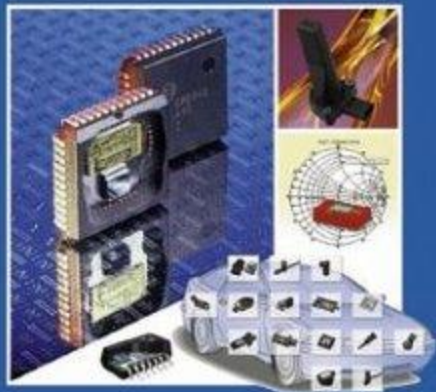


Sensors for Automotive Technology

Edited by J. Marek, H.-P. Trah,
Y. Suzuki, I. Yokomori

Series Editors:
J. Hesse, J. W. Gardner, W. Göpel

With a Foreword
by H.-J. Queisser



Sensors
Applications

Volume 4

Sensors Applications

Volume 4

Sensors for Automotive Applications

Sensors Applications

Upcoming volumes:

- **Sensors in Medicine and Health Care**
- **Sensors in Aerospace Technology**
- **Sensors in Environmental Technology**

Related Wiley-VCH titles:

W. Göpel, J. Hesse, J. N. Zemel

Sensors Vol. 1–9

ISBN 3-527-26538-4

H. Baltes, G. K. Fedder, J. G. Korvink

Sensors Update

ISSN 1432-2404

T. C. Pearce, J. W. Gardner, S. S. Schiffman, H. T. Nagle

Handbook of Machine Olfaction

ISBN 3-527-30358-8

Sensors Applications

Volume 4

Sensors for Automotive Applications

Edited by

J. Marek, H.-P. Trah, Y. Suzuki, I. Yokomori

Series Editors:

J. Hesse, J. W. Gardner, W. Göpel †



WILEY-
VCH

WILEY-VCH GmbH & Co. KGaA

Series Editors

Prof. Dr. J. Hesse

formerly of
Carl Zeiss, Jena
Bismarckallee 32c
14193 Berlin
Germany

Prof. J. W. Gardner

University of Warwick
Division of
Electrical & Electronic
Engineering
Coventry CV 7AL
United Kingdom

Volume Editors

Dr.-Ing. Jiri Marek

Robert Bosch GmbH
Sensore-
Entwicklungsleiter
Postfach 1342
72703 Reutlingen
Germany

Dr. Hans-Peter Trah

Robert Bosch GmbH
FV/FLD
Robert-Bosch-Platz 1
70839 Gerlingen
Germany

Prof. Dr. W. Göpel †

Institut für Physikali-
sche und Theoretische
Chemie
Universität Tübingen
Auf der Morgenstelle 8
72076 Tübingen
Germany

Yasutoshi Suzuki

Denso Corporation
Manager of Research
and Development
Section IC
1-1 Showa-cho,
Kariya-shi
Aichi-ken, 448-8661
Japan

Iwao Yokomori

Denso Corporation
IC Engineering
Department 2
1-1 Showa-cho,
Kariya-shi
Aichi-ken, 448-8661
Japan

This book was carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published

by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available on the Internet at <<http://dnb.ddb.de>>

© 2003 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation in other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany

Printed on acid-free paper

Composition K+V Fotosatz GmbH, Beerfelden

Printing Betz-Druck GmbH, Darmstadt

Bookbinding Litges & Dopf GmbH, Heppenheim

ISBN 3-527-29553-4

Preface to the Series

As the use of microelectronics became increasingly indispensable in measurement and control technology, so there was an increasing need for suitable sensors. From the mid-Seventies onwards sensors technology developed by leaps and bounds and within ten years had reached the point where it seemed desirable to publish a survey of what had been achieved so far. At the request of publishers WILEY-VCH, the task of editing was taken on by Wolfgang Göpel of the University of Tübingen (Germany), Joachim Hesse of Carl Zeiss (Germany) and Jay Zemel of the University of Philadelphia (USA), and between 1989 and 1995 a series called *Sensors* was published in 8 volumes covering the field to date. The material was grouped and presented according to the underlying physical principles and reflected the degree of maturity of the respective methods and products. It was written primarily with researchers and design engineers in mind, and new developments have been published each year in one or two supplementary volumes called *Sensors Update*.

Both the publishers and the series editors, however, were agreed from the start that eventually sensor users would want to see publications only dealing with their own specific technical or scientific fields. Sure enough, during the Nineties we saw significant developments in applications for sensor technology, and it is now an indispensable part of many industrial processes and systems. It is timely, therefore, to launch a new series, *Sensors Applications*. WILEY-VCH again commissioned Wolfgang Göpel and Joachim Hesse to plan the series, but sadly Wolfgang Göpel suffered a fatal accident in June 1999 and did not live to see publication. We are fortunate that Julian Gardner of the University of Warwick has been able to take his place, but Wolfgang Göpel remains a co-editor posthumously and will not be forgotten.

The series of *Sensors Applications* will deal with the use of sensors in the key technical and economic sectors and systems: *Sensors in Manufacturing, Intelligent Buildings, Medicine and Health Care, Automotive Technology, Aerospace Technology, Environmental Technology* and *Household Appliances*. Each volume will be edited by specialists in the field. Individual volumes may differ in certain respects as dictated by the topic, but the emphasis in each case will be on the process or system in question: which sensor is used, where, how and why, and exactly what the benefits are to the user. The process or system itself will of course be outlined and

the volume will close with a look ahead to likely developments and applications in the future. Actual sensor functions will only be described where it seems necessary for an understanding of how they relate to the process or system. The basic principles can always be found in the earlier series of *Sensors* and *Sensors Update*.

The series editors would like to express their warm appreciation in the colleagues who have contributed their expertise as volume editors or authors. We are deeply indebted to the publisher and would like to thank in particular Dr. Peter Gregory, Dr. Jörn Ritterbusch and Dr. Claudia Barzen for their constructive assistance both with the editorial detail and the publishing venture in general. We trust that our endeavors will meet with the reader's approval.

Oberkochen and Coventry, November 2000

JOACHIM HESSE
JULIAN W. GARDNER

Foreword

H.-J. QUEISSER

Little sensors make great sense! Biological systems can survive sudden external changes only if their sensory organs can speedily detect the altered environment and thus trigger the necessary responses. Technical systems are now beginning to behave in a similar quick, efficient mode. Distributed intelligence has become available and affordable and is making remarkable inroads into many applications. This voluminous tome is a most impressive example of current trends in utilizing modern materials, novel simulation techniques, and, in particular, the new and commanding technology of microminiaturization.

Richard Feynman, that imaginative and farsighted theorist, presented the after-dinner speech at the Pasadena meeting of the American Physical Society in 1959; its memorable title was "There is plenty of room at the bottom!" I consider myself most fortunate to have been a witness to this fascinating speech. Feynman showed, profoundly but amusingly, that no basic laws of physics prohibit technical systems from being reduced to very small sizes, even to atomic dimensions. Neither thermodynamics nor quantum mechanics forbid such ventures into smallness, although economic considerations might be limiting. This speech encouraged and motivated me enormously in my own work on semiconductors. I had just joined Shockley's laboratory in Mountain View, California to work on silicon, the material that was later to lend its name to the Santa Clara Valley. Shockley's lab was housed in an old apricot barn on San Antonio Avenue, now recognized as the 'cradle' of Silicon Valley.

Microminiaturization of semiconductors had been the driving motivation all along; it began in seriousness in the 1960s.

The success in size reduction that we have attained today was unimaginable in the beginning. Submicron control of design rules seemed hopeless and far too expensive. But the economics of the integrated circuit overcame all obstacles. Expensive new tools and methods were developed, and they paid off. Later on, in the 1980s, the remarkable tools in the arsenal of semiconductor processing gave birth to the new technology of micromechanics, with its small sensors of high speed and sensitivity, yet low cost and wide applicability. Silicon is now extremely well understood and under control. This material is therefore the mainstay for sensors and actuators of microscopic dimensions. Recent developments, however, clearly show that new materials and new technologies with specific applications in micro-

Sensors for Automotive Applications. (Sensors Applications Volume 4).

Edited by J. Marek, H.-P. Trah, Y. Suzuki, I. Yokomori

Copyright © 2003 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ISBN: 3-527-29553-4

mechanical sensors are still evolving. Technologies like thin-film technology, thick-film technology, and ceramics technology also turn out to be of essential importance in automotive sensors. The right choice of technology always depends on the special application, but all these innovative technologies are driven by common challenges: high-volume batch fabrication, miniaturization, functional improvements, high reliability, and low cost.

This book covers one aspect of sensors: automotive applications. Here lies a steady, huge market with many millions of end consumers. And those consumers are willing to pay for the novel features of these microdevices, although they will never actually see the devices. The tiny chips are well hidden within and distributed throughout the vehicle. Yet it is obvious to everybody that two essential properties can be enhanced by sensor technology: safety and environmental protection. Antiskid systems and the air bag would be unthinkable without highly sophisticated sensors. The well-controlled management of engine combustion has been a blessing for reducing exhaust reductions; legal authorities recognized the technical capabilities and responded by making stringent requirements, thus reinforcing the development and mass production of sensors.

Modern sensors are remarkable in many ways. Their small dimensions open up new areas of mechanics, flow control, friction, and oscillation. Force measurements are just one example. The once somewhat obscure classical Coriolis force is now the principle means of sensing rotation. And the even more obscure miniscale quantum-mechanical Casimir force, arising between two close interfaces, is now also accessible to sensor structures. Sensitivities are astonishing even now, but will most probably continue to be enhanced. Very many external parameters, such as temperature, pressure, and electromagnetic fields, can be accurately and quickly measured. What a wonderful area of activity for physicists, chemists, engineers – and salespeople alike! The prospect of protecting humankind as well as the environment is gratifying.

This volume contains a wealth of information on all facets of sensors in the automobile. I feel privileged and proud to have been approached by the two editors, Jiri Marek and Hans-Peter Trah, both former doctoral students of mine, to contribute this preface. Thanks to the entire international team of contributors and hopes for the successful use of its information by all its readers!

Stuttgart, April 2003

HANS-JOACHIM QUEISSER

Contents

Preface to the Series V

Foreword VII

Abbreviations XIII

List of Contributors XVII

1	Overview	1
	<i>Jiri Marek and Hans-Peter Trah</i>	
1.1	Introduction	1
2	Automotive Sensor Market	5
	<i>Karl-Franz Reinhart and Matthias Illing</i>	
2.1	Introduction	5
2.2	Automotive Sensor Market Overview	6
2.3	Incentives and Current Trends in Automotive Sensors	9
2.4	Regional Trends	11
2.5	Market for Individual Sensors	12
2.6	Outlook: New Sensors and Market Opportunities	16
2.7	Summary	19
2.8	References	20
3	Measurement Principles: Basic Considerations about Sensing	21
	<i>Hans-Peter Trah, Kersten Kehr and Roland Müller-Fiedler</i>	
3.1	Basic Considerations and Definitions	22
3.2	Classification of Sensor-Input Signals	24
3.3	Signal Conversion and Signal Extraction (Signal Path)	28
3.4	Sensor Output Signals	32
3.5	Summary and Outlook	36
3.6	References	37

4	Design Methodology	39
4.1	Methodology	39
	<i>Reinhard Neul and Hans-Reiner Krauss</i>	
4.2	CAD Tools for MEMS	58
	<i>Gunar Lorenz and Joe Repke</i>	
5	Technology	73
5.1	Bulk Micromachining	73
	<i>Daniel Lapadatu and Henrik Jakobsen</i>	
5.2	Integrated Surface Micromachining Technology	92
	<i>Bob Sulouff</i>	
5.3	Surface Micromachining – Discrete	103
	<i>Jan Peter Stadler, Michael Offenberger, and Franz Lärmer</i>	
5.4	Thin Films on Steel	123
	<i>Stefan Finkbeiner</i>	
5.5	Thin Films on Silicon	141
	<i>Matthias Fürtsch and Stefan Finkbeiner</i>	
5.6	Ceramic Materials and Technologies for Exhaust Gas Sensors	159
	<i>Harald Neumann, Hans-Martin Wiedenmann and Johann Riegel</i>	
5.7	Magnetic-Field Sensor Technologies	172
	<i>Henrik Siegle</i>	
5.8	Assembly, Packaging	191
	<i>Kurt Weiblen and Jan Peter Stadler</i>	
5.9	Reliability	204
	<i>Roland Müller-Fiedler</i>	
5.10	Test of Automotive Sensors	224
	<i>Winfried Schönenborn, Frank Schatz and Gottfried Flik</i>	
6	Evaluation Circuits	237
6.1	Capacitive Evaluation Circuits	237
	<i>Bernhard Boser</i>	
6.2	Resistive Evaluation Circuits	256
	<i>Rainer Kienzler and Oliver Schatz</i>	
7	Applications	269
7.1	Accelerometers for Automotive Applications	269
	<i>David Monk, Dragan Mladenovic and Mark Skaw</i>	
7.2	Yaw-Rate Sensors	297
	<i>Oliver Schatz</i>	
7.3	Pressure Sensors	314
	<i>Iwao Yokomori and Yasutoshi Suzuki</i>	
7.4	High-Pressure Sensors	333
	<i>Jörg Gebers</i>	
7.5	Temperature Sensors	343
	<i>Aldo Bojarski and Werner Fichte</i>	

- 7.6 Mass-Flow Sensors 360
Uwe Konzelmann
- 7.7 Radar Sensors 372
Peter Knoll
- 7.8 Video Sensors 386
Peter Knoll
- 7.9 Wheel-Speed Sensors 403
Matthias Mörbe and Günter Zwiener
- 7.10 Cam and Crank-Angle Sensors 416
Iwao Yokomori and Yasutoshi Suzuki
- 7.11 Steering-Angle Sensors 428
Peter Kofink
- 7.12 Force and Torque Sensors 450
Matthias Mörbe and Cord von Hörsten
- 7.13 Light Sensors 462
Iwao Yokomori and Yasutoshi Suzuki
- 7.14 Rain Sensors 474
Andreas Schneider
- 7.15 Chemical Sensors for Oxygen Detection and Air/Fuel Ratio
Control 480
Johann Riegel, Hans-Martin Wiedenmann and Harald Neumann
- 7.16 Chemical Sensors for Emission Control 500
Dietmar Schmitt
- 7.17 Chemical Sensors for Air Quality 509
Klaus Dieter Frers
- 7.18 Chemical Sensors for Liquid Media 516
Bernhard Jakoby and Falk Herrmann
- 7.19 Electric-Current Sensors 527
Henrik Siegle
- 7.20 Tire-Pressure Sensors 536
Norbert Normann, Gunter Schulze and Wolfgang Keller

Subject Index 545

Abbreviations

A/D	analogue to digital
ABC	active body control
ABS	antilock braking system
ac	alternating current
ACC	adaptive cruise control
AFS	active front & rear steering
APS	active pixel sensor
ASIC	application-specific integrated circuit
ASC	anti skid control
AQS	air-quality sensor
BBP	bridge base potential
BCD	body control damping
BITE	built-in test
BSI	bit synchronous interface
CAN	control area network
CCD	charge coupled device
CEA-LETI	le Commissariat à l'Energie Atomique – le Laboratoire d'Electronique, de Technologie et d'Instrumentation
CMOS	complementary metal-oxide semiconductor
C-V	capacitance to voltage conversion
CW	continuous wave
dB	decibel
dc	direct current
DLI	distributor-less ignition
DR	dynamic range
DRL	daylight running lamp
DSP	digital signal processing
ECU	electronic control unit
EHB	electrohydraulic braking
EMC	electromagnetic compatibility
EMI	electromagnetic interference
EPROM	erasable programmable read-only memory
EPS	electric power steering

ESD	electrostatic discharge
ESP	electronic stability program
FEM	finite element method
FIR	far infrared
FMCW	frequency modulated continuous wave
FSD	full-scale deflection
GDI	gasoline direct-injection
GMR	giant magnetoresistance
GPS	global positioning system
HUD	head-up display
IC	integrated circuit
L	litres
LDW	lane-departure warning
LIN	local interconnect network
LPCVD	low-pressure chemical vapor deposition
LRR	long-range radar
MEMS	microelectrical mechanical systems
μ C	microcontroller
MI	magnetoimpedance
MOS	metal-oxide semiconductor
MPU	magnetic pick-up coil
MR	magnetoresistive effect
MRE	ferromagnetoresistive element
MST	micro system technology
NIR	near infrared
OBD	on-board diagnosis
PAS	peripheral acceleration sensor connection
PD	photodiode
PID	proportional integral differential
PLCC	plastic leadless chip carrier
ppm	parts per million
PROM	programmable read-only memory
RIE	reactive ion etching
RME	rapeseed oil methyl ester
rms	root mean square
ROSE	roll-over sensing
SAS	steering-angle sensor
SbW	steer-by-wire
SMD	surface mounted device
SNR	signal-to-noise ratio
SOI	silicon on insulator
SPI	serial peripheral interface
SRR	short-range radar
TCM	top (of steering) column module
TCO	temperature coefficient of offset

TCS	temperature coefficient of sensitivity
TFA	thin film on ASIC
TMR	tunnel magnetoresistance
TWC	three-way catalyst
VDC	vehicle dynamics control
WSS	wheel-speed sensors

List of Contributors

ALDO BOJARSKI
 AB Elektronik Sachsen GmbH
 Salzstr. 3
 01738 Klingenberg
 Germany
 Aldo.Bojarski@ab-sachsen.de

Prof. BERNHARD BOSER
 University of California at Berkeley
 Department of Electrical Engineering
 and Computer Science
 572 Cory Hall
 Berkeley, CA 94720-1770
 USA
 boser@eecs.berkeley.edu

WERNER FICHTE
 AB Elektronik Sachsen GmbH
 Salzstr. 3
 01738 Klingenberg
 Germany
 Werner.Fichte@ab-sachsen.de

DR. STEFAN FINKBEINER
 Robert Bosch GmbH
 AE/STZ2
 Postfach 13 42
 72703 Reutlingen
 Germany
 stefan.finkbeiner@de.bosch.com

Dr. GOTTFRIED FLIK
 Robert Bosch GmbH
 FV/FLD
 Postfach 10 60 50
 70049 Stuttgart
 Germany
 gottfried.flik@de.bosch.com

KLAUS DIETER FRERS
 Paragon AG
 Schwalbenweg 29
 33129 Delbrück
 Germany
 klaus.frers@paragon-online.de

Dr. MATTHIAS FÜRTSCH
 Robert Bosch GmbH
 AE/SPP2
 Postfach 13 42
 72703 Reutlingen
 Germany
 matthias.fuertsch@de.bosch.com

JÖRG GEBERS
 Robert Bosch GmbH
 GS-SE/ENG-G
 Postfach 30 02 20
 70442 Stuttgart
 Germany
 joerg.gebers@de.bosch.com

DR. MATTHIAS ILLING
Robert Bosch GmbH
AE/PMS
Postfach 13 42
72703 Reutlingen
Germany
matthias.illing@de.bosch.com

Prof. Dr. HENRIK JAKOBSEN
Senso Nor asa
Boks 196
3192 Horten
Norway
henrik.jakobsen@sensor.no

Prof. Dr. BERNHARD JAKOBY
Institut für Industrielle Elektronik
und Materialwissenschaften
TU Wien
Gusshausstr. 27/366
1040 Wien
Austria

DR. KERSTEN KEHR
Robert Bosch GmbH
FV/FLD
Postfach 10 60 50
70049 Stuttgart
Germany
kersten.kehr@de.bosch.com

WOLFGANG KELLER
Beru Electronics GmbH
General Manager
Gewerbestraße 40
75015 Bretten-Gölshausen
Germany
Wolfgang.Keller@beru.de

DR. RAINER KIENZLER
Robert Bosch GmbH
AE/EIC
Postfach 13 42
72703 Reutlingen
Germany
rainer.kienzler@de.bosch.com

Prof. Dr. PETER KNOLL
Robert Bosch GmbH
AE-RU/EL
Daimlerstr. 6
71229 Leonberg
Germany
peter.knoll@de.bosch.com

PETER KOFINK
Valeo Schalter und Sensoren GmbH
Stuttgarter Str. 119
74321 Bietigheim-Bissingen
Germany
peter.kofink@valeo.com

DR. UWE KONZELMANN
Robert Bosch GmbH
GS-SE/ENG
Postfach 30 02 20
70442 Stuttgart
Germany
uwe.konzelmann@de.bosch.com

DR. HANS-REINER KRAUSS
Robert Bosch GmbH
AE/STZ4
Postfach 13 42
72703 Reutlingen
Germany
hans-reiner.krauss@de.bosch.com

DR. FRANZ LÄRMER
Robert Bosch GmbH
FV/FLD
Robert-Bosch-Platz 1
70839 Gerlingen
Germany
franz.laermer@de.bosch.com

GERD LORENZ
Robert Bosch GmbH
AE/SPP5
Postfach 13 42
72703 Reutlingen
Germany
gerd.lorenz2@de.bosch.com

Dr. GUNAR LORENZ
Coventor, Inc.
3, Avenue de Quebec
91951 Courtaboeuf Cedex
France
gunar.lorenz@coventor.com

Dr. JIRI MAREK
Robert Bosch GmbH, AE/EL4
Postfach 13 42
72703 Reutlingen
Germany
jiri.marek@de.bosch.com

DRAGAN MLADENOVIC
Motorola
Sensor Development Engineering
2100 East Elliot Road
Tempe, Arizona 85284
USA
dragan.mladenovic@motorola.com

Dr. DAVID J. MONK
Motorola
Chief Engineering Manager
Sensor Development Engineering
2100 East Elliot Road
Tempe, Arizona 85284
USA
dave.monk@motorola.com

MATTHIAS MÖRBE
Robert Bosch GmbH, CS-EM/ECS
Siemensstr. 31
71254 Ditzingen
Germany
matthias.moerbe@de.bosch.com

Dr. ROLAND MÜLLER-FIEDLER
Robert Bosch GmbH
FV/FLD
Postfach 10 60 50
70049 Stuttgart
Germany
roland.mueller-fiedler@de.bosch.com

Dr. REINHARD NEUL
Robert Bosch GmbH
FV/FLD
Postfach 10 60 50
70049 Stuttgart
Germany
reinhard.neul@de.bosch.com

Dr. HARALD NEUMANN
Robert Bosch GmbH
GS/ESV
Postfach 30 02 20
70442 Stuttgart
Germany
harald.neumann@de.bosch.com

NORBERT NORMANN
Beru Electronics GmbH
Gewerbestraße 40
75015 Bretten-Gölshausen
Germany
Norbert.Normann@beru.de

Prof. Dr. HANS-JOACHIM QUEISSER
Max-Planck-Institut
für Festkörperforschung
Heisenbergstr. 1
70569 Stuttgart
Germany
H.Queisser@fkf.mpg.de

Dr. KARL-FRANZ REINHART
Robert Bosch GmbH
FV/FLD
Postfach 10 60 50
70049 Stuttgart
Germany
karl-franz.reinhart@de.bosch.com

Dr. JOHANN RIEGEL
Robert Bosch GmbH
Postfach 30 02 20
70442 Stuttgart
Germany
johann.riegel@de.bosch.com

DR. FRANK SCHATZ
Robert Bosch GmbH
FV/FLD
Postfach 10 60 50
70049 Stuttgart
Germany
frank.schatz@de.bosch.com

OLIVER SCHATZ
Robert Bosch GmbH
AE/SPP
Postfach 13 42
72703 Reutlingen
Germany
oliver.schatz@de.bosch.com

DR. DIETMAR SCHMITT
NGK EUROPE GMBH
Westerbachstraße 32
61476 Kronberg im Taunus
Germany
dschmitt@ngk-e.de

ANDREAS SCHNEIDER
Robert Bosch GmbH
BE-WS/ETW 2
Postfach 11 63
77813 Bühl
Germany
andreas.schneider@de.bosch.com

WINFRIED SCHÖNENBORN
Robert Bosch GmbH
AE/STZ 4
Postfach 13 42
72703 Reutlingen
Germany
winfried.schoenenborn@de.bosch.com

GUNTER SCHULZE
Beru Electronics GmbH
Gewerbestraße 40
75015 Bretten-Gölshausen
Germany
Gunter.Schulze@beru.de

DR. HENRIK SIEGLE
Robert Bosch GmbH
FV/FLT
Postfach 10 60 50
70049 Stuttgart
Germany
henrik.siegle@de.bosch.com

MARK SKAW
Motorola
Sensor Development Engineering
2100 East Elliot Road
Tempe, Arizona 85284
USA
mark.skaw@motorola.com

BOB SULOUFF
Analog Devices Inc.
Micromachined Product Division
21 Osborn Street, MS-731
Cambridge, MA 02139-3556
USA

DR. JAN PETER STADLER
Robert Bosch GmbH
AE/STZ
Postfach 13 42
72703 Reutlingen
Germany
janpeter.stadler@de.bosch.com

YASUTOSHI SUZUKI
Denso Corporation
Manager of Research and Development
Sect. IC
1-1, Showacho, Kariya-shi
Aichi-ken, 448-8661
Japan
suzuki@ice2.denso.co.jp

DR. HANS-PETER TRAH
Robert Bosch GmbH, FV/FLD
Robert-Bosch-Platz 1
70839 Gerlingen
Germany
hans-peter.trah@de.bosch.com

KURT WEIBLEN
Robert Bosch GmbH
AE/STZ3
Postfach 13 42
72703 Reutlingen
Germany
kurt.weiblen@de.bosch.com

IWAO YOKOMORI
Denso Corporation
IC Eng. Dept. 2
1-1, Showacho, Kariya-shi
Aichi-ken, 448-8661
Japan
iwao_yokomori@mgtnet.denso.co.jp

FRANKLIN WEIGOLD
Analog Devices
Vice President and General Manager
Micromachined Product Division
21 Osborn Street, MS-731
Cambridge, MA 02139-3556
USA
franklin.weigold@analog.com

1

Overview

JIRI MAREK and HANS-PETER TRAH

According to market studies, the automotive market is the largest segment for sensors. At the same time, sensors are playing an important role in the performance of modern electronic systems in the automobile. In this book we would like to give you an overview of this field.

First (Chapter 1) we explain the automotive sensor market: the market size, important technical issues, business models, and major players are discussed. Current trends and the outlook for new sensors and market opportunities are reviewed.

The production of automotive sensors in high quantities at very low failure rates requires elaborate design methods and simulation tools, as well as improvements in production processes. The second part (Chapters 2–6) focuses on these design and manufacturing issues.

In the third part of the book (Chapter 7) we present the different sensor types and show the various fields of application, ranging from engine management to safety systems and all the way to comfort systems.

The purpose of the book is to give you an overview of the various sensors used in the car today, as well as of the technologies required to manufacture these components. We hope to give interesting insights into this exciting field.

1.1

Introduction

The use of electronic control systems in automobiles has grown rapidly in recent decades. Fifty years ago, most systems were based on (electro-)mechanical principles; today almost all systems are electronic. Without these systems, today's cars simply would not be operational.

The first electronic systems to enter the car were for engine management. In the 1960s electronic ignition systems became widespread, followed soon after by electronic fuel injection. To meet emission regulations, electronic engine management was implemented in the 1970s. Due to these systems' complexity, analog control was replaced by digital electronics soon after. On-board diagnostics are now required to assure the functionality of these complex systems over a car's lifetime. Today's developments focus on the creation of ULEV (ultra low-emission vehicles).

