factory photo: PerkinElmer Optoelectronics)

14.3.3 Image Pickup Tubes

The basic principle of the image pick-up tube (Fig. 14.5) is the *conversion of the optical image* into an *electrical image* on a storage layer and the readout of this image with an electron beam. The image is generated either by means of the external photoelectric effect in the case of the iconoscope and the orthicon (both are insignificant today) or by means of the internal photoelectric effect (following section) in the case of the vidicon, endicon, vanicon or plumbicon. In the *Vidikon*, the semiconducting photosensitive layer is used to generate and store the electrical image, which is then locally made conductive and discharged with an electron beam. This discharge is tapped capacitively and corresponds to the image information at the respective position.

Vidikon image capture tubes are still manufactured today, although powerful CCD image sensors have replaced them in many areas. The radiation resistance, however, still predestines image pickup tubes for use under extreme conditions such as in nuclear facilities and in space. Image pickup tubes are suitable for the wavelength range from about 350–700 nm.



Fig. 14.5 Vidikon imaging tube with a storage layer made of SeAsTe (Saticon©) (Factory photo: PerkinElmer Optoelectronics)

14.4 Internal Photoelectric Effect

In the internal photoelectric effect, electrons are excited from the valence band into the conduction band in a semiconductor material. The photon energy must overcome the bandgap, that is, the energy difference between the bands. The excitation leaves a mobile positive charge in the valence band, while the energetically excited electron can move freely in the conduction band. These charge carrier pairs can either be separated in a *photodiode* in the electric field and lead to a photocurrent or directly as free charge carriers in the *photoresistor* to reduce the resistance of the semiconductor material between two electrodes.

The spectral sensitivity of a photodiode or a photoresistor results on the one hand from the bandgap of the semiconductor material used, since this determines the minimum energy of the photons. On the other hand, the electrons have to be excited at the correct position in the semiconductor component (in the case of a photodiode in the depletion layer) in order to contribute to the current flow. Different semiconductor materials are therefore used for different wavelength ranges to be detected (Table 14.1). To detect photons with lower energies, semiconductor detectors must be cooled depending on the measurement requirements. MCT detectors are always cooled (often with liquid nitrogen).

14.4.1 Photoconductor

Photoconductors have numerous advantages over photodiodes: they can be operated at up to 400 V (depending on the type) and can therefore be used directly on the mains circuit, for

Table 14.1 Different semiconductor materials and their approximate areas of application in wavelength and photon energy

Material	Detectable wavelength $\lambda/\mu\text{m}$		Detectable photon energy W/eV	
	Min.	Max.	Min.	Max.
Various alloys of mercury-cadmium-teluride (MCT)	5.5	22	0.056	0.23
Indium antimonide (InSb)	1.5	5.5	0.23	0.83
Lead sulphide (PbS)	1.2	4	0.31	1.0
Indium Arsenide (InAs)	1.2	3	0.41	1.0
Indium-Gallium-Arsenide (InGaAs)	0.8	1.7	0.73	1.6
Germanium (Ge)	0.88	1.4	0.9	1.4
Silicon (Si)—Various detector geometries	0.3	1.1	1.1	4.1
Silicon carbide (SiC)	0.21	0.4	3.1	5.9

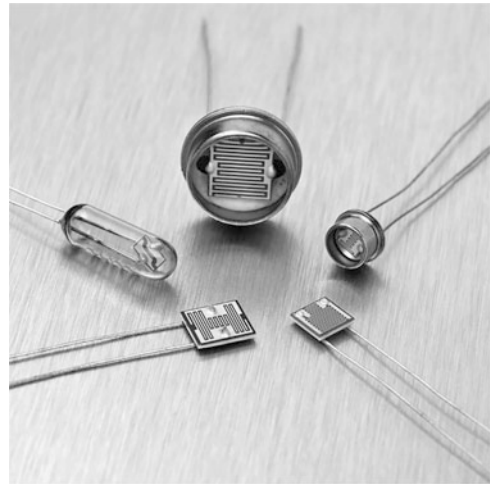
example, to control a relay. The semiconductor material is, depending on the required spectral sensitivity, differently doped cadmium sulfide. The spectral sensitivity of photoconductors has a maximum at 500–600 nm and is similar to the sensitivity of the human eye. They are also suitable as optically controllable resistors, for example, in analog optocouplers.

The reaction of photoconductors to illumination is determined by the generation and recombination of charge carriers. Since the *recombination rate* is described by the law of mass action and is thus proportional to the number of positive charge carriers as well as proportional to the number of negative ones, the average lifetime of the charge carriers decreases with their concentration in the semiconductor. Conversely, the average lifetime of the charge carriers increases when their number in the material decreases. Therefore, photoconductors are very *non-linear* and react much *more sensitively* (often desired) and more *slowly to low illuminance levels*. For example, the resistance after illumination is up to three times higher after 5 s of subsequent absolute darkness than after 1 s.

Despite the disadvantage of slow response at low illuminance levels, there are numerous applications for photoconductors. However, due to the environmental impact of the semiconductor material cadmium sulfide and the introduction of corresponding voluntary commitments by industry and legislation¹ in many countries worldwide, they have lost ground compared to photodiodes. Figure 14.6 shows some designs for photoconductors.

¹For example, in Europe: “Restriction of the use of certain hazardous substances in electrical and electronic equipment” (RoHS); in China: “Measures for Administration of the Pollution Control of Electronic Information Products” so-called “China-RoHS”; in Japan the industry’s self-obligation, the so-called JGPSSI (Japan Green Procurement Survey Standardization Initiative). Further initiatives exist worldwide, for example, in California and other US states.

Fig. 14.6 Photoconductors based on the internal photoelectric effect (photograph by PerkinElmer Optoelectronics)



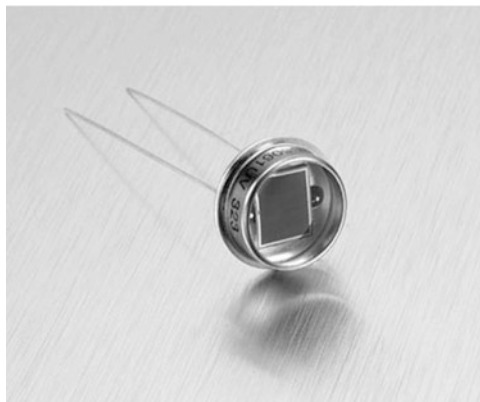
14.4.2 Photodiodes

Photodiodes contain a pn junction, which is arranged parallel to the surface of the diode. The heavily doped uppermost plane is used to divert the charge carriers to the metallized edge of the diode and is itself not very light sensitive. Below this is the depletion layer, from which all charge carriers are removed by an electric field when the diode is operated without voltage or in the reverse direction. If charge carriers are generated in the depletion zone by photons, they flow off towards the top and bottom of the diode and contribute to the photocurrent. Figure 14.7 shows a photodiode.

Photodiodes can be operated either in photovoltaic mode (*short circuit*) or with *negative bias voltage* (photoconductive mode). The advantage of the photovoltaic mode is the fact that the only energy source for a possible current flow comes from the light to be measured and therefore no current can flow in the absence of light. Thus, the zero point is offset-free and can be measured very accurately. This circuit is suitable for measuring very low light intensities with low frequency. The circuit design uses an operational amplifier with low input offset current (bias current), such as the OPA128 from BurrBrown (Texas Instruments).

If the photodiode is operated with negative bias voltage, a temperature-dependent reverse current always flows. As a result, the zero point is no longer exactly defined and low light intensities of low frequency can no longer be measured so accurately. The advantage of this circuit, however, is that the depletion zone in the diode is increased due to the reverse voltage and thus the diode capacitance is reduced. Signals of high frequency can thus be measured better.

Fig. 14.7 Silicon photodiode
(Factory photo: PerkinElmer
Optoelectronics)



PIN diodes

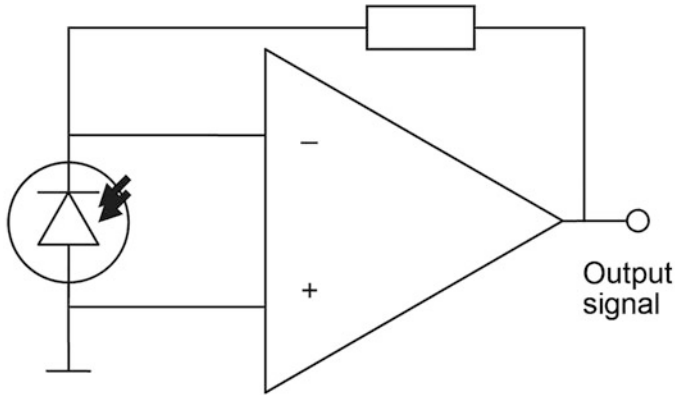
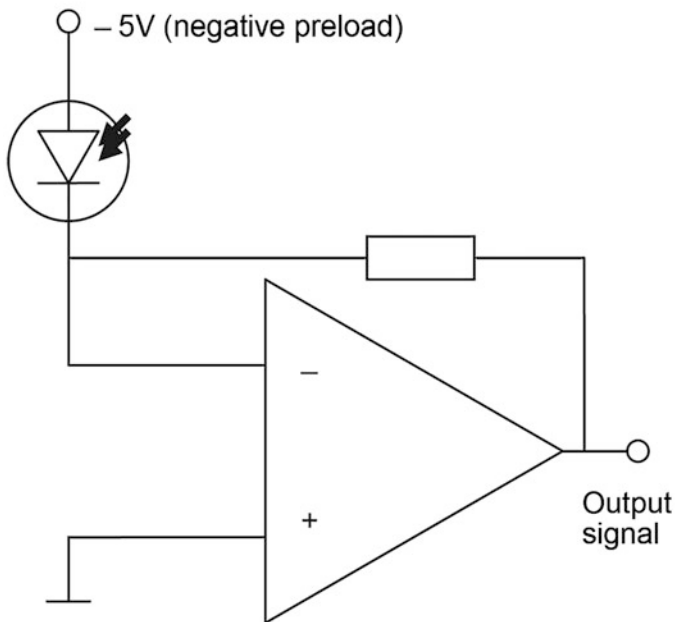
PIN photodiodes still have an undoped intrinsic layer between the p- and n-doped layer. This *increases the depletion zone* by the thickness of the intrinsic layer. The advantages are the lower capacitance and thus the higher reaction speed of the diode as well as the greater thickness of the photosensitive layer. The latter increases the sensitivity of the diode at long wavelengths, whose penetration depth is correspondingly greater.

Avalanche photodiodes

In avalanche photodiodes (Fig. 14.9, *APDs*), a very high reverse voltage in the order of 150–400 V for silicon APDs and 40–90 V for InGaAs APDs is applied to the diode. As a result, electrons generated by light quanta are accelerated in the component in such a way that they generate free electrons again by means of shock ionization in the semiconductor material itself, thus multiplying the photocurrent.

The voltage at which a charge avalanche no longer comes to a standstill on its own, but leads to a continuous breakdown voltage, is called break down voltage. Above the breakdown voltage, suitable avalanche photodiodes can be operated in the so-called “*Geiger mode*”, below it they are operated in the so-called “*linear mode*”. The respective operating voltage refers to the breakdown voltage. Since the latter is temperature-dependent, the operating voltage of the avalanche photodiode must be regulated according to its temperature.

Avalanche photodiodes can be operated in the so-called *Geiger mode* (named after the functional principle of the Geiger-Müller counter tube) and *individual photons can be counted*. The reverse voltage at the diode is set so high that a single photon can ignite a permanent avalanche breakdown in the diode. This avalanche breakdown must be extinguished immediately by switching off the voltage so that the component does not overheat. With each avalanche breakthrough, a single photon is counted. The Geiger mode is suitable for detecting extremely *small amounts of light at very low frequencies*.

Photovoltaic mode:**Operation with negative preload:****Fig. 14.8** Photodiode operating modes

In the so-called *linear mode*, the avalanche effect only amplifies the photocurrent of the diode without a continuous breakdown. The amplification of the photocurrent by the linear avalanche effect is in the order of magnitude range between 1 and 500, depending on the



Fig. 14.9 InGaAs APD with attached glass fiber (left), silicon APD in plastic housing (right) (Factory photos: PerkinElmer Optoelectronics)

photodiode (especially the semiconductor material used), the temperature of the component and the applied reverse voltage. For silicon, two to *three-digit gains* are achieved, while for InGaAs only one- to two-digit gains are possible (Fig. 14.9).

In linear mode, small amounts of low-frequency light are difficult to detect due to the high offset current, but in linear mode, an APD responds extremely well to rapid changes in the amount of light received. Fast, small signals such as the laser flash scattered by the target during distance measurement with *LIDAR* (Light Detection and Ranging) are reliably detected by an APD in linear mode.

14.4.3 Phototransistor, Photothyristor and Photo-FET

The *phototransistor* uses the same physical effect as the diode—charge carriers are generated in a semiconductor layer. The resulting charge carriers have the function of the base current. Due to the amplifying effect of the transistor structure, the *higher output current can be* generated than with a diode. The disadvantage is the *low cut-off frequency* in the kHz range due to the high junction capacitances.

Phototransistors are mainly used in optocouplers for switching digital signals. The amplification compensates for the low efficiency of the light-emitting diodes so that a current transmission of 1:1 is possible.

In the *photothyristor*, the charge introduced is used to ignite the component. It is mainly used in power electronics. The advantage here is the galvanic separation of control and load circuit.

In the *photo-FET*, the gate electrode is designed as a photosensitive layer. Here too, the amplification of the FET is used to switch larger currents. The most common application is *optical relays*. However, photo-FETs are also used in imaging systems. Here the high amplification of the component is used to realize cameras with a low dark threshold.

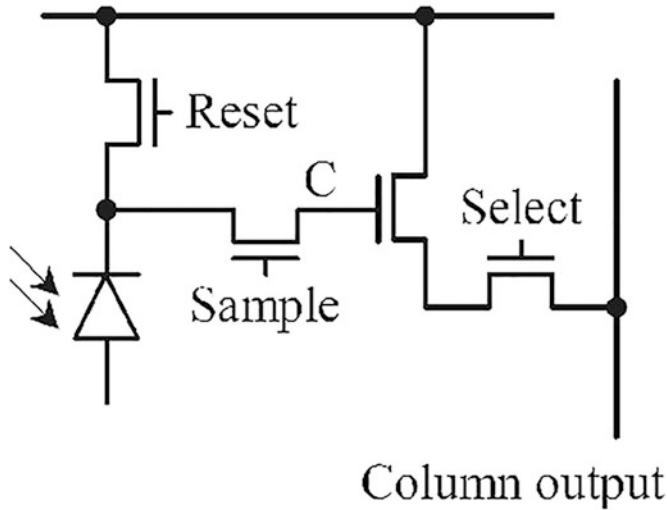


Fig. 14.10 Pixel circuit of a CMOS sensor (Factory photo: Fillfactory IBIS5)

14.4.4 CMOS Image Sensors

In contrast to the CCD sensors described in Sect. 14.5, CMOS image sensors are based on photodiodes. These are arranged in a row or matrix and can be addressed directly and in any order by means of row and column addressing. Figure 14.10 shows a convenient pixel cell.

The barrier layer of the photodiode is charged by the incident radiation (element operation). After exposure time X , this charge is transferred via the “sample transistor” to the gate of the third transistor. With the “select” signal (row line) the storage transistor is switched to the column line, which ends at a multiplexer. After the charge has been sampled, the residual charge on the photodiode is discharged so that a new exposure cycle can begin. CMOS sensors are inferior or equivalent to CCD sensors in terms of their optical properties, if the highly dynamic sensors are not considered. The advantage is the integration of other logical components on one chip, such as AD converters, timing generators and random access to the pixels.

14.4.5 High Dynamic CMOS Image Sensors

The human eye can process a *dynamic range* of up to 140 dB in an image. This means that the brightest point in the image is 10^7 times brighter than the darkest point. A normal CMOS or CCD sensor can cover about 60–70 dB due to the electrical circuitry used on and around the chip. This can be increased up to 90 dB for selected scientific sensors.

The basic brightness in an image is adjusted by selecting the exposure time. However, in many applications, there will always be the case that parts of the image are under- or overexposed. This is especially the case when capturing images in natural surroundings. Typical examples are driver assistance systems of all kinds or the cutting edge measurement of combine harvesters. Here, the image dynamics are particularly stressed when hard shadows, backlighting or rapid changes in brightness occur (e.g. when driving into or out of a tunnel).

Since it is possible to integrate linear and digital circuits with the CMOS sensor, several solutions for the construction of image sensors with increased dynamics have been developed, especially with this sensor. The most important approaches are:

- *Logarithmic characteristic curve*

The voltage coming from the sensitive element is further processed by an amplifier with logarithmic characteristics. This means that the gain decreases with increasing input voltage.

- *Approximated non-linear characteristic*

The steepness of the charge is influenced by switched auxiliary currents depending on the charge state of the sensor element. This allows the transfer characteristic to be adapted to the target process. A representative of this type of sensor was the LM9618 from National Semiconductor, whose characteristic curve is shown in Fig. 14.11. With this principle, 62 dB are achieved in linear mode, 110 dB can be measured with a non-linear characteristic.

- *Multiple exposures*

With multiple exposure, multiple images of the same scene are taken in quick succession. Each image has a different exposure time. These images are then “superimposed” to form an image with a greater dynamic range. The technical limit here is mainly the readout speed and the movement dynamics of the object.

- *Stepped exposure*

A simple pixel is used to detect the radiation, which causes the voltage to increase. After a short exposure time, a query is made as to whether the photovoltage has exceeded half the working range. If this is the case, the integration is terminated at this pixel—the time stage is stored at the pixel. All other pixels continue to integrate. After the 2, 4, 8 and 16 times time, this process is repeated. In the end, all pixels have a voltage level which must be divided by the integration time, because the brightest pixel was finished first (i.e. by 1) and the darkest pixel took the longest time (by 16). Figure 14.12 shows a signal curve for this.

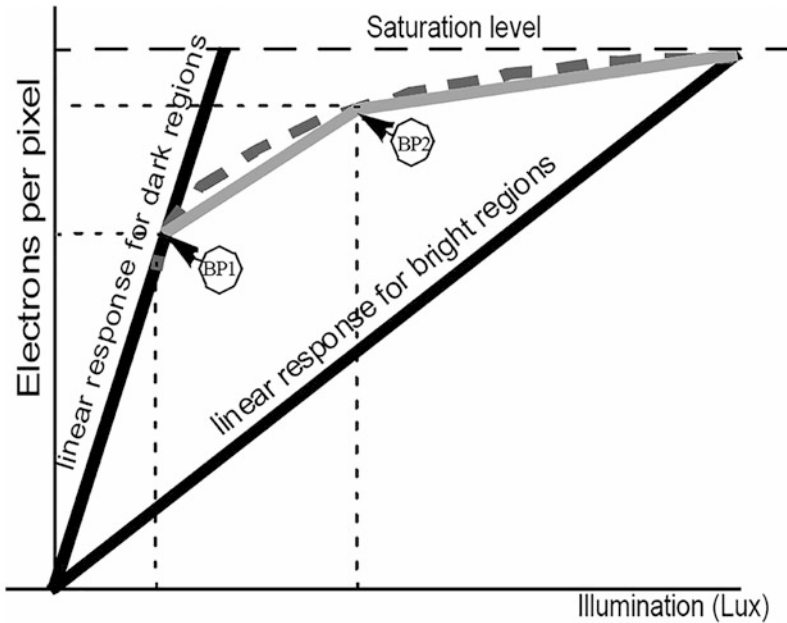


Fig. 14.11 Exposure curve of the LM9618 (Source: NSC)

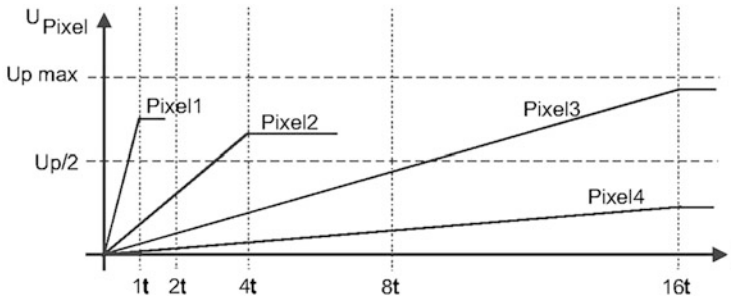


Fig. 14.12 Voltage curve with stepped exposure (Source: Silicon Vision—LARS)

14.5 CCD Sensors

“CCD” (*Charge Coupled Device*) describes the type of *controlled charge transport with the help of migrating electrical fields* in a semiconductor layer. However, “CCD” has become a synonym for image sensors in which the *charge distribution* on a silicon chip generated by an optical image is read out with the help of this technique.

In the CCD image sensor, an electric field acts on a silicon surface by means of transparent electrodes on the top of the sensor. Photon generated charge carrier pairs in the silicon are separated by the field and remain trapped in the potential gradient under the respective electrode until they are read out by the CCD technology. For detection, the potentials at the electrodes are switched over in such a way that the charges continue to move step by step until they can be read out serially at a suitable point.

14.5.1 Line Sensors

Figure 14.13, which shows the beginning of the charge transfer, is used for clarification. In drawing file A) the charge is collected. This state also corresponds to the exposure time. By adding P2, the charge well is enlarged and the charge carriers are pulled under the cover (B). When P1 is switched off, the entire charge is concentrated above the electrode at P2 (C). This process is continued until the charge has reached the edge of the arrangement and is evaluated there. Since moving the charges past the subsequent optically active surfaces leads to a change in the charge, the first step is to shift the charge to a parallel path which is completely darkened.

The block diagram of a line sensor in Fig. 14.14 clearly shows the parallel structure for reading the image data. To start the image acquisition all pixels are discharged via the shutter gate. Then the exposure time starts, which ends with the transfer of the charge from the pixels to the analog shift register. From here the charges are shifted to the left to the output where they are amplified.

Line sensors are available from 64 pixels to over 10,000 pixels. They are widely used in industrial applications, especially where the second dimension is created by the movement of the object itself (Fig. 14.15). These include transport processes of all kinds, such as thread monitoring on weaving machines to error analysis of boards during furniture production. An example of static processes is the bar code scanner.

14.5.2 CCD Matrix Sensors

In order to capture static, two-dimensional images, the sensor must also become two-dimensional. This is achieved by a matrix-like arrangement of the sensitive elements, coupled with a readout logic according to the CCD principle. An example of this is shown in Fig. 14.16. The readout process is first carried out in the vertical register, which is shifted one line down, and in the horizontal register to the left to the output. The disadvantage of the CCD matrix sensor is that only *complete images* can be read out. The query of an image section is not possible or only with considerable effort in the control.

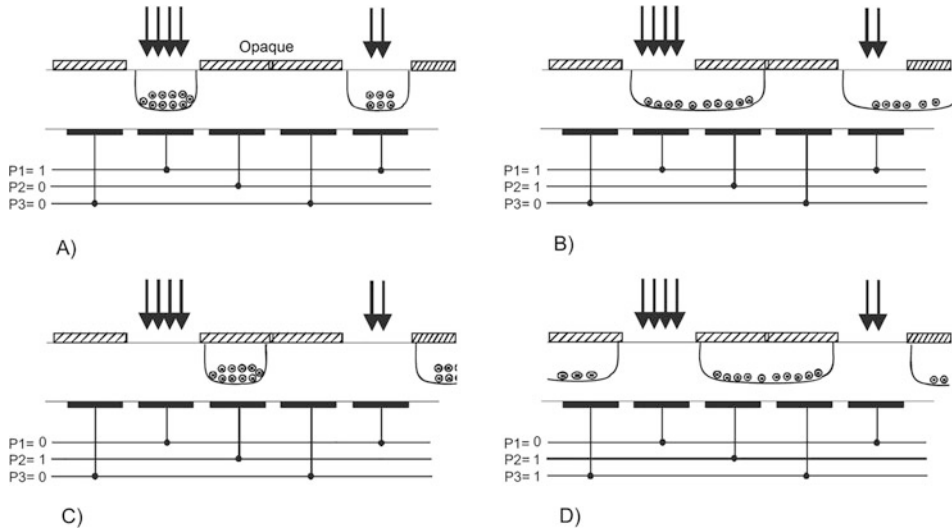


Fig. 14.13 Phases a charge shift in the CCD sensor

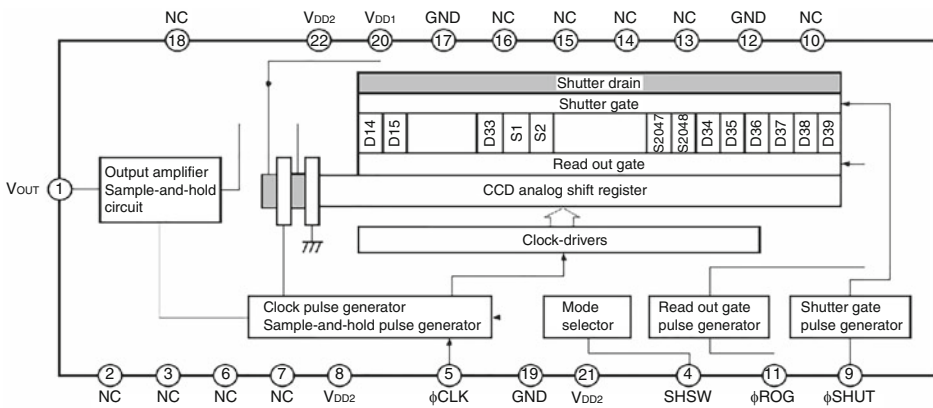


Fig. 14.14 Block diagram of a line sensor (Source: Sony ILX751)

14.6 Quantum Well Infrared Photodetector QWIP

In QWIP, thin layers of a semiconductor with a smaller bandgap (order of magnitude 50) are embedded between layers of a semiconductor with a larger bandgap. Semiconductor alloys are used that allow the lattice constants to be adjusted, such as GaAs/AlGaAs (gallium arsenide: low bandgap, aluminum gallium arsenide: higher bandgap) or GaInAs/InP (gallium indium arsenide: low bandgap, indium phosphide: higher bandgap).