After power train control, the area of safety received a strong market push. In the 1980s airbag passenger-protection systems achieved high market penetration in response to legislation in the USA; Japan and Europe followed soon after. In European cars, the use of both side and front airbags has led to further improvements in passenger protection. Vehicle stability was first improved with ABS (antilock braking system) systems in the 1980s. ABS brakes dramatically improve vehicle performance on slippery or icy roads. In 1995, the electronic stability program (ESP) began to be produced in volume; by comparing the steering wheel's angle with the vehicle rotation rate, vehicle stability is improved due to individual braking of the four wheels. Driver-assistance systems using ultrasound, RADAR, LIDAR, video the will further improve passive and active safety in the future.

The third major area for automotive sensors is in comfort systems. Besides entertainment (radios, CD-players, cassette-players), climate-control systems, adaptive cruise control, electronic positioning of mirrors, seats, etc., as well as navigation systems, are gaining in popularity. Figure 1.1 shows the development of automotive electronics from an historical perspective.

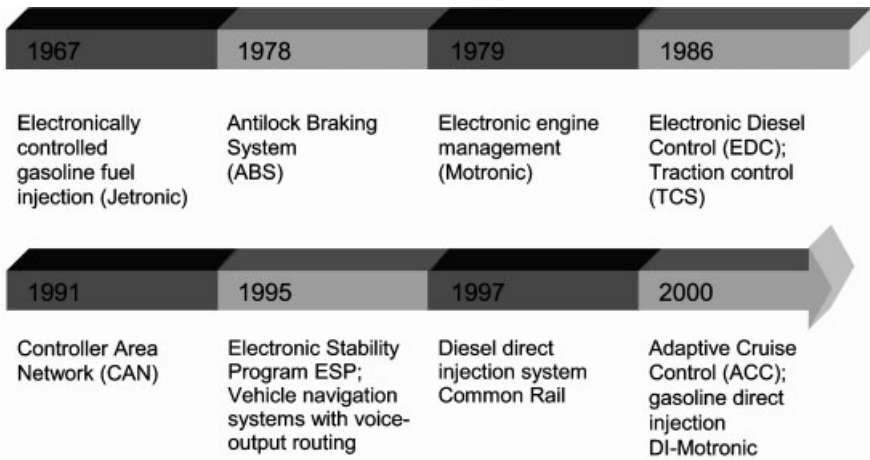


Fig. 1.1 Milestones in automotive electronic control systems

Electronic monitoring and control systems follow the same basic concept (Figure 1.2): environmental and engine parameters such as

- temperature of air intake
- pressure at air intake
- pressure of brake fluid
- air mass intake at the manifold
- fuel tank pressure
- chemical composition of the exhaust
- wheel speed

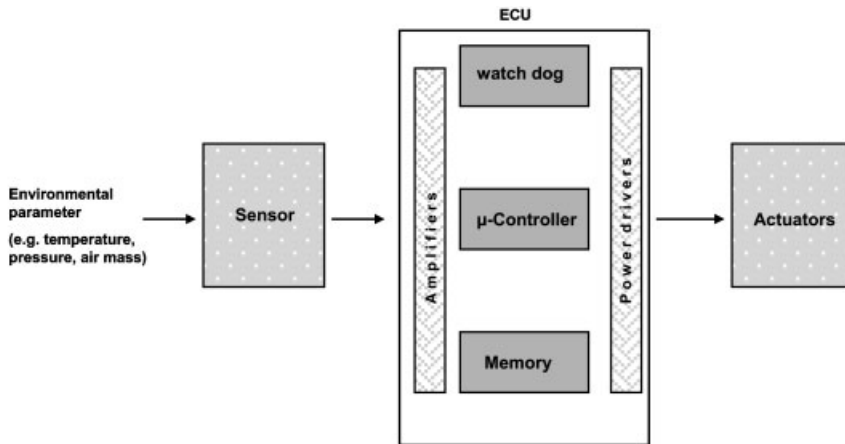


Fig. 1.2 Schematic principle of electronic control systems

are needed as input values for the control system. The appropriate sensor converts the mechanical or chemical parameter to an electrical signal. The electronic control unit (ECU) takes this value and performs the appropriate calculations using the software and algorithm embedded within the microcontroller. The output of the ECU then drives the actuators of the system, for example:

- injector valves for fuel injection
- spark ignition
- valves of the braking system
- motors for climate control

Automotive sensors need to work in a very demanding environment. Depending on the application, the sensor has to function in a temperature range of -40°C to 85°C (passenger compartment) or 125°C (engine compartment). Exhaust sensors must withstand temperatures up to 800°C . The vibration levels affecting sensors mounted on the engine can reach 100 times the gravitational force. Sensors also have to withstand exposure to gasoline, salt, detergents, and other chemicals. All these requirements have to be met at a very low price, as well as with a reliability of a few ppm!

The development of sensors for automobiles has generally followed progress in electronics. The first sensors used electromechanical working principles. In the past two decades, semiconductor-based technologies, like micro-machining, thin film technology, and hall/AMR (anisotropic magnetic resistance), took over. However, the second important type of sensors – chemical transducers – are still dominated by ceramic technologies, due to the high temperature requirements.

Major achievements in the development of automotive sensors are following the system roadmap. Often, however, innovations in sensor technology have been the enabling technology for making an automotive system economically feasible, and on several occasions a breakthrough on the sensor side decreased system costs significantly, leading to a high market penetration. Therefore, the question of whether sensor technology has pulled or pushed improvements in automotive sys-

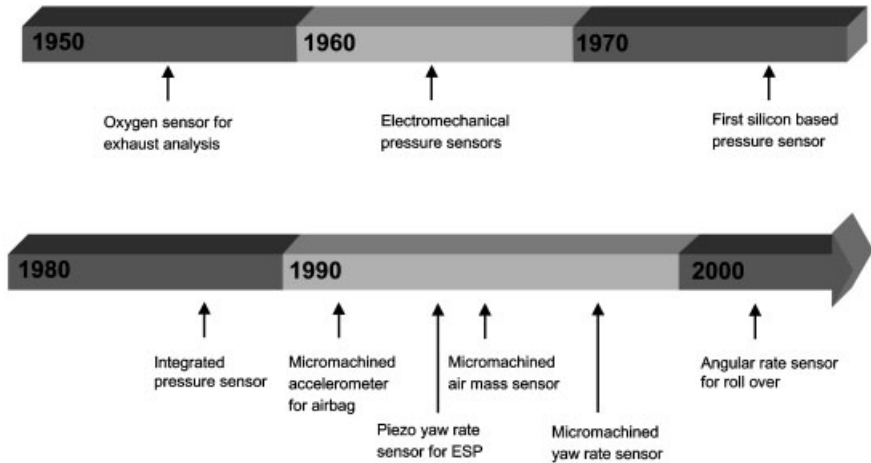


Fig. 1.3 Important milestones in automotive sensor development

tems has to be discussed case by case. Figure 1.3 shows the history of important developments in the field of sensors.

The first market pull came in the 1960s, when the demand for emission control necessitated sensors for measuring engine load. This measurement task was performed first with vortex air-flow meters and then with hot wire air-mass sensors. Micromachining technology was applied in this field fairly late; in 1995 a micromachined air-mass sensor began being produced in volume. The other sensing alternative, measuring the pressure at the manifold air intake, was also implemented. These pressure sensors were first based on an electromechanical principle and were later realized using silicon micromachining. The next step was the integration of the sensing element and the associated evaluation and trimming circuit into one device.

Airbag systems, first developed in the 1980s, have a critical need for accurate measurement of acceleration during a collision. In 1984 the first electromechanical accelerometer started in volume production; it used a spring-mass system and piezoresistive signal transformation. About five years later, the piezoelectric principle was applied using PZT (Pb-Zr-Ti) ceramics. In the beginning of the 1990s, the first micromachined accelerometer came to market.

The next safety system having very demanding sensor requirements was the electronic stability program. In 1995, development of the electromechanical yaw-rate sensor using piezoelectric excitation and pickup made the ESP possible. Only three years later, a micromachined yaw-rate sensor came to market; it both reduced manufacturing costs and enabled high market penetration of ESP. In 2001, the angular rate sensor started volume production, enabling passenger protection in rollover accidents.

The above sensors enabled the automotive industry to implement the very demanding systems of today's vehicles. Many more innovative automotive systems are being developed, and sensors will contribute significantly to the development and performance of these systems in the future.

2

Automotive Sensor Market

KARL-FRANZ REINHART and MATTHIAS ILLING

2.1

Introduction

According to the association of the German automotive industry (VDA), 55.6 million cars and trucks were manufactured worldwide in 2001 [1]. This figure will not change significantly in the coming years. The automotive market is rather mature – if not saturated – in most countries; average growth rates on the order of just 2% per year can be expected.

However, the picture is different with automotive components. Since the late 1970s, electronic systems have been increasingly used in the automobile. The relative value of electronic components in cars has increased from near zero in the 1970s to almost 20% in a 2000-model-year midclass car and even higher in luxury cars. In addition, electronics are responsible for the majority of innovations and thus contribute strongly to differentiating a car brand from its competitors.

Essentially all electronics systems in use – be it for engine management, in safety systems, or as convenience features – need one or more sensors as input to their signal processing. An overview by Fleming [2] counted 107 different sensor applications in the car; a luxury car typically contains 100 or more sensors. Therefore the automotive sensor market has grown at least as well as the electronics market over the past two decades and will continue to do so for the foreseeable future.

This article gives an overview of the automotive sensor market. It starts with a general overview of market size and growth and puts it into perspective with other relevant markets, such as the general sensor market and the electronics market. The major distinctive features of the automotive market are explained. Another section details some of the major trends in technologies and products and within geographical regions. Finally, various market segments are highlighted in more detail. The conclusion includes an outlook on possible future developments.

2.2

Automotive Sensor Market Overview

2.2.1

Market Size

Sensors are used in a multitude of markets and applications and perform an innumerable variety of monitoring and control functions. To put the automotive sensor market into perspective, we briefly mention other sensor markets here. According to a study from Intechno Consulting [3], sensor markets can be categorized as follows:

- machinery manufacturers and suppliers
- processing industries
- aircraft and ship building
- construction sector
- consumer and other electronics
- automotive
- others

According to the same study, in 2003 the overall sensor market will be \$42.2 billion, and the automotive sensor market will be about \$10.5 billion, or 25% of the overall sensor market. This is by far the largest of the market segments mentioned above. Taking into account the tendency of automotive sensors to be lower in cost than sensors in other segments (with the possible exception of the consumer electronics market), the large share of the automotive sensor market is even more pronounced if one looks at production volumes.

The relative importance of the automotive sensor market will continue to grow in the future. The average annual growth rate (CAGR) of the total sensor market from now until 2010 is estimated as 4–5%, and expectations for the automotive sensor market range from 5.1% [4] to 7.5% [3].

It is also interesting to compare the automotive sensor market with the automotive electronics market, since they share the same driving forces and thus have similar dynamics. In comparing market figures, we should be aware that sensors do contain a significant and increasing amount of electronics, so some double counting may occur in the literature.

According to a report of the German association of electrical engineers ZVEI, the automotive electronics market had a volume of \$14.9 billion in 2000 and should grow to \$23.9 billion in 2005. This is equivalent to a growth rate of 10% per year [5].

2.2.2

Technical Issues

As indicated before, about 55.6 million cars and light trucks are manufactured each year. In consequence, each broadly distributed sensor type has a market of

Tab. 2.1 Environmental conditions for automotive sensors

Chemicals and materials	Fuel Oil Water Salt spray Ozone Dust and sand Cleaning materials
Temperature	-40°C to 125°C
Vibration and shock	Up to several 100 times gravitational acceleration

about the same magnitude. Due to the high degree of standardization and the platform strategies of original equipment manufacturers (OEMs), typical yearly production volumes for a sensor variant start at about 100 000 units and can easily reach several million units per year. In summary, automotive sensors are almost always high-volume products, which sets this market apart from most other sensor markets, except for some consumer applications.

The cost pressure on all automotive components is high. As a result, low-cost, high-volume manufacturing processes are a necessary prerequisite for successfully introducing a sensor into vehicles. This is the reason why certain technologies, such as silicon-based sensors and microsystems, have been particularly successful recently (see Section 2.3).

The manufacturing processes need to be reliably controlled to meet high quality standards. To ensure meeting overall automotive quality goals (e.g., warranty periods of several years or 100 000 miles), the failure rates of sensors must not exceed 10 ppm over their lifetime. As the number of sensors is increasing, these failure limits will be tightened even more. Reliability is especially important in safety-critical applications such as airbags and braking systems. Additionally, in these applications fail-safe or failure-warning mechanisms must be designed into the sensor.

Both the technology and design of sensors thus need to be robust, specially in view of the typically harsh conditions in which sensors operate (Table 2.1). To summarize: automotive sensors have to be developed to aerospace specifications and manufactured at consumer prices.

2.2.3

Business Models and Players

The automotive sensor business is a typical OEM business; there is virtually no consumer aftermarket. The sensors are specified by either the car manufacturer or the tier-one supplier, depending on who has responsibility for a certain system. Car manufacturers and tier-one system suppliers form the customer base for the sensor supplier.

On the supply side, the market can be roughly grouped into three categories:

Tab. 2.2 Nonexhaustive list of major suppliers of sensors to the automotive industry

System suppliers	Denso, Delphi, Siemens, Bosch
Semiconductor manufacturers	Texas Instruments, Motorola, Infineon, Philips, Analog Devices
Sensor specialists	VTI-Hamlin, Sensoror, Murata, Matsushita, Systron-Donner, Hitachi, Kavlico, Nagano Keiki

- system suppliers with in-house sensor component manufacturing
- semiconductor manufacturers who also have a sensor business
- sensor specialists

No single sensor supplier offers all – or even nearly all – the 100+ sensor types mentioned above. A few suppliers, however, do offer a broad, comprehensive spectrum of products. Usually these are the system suppliers. Semiconductor suppliers and specialists typically focus on a narrow portfolio, often determined by the technologies or know-how available or necessary for a certain sensor type. As a result, the competitive landscape looks very different, depending on the particular sensor considered.

A (nonexhaustive) list of players is shown in Table 2.2.

In terms of business models, system suppliers generally develop and manufacture sensors to support their systems business (captive market). Customized sensors provide a technological and functional advantage to their system. However, due to intense price pressure in the system business, sensors have become an attractive market opportunity on their own. Thus, some systems houses have set up component businesses, sometimes targeting not only the automotive market but nonautomotive markets as well.

For semiconductor specialists, the automotive sensor market provides an extension to their traditional business, with the benefit that it is rather stable and not as cyclical as other semiconductor markets typically are.

The strategy of sensor specialists is to place themselves where specific technologies or application know-how is required.

Generally, the most important success factors in the automotive sensor market are:

- ability to combine mechanics and electronics using a variety of technologies
- application know-how
- long-term strategy: typical development times of 3–5 years or even longer when new technology is needed
- ability to make significant investments in R&D and manufacturing environment
- excellent quality management.

2.3

Incentives and Current Trends in Automotive Sensors

The applications of automotive sensors can be categorized into the areas of drivetrain, safety, and comfort/convenience. Driving forces for the use of a certain sensor may be market pull (consumer demand), technology push (OEM product differentiation by new feature), or legislation. In the drivetrain and safety areas all three factors are present. However, comfort and convenience applications are driven by market pull and technology push, with legislation usually playing a minor role. Table 2.3 gives an overview of applications and drivers for some of the major sensors used in the automobile.

The typical evolution of a sensor over its market lifetime is, first, it is introduced as a technology-push product in a high-end model produced by an OEM. Later, this state-of-the-art technology is recognized as a major improvement in terms of environmental impact or passenger safety and becomes required by law (e.g., airbags in the US). Or it is recognized by the customer as an important ‘must-have’ feature, for example, ESP in Europe. These systems and corresponding sensors are then introduced into all car classes and eventually become a high-volume product.

To be able to introduce sensors into all vehicle classes, it is absolutely necessary to manufacture them in high volume at very low cost. This requirement has promoted the development of microsystems technologies, which are often silicon-based technologies. The reason is that silicon technology is a batch technology that offers high volume capability as well as low cost. Developers of such processes benefit from more than 30 years of experience in the manufacturing of integrated circuits, as well as from already having production infrastructure. Sen-

Tab. 2.3 Driving forces for sensor use in various applications

	<i>Market pull</i>	<i>Technology push</i>	<i>Legislation</i>
Drivetrain	– pressure sensor (direct injection)	– pedal angle (drive-by-wire)	– lambda sensor – pressure (MAP) – airflow – tank pressure
Safety	– speed (ABS) – acceleration (airbag, Europe) – pressure – angular rate (ESP, Europe)	– pressure – angular rate (ESP, USA) – radar/video – surround sensing – acceleration (active suspension)	– acceleration (airbag, USA) – weight sensor (USA) – tire pressure (USA)
Comfort/ convenience	– angular rate (navigation)	– rain sensor (Europe) – light sensor – air quality (HVAC)	

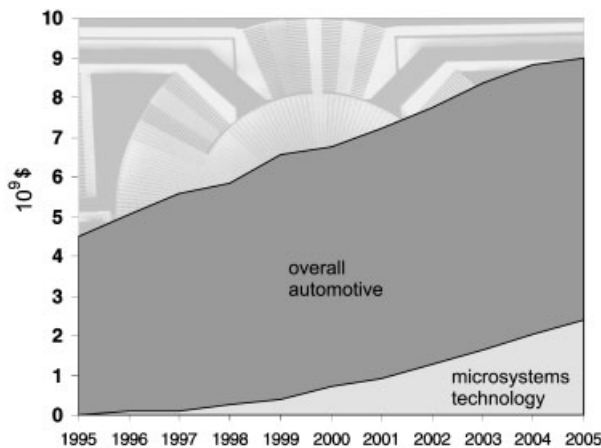


Fig. 2.1 Market evolution of automotive sensors. Silicon-based sensors are capturing an increasing fraction of the sensor market (adapted from 5])

sors that combine electrical and nonelectrical functions in one system are often called microelectromechanical systems (MEMS) or microsystems (MST). Generally, the following trends are supported by these microsystems technologies:

- *Miniaturization.* Typical sizes of functional elements in microsystem-based sensors are on the order of micrometers. A trend to continued shrinkage of feature sizes exists, similar to but not as pronounced as in microelectronics. The overall system size is meanwhile often determined by its packaging and the need for interconnections, rather than by the sensor element itself.
- *Cost reduction.* The requirement to introduce more sensors into all vehicle classes can be met only when the cost of the sensor decreases. A typical example is the acceleration sensor in airbag systems. In the early 1990s an acceleration sensor based on conventional technology cost about \$10. Consequently, only high-end cars were equipped with airbags, and side airbags were virtually nonexistent. Introduction of microsystems-based sensors in the mid-1990s offered the same functionality at prices as low as \$2. Today, airbags are standard equipment in all vehicle classes.
- *Increased functionality.* New manufacturing methods offer new opportunities for sensor functionality. Close integration with electronics enables 'smart' sensors; examples include data-preprocessing, continuous self-testing, and digital interfaces. Because of shrinkage of the components, various functionalities can be integrated into one sensor. Commonly combined functions include sensing temperature and pressure or acceleration and angular rate.
- *Quality.* Semiconductor processing techniques enable integration and thus a reduction in the numbers of elements and interconnections, resulting in improved quality. Furthermore, proven batch processes for quality assurance and testing of silicon-based sensors have been adopted from processes for integrated circuits.

The positive features of microsystems technology have led to considerable market success. Although the market for 'traditional' sensors is more or less leveling off, the microsystems sensor market is growing strongly, with predicted growth rates of about 20% per year. By 2005, sensors using these technologies are expected to capture almost 30% of the overall automotive sensor market (Figure 2.1).

2.4 Regional Trends

Certain regional differences exist. Traditionally, the North American market was a driving force for development of automotive sensors, as well as being the largest market. This trend was due to early US legislation, in particular, for reduction of emission reductions. In consequence, in the US, drivetrain-related sensors play an important role among the sensor types used in the US.

In recent years, the automotive sensor market in Europe has grown strongly and is now of equal importance to the US market. Emission-related legislation has now reached comparable standards as in the US. In addition, the European consumer has demanded safety-related features, many of which, such as ABS and the airbag (including side airbags), are now considered standard equipment.

For the immediate future, several regional trends can be seen. In the US, for example, occupant sensing is an important topic for passive safety systems. Legislation (taking effect in 2003) requires detection of the type of passenger and appropriate modulation of the activation of restraint systems (suppression or low-risk activation of airbags, for example). For this, new sensing systems are required, such as weight sensors, infrared sensors, or video detectors. However, which of these technologies will prevail is still unclear. Other sensors driven by the US market are tire-pressure sensors, for which legislation is currently (2002) being prepared. Another important feature may become rollover detection, using angular rate sensors for passive safety systems, which may be of particular interest for high center-of-gravity sports utility vehicles.

In Europe the ESP has been successfully introduced into the upper and middle vehicle classes. The penetration rate is expected to further increase, so that yaw rate sensors will experience strong growth rates. Another hot topic in the European market is direct-injection diesel (and, more recently, gasoline) engines, which require (among others) high-pressure sensors on the fuel rail and air-flow sensors in the air intake. Another trend is towards higher rates of installation of HVAC systems and new technology within these systems. These will require additional pressure, CO₂, and air-quality sensors.

In Japan, driving forces for sensor use are the high penetration of navigation systems. Another area in which Japan is taking the lead are telematics systems, for which additional sensors will be needed.

2.5

Market for Individual Sensors

This section describes the particular markets for major established automotive sensors. Market trends for new sensor types are described in Section 2.6.

2.5.1

Acceleration Sensors

The use of acceleration sensors has traditionally been driven by passive safety systems, such as airbags and seatbelt tensioners. These systems came onto the market in the early 1980s, with vehicles typically being equipped with one driver-side airbag and a corresponding acceleration sensor. Now, the airbag system in a typical new passenger car includes a front passenger airbag, side airbags on both sides, and window bags and contains 3–5 acceleration sensors.

Although airbags are still the main market for acceleration sensors, other applications require them as well:

- vehicle dynamics control systems
- antilock braking systems
- anti-theft systems
- active suspension control
- headlight leveling

The total market for acceleration sensors in 2000, according to NEXUS [7], is \$467 million. The market will grow to \$687 million by 2005. The main drivers for this growth are the increasing rates of equipment with side airbags. Active suspension control, which typically uses 3–5 sensors per system, has the potential to spur further growth if more vehicles are equipped with these systems in the future.

2.5.2

Pressure Sensors

Pressure sensors have numerous applications in the automobile. Historically, the first and still most important is to control the air intake. Other pressure-sensor applications are

- turbocharger pressure
- oil pressure
- atmospheric pressure
- fuel-tank pressure
- brake pressure
- climate control
- fuel pressure

The pressure ranges and environmental requirements differ from application to application; therefore, various technologies coexist: silicon micromachining and ceramic technologies are mostly used in the low-pressure range (1 kPa to 2 Mpa), and steel membranes are often used for higher pressures.

The market for pressure sensors is still growing, from \$700 million in 2000 to \$1350 million in 2005, according to a study by Strategy Analytics [4]. The main drivers are fuel-pressure sensors for direct injection engines and higher rates of equipment with advanced air conditioning systems. A strong push for pressure sensors is also coming from tire-pressure monitoring systems, which may well constitute the largest field of application for pressure sensors by the end of the decade.

2.5.3

Mass-Flow Sensors

Mass-flow sensors are used to measure air intake. Their advantage over pressure sensors in the same application is primarily higher accuracy. The growth of this sector is moderate and mainly due to increasingly strict environmental legislation.

A wide range of manufacturers use various technologies (heated wire, ceramics, silicon). The market for mass-flow sensors was estimated as about \$800 million in 2000.

2.5.4

Angular-Rate Sensors

Angular-rate sensors are relatively new in automotive applications. They were first used in 1995 in vehicle dynamics control systems. More recent applications include navigation systems and rollover systems. Since all these systems are still in a growth phase, these sensors have a large market potential in coming years. For the year 2000, the market is estimated to be approximately \$250 million, growing to more than \$700 million in 2005.

2.5.5

Rotational-Speed and Phase Sensors

Rotational-speed and phase sensors represent the largest market by unit of all sensor types. The most important applications are:

- wheel speed for ABS and vehicle dynamics control
- engine speed, camshaft, and crankshaft phase for motor control
- gear shaft speed for transmission control

Most sensors are based on magnetic field detection. Although previous generations of rotational-speed and phase sensors were passive sensors mainly using variable reluctance principles, current sensors are mostly active, based on the Hall-effect or anisotropic magnetoresistance, or even (in the future) on giant magnetoresistance technology, due to increasing performance requirements.

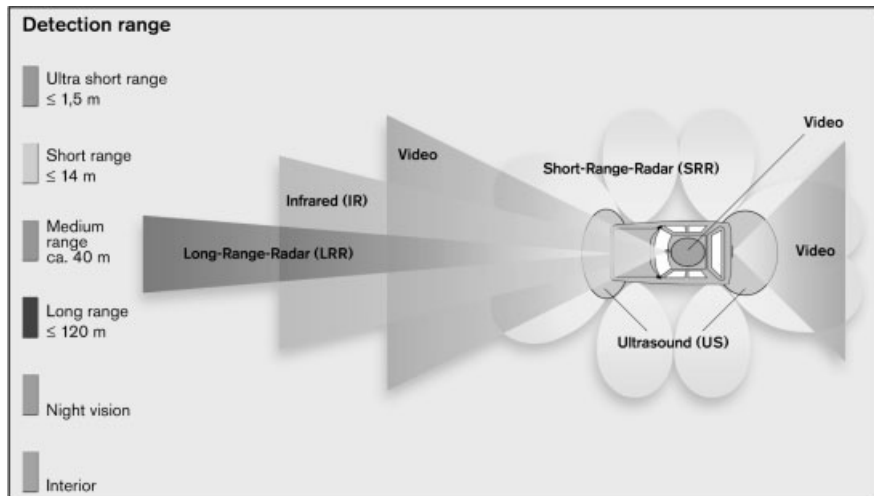


Fig. 2.2 Sensors for surround sensing

The total market for rotational-speed and phase sensors is estimated to be 310 million units in 2002, increasing to 390 million units in 2008, and representing a market value of more than \$1600 million. The total market is almost saturated, so new sensor types will replace the older, passive technology.

2.5.6

Angle and Position Sensors

Similar to rotational-speed sensors, angle and linear position sensors can be found in a wide variety of applications and systems. Sensor requirements and associated techniques are as manifold as these applications. Some of the most important sensors measure

- throttle valve angle
- pedal angle or position
- seat position
- gear lever position
- steering angle

Magnetic principles are used, as well as inductive, capacitive, and optical sensors. In many applications, where cost requirements predominate over performance and reliability issues, simple potentiometers are often found. The market volume in 2002 is about 150 million units and will approach 300 million units, worth more than \$1700 million, in 2008.