Fig. 14.15 CCD line sensor for use in spectroscopy (Factory photograph: PerkinElmer Optoelectronics)

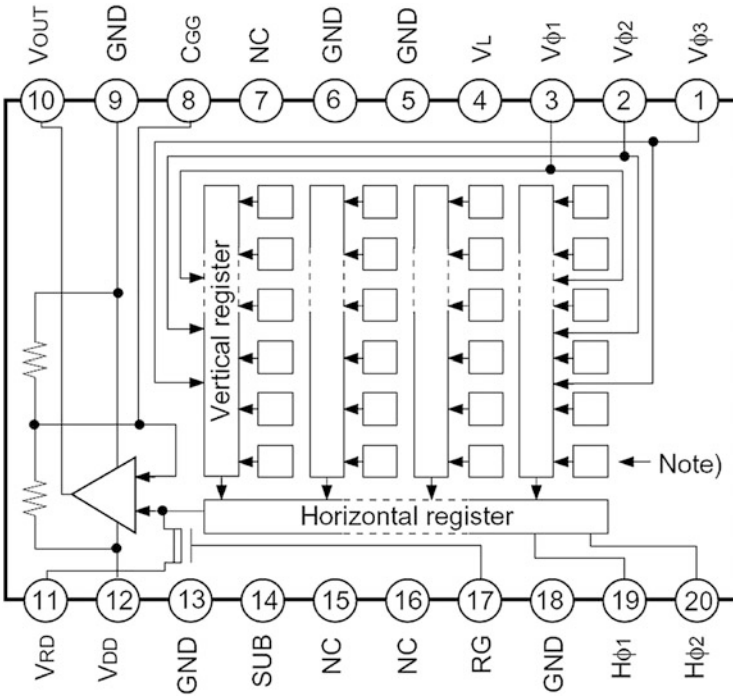


Fig. 14.16 Block diagram of a CCD matrix sensor (Source: Sony ICX085AL)

In the material with the smaller bandgap, *one-dimensional quantum wells* are formed in the conduction band for electrons as well as one-dimensional quantum wells for holes in the valence band. The layer thickness of the material with the low bandgap corresponds to the width of the quantum well and thus determines the position of the quantised energy levels. In the following, only the quantum well for electrons formed in the conduction band

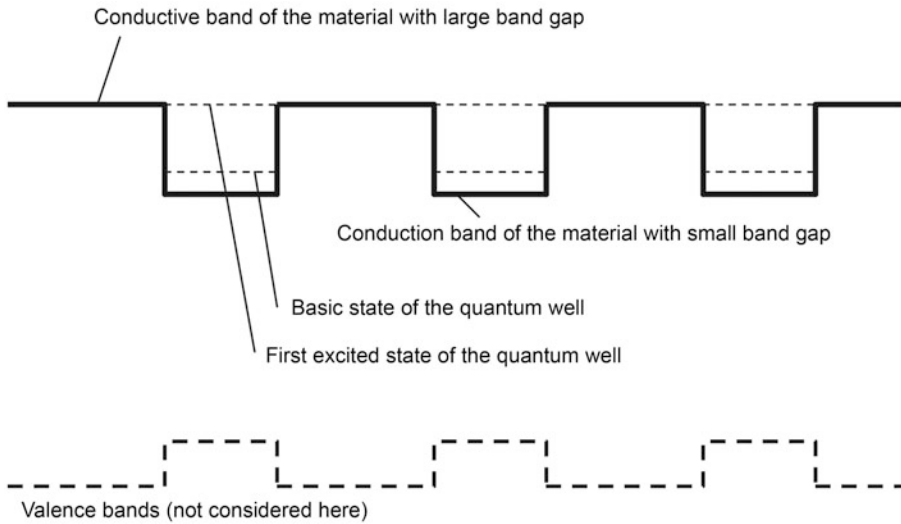


Fig. 14.17 Band diagram in the QWIP

is considered. Its width is usually designed so that its first excited state is at the energy level of the conduction band of the material with the larger bandgap (Fig. 14.17).

The semiconductor material is doped so that the ground states of the quantum wells are filled with electrons. By means of photons with a sufficiently high energy, electrons can be lifted from the ground state to the first excited state of the quantum well and then move freely in the semiconductor. This photoconductivity is used for the detection of *infrared light quanta*. QWIP detector arrays are generally used as arrays for *infrared imaging (thermography)* with high resolution. Accordingly, photoconductivity of holes in the valence band can also be achieved. However, due to the lower mobility of the holes, this effect is mostly not used.

14.7 Thermal Optical Detectors

In contrast to the quantum detectors described above, thermal detectors do not react to individual photons, but the energy transported by the optical radiation. The IR radiation is converted into heat on an absorber surface in the sensor. This heat is in turn detected and thus enables indirect radiation detection. This makes thermal detectors *much slower* and *less sensitive on the one hand*, but independent of the quantum energy or the wavelength of the photons on the other. Their signal is proportional to the total energy of the measured radiation and independent of its wavelength. Especially in the mid and far-infrared range, where quantum detectors have to be cooled, thermal detectors are often used, because they do not have to be cooled in principle.

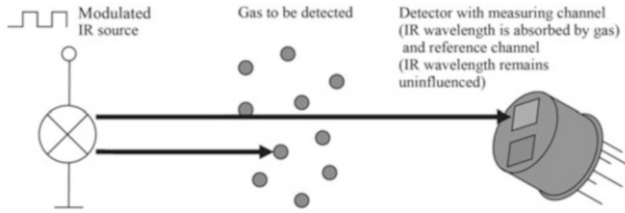


Fig. 14.18 Functional principle of photoabsorptive gas measurement with thermal detectors (pyroelectric detectors or thermopiles)

Applications

Applications of thermal detector (Fig. 14.19) are *contactless temperature measurement* and *IR imaging* in the lower price segment, *gas detection*, *spectrometry* and wherever the intrinsic advantages such as *large optical bandwidth* and the possibility to dispense with cooling compensate for the disadvantages such as slow response and low detectivity. Advantages of thermal IR detectors are:

- *Cooling is not necessary in principle* (in most cases uncooled detectors are used).
 - Sensitivity is *independent of the wavelength*,
 - Passive detection principles for thermopile sensors and pyroelectric detectors (output signal is zero at zero input signal). This allows *sensitive detection of small signals*.
- Disadvantages are:

- slow reaction and
- low detective power.

Table 14.2 provides an overview of typical applications.

14.7.1 Thermal Piles

A modern thermopile sensor (Fig. 14.20) is based on a thin, poorly heat-conducting membrane with a black absorber layer that is heated by the incident IR radiation. The membrane is held in thermal contact with the environment by a frame made of good heat conducting silicon. Its temperature is not significantly affected by possible IR absorption.

Between the center (membrane) and the periphery (silicon), numerous (order of magnitude 100) thermocouples are arranged in such a way that the *warm contacts* are in the center, while the *cold contacts* measure the reference temperature of the periphery or the environment (Fig. 14.20). The power transported with the absorbed IR radiation is converted into heat in the absorber on the membrane, which is converted into a *temperature difference* between the hot and cold contacts of the thermopile due to the thermal resistance of the membrane. The *thermopile* in turn converts this temperature difference into a *voltage* (Sect. 2.10).

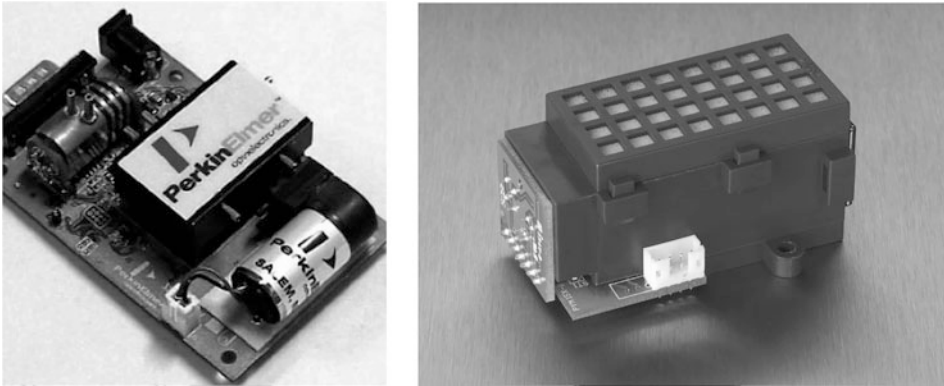


Fig. 14.19 Optical gas sensors based on thermal IR detectors. Left: for medical applications, right: low-cost CO₂ sensor (Factory photos: PerkinElmer Optoelectronics)

Table 14.2 Application areas of thermal IR detectors

Applications	Sensor	Comment
Temperature measurement	Thermopiles, pyroelectric detectors (with interrupter)	Thermopiles: Detailed description in Sect. 14.7.1 Pyroelectric detectors: Since these detectors cannot measure constant signals, an optical interrupter wheel “chopper” is used for temperature measurement.
Spectrometry	Thermopiles	Because of their wavelength-independent sensitivity in the (near.) middle and far IR.
Thermal imaging	Coarse: thermopile, fine: bolometer	Advantageous because of the size of each sensor element, thermopile arrays larger than 10 × 10 pixels are not practical and arrays larger than 4 × 4 pixels are not commercially available. Microbolometer arrays are typically used for IR imaging.
Person detection	Pyroelectric detectors	Applications: Building surveillance, automatic light switches, automatic water taps
Gas measurement	Thermopiles, pyroelectric detectors	Description, Figs. 14.18 and 14.19

Modern thermopile sensors (Fig. 14.21) are manufactured in a *MEMS process*. In this process, a large-area hole (1 mm²) is etched on the back of silicon using anisotropic (alkaline) etching, or “Reactive Ion Etching (RIE)”. On the front side, there is a surface of silicon oxide/silicon nitride and the thermopile. The thin oxide/nitride membrane remains after etching. On it are the warm contacts of the thermopile, while the cold contacts are located above the remaining silicon in the periphery next to the hole. Typical materials for the thermocouples are bismuth/antimony or polysilicium/aluminum.

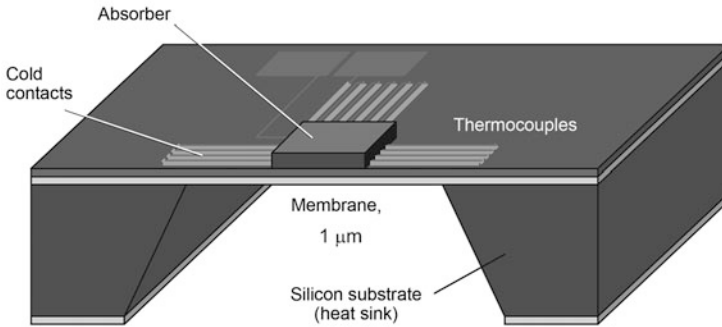


Fig. 14.20 Schematic section through a MEMS thermopile sensor

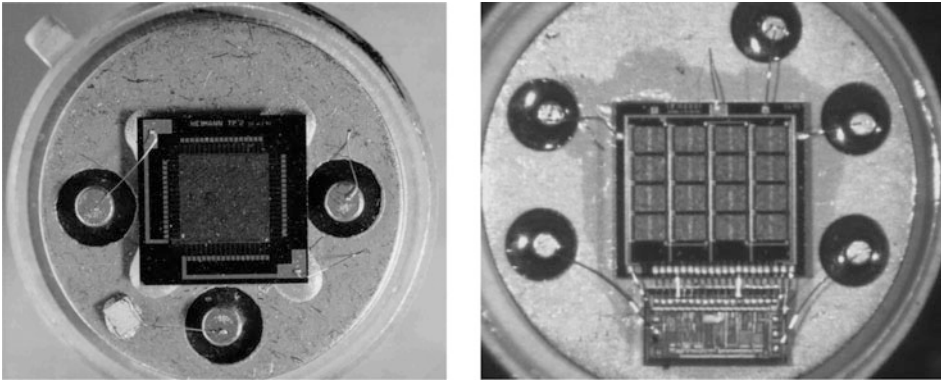


Fig. 14.21 Left: Thermocouple sensor. Right: Thermal column sensor array consisting of 4×4 detectors (middle of the picture) with ASIC for signal processing (bottom of the picture) (Factory photos: PerkinElmer Optoelectronics)

When using a polysilicon/aluminum thermopile, the manufacturing process is CMOS-compatible. Such a sensor typically achieves 55 V/W sensitivity with a thermopile resistance of 75 k Ω .

14.7.2 Pyroelectric Detectors

In contrast to a thermopile, a pyroelectric detector (Sect. 2.13) does not react to a continuous signal but to the *change in* the received *IR radiation*. Infrared radiation is absorbed and heats pyroelectric materials, generally lead zirconium titanate (PZT ceramics) or lithium tantalate crystals, which react to heating with a strong change in electrical polarization. This is capacitively removed and amplified by means of electrodes on the material.

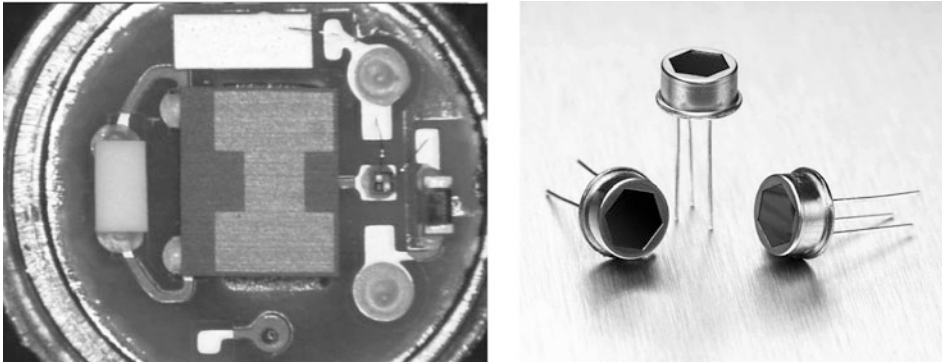


Fig. 14.22 Infrared detector based on the pyroelectric properties of PZT ceramics (Factory photos: PerkinElmer Optroelektronics)

Figure 14.22 shows such an IR detector. Several fields are defined by the electrodes in such a way that an IR image moving on the detector results in a signal. Such sensors are generally used in motion detectors.

The module contains four lithium tantalate crystals (Fig. 14.23), two of which are assigned to each optical channel. *One crystal* is provided with an *IR radiation-absorbing surface* in each case and is used for *radiation detection*. The respective other crystal is provided with an *IR-reflecting coating* and is connected opposite the IR-measuring crystal. It serves to *compensate for fluctuations in ambient temperature*. The shown sensors contain two such channels each for different IR wavelengths.

14.7.3 Bolometer

In a bolometer, *infrared radiation* is *absorbed* and heats a temperature-dependent resistor, for example, a thermistor, which in turn generates an electrical signal by changing its resistance. In contrast to thermopiles or pyroelectric detectors, thermistors can be *miniaturized* very well, so that a large number of thermistors can be integrated on a sensor as a microbolometer array for infrared imaging (*thermography*). Bolometer arrays can operate at room temperature, but must be temperature-stabilized and use an *optical chopper* (shutter) to measure and compensate for the zero drift of the pixels.

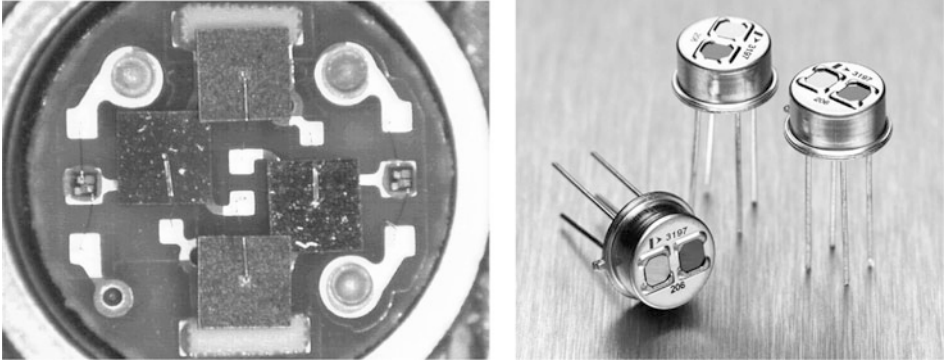


Fig. 14.23 Infrared detector based on the pyroelectric properties of lithium tantalate (Image material: PerkinElmer Optoelectronics)



Gert Schönfelder

Sensors convert the physical quantity to be measured into an electrical signal. There are three points to consider:

1. The sensor signal is usually *very small*. It must therefore be amplified to a *standard level* so that it can be processed by a controller.
2. The influence of *disturbance variables* must be taken into account (e.g. influence of temperature on the characteristic curve of a humidity sensor).
3. Characteristic curves must be *filtered* and *linearized* very often.

Temperature measurement has a special position here. There are measuring elements with *undisturbed* and *linear* character (e.g. PT100 and thermocouple). These elements can often be connected directly to special inputs of control systems. In this sense, they form their *own signal standard*.

15.1 Signal Processing

With a large number of sensor elements, the output signal is generated by relatively few physical effects, which can be summarized as follows

- Generation of a *charge* or a *voltage* (e.g. piezoelectric effect, thermocouple, photoelement).
- Change of *resistance* or a *conductance* (e.g. strain gauge, photoresistance, change of the channel resistance of FETs by ions).

G. Schönfelder (✉)
Prignitz Mikrosystemtechnik, Wittenberge, Germany

- Change in *capacitance* or *inductance* and resulting *AC resistance* (e.g. humidity, distance and distances).

There are many specific methods of signal acquisition for discrete sensor systems. These range from pure *binary information* to the formation of *pulses* and *pulse groups* and *complex data sets* for image evaluation. In this group, time or frequency as an information-carrying variable is very common.

The following sections cover the basic possibilities of signal processing. These considerations are carried out exemplarily on a resistive sensor element since in this case the methods can be clarified very well.

15.1.1 Analog (Discrete) Signal Conditioning

The classic, analog signal conditioning still has its importance when it comes to *low-noise sensor signals*. Due to the simple circuit technology there are hardly any additional noise sources. This method is also suitable when a signal of low quality is required, for example for switching purposes.

In the circuit, in Fig. 15.1 the sensitive element is formed by the resistors R1 to R4. The available output signal of the bridge is tapped symmetrically with the differential amplifier IC1A and amplified asymmetrically to the standard level with IC1B. Zero adjustment is carried out with resistor R5, full-scale adjustment with resistor R7.

15.1.2 Signal Conditioning with System Circuits

In connection with frequently occurring tasks, system circuits have been developed by the industry which allow the coupling of different sensor elements and convert them to a standard signal. In Fig. 15.2, an IC from Texas Instruments was selected as an example, which is used to convert a thermal resistance to a current loop signal.

As is typical for system circuits, it contains not only the amplifiers but also the necessary additional components such as current sources and multiplexers. In the example, the necessary settings are stored in an external EEPROM.

15.1.3 Signal Conditioning with ASICs

For standard measuring tasks such as temperature, pressure or displacement measurements, it is advisable to develop circuits that are adapted to the specific measuring tasks of the individual sensor elements. This results in highly complex components that, thanks to a large number of adjustable parameters, cover the complete range of tasks from signal acquisition, processing and correction to analog or digital data output.

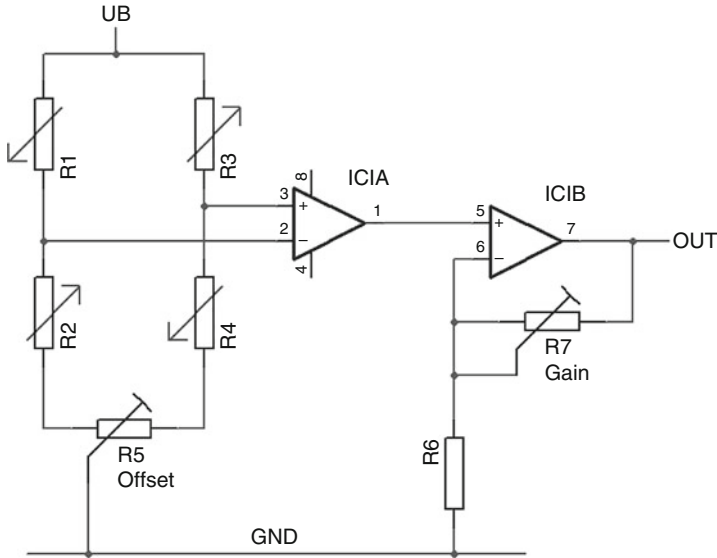


Fig. 15.1 Basic circuit for analog signal conditioning

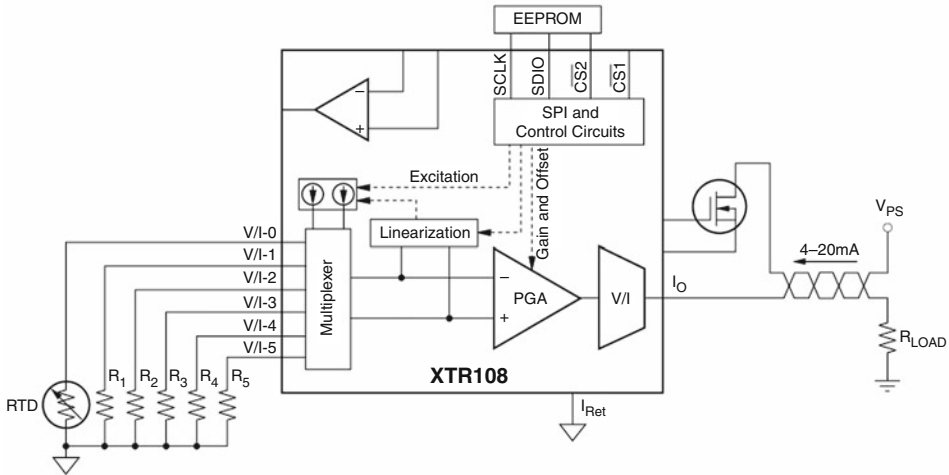


Fig. 15.2 System circuit XTR108 (Source: Texas Instruments)

As an example, the ZMD31050 was picked out, which is shown in Fig. 15.3 in the principle circuit for a ratiometric pressure transmitter. It has a 16-bit AD converter, an input filter and an 11-bit DA converter. Diode D1 is used to determine the temperature of the bridge which is required for compensation of the pressure characteristic curve. The output signal is calculated by solving a third-degree equation with both input variables.

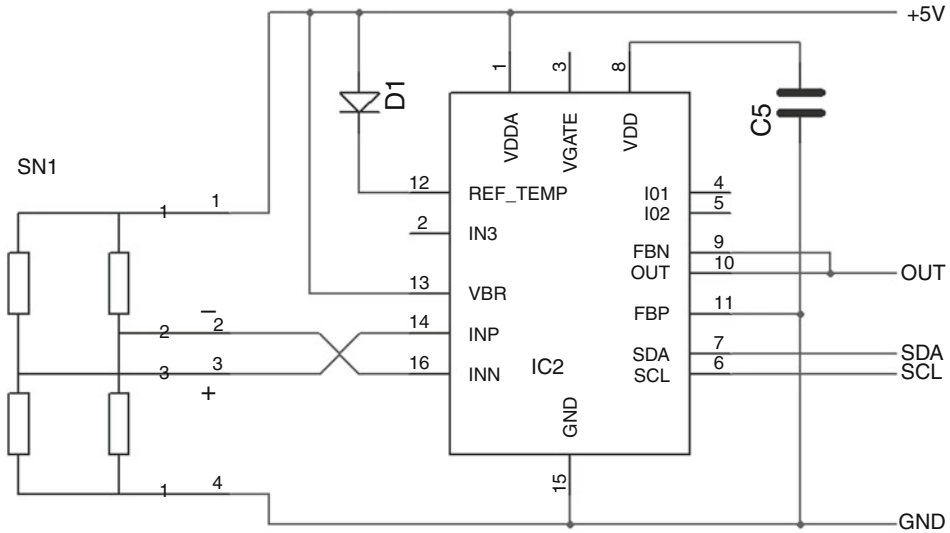


Fig. 15.3 ASIC for pressure transmitter

15.1.4 Signal Conditioning with Microcontrollers

The sensor signals can also be processed directly with a microcontroller. The most important requirement here is an analog-to-digital converter with a corresponding accuracy and the possibility of outputting an analog signal. This can be realized by a DA-converter or a PWM-unit. These units do not necessarily have to be integrated into the controller.

If the signal acquisition and processing is done with a processor, it is always in connection with an adjustment.

The advantage of this processing path is that characteristics on the input and output side that cannot be converted by analog means can also be realized. This is made possible by the use of the software. Figure 15.4 shows a circuit section for a sensor with a resistance bridge, which realizes a voltage output of, for example, 0–10 V.

However, this path has the disadvantage that signal processing with software is always slower than in an analog signal path. It becomes slower the more complex the correction algorithm and signal processing becomes (Sect. 15.2.3).