2.5.7

Exhaust-Gas Sensors

Pollution legislation continues driving further reductions in engine emissions. The dominant exhaust-gas sensors today and in the near future are oxygen partial pressure sensors – also called lambda sensors. Due to the high temperature of exhaust gas, these sensors are made by ceramics technology in combination with thick-film processing.

Only moderate growth is expected in the market for oxygen sensors, because most cars are already equipped with them. New after-treatment systems for exhaust gas, often include more than one oxygen sensor, for example, a lambda point sensor together with a broadband universal oxygen sensor. In the next few years, new concepts for decreasing emissions will increasingly lead to the integration of additional exhaust-gas sensors, such as temperature sensors or partial-pressure sensors sensitive to other exhaust gas components like NO_x or hydrocarbons. Particulate measurements for diesel engines may also result in new types of exhaust-gas sensors, either controlling the function of the particulate filter or being part of a motor component that controls the regeneration of a particulate filter.

Whether use of all of these sensors will become widespread in the near future depends on progress in developing new motor management and injection concepts, which is always struggling to meet the next step of exhaust-gas legislation without making use of expensive additional sensors.

Today's market for exhaust-gas oxygen sensors is about 90 million units, growing to more than 110 million units and to a market value of about \$1.5 billion by 2009. The new types of exhaust-gas sensors mentioned above are too new to allow estimating their market penetration in the next generations of systems with any certainty – nevertheless, these new sensor types will finally generate their market.

2.6

Outlook: New Sensors and Market Opportunities

Several new sensor types and applications, which can be expected to have significant market potential in the future, are presently in the R&D stage. For most, only qualitative forecasts for market relevance and opportunities can be given, because quantitative estimates in this early development stage suffer from several uncertainties. For example, competing solutions might emerge, or it may turn out that the new parameters can be indirectly derived from existing system data by the use of models. Some of the most interesting new sensor applications in our opinion are mentioned in the following sections.

2.6.1

Sensing the Surroundings for Comfort and Safety Systems**2.6.1.1 Radar and Video Sensors**

Radar and video systems for sensing the surroundings will create an attractive new sensor market in future vehicles and will supplement today's ultrasonic parking aids. The combination and (often) knowledge-based evaluation of various sensor signals as shown in figure 2.2 is a required characteristic for sensing the surroundings of a vehicle. The sensors are not simply detecting a single physical quantity, but, rather, they represent a sensing subsystem aiming to evaluate complex data, as in traffic situations.

A starting point for these new sensor functions is the adaptive cruise control system (ACC), which has already been introduced to the market. ACC systems monitor objects that are far in front of a vehicle and their relative speeds by either RF-radar (at 77 GHz) or laser-radar sensors. Due to the current high price of this system, the market will possibly be restricted mainly to luxury cars in the next few years. Volumes are expected to grow from several hundred thousand units in 2002 to a market of more than two million systems worldwide in 2005 [6].

Presently, short-range radar systems (SRR) are being developed by several suppliers. Their purpose is to monitor the close environs of the car not covered by ACC systems. Knowledge about nearby objects is required to enhance stop-and-go driving as well as to deal with complex situations in dense city traffic. Presently, short-range detection is done mostly with ultrasonic sensors used to aid parking. Radar sensors at frequencies of 24 GHz will increase the distance range. A system of sensors will allow a new diversity of functions in the fields of passive and active safety, driver support, and vehicle guidance. A significant market for short-range systems is not expected before the year 2005. However, in 2005 about 1 million shipments are expected, representing a considerable market value, due to the high cost of the monolithic microwave ICs (MMICs) to be used in SRR sensors.

Video sensing will develop as a complementary technique for sensing a vehicle's surroundings and can also be integrated with other sensing methods such as radar and ultrasound. Additionally, applications in the passenger compartment, to detect seat occupancy and occupant classification, are possible. Video sensing is encouraged by the considerably decreasing costs of semiconductor video imagers as they become widespread in consumer products and personal computers, as well as by new features of imager chips relating to logarithmic sensitivity and fast random access.

2.6.1.2 Night-vision Systems

Night-vision systems have already been introduced in very low volumes in some upscale automobiles. Their function is based on either infrared-sensitive array detectors or a CMOS camera combined with active near-infrared illumination of the scene. The night vision approach further completes the idea of a 'sensitive car'.

2.6.2

New Sensors for Drivetrain and Motor Control

2.6.2.1 Engine Combustion Sensing

Diesel and gasoline combustion engines can be optimized by using new concepts for controlling combustion in single cylinders. Several sensing principles are under development to monitor combustion, for example, sensing of cylinder pressure or ion current inside the cylinder, as well as more indirect methods, such as external piezoacoustic sensing. The most precise and direct measurement, however, is achieved by direct pressure sensing in the combustion chamber. Successful development of the corresponding concepts will enable mass production of pressure sensors that are both accurate and able to withstand the extreme environment in a combustion chamber throughout a car's lifetime. This would open a considerable new market, especially considering that one sensor is required for each cylinder. The emergence of combustion-pressure sensors, together with new engine-control concepts, would certainly have a great impact on the market for today's motor-control sensors, because some signals now provided by other sensors could also be obtained from the combustion pressure signal.

2.6.2.2 Sensing of Liquid Media Conditions

Another sensor type – the oil-condition sensor – to be integrated into motor management systems is already being introduced into the market. The frequency of routine maintenance tends to increase steadily with each new model year. Making sure that the motor oil remains in good condition for a longer time (e.g. 50000 km) forces the car makers to actually measure the oil's condition rather than estimate it from the miles driven and load data. Oil-condition sensors measure physical parameters such as the dielectric constant, viscosity, and temperature to assess the fluid's condition. By 2008, equipment rates as high as 50% can be expected. Engine oil-condition sensors are probably the first members of an emerging new class of fluid-condition sensors for monitoring fluids in general. Applications being investigated include monitoring, for example, fuel type and quality or hydraulic brake fluids. Such sensors could also be used to determine whether conditioners need to be added to an exhaust gas after-treatment reduction system.

2.6.3

New Sensors for Chassis Systems

The vision of braking and steering 'by wire' will demand new, extremely reliable sensors. Even in early implementations of steer-by-wire systems, in which manual control can override any system failure, more than one sensor is normally used for the sake of redundancy. Many of the sensor principles required are already established in the market, including steering-angle sensors (e.g., for vehicle dynamics control) and pedal-position sensors. Mechanical action or feedback control, however, will drive the emergence of torque and force sensors.

Steering-torque sensors will be needed in new electromechanical steering systems such as active power steering and will therefore have high annual market growth rates in the near future. Several technical solutions are feasible. Torque is measured either directly (e.g., by magnetoelastic means or with a strain gauge) or in a more indirect way, by measuring the torque angle of a torsion bar by optical, magnetic, or inductive means. The total market for torque sensors in general can be estimated as approximately 30 million units in 2008. Steering sensors will be the predominant application of torque sensors during the next few years. The market for other torque sensors, such as those intended for drivetrain applications, will be smaller and grow more slowly.

2.6.4

Other Interesting Future Automotive Sensor Markets

Several other sensors not described above will enter the market within the next decade, but cannot be discussed here in detail. Just to mention a few:

- Biometric sensors for driver identification (mainly fingerprint for passive enter-and-go systems or for personalized comfort control).
- Gas sensors (e.g., for sensing air quality, for measuring carbon dioxide in new types of air conditioning systems, for detecting ozone or fuel-cell gases such as H₂).
- Telemetric sensors (active or passive).
- Electric current sensors (potential-free sensors, e.g., for alternators, electric motors, and starters or for battery power management in future 42-V systems).
- Several new sensors might emerge to be integrated into mechanical components, in which the mechanics are part of the signal-generation path but are not produced by the sensor manufacturer. The premise is that business barriers for market entrance, such as product responsibility, or interfacing issues, such as sensor element testing and adjustment procedures, can be removed. Examples include integrated tire sensors or torque and force sensors in gear shafts or on wheel bearings.

2.6.5

Multiple Sensor Use, Sensor Platforms, and Sensor Fusion

Presently, automotive sensors are mostly dedicated solely to a particular system. Future sensors will have to be designed as multiple-use sensors for incorporation into different automotive systems. Each sensor then either has to include a bus controller or will be combined – where possible – with other sensors into an intelligent sensor-subsystem that performs a certain level of information preprocessing, for example, an inertial sensor cluster that delivers vehicle dynamics data to the bus. Particularly in the field of inertial sensors will such sensor subsystems emerge in the future – this trend is already being seen in the emergence of multi-axis inertial sensors.

Multiple use of sensors does not necessarily mean that fewer sensor elements are required, because overall sensor performance has to be improved at the same time. The strong demand for fail-safe sensors in safety-related applications will lead to systems using additional sensor signals for increased reliability and based on partial redundancy or on systems to crosscheck the plausibility of the individual signals. Another way of increasing a sensor's accuracy is by using information from multiple sensors.

Sensor fusion and multiple-sensor use will certainly change the long-term sensor market because even more electronics for preprocessing and bus compatibility will be integrated and also because the sensors will become standardized and more easily interchangeable.

2.7

Summary

The automotive sensor market is an attractive and still-growing multibillion-dollar market characterized by very high production volumes of sensors that must be extremely reliable and low in cost. Microsystem sensor technologies meet these needs and are thus widely accepted in automotive sensor applications. Growth in the sensor market continues to be driven by increasing demands for more safety and comfort and greater reductions in emissions. Existing sensors will continue to find new applications, and new sensors will emerge to improve functionality.

2.8

References

- 1 Verband der Automobilindustrie, <http://www.vda.de> (accessed Oct 2002).
- 2 FLEMING, W.J. *IEEE Sensors Journal* **2001**, *1*, 296–308.
- 3 Intechno Consulting Basel, *Sensor Markets 2008, 1999*. See also Intechno Consulting, <http://www.intechnoconsulting.com/> (accessed Oct 2002).
- 4 Strategy Analytics Boston, *Automotive Sensor Demand 1999 –2008*, **2001**.
- 5 Zentralverband Elektrotechnik und Elektronik Industrie, Frankfurt, *Mikroelektronik-Trendanalyse bis 2005*, **2001**.
- 6 Strategy Analytics Boston, *Automotive Radar Systems*, **2001**.
- 7 Nexus!, *Market Analysis for Microsystems II*, **2002**.

3

Measurement Principles: Basic Considerations about Sensing

HANS-PETER TRAH, KERSTEN KEHR and ROLAND MÜLLER-FIEDLER

Sensing is of essential importance for all kinds of systems – whether they are natural like the human body or artificial like the automobile. Without sensing, necessary information about the status of any process (the surrounding world) could not be provided and, accordingly, it would be impossible to respond to these processes – or at least it would be difficult to do so in a sensitive or intelligent way. Sensing, however, is rarely recognized as a special procedure, since human beings take it for granted that they can handle billions of bits of information every second without even needing to be aware of it. For technical systems however, the state of the art of sensing and controlling is orders of magnitude below what can be achieved with natural intelligence.

Nevertheless, more and more systems within the automobile have started to acquire some ‘intelligence’: there is plenty of sensor information and sensor signals, as well as fast and sophisticated algorithms within the systems’ electronic control units. But the output of the algorithms can only be as conclusive as the signals provided by the sensors, which inform the control units about actual conditions within and around the system. The meaning of *system* is twofold: On the one hand, it comprises a higher-level system within the car, which includes the sensor as a component that provides the necessary information for controlling the system. On the other hand, the sensor itself may be treated as a special type of transducer, and a transducer can be treated as a special type of system. Therefore, the respective interpretation of the terms ‘sensor input signal’ and ‘sensor output signal’ depends very much on the domain to which these signals can be assigned to.

With this background, it is worth considering the fundamentals of sensing, signal extraction, and signal conversion in more detail. This is the subject of this chapter. After a short introduction, the topic is divided into three major parts:

- The nature of the input signals (physical classifications and consequences).
- The means of signal transformation (appropriate physical conversion processes).
- The type of output signals (accuracy estimations).

Finally, a summary and an outlook are presented.

For scientific understanding, we can get information only from what can be measured. The first question of interest has to be: “What can be measured?” But

questions like ‘how?’, ‘how accurate?’, and ‘how reliable?’ are also extremely important for selecting measuring principles for technical systems. Other chapters in this book give evidence that these questions are of extraordinary importance, especially in automotive applications, because several challenging requirements must be met simultaneously:

- high reliability under harsh environmental conditions
- high accuracy
- low weight, volume, and power consumption
- high production volumes.
- low cost

These requirements have been summarized as “automotive sensors have to be developed to aerospace specifications and manufactured at consumer prices” [1]. To meet this challenge it is essential to consider one more question: “Which approach is best for measuring a certain type of signal?” In this chapter, we deduce an answer to this question for the specific boundary conditions of automotive sensors.

3.1

Basic Considerations and Definitions

The surrounding world may be divided into three domains [2, 3]:

- matter
- energy
- information

The relationships between these domains are described by physical laws. In consequence, ‘construction guidelines’ for technical systems make use of these domain definitions for preliminary system approaches. The resulting overall approach to the principle of operation of a system is depicted in Figure 3.1.

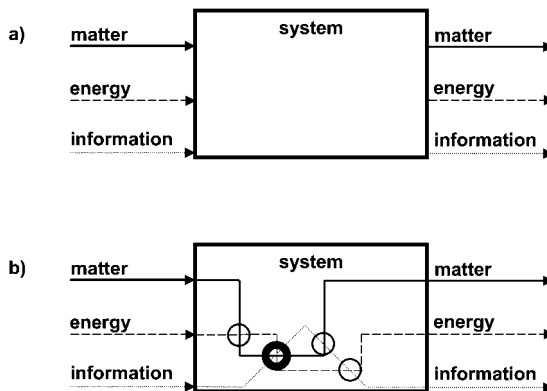


Fig. 3.1 a) Schematic view of a system with general representation of occurring domain fluxes, b) Interaction points of domain fluxes

Here, a system is shown without any concrete specifications; nevertheless, some fundamental conclusions may be derived from the laws of conservation. These laws – valid only for thermodynamic equilibrium *within* a system – have been stated for matter and energy as the laws of mass conservation and energy conservation and can be described by the following equations:

$$m = \Sigma m_i = \text{constant} \quad (3.1)$$

$$E = \Sigma E_i = \text{constant} \quad (3.2)$$

where m is the mass (matter) and E is the energy.

Concerning information, however, the situation is more complicated: Everyone ‘knows’ that sending information from one point to another appears to double that information, because it is not lost at either the source or the point of delivery. So the question is: “Is there also a law of conservation for information?” Nowadays, of course, it is well known that information may be included in the framework of energy conservation, as shown by the thermodynamic theory of information [4, 5]. Information may be described by an equation such as

$$I = \text{constant} \times (-\Delta S) \quad (3.3)$$

where I is the information, and S is the entropy (S is proportional to the logarithm of the probability finding the system in a given macrostate). This equation implies that information may be explained as a decrease in entropy, or an increase in order. Because entropy in thermodynamic theory is strongly related to the energy domain and can be decreased only by providing external energy (combination of the first and second laws of thermodynamics; that is, a nonequilibrium system), the mass-energy-information loop can be closed (Eqs. 3.1–3.3): information must always be connected to the domain of energy (or internal energy, when stored within matter). In other words: information and signals are always part of an energy domain – they are ‘carried’ by energy.

In this context, it is important to distinguish between ‘information’ and ‘signal’: *information* comprises all kinds of changes – even unwanted ones – connected with the domains of energy and matter, whereas *signal*, in the special sense of sensor input or output signal, includes only that part of the information that is of interest for the desired measurement application. In other words, signal formation or signal extraction (sensing, measuring) is a filtering process for information in which the signal of interest is selected from the entire information carried by the different energy domains involved.

Therefore, signals can be sent only if changes in energy and/or matter are involved. Consequently, Figure 3.1a should be rendered more precisely as shown in Figure 3.1b. As a fundamental implication of these indispensable changes, signal conversion requires the existence of ‘domain-flux crossings’ or ‘interaction points’ between information, energy, and/or matter [6]. These interaction points represent the physical components of transducer devices, where signals are converted with-

in or among the domains. In Figure 3.1b the domain-flux crossings are indicated by bold circles; table 3.1 shows typical examples of sensors, where signal conversion by changes in energy is involved. On the other hand, however, all kinds of interference can occur at these interaction points (light circles in Fig. 3.1b), and in fact, exclusively at these points.

3.2

Classification of Sensor-Input Signals

Now the question “Which approach is best for measuring a certain type of signal?” can be expressed in concrete terms:

- Which kinds of interactions between signal, matter, and energy should be preferred when selecting a transducer principle?
- How should the corresponding physical component be designed for optimum signal conversion with respect to all the above requirements?

Asking these questions may pose a new challenge for the designer, because there is a huge variety of possible technological approaches, covering a broad range of physical domains (mechanical, electrical, thermal, magnetic etc.). To provide an orientation guide within this maze, a useful approach may be to classify systems by using sets of ‘intensive’ and ‘extensive’ thermodynamic parameters. An additional, ‘inertial’, class of sensor input parameters can also be defined. So we first have to take a detailed look at the specific features of these input parameters. This is done in the following sections [see also 7].

3.2.1

Relations between Intensive and Extensive Signals

Certain relations between intensive and extensive signals are of fundamental significance for system (or transducer) design. According to [3], these relations can be summarized as follows:

$$x_i \times x_e \sim E_{\text{internal}}^{\mathcal{G}} \quad (3.4)$$

$$\text{grad}x_i \sim F^{\mathcal{G}} \quad (3.5)$$

$$dx_e/dt \sim I^{\mathcal{G}} \quad (3.6)$$

where x_n ($n = i, e$) are the corresponding intensive or extensive parameters, respectively, $E_{\text{internal}}^{\mathcal{G}}$ denotes the generalized internal energy stored in the matter involved, $F^{\mathcal{G}}$ is a generalized force, and $I^{\mathcal{G}}$ a generalized current. Examples of these relationships can be found in all physical theories: the internal energy of an elastic material is proportional to the product of stress (intensive) and strain (extensive), the internal energy of a gas is proportional to the product of pressure (intensive) and volume (extensive) or to the product of temperature (intensive) and the number of moles (extensive), etc.

3.2.2

Intensive Sensor-Input Signals

Intensive signals behave in a nonadditive manner, as is true for pressure and temperature. When the size of a system is increased, the signal value remains constant and does not scale with system dimensions. Intensive signals can therefore be measured within small parts of a system – but only when it is at (local) thermodynamic equilibrium – and so miniaturization is limited only by resolution aspects and by mounting requirements.

To measure intensive signals such as pressure, temperature, or chemical composition it is often (but not always, as shown below) necessary to bring the sensing element into direct contact with the medium to be measured, that is, the matter carrying the energy or signal. If the sensor needs to operate in a harsh environment or at very high temperatures, reliability requirements usually demand that the electrical or electronic part of the sensor be separated from the mechanical, fluidic, or high-temperature part.

As described in detail in Chapter 5.1, microsystem technologies commonly use silicon wafers as source material for the manufacturing of sensors. Bulk micromachining affects both sides of a silicon wafer and thereby facilitates separating the electronic and mechanical functions, which is not possible with surface micromachining, because of simple topological or reliability requirements. Every sensor must be mounted in a package, that among other things provides the electrical interconnection and protection against environmental influences but on the other hand may unintentionally interact with the signal conversion and thereby impair the sensor's functionality.

Most sensors for intensive signals use resistive (piezoresistive, thermoresistive, chemoresistive) sensing principles because they can be easily implemented either in bulk or in surface micromachining. However, resistive signal formation cannot be reversed for sensor self-test. So it is not possible to monitor the functionality of a sensor by applying an electric current and measuring a resulting mechanical, thermal or chemical reaction. For these sensors, testability in the sense of monitoring the functionality during operation is restricted considerably and can only be accomplished using sophisticated and costly measures.

3.2.3

Extensive Sensor-Input Signals

Extensive signals behave in an additive manner. In consequence, it is not possible to measure extensive signals such as mass or volume in small parts of a system, and therefore sensors for extensive signals cannot easily be miniaturized. For sensors based on microtechnologies, which are normally preferred for mass production due to cost considerations, extensive signals are difficult to handle because there is often no simple level of standardization for batch-oriented fabrication processes. How this problem may be solved is shown in the next section.

3.2.4

Conversion of Extensive Signals to Intensive Signals

If an extensive signal is to be measured with a low-cost microdevice that can be batch-manufactured, the signal has to first be converted into an intensive signal. For example, the mass flow (extensive) in a tube may be represented by the mass-flow density (intensive) at a small region within the tube. To ensure correct correlation between mass-flow density and mass flow, however, the tube itself becomes an essential part of the measuring setup. Consequently, such sensors are often not testable before their manufacturing process is entirely complete. The stage of production at which a sensor can be tested is crucial, because it can be ten times cheaper to discard a sensor chip than to discard the complete sensor.

3.2.5

Inertial Sensor-Input Signals

Inertial signals have to be treated separately. Newton's second law states:

$$F = m \times a \quad (3.7)$$

That is, in an accelerating system, where a is the acceleration and m is the mass, a force F acts on that mass. The result of this force – typically a displacement of the mass – is usually detected to measure acceleration or yaw rate. Although force is an extensive parameter, the devices used to detect an inertial signal behave differently than sensors for extensive quantities, because the forces can be adjusted by designing the mass m in accord with Eq. 3.7.

Inertial signals always act on the total system. The signal has the same magnitude at every place, even in very small parts of the system. The sensing elements for measuring inertial signals can therefore be miniaturized, and the degree of miniaturization is limited only by the resolution requirements for the desired application. Additionally, the sensing element does not need to be in direct contact with any medium, because inertial signals act through the walls of a system. Since there is no media contact, it is not necessary to fall back on the robust and intensive structures of bulk micromachining. Rather surface micromachining has become the most successful technology for high-volume production of inertial sensors. In addition, sensors for inertial signals – if they can be small enough – can be mounted in standard low-cost packages that are also used in conventional IC technology.

Most modern inertial sensors use capacitive sensing principles, because they can be easily manufactured by surface micromachining. Therefore, most inertial sensors are more-or-less testable at the wafer-scale level, because capacitive sensing may be reversed to electrostatic actuation so as to incorporate built-in testing functions. The possibility of using standard packaging processes, which normally show little influence on the signal path for capacitive sensors, leads to an optimum situation for microsensor devices: inertial sensing elements can be treated

almost like standard integrated circuits. One difference, of course, remains: for nonmonolithically integrated sensing elements, two chips have to be mounted in the same package instead of one.

3.2.6

Section Summary

The physical nature of sensor input signals (i.e., intensive, extensive, inertial) is of major importance for many fundamental questions concerning the choice of an appropriate transducer principle. Most issues related to

- appropriate technology for reliable, low-cost devices
- miniaturization
- modularization and scalability
- signal conversion
- testability

may be directly deduced from the type of physical signal, as summarized in Table 3.1.

Tab. 3.1 Physical categories of sensor input signals together with the consequences for miniaturization, modularity, scalability, standardization, and testability. *Sensing element* means the sensing device (i.e., a silicon chip on a wafer); *sensor module* means a sensing element already connected to the surrounding world at least to some extent, where the primary input signal may be supplied to the sensing element; *sensor* means the final product (completely packaged and ready for delivery to the customer).

Category of input signals	<i>Inertial</i>	<i>Intensive</i>	<i>Extensive</i> → <i>Intensive</i>	<i>Extensive</i>
Miniaturization of the sensing element	easy	easy	possible in principle	sometimes
Miniaturization of the sensor	easy	possible in principle	difficult	difficult
Potential degree of standardization	easy	possible in principle	difficult	difficult
Optimum level for product standardization	sensor	sensor module	sensing element	sensing element
Adjustment of measuring range (scalability, modularization)	sensing element	sensing element, sensor module	sensing element, sensor module	sensor
Most common method of signal conversion	capacitive	resistive	resistive	resistive
Testability at low level of added value	easy	possible in principle	difficult	difficult

3.3

Signal Conversion and Signal Extraction (Signal Path)

For the formation of an output signal, some kind of interaction between the input signal carrying the information of interest and a construction element or device made of matter is necessary [6]. The signal itself may be carried by energy of one of these six types [2]:

- radiant
- mechanical
- thermal
- chemical
- magnetic
- electrical

Thus the input signals may be categorized by the same six domains. Examples of devices converting one type of energy (or signal) to another are shown in Table 3.2.

These devices are called transducers (Figure 3.2). For sensors, the column in Table 3.2 showing electrical output signals is of special interest, since sensors typically have to supply the output to an electronic processing unit. Possible effects that can be employed for signal (or energy) conversion are shown in Table 3.3 [2].

Transducers based on diverse effects differ in efficiency, static output characteristics (linear, exponential, etc.), dynamic behavior (time constants, bandwidth, etc.), and cross sensitivity to the influence of interfering factors (e.g., tempera-

Tab. 3.2 Examples of devices for interconverting various types of signal and energy (after Middelhoek and Audet [2]).

Output	Rad.	Mech.	Therm.	Elect.	Magn.	Chem.
Input						
Rad.	optical filter	Golay cell detector	solarimeter	solar cell		photographic processors
Mech.	flint	gearbox		electric generator		
Therm.	temperature-sensitive LCD	bimetal	heat exchanger	thermocouple		
Elect.	LED	loudspeaker	cooling element	MOSFET	coil	battery
Magn.	magneto-optical modulator	magnetic clutch	adiabatic demagn.	Hall plate	magnetic circuit	
Chem.	candle	Combustion engine	gas heater	pH meter		chemical reactor

Tab. 3.3 Examples of effects utilized for interconverting various types of signals and energy (after Middelhoek and Audet [2]).

<i>Input</i>	<i>Output</i>	<i>Rad.</i>	<i>Mech.</i>	<i>Therm.</i>	<i>Elect.</i>	<i>Magn.</i>	<i>Chem.</i>
Rad.	photoluminescence	radiation	radiation	heating	photoconductivity	photomagnetism	photochemistry
Mech.	photoelasticity	transfer of momentum	friction	heat	piezoelectricity	magnetostriction	pressure-induced explosion
Therm.	incandescence	thermal expansion	heat	conduction	Seebeck effect	Curie-Weiss-law	endothermic reaction
Elect.	inject. luminescence	piezoelectricity	Peltier effect	pn junction effect	Ampere's law		electrolysis
Magn.	Faraday effect	magnetostriction	Ettinghausen effect	Hall effect	magnetic induction		
Chem.	chemiluminescence	explosive reaction	exothermic reaction	voltaic effect			chemical reaction

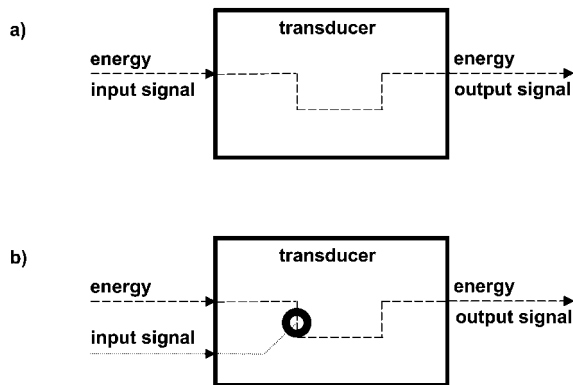


Fig. 3.2 a) Self-generating transducer, b) modulating transducer according to [2]

ture). As shown in Section 3.4, these aspects are of fundamental importance with respect to the achievable sensitivity, time resolution, and accuracy of output from the sensor devices.