15.2 Sensor Calibration

Although the term sensor calibration has become established in linguistic usage, it is wrong in the technical sense. In *calibration*, the *deviation* of a measured variable from a *standard* is documented. If the *signal* of a sensor is adjusted to a *setpoint value*, this is called an *adjustment*. The sensor adjustment has the task of freeing the measurement signal from

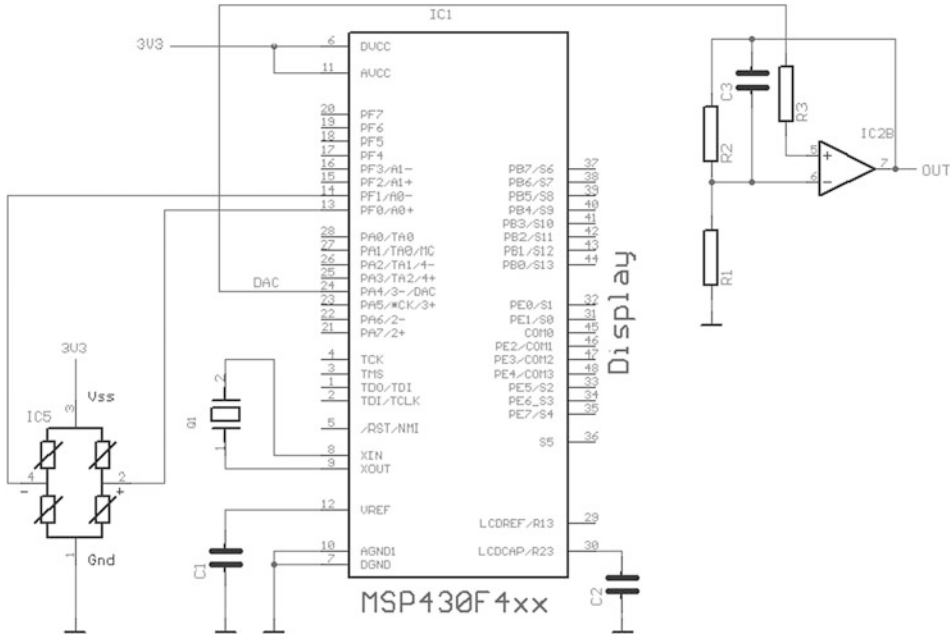


Fig. 15.4 Sensor circuit with a microcontroller

nonlinearities and disturbance variables by technical or mathematical methods. The main disturbance variable is usually the temperature.

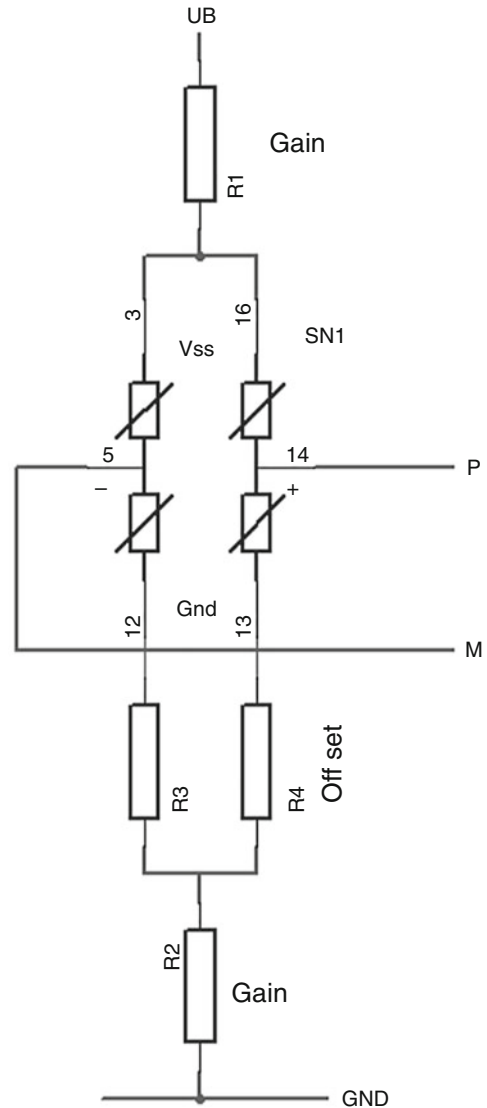
15.2.1 Passive Compensation

For simple applications, for example in connection with the circuit in Fig. 15.5, a passively compensated measuring bridge can be used. This compensated sensor element then generates a defined output signal and can thus be conditioned with a permanently set simple amplifier. The adjustment of the resistors R1 to R4 is carried out within the scope of sensor manufacture. Figure 15.5 shows the schematic.

The resistors R3 and R4 are used to zero the bridge. Depending on the sign of the offset, one or the other resistor must be detuned.

The total resistance of the bridge and the change in resistance due to the measured variable results in a *slope*, which is specified in mV/V(U_b). This ratio can be changed by the resistors R1 and R2 so that the slope takes on a defined value. This target value is always smaller than the possible slope. The resistor value determined for this purpose is applied to R1 and R2 half each so that the output voltage of the bridge remains at U_b/2.

Fig. 15.5 Passive compensated measuring bridge



15.2.2 Adjustment with Analog Signal Processing

If the required corrective measures and properties are known for a sensing element, the required equation can be simulated as a sequence of linear and non-linear amplifier elements. In Fig. 15.6 an older analog ASIC is used as an example. The whole circuit is composed of amplifiers and multiplying DA-converters. As a result, a linear characteristic

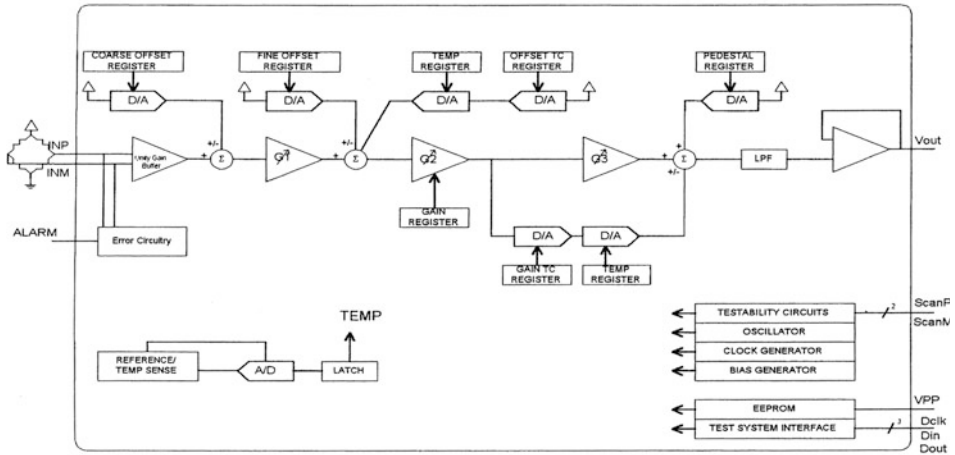


Fig. 15.6 ASIC with analog signal processing (SCA2096)

with offset and gain correction is realized for the useful signal, which is influenced in a second path by the measured temperature with a second-degree characteristic.

The adjustment procedure is carried out in three steps:

1. At room temperature, the offset (left two steps) and gain (middle step) are set as a direct relationship between input and output signal.
2. Then the adjustment is carried out at a higher temperature when the input conditions are repeated by changing the offset and gain of the temperature branch.
3. This process is repeated at a lower temperature.

If higher accuracies are required, these adjustments have to be run through several times, since the settings of the individual registers are not without effect on the other settings. The found settings are permanently stored in a memory in the ASIC. The ASIC itself generally works ratiometrically, that is, the output signal also behaves proportionally to the supply voltage.

15.2.3 Adjustment with Digital Signal Processing

A wide range of solutions exists for digital signal processing. Between the ASIC with a pure analog signal path and the pure computer solution, there is a wide range of ICs, which are equipped with an analog front end and have a special, fixed programmed computer structure internally. One of these is the circuit shown in Fig. 15.3, which is used to realize an arithmetic adjustment.

The basic structure in digital signal processing always consists of an input stage with offset correction and a high-resolution AD converter and an output stage as a DA converter.

It is irrelevant whether this is formed by separate components or integrated into an ASIC or processor module. The starting value for the processing is always the value of the input converter, which is supplemented by any interference signals measured on other channels. The result of the operation must be a value for the output DA converter, for the determination of which there are several methods.

Arithmetic method

To use this method, the *correction function* must be available as an *equation* through a closed mathematical representation. During operation, the processor calculates this instruction cyclically, using the input variables and the coefficients determined during calibration.

To determine the coefficients, a system of equations with n unknowns must be solved according to the degree of the functions. In the example of the ZMD31050, these are 7–9 equations for which a corresponding number of measured values is necessary. These measured values are obtained by setting the complete transmitter, for example, to different temperature/pressure combinations, and reading out the corresponding AD values for pressure and temperature. The equation system is then solved in an external computer and the results are stored in the ASIC. Figure 15.7 shows the equations of the ZMD31050. The computational effort can easily be estimated here.

In the example case, this method allows a high measuring rate (up to 4000 measurements/s), since the computer structure is adapted to the equation. If this path is taken by means of general processors, very long computing times can result, which drastically reduce the measuring rate. Some time can be gained by dispensing with floating-point arithmetic, but this requires exact knowledge of the measuring elements and their scattering.

Characteristic curve interpolation

If the arithmetic path cannot be traversed, for example, because no closed equation can be created, the path of characteristic curve interpolation is possible. Here, several interpolation points are determined from the input characteristic curve. The positions of the interpolation points (X-axis) are stored in the computer and the rise and offset of the characteristic curve range between them is determined from the measured values of neighbouring values. By using several interpolation points, the characteristic curve section can also be represented by a function of a higher degree. During the execution of the program, it is determined between which interpolation points the input signal is located and the corresponding set of coefficients is selected for processing. The calculation is then relatively simple.

If disturbance variables are still included in the processing, the characteristic curve expands to the surface or the room. A set of coefficients is then stored for each direction and each pair of interpolation points and the dimensions are processed sequentially.

If the interpolation points are closely spaced, linear interpolation is generally sufficient to determine the output value.

<p>⇒ Range definition of inputs</p> $Z_p \in [-2^R; 2^R] \quad Z_{T1} \in [-2^{R-1}; 2^{R-1}]$ $c_i \in [-2^{15}; 2^{15}]$ <p>⇒ Conditioning Equations</p> $Y = \frac{Z_p + \text{Offset}(Z_{T1})}{\text{Gain}(Z_{T1})} \quad Y \in [0; 1)$ $Y = \frac{Z_p + c_0 + 2^{-(R-1)} \cdot c_4 \cdot Z_{T1} + 2^{-2(R-1)} \cdot c_5 \cdot Z_{T1}^2}{c_1 + 2^{-(R-1)} \cdot c_6 \cdot Z_{T1} + 2^{-2(R-1)} \cdot c_7 \cdot Z_{T1}^2}$ $P = Y \cdot (1 - 2^{-15} \cdot c_2 - 2^{-15} \cdot c_3) + 2^{-15} \cdot c_2 \cdot Y^2 + 2^{-15} \cdot c_3 \cdot Y^3$ $P \in [0; 1)$	<p>R ... Resolution of A/D-Conversion</p> <p>Z_p ... Raw A/D-result for pressure (auto-zero compensated)</p> <p>Z_{T1} ... Raw A/D-result for temperature 1 (auto-zero compensated)</p> <p>Conditioning coefficients stored in EEPROM Register 0 to 7; c_i ∈ [-2¹⁵; 2¹⁵] complement on two:</p> <p>C₀ ... Bridge Offset</p> <p>C₁ ... Gain</p> <p>C₂ ... Nonlinearity 2nd order</p> <p>C₃ ... Nonlinearity 3rd order</p> <p>C₄ ... Temperature coefficient Bridge Offset 1st order</p> <p>C₅ ... Temperature coefficient Bridge Offset 2nd order</p> <p>C₆ ... Temperature coefficient Gain 1st order</p> <p>C₇ ... Temperature coefficient Gain 2nd order</p>
--	--

Fig. 15.7 Equations for adjusting the ZMD31050

The characteristic curve interpolation is faster in execution than the arithmetic method because it is based on simpler operations. A disadvantage is the large number of necessary measuring points and the relatively large coefficient memory.

Look-up table

The fastest and most memory-consuming way is to correct the data using a look-up table (LUT). For the simple characteristic curve, a corresponding output value is stored in a table for each possible input value. This assignment can also be multidimensional. In this way, very high accuracies can be achieved. However, since this requires many tables, a *reduced table* is often created in which intermediate positions are obtained by interpolation. This can also be regarded as a *characteristic curve interpolation* with a *high number of interpolation points*.

The clear advantage, however, is the *processing speed*, since only memory accesses are involved. If the interpolated LUT is set to binary values, for example, to every fourth input value, the interpolation can be limited to addition, subtraction and shift operations, so that high processing speeds are achieved here as well.

15.3 Energy Management for Sensors

Energy management is a current topic not only because of the increased environmental awareness but also because of a multitude of new applications. Today, sensor applications are increasingly used in mobile and self-sufficient systems that need to be supplied with

energy. To keep energy consumption as low as possible, the power consumption in the sensor must be as low as possible.

There are two basic ways to manage energy: Either you build an *extremely energy-saving* sensor (e.g. measuring consumption on the radiator) or you *reduce the active time* of the sensor. Especially the latter variant requires a processor to realize this control process.

To produce a low energy sensor, several factors and influences must be taken into account. In the considerations, it is assumed that a processor core is present in the signal processing.

Circuitry measures and quiescent current

The processor used must itself have internal power management, that is, the peripheral units must be *switchable*. This means that only the required parts can be activated. In addition, the power consumption should be as low as possible in a sleep mode, although basic functions must still be active in this state, such as the clock and possibly a display. These design requirements also apply to the same extent to the external circuitry. For example, the sensor element, any external amplifiers and the areas of data communication must also be switchable.

The values for the quiescent current achievable with current processors are in the range of less μA . As an example, the circuit shown in Fig. 15.4 is used to achieve 13 μA in the quiescent state with the display is running.

Measuring time and active current

The two variables are related because they influence each other directly. In CMOS circuits, the *current consumption* is always related to the *clock speed*. The faster a processor works, the higher its power consumption. On the other hand, the measurement process is completed faster with a faster processor. The energy consumption is calculated as the product of time and current. Therefore, other criteria must be found that enable fast execution at a lower clock frequency. These include, for example:

- The influence of the selected *software* or *programming language*.

The programming of the functions should be done in a minimalist form. The use of common high-level languages is rarely optimal because too many unnecessary functions are dragged along. Since the signal processing tasks are compact, one should also consider assembler programming.

- The influence of the *correction algorithms*.

As already mentioned in Sect. 15.2, the choice of the correction path has a strong influence on the computational effort and the execution speed.

- The choice of *number format*.

Calculating a signal correction in floating-point format is the easiest way, but also the slowest. Therefore, it should be considered to perform the calculations without a floating-point in an integer format. However, this requires more effort, since the

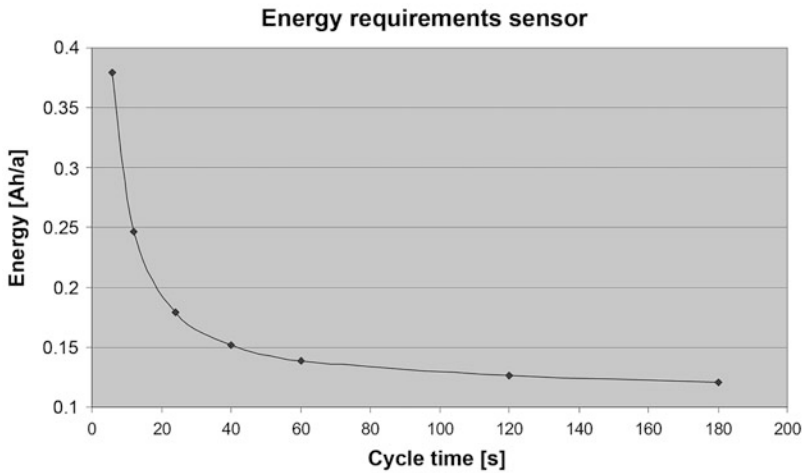


Fig. 15.8 Energy requirement as a function of the active time

estimation of the number range in connection with the sequence of arithmetic operations must be made in advance.

- *Place of processing.*

A part of the correction tasks can also be shifted to the superordinate computer system in favour of active time. This allows time-consuming functions to be delegated to a location without energy problems.

Measuring cycle

The frequency of measurement also influences the *energy balance*, as it determines the ratio of active time to resting time. For this reason, it is necessary to define exactly how often measurements must be taken in a specific case. The interval between measurements can range from seconds for technical applications to hours for climatic sensors.

Figure 15.8 shows the energy requirement of a digital manometer for 1 year as a function of the measuring cycle. The quiescent current is 13 μA , the active current 3.7 mA and the active time 50 ms.

Bibliography

1. Beschreibung EVAL-Kit zum ZMD31050: ZMD31050 Application Kit Description; Seite 27 über: <http://www.zmd-gmbh.de/products.php>
2. Datenblatt „Two-Wire Transmitter XTR108“; über: <http://www.ti.com/>
3. Datenblatt „ZMD31050“; über: <http://www.zmd-gmbh.de/products.php>



Gert Schönfelder

Standardized interfaces are used to transmit the measured values from the transmitters to the evaluation unit. These allow the user to easily connect transmitters from different manufacturers. The possibilities for analog signal transmission are based on a few physical variables. Transmission on a digital basis is limited to the most important representatives, which are decisive for sensor technology.

16.1 Analogue Interfaces

Voltage, current, frequency or pulse width are used for signal transmission in analog form. These output variables are easy to control in practice and can be transmitted by means of two or three connecting wires. With analog signals, *the measured variable* is always *mapped* to the *specified signal range*. This can mean, for example, that a pressure of 0 bar to 20 bar or a temperature of $-20\text{ }^{\circ}\text{C}$ to $50\text{ }^{\circ}\text{C}$ is mapped to a signal of 0 V to 10 V. This shows the first fundamental disadvantage: At the other end of the line, there is no more information about the measured variable.

The second major disadvantage of the analog transmitters is the *cabling* in a *star topology*, *that is*, each measured value is routed individually to the processing unit. The resulting cabling effort is easily imaginable with the increasing use of sensors.

The most important properties of the analog interfaces are summarized in Table 16.1.

The analog variables frequency and phase stand out from voltage and current in that they do not represent static signal transmission. Although their resolution is infinitely

G. Schönfelder (✉)
Prignitz Mikrosystemtechnik, Wittenberge, Germany

Table 16.1 Characteristics of analog transmission paths

Information media	Reach	Resolution	Advantages	Disadvantages
Voltage	Several metres	Up to 16 bit	Simple generation relatively low power consumption	Voltage drop on transmission link (error due to transmission)
Current 2-wire	Up to km	Up to 16 bit	Large transmission distance, since no error due to transmission Only double pole connection	High power dissipation Only sensors with low own consumption
Current 3-wire	Up to km	Up to 16 bit	Large transmission distance, since no error due to transmission More current available for the sensor element	High power dissipation
Frequency	Media dependent	Up to 20 bit	The high dynamic of the measured variable Stable reference value (time) alternative transmission media possible	High transmission bandwidth required
Phase/PWM	Media dependent	Up to 12 bit	Stable reference value (time) alternative transmission media possible	High transmission bandwidth required Error due to runtime and bandwidth

variable, which characterizes them as an analog signal, they behave like a high-frequency signal during transmission.

16.1.1 Voltage Output

The standard signal from 0 V to 10 V represents the basic type of voltage output. This is independent of the supply voltage used for the transmitter, which typically ranges from 15 V to 30 V. The load resistance is usually more than 10 k Ω .

The main disadvantage of voltage output is that the output signal is distorted by long transmission paths, as their resistance causes a voltage drop. In the same way, an offset voltage can be caused by the voltage drop on the ground line caused by the operating current.

The advantage of this interface is the *simple signal evaluation* because this voltage range can be digitized and processed with simple means.

There are still some special versions on the market which differ in the range of the output voltage. For example, the 0 V to 5 V variant is occasionally found, which then manages with a lower supply voltage. By shifting the zero point to, for example, 1 V, problems with the zero point can be avoided, which can occur with simple output stages in the transmitters.

16.1.2 Ratiometric Voltage Output

The ratiometric voltage output represents a separate form. It has the special feature that the measured value is assigned to a *percentage of the supply voltage* U_v . This type of transmitter is generally operated with 5 V. The available range of the output voltage is then 5 % to 95 % U_v (0.25 V to 4.75 V) or 10 % to 90 % U_v (0.5 V to 4.5 V). This already avoids problems of the output stages in the area of supply voltage and ground.

The advantage of this interface is the *uncomplicated analog-to-digital conversion* since the supply voltage can be used as a reference for the converter. Fluctuations in the supply voltage cancel each other out. This interface type is often used for compact systems within devices. As it is technically simple, it is the cheapest type.

16.1.3 Current Output

The most important advantage of this interface type should be mentioned first: It can be used with almost any length of cable. Although the current generated by the transmitter causes a voltage drop on the connecting lines and thus increases the necessary supply voltage of the system, the measured variable itself is not affected.

At the control unit, from where the power is supplied, the flowing current is converted into a voltage via a resistor, which is called a *load*. Typical values for this resistor are 100 Ω or 200 Ω . The maximum value is generally 600 Ω , depending on the system.

In practice, two variants of the current interface are used: a two-wire and a three-wire circuit.

Two-wire interface

With the two-wire interface, only the supply voltage is connected and the current consumption of the transmitter is modulated (Figs. 16.1 and 16.2). As the electronics for signal processing have their own power consumption, the current for the smallest measured value cannot, therefore, be zero. A value range from 4 mA to 20 mA for the measured value has become the standard. However, this type can only be used if the electronics can manage with the remaining base current of less than 4 mA.

Fig. 16.1 Current transmitter with current measurement

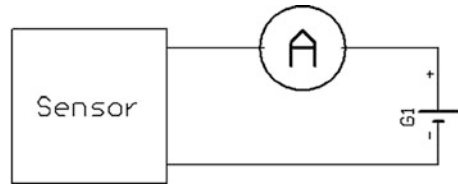
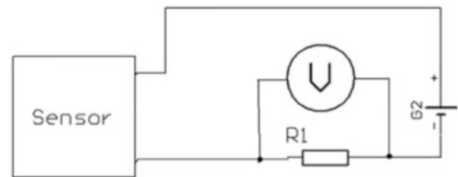


Fig. 16.2 Current transmitter with voltage measurement at load



Three-Wire Interface

The three-wire interface is normally connected to a supply voltage. The current driving output is connected with the load to the ground (Figs. 16.3 and 16.4). Due to this connection version, the transmitter itself may have higher internal consumption. Here, the current range from 0 to 20 mA has become the standard. However, the version with 4 mA to 20 mA is often also offered here to ensure a uniform system design.

16.1.4 Frequency Output and Pulse Width Modulation

The use of time-based sizes has the advantage that the simple transmission is accompanied by easy digitization and a good resolution. Frequency information can be easily reconstructed after a long transmission distance.

Frequency output

In the case of information transmission with an analog signal, there is a technically induced limitation of the possible resolution of the measurement (Fig. 16.5). This may be due to the loss resistances of the transmission path or signal noise, for example. The ohmic loss during transmission of a frequency-modulated signal can be eliminated by a trigger at the endpoint since the amplitude of the signal has no influence on the information. This also means that there is no error due to noise.

Due to the advantages described above, a frequency is used as an information carrier, especially for sensors with high dynamics. An example are light sensors with a dynamic range of 10^4 and more.

Pulse width modulation

In pulse-width modulation (*PWM*), a square-wave signal is transmitted in which the *ratio of period duration* (rising edge to rising edge) and *high duration* (rising edge to falling

Fig. 16.3 Three-wire interface with current measurement

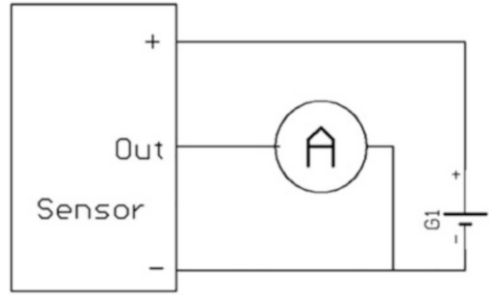


Fig. 16.4 Three-wire interface with voltage measurement

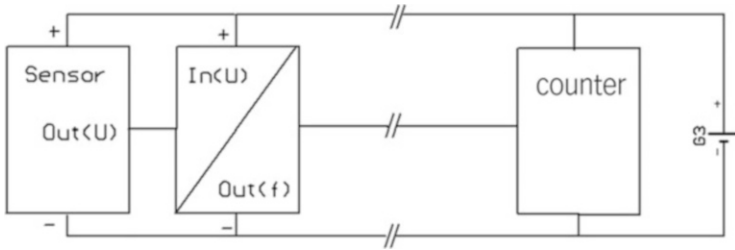
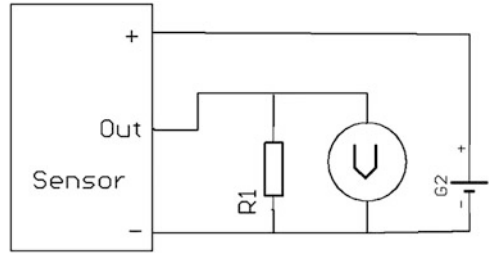


Fig. 16.5 Sensor with frequency output by U/f converter

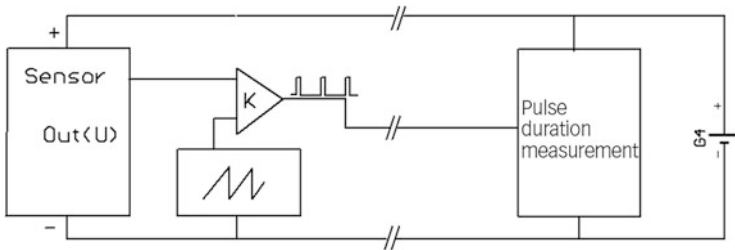


Fig. 16.6 Generation of a PWM signal

edge) is used as the information carrier. This signal form is independent of the frequency used and its stability (Fig. 16.6).

The *advantage* of this interface is the *simple implementation* of the transmitter and receiver. Most microcontrollers have a PWM unit.

The *disadvantage* lies in the *square-wave signals* used, as these generate a large number of harmonics and thus place high demands on the transmission path. Since the time between two signal edges carries the information, sloped edges, caused by limited bandwidths, lead to a transmission error. This becomes particularly extreme when small-signal values are transmitted. In this case, the remaining needle pulse has a 1000 times larger bandwidth than the working frequency. For these reasons PWM interfaces operate in the frequency range from a few 100 Hz to a few kHz and usually have a resolution of 10 bits to 12 bits.

16.1.5 4-/6-Wire Interface

The 4- and 6-wire interface occupies a special position in sensor technology, as it is not assigned to level ranges, but represents a schematic method (Fig. 16.7). It occurs mainly in connection with passive measuring elements (e.g. temperature sensors, Chap. 6).

The basis is the *separation of supply and measuring lines* on one sensor element. The aim is to compensate for the voltage drop on the supply lines for the measurement by measuring directly at the sensor element using separate measuring lines. As the measuring amplifiers are high-impedance, no current flows on the measuring lines and thus no voltage drop occurs. This makes it possible to measure small voltages very accurately even on long lines. This circuit technology is therefore used when very small signals are to be measured and transmitted with a very high resolution. Measurement resolutions of up to 24 bits are thus possible, which is practiced, for example, when using strain gauges.

With the 6-wire circuit, additional reference lines are added to the supply lines and the measuring lines, which must also not be loaded. This is necessary if, in a bridge circuit, the voltage of the bridge and the measured variable must be transmitted. The bridge voltage is then used as a reference voltage for the A/D conversion (Fig. 16.8).

16.2 Digital Interfaces

With the digital interface, the transmission is also based on voltage or current modulation. The free design of the protocol allows a large number of variants. Digital interfaces are often designed as a bus, which simplifies cabling. In practice, however, all structures known from network technology, such as star, ring and bus topology, are used.

The most important advantage of digital interfaces is the transmission of *more extensive information* for each sensor. Thus, the processing unit not only has the standardized measured value at its disposal but also, for example, *information* about the *measuring*

Fig. 16.7 Arrangement for a four-wire interface

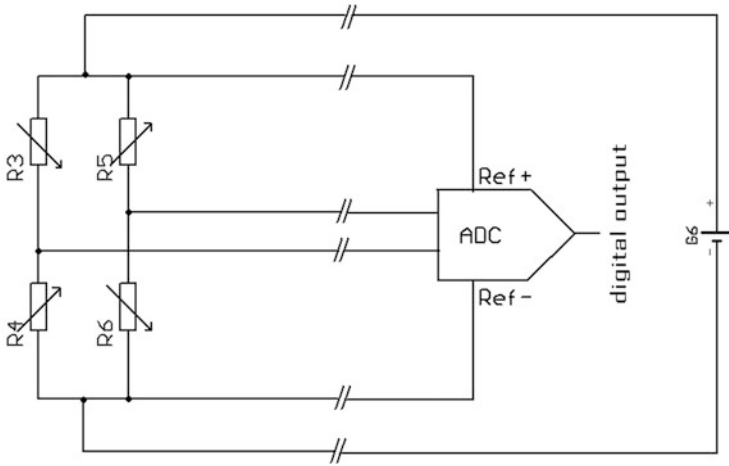
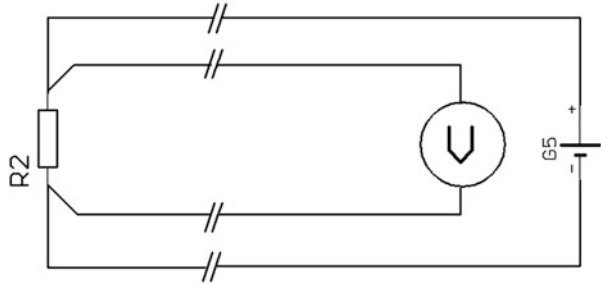


Fig. 16.8 Use of a 6-wire interface for bridge applications

range, the measured variables, possible limit values of the sensor and much other information. Only this information allows the automatic administration and maintenance of a network of several hundred sensors.

With digital transmission, features such as secure data transmission, error correction and bidirectional communication with the sensor should also be mentioned.

Table 16.2 gives an overview of selected interfaces.

The reference in Table 16.2 to required licenses refers to the manufacturer of the corresponding components and not to the user.

The digital interfaces generally require a *comprehensive software protocol* which secures the data traffic between sensor and application. This requires a corresponding computer structure both on the transmitter side and in the higher-level control system.

The types selected here represent a small selection from the range of applications for sensor technology.

Table 16.2 Properties of digital transmission paths

Interface	Reach	Data rate	Advantages	Disadvantages
CAN	Up to 1500 m	to 1 MBit/s	Good range Simple drivers Implemented in modern processors	<ul style="list-style-type: none"> • Extensive protocol software required • Numerous sub-protocols and special versions • License required • Small data packets of 8 bytes
LON	to 2700 m	Up to about 1 MBit/s Type 78 kBit/s	Very large networks Hierarchically structured	<ul style="list-style-type: none"> • Extensive protocol software required • High hardware expenditure • License required—also for software per node
HART	Several 100 m	About 100 byte/s	Only two wires for connection Old cabling can be used	<ul style="list-style-type: none"> • Extensive protocol software required • Very slow only • Star topology • License required
RS485	Several kilometers	1...15 MBit/s	The technical basis for other interfaces Simple implementation (UART in CPUs usually present)	<ul style="list-style-type: none"> • No standardised software
Profibus	Up to 9 km	1...12 MBit/s	Simple hardware Large expansion High data rate	<ul style="list-style-type: none"> • Extensive protocol software required • Small data packages • License required
IO-Link	20 m	4800 Bd 38400 Bd 230400 Bd	Use of simple (existing) cables Mixable with “normal” sensors	<ul style="list-style-type: none"> • Only point-to-point connection • Extensive protocol software • Required License required
I2C	Board	to 400 kBit/s	No overhead defined—therefore easy implementation Implemented in many microcontrollers Bus topology	<ul style="list-style-type: none"> • Only a local use • License required for hardware implementation

(continued)

Table 16.2 (continued)

Interface	Reach	Data rate	Advantages	Disadvantages
SPI	Board	to 10 MBit/s	No overhead defined—therefore easy implementation In many microcontrollers implemented No license required	<ul style="list-style-type: none"> • Local use only • Point-to-point connection only
IEEE 1451	Board	from 2 kBit/s	Supported by many microcontrollers, since SPI-based “bus-capable” SPI variant License-free	<ul style="list-style-type: none"> • Local use only • Simple protocol must be implemented

16.2.1 CAN Group

The generic term “CAN” covers several transmission protocols. The CAN bus itself is mostly used by them only as a data transmission and protection element. Analogous to the ISO-OSI layer model, one speaks of higher-level or application layer protocols.

CAN

The CAN protocol is specified in the ISO 11898 standard. It is a *message-oriented protocol*. In principle, all participants have equal rights for bus access. Messages are prioritized via an 11-bit or 29-bit identifier. Up to 8 bytes of user data are transported per transmission packet. The expansion of the network is inversely proportional to the transmission speed. It can range from 100 m at 500 kBit/s to 1500 m at 50 kBit/s or even more. The maximum transmission speed is 1MBit/s. The number of participants in the network is limited by the selected driver modules. Usually, there are 64 participants per network today. Twisted two-wire lines are often used as a transmission medium. Fibre optic cables are also used occasionally.

By prioritizing using identifiers, important messages can be transmitted with short latency despite high busload. This enables the system integrator to handle errors despite overload.

CAN open

The CAN in Automation e.V. (registered association) user association (CiA) specifies several mechanisms for controlling and implementing network nodes. Network management messages are used to adjust the transmission and reception behavior of the individual nodes. The associated mechanisms are summarized in the CiA DS 301 regulation. Using the LSS procedure (CiA DS 305), the individual nodes are configured in terms of transmission speed and identifier.

An electronic datasheet (EDS) provides information about the functionality of the individual participants. It is carried separately as a file or read out directly from the device.

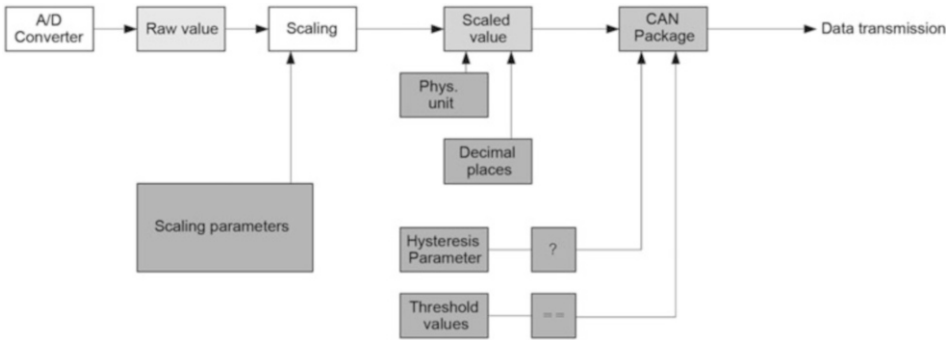


Fig. 16.9 Structure of a CANopen sensor

The individual application areas are served by so-called device profiles. The CiA DS 404 profile has been established for the area of sensor technology. The specific characteristics for the digital transmission of analog measured values are fixed in this profile. In addition to the coding of the measurement data, parameters for digital post-processing of the measurement signal are also available. Figure 16.9 provides a rough overview of the standardized options.

There is a multitude of other device and application profiles. New profiles are constantly being defined by the user association. This enables a wide distribution and comparatively universal application of the CANopen protocol. The conformity of a device with the CANopen protocol is confirmed by the CiA in a conformity test.

SAE J1939

With the J1939 protocol, the Society of Automotive Engineers (SAE) has introduced a CAN-based application protocol specifically for use in commercial vehicles.

It basically uses the 29 bit CAN identifier. The transmission speed is currently fixed at 250 kBit/s. This automatically results in the physical boundary parameters of the CAN network. The concrete identifier assignment is integrated into an automatic process, which is supported by every device.

Due to the specific focus on the commercial vehicle sector, the SAE does not provide for conformity testing. The individual participants are usually configured manufacturer-specific. The coding of the user data is also based on the manufacturer's specifications. Conformity with the automatic ID assignment system is checked at the so-called "PlugFest" events. There, J1939 devices are randomly connected and commissioned.

Based on the functions of the J1939, further functionalities were specified. Especially for commercial vehicles in the agricultural sector, the ISOBUS according to ISO11783, the "Truck & Trailer Interface" according to ISO 11992 and especially for vehicles in the maritime sector the NMEA2000 attachment are well-known. Here too, all participants are configured manufacturer-specific.

Further CAN-based systems

Due to the widespread use of the CAN interface in a large number of microcontrollers and its comparatively easy handling, a large number of other protocols have established themselves on the market. Many manufacturers also use the CAN bus for proprietary transmission protocols. DeviceNet is another popular standardized protocol. This is maintained and managed by the Open DeviceNet Vendor Association (ODVA) and is widely used in the American economic area.

16.2.2 LON

The LON bus (local operating network) is designed for the connection of large quantities of sensors and actuators. The network is structured hierarchically. The levels are divided into domains, sub-networks and nodes, whereby a domain can contain 32,385 nodes. It is possible to connect up to 248 domains in one system.

The total extent of the network can be several kilometres. The possible transmission rate depends on the topology and extent of the network and can be between 2 kBit/s and 1.25 MBit/s.

This comprehensive structure requires appropriate software in the nodes. For this reason, the company Echelon has developed a “Neuron” processor that contains these functions.

The interests of the LON network are represented in Germany by the LNO (www.lno.de).

16.2.3 HART

The HART interface represents a compromise between the analog world and the desire for more information from the sensor. The basis is a transmitter with a current interface, which is normally designed as a two-wire system. The analog measured value is transmitted unchanged via this path (Fig. 16.10).

Digital data transmission is achieved by *modulating* an AC voltage to the *analog current value*. Since this AC voltage has a zero mean value, it does not influence the analog measurement result. A HART modem is installed at both ends of the transmission path, which carries the modulation or filters out the AC voltage for analysis. Due to this setup, commands can be sent to the sensor and response can also be generated by the sensor.

The protocol is an *asynchronous half-duplex procedure*. The data is coded according to a *frequency shift procedure (FSK)*, where the “0” is assigned the frequency 2.2 kHz and the “1” the frequency 1.2 kHz. Data transmission is relatively slow at about 100 byte/s. The transmission is always carried out as a *point-to-point connection* (star topology). Operation

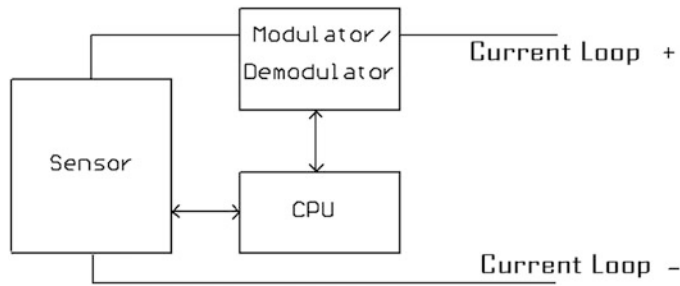


Fig. 16.10 Sensor structure for a HART sensor

as a bus is possible, but in this case, the analog measured value is not required. The interests of HART users are represented internationally by the HART Foundation.

16.2.4 RS485

Data transmission according to RS485 is an old standard that defines more the physical conditions than the content and form of the transmission. It is an *asynchronous* transmission in *half-duplex mode*, that is, it is sent or received one after the other by the master. The bus is a twisted two-wire line with a cross-section of 0.75 mm^2 , terminated at the ends with 120Ω each. The data is sent as a differential voltage signal and is therefore resistant to interference. Typically, up to 32 participants can be connected, which can be distributed over several hundred meters of bus length. With modern driver circuits, higher numbers of participants can also be achieved.

This interface is used in industrial automation and also forms the basis for other interfaces, such as Profibus.

In the driver stage shown in Fig. 16.11, resistor R11 assumes the function of line termination, which must be activated at the end of the bus. Its value is 120Ω and thus corresponds to the characteristic impedance of the twisted two-wire line.

In selected industrial applications, for example, a Siemens S7, this output stage is modified with resistors R9 and R10 to improve the levels. In this case $R11 = 220 \Omega$ and $R9 = R10 = 390 \Omega$ are used. The termination resistor is then also close to 120Ω .

16.2.5 IO-Link

The IO-Link interface arose from the need to parameterize sensors, such as pressure switches, via their *actual process interface*, without the need for an additional interface or control elements on the sensor. In this context, bidirectional communication with the control level was required, which is also suitable for communication with measuring sensors and actuators as a logical consequence.

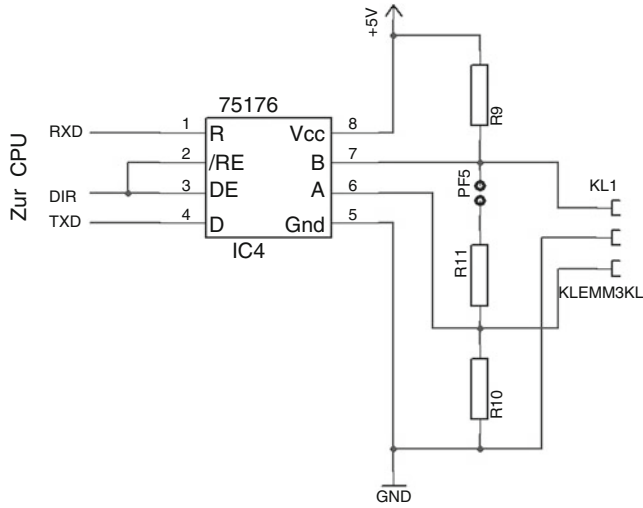


Fig. 16.11 Driver stage for an RS485 interface

The interface has the advantage in its dimensioning that it is based on the existing three-pin cabling of sensors. For actuators with separately disconnectable supply, it is the existing five-pole wiring. As already mentioned, bidirectional IO-Link communication is also ideally suited for transmitting digitized measured values. This enables a cost-effective changeover from analog to digital sensors. Like the supply, communication is fixed to the 24 V that is common in plant construction, which also leads to large signal-to-noise ratios due to the high level. An IO-Link sensor has at least the supply connections (pin 1 and pin 3) and a combined signal input/output (pin 4). The sensor can be operated in the so-called SIO mode, that is, the standard IO mode. In this case, pin 4 behaves like the switching output of a binary sensor. In IO-Link mode, pin 4 handles bidirectional communication with 3 different baud rates. The most common connection plugs are M12 connectors (Fig. 16.12), however, M8 and M5 connectors are also specified. For actuators with a supply voltage that can be switched off separately, pin 2 and pin 5 are used for this purpose.

The IO-Link is a *point-to-point connection*. Integration into the various bus systems is performed by IO-Link masters, which typically have several IO-Link ports. The IO-Link master is thus also simultaneously a gateway from the IO-Link communication to the respective bus system.

The IO-Link communication between IO-Link master and IO-Link device is based on the transmission format of a UART and consists of a start bit, 8 data bits and parity and stop bits. The baud rates 4800, 38,400 and 230,400 are specified for the speed.

When an IO-Link device is powered on, it always goes into standard IO mode (SIO) first. This ensures that it can always be operated as a standard binary sensor. If SIO is defined in the port configuration of the master for this port, the master leaves the device in SIO mode and only transmits the state of the switching output as 1 process data bit. If the

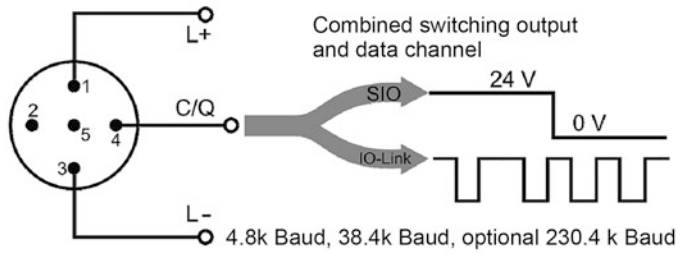


Fig. 16.12 Pin assignment of an IO-Link sensor with M12 connector (Source: IO-Link Konsortium)

IO-Link communication mode was specified in the port configuration of the master for this port, the master performs a wake-up procedure with the IO-Link device to set it into communication mode. Device data is then read out (e.g. device identification, diagnostic data) and the device is parameterized as required. Then the cyclic data exchange of process data is started. The master can reset the device to SIO mode at any time after parameterisation. This is useful for devices that have only one bit of process data, which can be easily transferred in SIO mode. If the communication is interrupted—for example, by disconnecting the sensor—the master tries again to establish a data connection by means of the wake-up procedure.

A comprehensive software system must be implemented in the sensors, which provides the necessary data structures and secures the communication rules. For the operation of an IO-Link device, a vendor ID is required, which is assigned by the IO-Link consortium.

16.2.6 Profibus

Profibus is widely used in the industrial sector because it allows very long cable lengths and allows the easy physical connection. It is technically based on the RS485 interface, to which several software levels according to the OSI model are superimposed to ensure network functionality (Fig. 16.13).

The transmission rate is typically in the range of 1 Mbit/s and can be up to 12 Mbit/s with reduced expansion. A maximum of 31 subscribers can be connected to each segment, whereby the network can extend over up to four segments in series. The maximum distance between two subscribers can be up to 1.2 km, the total extension of the network can reach over 9 km by means of repeaters.

The interests of Profibus users are represented internationally by the Profibus-Nutzer-Organisation (PNO).

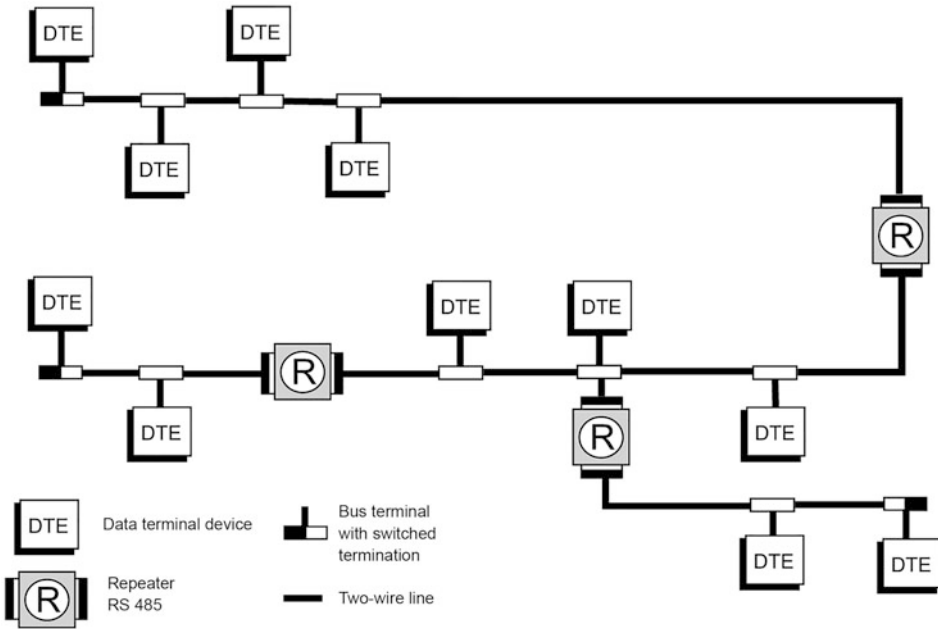


Fig. 16.13 Example of a Profibus network (Source: M. Felsler)

16.2.7 I²C

The I²C, also written as I2C, is a *serial, byte-oriented interface* that was developed to connect circuits on the board in the field of audio and video technology. It is a *synchronous two-wire interface* consisting of a data and a clock line. Both lines can be used bidirectionally and can be set up as a bus. The number of possible bus participants is limited to 128 by the defined address range of 7 bits.

Due to the synchronous data transmission, this interface does not make any demands on the stability of clocks, which makes it well suited for a pure software implementation. Many microcontrollers also offer this interface as a hardware unit. Since this bus works without driver circuits, its transmission length is limited to a few 10 cm.

With the I2C, data transmission is carried out using a simple hardware protocol with up to 400 kbit/s. The data set is framed by a start- and stop-signal game. The first byte after the start contains the target address and as 8th bit the direction information in the form of the read-write bit. The subsequent bytes contain the data to be transmitted, the structure of which is determined by the respective block. All transmitted bytes are acknowledged by an acknowledgement bit (ACK/NACK). In theory, each bus device can operate as a master or slave, but in practice, most components such as sensors, AD converters or DA converters are configured as a slave (Fig. 16.14).

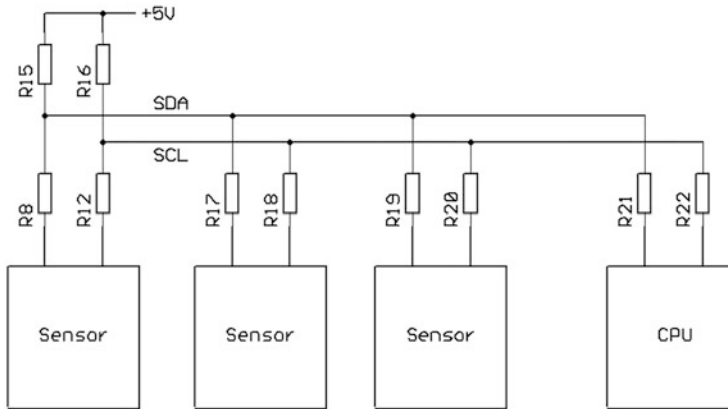


Fig. 16.14 Sensor system with I2C bus

Since the interface is uncomplex and power-saving, it is often used as a communication path in highly integrated sensor elements.

Since the output stages of an I2C interface are designed as open drains, the bus must have a working resistance (R15 and R16 in Fig. 16.14). The resistors at the sensors protect the output stages, especially for the time of the direction change on the bus. This is especially necessary if the sensor side is also represented by a CPU, which switches a full-fledged push-pull output stage to the bus in case of transmission.

16.2.8 SPI

The SPI interface was also developed for connecting peripheral components to a processor within a board. The starting point was the *fast data exchange* via coupled shift registers. This eliminates any overhead in hardware and software, thus achieving high data rates.

The data is transmitted synchronously via a clock and two data lines. The data set is synchronized by a fourth line as CE (chip enable) or SE (slave enable). With the two data lines, one leads from the master to the slave and the second back. The direction of the lines remains stable so that drivers can be used. In the case of a master-slave change, the lines are swapped internally. The clock is always generated by the master. Figure 16.15 shows the block diagram of an SPI interface.

The SPI interface can also be constructed as a bus if the output stages are designed accordingly. In this case, however, the master-slave structure must be defined. An example is the generic SPI described in Sect. 16.2.9.