A basic representation of a transducer is shown in Figure 3.2, which explains the difference between a self-generating transducer and a modulating transducer.

In modulating transducers a signal is created by means of the modulation of one energy type (energy carrier) by a second type (signal carrier), whereas in self-generating transducers a common carrier contains the energy and the signal as

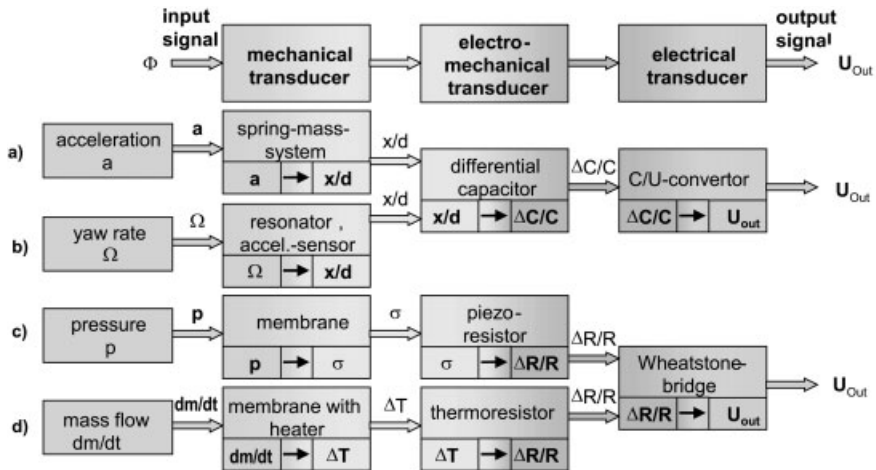


Fig. 3.3 Examples of multiple step signal conversion for different physical input signals to electrical output signals: a) acceleration as an inertial input signal, b) yaw rate as an inertial input signal, c) pressure as an intensive input signal, d) mass flow (density) as an extensive input signal

well. Self-generating transducers may be less sensitive to interfering parameters than modulating transducers, because they are simpler in structure.

In general practice, signal conversion cannot be performed in a single step, because conversion requires appropriate effects (Table 3.3). Examples of multiple-step signal conversions are shown in Figure 3.3.

At this point we may state one very important conclusion: any signal path has to result in electrical output, as has been pointed out already. Because only a few electrical quantities can be used as output signals (voltage, current, frequency), the final transducer in a signal path *must* produce exactly this intended electrical output signal, with high accuracy in relation to the input signal. To achieve the desired electrical output, only a few conversion principles are suitable:

- change in resistance
- change in capacitance or charge
- change in inductance
- change in frequency

Considerations of costs for high-volume production, reuse of know-how, as well as time to market, necessitate standardization and modularization. Therefore, the sensor designer will often try to use similar or even (almost) identical transducers when setting up a signal path (Figure 3.3).

In Table 3.3 the first row is of special interest: the conversion of a radiant signal to another type of energy. In contrast to all other types, transfer of radiant energy is the only one that works without direct material contact. Therefore this domain is very appropriate for certain types of measurements:

- Long-distance measurements, because long-range matter-based conversion principles are impossible in practice.
- High-temperature applications, because matter-based conversion principles would suffer from the effect of temperature on the long-term stability of the materials used.
- Applications in corrosive environments, for reasons of long-term signal stability and reliability.
- Applications for measurements of rapidly moving, vibrating, or rotating parts, because interconnecting problems would arise for matter-based principles.
- Applications for which the components of a sensor system are manufactured and delivered by different suppliers, due to reasons of functional testability and legal responsibility.

Every system is partitioned in various respects inevitably:

- Functional partitioning: input, output, internal states
- electronic circuit and mechanical parts

Therefore it makes sense to partition the system in such a way that it can be composed of independent components with well defined interfaces. In this case the components can easily be manufactured and delivered by different suppliers. If radiation of energy is involved for signal transmittance, transmitting and receiving components can easily be separated and thereby establish an ideal interface for system partitioning (Figure 3.4).

However, this approach can result in a complicated overall signal path configuration, with many signal-conversion steps before the final electrical output signal is produced. As pointed out already, interference may increase at each step of signal conversion.

In general, each signal-conversion step may be expressed by equations such as

$$V_{\text{out}} = f(\Phi, Y_1, Y_2, Y_3(Y_1), \dots) \quad (3.8)$$

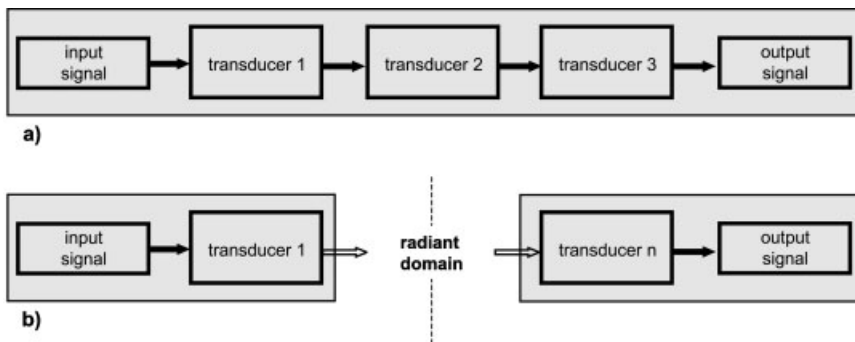


Fig. 3.4 Examples of multiple step signal conversion from input signals to output signals: a) all steps of energy conversion are “non-radiant” (which means that they are matter-based), b) one conversion step makes use of the radiant energy domain

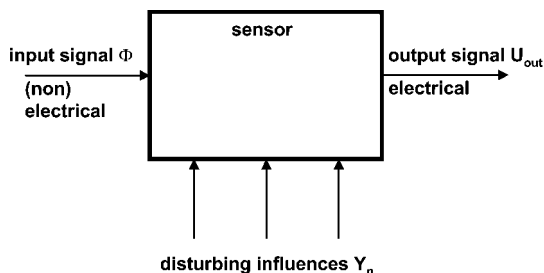


Fig. 3.5 Basic sensor function

for the transducer output signal V_{out} , where Φ is the input signal and Y_n are interfering parameters (see Figure 3.5). Solving Eq. 3.8 yields the value of the input signal of interest:

$$\Phi = g(V_{out}, Y_1, Y_2, Y_3(Y_1), \dots) \quad (3.9)$$

The functions f and g are called *sensor models*. These models can be complicated, especially when multistep signal conversion and cross interactions between different components and types of energy have to be taken into account. Therefore we need to deal with the output signals of different transducing elements within a signal path in more detail (next section).

3.4

Sensor Output Signals

To obtain conclusive data about the environment, technical systems rely on information provided by sensors via electrical output signals. For valid evaluation of the sensor data, it is essential to know the output signal characteristics:

- static output characteristic
- dynamic behavior
- cross sensitivity to interference

The fundamental transfer function of a sensor is determined by the properties of the transducer principles chosen when setting up a signal path. Static sensor performance is defined by the static transfer function that describes the relationship between the output signal U_{out} and the input signal Φ (Eqs. 3.8 and 3.9) and additional parameters such as the measurement range, attainable sensitivity, resolution, and desired accuracy [8].

The ideal parameters can be theoretically predicted from knowledge of the physical properties of each transducer element in the sensor device and their interactions.

Some examples of static output signals are shown in Figure 3.6a; besides an ideal linear (proportional) behavior, real sensors often show inverse-proportional, exponential, parabolic, higher-order polynomial, or much more complicated dependencies. For reasons of simplicity, our considerations below refer to linear behavior.

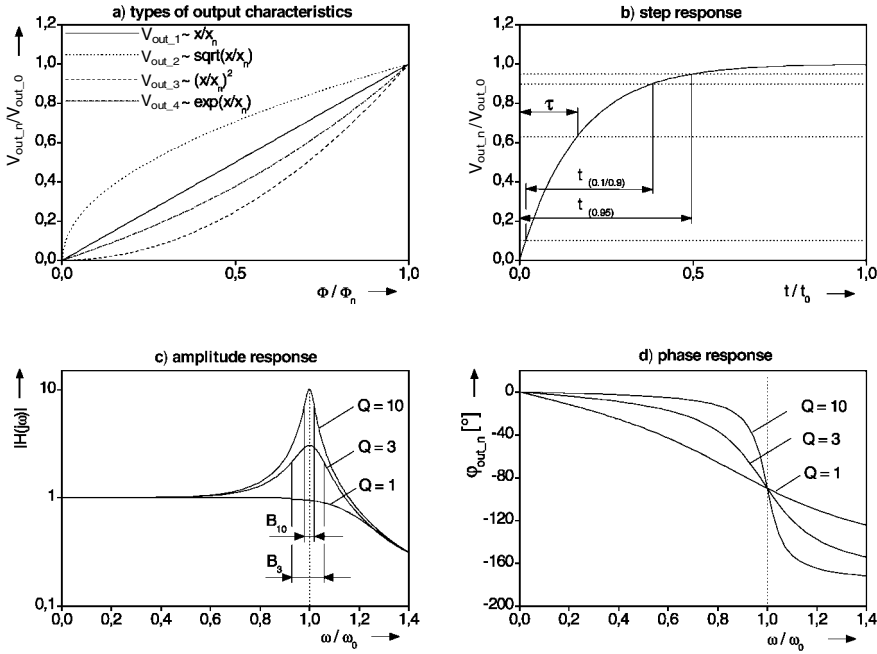


Fig. 3.6 Static and dynamic characteristics of sensors

Besides static parameters, the time dependency of a sensor's behavior characterized by the transfer function, the impulse or step response, has to be taken into consideration. The frequency-transfer function $H(s)$

$$H(s) = L\{V_{\text{out}}(t)\} / L\{\Phi(t)\} \quad (3.10)$$

describes the relation between the Laplace transforms of the time-dependent output signal $V_{\text{out}}(t)$ and the input signal $\Phi(t)$, with $s = j\omega + \sigma$ being the Laplace operator and ω being the angular frequency. This function leads to the dynamic system properties – the natural frequencies f_0 , bandwidth B , damping coefficients c , quality factor Q , and phase lag φ – within the specified frequency range.

Figure 3.6 c and d illustrate amplitude and phase responses of oscillators having different damping coefficients. The step response of a sensor is usually determined by the time constant as well as by the typical rise and response times of the system. Figure 3.6b shows the response of a critically damped system to a step-like change in the input signal Φ . The time constant τ (as defined for an exponential response), the 10% to 90% rise time $t_{(0.1/0.9)}$, and the 95% response time $t_{(0.95)}$ are marked.

However, this 'ideal' transfer function is strongly affected by a sensor's inherent nonideality and by the cross-sensitivity of a sensor to interfering environmental parameters (Y_n in Eqs. 3.8 and 3.9). Hence, these cross-sensitivity and interference effects, including those occurring in the sensor device's signal-processing unit

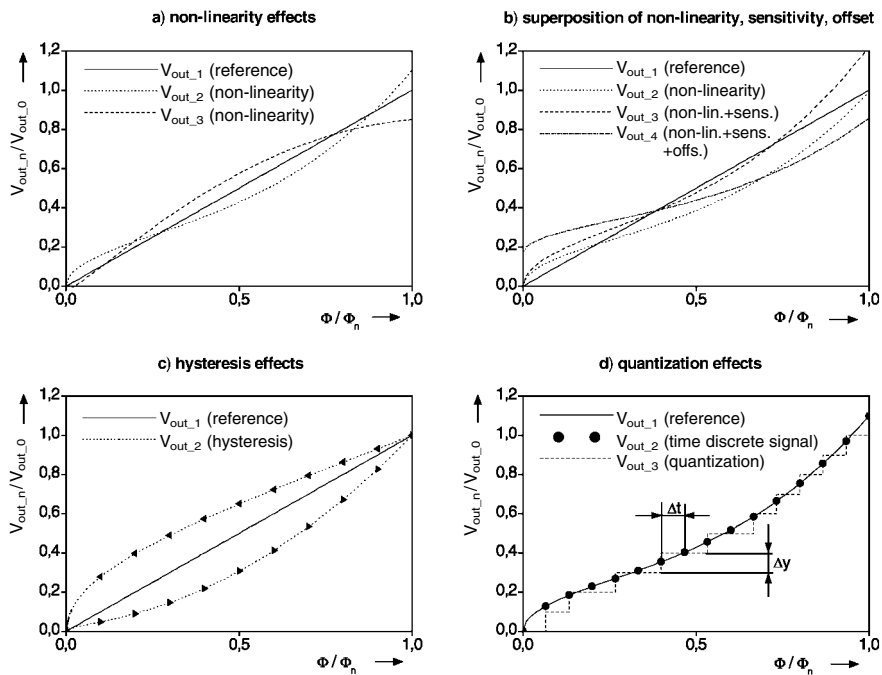


Fig. 3.7 Classification of disturbing interfering parameters

(e.g., quantization error, electronic noise) reduce the attainable sensitivity, resolution, and accuracy of a sensor device or modify its dynamic characteristics.

The resulting deviations of the output signal from ideality can be differentiated as random and systematic errors. The systematic errors are usually classified as sensitivity, offset, nonlinearity (Figure 3.7a and b), hysteresis (Figure 3.7c), and quantization errors (Figure 3.7d). Furthermore, the ‘ideal’ transfer function is also affected by the time dependency of the sensor characteristics at very low frequencies, resulting in effects such as long-term instability, drift, or creeping of mechanical parts.

A more detailed overview of signal characteristics, including the definition of fundamental terms describing sensor characteristics, is given for instance in [8] and [9], and with respect to the special behavior of pressure sensors in [10].

The demand for high sensor performance on the one hand and reliability and long-term stability on the other necessitates more and more thorough understanding of the error characteristics of specific transducer elements becomes more and more important. This knowledge is crucial for device optimization and can be used for error correction by compensation within the transducer or by calibration of the output signal to increase sensor performance to the desired level.

We now discuss the feasibility of using compensation and calibration to reduce the cross-sensitivity of interfering inherent and environmental parameters on a sensor’s output signal in more detail, using temperature effects as an example. Cross-sensi-

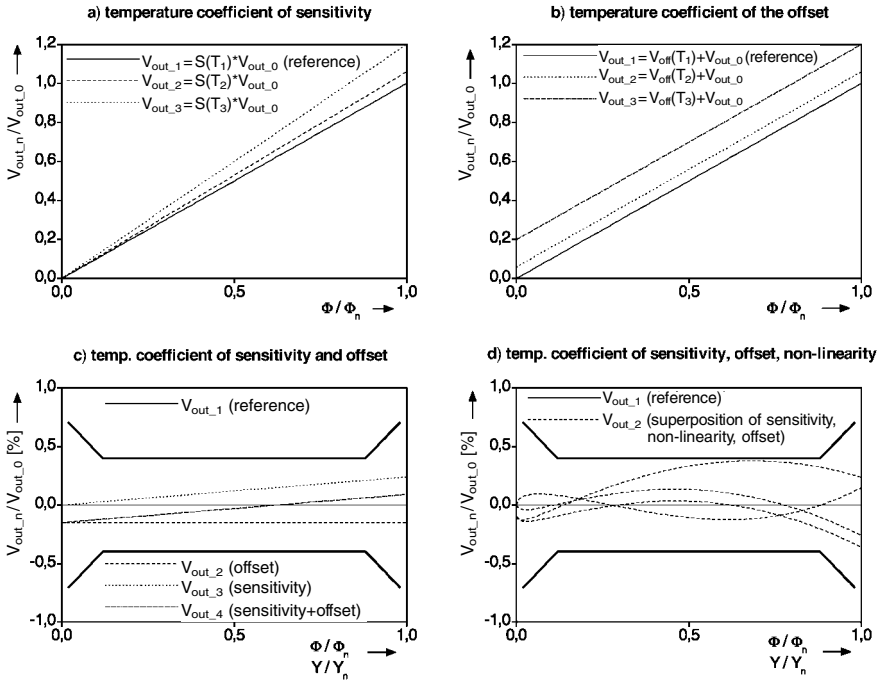


Fig. 3.8 Effect of temperature on sensor characteristics

tivity to temperature changes due to self heating and to environmental conditions occurs in most sensors, especially in automotive applications, in which the specified operating temperatures range from -40°C to $+125^\circ\text{C}$ or even higher.

Two fundamental effects of temperature change that impair the output signal, the effect of temperature on the sensitivity $S(T)$ and on the offset $V_{off}(T)$ are illustrated in Figure 3.8a and b.

A feasible approach to specifying the allowable deviation of the output signal from ideal behavior makes use of the error-band concept. The error band indicates the deviation of the actual signal from the ideal sensor output signal, caused by interfering environmental parameters, such as temperature, within defined ranges. For a complete sensor device, the combined effect of all deviations (errors) has to be within the specified tolerance band. The curves in Figure 3.8c depict the influence of sensitivity and offset errors shown in Figure 3.8a and b applying the error-band concept. Figure 3.8d summarizes the superposition of different errors including nonlinear effects.

In practice, polynomial approximations including linear and parabolic terms are used to calculate the temperature dependency with sufficient accuracy. The following equations describe the temperature dependency of the output signal $V_{out,n}$ and include the sensitivity S , signal offset V_{off} , the nonlinearity of the characteristic curve, as well as their respective temperature coefficients TC_{xi} :

$$V_{\text{out}_n} = S \times V_{\text{out}_0} + V_{\text{off}} + \text{NL} \times f(V_{\text{out}_0}) \quad (3.11)$$

$$S = S_0 \times \Sigma \text{TCS}_i \times T^i \quad (3.12)$$

$$V_{\text{off}} = \Sigma \text{TCV}_{\text{off}i} \times T^i \quad (3.13)$$

$$\text{NL} = \Sigma \text{TCN}_i \times T^i \quad (3.14)$$

As already mentioned, compensation and calibration methods may be used to improve the accuracy of the output signal. From Figure 3.8a and b and Eqs. 3.11–3.14, it is obvious that a sensitivity or an offset error can be calibrated by identifying the input/output characteristics at a single point of the transfer function at several temperatures; the appropriate temperature coefficients can then be deduced. In case of a superposition of sensitivity and offset deviations, calibration can also be done by measuring the input/output characteristic at at least two points.

However, this method is not applicable to all situations. Figure 3.8d shows that detecting the input/output characteristics at several points is not sufficient to calibrate sensitivity and offset errors if there are nonlinear effects involved. Here, a complete system identification might be necessary, depending on the order of the effects. Calibration of this kind of error is possible in theory, but not practicable, considering the expense in time and money.

The most suitable solution is to reduce the limiting effects by design optimization, for instance by changing the transducer principle and the signal path or by optimizing their structural setup (including the technological steps used for fabrication) on the basis of detailed knowledge of the physical characteristics. It is also feasible to use additional design elements to compensate the effects internally, within the transducer element.

Accurate estimation of error effects is one of the most crucial factors in the design of sensor devices.

3.5

Summary and Outlook

This chapter has discussed several basic considerations about sensing. The relationships among input signals, signal conversion, and output signals of sensor devices were deduced from fundamental physical principles. With respect to the specific boundary conditions in automotive applications, these considerations lead to important consequences concerning appropriate technology, reliability, testability, signal resolution, signal accuracy, etc. Therefore, many of the conclusions of this chapter are present (directly or indirectly) in other chapters of this book: design-related questions in Chapter 4; technology, reliability, and testability in Chapter 5; evaluation circuits in Chapter 6; and finally, all application-specific questions in Chapter 7.

The field of automotive sensing is still growing rapidly. New design approaches, new technologies, new transducer principles, new applications, and even new physical conversion principles arise frequently in the sensor developer community. We hope this chapter will support sensor designers in systemizing this very diverse field. And last but not least – the economic success of products and technologies depends to a great extent on decisions made as to the design approach and preferred technology to employ. These decisions will thus eventually determine which products and technologies survive in the automotive sensor market.

3.6

References

- 1 ILLING, M., REINHART, K.F., Chapter 2 of this book.
- 2 MIDDELHOEK, S., AUDET, S.A., *Silicon Sensors*, Academic Press, London, UK, 1989.
- 3 MIDDELHOEK, S., The sensor cube revisited, *Sens. Mater.*, 10(7), 1998, 397–404.
- 4 BRILLOUIN, L., *Science and Information Theory*, Academic Press, New York NY USA, 1956.
- 5 WIENER, N., *Cybernetics*, M.I.T. Press, Cambridge, MA, USA, 1965, and Wiley, New York NY USA, 1961.
- 6 VAN DUYN, D.C., *Multi Signal-Domain Modeling of Solid State Transducers*, Delft University Press, Delft, Netherlands, 1993.
- 7 TRAH, H.-P., FRANZ, J., MAREK, J., Physics of semiconductor sensors, *Advances in Solid State Physics* 39, ed. B. KRAMER, Vieweg Verlag, Braunschweig, Germany, 1999, 25–36.
- 8 GARDNER, J.W., *Microsensors: Principles and Applications*, John Wiley & Sons, West Sussex, England, 1994.
- 9 HART, H., LOTZE, W., WOSCHNI, E.-G., *Meßgenauigkeit*, R. Oldenbourg, Munich, Germany, 1997.
- 10 PFEIFER, G., WERTSCHÜTZKY, R., *Drucksensoren*, VEB Verlag Technik, Berlin, Germany, 1989.

4

Design Methodology

4.1

Methodology

REINHARD NEUL and HANS-REINER KRAUSS

The development process of today's products is dominated by demands such as product performance, quality, time to market, development cost, development risk, etc., while the products' complexity is continually increasing. The use of microelectromechanical systems (MEMS) for general applications and also in the particular field of automotive sensors is growing notably all over the world. This market is expected to continue growing at least in the short- and medium-range future, before reaching saturation at a high level [1, 2]. To keep pace with the market, improving the efficiency of developing MEMS is very important. To improve and accelerate the design and development process, the methods used need to optimize performance and yield of the devices, while minimizing development time and cost by avoiding unnecessary design cycles and production runs. These objectives are achieved by extensively using modeling and simulation techniques, particularly in the early phases of development.

4.1.1

Development of MEMS

Development of MEMS sensors involves tradeoffs between the system's specifications, the properties and capabilities of the sensing element, and the signal evaluation circuit. For MEMS devices, this tradeoff is influenced by the available technology's properties (Fig. 4.1.1). Every aspect depends on and is influenced by the others. In most development processes, the specifications of the sensor systems and the properties of the available technologies are more or less given at the beginning. Then the designs of the sensing element and the signal evaluation circuit have to be set up so as to achieve optimum performance of the complete system.

The development process for any technical system is generally described by a v-shaped diagram (Fig. 4.1.2; see also [3]). This diagram originates from software engineering (see [4]) and has been adopted by people working on mechatronic sys-

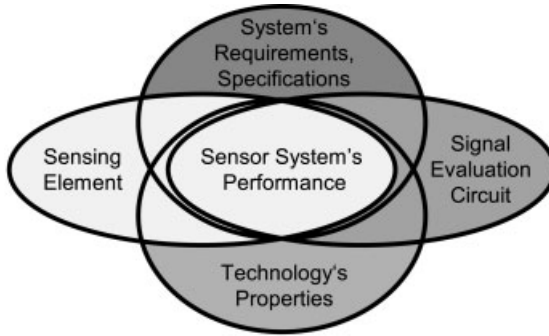


Fig. 4.1.1 Tradeoffs in the development process of MEMS sensors

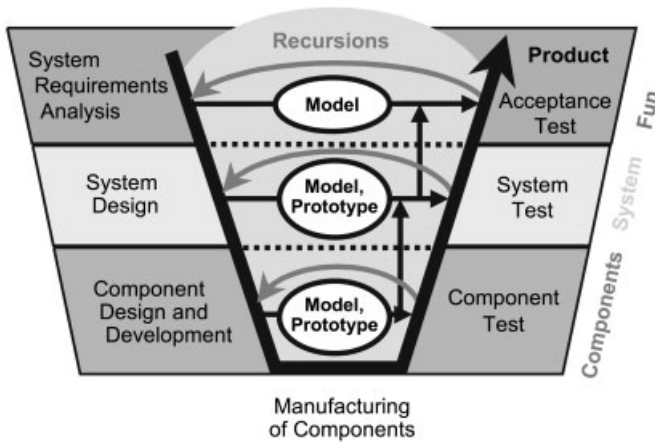


Fig. 4.1.2 The development process for mechatronic systems in general and for MEMS devices

tems (e.g., [5, 6]). Because MEMS are a special kind of mechatronic system, it also holds for their development process.

In general, a development process starts with an analysis of system requirements, followed by design on system level. Subsequently, the system is decomposed into its components and the components are designed and developed. At this stage, manufacturing of first samples of the components is started. If the components pass function tests, the parts are assembled and the complete system is tested. After a subsequent acceptance test, this process leads to the final product.

A real development process is not as straightforward as this, because tests of components and the system and acceptance tests usually reveal problems, which need to be solved. To do so, the designers have to go back to the relevant level in the component design or system design, or even to the system requirements analysis. This is an iterative process that has to be repeated until all tests are satisfied.

Usually only digital integrated circuits can be developed following a single pass through the scheme in Fig. 4.1.2. The development of most other technical systems, including MEMS devices, needs several passes through such a scheme, resulting each time in an engineering sample of the product under development, at a further-improved stage of maturity.

The loops within the design process are time-consuming and add to the cost. Especially in designing MEMS structures, the process commonly requires many fabrication runs and time-consuming experimental setups to obtain a detailed overview of the sensor's behavior. The only way to decrease the iterations to a reasonable number is to use simulation techniques during the early phases of the design process and to do as much testing as possible before manufacturing the components. Therefore, modeling and simulation methods are key factors in the design and development process.

4.1.2

Modeling and Simulation

Modeling and simulation methods can be subdivided into three levels of abstraction (Fig. 4.1.3).

The lowest level of abstraction, here called the *geometry level*, is the closest to physical reality, in which the physics is described by partial differential equations. This level is the domain of finite-element, boundary-element or related methods (e.g., [7–9]). Due to their high accuracy, these methods are well suited for calculating, for example, the distribution of stresses, distortions and natural resonant frequencies of MEMS structures. But they also entail considerable computational effort. Thus, these methods are used to solve detailed problems only when needed, whereas simulations of complete sensor systems and, in particular, transient analyses are carried out using methods at higher levels of abstraction.

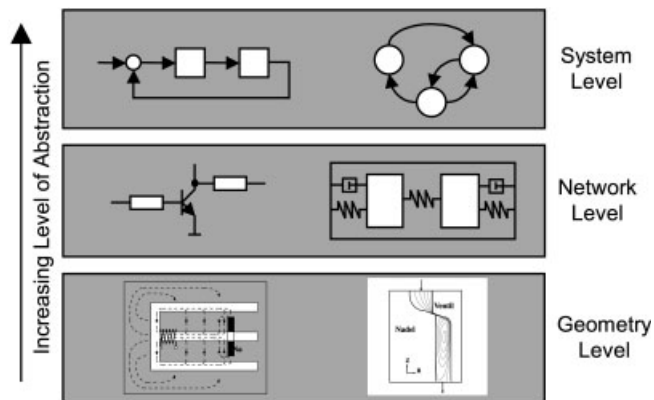


Fig. 4.1.3 Levels of abstraction in modeling and simulation

The *network level* is the domain of, for example, circuit and multi-body simulations. On this level, the physics is described by systems of differential algebraic equations. In spite of the higher level of abstraction, the modeled components are close to physical reality, because we can still recognize physical parts such as resistors and transistors or masses and springs (e.g., [10–15]).

The highest level of abstraction, here called the *system level*, is the domain of system-level simulators, which use block diagrams from control engineering and signal processing or state charts to describe digital systems. System-level simulation requires the least computational effort and is therefore well suited for testing the overall performance of a sensor system. Many examples of system level-modeling can be found in [16].

The simulation level to be used depends on the particular question that arises during the design of MEMS. Each level has its own advantages and disadvantages and cannot be omitted.

4.1.3

Work Flow for Developing MEMS Sensors

The two basic approaches to organizing a design process are bottom-up and top-down. Very early in development, when we are still mainly doing research, most design procedures are bottom-up. But if the technologies being developed are getting more and more mature and we understand its capabilities and properties, we use a top-down approach in developing products, because this approach is better suited to an industrial development process, where we usually start with the specifications required of the product.

A brief description of the work flow is outlined in Fig. 4.1.4. (This work flow is just the left side of the development scheme represented by the v-shaped diagram in Fig. 4.1.2.) Development work starts in a top-down style, with system require-

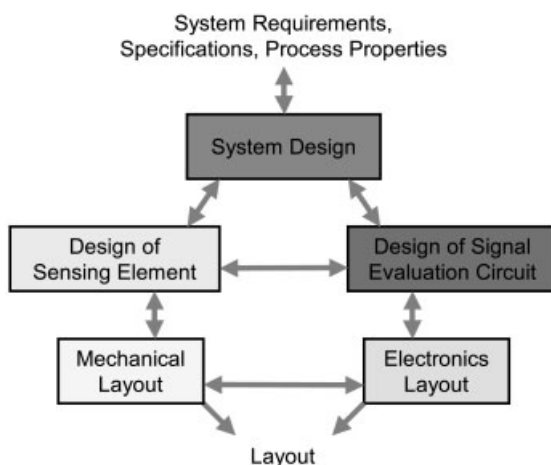


Fig. 4.1.4 Work flow in the development of MEMS sensors

ments and specifications, bearing in mind the properties of the available manufacturing technologies.

The system design begins with analytical investigations of the system's behavior and performance and then uses methods and tools from system-level modeling and simulation (Fig. 4.1.3), ensuring the overall performance of the product under development and taking into account all influences and dependencies between the subsystems. In general, system-level methods and tools can deal with complete systems. Taking testability issues into consideration already at this stage is strongly recommended.

After finishing the system-level design, the system is partitioned into subsystems. In designing MEMS sensors, single-chip or dual-chip (or even more) approaches can be used. All approaches have their benefits and drawbacks.

After partitioning into subsystems (Fig. 4.1.4), the micromachined and the electronic subsystems are designed. The methods and tools usually applied are from the network and geometry levels, respectively (Fig. 4.1.3), which generally cannot deal with complete sensor systems. Therefore, a *back-annotation* and *re-calculation* step is necessary at the system level again, after the sensing element and the signal-evaluation circuit are designed, to assure that none of the changes made to the components compromise the complete system's performance. For this back-annotation step, we need appropriate behavioral models for system simulation, which can be obtained either by back annotating parameters to system level models or by generating reduced-order models.

Next is the layout of the micromechanical and electronic parts of the system. Even at this stage changes are being made, which need to be back-annotated to the component level and also to the system level, if necessary. In electronics, for example, the back-annotated parameters are the specific capacitances, which are not known before layout. In mechanical micromachining, the specific parameters include capacitances as well as masses, moments of inertia, thermal capacities, and conductivities.

4.1.4

Example: Rotationally Vibrating Angular-Rate Sensor

As an example, we describe the work flow for developing an angular-rate sensor, which is implemented in automobile navigation systems as a supplementary inertial-detection system in addition to GPS. (Sensors of similar design are also used in passenger-restraint systems for rollover detection.)

4.1.4.1 Description of the Angular-Rate Sensor System

Let us begin by briefly introducing the angular-rate gyro, before we start on the design process. A scanning electronmicrograph (SEM) of a sensing element is shown in Fig. 4.1.5 ([15, 17–19]).

This sensing element is manufactured by surface micromachining. The mechanical sensing element consists primarily of a flat movable polysilicon disk that is suspended on flexible beams in the center of the element. This disk-shaped ro-

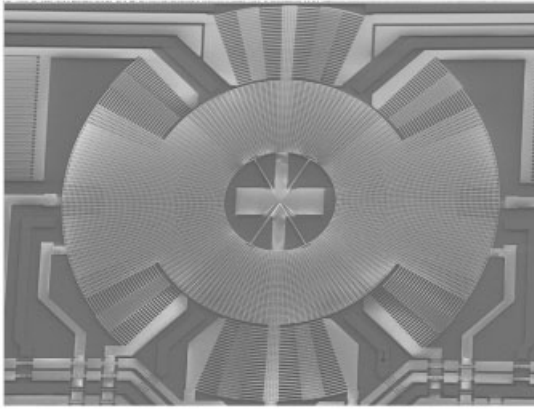


Fig. 4.1.5 SEM image of sensing element in Bosch's MM2 rate gyro

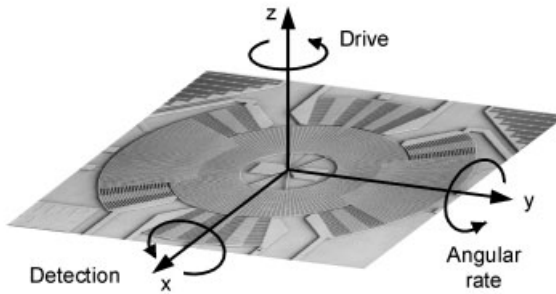


Fig. 4.1.6 Principle of operation of Bosch's MM2 rate gyro

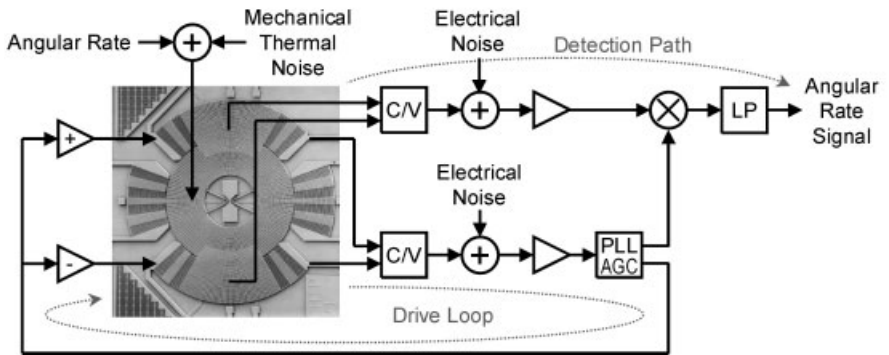


Fig. 4.1.7 System view of Bosch's MM2 rate gyro

tor is driven into resonant vibration about its vertical axis by comb drives located on the circumference of the structure.

If an angular velocity is applied perpendicular to the vertical axis, the rotor responds with a vibrating precession, due to the principle of conservation of angular momentum (Fig. 4.1.6). This tilt oscillation has the same frequency as the in-plane rotational vibration and an amplitude proportional to the angular rate to be measured. This motion is detected by electrodes located beneath the structure as a change in capacitance.

A block diagram of the MM2 angular-rate sensor system (Fig. 4.1.7) shows its controlling and sensing components. Some of the comb structures on the circumference of the disk-shaped sensing element drive the probe mass into rotational vibration, and others are used for detecting the drive motion. Together with part of the sensor electronics, they form the drive loop. A phase-locked loop (PLL) provides any necessary phase shift, and an automatic gain control (AGC) stabilizes the oscillation at constant magnitude. In the detection path, the very small signal due to changes in capacitance (down to the attofarad range) is amplified and, after synchronous demodulation and low-pass filtering, the angular rate signal is output. Figure 4.7 also shows the 3 fundamental noise sources that limit the resolution of the device.

4.1.4.2 Design of the Angular-Rate Sensor System

According to the typical work flow for developing MEMS sensors (Fig. 4.1.4), design of the angular-rate sensor system starts from the required specifications, with system-level design, and comprises the steps described in this section.

4.1.4.2.1 Transducer Principle

The principle of operation of transducers is based on the conservation of either linear (i.e., Coriolis effect) or angular momentum, making a transducer well suited for micromachined rate-sensing gyros. One or more linearly or rotationally vibrating probe masses are required, for which the input motion stimulus and the output signal can be accomplished by various physical effects (electrostatic, electromagnetic, piezoresistive, etc.). Usually the drive motion is resonant, so the detection motion can also be resonant; or the two natural frequencies are separated by a certain frequency shift. Drive and detection motion can be excited by in-plane motions or by a mixture of in-plane and out-of-plane motions.

In designing Bosch's MM2 rate-sensing gyro, a 'semi resonant' rotational principle was chosen; that is, in-plane stimulation is resonant and electrostatic, and out-of-plane detection is also electrostatic, but off-resonant. These decisions can only be made after several studies and assessments of alternative schemes. Here, we focus only on the actually implemented principle of operation, as described in Section 4.1.4.1 (see also [15, 17–19]).

4.1.4.2.2 Analytical Modeling and System Design based on Process-Dependent Parameters

The physics of the transducer is described by the principle of conservation of angular momentum:

$$\begin{aligned}\mathbf{M} &= \mathbf{J} \mathbf{a}_{0,2} + \boldsymbol{\omega}_{0,2} \times \mathbf{J} \boldsymbol{\omega}_{0,2} \\ \mathbf{a}_{0,2} &= \mathbf{a}_{0,1} + \mathbf{a}_{1,2} + \boldsymbol{\omega}_{0,1} \times \boldsymbol{\omega}_{1,2} \\ \boldsymbol{\omega}_{0,2} &= \boldsymbol{\omega}_{0,1} + \boldsymbol{\omega}_{1,2}\end{aligned}\quad (4.1)$$

Eq. 4.1 can be obtained from several mechanics textbooks (e.g., [20, 21]) and is shown here in 3-dimensional vector notation under consideration of relative motion. \mathbf{J} is the matrix of the probe mass's moment of inertia; $\mathbf{a}_{0,2}$ and $\boldsymbol{\omega}_{0,2}$ are the angular acceleration and velocity of the probe mass's coordinate system (index 2) with respect to the inertial frame (index 0); $\mathbf{a}_{0,1}$ and $\boldsymbol{\omega}_{0,1}$ are the angular acceleration and velocity of the reference system (index 1) with respect to the inertial frame. The reference system is attached to the sensing element's substrate and therefore also to the vehicle whose motion is to be measured. $\mathbf{a}_{1,2}$ and $\boldsymbol{\omega}_{1,2}$ belong to the probe mass with respect to the reference system. \mathbf{M} is the torque applied to the probe mass and is composed of the driving stimulus as well as the stiffness of the suspension beams and the damping of the mechanical resonator.

From Eq. 4.1 we can derive the sensitivity of the transducer. The measurement effect in terms of the tilt angle's magnitude a around the x axis (Fig. 4.1.6) of the disk-shaped probe mass is given by

$$a = \frac{2}{\omega_z} \cdot \frac{\gamma \Omega_y}{\sqrt{\left(\frac{\omega_x^2}{\omega_z^2} - 1\right)^2 + \frac{\omega_x^2}{Q_x^2 \omega_z^2}}}\quad (4.2)$$

where γ is the magnitude of the drive motion's in-plane rotation about the z axis, ω_z the in-plane resonance frequency, ω_x is the out-of-plane resonance frequency, and Q_x the out-of-plane quality factor. Ω_y is the angular velocity around the y axis, which is to be measured.

Starting from the analytical description, we can create a first rough design of the sensor system, including the resonant frequencies for drive and detection and the sizes of the sensing element, the electrostatic comb structures, and the detecting capacitors. While this work is done, the physical properties of the technology already have to be considered.

Design of the signal-evaluation circuit is handled in the same way. This first design is verified using system simulation, which relies on Eq. 4.1 and the block diagram in Fig. 4.1.7. To perform the simulation based on process-dependent parameters, the coefficients in Eq. 4.1 must be expressed in terms of parameters obtained from real manufacturing processes, for example, edge losses in the geometry of the probe mass and the flexible beams, layer thickness, residual stresses, encapsulated gas pressure, etc. For each process-dependent parameter, the statistical deviation

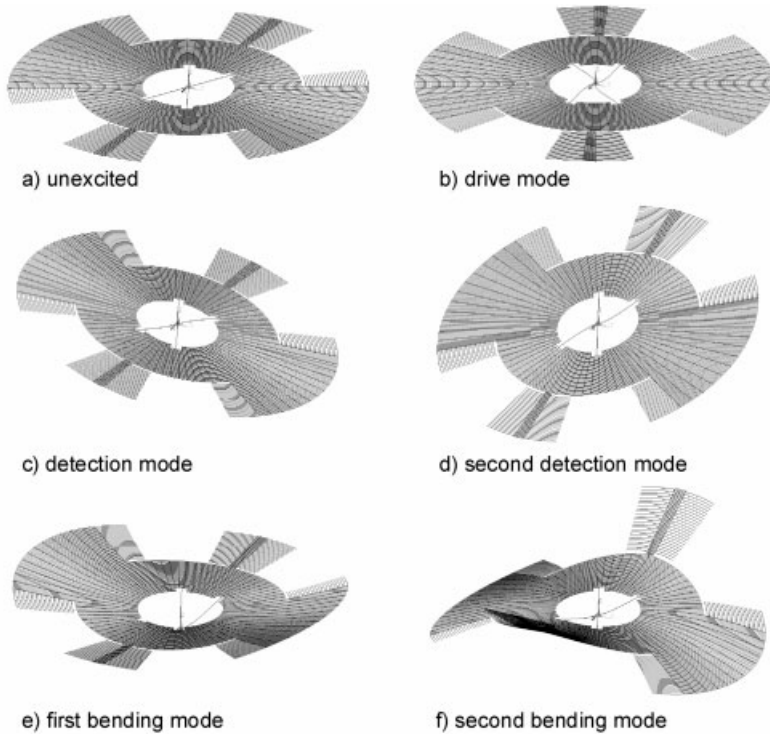


Fig. 4.1.8 Modal analysis of sensing element in Bosch's MM2 rate gyro

has to be obtained from in-process or electrical monitors and expressed as their process capability values c_{pk} (for a definition of c_{pk} values see, e.g., [22, 23]).

4.1.4.2.3 Detailed Design and Optimization of Sensing Element and Signal-Evaluation Circuit

System design is followed by detailed design of the sensing element and the signal-evaluation circuit (Fig. 4.1.4). Electronics design is outside the scope of this section and is covered in Chapter 6. Here, we focus on the sensing element. Detailed elaboration of the sensing element's geometry is creative engineering work, which in the first instance must be verified by finite-element simulation. Fig. 4.1.8 shows a modal analysis of the rate gyro's sensing element.

The engineering task is to find geometrical properties that lead to appropriate resonant frequencies for the drive and detection modes and to well separated perturbation modes, which are nearly insensitive to normal operation of the device or to external disturbance (e.g., acceleration or vibration). We also have to minimize the influence of variations within manufacturing tolerances (this issue is mainly treated in Section 4.1.5).

The next step is to verify the details of the electrostatics used in the drive and detection, which can be done by finite-element or boundary-element analysis. Cer-

tainly the computational burden is very high, especially if we handle electrostatics and mechanics simultaneously. If the probe mass can be approximated by a rigid body and if we are interested only in the fundamental modes of operation and if the perturbation modes are well separated, we can use network-type models of the sensor. This approach enables us to incorporate mechanics and electrostatics into one network-type simulation and to apply an arbitrary 3-dimensional motion of the sensor system, concurrent with a considerable reduction in computational effort (see Section 4.2 and [10, 13–15]).

After detailed design and verification of both the sensing element and the signal-evaluation circuit, we have to back-annotate all changes made during the detailed design procedures to system level, to ensure that the changes do not prevent the whole sensor system from operating correctly. The back-annotation step can be accomplished simply by parameter exchange, if possible, or by generating reduced-order models where needed. This leads to a refined system-level model of the whole sensor system, which for instance allows us to study the implemented control loops' behavior, the signal scaling in the signal-processing path, the signal-to-noise ratio, and thus the resolution of the device.

Subsequently, the mechanics and electronics are laid out. During layout, electrical and mechanical properties may change once more. Therefore, we have to perform back-annotation again to verify the system's behavior.

This angular-rate sensor example is continued in Section 4.1.6, where we take into account additional considerations necessary for obtaining robust designs. But first, we need to give some background on robust design.

4.1.5

Robust Design of the Mechanical Sensing Element

The methods described in this section are used particularly during design of the sensing element (Fig. 4.1.4) and also to some extent in the system-design phase. (Similar approaches are applicable to the design of the signal-evaluation circuit, but we will focus here on the micromechanical aspects.)

The goal of the work flow, generally described in Section 4.1.3 and applied to the rate gyro design example in Section 4.1.4.2, is to obtain a nominal set of design parameters for the MEMS being developed. In this section we are additionally interested in the effects of parameter variations caused by manufacturing tolerances on the device's performance. This view of the design process of MEMS is outlined in Fig. 4.1.9 (see also [24–26]).

MEMS devices have to meet certain criteria with respect to their functional parameters, for example, a scale factor, offset of the output value, temperature coefficient, nonlinearity, hysteresis, noise, resolution, and cross sensitivities, which characterize the system's performance. In addition, we are interested in the reliability, yield, and cost of the devices. The set of functional parameters depends on a set of model parameters, consisting of processing, material, and geometrical parameters. All model parameters act as input parameters for the design procedure as well as for the manufacturing process. Material parameters are influenced by

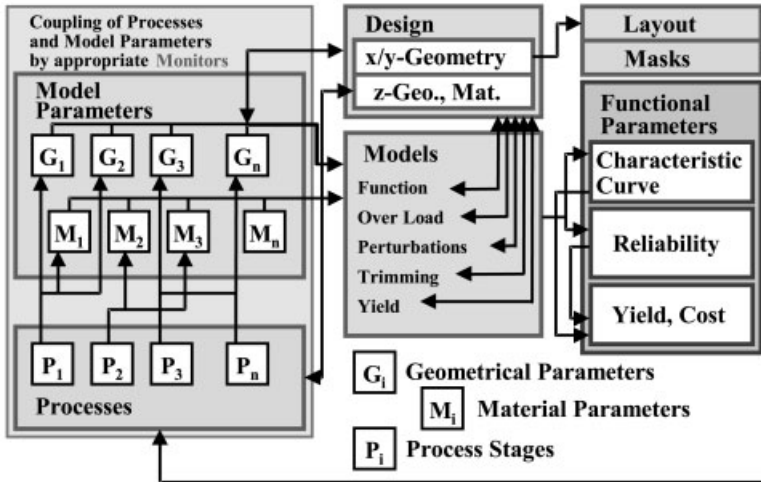


Fig. 4.1.9 Schematic chart of the dependencies relevant to robust design. The only allowed input model parameters are geometry and material parameters, which are correlated with processing effects

the processing steps and their parameters, whereas geometrical parameters are determined by processing parameters and the actual design.

The advantage of this approach to a robust design process is the possibility of performing a special kind of sensitivity analysis, as described in the following section. We determine certain quantities for the influence of each model parameter on each functional parameter, leading to the identification of the influences that are critical for a particular design. Knowledge of these critical influences at an early development phase is essential for keeping recursion loops small (Fig. 4.1.2).

4.1.5.1 Influence Strengths

First, we need to know the dependencies of the functional parameters on the model parameters in terms of mathematical functions. All functional parameters (FP) of the device have to be expressed in terms of process-dependent model parameters (MP):

$$FP_i = f(MP_j) \quad (4.3)$$

Functional parameters FP_i of a sensor include the sensitivity S , cross sensitivities, temperature coefficient TC , temperature coefficient of sensitivity TCS , offset O , and corresponding TCO . Nonlinearities of TCS and TCO are $NLTCS$ and $NLTCO$. And we also have hysteresis, burst pressure, hermeticity, and other parameters. These functional parameters need to be described by model parameters MP_j , $j=1\dots n$, which are appropriate for the processes used to fabricate the device (layer thickness, etching profiles, residual layer stress, etc.).

The models used for Eq. 4.3 may range from simple to complicated. Most of these functions are already known from system design and from the detailed design of the sensing element (see Section 4.1.3). If there are no analytical models available or if the physical relationships are too complicated for analytical description, finite element analysis, network-type analysis, or empirical studies need to be used to obtain the relationships summarized in Eq. 4.3.

To obtain the influence strengths EFS_{ij} (EFS stands for german *Einflussstärke*, ([24–26]) the matrix formed by the elements in Eq. 4.4 has to be calculated:

$$EFS_{ij} = \frac{\frac{\partial}{\partial MP_j} FP_i |_{MP_{1..n}=OP} \cdot \Delta MP_j}{ATB_{FP_i}} \quad (4.4)$$

Because only the first derivatives at the actual design (with model parameters at some operating point OP) are used in Eq. 4.4, we have to carefully ensure that the functions behave almost linearly close to the OP; otherwise, higher-order derivatives have to be taken into account. This gradient is multiplied by ΔMP_j , which is the difference of the upper and lower limits of the allowed values for the model parameter. For yield calculations using Monte Carlo methods this span may be interpreted as $\pm 3\sigma$ of the distribution of the MP, if Gaussian behavior can be assumed. Using this method we obtain the maximum possible scattering of the FP from a single independent MP. Therefore, the EFS value represents the capability of the design with respect to the maximum possible scattering due to process tolerances as well as with respect to the allowed tolerance band (ATB) of the functional parameter FP_i . If FP_i is an adjustable parameter, the ATB is given by the capture range of the trimming procedure. If FP_i is not adjustable – due to, for example, physical or cost reasons – the ATB is a certain proportion of the maximum tolerated error specified for the device. The EFS_{ij} directly show the contribution of MP_j scattering to the ATB of the corresponding functional parameter FP_i .

We classify the EFS_{ij} values as follows:

- $EFS_{ij} > 1$ hypercritical
- $1 > EFS_{ij} > k$ critical; $k = 1/\sqrt{n+1}$ where n is the number of model parameters
- $EFS_{ij} < k$ uncritical
- $EFS_{ij} = 0$ no influence

To get an idea about the complexity of an EFS matrix we have to consider that a typical microsensor has many MP (sometimes more than 100), and the number of FP to be taken into account is usually between 10 and 50. Therefore, it is necessary to focus on the ‘important’ parameters. The EFS matrix is a very useful tool for finding these important parameters. Examples of such EFS matrices are given in Sections 4.1.6 and 4.1.7. For further EFS examples please see [24–27].

Regarding any particular FP, the sum

$$EFS_i = \sum_{j=1}^n \sqrt{EFS_{ij}^2} \quad (5)$$

EFS c_{pk}	+ (high)	- (low)
+ (low)	no action required	extend range of process limits
- (high)	reduce range of process limits	<ul style="list-style-type: none"> • design optimization • process improvement • system optimization

Fig. 4.1.10 Optimization measures in response to possible combinations of c_{pk} and EFS values

is the overall scattering caused by all independent MP that contribute to the corresponding FP. An $EFS_i=1$ means that $\pm 3\sigma$ of all pieces (99.73%) are 'designed' to be within this FP's ATB. If $EFS_i \leq 1$, more pieces are within the ATB and therefore, such a design is called a *robust design* with respect to the FP under consideration. If $EFS_i \geq 1$, the design is not robust with respect to the allowed process tolerances. Such a design and/or the processes in use will have to be improved by optimization or by compensation techniques. Examples are given in Section 4.1.6.

4.1.5.2 Optimization

The EFS matrix shows all influences of the process tolerances on all functional parameters at a glance. Therefore, all consequences of changes in design parameters or in process tolerances can be forecast and, vice versa, process problems can be detected easily with this matrix. In addition, the EFS matrix is a powerful tool for the development of optimization algorithms for function, yield, and cost.

Processes are characterized by their c_{pk} values and designs are characterized by EFS values. If the c_{pk} value is sufficiently large, the process is well controlled. If the EFS value is small, the design is robust with respect to the selected process. Knowing about their connections, we can vary these parameters to optimize a design using a given process. If this is not sufficient to meet the specifications and to achieve an acceptable yield, either we have to improve the process, thus changing the specifications of the sensing element and/or the signal-evaluation circuit, or we may even have to reenter the system-design stage and think about a better transducer principle.

Fig. 4.1.10 summarizes the actions that have to be taken for various combinations of c_{pk} and EFS values. Examples of applying these measures are shown in Section 4.1.6.

4.1.6

Example: Angular-Rate Sensor (Continued)

Section 4.1.4.2 outlines the system design of the rate gyro and the design of the sensing element, leading to nominal parameters (the operating point [OP] of the design). To prepare for EFS analysis, we have to define model and functional parameters, set up the model equations (Eq. 4.3), and perform the analysis according to Eq. 4.4 as described in Section 4.1.5.1. A part of the EFS matrix for Bosch's MM2 rate gyro is shown in Fig. 4.1.11.

The examples in the next two sections show some capabilities of the EFS method.

Model parameters (MP)		Process limits				Functional parameters (FP)					
		typ.	lower limit	upper limit	cpk-value	Unit	S	Ph	Res	B	
						-	°	°/s	V		
						Op. point	0,37	3,70	0,85	0,80	
						lower limit	0	0	0	0,31	
						upper limit	1	15	2	1,10	
Process parameter:											
Trench edge loss	μm	0,7	0,2	1,2	-		0,45	0,07	0,29	0,39	
Epi thickness	μm	10,6	9,1	12,1	-		0,47	0,10	0,31	0,21	
Stress gradient	MPa/μm	1,2	-1	3	-		0,49		0,32	0,52	
Oxide thickness	μm	1,6	1,4	1,8	low		0,25		0,17	0,15	
		1,6	1,3	1,9	high		0,33		0,22	0,21	
							$\sqrt{\Sigma \text{EFS}^2}$	0,85	0,12	0,56	0,70
								0,88	0,12	0,58	0,71

Fig. 4.1.11 Part of the EFS matrix of Bosch's MM2 rate gyro. Functional parameters are: normalized sensitivity S , demodulation carrier phase Ph , resolution Res , self-test signal magnitude B

4.1.6.1 Design Optimization by Variation of Process Parameters

We show the effect of the allowed processing tolerance with the EFS method, using as an example the thickness of the sacrificial oxide layer (please refer to Section 5.3). Oxide layer thicknesses in the range of 1.4–1.8 μm result in a low c_{pk} value, meaning that real processing deviations are higher, according to the definition of c_{pk} values. To achieve a better yield from the process for the chemical vapor deposition of the sacrificial oxide layer, the EFS matrix was recalculated using broader processing limits (1.3–1.9 μm). The new limits provide for a higher c_{pk} value, but also for increasing EFS values; thus, the quality of the design is a more critical issue now. Therefore, we have to check that this change in processing parameters doesn't compromise the robustness of the design. From the ESF matrix (Fig. 4.1.11) it is obvious that most functional parameters are affected by a variety of model parameters. Because individual processing steps are assumed to be independent, and the functions in Eq. 4.3 are monotonic for each model parameter, the total EFS values can be calculated from Eq. 4.5. The sensitivity, resolution, and magnitude of the test signal are affected according to the EFS matrix (bold numbers in Fig. 4.1.11), but the total EFS values are only slightly increased and are still well below 1, so we can accept the new processing limits without causing an undue decrease in yield.