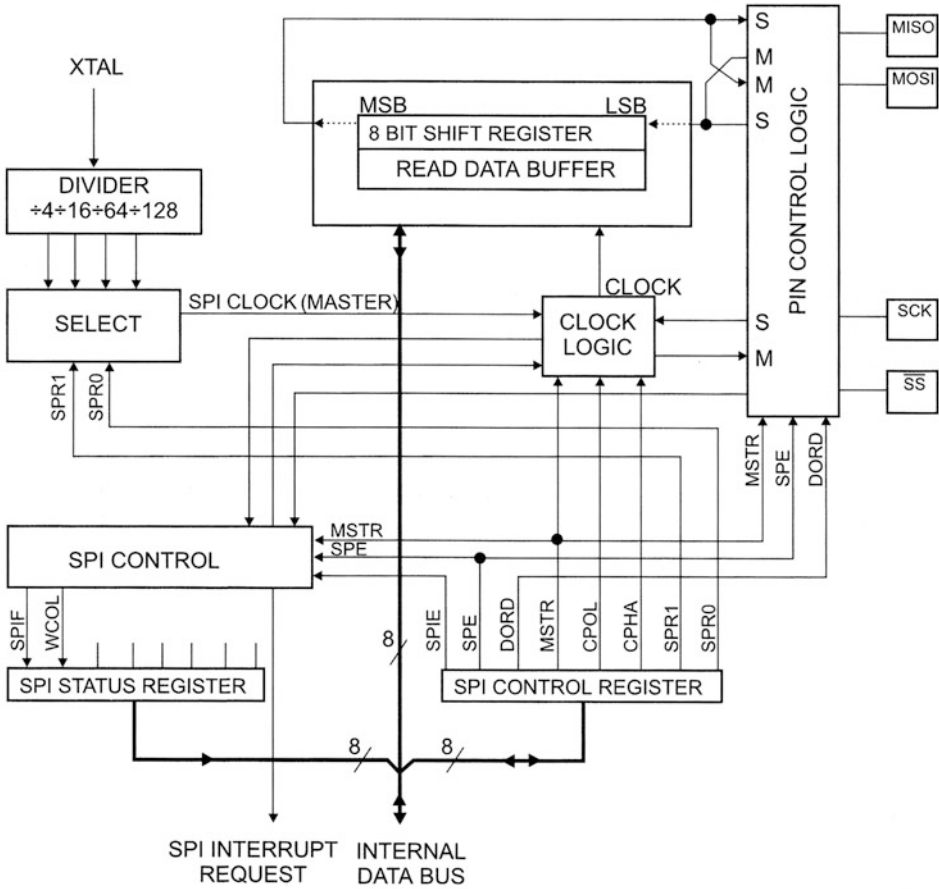


Fig. 16.15 Block diagram of an SPI interface in the microcontroller (Source: AT-mega128 datasheet from Atmel)

16.2.9 IEEE 1451

The IEEE 1451 is a family of standards that define the connection between a sensor or actuator and a network. The core is an electronic datasheet, the TEDS (*Transducer Electronic Data Sheet*), which is independent of the interface used. It consists of information on *identification, calibration data, measuring range* and manufacturer-specific data. The following seven substandards are currently distinguished:

IEEE 1451.0	commands and operations and the TEDS as basic communication between the transducer and a network node (NCAP—Network Capable Application Processor).
IEEE 1451.1	Description of the interface as objects for a client-server structure.

(continued)

IEEE 1451.2	Defines a sensor-NCAP interface for a point-to-point connection based on an extended SPI interface.
IEEE 1451.3	Like IEEE 1451.2, but as a bus.
IEEE 1451.4	Defines the TEDS and communication for analog sensors that do not contain a processor.
IEEE 1451.5	Describes the sensor-NCAP interface and TEDS, especially for the cooperation with wireless standards.
IEEE 1451.6	Describes the sensor-NCAP interface and TEDS for a special use in connection with high-speed CANopen.

The availability of information from the TEDS makes it possible to *manage* large stocks of sensors and actuators in a *network*. The possibilities here range from *function tests* and *parameterization* to *checking expiration dates* (e.g. when substances are consumed in gas sensors) and generating *maintenance plans*.

There is also the advantage that the standardized interface divides the production of the systems into two parts: *sensors* and *network nodes*. This simplifies inventory management for the user since he does not need every type of sensor for every type of network, and the manufacturer, since he only has to produce one type of interface. The following standards are typical for applications in sensor technology.

IEEE 1451.4

The core of the IEEE 1451.4 is a mostly passive measuring element, for example, a resistive bridge. A memory circuit is connected to this measuring element, which contains the data relevant to the measuring element. In order not to use any additional lines for this purpose, the switchover between measuring element and memory is done, for example, by reversing the polarity of the operating voltage. Here, the standard mainly defines the structure of the data in the memory and their transmission (Fig. 16.16). The memory size is defined as 256 bits only. It is used, for example, in connection with strain gauges.

IEEE 1451.2

With IEEE 1451.2 it is assumed that a controller structure is located on the sensor element. This allows data transmission via a protocol and is also capable of managing a somewhat more extensive TEDS. The sensor part is called the *STIM (Smart Transducer Interface Module)*. It communicates directly or via a bus with the *NCAP (Network Capable Application Processor)*, which converts the IEEE protocol to the desired target interface (Fig. 16.17).

The used interface of the STIM is based on an extended SPI interface. The additional four lines are used for bus arbitration, among other things. Both sensors and actuators can be connected to the bus.

IEEE 1451.4

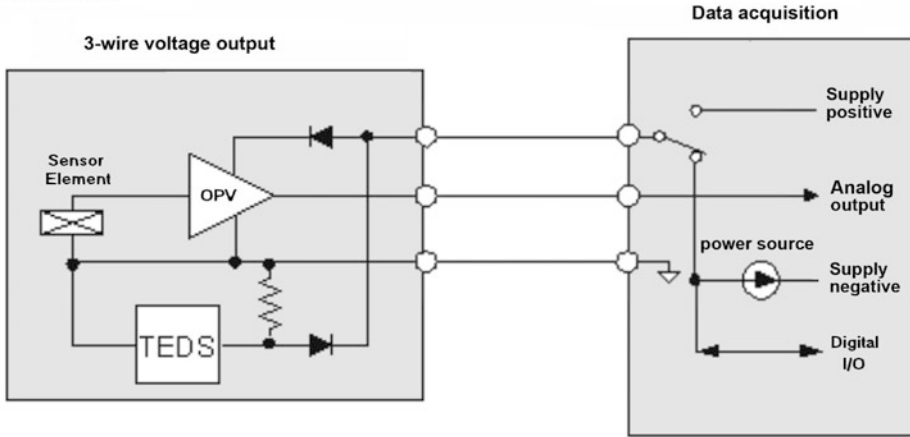


Fig. 16.16 Example of an IEEE 1451.4, Class 1, two-wire interface (Source: Steven Chan: “Update on the IEEE 1451 Smart Transducer Interface Standard”; AEPTEC Microsystems/3E Technologies)

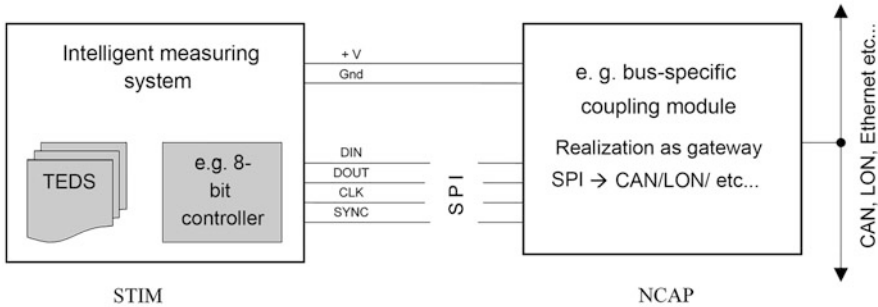
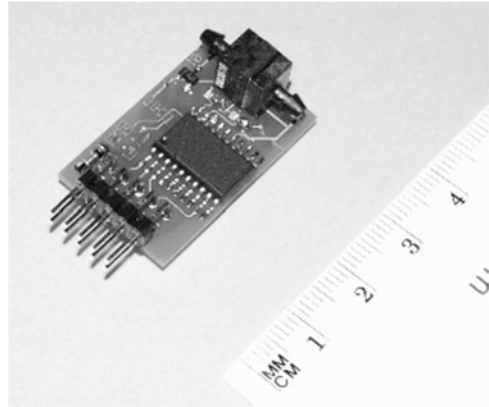


Fig. 16.17 Principle of a sensor structure for an IEEE 1451.2 interface (Source: Leitfaden generisches SPI-Interface)

The standard contains a command list, which serves as the basis for communication. This is used to access the measurement data and the objects of the TEDS. The TEDS is very detailed here and can assume the size of several gigabytes in the following seven objects

• Meta-TEDS	contains global data of the sensor,
• Address TEDS	contains address-specific data formats and time conditions,
• Meta-Ident-TEDS	contains information on the manufacturer and type designations,
• Address ID TEDS	contains address-specific manufacturer information,
• Calibration TEDS	contains calibration instruction(s),
• Calibration ID TEDS	contains a description of the calibration data,
• End-User TEDS	contains data fields defined by the manufacturer.

Fig. 16.18 gSPI-capable pressure sensor with 8-bit controller (memory 2 kB). The controller contains the interface and a characteristic curve correction (Source: Prignitz Mikrosystemtechnik GmbH)



Several measurement channels can be implemented in one STIM, which then leads to multiple occurrences of individual TEDS objects.

Generic SPI (gSPI)

The gSPI is a reduced form of the IEEE 1451.2. Figure 16.18 shows a pressure sensor with gSPI capability. This implementation can manage with minimal resources and is still compatible with the standard. The required reduction was achieved by limiting the size of the TEDS objects to a maximum of 252 bytes each. This is sufficient in practice since most of the volume in TEDS is created by descriptive texts, which can also be stored in several languages. A practical implementation then requires about 370 bytes for a single-channel sensor, with a second channel the value increases to 480 bytes.

Reference:

This chapter was taken from the book “Messtechnik mit dem ATmega” (ISBN: 978-37723-5927-9). Published by FRANZIS Verlag GmbH, 85540 Haar near Munich.

Bibliography

1. Bittkow, Michael: „Wer ist der beste Bus?“. In: www.elektronik-kompodium.de
2. Drunk, Gerhard; Feinäugle, Albert: „IO-Link – die USB-Schnittstelle für die Sensor-Aktor-Ebene“; über: www.xpertgate.de/magazin/report/Branchenreport-IO-Link.pdf
3. Felser, Max: PROFIBUS Kompetenzzentrum (PICC); Berner Fachhochschule; Juni 2007; <http://www.profibus.felser.ch/einfuehrung/profidp.pdf>
4. Informationen zum IEEE 1451; über: <http://ieee1451.nist.gov/>
5. IO-Link-Konsortium: <http://www.io-link.com/de/index.php>
6. Leitfaden „Generisches SPI-Interface“, © AMA Fachverband für Sensorik e.V., April 2004; über <http://www.ama-sensorik.de/download/AMA%20Leitfaden.pdf>
7. Schäppi, Urs: Das HART Protokoll, Berner Fachhochschule Hochschule für Technik und Informatik HTI; über: <https://prof.hti.bfh.ch/uploads/media/HART.pdf>



Gert Schönfelder and Sorin Fericean

17.1 Features for Function Monitoring

The decentralization of data processing requires the transfer and increase of intelligence on the lowest level, the sensor and actuator level. This requires additional functions for parameterization and diagnosis of the systems from a central location.

Regardless of the level of intelligence, sensors for industrial applications must contain certain protective functions. Here, the proximity switch (PS) is to serve as a technical example. However, these protective functions apply in principle to all types of sensors, regardless of their functionality and the type of output signal.

Main protection functions

The basic function of a sensor must be to ensure its own “survival”. This is independent of its functionality. The most important main protective functions for sensors are:

- a) The *short-circuit protection function* protects the sensor in case of a short circuit in the output circuit. In practice, there are two types of such protective functions. The simplest version contains an NTC thermistor in the PS output circuit. The short-circuit load current rapidly increases the temperature of the NTC semiconductor. Its resistance increases and limits the load current to harmless values. The disadvantage is the resistance of the NTC in the load circuit.

G. Schönfelder (✉)
Prignitz Mikrosystemtechnik, Wittenberge, Germany
S. Fericean
Balluff GmbH, Neuhausen a.d.F., Germany

The mostly implemented solution shows an astable behaviour. In the case of a short circuit, the output is switched off immediately. After a certain waiting period (less than 100 ms), the switching output is switched on again for a short test period (at 100 μ s) to determine whether the short-circuit case still exists. This time is short enough for the output to withstand the interference. If the short-circuit has disappeared or been removed, the output immediately resumes its normal function.

- b) The *wire breakage protection function* has the task of preventing faulty sensor function in the event of the undesired interruption of one or more sensor supply lines. In such cases of emergency, parasitic supply paths and thus possible incorrect switching of the load must be prevented.

A wire break in the positive supply line or the load supply line at the positive-switching output or the negative supply line or in the load supply line at the negative-switching output leads to an intrinsic protective function: the sensor or the load is de-energized (Fig. 17.1).

On the contrary, if the negative supply line is interrupted at the positive switching output, the sensor could be supplied parasitically via the two intact supply lines and control the load incorrectly. The same applies adequately to the negative switching output. This danger is eliminated by implementing a wire breakage protection function.

- c) The *reverse polarity protection function* prevents the sensor from being disturbed or destroyed as a result of connections to the supply source and/or load being reversed.

For a system with three connections, there are $3! = 6$ connection possibilities. These are shown for the positive switching output in Fig. 17.2. In addition to the correct polarity, five single or double reverse polarity cases can be recognized. Adequate protection ensures protection against sensor destruction (minimum requirement) and—if possible—also negligible reverse polarity currents.

Diagnostic Information

The concept of intelligent sensors requires the following three diagnostic functions:

- a) *Monitoring* the complete functionality of the sensor. Not only the status of the sensor output is reported, but also the plausibility of the status.
- b) *The message of an occurred sensor failure*. Sensor failure should be signalled. It must be ensured that the displayed output status is not caused by a defective sensor.
- c) *Forewarning function*. The reduction of the performance reserve, for example, due to contamination or misalignment of a sensor, is intended to generate a warning message at the control level.

The first two levels describe a minimum requirement for *intelligent sensors with diagnostic function*. The additional implementation of the advance warning function is

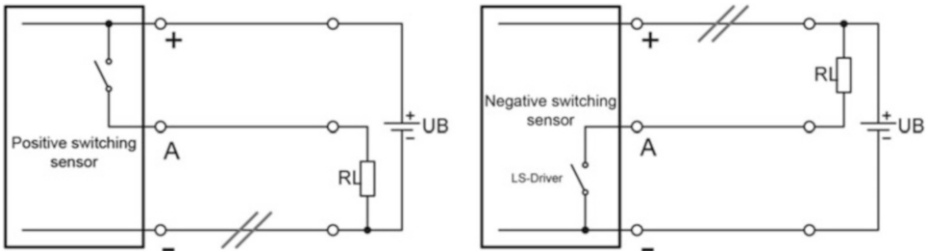
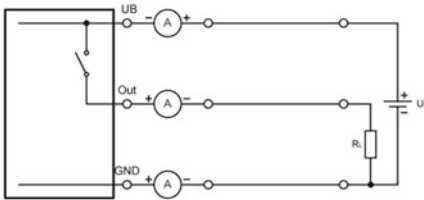
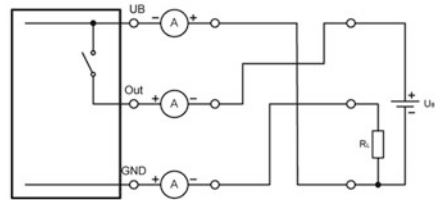


Fig. 17.1 Wire breakage protection functions

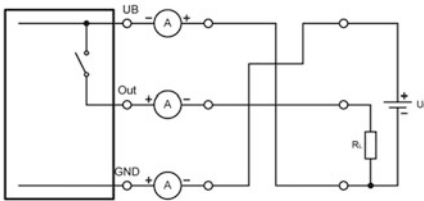
0. Correct polarity



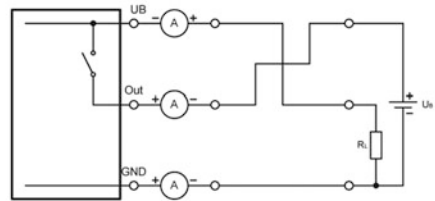
3. Multiple polarity reversal



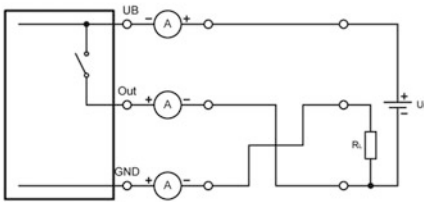
1. Simple polarity reversal



4. Simple polarity reversal



2. Simple polarity reversal



5. Multiple polarity reversal

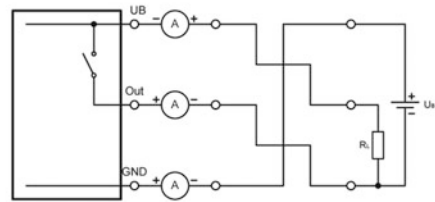


Fig. 17.2 Illustration of the correct or reverse-polarity connections of a positive-switching switch

increasingly demanded. The circuit implementation of these functions in modern sensor electronics is unproblematic nowadays.

There are basically two possibilities for the output of the diagnostic information:

1. Use of an *extra output* on the sensor. The results of the diagnostic functions are coded and output by static or dynamic signals. The versions are easy to implement and evaluate. The only disadvantage is the presence of the second diagnostic output and the additional input required on the control side.
2. *Multiple function assignment* of the only conventional sensor connection. This solution does not require any addition or deviation from the standard output version (only three leads as in Fig. 17.1) and is intended to enable both standard applications and applications with additional diagnostic messages.

A well-known diagnostic concept is implemented by the *Balluff basic diagnostic version*, which meets the first two diagnostic requirements in an intelligent patent-protected manner: Monitoring and reporting sensor failure. As can be seen from its connection diagram (Fig. 17.3), the proximity switch (PS) has three standard connections. Its output signal with modulated pulses is specific (Fig. 17.4). Like a classic PS, the sensor provides an output signal with two levels. New is the presence of short pulses with 300 μs pulse duration and 6.6 ms repetition period, superimposed by the static output signal. These pulses have the opposite level. They are generated by an internal artificial actuation simulation in the PS sensor element, propagate through the sensor and its output leads and can be evaluated in the application. The pulses are present as long as the LV functions without error and are extinguished in the event of an error.

The sensor supports two possible circuits at its output:

- A *conventional consumer*, such as a relay, can be connected directly to the output (resistance symbol R_L in Fig. 17.3). The timing of the superimposed pulses has been set so that all relays do not react to these short pulses; the relay only evaluates the *static levels*.
- If you want to apply the *functional diagnosis*, an evaluation device is inserted between the PS and the consumer (Fig. 17.3). The device evaluates the complex output signal of the PS and supplies two static signals: a *sensor output signal* which corresponds to the sensor actuation state and a *diagnostic output signal* whose state signals the presence/absence of the superimposed pulses and thus supplies the information sensor functional or faulty.

A higher-value diagnostic concept is the *Balluff DSC* (Dynamic Sensor Control) *concept*, which includes a full version of diagnostics. As shown in Fig. 17.3, the DSC-capable PS still only has three standard connections. Furthermore, its output signal has modulated pulses with different repetition periods. As additional information, more detailed error information is available, which is coded by repeat periods.

Fig. 17.3 Wiring diagram of the PS with integrated functional diagnostics with superimposed pulses

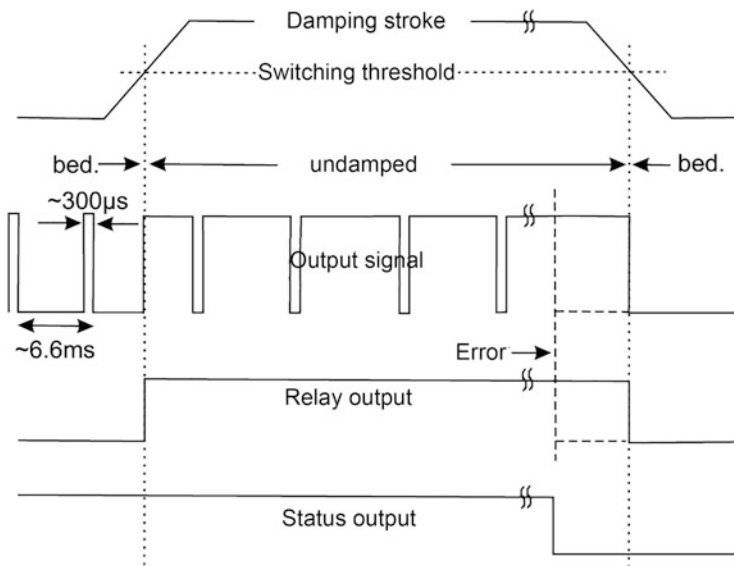
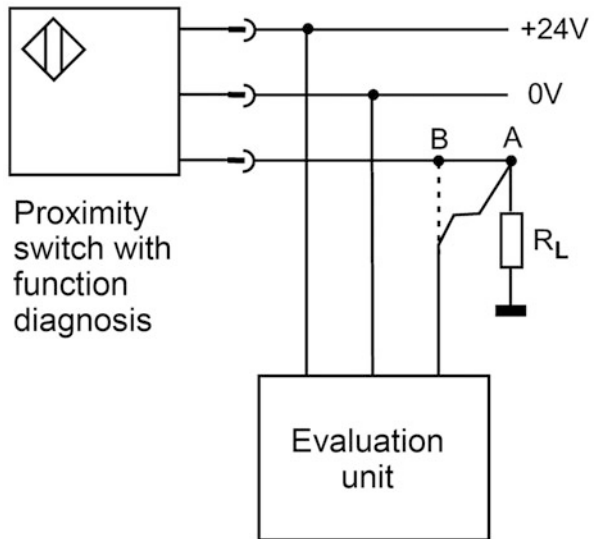


Fig. 17.4 Pulse diagrams of a proximity switch with functional diagnosis

17.2 Electromagnetic Compatibility (EMC)

In connection with electronics that are becoming increasingly complex and sensitive, the question of the influence of external interference arises. A sensor should deliver an undisturbed measured value in an increasingly disturbed environment. This includes intentional disturbance variables such as those caused by radio operation, but also unintentional ones that can arise between switching processes and cosmic radiation.

Based on these basic ideas, EMC includes two question directions:

1. *Securing and testing* a sensor against the effect of external disturbances of electromagnetic and electrostatic origin (immision) and
2. The *prevention and testing of* a sensor against the emission of such fields (emission).

The problems of EMC are regulated by a comprehensive body of standards. It must be defined by the sensor element under consideration and the concrete requirements in use, which of these standards are to be applied in practice.

In the following section, some selected relationships are shown using the example of proximity switches.

The standard specifications or the typical current time values (shown in bold) of these parameters for sensors can be found in Table 17.1.

Opening credits

The proof of electromagnetic compatibility in the accredited EMC test laboratory includes the following tests:

- Immunity to *discharge of static electricity* (ESD = Electrostatic Discharge),
 - Immunity to *electromagnetic fields* (RFI = Radio Frequency Interference),
 - Immunity to *fast, transient pulses* (Burst, EFT = Electrical Fast Transient),
 - Immunity to *surge voltages* (Surge),
 - Immunity to *conducted interference induced by high-frequency fields* and
 - Immunity to *magnetic fields with power engineering frequencies*.
- The test conditions with regard to actuation, supply or installation are specified very precisely. Destruction of the sensor is not permitted; the following three acceptance criteria (ANK) apply to the overall behaviour of the sensor:
 1. *Criterion A*: During the test, no false pulses longer than 0.1 ms may occur at the sensor output.
 2. *Criterion B*: During the test, no false pulses exceeding 1 ms may occur at the sensor output.

Table 17.1 Summary of the most important EMC requirements and test conditions (standard requirements for PS—see Sect. 3.4.1—are marked with bold and underlined)

Immunity to interference:	Origin, characteristics	Applicable standard	Test setup or test signal characteristics and test frequency	Standard test levels and requirements for NS
Discharge of static electricity (ESD)	Discharge occurs when people or objects become charged and discharge on contact with conductive materials	IEC 61000-4-2	<ul style="list-style-type: none"> – Short and one-off discharge in the ns range with voltages up to 30 kV – 10 single pulses at intervals of 1 s – Directly by contact or air discharge or indirectly by a coupling plate 	Contact/ Air discharge: 1: 2 kV/2 kV 2: 4 kV/4 kV 3: 6 kV/8 kV 4: 8 kV/15 kV <u>ANK: B</u>
Electromagnetic fields (RFI)	Electromagnetic radiation from radio transmitters, mobile phones, radar systems and also devices in close range	IEC 61000-4-3	<ul style="list-style-type: none"> – Narrow frequency band tuned up to 2.7 GHz, often directional energy, frequency or amplitude modulated – Test intensity ≤ 100 V/m – AM modulation: 1 kHz, sinusoidal, 80 	1: 1 V/m 2: 3 V/m 3: 10 V/m <u>ANK: A</u>
Fast, transient pulses (burst)	Switching inductive loads (e.g. relays, motors) on and off. Low-energy interference with broadband spectrum up to 300 MHz and amplitudes up to several kV.	IEC 61000-4-4	<ul style="list-style-type: none"> – Triangular pulses with 50 ns width – 5 kHz repetition rate of $\Rightarrow 75$ pulses/packet – Packet width: 15 ms – Repetition period of the packets: 300 ms – Test voltage capacitively coupled between grounded NS housing and all lines 	1: 0.5 kV 2: 1.0 kV 3: 2.0 kV 4: 4.0 kV <u>ANK: B</u>
Surge voltages (Surge)	Lightning, short circuits and switching operations in high-energy networks. High energies, voltages and	IEC 61000-4-5	<ul style="list-style-type: none"> – Triangular pulses with 50 μs width – Single pulses with positive and negative polarity and alternating pulse sequences 	1: 0.5 kV 2: 1.0 kV 3: 2.0 kV 4: 4.0 kV ANK: no destruction

(continued)

Table 17.1 (continued)

Immunity to interference:	Origin, characteristics	Applicable standard	Test setup or test signal characteristics and test frequency	Standard test levels and <u>requirements for NS</u>
	currents. Rise times in the μs range, broadband spectrum up to several MHz.		– Test voltage applied between two respective LV terminals – Generator internal resistance 500 Ω	
Conducted disturbances induced by high-frequency fields (LGS)	Interference impulses that are caused by electromagnetic fields being scattered via cables. Narrow frequency band, frequency or amplitude modulated, continuous interference.	IEC 61000-4-6	– Frequency range: 150 kHz to 80 MHz, in 1% steps – AM modulation: 1 kHz, sinusoidal, 80 – Interference coupled in via a coupling/decoupling network	1: 1Vrms 2: 3Vrms 3: 10Vrms <u>ANK: A</u>
Magnetic fields with power engineering frequencies	Created by currents in power connections, busbars, equipment with high power. Constant magnetic fields up to 100 A/m or short-term magnetic fields up to 1000 A/m	IEC 61000-4-8	– Field generated by an induction coil square 1 m \times 1 m – Frequency 50 Hz	1: 1 A/m, continuous 2: 3 A/m, continuous 3: 10A/m, continuous 4: 30A/m, continuous 5: 100A/m, continuous <u>ANK: A</u>

3. *Criterion C*: if the switching state changes for more than 1 ms or a loss of behaviour requires a sensor reset.