4.1.6.2 Improvement in Sensor Performance

What can be done if, for instance, we want to specify the magnitude B of the self-test signal within closer limits? The EFS matrix shows that, for example, the sacrificial oxide thickness has little influence on the test signal's magnitude, whereas the stress gradient of the epi polysilicon layer (see Section 5.3) is much more critical (italic numbers in the last column of Fig. 4.1.11). This identifies the corresponding processing step to have the maximum influence on the test signal's magnitude. Therefore, the stress gradient in the epipolysilicon layer is the most promising candidate for optimization, and thus, process engineering needs to focus only on improving that single processing step. Knowing exactly where im-

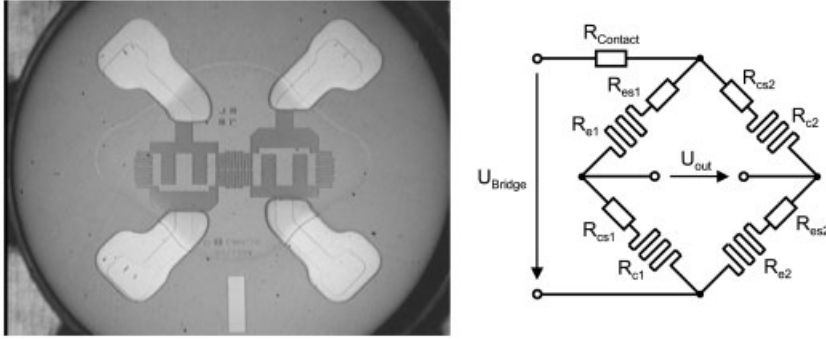


Fig. 4.1.12 Thin-film metal structures on the steel membrane of Bosch's high-pressure sensor and the equivalent circuit of a Wheatstone bridge [27]

provements have to be made is also important for decreasing development costs and time-to-market for a new sensor.

4.1.7

Example: High-Pressure Sensor

Here we show an example of applying the EFS method to a non-silicon-based pressure sensor to operate at high pressure ranges. The membrane of many high-pressure sensors (Bosch, WIKA) is manufactured of steel, with thin-film metal resistors as a measurement signal pickup (Fig. 4.1.12).

The sensor's transducer principle combines a membrane, as a mechanical transducer, with thin-film metal resistors forming a Wheatstone bridge, as an electro-mechanical transducer. The output voltage of the Wheatstone bridge is given by

$$U_{\text{out}} = U_{\text{bridge}} \frac{R_{C1}(R_{C2} + R_{E2}) - R_{E2}(R_{C1} + R_{E1})}{R_{\text{contact}}(R_{C1} + R_{E1} + R_{C2} + R_{E2}) + (R_{C1} + R_{E1})(R_{C2} + R_{E2})} \quad (4.6)$$

$R_{C1,2} = R_{c1,2} + R_{cs1,2}$ and $R_{E1,2} = R_{e1,2} + R_{es1,2}$ are the resistances of the strain gages located in the center (R_C) and at the edge (R_E) of the membrane, respectively. $R_{cs1,2}$ and $R_{es1,2}$ (which do not change with pressure) are resistances in series to the strain gages. R_{contact} includes the bonds and their connections to the Wheatstone bridge. Each resistor R_i in Eq. 4.6 can be expressed as

$$R_i = \rho_i (1 + (T - T_0)TC) \frac{l_i}{w_i t_i} \quad (4.7)$$

Processing-dependent parameters are:

- the specific resistances ρ_i of $R_{C1,2}$ and $R_{E1,2}$ and R_{contact}
- the metal film thickness t_i

Model parameters (MP)		Process limits			Functional parameters (FP)			
		typical	lower limit	upper limit	Unit	Sensitivity	Rin	TK Rin
					mV	Ohm	ppm / K	
					Op. point	0,50	3500,00	0,00
					lower limit	0,00	2800,00	-50,00
					upper limit	1,00	4200,00	50,00
Membrane thickness	mm	0,30	0,28	0,32		0,62		
Membrane radius	mm	0,30	0,25	0,35		0,22		
Bore radius	mm	2,01	2,00	2,03		0,10		
Y-modulus	GPa	196	186,2	205,8		0,20		
Square resistance NiCrSi	Ohm/sq.	25,50	23,00	28,00			0,42	
TKR NiCrSi	ppm/K	-15	0	-30				0,01
Edge loss	μm	3,5	2	5			0,11	
TKR contact	1/K	0	-0,01	0,01				0,07
$\sqrt{\sum \text{EFS}^2}$						0,69	0,43	0,07

Fig. 4.1.13 Part of the EFS matrix for Bosch's high-pressure sensor, showing the influence of selected model parameters on 3 functional parameters

- different tolerances of the photolithographic process for the lengths l_i and widths w_i of R_C and R_E

Part of the resulting EFS matrix is shown in Fig. 4.1.13.

4.1.8

Characterization and In-process Measurement of Model Parameters

Performing EFS analysis requires knowledge of the values of the model parameters and their statistical properties. However, many of the process parameters that influence a certain functional parameter are not directly measurable using in-process tests. Or the applied measurement methods may provide no information about a particular device on the wafer, because these measurement methods need too large a wafer area and therefore only a few spots on the wafer are used for measurement. For instance, the stress gradient can be measured at the wafer level, but the stress in the rotationally vibrating mass in the MM2 rate gyro can be different from the value measured on the wafer. To use the EFS method, it is important to characterize the process parameters for each single chip, to obtain their statistical properties.

Additionally, some parameters, such as sensitivity, are influenced by electrical and mechanical effects. This means that designing the sensor has to include not only the sensor's geometry, but also the design of tests that can distinguish between electrical (e.g., capacitance) and mechanical (e.g., stress) properties. Thus, designing for testing is an important issue.

The only way to obtain all necessary parameters is to apply model-based measurement techniques and electrical stimulation during wafer-level tests for each single device or by using test structures designed especially to obtain in-process measurement data (see [28, 29] and Section 5.10). This requires that the sensor's behavior is well understood and the dependencies of the various parameters can be expressed analytically according to Eq. 4.3.

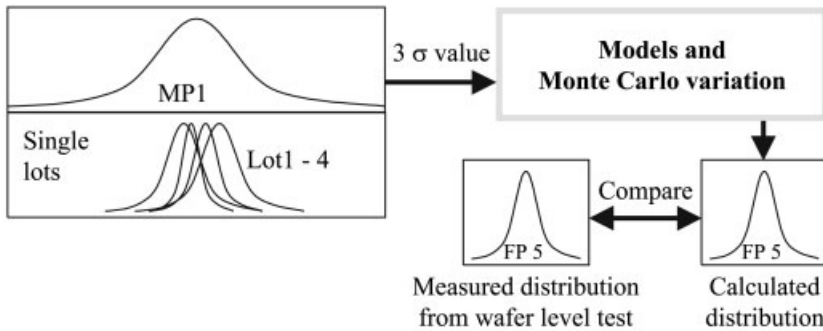


Fig. 4.1.14 Monte Carlo calculation of a functional parameter's distribution. In this example it is assumed that only one model parameter affects the functional parameter being considered

4.1.9

Yield Investigation by Monte Carlo Methods

Here we again assume that the sensor's behavior is described by analytical functions. Input parameters are only the processing-, material-, and geometry-dependent parameters. Then we can examine the distribution of the functional parameters by simulations based on these equations. Two conceivable ways of calculating the yield for a new design that uses a given process follow.

The first method is to do a worst-case calculation, which means that only the process limits of each processing step are taken into account. Unfortunately, this method is not applicable to most sensors in high-volume production today, because it is not realistic to treat a process flow by setting all parameters simultaneously to their process limits.

The second method is to vary the parameters by Monte Carlo methods to obtain realistic predicted yields. This means that we have to vary the processing parameters independent of each other, using their realistic distributions.

The Monte Carlo method is outlined in Figure 4.1.14. For each lot, the distributions of the model parameters (shown here for one model parameter, i.e., the epilayer thickness of the MM2 rate gyro) are used as input to the Monte Carlo method, which results in a calculated distribution of the functional parameters (e.g., the sensitivity of the MM2 sensor). This procedure has to be done simultaneously for all model parameters.

The predicted results can be verified by comparing simulation results with measured distributions of functional parameters. Fig. 4.1.15 shows a Monte Carlo simulation of the magnitude of the detection signal output of Bosch's MM2 rate gyro. This functional parameter depends on the sacrificial oxide layer's thickness, the thickness and stress gradient of the epi layer, trench edge loss, etc. The Monte Carlo simulation for the rate gyro is compared with measurement data from 145 devices.

Figure 4.1.16 shows the same comparison for the offset of Bosch's high-pressure sensor.

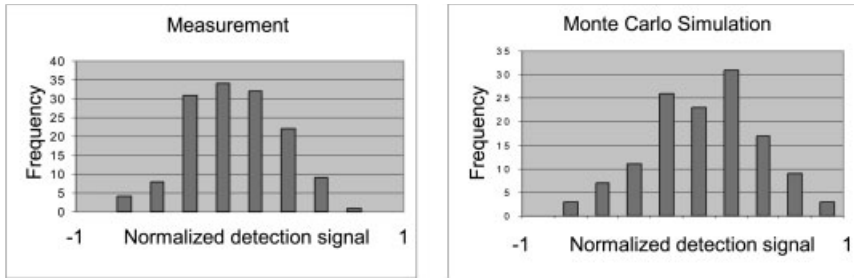


Fig. 4.1.15 Comparison of measured distribution (left) and results of Monte Carlo simulation (right) for the magnitude of the detection signal output of Bosch's MM2 rate gyro

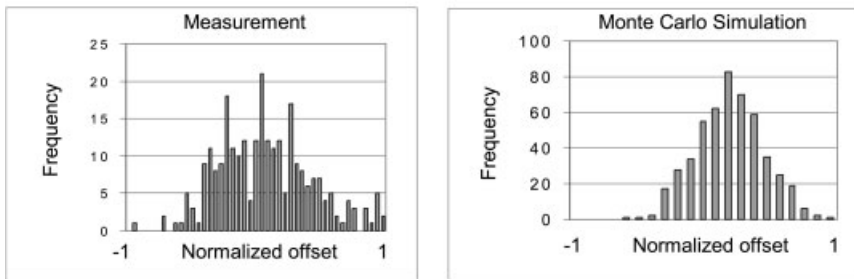


Fig. 4.1.16 Comparison of measured distribution (left) and results of Monte Carlo simulation (right) for the offset of Bosch's high-pressure sensor (adopted from [27])

The model parameters for the Wheatstone bridge's offset are the lengths and widths of the single resistors as well as the thicknesses and specific resistances of the thin metal films. Subtracting the randomly distributed edge losses from the original layout's geometry parameters yields the real geometrical variations for calculating the resistances. Additionally varying ρ_i and t_i according to their distributions leads to the results shown in Fig. 4.1.16.

4.1.10

Conclusions

Application of the methods described in this section in the early stages of the development process helps designers to verify their designs before real hardware is manufactured, thus saving time and cost by avoiding unnecessary loops in the development process.

The requirements, in addition to powerful software for precise simulation results that also need to be accurately predictive, are accurate and reliable models. The necessary investigations start with modeling of MEMS-relevant physical effects and material properties and the properties of technological processes. Due to the fact that MEMS devices consist of parts belonging to different physical domains, solvers able to handle these domains simultaneously are needed.

4.1.11

References

- 1 Intechno Consulting, *Intechno-Weltreport Sensormärkte 2008*, Basel, Switzerland, 1999.
- 2 Strategy Analytics, Ltd., *Market Report Automotive Sensor Demand 1998–2007*, Luton, UK, 2001.
- 3 R. NEUL, G. LORENZ, S. DICKMANN, Modelling and simulation of micro electromechanical sensor systems, *Proc. MICRO.tec 2000*, 2, Hannover, Germany, 2000, 67–72.
- 4 J. BUXTON, J. McDERMID, Architectural Design, in *Software Engineer's Reference Book* (ed. J. McDERMID), Butterworth–Heinemann, 1991, Ch. 17, pp. 1–33.
- 5 R. BATE et al. (ed.) *A Systems Engineering Capability Maturity Model*. Carnegie Mellon University, Pittsburgh PA, USA.
- 6 R. ROTHFUSS, M. LASA, H.-M. HEINKEL, P. TIRGARI, Systems engineering in the design of mechatronic systems, *Int. J. Vehicle Design* 28, 2001, 19–36.
- 7 J. ARGYRIS, H.-P. MLEJNEK, *Die Methode der Finiten Elemente in der elementaren Strukturmechanik*, Vieweg & Sohn, Braunschweig, Germany, 1986.
- 8 K.-J. BATHE, *Finite-Element-Methoden*. Springer, Berlin, Germany, 1990.
- 9 J.S. PRZEMIENIECKI, *Theory of Matrix Structural Analysis*, Dover, New York NY, USA, 1985.
- 10 G. LORENZ, R. NEUL, Network-type modeling of micromachined sensor systems, *Proc. Int. Conf. Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators, MSM98*, Computational Publications, Cambridge MA, USA, April 1998, 233–238.
- 11 N. ZHOU, J. V. CLARK, K. S. J. PISTER, Nodal analysis for MEMS design using SUGAR version 0.5, *Proc. Int. Conf. Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators, MSM98*, Computational Publications, Cambridge MA, USA, April 1998, 308–313.
- 12 J. E. VANDEMEER, M. S. KRANZ, G. K. FEDDER, Hierarchical representation and simulation of micromachined inertial sensors, *Proc. Int. Conf. Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators, MSM98*, Computational Publications, Cambridge MA, USA, 1998, 540–545.
- 13 G. LORENZ, R. NEUL, S. DICKMANN, Modeling of electrostatic MEMS components, *Proc. 2nd Int. Conf. Modeling and Simulation of Microsystems, MSM99*, Computational Publications, Cambridge MA, USA, April 1999, 128–131.
- 14 G. LORENZ, *Netzwerksimulation mikromechanischer Systeme*, Dissertation, Universität Bremen, Germany, 1999.
- 15 D. TEEGARDEN, G. LORENZ, R. NEUL, How to model and simulate microgyroscope systems. *IEEE Spectrum*, July 1998, 66–75.
- 16 S. D. SENTURIA, *Microsystem Design*, Kluwer, Boston, USA, 2001.
- 17 K. FUNK, *Entwurf, Herstellung und Charakterisierung eines mikromechanischen Sensors zur Messung von Drehgeschwindigkeiten*, Dissertation, Universität München, Germany, 1998.
- 18 R. SCHELLIN, A. THOMAE, M. LANG, W. BAUER, J. MOHAUPT, G. BISCHOPINK, L. TANTEN, H. BAUMANN, H. EMMERICH, S. PINTÉR, J. MAREK, K. FUNK, G. LORENZ, R. NEUL, A low cost angular rate sensor for automotive applications in surface micromachining technology, *Proc. Advanced Microsystems for Automotive Applications 99* (eds D. E. RICKEN, W. GESSNER), Springer, Berlin, Germany, 1999, 239–250;
- 19 A. THOMAE, R. SCHELLIN, M. LANG, W. BAUER, J. MOHAUPT, G. BISCHOPINK, L. TANTEN, H. BAUMANN, H. EMMERICH, S. PINTÉR, J. MAREK, K. FUNK, G. LORENZ, R. NEUL, A low cost angular rate sensor in Si-surface micromachining technology for automotive applications; *SAE Technical Digest Series 990931*, 1999.
- 20 W. SCHIELEN, *Technische Dynamik*, Teubner, Stuttgart, Germany, 1986.
- 21 K. MAGNUS, H. H. MÜLLER, *Grundlagen der Technischen Mechanik*, Teubner, Stuttgart, Germany, 1990.
- 22 H. BAUER (ed.) *Bosch Automotive Handbook*, 5th edition, SAE Society of Automotive Engineers, Warrendale, USA., 2000, p. 155.

- 23 J. M. JURAN, F. M. GRZYNA (eds.) *Juran's Quality Control Handbook*, McGraw-Hill, New York NY, USA, 1998.
- 24 S. FINKBEINER, J. FRANZ, S. HEIN, A. JUNGER, J. MUCHOW, B. OPITZ, W. ROMES, O. SCHATZ, and H.-P. TRAH, Simulation of nonideal behaviour in integrated piezoresistive silicon pressure sensors, *Design, Test, and Microfabrication of MEMS and MOEMS, DTM 1999*, SPIE Proc. 3680, Bellingham WA, USA, 1999, 188–199.
- 25 W. ROMES, J. MUCHOW, S. FINKBEINER, J. FRANZ, O. SCHATZ, H.-P. TRAH, Simulation of nonideal behaviour in integrated piezoresistive silicon pressure sensors, *Proc. 2nd Int. Conf. Modeling and Simulation of Microsystems, MSM99*, Computational Publications, Cambridge MA, USA, April 1999, 624–627.
- 26 H.-P. TRAH, J. FRANZ, J. MAREK, Physics of semiconductor sensors, *Advances in Solid State Physics, Arbeitskreis Festkörperphysik, Deutsche Physikalische Gesellschaft* 39, 1999, 25–36.
- 27 J. MUCHOW, A. KRETSCHMANN, R. HENN, K. SKROBANEK, S. FINKBEINER, H.-R. KRAUSS, Influence of process variation on the functionality of a high pressure sensor, *Design, Test, Integration and Packaging of MEMS/MOEMS, DTIP 2002*, SPIE Proc 4755, Bellingham WA, USA, 2002, 143–148.
- 28 M. MAUTE, S. KIMMERLE, J. FRANZ, J. HAUER, D. SCHUBERT, H.-R. KRAUSS, D. P. KERN, Parameter extraction for surface micromachined devices using electrical characterization of sensors, *Proc. 4th Int. Conf. Modeling and Simulation of Microsystems, MSM2001*, Computational Publications, Cambridge MA, USA, April 2001, 406–409.
- 29 M. MAUTE, *Oberflächenmikromechanik-Sensoren als elektrische Teststrukturen zur Charakterisierung ihrer Herstellungsprozesse*, Dissertation, Universität Tübingen, Germany, 2002.

4.2

CAD Tools for MEMS

GUNAR LORENZ and JOE REPKE

4.2.1

Introduction

MEMS researchers have made remarkable advances in recent years, fabricating motors, levers, gears, pumps, pressure and movement sensors, switches, and fields of movable mirrors too small to be seen with the naked eye. However, transitioning this device research into large-scale production has proven more challenging than expected. Whenever moving toward mass production, designers are plagued with all sorts of MEMS-specific problems, including such disparate phenomena as thermomechanical noise, stress gradients, contact stiction, and gas damping. State-of-the-art modeling and simulation techniques can uncover likely fabrication problems at an early stage and thus help accelerate design iterations and drastically shorten development cycles.

4.2.2

Modeling Tools

A wide range of simulation tools and techniques are commonly used to support and accelerate the various steps of the MEMS design process. The different modeling techniques have their tradeoffs in simulation time, skill level requirements, limita-

tions on the scope of the design problem that can be analyzed, and restrictions on the types of simulations that can be performed. In a largely simplified and rather idealized view, three major categories of tool classes can be identified, including signal flow simulators, circuit simulators, and partial differential equation (PDE) solvers.

4.2.2.1 Signal Flow Simulators

The most abstract level of MEMS design is the use of block diagrams and mathematical spreadsheets to lay out the system architecture and evaluate the designer's first ideas and case studies.

A convenient way of creating and organizing functional-level models is to use signal flow simulators with graphical user interfaces, such as MATLAB Simulink [<http://www.mathworks.com>]. Signal flow simulators offer the user a simple way to express physical behavior in diagrams of directed and nonfeedback building blocks.

The large number of blocks available in the software, ranging from basic transfer functions to filters and controllers, make signal flow simulators the perfect choice in the design of MEMS control systems and digital signal processors (DSP's). Their simplicity in use, as well as their robust and speedy simulation performance, comes at the price of a high level of model abstraction. The effectiveness of block diagrams depends largely on the user's ability to identify subsystems having causal dependency between input and output signals. Although these block diagrams are straightforward for modeling many system components, such as sigma-delta modulators, filters, and position control systems, they quickly become clumsy for modeling energy-conserving systems such as electronic circuits and mechanical structures. In block diagrams, the actual electromechanical device is often assembled from simple voltage-force converters and one or more force-position oscillator blocks. These models are usually parameterized with abstract and nongeometrical system quantities such as mass, stiffness, and Q factors. These models are useful for determining the needed behavior (specifications) for a MEMS device and for designing the system surrounding the MEMS but cannot be used for detailed design of the devices themselves.

4.2.2.2 Circuit Simulators

Circuit simulators have their roots in the modeling and simulation of electronic circuits. After the first time-domain circuit simulators began to appear in the late 1960's [1], they evolved into comprehensive design environments, becoming essential to every electronic circuit designer. Modern IC's would be virtually impossible without state-of-the-art time-domain circuit simulators like Spectre, ELDO, SABER, and the many flavors of SPICE. In a parallel path, frequency-domain simulators became the dominant tool for simulation of radio frequency (RF) and microwave components and circuits. Today, Agilent's Advanced Design System (ADS), Applied Wave Research's (AWR's) Microwave Office, and Cadence's Analog Design Environment, among several others, arm the RF designer with a complete arsenal of analysis tools [2].

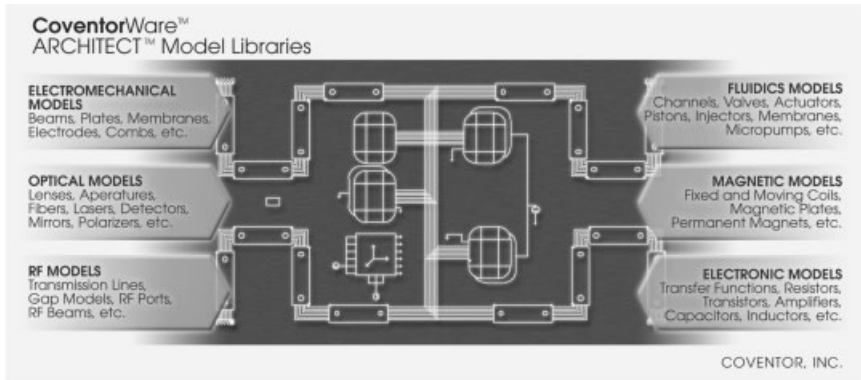


Fig. 4.2.1 Coventor's Architect™ libraries for MEMS system modeling in SPECTRE or SABER

With the development of SABER in 1986, the first analog hardware description language (AHDL), MAST [3], became available. Hardware description languages, like MAST and more recently VHDL-AMS [<http://www.eda.org/vhdl-ams>] and Verilog-AMS [<http://www.eda.org/verilog-ams>] enable designers to create models independent of the simulation program and based on ordinary differential equations (ODE's). The separation of models and simulators opened the scope of circuit simulators from their pure electronic origins to multi-physics behavioral modeling. Today, most circuit simulators are capable of handling networks of different physical domains as well as digital and signal flow models in the same schematic.

Circuit simulators were initially used to model MEMS as single-degree-of-freedom, mass-string-damper systems using electrical equivalent circuits. In the late 1990's, MEMS designers started to realize the full potential of circuit simulators for the MEMS design process. Around 1998, three independent workgroups from Carnegie-Mellon University, the University of California at Berkeley, and Robert Bosch GmbH started to publish their work on the first MEMS libraries for circuit simulators [4–7]. After 3 additional years of development, the first commercially available MEMS library appeared in 2001 as part of Coventor's MEMS development software. The initial library included models of beams, plates, comb structures, and electrodes. Later versions include other physical domains, such as magnetics, thermodynamics, microfluidics, optics, and RF components, turning traditional electronic simulators into full-blown MEMS design environments (Figure 4.2.1) [8].

In a circuit simulator, MEMS designers work in a schematic-driven environment using symbols that represent behavioral models of individual components or elements of a component. These symbols are connected and configured in the schematic to represent the design of the MEMS device. MEMS behavioral model libraries include underlying code that expresses how the individual components behave when subjected to electrical or mechanical stimuli or stimuli from other domains. The complex mathematical descriptions used with high-level models leads to a small number of degrees of freedom, making it practical to investigate MEMS devices in a much more spontaneous manner. Complete systems, includ-

ing peripheral subsystems (for example, sensing and driving circuits) can be built in a common design environment to provide an accurate system-level evaluation.

4.2.2.3 PDE Solvers

MEMS designs have traditionally been simulated with the finite element method (FEM), boundary element method (BEM), or finite-difference method, where individual components are analyzed by solving partial differential equations. A variety of single-domain and cross-domain solvers are available, ranging from traditional engineering tools like ABAQUS, ALGOR, ANSYS, COSMOS, FIDAP, FIRE, HFSS, and MAFIA to products specialized for the needs of MEMS designers, such as ANSYS/Multiphysics [www.ansys.com], CFD-ACE+ [www.cfdarc.com], CoventorWare Analyzer [www.coventor.com], and IntelliSuite [www.intellisence.com].

PDE solvers are used at the component or subassembly level to provide detailed simulation results for displacement, stress, resonant frequency, damping, and electromagnetic field properties. Their ability to handle virtually any geometry makes them the perfect tools for design verification and when behavioral modeling reaches its limits. Although PDE solvers have much more geometric flexibility than behavioral modeling techniques, they are largely restricted to simulations of a single physical domain. The usability of PDE solvers is limited by their demands on computer memory and simulation time. Directly coupling physical domains within PDE solvers demands even more resources. The intelligent use of numerical solvers ultimately depends on a solid understanding of the underlying principles of the device and the solver operations. Thus, a high level of expertise is needed to operate the tools and to interpret the simulation results correctly.

4.2.3

Design Flow Alternatives

Sophisticated software-driven design flows allow the user to move seamlessly between models at different levels of abstraction and guarantee a maximum amount of data reuse from previous simulations. Two different design flow methodologies are usually applied, including a conventional, bottom-up design and a system-level, structured custom-design methodology (Fig. 4.2.2).

With a conventional, bottom-up MEMS design methodology, the designer enters the design flow by generating a two-dimensional (2D) layout and defining the process. The layout is usually derived from high-level considerations with signal flow type simulators, mathematical spreadsheets, or simple analytical formulas. The layout and process are used to generate a 3D model, either by simply mapping process and layout files (CFD-ACE+, CoventorWare™) or by using a process simulator from a given manufacturer (IntelliSuite™) (Figure 4.2.3).

3D models also can be created with a 3D layout tool, but this is generally not recommended, because designs created outside the context of fabrication processes may not be capable of being fabricated.

In the next step of the bottom-up design flow, the 3D model is meshed for PDE type analysis. Pertinent simulations for all relevant physics can be performed to

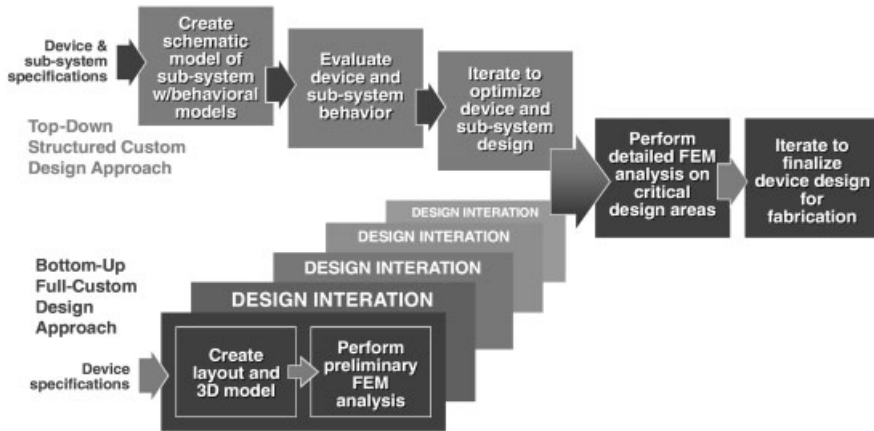


Fig. 4.2.2 Top-down and bottom-up simulation approaches

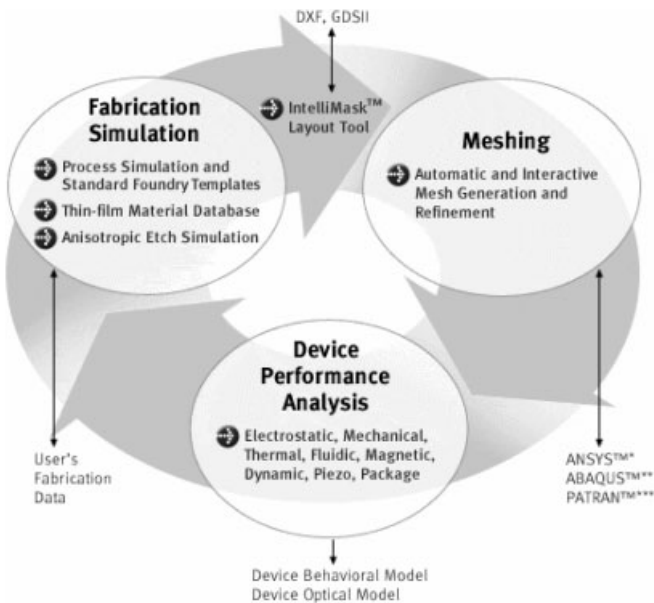


Fig. 4.2.3 IntelliSuite™ design flow for MEMS as presented by Corning's Intellisense

examine the operation of the device. Once the desired behavior is obtained, a set of simulations covering the intended operational range can be used to extract a detailed behavioral model of the MEMS device for use in higher-level system simulations.

The main drawback of this approach is that the initial simulations of these devices to determine their basic dimensions require computationally intensive, and

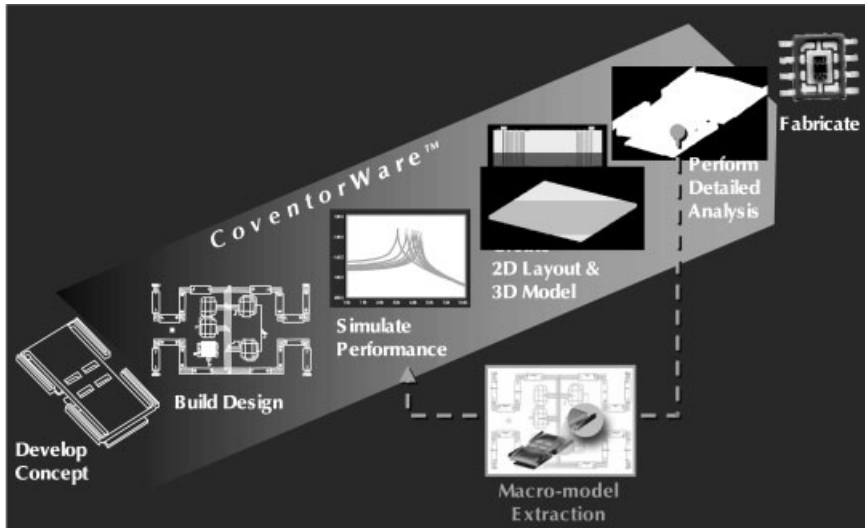


Fig. 4.2.4 CoventorWare™ design flow for MEMS, as presented by Coventor, Inc.

therefore time-consuming PDE simulations. It is also likely that additional passes through this sequence will be required after the initial design, to include the effects of system interactions analyzed with the extracted behavioral models. Although this analysis can be time-consuming, simulation can greatly improve the probability of success and reduce the number of fabrication cycles required to complete a MEMS design.

An alternative process, known as system-level, top-down MEMS design methodology (Fig. 4.2.2) uses the initial set of MEMS system requirements to select a design and fabrication approach. Like the bottom-up approach, initial considerations normally include calculations with signal-flow type simulators. Instead of using a layout-drawing tool to create a 2D model, high-level design techniques require the use of a graphical schematic capture tool to position and connect the model symbols that represent the reduced-order building blocks (Fig. 4.2.4).

Procedurally, the user selects the desired functions from available libraries, configures the parameters for the MEMS application, and then interconnects their ports to create the entire function (Fig. 4.2.1). The procedure is analogous to digital circuit capture, where symbols represent gates, counters, registers, and arrays that have individual input and output function pins connected to create a desired Boolean or sequential digital function. For high-level capture in designing MEMS, electromechanical library building blocks may represent masses, plates, and electrodes, parameterized as to both geometry and material properties. It is force, rather than current, that is transferred between pin connections for the mechanical elements. The symbol ports represent inputs for electrical, mechanical, or magnetic sources or control input or output pins for degrees of freedom (translational and rotational) that may be selected for simulation. A complete captured model can include

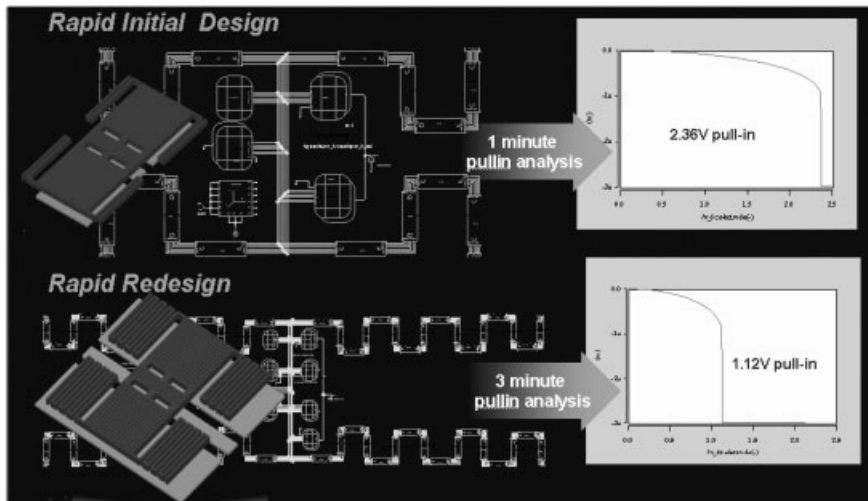


Fig. 4.2.5 Fast design iterations using library-based design methodology

the MEMS device to be fabricated as well as the surrounding system interface, allowing accurate analysis of the device within its system environment.

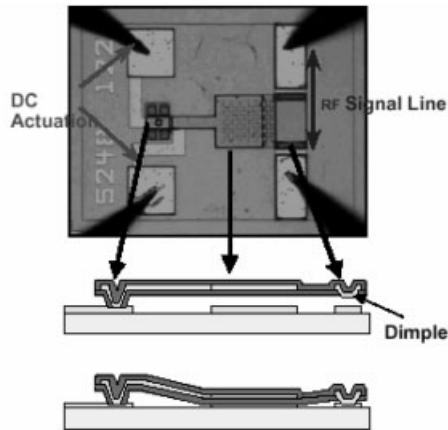
A first-pass, accurate simulation of the MEMS device, electronic control circuitry, environmental effects, and signal conditioning is possible in a matter of minutes. Design investigation with sensitivity and Monte Carlo analysis for optimization and statistical performance analysis may easily be performed (Fig. 4.2.5).

Because the simulations use code-based, reduced-order models instead of FEM-based or BEM-based partial differential equation models, the simulation time is reduced by orders of magnitude. The compelling benefit of this new paradigm is twofold: (1) designers can capture complete device and subsystem behavior across the different physics domains required for sensor, optical, and RF MEMS, and (2) accurate, comprehensive simulations take only minutes instead of days, enabling rapid exploration of wide-ranging design spaces.

Once the initial design is complete to the designer's satisfaction, a device layout can be generated from the high-level description. If the design is simple enough, this may be sufficient to tape out for fabrication. The more likely path involves detailed 3D PDE analysis to verify that the circuit model approximations are acceptable for a given design. If modifications to the system model appear necessary, the initially created circuit model can be modified with extracted model components or enhanced with extracted parameters to account more accurately for effects like air damping, anchor flexibility, etc.

System models built with parametric and extracted libraries provide a straightforward way to encapsulate and document design problems and solutions for future developments. Continuous updating of existing library components helps designers leverage previous experience and increase the likelihood that less time is needed in future development cycles. Furthermore, by working with circuit li-

Fig. 4.2.6 Scanning electronmicrograph and diagram of the cantilever-beam series RF switch from HRL Laboratories (Schaffner et al. [9]). © 2000 HRL Laboratories, LLC. All rights reserved



braries, designers are forced to think in terms of verified MEMS components rather than in custom layouts, making it much easier to enforce design rules and ensure proper operation of the actual device.

4.2.4

RF Design Example

MEMS components for RF applications have attracted a significant amount of attention over the past several years [10–13]. This has in part resulted from the established MEMS design community branching their technology into the RF components marketplace, as well as from the wireless and automotive industry's search for RV technologies that can achieve levels of performance, cost, and size that are less achievable, or even unattainable, with more traditional technologies. Consequently, a new demand for MEMS design tools that integrate RF simulation capability is beginning to emerge.

Both qualitative and quantitative tradeoffs among the system subdomains are demonstrated for such a design flow below, using the example of a cantilever beam RF switch design, similar to that published by HRL Laboratories [9] (Fig. 4.2.6).

A scanning electronmicrograph of such a switch shows the series switch plate suspended over a microstrip transmission line, with a large actuation electrode located between the anchor and the series switch conductor (Fig. 4.2.6). Exaggerated sideview animations of the switch demonstrate the operation of the cantilever beam RF switch.

In the first step of the design flow, the system model of the electromechanical structure can be quickly constructed from parts in the parametric electromechanical library (Fig. 4.2.7)

Only three standard components – an anchor, a beam, and a beam with an electrode – are needed to represent the electromechanical switch's behavior on the circuit level. The beam models simulate the 3D nonlinear mechanical behavior, based upon a complete set of parameters describing the dimensions and material

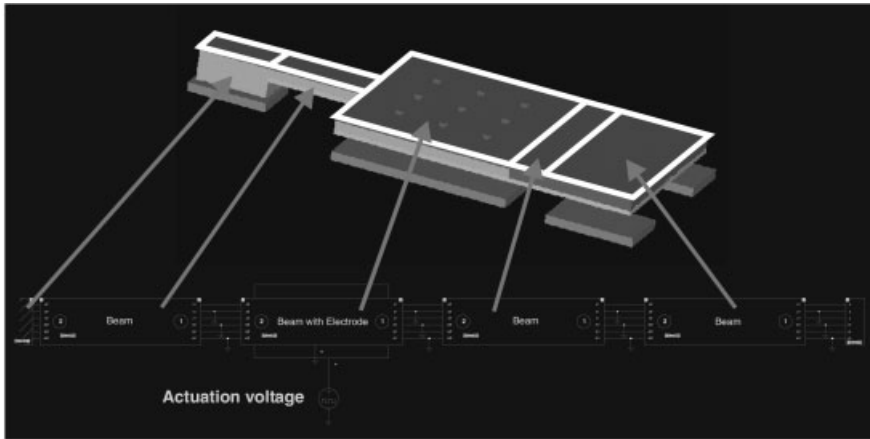


Fig. 4.2.7 Representation of an RF switch at the circuit level

data for the three mechanical layers. The ‘beam with electrode’ component includes the electrostatic forces acting on a buried conductor suspended above the planar electrode. All library symbols are linked by six ‘mechanical wires’, representing the motion and rotation of the connected beam ends. Some of the wires in the schematic are grounded, thus restricting the model to only vertical motions in this initial simulation.

In this system model, important design and process parameters, such as the dimensions of the actuation electrode, the width of the cantilever beam, layer thickness, and material stress and strain, can be defined globally for easy manipulation during the design process. Design studies including frequency response in vacuum, pull-in analysis, contact-force investigations, sensitivity analyses, and Monte Carlo statistical simulations are easily performed in this environment (Fig. 4.2.8). With the addition of damping parameters (normally extracted from PDE solver results), dynamic design behavior can be investigated with AC and transient simulation (Fig. 4.2.9).

For RF switches, a parameter of primary concern is the *insertion loss*, which defines the portion of RF signal power that is lost within the device as thermal energy dissipated in the materials. Also contributing to the total amount of lost power is the amount of signal power reflected by the device back towards the input signal source, defined as the *return loss*. To minimize the reflected power, the switch must be designed to properly match the characteristic impedance of the transmission line structures that connect the signal path to and from the switch. Transmission lines are an integral part of the circuit, and significant power can be lost within them if proper attention is not paid to their design.

Much of the literature on modeling RF MEMS devices focuses on equivalent-circuit models, although the circuit values are often obtained by curve-fitting to measured RF performance. This is especially true for obtaining the s-parameters (for a discussion of s-parameters, see [14, 15]). These equivalent circuits, however, are

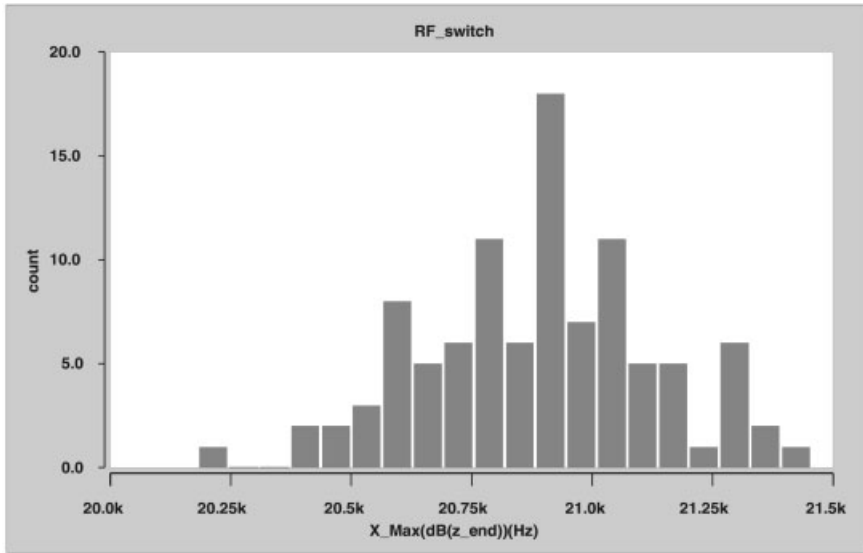


Fig. 4.2.8 Statistical probability of the first resonance frequency in the presence of fabrication process tolerances

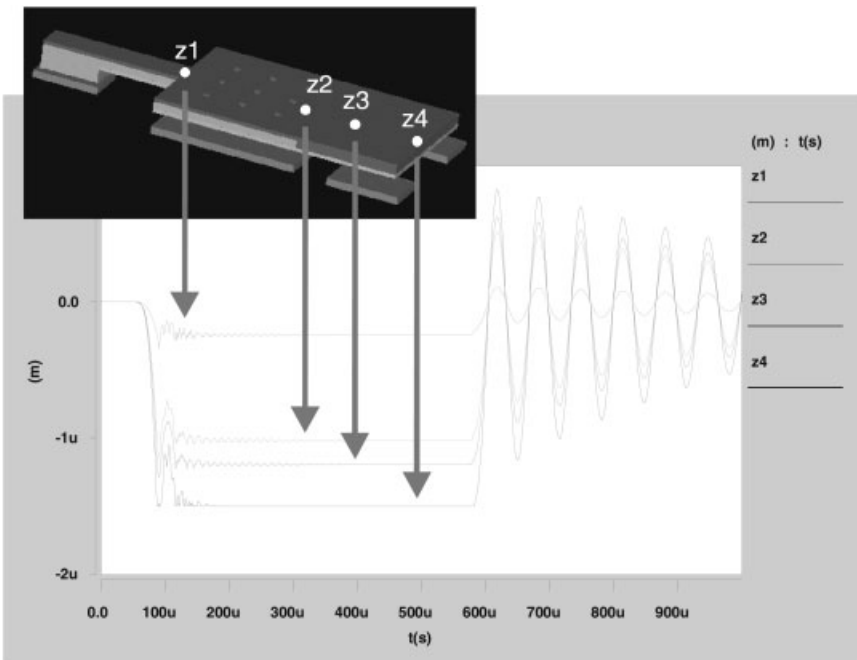


Fig. 4.2.9 Transient results showing the z position of 4 different points during a switch-release cycle

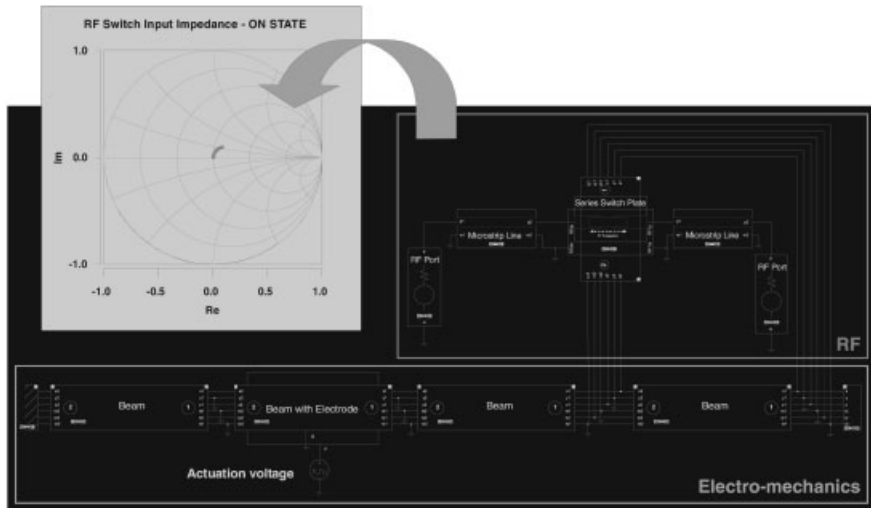


Fig. 4.2.10 Switch schematic including RF components

initially created from cognitive inspection of the MEMS structure, in which individual features of the structure are assumed to represent an equivalent circuit component due to their geometries and material composition.

Encapsulating available knowledge into parametric models for electrical performance at radio frequencies (RF) enables the designer to predict insertion loss, return loss (reflection), and insertion phase and delay of the MEMS switch without time-consuming full-wave analysis in the same environment as used for the initial electromechanical analysis.

In the schematic shown in Figure 4.2.10, the RF path is visible between the two signal sources (RF ports) used for extracting the S parameters, and is composed of a length of microstrip transmission line from each port connected to a model for a ‘series-switch plate’. Driven by the 6 mechanical wires at each side, which control its position, the switch plate is internally modeled as an equivalent circuit including transmission line, frequency-dependent resistance, and variable capacitance between the conductor on the plate and the underlap of the ends of the microstrip lines separated by the gap for the switch isolation. As with the beams, this model is defined by a complete set of parameters, such as the dimensions and material properties. Parameters can be adjusted quickly to achieve the desired RF performance for different closing states of the electromechanical structure.

The design iterations at the circuit level are accompanied by a series of bottom-up verifications, taking advantage of the integrated design flow between the system model and the 3D model generation (see Fig. 4.2.4). Simplifications previously made in the behavioral model can be verified using conventional PDE analysis. Furthermore, additional device parameters such as damping values and stress distributions can be analyzed (Fig. 4.2.11).

FEM analysis examples from two domains:

Fig. 4.2.11 FEM electromechanical stress analysis

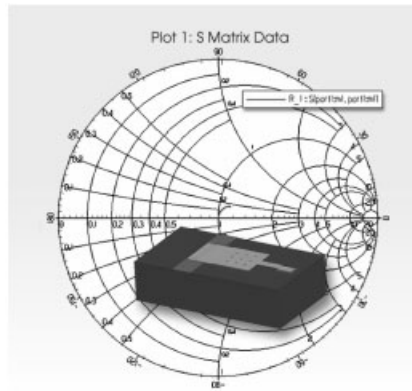
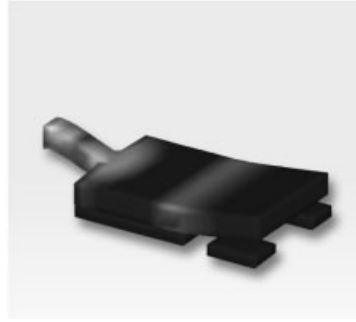


Fig. 4.2.12 FEM electromagnetic analysis

Figure 4.2.12 shows a 3D model analyzed by Ansoft's HFSS, back-dropped by the results of a detailed high-frequency electromagnetic (EM) simulation displaying the input impedance of the closed RF switch. The finite-element 3D field-solving approach is the most accurate EM analysis for an arbitrarily shaped structure, assuming that the designer knows the values for the material properties. From an arbitrary 3D structure, such as one contained in a deformed-geometry, SAT file output from Coventor's AnalyzerTM tool, a FEM electromagnetic solver can produce S parameter files (data-based models) for use in a system or subsystem simulator.

Demonstrating the system-level portion of the design flow, Figure 4.2.13 presents a schematic diagram of a simple one-bit microwave phase-shifter circuit within Agilent Technologies' Advanced Design System (ADS). FEM or system simulation-extracted models representing the open and closed states of the cantilever-beam RF switch supply the RF performance for an idealized switch model. Four instances (symbols in the corners of the schematic) of these switches are arranged to form the 2 signal paths. Depending on the polarity of the control signal, either the 'through-path' (0° path) or the 180° path is selected. A phase change is introduced by the additional length of the transmission line in the 180° path, effecting a time delay, which produces the 180° phase change in the desired frequency band (Fig. 4.2.14).

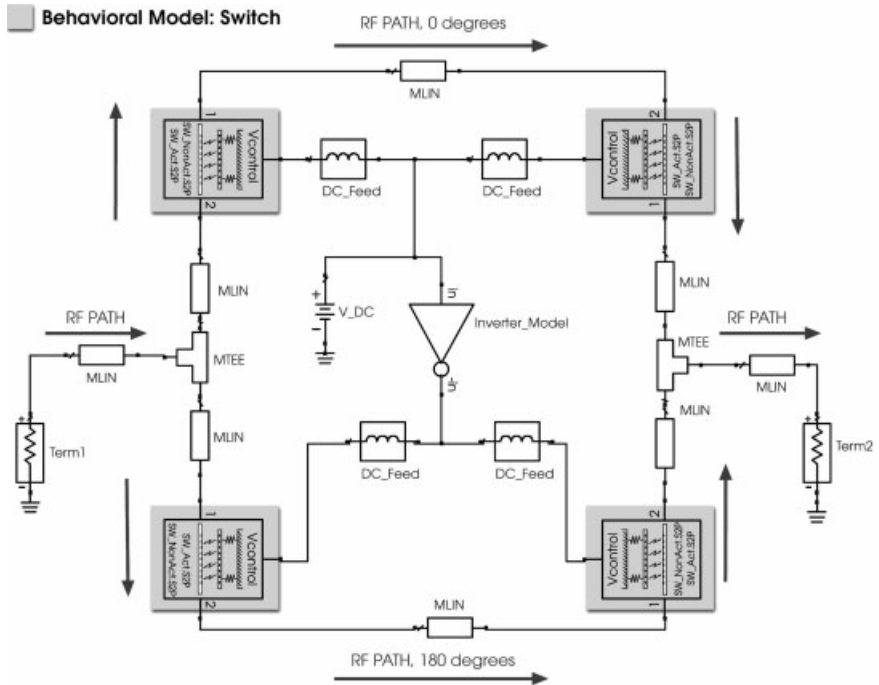


Fig. 4.2.13 One-bit phase shifter circuit model created in Agilent's Advanced Design System (ADS) using extracted switch models based on S parameter data

4.2.5

Conclusion

The unique characteristics of MEMS devices, such as multiple physical domains, manufacturing processes, and scaling, create unique challenges in the race to develop MEMS systems for automotive applications. Choosing software tools developed specifically for MEMS design can provide substantial advantages relating to time to market, development costs, and risk reduction. To master the challenge of designing commercially successful automotive MEMS products at a very low cost, the designer must be able to simultaneously model all of the relevant subsystem and their interactions in different but sufficient detail. This enables trade-off decisions to be made without requiring excessive computation and, especially, minimizing expensive manufacturing iterations. A software-driven design flow that supports both the top-down and the bottom-up design methodology strikes an ideal balance between the freedom and speed of system-level design and the detailed investigations of PDE solver numerical analysis, leading to an integrated design process that is more effective and significantly less time-consuming than previous methods.

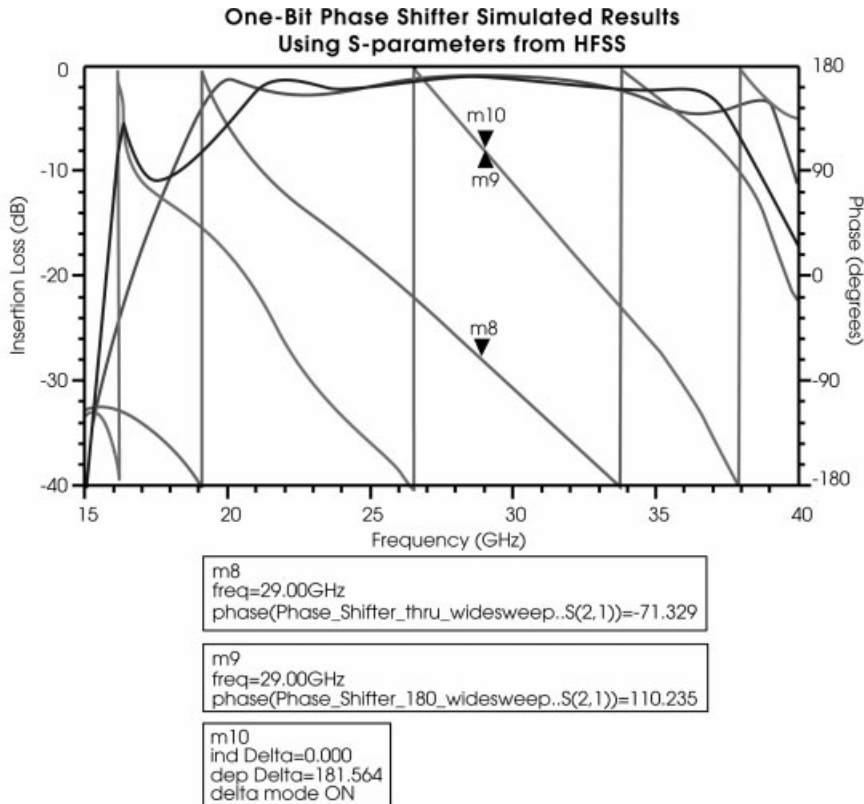


Fig. 4.2.14 ADS simulation results for the one-bit phase-shifter circuit

4.2.6

References

- 1 K. S. KUNDERT, *The Designer's Guide to Spice and Spectre*, Kluwer, 1995.
- 2 G. BREED, *A Heritage of Technology and Innovation: The History of Agilent EEs of EDA*, http://eesof.tm.agilent.com/pdf/history_2001.pdf (accessed Oct 2002), Agilent Technologies, Palo Alto, CA, USA.
- 3 H. A. MANTOOTH, M. FIEGENBAUM, *Modeling with an Analog Hardware Description Language*, Kluwer, 1995.
- 4 N. ZHOU, J. V. CLARK, K. S. J. PISTER, Nodal analysis for MEMS design using SUGAR version 0.5, *Proc. Int. Conf. Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators (MSM)*, Santa Clara, CA, USA, 1998, 308–313.
- 5 J. E. VANDEMEER, M. S. KRANZ, G. K. FEDDER, Hierarchical representation and simulation of micromachined inertial sensors, *Proc. Int. Conf. Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators (MSM)*, Santa Clara, CA, USA, 1998, 540–545.
- 6 G. LORENZ, R. NEUL, Network-type modeling of micromachined sensor systems, *Proc. Int. Conf. Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators (MSM)*, Santa Clara, CA, USA, 1998, 233–238.