The test conditions given here are examples of a concrete application. For example, a generator internal resistance of 500 Ω is specified for the surge test. In other applications, this can also be only 40 Ω or 4 Ω , which requires other protective circuits. The coordination of the design and testing must therefore always be carried out in connection with the intended application.

An example of an EMC circuit is shown in Fig. 17.5, where the simple case of a protective circuit for a two-wire current transmitter is shown.

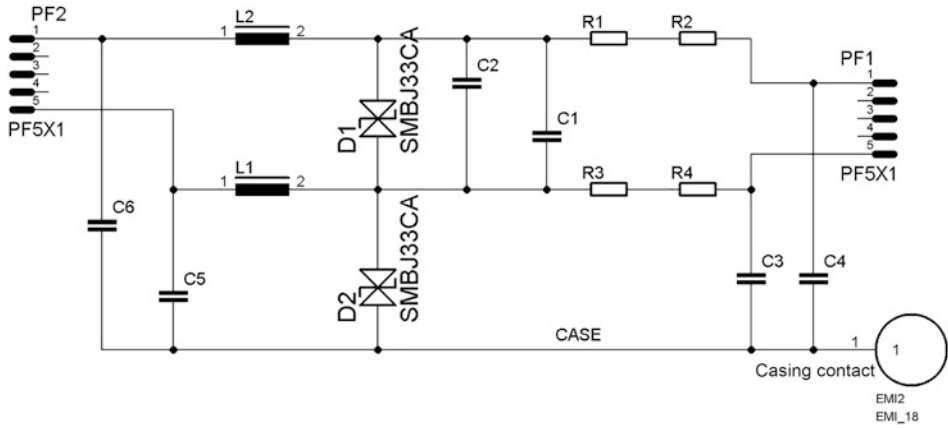


Fig. 17.5 EMC protection circuit for a two-wire current transmitter

On the right side of the picture, the supply is connected—on the left side the electronics of the transmitter. The resistors in the input limit the pulse current and, together with the capacitors, also form a low-pass filter, which also bridges the response time of the suppressor diodes. Diode D1 must absorb the entire interference energy. The then almost unloaded path to the transmitter is filtered by a further low-pass filter.

Here in the example, diode D2 has the task of providing the function of D1 for the insulation to the housing. If not only 30 V but 500 V are required as insulation voltage, there will be extensive changes in the circuit, because the capacitors C3 to C6 must be designed for these voltages. In addition, a very high-impedance discharge circuit must be provided so that these capacitors do not remain charged to the interference voltage.

The entire protective circuitry quickly outperforms the actual sensor in terms of complexity.

17.3 Functional Safety (SIL)

Functional safety is about *reducing risks* in the use of technical equipment. The aim here is to assess the existing risk in the operation of a facility and subsequently reduce it to an acceptable level. A risk is defined as:

$$\text{Risk} = \text{probability of a hazardous event occurring} * \text{costs caused by this event}$$

The functional safety (SIL = Safety Integrity Level) comprises several sets of standards, for example

DIN EN IEC 61508	as a basic standard,
DIN EN IEC 61511	as a specific standard for the process industry,
DIN EN IEC 62061	as a specific standard on machine safety,
DIN EN IEC 61513	as a specific standard for the nuclear industry.

It shall be applied to safety-relevant systems if these

- an electrical device,
- an electronic device, or
- a programmable electronic device

contained. It is therefore applicable for sensors in the safety area. The considerations are always necessary for the entire system, that is, from the sensor to the signal processing to the actuator.

The SIL levels describe *different failure probabilities* for a system. In the simplest case, this ranges from a *safety-oriented design* to a *multiple redundant system* for the highest requirements. A safety-oriented design starts with the adherence to component parameters, reliability tests of components and technology and production traceability.

The failure probabilities assigned to the SIL levels are additionally differentiated between high and low requirements. Here, “low” means that the safety device does not become active more than once a year. The case of high requirements means that the safety device is active continuously or more often than once a year.

Figure 17.6 shows an example of a risk graph for determining the required SIL level. It must be adapted to the specific situation.

The following assignments apply:

- C Extent of damage
 - C1 Slight injury to a person or minor damage to the environment.
 - C2 Serious, irreversible injury to one or more persons or death of a person or temporary major adverse environmental effects.
 - C3 Death of several persons or long-term major adverse environmental effects.
 - C4 Catastrophic environmental damage, very many dead.
- F Duration of stay
 - F1 Rarely to often,
 - F2 Frequently to continuously.
- P Danger prevention
 - P1 Possible under certain conditions,
 - P2 Hardly possible.
- W Probability of occurrence
 - W1 Very low,
 - W2 Low,
 - W3 Relatively high.

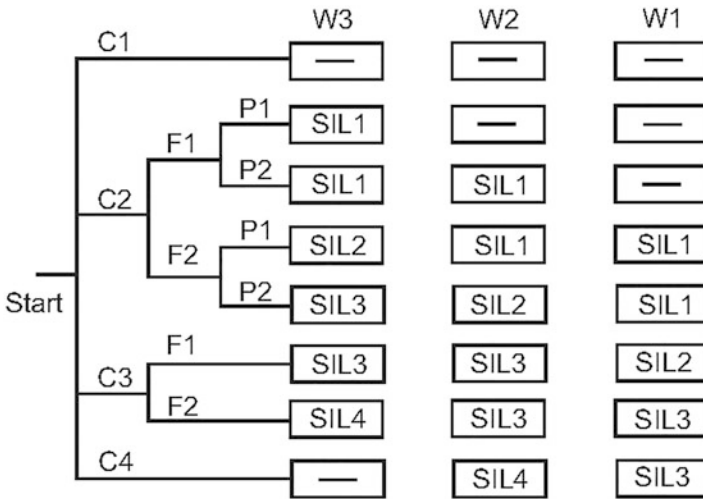


Fig. 17.6 Risk graph according to IEC 61508/61511

Technical requirements for the sensor then result from the specific safety level. The technical product will be certified by an approved body after completion of the procedure. Approval for the hazardous area (ATEX) can also be part of a SIL approval.

17.4 Sensors in Explosive Environments (ATEX)

If sensors are used in an *explosive*, that is, an *ignitable* environment, special regulations apply. At this point, there is a large overlap with the requirements of SIL. However, it is not only a matter of minimizing the risk of injury to persons and equipment but also of eliminating it. This can be achieved both by structural engineering measures and by appropriately designed electronic means.

There are two fundamentally different approaches to solving the task in explosion protection:

1. *The dimensioning of the assembly* shall ensure that no spark can occur or that no component heats up so that a surrounding mixture explodes, or
2. if sparking or a hotspot cannot be technically avoided (e.g. in circuit breakers), then the sensor must be *encapsulated* in such a way that this ignition source has no effect on the environment. Localized explosions are also conceivable.

The following section gives a brief introduction to this issue. Here too, the requirements are described in detail in a comprehensive body of standards, which have now been harmonised at the European level. Corresponding standards can be found for example in

Table 17.2 Assignment of device group and category

Device group	Equipment category	Probability of the EX atmosphere	Degree of safety to be ensured	Sufficient security for	Zone classification
Group II All sectors except mining	1	Permanent, long-term or frequent presence	Particularly high	Protection against two independent faults	Zone 0 Zone 20
	2	Occasionally available	High	Avoidance of ignition sources during operation and malfunctions	Zone 1 Zone 21
	3	Only rarely and then briefly present	Normal	Normal operation	Zone 2 Zone 22
Group I Mining (especially firedamp and dust)	M1	Available	Particularly high	Protection against two independent faults	
	M2	Upon the occurrence of EX atmosphere can be switched off	High	Ignition sources ineffective even under difficult conditions	

UL (USA) and CSA (Canada). European approval is also recognised in many non-European countries or can serve as the basis for a national approval there.

17.4.1 Basic Principles of ATEX

The extensive standards and accompanying regulations within the framework of the EC directives on explosion protection and their implementation in European standards cannot be dealt with in this work. However, a few basic terms are conveyed.

The first step is to classify the operational environment according to the frequency and type of occurrence of the hazard. This results in a *device group* and a *device category as shown* in Table 17.2.

The basic standard is *DIN EN 60079-0*, which has been valid since 01.12.2004. Several substandards are based on this standard, which define the different types of protection.

Risk classification is thus made, from which a catalogue of design measures and electrical dimensioning is derived. Within the category, a further distinction is made between *Zone x* and *Zone 2x*. Zone x applies to air, gases and vapours and zone 2x to dusts. Different test requirements are derived from this distinction since dust can deposit on

a transmitter, for example, and thus cause short circuits or sources of fire, whereas a gas mixture can lead directly to an explosion.

The second step is the assignment to an *ignition protection type*. This already describes the path taken for protection. The standards differentiate between 11 different types of protection, whereby protection through *intrinsic safety* and *flameproof encapsulation* are of particular importance for sensor technology. Both types of protection are dealt with in more detail in the next section.

The third part is the classification of *thermal hazard*. Temperature classes are defined which specify a temperature limit that the assembly must not exceed even in the event of a fault. Together with the three explosion subgroups, this results in a matrix of substances that may occur in the environment. For sensors, the classes T4, T5 and T6, which cover a range up to 200 °C surface temperature, are of importance. In general, T4 is permitted because there is only one substance (carbon disulphide) which requires T6. However, attention must be paid to whether a sensor principle generates higher temperatures internally, as is occasionally the case with gas or flow sensors.

The marking of an assembly for use in potentially explosive atmospheres is also made up of these discussed parts:

EX 2G ia IIC T4 meaning

EX	An ATEX component. Formerly also EEX with the first “E” for European, which is no longer necessary as there are no longer any national variants.
2G	Device category 2 for gases
ia	Protection class intrinsically safe with guaranteed protection for two independent faults. With ‘ib’ the protection only exists after one fault.
II	Device group II
C	Explosion subgroup C for gases
T4	Temperature group T4, which corresponds to a maximum surface temperature of 135 °C to 200 °C

The testing and confirmation of the applicability of an assembly for use in EX areas is carried out by an authorised institution. This institution issues a test number and corresponding certificates. The manufacture of EX modules requires an additional mandatory QA approval, which is based on the ISO 9001 certification.

17.4.2 Ignition Protection Type Intrinsic Safety

The type of protection “intrinsically safe” is defined in the standard *DIN EN 60079-11* (formerly DIN EN 50020). It is based on the *avoidance of spark-forming energy quantities and voltages in the sensor*, as well as a *limitation of the occurring power losses*. It is often found in the area of sensors, as the limited amounts of energy are sufficient for them.

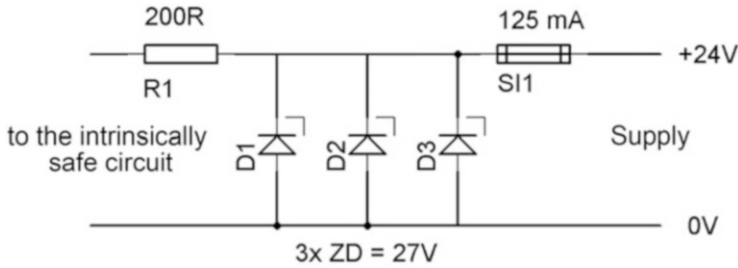


Fig. 17.7 Circuit of a Zener barrier

An intrinsically safe module always occurs in interaction with assigned equipment. The intrinsic safety of the sensor alone is not sufficient—secure data transmission and power supply to the circuit is required.

An intrinsically safe circuit is supplied via a *supply isolator*, which isolates the data circuit and limits the energy that may occur in the circuit. The simplest case is the use of a *two-wire current transmitter* which modulates the measured value to the operating signal in the range of 4–20 mA. There are therefore no separate data lines that have to be secured.

The supply thus consists only of Zener diodes and resistors, as shown in Fig. 17.7. It is also referred to as a *Zener barrier*. The three Zener diodes suppress short-term overvoltages and destroy the fuse in the event of longer faults. The parallel connection is used for failure redundancy if a diode reacts to overload with an interruption. The series resistance limits the maximum current flow in the system.

With the specified dimensioning of the Zener diodes, a maximum of 27 V can occur. The resistor of 200 Ω limits the output current to 125 mA in the case of a short circuit. The components used may only utilise a maximum of 50 % of their permissible power loss.

This arrangement is a supply with linear characteristic.

On the side of the sensor, it must also be ensured that possible component failures do not lead to impermissible states. This is supplemented by the requirement for 500 VAC isolation between the sensor signal and housing.

In addition, an insulation distance of at least 0.7 mm between all (critical) live parts must be ensured for the electronics themselves. Since this is not possible with modern semiconductor components, these are to be regarded as short circuits in the circuit in principle.

A second point to limit the ignitable energies is done by avoiding energy storage devices such as inductors and capacitors. These are subject to severe restrictions when used.

Figure 17.8 shows an extract of the output stage of a transmitter with 4–20 mA.

The most critical current path (bold) is the one at a short circuit of T2 or the OPV. In this case, the path only contains resistors R1 and R15. The total current in the circuit is then determined by the maximum 27 V supply, the 200 Ω in the Zener barrier and the approximately 300 Ω in the sensor. In addition, 100 Ω are also included as measuring resistors for decoupling the measured value. This results in a short-circuit current of 45 mA and a proportionate power loss at the sensor of about 610 mW.

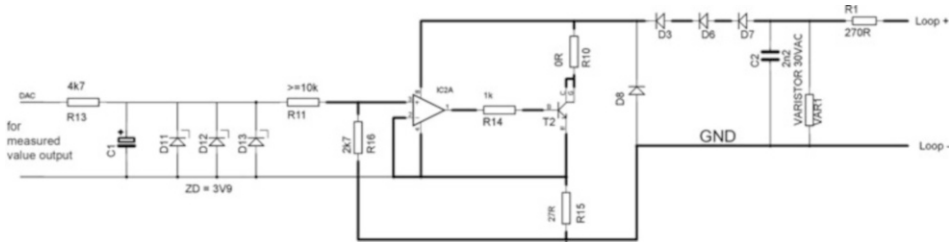


Fig. 17.8 Circuit of an output stage for current loop

The filter capacitor C1 must not be allowed to charge to a dangerous voltage if the OPV is defective. For this reason, it must also be separately limited by means of a Zener barrier. With the 3.9 V selected here, the capacity can be in the μF range. In the normal case at 27 V, it may only be in the range of 50 nF, whereby the connecting cables must be included in this calculation.

17.4.3 Type of Protection Flameproof Enclosure

The type of protection “flameproof enclosure” is defined in the standard *DIN EN 60079-1* (formerly DIN EN 50018). The alternative approach is to *avoid* an *ignition source*—and should it occur, this has no effect on the environment.

With flameproof encapsulation, it is assumed that the *electronics* are housed in a *sealed container*. This container must not have any exchange of gases inside with the environment or only an exchange of gases with the environment due to narrow gaps.

If an explosive mixture should nevertheless accumulate inside, the container must be safe enough to ensure that it is not destroyed by the explosion itself and that no flames or hot gases escape.

The properties of the electronics are largely subordinate to this. However, it may be necessary to ensure that no sparks are generated when connecting or disconnecting cables.

This type of protection is well suited for sensors since the electronics can be freely designed and the protection in case of an explosion is easy to handle with the usually small volumes.

Bibliography

1. BARTEC GmbH, Bad Mergentheim, <http://www.bartec.de>
2. Endress+Hauser, www.de.endress.com/SIL
3. Fritsch, A.: „Funktionale Sicherheit und Explosionsschutz; Firmenschrift der Firma Stahl; 2005
4. Leuze Electronic: „Ex-Schutz Grundlagen“, http://www.leuze.de/news/atex/p_05_de.html
5. Pepperl + Fuchs: „Handbuch Exschutz“, Firmenschrift 2007, www.pepperl-fuchs.com



Measurement Errors, Measurement Accuracy and Measurement Parameters

18

Gert Schönfelder

The measurement of physical, chemical and biological quantities by sensors is subject to deviations. The *deviation* d is the difference between a measured value x_i and a *true value* x_0 . This deviation represents an error. It applies:

$$d = x_i - x_0.$$

The true value can either be a *normal* or a *probable value* determined by a large number of measurements. Standards can be predefined measured quantities or they can refer to the basic physical quantities for time (s), length (m), mass (kg), electric current (A), temperature (K), luminous intensity (cd) or the quantity of substance (mol). All physical measurands can be traced back to these basic units of the SI measurement system (SI: Standard International).

A *probable value* is determined by *statistical* methods. Under certain conditions, this is the *arithmetic mean value* (Sect. 18.2).

18.1 Classification of Measurement Errors According to Their Cause

If the measurement errors are classified according to their cause, there are four basic types:

- the *systematic* measurement error,
- the *random* measurement error,
- the *gross* measurement error and

G. Schönfelder (✉)
Prignitz Mikrosystemtechnik, Wittenberge, Germany

- the *methodical* measurement error.

The *systematic measurement error* has its cause in the measuring circuit, the measuring method, the properties of the reference measuring point and the properties of the measuring components used. A *systematic* measurement error always shows *deviations in one direction* (too little or too much; the Gaussian distribution does not apply). It gives reproducible results under the same conditions. It can therefore be technically taken into account in the circuit or signal processing and can therefore *be corrected*.

The classic example here is the voltage measurement at a source with an internal resistance greater than zero and a measuring instrument with an internal resistance less than infinite. However, the thermal properties of the components used are just as important.

This problem is often neglected since today's measuring instruments have very high input resistances. In the field of sensors, however, very practical quantities are dealt with. For example, a measuring bridge has an internal resistance of several k Ω and the AD converter as a measuring device has a few hundred k Ω input resistance. This constellation quickly results in systematic errors in the range of percent, which must be taken into account.

The *random measurement error* occurs during a series of measurements and shows itself in fluctuating measured values. A *Gaussian distribution* is present here (Sect. 18.2.1; Fig. 18.1). The cause can be statistical processes in components or the environment (e.g. radiation or temperature fluctuations). The cause of random errors can also be changing measurement conditions, the experimenter or noise.

Random measurement errors can be determined and reduced with statistical methods. If noise is determined as the cause, circuit measures such as filters or low-noise components can also contribute to the reduction.

The *gross measurement error* is shown in the form of individual *outliers* in a measurement series. These can be caused by external events, such as switching operations in the environment. They are best countered by *filtering* the measured values, whereby the *median filter* is particularly suitable here.

The *methodical measurement error* is one of the systematic errors in terms of its effect but has a random component. In contrast to this, it cannot be corrected easily. The methodical measurement error occurs mainly in the field of digital technology. In addition to the *quantization error*, it also includes *errors in the transmission* of the measured values. The best-known representative is the *digital residual error* (Sect. 5.2). It can be defined exactly, but its occurrence is subject to random correlations. Elimination is only possible at an unreasonably high cost.

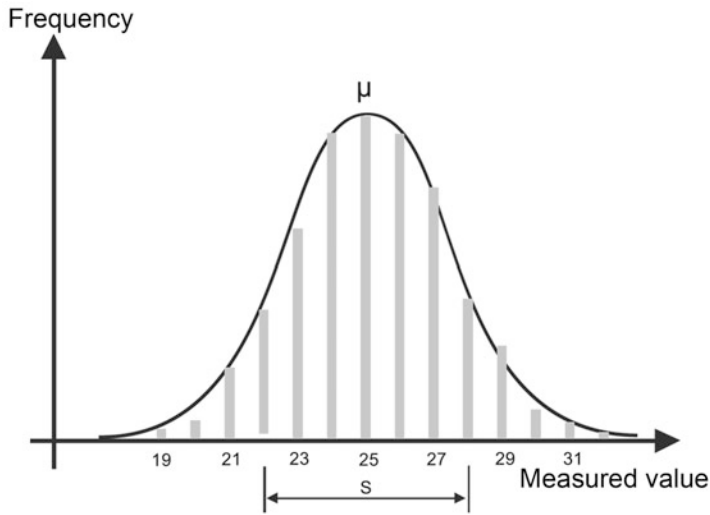


Fig. 18.1 Gaussian bell-shaped curve of the frequency distribution of measured values

18.2 Display of Measurement Errors

18.2.1 Arithmetic Mean, Error Sum and Standard Deviation

The following relationships apply to random errors. If the measured values x_i are graphically evaluated in a *histogram* according to their *frequency*, the frequency distribution is the *Gaussian distribution* (Gaussian bell curve; Fig. 18.1). The frequencies are symmetrical to a value measured most frequently, the so-called *expected value* μ .

The expected value μ as the most frequent measured value is determined by the *arithmetic to mean value*. It applies:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

with N : number of measurements and x_i : the measured value of measurement i .

Error is the deviation d of the current measured value x_i from the arithmetic mean value \bar{x} . It applies: $d = x_i - \bar{x}$.

The *error sum* FS is calculated as the sum of the *quadratic deviations* from a fixed value x_0 : $FS = (x_i - x_0)^2$. If the arithmetic mean is used \bar{x} for the fixed value x_0 , the error sum is *minimal* (FS_{\min}). It applies:

$$FS_{\min} = \sum_{i=1}^N (x_i - \bar{x})^2.$$

The *standard deviation* s is a measure for the width of the Gaussian bell curve. It is calculated:

$$s = \sqrt{\frac{FS_{\min}}{N-1}}.$$

At a distance s from the arithmetic mean value (Fig. 18.1) there are about 2/3 of all measured values. This means that a *narrow* bell-shaped curve with a smaller standard deviation has *more accurate* measured values than a *wide* bell-shaped curve with a larger standard deviation.

18.2.2 Absolute Error

For the measurement, it is ultimately of interest what effect the errors have on the result. The statement that there is an error in the voltage measurement of 2 mV does not yet represent usable information. It is therefore necessary to relate the error quantities to the measurement in order to obtain a statement about the measurement system. The absolute error F_{abs} indicates the deviation between the nominal measured value MV_{so11} and the actual value of the measured quantity MV_{mess} . It carries the unit of the measurand. It is only valid for the designated measuring point. If its value is the same over the entire characteristic curve, it is based on an offset error of the measuring arrangement and applies:

$$F_{\text{abs}} = MV_{\text{mess}} - MV_{\text{soll}}.$$

The absolute error is used less frequently because it does not allow comparisons to be made due to its link to the concrete measured value. The error of 100 mV represents a high accuracy for a measured value of 100 V and low accuracy at 5 V measured value.

A speedometer in a motor vehicle usually has a preset absolute error of a few km/h. This ensures that you do not exceed the permitted speed with the larger winter tyres, as the speed is measured via the engine speed and not via the tyre circumference.

18.2.3 Relative Error

The frequently calculated relative error relates the measurement deviation to the measuring range. This error representation thus also allows different sensors to be compared with each

other. The error has no unit of measurement and is practically given as a percentage. Three representations are commonly used for the relative error.

The relative error in relation to the final value (full scale)

The relative error $F_{\text{rel(FS)}}$ can be related to the maximum value of a measurement MV_{max} . This means, for example, that a voltage measurement with a range of 10 V and an absolute error of 100 mV has a relative error of:

$$F_{\text{rel(FS)}} = \frac{MV_{\text{mess}} - MV_{\text{soll}}}{MV_{\text{max}}} = \frac{0.1 \text{ V}}{10 \text{ V}} = 0.01 = 1\% \text{FS}$$

has. This value appears in the data sheets as 1% F.S. (full scale) or also as 1% final value. If there is no further designation behind the percentage, a reference to the final value can always be assumed.

In addition to this information, a reference to “the range” may also be made. This means that the error is referred to as the *range* and not to the final value of the output signal. As an example: For a transmitter with 4–20 mA output current, the maximum value is 20 mA, but the span is only 16 mA.

The relative error in relation to the measured value

The disadvantage of the reference to the final value is that the relative error used also applies to small measured values. This means that the above-mentioned measurement with a measured value of 200 mV and the same absolute error of 100 mV can then already represent a deviation of 50%.

For this reason, with more demanding measuring arrangements, the error is related to the *current measured value* MV_{mess} , which is then indicated in % MV.

$$F_{\text{rel(MV)}} = \frac{MV_{\text{mess}} - MV_{\text{soll}}}{MV_{\text{mess}}} = \frac{0.1 \text{ V}}{0.2 \text{ V}} = 0.5 = 50\% \text{MV}. \quad (18.1)$$

From this formula, it can be seen that especially offset errors lead to drastic error values here. However, the reverse conclusion to this error specification is also that the measuring arrangement must offer a considerably higher precision than with an error specification related to the final value.

The relative error in relation to the smallest value (BFSL)

With the relative error according to FS, a *reference line* is regarded as the reference variable, which passes through the zero point and end value of the reference variable. This means that the greatest error occurs at the points of greatest non-linearity of the measurement characteristic.

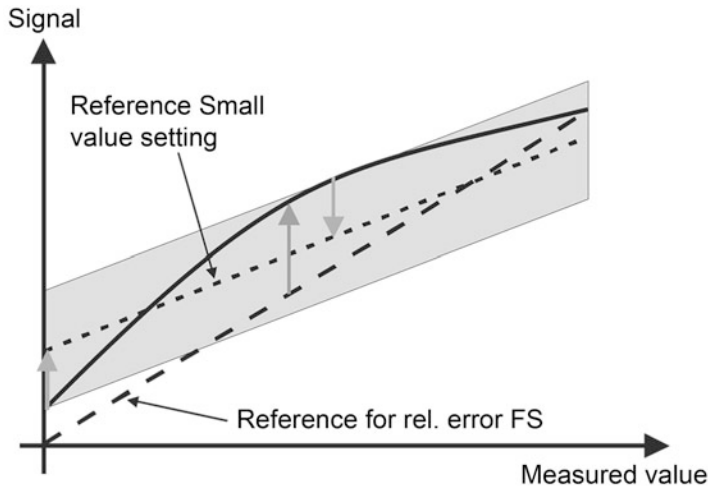


Fig. 18.2 Measurement error according to BFSL

With the *BFSL method* (Best Fit Straight Line), the reference line is laid over the measurement characteristic curve in such a way that the maximum positive and negative deviation is *equal* (Fig. 18.2).

In practice, the resulting error value is smaller by a factor of 2–4 than a specification with reference to the final value. The reason for this is that the characteristic curve is optimized so that the deviations reach a minimum and the offset error is not taken into account. In the application, this requires that the subsequent processing stage allows an adaptation to the imaginary reference characteristic curve.

Error representation in data sheets

The presentation of measurement errors is not clearly regulated in practice and every manufacturer tries to put his products in a favourable light. For this reason, it is often difficult to determine the actual error from the information in the datasheet.

Table 18.1 shows an extract from the datasheet of a pressure transmitter.

This table can be interpreted as follows:

- The basic specifications always apply to room temperature and are indicated here in the first three lines with 0.1% FS. No information is given for the zero point error.
- The total error in the range +10 °C to +40 °C is given as 0.15%FS, which should cover the sum of all errors in this temperature range.
- If the temperature falls below –20 °C, there is no longer any indication of an error, as compensation was only made up to this point.
- In the range from –20 °C to +10 °C and from 40 °C to 80 °C, an error of 0.1%/10 K is added, so that at 70 °C, for example, an error of maximum 0.45%FS results.

Table 18.1 Error information in a datasheet

Characteristic curve deviation (includes linearity, Hysteresis and Reproducibility)	% of span	<0.10 (<0.3 for measuring ranges >1000 bar)
Hysteresis	% of span	<0.04
Reproducibility	% of span	<0.05
Stability per year	% of span	<0.1 (for reference conditions)
Permissible temperatures		
Measuring material	°C	−40 ... +105
_environment	°C	−40 ... + 80
_Storage	°C	−40 ... + 85
Total error at +10 ... +40 °C	%	<0.15 (<0.6 for pressure ranges above 1000 bar)
Compensated temperature range	°C	−20 ... +80
Temperature coefficients in		
Compensated temperature range:		(Temperature error in the range +10 ... +40 °C included in the total error)
_ medium TC	% of span	<0.1 per 10 K (of zero point and span)

The influence of the operating voltage and the load of the signal output on the measurement must be added to the error information in the table.

When selecting a transmitter, the datasheet information must be compared and evaluated very thoroughly with the actual operating conditions. It is not uncommon for the individual sources of error, such as offset, gain, linearity and temperature, to be specified separately. Since it is not known whether the individual parameters are interdependent, they must be added up linearly.

18.3 Measurement Parameters

The output signals of sensors or sensor elements consist of a signal which is superimposed by various disturbances and error components. In order to identify these in the datasheets, *statistical methods* are used to describe the properties.

The following section describes some important values. Analogue-digital converters are used here as an example of the measurements.

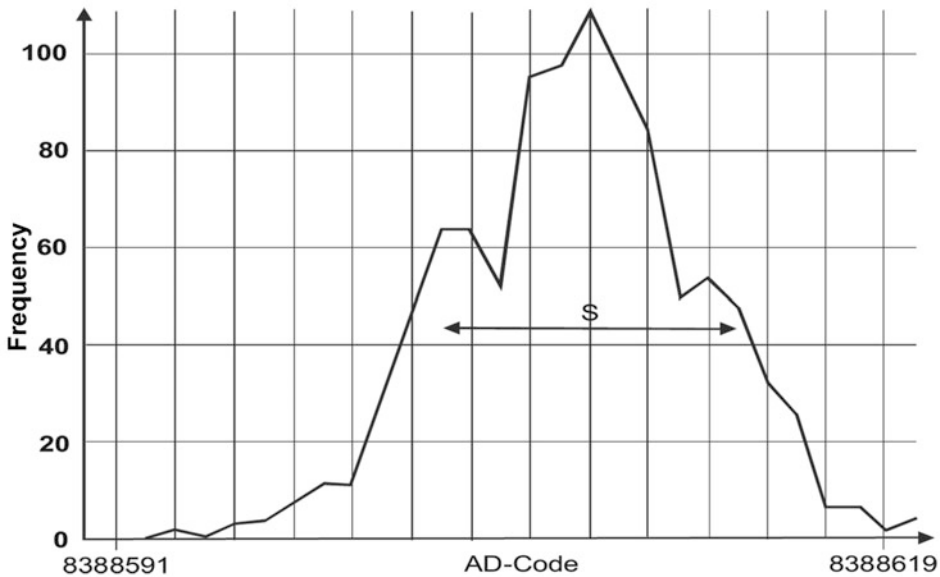


Fig. 18.3 Histogram of conversion results at 0 V input signal (center position) of an AD7791 (Source: Analog Devices)

18.3.1 Scattering of Measured Values

Due to the noise of a measuring arrangement, the measured value is not fixed but “moves” around the true value, the arithmetic mean value (Sect. 18.1; Fig. 18.1). The graphical representation of the measured values over the samples in a histogram normally produces a Gaussian distribution. Figure 18.3 shows this histogram for a converter with a 24-bit resolution.

In the example used, the measurement result is spread over a width of 28 code levels. The *scattering* d is calculated with

$$d = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})$$

with	n	Number of measured values
	x_i	Measured value i
	\bar{x}	Arithmetic mean value of all measured values.

If the result is related to the resolution of the AD converter, this results in a loss of 4.5 bits, which is why the datasheet only specifies an effective resolution of 19.5 bits.

18.3.2 Resolution of Measured Values

The resolution describes the smallest distinguishable change in the measured variable that can be detected in the output signal. With an AD converter, this would theoretically be a 1-bit stage. However, the above considerations on scattering can lead to the result that the smallest distinguishable signal can also be larger than the one-bit stage.

$$R_{\max} = \frac{\Delta x_{\min}}{x_{\max} - x_{\min}} \cdot 100\%$$

with	R_{\max}	maximum resolution of the measuring range/span
	x_{\max}	largest measured value
	x_{\min}	smallest measured value
	Δx_{\min}	smallest detectable difference.

With the values of the example used above, the following results are obtained

$$R_{\max} = \frac{28}{2^{24} - 0} \cdot 100\% = 0.0001669\%. \quad (18.2)$$

No less often, the inverse of this calculation is used to determine the required *transducer resolution* from given measured variables.

If the ratio between the smallest possible measured value and span is very high, the value is also given in *dB as the measurement dynamic range*.

18.3.3 Signal-to-Noise Ratio and Dynamics of Measured Values

The measurement dynamics describes basically the same facts as the resolution—it is about the ratio of the smallest possible to the largest possible measured value. The limits are described by the *signal-to-noise ratio (SNR)*.

Following the Eq. (18.1) in Sect. 18.2.2, the ratio between the measuring span and the smallest measured value can also be given by

$$x_{\text{dyn}}[\text{dB}] = 20 \lg \left(\frac{x_{\max} - x_{\min}}{\Delta x_{\min}} \right). \quad (18.3)$$

Using the values from the example in the previous section, this results in a dynamic range of 115 dB. The theoretical value for a 24-bit A/D converter is 144.5 dB. A large part of the measurement stages is lost due to the noise of the conversion.

Since it is too complex to determine the smallest detectable measured value level, this can be reduced to the smallest value that can be measured at all. It is assumed that *nothing* can be measured *below* the noise band. The first detectable value must therefore be greater than the noise level. Thus the noise level is equal to Δx_{\min} . The Eq. (18.3) can thus be modified to

$$\frac{S}{N} [\text{dB}] = 20 \lg \left(\frac{U_{\text{spanne}}}{U_{\text{Rausch}}} \right),$$

where	S/N	the signal to noise ratio,
	U_{Spanne}	the measuring range and
	U_{Rausch}	is the noise voltage.

The result provides information about the highest possible accuracy with which a signal can be measured.

Bibliography

1. Beyer, Michele: „Wie genau ist Ihr Sensor? – Ein Wegweiser durch den Dschungel der Genauigkeitsangaben“; WIKA-Sonderdruck aus MSR Magazin 4/2008

Index

A

- A/B interface, 434–436
- Ablative process, 272
- Absolute angle information, 261
- Absolute counting systems, 266
- Absolute dosimetry, 702
- Absolute error, 796
- Absolute force, 377
- Absolute humidity, 562
- Absolute interfaces, 251
- Absolute magnetic measuring systems, 249–251
- Absolute measuring method, 212
- Absolute measuring system, 248
- Absolute pressure measurement, 390
- Absolute pressure value, 593
- Absolute value measurement, 524
- Absorption, 112–115
- Absorption of light, 581
- Acceleration, 439
- Accelerometers, 440
- Accuracy classes, 458
- Accuracy of the measurement, 490
- Accuracy values of encoders, 264
- Acoustic effects, 123–128
- Acoustics, 142
- Acoustic waves, 123–124
- Active RFID transponders, 350
- Active sensors, 2
- Active switching zone, 33
- Activity coefficient, 114
- AC voltage resistance measurement, 509
- Adjustable trigger threshold, 423
- Adjustment, 748
 - with analog signal processing, 750–751
 - with digital signal processing, 751–753
- AD598 module, 167
- Adsorption, 117–118
- Aerospace, 235
- Affinity analysis, 652
- Aft microphone, 556
- Airbags, 441
- Air conditioning technology, 390, 391
- Air flow, 595–599
- Air pressure, 593, 594
- Air pressure measurement, 390
- Algae test, 654
- All-solid-state reference electrodes, 617
- Alpha radiation, 692
- Alpha spectroscopy, 704
- Alternating voltage, 486
- Ambient climate values, 589
- Ampere (A), 495
- Amperometric, 103
 - Clark sensors, 627
 - sensor, 108–109, 628, 660
 - solid electrolyte gas sensors, 639
- Anisotropic Magnet Resistive (AMR), 248
- Analog signal conditioning, 746
- Analogue interfaces, 757–762
- Analogue oscillographs, 423–424
- Analysis sensor, 136
- Analyzer, 132
- Anemometer, 595
- Angle-dependent modulation, 269
- Angle measurement, 22
- Angle of attack in fluid mechanics, 243, 244
- Angle of rotation, 276, 433
- Angle sensors, 136
- Angular acceleration, 443–444
- Angular encoders, 433
- Anion activity, 615
- Anisotropic Magneto Resistance (AMR), 16–19, 283
- Antioxidants, 609
- Apparent power, 513
- Application Specific Integrated Circuit, 272

- Approximated non-linear characteristic, 733
 Apto-ASICs, 272
 Arithmetic mean, 795–796
 Artificial nanostructures, 83
 Aspiration psychrometer, 568
 Assmann psychrometers, 568, 585
 Astronomy, 140–141
 Atomic clock, 405
 Atomic force microscope, 404
 Atomic force microscope (AFM) systems, 362
 Atomic force microscopy, 400
 Attenuation, 126
 Audible sound, 123
 Auto-Ident cameras, 346, 348, 351
 Automation engineering, 235
 Automation technology, 36, 47
 Automotive industry, 36
 Avalanche diodes, 701
 Avalanche photodiode (APD), 90, 198, 710, 715, 729
- B**
- Backscattering of light quanta, 477
 Backscattering of photons, 479
 Baffle plate scales, 448
 Barcode scanner, 339–343, 345
 Barrier operation, 214
 Basic principles of ATEX, 788–789
 Baud, 406
 Bayer pattern, 545
 Beam path, 202
 Beat, 411
 Beaufort scale, 595
 Becquerel, 694
 Bending beam, 387
 Berkovich indenter, 403
 Best Fit Straight Line (BFSL), 798
 Beta radiation, 692, 715
 Bethe-Bloch equation, 699
 BET method, 118
 BIL displacement sensor, 191
 BIL sensor element, 188
 Bimetallic thermometers, 471
 Biological/biochemical effects, 651
 Biological damage, 695
 Bioluminescence, 654
 Biomechanics, 383
 Biomolecular recognition element, 675
 BiSS, 252
 Bit/s, 406
 Blocking voltages, 92
 Blood gas analysis, 665
 Bolometer, 742
- Bougert–Lambert–Beer’s law, 581
 Bourdon tubes, 391
 Brake systems, 390
 Breakdown voltage, 716
 Bridge, 508
 Brightness, 531, 532, 541
 Brightness sensors, 537
 Brinell, 399
 Bubble chamber, 703
 Building services engineering, 193, 391
 Bulk flow meter, 447
 Bundled sound lobes, 330
- C**
- Calcium carbide method, 579
 Calcium hydride method, 579
 Calibration, 748
 Calibration of the INS, 158
 Camera sensors, 136
 CAN, 252
 bus, 765
 open, 765
 Candela, 534
 Capacitance, 35–47
 Capacitance C in the alternating current circuit, 41
 Capacitive, 149
 distance measurement, 391
 encoders, 292–295
 humidity sensors, 572
 NS, 320
 Capacitively, 445
 Capacity C, 499
 Capacity digital converter, 503
 Capacity measurement by RC generators, 503
 Capture antibody, 679
 Capture unit, 410
 Cardioid microphone, 557
 Catalytic poisons, 610
 Cations, 623
 Cation-selective sensor, 624
 Cell-based sensors, 655
 Celsius, 452
 Centripetal force, 374, 377
 Cermet layer system, 228
 Channel photomultiplier, 724
 Characteristics of the NS, 325, 326
 Charge carrier mobility, 51
 Charge-coupled devices (CCD), 134, 717
 elements, 701
 lines, 199
 matrix sensors, 735
 sensors, 732, 734–735
 Charged particles, 691

- Charge Q, 498
 - Charge sharing, 503
 - Charge storage capacity, 512
 - Chemical bonds, 112
 - Chemical effects, 110–123
 - Chemical potential, 116
 - Chemical processes, 486
 - Chemiluminescence biosensor, 659
 - Chemisorption, 112, 117–118
 - Cherenkov detectors, 708
 - Chirp, 333
 - Chopper, 67
 - Chopper-stabilized amplifiers, 67
 - Chronic toxicity, 654
 - CIE1931, 543
 - CIE Lab colour space, 543
 - Clark electrode, 666
 - Clark sensor, 109
 - Classification of groundwaters, 608
 - Climate in electrical installations, 591
 - Climate in museums, 588–590
 - Club microphone, 556
 - CMOS image sensors, 134
 - CMYK is a colour space, 544
 - Code disk, 271, 272
 - Coil impedance, 148
 - Color, 541
 - codes, 71
 - correction matrix, 546
 - sensor, 547
 - Colorimetric observer, 541
 - Colorimetry, 652
 - Colossal Magneto Resistance (CMR), 19, 248
 - Colour filter, 545
 - Colour perception, 540–543
 - Colour recognition, 97, 540
 - Compass sensors, 311
 - Competitive immunoassays, 680
 - Compressed air technology, 391
 - Compressed gas storage, 391
 - Compton effect, 89, 696, 697
 - Computed tomography/positron emission tomography (PET), 703
 - Condensation measurement, 586–587
 - Conductive filling material, 39
 - Conductive plastic potentiometers, 222
 - Conductometric, 103
 - and impedimetric sensors, 109–110
 - sensor, 648
 - Cones, 540
 - Confocal microscopy, 362
 - Confocal scanning microscopy, 362
 - Constriction resistance temperature sensors, 461
 - Contact and thermoelectric voltages, 486
 - Contactless distance measurement, 152
 - Contactless position sensor, 33
 - Contamination indicator, 200
 - Continuous ultrasonic flow measurement, 602
 - Conversion into a current, 488
 - Conversion into an electromagnetic field, 488
 - Conversion of the physicochemical change, 651
 - Conversion to heat, 488
 - Coriolis force, 443
 - Correction function, 752
 - Correlation measurement technology, 437
 - Cosmological redshift, 141
 - Coulombmeter, 500
 - Coulometric determination of oxygen, 640–641
 - Coulometric principle, 578
 - Coulometry, 103
 - Counting the blade pulses, 62
 - Covalent immobilization, 663
 - Cross-coil instrument, 515
 - Cubic thorium oxide, 459
 - Cubic zirconia, 459
 - Cu-constantan, 464
 - Cup anemometer, 595
 - Curie temperature, 72
 - Current, 495
 - clamp, 522
 - dividers, 496
 - flow through a resistor, 495–496
 - measurement, 24, 497
 - output, 759–760
 - transformer, 522
 - Current-fed transformer, 496
 - Currentless voltage measurement, 630
 - Curve interpolation, 752
 - Cut-off frequency, 86, 92
 - CW Radar, 331, 333
 - Cyclic polarization, 611
 - Cytotoxicity tests, 653
- D**
- Daniel Bernoulli, 596
 - Data Acquisition System (DAS), 516
 - DCF77, 406
 - DCLVDT, 167
 - Dehumidifying, 593
 - Delay time, 415
 - Dependence on temperature, 124–125
 - Detect accelerations, 38
 - Detection of different materials, 159
 - Detection of inhomogeneous zones, 159
 - Determination of chemical elements, 142
 - Determination of enzyme activity, 673–674
 - Dew point, 563–564
 - mirror, 569–572
 - temperature, 563

- Diagnostic Information, 778, 780
 - Diagnostics, 390
 - Diaphragm, 396
 - Difference frequency, 433
 - Differential capacitor, 38
 - Differential pressure instruments, 595
 - Differential pressure measurement, 389
 - Differential Seebeck coefficient, 65
 - Differential transformers with sliding core (LVDT), 163–165, 167–169
 - Diffraction principle, 269
 - Diffraction optics, 272
 - Digital interface, 308, 762–776
 - Digital multi-turn encoders, 282
 - Digital residual error, 407
 - Digital Storage Oscilloscopes (DSOs), 424–425
 - Diode characteristic curve is temperature dependent, 463
 - Direct digital synthesis (DDS), 411
 - Direct flow measurement, 601
 - Direct force measurement, 377
 - Direct immunosensors, 679
 - Directional characteristic of a microphone, 556
 - Direction of rotation, 434
 - Direct piezoelectric effect, 6
 - Discrete Fourier transformation, 421
 - Displacement, 136, 148–255
 - immunoassay, 681
 - sensors, 148–193
 - Displacement sensors with magnetically coded measuring standard, 244, 257–308
 - Display of measurement errors, 795–799
 - Display period, 422
 - Distance, 136
 - Distance-coded measuring body, 247
 - Distance measurement, 418, 419
 - Distance sensors, 148
 - Distributed Temperature Sensing (DTS), 479, 481, 483
 - DNA-based sensors, 681–686
 - Domain wall generator, 287
 - Doppler effect, 137–143, 333, 437
 - Doppler effect of light, 139–140
 - Doppler navigation satellites, 141
 - Doppler shift, 603
 - Dosimeters, 700
 - Dosing and flow measurement tasks, 193
 - Double beam laser Doppler systems, 142
 - Double distance measurement, 437
 - Draw-wire encoders, 266
 - Drift chambers, 708
 - Drift velocity, 48
- E**
- Earthquake, 441
 - Eddy current, 33
 - effect, 55–63
 - inductive sensors, 152
 - IS, 153
 - Effective magnetic pole width, 276
 - Effective value, 486
 - Effective value of current and voltage, 517
 - Elastic force, 377
 - Electric field constant, 37
 - Electric field strength, 48, 132, 509
 - Electric polarization, 5
 - Electrical charge, 498
 - Electrical conductance, 644
 - Electrical polarization, 79
 - Electrical resistance thermometers, 453
 - Electricity meter, 515
 - Electroacoustic transducers, 555
 - Electrochemical effects, 103–110
 - Electrochemical pH value, 665
 - Electrochemical reference electrode, 613
 - Electrodynamic loudspeaker, 380
 - Electroforming, 272
 - Electrolytic inclination sensors, 317
 - Electrolytic sensors, 311, 313
 - Electrolytic trough, 510
 - Electromagnetic compatibility (EMC), 782–785
 - Electromagnetic force effect, 380
 - Electromagnetic oscillating circuits, 33–34
 - Electromagnetic radiation, 130, 691, 721
 - Electromagnetic wave, 128, 527, 691
 - Electrometer, 487, 500
 - Electromotive force (EMF), 104, 464
 - Electron plasmon, 659
 - Electro-optical detection, 655
 - Electro-optical effect, 98–102
 - Electrostatic optics, 510
 - Electrostatic voltage, 486
 - Ellipsometry, 652
 - Elongation, 384–388
 - Encoder, 258
 - Encoder with diffractive optics, 269–271
 - EnDat, 252
 - Endicon, 725
 - Energy, 485
 - density, 527
 - flux density, 528
 - management for sensors, 753–755
 - measuring heads, 84
 - meter, 515
 - spectroscopy, 694

- Enthalpy, 564
Environmental influences, 220
Environmental noise, 558
Enzymatic and immunological sensor, 664
Enzyme immunoassay (EIA), 679
Enzyme sensors, 669–675
Equilibrium vapour pressure, 562
Equivalent dose, 695
Equivalent Series Inductance L (ESL), 499
Equivalent Series Resistance (ESR), 499
Evaluation electronics, 1
Evanescent field sensors, 658
Evaporative cooling, 566
Evaporator, 591–592
Event counting, 409
Exit energy, 86
External magnetic field of the earth, 311
External photoelectric effect, 132, 723
Extrinsic sensors, 481
- F**
Fahrenheit (°F), 452
Faraday's law, 500
Fast Fourier transformation (FFT), 425, 486
Fast oxygen measurement, 633
Fe-constantan, 464
Ferraris meter, 515
Ferromagnetic toothed wheel, 32
Fiber Bragg grating, 387, 388
Fiber Optics, 136
Field-effect transistors, 104
Field plate, 50, 521
Figaro gas sensors, 641
Finite Element Method (FEM), 55
First Fick's law, 117
Fixed-contact glass electrodes, 618
Flameproof encapsulation, 789
Flameproof enclosure, 791
Flat punch, 401
Flowmeters, 446
Flow rate, 445–449, 601
Flow velocity, 602
Fluorescence immunoassay (FIA), 679
Fluorometry, 652
Flux concentrators, 523
Foil potentiometer, 232, 233
Following error, 254
Food/pharmaceutical and chemical industry, 47
Food/pharmaceutical industry, 36
Force, 376–378, 380, 381
Force measurement, 378
Fotomultiplier, 723
4- and 6-wire interface, 762
4-bit PRC code, 250
- Four-electrode measuring cell, 645
Fourier series, 486
4-peak measurement, 508
Free-running cyclical generation of measured values, 238
Frequency, 406, 407
 output, 760
 ratio measurement, 409
 shift, 137
 spectrum, 509
Frequency-to-voltage conversion, 411
Freundlich equation, 118
Frictional force, 377
Fringe projection, 365
Frontal detection, 159
Full scale, 797
Functional safety (SIL), 785–787
- G**
Gait analyses, 383
Galvanic sensor, 628
Galvanomagnetic effects, 48
Game consoles, 443
Gamma radiation, 129, 692
Gas solubility, 112–115
Gas thermometer, 474
Gaussian effect, 48–52
Gaussian-shaped intensity distribution, 102
Geiger mode, 716
Geiger-Müller counter, 721
Geiger-Müller range, 707
Gel-stiffened electrode types, 617
Geodesy, 141
Geometric quantities, 148–368
Giant Magneto Resistance (GMR), 18–19, 248
 effect, 287
 sensor, 521
Glow discharge anemometer, 598
Glucose detection, 674
Gold electrodes, 610
Gravimetric water content, 576
Gravitational force, 309, 374
Gravitational redshift, 140
Grey value shift, 204
Grinder systems, 231, 232
Gross measurement error, 794
Gyrocompass, 443
Gyroscope, 309, 443
- H**
Hair hygrometers, 572
Hall angle, 51
Hall effect, 52–56

- Hall plate, 523
 Hall resistance, 53
 Hall sensor, 54, 521, 522
 Hall switches, 522
 Handling, 193
 Hardness, 398–404
 Hardness determination by nanoindentation, 400–403
 HART, 767–768
 Hearing threshold, 551
 Heat radiation, 67
 Henry constant, 114
 High dynamic CMOS image sensors, 732–733
 High-speed photography, 100
 High-temperature ion conduction, 459
 Hipface, 252
 HIV Gene Chip, 684
 Hooke's law, 385
 Hot/cold-hardening adhesives, 11
 HSB colour model, 544
 Humidification, 591
 Humidity, 561
 Humidity measurement in closed rooms, 587–593
 Hybrid interfaces, 261
 Hydraulics, 390
 Hydrology, 600
 Hydrometric vane, 604
 Hypersonic, 123
 Hysteresis, 226, 227, 253, 323, 324
- I**
- I2C, 771–772
 IEEE 1451, 773–776
 Ignitable environment, 787
 Ignition protection type, 789
 Illuminance, 90, 535
 Image pickup tubes, 725
 Imaginary reactance X_C , 41
 Immobilisers, 441
 Immunoanalysis, 677
 Immunosensors, 675
 Impedancemetric sensors, 637
 Impedimetric sensors, 103, 637
 Inclination, 308
 Inclination sensor, 442
 Inclined plane, 159
 Inclinator, 442
 Incremental analog signals, 261
 Incremental encoder, 307, 308, 434–436
 Incremental interfaces, 308, 309
 Incremental magnetic measuring systems, 248, 249
 Incremental measuring system, 248
 Indenter, 398
 Indirect force measurement, 378
 Indirect immunosensors, 679
 Indirect ionizing, 694
 Induced fit reaction, 652
 Inductance, 30, 148, 509, 517
 Inductive, 149
 and capacitive LV, 316
 conductivity measurement, 646
 distance, 148–193
 NS, 320
 sensors with variable reluctance, 152
 Inductively, 445
 Inductive Proximity Sensors (INS), 152
 Industrial Ethernet, 265
 Infrared, 129
 Injection technology, 390
 Inner photoelectric effect, 84, 87–88
 Inorganic-non-metallic thermocouples, 469, 470
 Insulation resistance, 92
 Insulation voltage, 92
 Integrating light pulses, 419
 Interference, 132, 269
 phenomena, 212
 resistance, 237
 Interferometric length measurement, 212, 213
 Internal photoelectric effect, 132, 726
 International altitude formula, 594
 International System of Units (SI), 451
 Intraoperative gamma probe, 718
 Intrinsic safety, 789–791
 Inverse piezoelectric effect, 5
 IO-Link, 768–770
 Ion-conductive glass, 613
 Ion conductors, 459
 Ion dose, 704
 Ion dose rate, 704
 Ionising radiation, 691, 700
 Ionization chamber, 396, 703
 Ionophores, 624
 Ion-selective conductometric microsensors (ISCOM), 624
 Ion-selective field-effect transistor (ISEFT), 104, 107, 662
 Irradiance, 528
 IWS, 160
- J**
- Josephson tunnel element, 486
 Joule, 485
- K**
- Karl Fischer titration, 577
 Kelvin (K), 451
 Kerr effect, 99

- Kerr constant, 101
Kohlrusch measuring cell, 644
- L**
- Lab room, 543
Lactate dehydrogenase, 654
Lambda sensors, 634, 636–638
Lambert radiator, 530
Lambert's law of cosine, 530
Land surveying, 141
Langmuir adsorption isotherm, 117
Langmuir-Hinshelwood mechanism, 118
Laser, 134
 scanners, 136
 triangulation, 357
Laser Doppler Anemometry (LDA), 142
Laser Doppler method, 599
Laser Doppler vibrometers, 142
Lateral approach, 159
Law of induction, 28–32
LCh color space, 544
LC resonant circuit, 33
Length measurement, 23
Level measurement, 38, 103, 390, 593
Level measurement in waters, 390
Level probes, 391
Light barriers, 93–94
Light curtain, 96, 136
Light-emitting diodes (LED), 134, 195
Light grid, 136
Light measurement, 96–97
Light meter, 537
Light scattering, 659
Light sensor, 136
Limiting wavelengths, 133
Line sensors, 735
Linearity deviation, 251, 253, 276
Linear, magnetostrictive displacement sensors, 234
Linear variable differential transformer (LVDT), 31
Liquid-glass thermometers, 472
Liquid spring thermometers, 472
Liquid thermometers, 453
Lissajous figures, 428
Lithium enriched silicon, 714
Logarithmic characteristic curve, 733
Logic analyzer, 426
Longitudinal waves, 549
Look-up table (LUT), 753
Lorentz force, 15, 48, 52
Lorentz transformation, 510
Loudness, 554
Loudspeakers, 555
Lumen (lm), 532
Luminance, 534
Luminance meter, 537
Luminous efficacy, 535
Luminous flux, 532
Luminous intensity I_v , 534
Luv color space, 544
Luxmeter, 535
LVDT differential transformer sensors, 152
- M**
- Magic eye, 510
Magnetically detectable, 433
Magnetically encoded rotary encoder, 275–279, 281
Magnetic fields, 50
 current flow, 496
 NS, 320
 strength, 132, 520
Magnetic flux density B, 520
Magnetic-inductive measuring method, 604
Magnetic resonance imaging (MRI), 521
Magnetic sensors, 309, 496
Magneto, 149
Magneto-electric scanning, 150
Magneto-inductive BIL, 193
Magneto-resistive (MR) effect, 15–24, 248
Magneto-resistive inclination sensors, 310, 317
Magnetostrictive displacement sensor, 234, 236, 238, 239
Magnetostrictive effect, 25–29
Maltese cross transmission, 283
Mass, 373, 374, 376, 445
Mass flow, 445
Material flow, 116
Material moisture, 574
Material moisture analyzer, 577
Material radiation, 721
Matthiesen rule, 458
Maximum amount of humidity, 562
Mean value of the power, 512
Measured value, 797
Measurement by comparison, 488–490, 492
Measurement errors, 801
Measurement of capacitances, 501
Measurement of thermal conductivity, 584–585
Measurement through integration, 491
Measuring equipment monitoring, 382
Measuring principles for resistance, 506
Measuring principles for the electric field strength, 510–511
Mechanical, 235
 deformation, 391
 engineering, 36, 47
 oscillations, 374

- Medical technology, 47, 142, 235, 383
 Membrane deformation, 393
 Micro-electro-mechanical system (MEMS), 317, 441
 inclinometers, 311
 sensors, 314
 Metabolic cell chip, 687
 Methodical measurement error, 794
 Michaelis constant, 671
 Michaelis-Menten-Theory (MM-Theory), 671
 Microgravimetric, 655
 Microphones, 555
 Microstructured discs, 272
 Microwave radiation, 129
 Milliohmmeters, 509
 M-In-Track and R-In-Track[®] Sagentia technologies, 152
 Mixed-phase disorder, 121–123
 Mixed potential sensors, 637
 Mixed Signal Oscilloscopes (MSOs), 426
 Mößbauer spectroscopy, 142
 Mobile working machines, 235
 Modulated light, 421
 Modulator, 258
 Mohs' hardness scale, 398
 Moiré effect, 269
 Moisture, 561–574
 Moisture analysis using electromagnetic fields., 580
 Mollier diagram, 564
 Monoclonal antibodies, 676
 Monoflop, 410
 Monte Carlo method, 494, 517
 Morphological modular system of inductive sensors, 154
 Motion analysis, 441
 Motion detectors, 81
 Motor feedback systems, 265
 Moving Target Indicator (MIT), 141
 Multimeters, 491
 Multiple exposures, 733
 Multiple positions—one sensor, 241, 242
 Multiple redundant system, 786
 Multi-position measurement, 239
 Multi-turn angle sensor, 284
 Multi-turn encoders, 266
 Multi-turn potentiometer, 284
 Mutual inductance, 296
- N**
- Nano-electro-mechanical system (NEMS), 374
 Nanoindentation, 400
 Navigation satellites, 406
 Near-field navigation, 443
 Nebulizer, 592
 Negative switching output (NPN output), 324
 Negative temperature coefficient (NTC), 51, 73, 460
 Nernst equation, 112, 607
 Neurochips, 687
 Neuro-transistor, 687
 Neutral particles, 692
 Newton's law of action, 373
 Newton's second axiom, 376
 Nikolsky equation, 623
 Noise voltage, 486
 Nominal response ramp, 75
 Non-contact magnetoinductive displacement sensors (smartsens-BIL), 184–193
 Non-contact measuring methods, 357
 Non-contact temperature measurement, 453
 Non-mechanical magnetic NS, 316
 Non-self-emitting devices, 543
 Novel GMR system for detection and storage of rotation information, 287, 289, 291
 NS (proximity switches), 316
 NS with diagnostic information, 328
 NS with several switching outputs, 326, 328
 NTC resistors, 73
 Nuclear magnetic resonance (NMR), 583–584
 Nuclear measurement methods, 582
 Nuclear photoelectric effect, 696
- O**
- Object detection, 102
 and distance measurement with ultrasound, 330, 331
 with laser scanner, 337
 with radar, 331, 332, 334
 Object influences, 219, 220
 Ohmic resistance, 644
 Ohm's law, 48, 495
 Omnidirectional microphone, 556
 1-bit ADC, 491
 Open sensor solution, 233
 Operating principle of sensors, 2
 Optical belt weigher, 449
 Optical coherence tomography (OCT), 362
 Optical effects, 128, 129, 132–134, 136, 139–142
 Optical encoders, 267–275
 Optical fibers (OF), 129
 Optical interferometry, 149
 Optical light propagation time, 149
 Optical probing 3D measuring methods, 360–363
 Optical sensors, 133–135
 Optical, switching single point sensor, 360, 361
 Optical systems, 453

- Optical 3D measurement (grid and line projection), 363, 364
- Optical triangulation, 149
- Optocoupler, 84, 91–93
- Optoelectronic components, 195–205
- Optoelectronic distance and displacement sensors, 193–213
- Optoelectronic NS, 320
- Oscillographs, 422
- Oscillograph tube, 510
- Oscilloscopes, 416, 422
- Outer photoelectric effect, 84–87
- Output type, 324, 325
- Overload protection, 393
- Oxygen ion conducting, 459
- P**
- Pain threshold, 551
- Pairing effect, 699
- Palpable 3D measurement methods, 357, 358, 360
- Pantone colours, 545
- Parallel coding, 249
- Partial saturation, 185
- Partially selective redox electrodes, 626
- Particle flow, 691
- Passive compensation, 749
- Passive infrared sensors (PIR), 82
- Passive RFID transponders, 350
- Passive sensors, 2
- Path dependent change of the coil inductance, 186
- Path-time measurement, 437
- pCO₂ value measurement, 667
- Peak voltage, 486, 494
- Pellistors, 643
- Perception of loudness, 553
- Perception of sound, 549
- Period duration, 409, 486
- Period duration measurement, 415
- Permeability, 30
- Permittivity number, 37
- Personal protection, 196
- Peters equation, 607
- Phase accumulator, 412
- Phase/frequency delay method, 208, 210
- Phase-locked loop (PLL), 415
- Phase position, 434
- Phase shift, 416, 428
- pH electrodes, 613
- pH glass electrode, 106
- pH-ISFET, 617
- pH measurement, 618–622
- Phon, 553
- Photoconductivity, 84
- Photoconductor, 87–88
- Photodiode, 90, 726, 728
- Photoelectric effect, 84–97, 696
- Photoelectric NS, 316
- Photoelectric properties, 721
- Photoelectric scanning, 149, 211, 212
- Photoelectric sensor elements, 90
- Photoelectric sensors, 721–742
- Photoelectric switches, 95–96
- Photo element, 90
- Photo-FET, 731
- Photoionization, 88–89
- Photometric, 527
- quantities, 531
- radiation equivalent, 533, 534
- Photomultiplier, 85, 90–91, 709
- Photon detectors, 721
- Photons, 85
- Photoresistor, 90, 726
- Photothyristor, 90, 731
- Phototransistor, 90, 731
- Photovoltaic effect, 84
- Photo winch diode, 90
- Physisorption, 112
- Piezo crystal, 396
- Piezoelectrically, 445
- Piezoelectric ceramics, 8
- Piezoelectric coefficient, 6
- Piezoelectric crystals, 7
- Piezoelectric effect, 5–8, 378
- Piezoelectric force washer, 378
- Piezoelectric interference signals, 80
- Piezoelectric materials, 80
- Piezoelectric measuring principle, 379
- Piezoelectric system, 657
- Piezo-resistive bridges, 391
- Piezoresistive constants, 13
- Piezoresistive effect, 378, 379
- Piezoresistive inclination sensors/DMS bending beam sensors, 313
- Piezoresistively, 445
- Piezo resonators, 396
- PIN diode, 701, 710
- PIN photodiodes, 729
- Planckian quantum of action, 132
- Planck's quantum, 86
- Planck's quantum of action, 694
- Platinum temperature sensors, 457
- PLCD displacement sensors (Permanent Linear Contactless Displacement Sensor), 178–184
- Pneumatics, 390
- Pockels effect, 99
- Pointer instruments, 488
- Polarization, 334

- Polarizer, 132
 - The Polarogram, 618
 - Polarographic, 617
 - Pole width, 276
 - Polymer coating systems, 230
 - Position-controlled hydraulic drive, 242
 - Position detection and contour tracking, 159
 - Position Sensitive Detector (PSD), 360
 - Position-sensitive photodiode, 94
 - Position sensor, 175, 238
 - Positive switching output (PNP output), 324
 - Positive temperature coefficient (PTC), 73, 460
 - Potentiometer circuit, 225
 - Potentiometric, 103, 150, 617, 627
 - displacement and angle sensors, 222–234
 - electrodes, 608
 - oxygen measurement, 630–633
 - sensors, 104–108
 - Power, 511
 - AC, 513–515
 - DC, 512–513
 - Power factor $\cos\phi$, 514
 - Precious metal redox electrodes, 610–611
 - Precipitation measurement, 585–586
 - Precision measuring resistors, 507
 - Presentation of measurement errors, 798
 - Pressure, 388–393
 - fluctuations, 549
 - transmitters, 393
 - Principles of tachometers, 62
 - Process control, 381
 - Process engineering, 36
 - Profibus, 252, 770
 - Protection functions, 777, 778
 - Proximity/limit switch, 34
 - PSD element, 198
 - Psychrometer, 566
 - Psychrometric measuring principle, 568
 - Pt100, 455
 - Pt1000, 455
 - PTC resistors, 73
 - PTC thermistors (PTC resistors), 73–75, 77
 - Pulse and phase measurement, 409
 - Pulse delay method, 206, 208
 - Pulsed inductive linear position sensor (Micropulse® BIW), 168–173
 - Pulse duration measurement, 415
 - Pulse integration, 421
 - Pulse length, 409
 - Pulse radar, 331
 - Pulse ratio, 415
 - Pulse wave (PW), 142
 - Pulse width, 412
 - Pulse width modulation (PWM), 412, 431, 760
 - PW-Doppler, 142
 - Pyroelectric capacitance, 80
 - Pyroelectric coefficient, 80
 - Pyroelectric detector, 741
 - Pyroelectric effect, 79–84
 - Pyroelectric sensitivity, 80
 - Pyroelectric sensors, 82
 - Pyroelectric sensors for motion and presence detection, 334, 335, 337
 - Pyrometers, 479
- Q**
- The Quadratic deviations, 795
 - Quantity, 445
 - Quantum Well Infrared Photodetector (QWIP), 736–738
 - Quartz crystals, 406
 - Quartz resonators, 396
 - Quartz thickness oscillators, 657
- R**
- Radially scanned systems, 270
 - Radial movement, 159
 - Radiant intensity, 529, 530
 - Radiant power, 528
 - Radiation diagnostics, 703
 - Radiation energy, 694
 - Radiation thermometers (pyrometers), 477
 - Radio Detection And Ranging, 331
 - Radio frequencies, 129
 - Radioimmunoassay (RIA), 679
 - Radiometric, 527
 - RAL colours, 544
 - Raman effect, 477
 - Random measurement error, 794
 - Rated operating distance, 321
 - Rated switching distance, 42
 - Ratiometric measurement, 507
 - Ratiometric voltage output, 759
 - Reactance, 148
 - Reaction heat, 83
 - Reactive power, 513
 - Real Aperture Radar (RAR), 141
 - Real operating distance, 322
 - Real power, 513
 - Real signal, 423
 - Receiver, 258
 - Recipient, 272–274
 - Red laser light, 196
 - Redox balance, 607
 - Redox glass electrodes, 611–613
 - Redox potential, 609

- Redox recycling, 661
Reflecting surface, 129
Reflection, 127–128
Reflective surface, 204
Reflector light barrier, 94
Reflex light barriers, 94
Refraction, 129
Refractive index, 98, 131
Relative biological effectiveness, 695
Relative dosimetry, 709
Relative error, 796–799
Relative humidity, 562
Relative pressure measurement, 390
Relative propagation time measurement., 416
Repetition frequency, 44
Resistance measurement through compensation, 508–509
Resistance thermometers, 70
Resistive and piezoresistive effect, 9–14
Resolution, 226, 227
Resolution of measured values, 801
Resolvers, 295, 296, 298, 299, 302–304
Resolver systems, 304
Resonant circuit, 381, 519
Retro-reflective photoelectric sensor, 136
Reverse magnetostrictive effect, 237
Reverse polarity protection, 778
RFID frequency ranges, 351, 352
RFID readers, 339
RFID systems and readers, 348, 349
RGB color model, 543
RGB colour locations, 543
Rheostat mode, 228
Rhermodynamic equilibrium states, 451
Rise times, 428
Risk classification, 788
RMS, 425
Robotics, 193
Roche Diagnostics, 674
Rockwell, 399
Rod, 540
 designs, 240
 dosimeter, 707
Rolling friction, 377
Röntgen, 692
Room acoustics, 560
Rotary measuring standards, 211
Rotational movement of an axis, 280
Rotational position, 434
Rotation-counting angle sensors, 281–291
Rotation-counting method on inductive basis, 284, 286
Rotation sensors, 309
rpm, 406
RS485, 768
RSM2800 series, 291
Running time, 418
- S**
SAE J1939, 766
Safe operating distance, 42, 323
Safety aspects for sensors, 777–791
Safety-oriented design, 786
Sampling method, 426
Sampling Principle, 426–427
Sandwich immunoassay, 679
Saturated calomel electrode (SCE), 615
Saturation humidity, 562
Sawtooth generator, 422
Scanning of a rotation, 434
Scanning of rotating objects, 159
Scattering of measured values, 800
Scintillation detectors, 708
Scintillators, 721, 722
Seat belt tensioners, 441
Second, 405
Secondary electron multiplier (SEV), 90
Second Fick's law, 117
Seebeck coefficient, 64
Seebeck effect, 63, 464–470
Seeger cones, 476
Seismic mass, 440
Self-emitting lamps, 544
Self-heating coefficient, 70
Semi-active RFID transponders, 351
Semiconducting ceramic resistors, 73
Semiconductor components, 195
Semiconductor industry, 47
Semiconductor laser diodes, 196
Sensitivity curve $V(\lambda)$, 534
Sensor electronics of the INS, 155, 156, 158
Sensor element, 1
Sensor operating frequency, 156
Sensors for angle and rotation, 257–308
Sensors for automatic identification (Auto-Ident), 337–355
Sensors for object detection, 316–355
Sensors in explosive environments (ATEX), 787–791
Sensor systems, 1, 2
Serial code, 212
Serial interfaces, 261
Servo converter, 492
Servo-inclinometer, 315–317
Severinghaus electrode, 667
Severinghaus principle, 627
Shaft couplings, 262

- Shear beam load cell, 376
Shipbuilding, 235
Shore durometer, 399
Short ASIC (application specific integrated circuit), 156
Short-circuit protection, 777
Sievert, 695
Sigma-delta converter, 491
Signal conditioning with ASICs, 746–747
Signal conditioning with microcontrollers, 748
Signal conditioning with system circuits, 746
Signal processing, 745–748
Signal-to-noise ratio (SNR), 801
Silicon pressure sensor, 394
Silver chloride reference electrode, 615
Sin/cos resolver (vector decomposer), 298–300
Single beam laser Doppler systems, 142
Single-head systems, 213
Single nucleotide polymorphisms (SNPs), 686
Singleturn sensor system, 281
Sinusoidal signals, 258
Sliding friction, 377
Slingshot psychrometer, 568
Slope, 309
Smart-Level 50 technology, 44
Smart sensors, 2
Snellius' law, 131
Solid angle, 528
Sone, 554
Sound absorption, 126–127
Sound generators, 555
Sound impedance, 551
Sound insulation, 559
Sound intensity, 125–126, 550, 551
Sound perception of the human ear, 554
Sound power, 550
Sound pressure, 549
Sound quality, 560
Sound transducers, 555
Sound velocity, 123, 550
Spatial resolution, 443
Specific electrical conductivity, 504
Specific electrical resistance, 504–509
Specific humidity, 562
Specific light emission, 535
Spectral sensors, 136
Speed, 436–439
 and angle of rotation, 431–436
 measurement, 23, 142
 of sound, 550
SPI, 772
Spinel structure, 76
Spin-valve principle, 19
Spiral arms, 287
Spot welding, 11
Spring force, 377
Sputtered PZT films, 334
Square-wave, 258
Square-wave voltammetry, 617
SQUID, 521
SSI, 252
Standard chromaticity coordinates, 542
Standard deviation, 795–796
Standard hydrogen electrode (SHE), 613
Standardized interfaces, 757
Standard resolver (electrical resolver), 298, 302, 303
Standard spectral values, 541
Standard valence system, 541
Static friction, 377
Static luminescent bacteria test, 654
Statistical methods, 494
Stator coupling, 262
Steam humidifier, 592
Steep signal edges, 44
Stefan–Boltzmann's law, 477
Stepped exposure, 733
Steradian, 528
Stoichiometric point, 636
Strain gage (SG), 10, 379, 385
Strain measurement, 385–388
Strength of penetration, 400
Strength of the ionization, 696
Successive approximation, 492–493
Surface charges, 79
Surface conductivity, 641–643
Surface Near Field Optical Microscope (SNOM), 362
Surface plasmon resonance (SPR), 652, 658
Switching distance, 34, 321–323
Switching element function, 323
Switching lug, 33
Switching sensor, 357, 358
Switching times, 92
Synchronous triggering, 238
Systematic measurement error, 794
- T**
Tactile, 357
Tagushi, 641
Temperature and frequency, 474, 475
Temperature as physical state variable, 451–452
Temperature coefficient, 69, 227, 228, 455, 458
Temperature effects in semiconductors, 72–78
Temperature fixed points, 453
Temperature-independent component, 459
Tensiometric measurement, 582

- Tensiometry, 582
Thermal anemometers, 596
Thermal behaviour, 496
Thermal detectors, 83, 721
Thermal effect of current flow, 496
Thermal expansion of solid bodies, 470
Thermally activated ion conductivity, 459
Thermal optical detectors, 738–742
Thermal radiation, 81, 477
Thermal response times, 69–70
Thermistor, 72, 460
Thermistors (NTC resistors), 76–78
Thermochromic substances, 475
Thermocouple, 66, 452, 464, 468
Thermo-diffusion, 64
Thermoelectric effect, 63–68
Thermoelectric emf, 452
Thermoelectric voltage, 464
Thermography, 738
Thermogravimetric, 577
Thermoluminescence detectors (TLD), 708, 709
Thermopile, 67, 739
Thermoresistance effect, 68–71
Thickness measurement, 159
3D field simulations, 55
3D imaging measuring methods, 363, 365, 367
Three-dimensional measuring methods
 (3D measurement), 356–368
Three-electrode measuring cell, 611
Three-wire interface, 760
Threshold value comparator, 319
Throttle valve sensors (TPS), 235
Throughbeam light barrier, 94
Through-beam photoelectric sensor, 136
Time measurement, 407
Time-of-flight, 418
Time-of-flight measurement, 337
Times, 405
Time signal transmitters, 405
Titanate ceramic, 73
Toothed wheel sensors, 56
Torque, 395, 398
Torque wrench, 396
Torsional force, 396
Torsional structure-borne sound waves, 236
Torsional wave propagation, 238
Totally internal reflection fluorescence
 (TIRF), 659
Traffic engineering, 382
Transfer function, 190
Transmission, 127–128
Transmitter, 258
Transport equations, 116
Travel speed, 278
Triangulation, 204, 205
 principle, 94
 sensors, 199
Triggering, 422
True Power-On sensor (TPO), 281
Tunnel Magneto Resistance (TMR), 19, 248
Two-edge converter, 491
Two-element sensor, 82
2-phase switching sensor, 359, 360
Two-stage interaction mechanism, 184
Two-wire interface, 759
- U**
Ultrasonic Doppler principle, 603
Ultrasonic flow measurement, 601
Ultrasonic NS, 316
Ultrasonic sensor, 213–222
Ultrasonic transducer, 215–216
Ultrasonic wind sensors, 596
Ultrasound, 123, 149, 330
Ultrasound NS, 320
Upstream Fresnel lens, 83
Useful operating distance, 322
UV light, 129
- V**
Valinomycin, 624
Van der Waals forces, 112
Variable transformers (VT), 152, 295, 296, 298,
 299, 302–304
Velocity, 436
Venturi, 595
Vernier principle, 283
Vibration analysis, 142, 443
Vickers, 399
Vidicon, 725
Villari effect, 25
Virtual sensors, 2
Voltage, 485
Voltage output, 758–759
Voltammetric, 617
Voltammetry, 103
Volume flow, 446, 595
Volume per time, 445
Volumetric water content, 576, 578
1Vpp or sin/cos interfac, 304–306
- W**
Wandering phase shift, 430
Water flow, 600–605
Water jet cutting, 391

Water level dependent factor, 605
Water vapor partial pressure, 561
Water vapor saturation pressure, 561
Weighfeeders, 448
Weighing equipment, 446
Weight force, 376
Weight sensors, 376
Weston standard element, 486
Wetting sensors, 587
Wheatstone bridge circuit, 11, 386
White balance, 546
White light sensors, 362
WhitePoint sensor, 362
Wiedemann effect, 25
Wiegand rotation detection, 286
Wiegand sensors, 284
Wind direction meters, 595
Wind force, 595
Windward color space, 544
Wire breakage protection, 778

Wire potentiometer, 228
Work, 512
W-wire measurement, 507

X

XMR technology, 19
X-ray radiation, 129
X-rays, 692

Y

Yttria-stabilized zirconia (YSZ), 122

Z

Z coil quality QL, 148
Zoom option, 428