- 7 D. TEEGARDEN, G. LORENZ, R. NEUL, How to model and simulate microgyroscope systems, *IEEE Spectrum*, July 1998, 66–75.
- 8 G. LORENZ, A. MORRIS, I. LAKKIS, A top-down design flow for MOEMS, *Proc. Design, Test, Integration, and Packaging of MEMS/MOEMS 2001*, Cannes, France, 2001, 126–137.
- 9 J.H. SCHAFFNER, R.Y. LOO, C. QUAN, R. C. ALLISON, B.M. PIERCE, S.W. LIVINGSTON, A.E. SCHMITZ, T.-Y. HSU, D.F. SIEVENPIPER, F.A. DOLEZAL, G.L. TANGONAN, *Microwave Components with MEMS Switches*, HRL Laboratories, LLC, Malibu CA, USA, 2000.
- 10 H.J. DE LOS SANTOS, *Introduction to Microelectromechanical (MEM) Microwave Systems*, Artech, Norwood MA USA, 1999.
- 11 C.T.-C. NGUYEN, Microelectromechanical components for miniaturized low-power communications, *Proceedings, 1999 IEEE MTT-S International Microwave Symposium RF MEMS Workshop (on Microelectromechanical Devices for RF Systems: Their Construction, Reliability, and Application)*, Anaheim, CA, June 18, 1999, 48–77.
- 12 R.J. RICHARDS, H.J. DE LOS SANTOS, MEMS for RF/microwave wireless applications: the next wave, parts I (March, 2001) and II (July, 2001), *Microwave J.*, 2001.
- 13 J.J. YAO, RF MEMS from a device perspective (topical review), *J. Micromechanics Microengineering* 10, 2000, R9–R38.
- 14 Agilent Technologies, *Agilent AN 154 S-Parameter Design, Application Note*, <http://cp.literature.agilent.com/litweb/pdf/5952-1087.pdf> (accessed Oct 2002) 1990, 2000 .
- 15 G. GONZALEZ, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ, USA, 1984, 1997.

5 Technology

DANIEL LAPADATU and HENRIK JAKOBSEN

5.1 Bulk Micromachining

5.1.1 Introduction

Micromachining is the generally accepted term that describes microfabrication of mechanical structures using several microelectronics or micromechanics processing techniques. Anisotropic, isotropic, deep and selective etching techniques, combined with the well-established photolithography process, make high-precision, miniaturized three-dimensional (3D) structuring possible [1–4].

Two distinctive ways of patterning a material in an axis perpendicular to the substrate surface are *digging* in the substrate (i.e., *bulk* micromachining) and *building up* starting from the substrate surface (i.e., *surface* micromachining, which is discussed in Section 5.2 of this book).

Bulk micromachining relies on several etching techniques and creates projections of planar photolithographic masks in 3 dimensions. Surface micromachining relies on sacrificial layer and wafer bonding techniques. It creates true 3D structures as a stack of 2D patterned layers. Hence, it is more correct to refer to both micromachining techniques as two and a half dimensional ($2\frac{1}{2}$ D).

Bulk and surface micromachining are complementary techniques, and hence, using them in combination produces the most versatile devices [5]. The development of SOI wafers [6] offered such a combination, allowing the production of typical surface-micromachined structures by employing only bulk micromachining.

5.1.2 Wet Etching

Several chemical etchants can etch silicon isotropically or anisotropically, be dopant-dependent or not, and have a wide range of selectivity to silicon, which determines the appropriate masking material. Brief but clear descriptions of the silicon chemical etchants and their properties can be found in [2–4, 7].

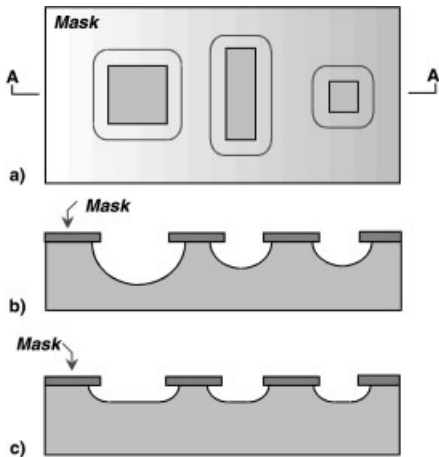


Fig. 5.1.1 Isotropic etching of silicon wafers: a) top view; b) cross section along AA' with agitation; c) cross section along AA' without agitation

Wet silicon etchants, due to their versatility and different behaviors, can be arranged into 3 distinct etching classes:

1) *HNA* (HF, HNO₃, CH₃COOH, and water [8]) is a complex etch system with highly variable etch rates and etch characteristics dependent on silicon dopant concentrations, the mix ratio of the three acids, the presence or absence of water, and even the degree of etchant agitation. The latter is typical of a diffusion-limited chemical reaction. For the same reason, HNA etches silicon isotropically.

2) *EDP* (ethylene diamine, pyrocatechol, and water [9]) etches silicon anisotropically and with a high selectivity, allowing a large variety of materials to be used as masks, such as silicon oxide and nitride, chromium, gold, and even aluminum.

3) *Aqueous alkaline* etchants, such as TMAH [10], KOH [7], NaOH [11], and CsOH [12], all etch silicon anisotropically with good selectivity to silicon nitride or oxide. They are all dopant dependent, etching silicon layers highly doped with boron slowly or not at all.

5.1.2.1 Isotropic Etching

Etchants with isotropic behavior, like HNA in water, act on silicon through a sequential process consisting of oxidation, dissolution, and diffusion steps, in which the overall chemical reaction is diffusion limited. Such a chemical reaction implies an insensitivity to surface phenomena (e.g., crystal orientation, defects, etc.), but also a strong sensitivity to local agitation, resulting in isotropic etching, smoothing, and rounding [8] (Fig. 5.1.1). The best mask for HNA etching is silicon nitride or gold. Silicon oxide is etched relatively rapidly, making its use as a masking material possible only for short etching times, i.e., shallow etchings.

5.1.2.2 Anisotropic Etching

The anisotropic behavior of some etchants, such as EDP, KOH, NaOH and TMAH, means that etching in the different crystallographic directions proceeds at different rates [7, 13, 14]. The most slowly etched directions are those in the $\langle 111 \rangle$

Fig. 5.1.2 Anisotropic etching on (100) silicon wafers: a) top view; b) cross section along AA'

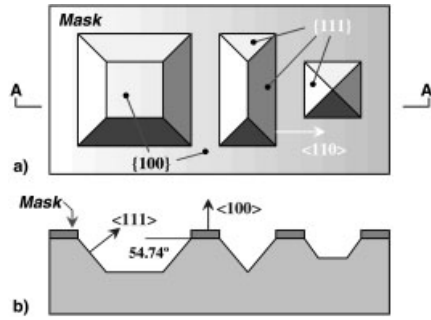
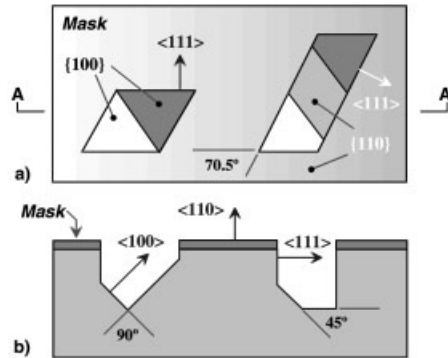


Fig. 5.1.3 Anisotropic etching on (110) silicon wafers: a) top view; b) cross section along AA'



family. The $\langle 100 \rangle$, $\langle 110 \rangle$, and all high-index directions are etched much faster. For this reason, if plenty of time is allowed, the resulting cavities are bound between $\{111\}$ planes. It is obvious that, for the same reason, the shape of the etched grooves in silicon depends on the wafer surface orientation. For example, on (100) wafers V-grooves, truncated pyramids, or full pyramids are obtained (Fig. 5.1.2); but on (110) wafers deep trenches with $\{111\}$ vertical walls and bottoms bounded between $\{100\}$ planes are obtained [2–4], if plenty of time is allowed to completely remove the (110) floor (Fig. 5.1.3).

5.1.2.3 Etch Stop Techniques

The behavior of all the wet etchants discussed above is also determined by the externally applied electric potential, if used. Etching in the presence of an electrical bias, either voltage or current, is known as electrochemical etching [12, 15].

No matter what etchant is employed, in order to achieve high predictability and reproducibility in the micromachining process, the etching step must be combined with some technique for stopping it automatically. Even the most casual examination of the present micromachining literature reveals that virtually all reported micromachined structures rely on at least one etch-stop technique. In its most comprehensive definition, an etch-stop process is any method that allows for selective removal of a specific material to produce a predefined relief.

The most suitable, successful etch-stop techniques are *electrochemical etch stop* (ECES), achieved by reverse biasing a *pn* junction [12, 15], and *high boron dose etch stop* [7, 16], which is explained below. Other etch-stop techniques have been reported [17], but they have not yet been used in mass production. A detailed review of etch-stop techniques for micromachining is given in [18].

The conventional high boron dose etch stop is based on the selectivity that some etchants display towards the doping level of silicon. The silicon etch rate in KOH [7], TMAH [19], and EDP [16] solutions, in particular, is reduced by over an order of magnitude at boron concentrations in excess of roughly 10^{20} cm^{-3} . The main benefits of the high boron dose etch stop are its independence from crystal orientation, the smooth surface finish, and especially the possibilities it offers for fabricating released structures with arbitrary lateral geometry in a single etch step. Double-sided etch work is possible, and many isolated *p*-islands can be etch-stopped at the same time in a single wafer. However, the required high boron doses introduce considerable mechanical stress into the material [2], which may even cause buckling in diaphragms or other multiple clamped structures. Moreover, introduction of electrical components into these microstructures for sensing purposes, such as piezoresistors, is inhibited by the excessive background doping. The latter consideration constitutes an important limitation to the applicability of the high boron dose etch stop technique.

In the basic form of ECES, the 2-electrode configuration, a *pn* junction, is reverse biased during etching by externally connecting a voltage source joining the *n*-layer, known as the work electrode, and an inert counterelectrode, usually platinum, that is immersed in the etchant. If the *n*-side of the wafer is protected from the etchant by a hermetically sealed etch holder or jig and biased *anodically* with respect to the passivation potential, etching occurs only in the *p*-substrate. In a 3-electrode ECES setup [15], a reference electrode is introduced for more accurate control over the potential of the etchant. A fourth electrode can be used to contact the *p*-type substrate, to gain direct control over the *pn* junction bias voltage [20]. The 4-electrode approach allows etch stopping on lower-quality wafers (larger leakage current) and better control of membrane thickness.

Fabrication of structures with arbitrary lateral geometry is possible by ECES if two or more junction levels are used, followed by an additional dry etching step, as is shown in a subsequent section.

A very important issue associated with anisotropic etching is *undercutting* of convex (or external) corners (Fig. 5.1.4). Because many different crystallographic planes are exposed at the corner, a non-90° angle on (100) wafers is always undercut until a 90° angle between the $\langle 110 \rangle$ directions is formed by the bounding $\{111\}$ planes. Hence, a multitude of fast-etching planes [21] can be observed around the convex corners. The undercutting rate depends strongly on the etchant used and the etching conditions.

To avoid undercutting, compensating structures must be used, generally simple polygonal extensions of the mask at the convex corners. A large variety of such convex corner-compensating structures have been proposed [22, 23].

Fig. 5.1.4 Undercutting of convex corners

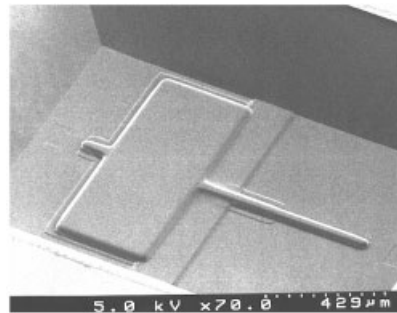
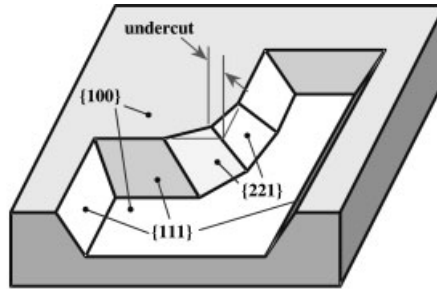


Fig. 5.1.5 Etch stop at several junctions with different depths

To fabricate creation of membranes of uniform, controllable thickness, an n -type epitaxial layer on top of a p -substrate is usually used to achieve an ECES against the resulting pn junction. This approach, however, has several limitations: (1) the fabricated membrane is always rectangular; (2) its lateral dimensions and position can vary significantly due to wafer thickness and the surface and flat-orientation tolerances; (3) the well known sharp transition from the $\{111\}$ to the $\{100\}$ crystal plane results in fragile clampings and suspensions; and (4) all released structures exhibit the same thickness.

On the other hand, by using more than one ion-implantation step followed by drive-in with or without a subsequent epitaxy step, adjacent etch-stop regions of different depths can be achieved [24, 25]. Shallower junctions can be created either by a shallow diffusion or by an epitaxial step.

Because the geometrical shape of the desired structures is defined by the implantation and drive-in processing steps, using etch stop on more than one junction, combined with a subsequent dry etching step, allows the fabrication of structures of any arbitrary shape without using the high boron dose technique. Undercutting of convex corners is also fully avoided. Moreover, the transition from a thick to a thin region has rounded contours, which are independent of crystal orientation, as can be seen in Figure 5.1.5.

Figure 5.1.6 shows a typical cross-section obtained by etch stopping on two junctions with different depths. The resulting profile has two important features:

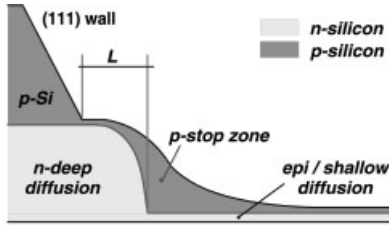


Fig. 5.1.6 Cross section showing the etching profile in transitions from regions of thick to thin membranes

the ECES *does not stop* at the metallurgical junction, but rather in the p -type substrate, and a transition zone from the thick to the thin membrane regions exists. The transition zone may extend over more than $50\ \mu\text{m}$. These features are discussed below.

Another major advantage of etch stopping on several junctions (Fig. 5.1.6) is the possibility of defining precisely the position and lateral extent of the resulting membrane. By using a deep n -type diffusion layer as the membrane's frame, in such a way that the bounding $\{111\}$ walls end up in these thicker n -regions, the edges of the membrane will be independent of the wafer thickness tolerance, surface and flat orientation tolerances, and front-to-back mask misalignment. The membrane's position is defined solely by the deep n -layer implantation mask, provided these tolerances do not add up to L , where L is defined (Fig. 5.1.6) as the separation between the diffused layer edge and the nominal termination point of the (111) side wall.

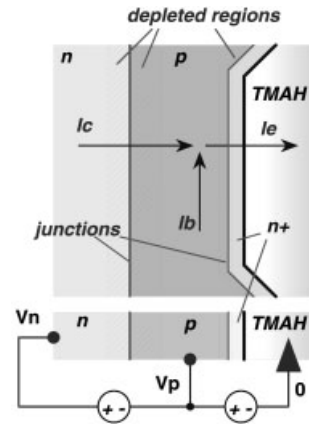
In most applications, accurate control of the thickness profile of the fabricated membrane or mass is required. Therefore, modeling the resulting etching profile and prediction of the etch stop's exact location are of a major importance in the sensor design field. Leakage diode currents, floating potential of the substrate toward anodic values, wafer taper, and bipolar effects in the n -layer/ p -substrate/liquid system were reported to generate premature etch stop [26, 27]. In the 4-electrode configuration, although all these effects are alleviated and membranes with thickness much closer to expected values are achieved, the etching often terminates *before* the metallurgical junction and even before the pn junction depletion region [20, 26–28].

An improved model [29] is based on the bipolar effect and on earlier conclusions [30] that n - and p -type silicon wafers immersed in alkaline solutions exhibit a depleted shallow electronic layer at their interface with the liquid phase.

Briefly, during the etching process, when the thickness of the neutral p -layer becomes thinner than the diffusion length of the electrons, a npn bipolar transistor is built up right under the silicon/liquid interface, as demonstrated in [27]. In such situations a large current, injected by the silicon/liquid interface, can flow through the reverse-biased pn junction – hence the typical current peak exhibited in all the records of the etching process.

Figure 5.1.7 presents the geometry of the induced npn transistor before the etch stop: the emitter consists of the induced shallow n^+ inversion layer together with

Fig. 5.1.7 Geometry of an induced bipolar npn transistor with biased base before etching



the charged layers residing in the liquid phase, the base is the p -substrate, and the collector is the n -layer.

As the etching proceeds, the substrate supply is cut off at the moment the emitter and the collector-base-depleted regions touch each other. This punch-through of the transistor results in a floating base potential, which rises suddenly toward more anodic values and instantly stops the etching. The thickness of the resulting stop zone is simply the sum of the two space-charge regions residing within the p -substrate.

The etching profile can be fully modeled by combining the etch stop model with the output of a process simulator, such as ATHENATM/SSUPREM3TM, which defines the location of the pn junctions and the electrical conductivity of the specific layers.

5.1.3

Dry Etching

Dry etching is a well established pattern-transfer technique in IC technology, characterized by a very good ability to pattern fine lines and by a high fidelity of pattern transfer [31]. It has become an important complementary technique for fabricating microstructures, after it has been adapted to the specific requirements of very deep etching.

The etchant is a gas plasma (Fig. 5.1.8), which can attack the silicon substrate in two ways: chemically (nondirectional etching), that is, by chemical reactions between silicon and the generated reactive radicals; or physically (directional etching), by material sputtering from the substrate as a result of collisions with the accelerated plasma ions.

Reactive ion etching is a combination of the two mechanisms: chemical reaction with electrically neutral radicals, which results in nondirectionality and high selectivity, and physical erosion due to accelerated ions, characterized by anisotropy and nonselectivity. The manner in which they interact varies with the gas/solid

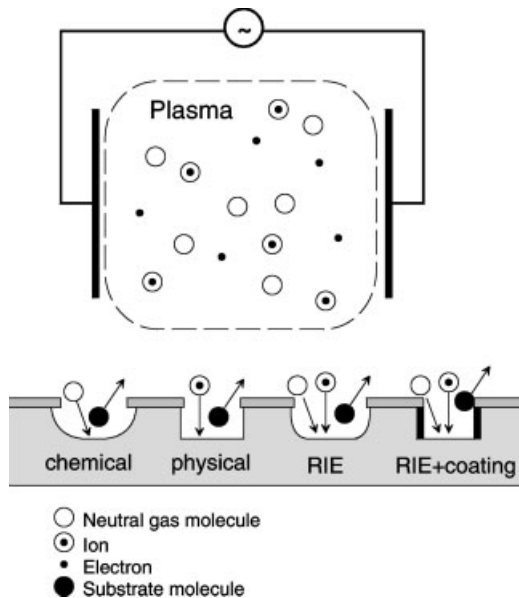


Fig. 5.1.8 The dry etching technique

system and etching conditions. As a result, simultaneous etching of exposed surfaces with both selectivity and anisotropy is possible.

A fourth possibility is *reactive ion etching with protective coating* of the lateral walls, which makes use of an inhibiting layer deposited on the lateral walls. This technique allows the unidirectional dry etching of very deep cavities [32].

Plasmas for etching are formed by applying a radio-frequency electric field to a gas held at low pressure in a vacuum chamber (upper part of Fig. 5.1.8). This makes plasma etching more expensive than wet etching.

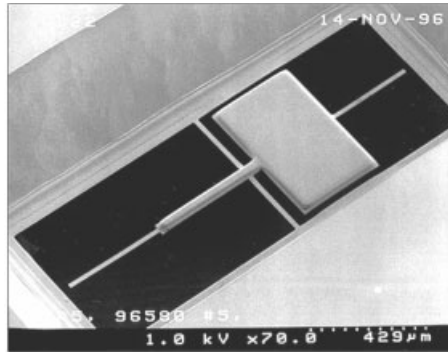
Plasma etchants are available for most of the films commonly used in micromachining. Etching of oxide, nitride, and silicon is usually done with fluorine-containing compounds such as SiF_4 , CF_4 , and SF_6 . Aluminum and other metals can be etched in chlorine plasmas such as SiCl_4 and BCl_3 . In certain conditions, better results can be obtained by combining several gases.

In the development of plasma etching processes, it is important to adjust the process parameters to meet requirements such as high etch rates for silicon, good selectivity over both mask and underlying layer, precise control of the line width, good uniformity, underetching as small as possible, and good control of the wall profiles.

Apart from the tunable parameters, such as RF power, gas composition and flow, and substrate temperature, several other factors affect the plasma etch rates, including impurities in or on the material to be etched, film stress, loading, elapsed time from the start of the etch, materials present in the plasma chamber, and wear of the electrode.

Although recognizing these many sources of etch-rate variations, the etch rates of the masking and underlying films for 317 different combinations of 16

Fig. 5.1.9 The result of dry etching performed on the structure shown in Figure 5.5



materials and 28 etches used in the fabrication of MEMS and integrated circuits are given in [14]. As an example, dry etching of silicon is used to perforate vertically structured membranes, such as the one shown in Figure 5.1.5, to define free-standing, movable mechanical structures with arbitrary lateral geometry (Fig. 5.1.9).

5.1.4

Micromachining of Glass

Similar to silicon, glass substrates can also be micromachined; however, the process is not straightforward, resulting in rather expensive processing sequences. Wet and dry etchants for glass, allowing the formation of deep recesses and even etch-through, have been developed. In addition, ultrasonic and laser drilling are alternative techniques.

5.1.5

Wafer Bonding

Wafer bonding techniques are developed to allow building of more complex 3D structures and to perform micropackaging on the wafer level.

The assembly of silicon microsensors into specially designed low-cost packaged units makes bonding on a wafer level necessary. Although derived from standard semiconductor and hybrid electronic-packaging technology, sensor packaging techniques must meet several specific microtechnology requirements: positioning and alignment in the micrometer range, the possibility of selective bonding, handling of fragile etched silicon wafers, and compatibility with the fabrication process of the sensor. The sensor package must allow the delivery of the measurand to the sensitive components, while protecting them from other influencing parameters and from the potentially hostile, destructive environment.

Microsensors are usually packaged by bonding the silicon wafer, which contains the sensitive devices, to dedicated support wafers; in some cases these support wafers become part of the sensor itself.

Several aspects must be taken into account when designing the sensor and its package [2]:

- Electrical characteristics, which must comply with the need of stable operation, minimization of interference and parasitic signals, high accuracy, and reliability.
- Thermal properties related to the need to preserve good thermal insulation or temperature uniformity.
- Mechanical considerations dealing with induced stresses and preventing mechanical failure.

Reducing package-induced stress to levels that do not affect the output of the sensor is by far the most difficult and challenging task. Thermal mismatch between the different materials, combined with the fact that wafer bonding usually takes place at elevated temperatures, generates a residual thermal stress in the active devices, which negatively affects their specifications. For instance, compressive stress can easily lead to buckling of clamped structures such as membranes and beams, making them nonfunctional. In some types of sensors, the transduction phenomenon itself is based on the dependence of some electrical properties (resistivity, resonance frequency) on the stress level within the structure. It is thus obvious that a ‘parasitic’ stress induced by the packaging is to be avoided or at least minimized. Hence, low-temperature bonding is desirable.

Cost considerations introduce additional restrictions. Sensor packaging can represent a substantial part of the total fabrication cost, being sometimes several orders of magnitude higher than the cost of the silicon part.

Different bonding techniques have been developed for use in microsensors, each owing its popularity to some specific advantages, although none of them meets all the requirements. Only a brief review is given below; for detailed information please see [33, 34].

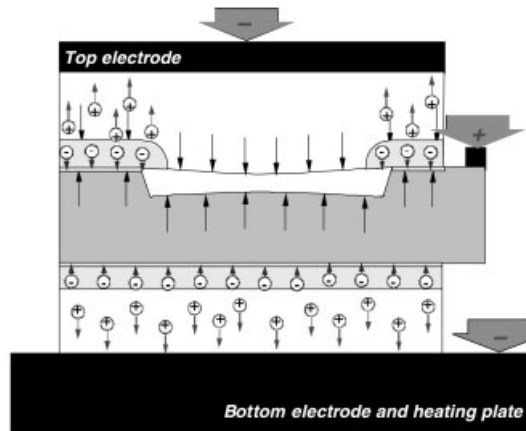
5.1.5.1 Anodic Bonding

Among the bonding techniques that have been developed for silicon sensors, *anodic bonding* has proved to be a reliable method for packaging at a wafer level and has therefore gained an important place in recent years. In its most-used form, anodic bonding consists in joining the silicon wafer with a borosilicate glass wafer, such as Pyrex #7740.

Anodic bonding is a relatively simple process, which provides hermetic seals by bonding silicon wafers to glass plates at rather low temperatures (300–450°C), with the assistance of an electrostatic field (400–800 V), such that the glass wafers are negative with respect to the silicon [35].

At the bonding temperature, the alkali oxides present in the glass dissociate into positive ions, primarily Na^+ , and negative oxygen ions. Under the action of the electric field, the mobile sodium ions quickly leave the region near the anode, and a depleted layer about 1 μm thick is formed (Fig. 5.1.10). The high electrostatic forces that develop across this layer move the nonbridging oxygen ions and direct them towards the silicon anode, where they oxidize the silicon and form strong silicon-oxygen bonds [34, 36]. The wafers to be bonded must be polished to

Fig. 5.1.10 Triple-stack anodic bonding setup and process; one silicon wafer is anodically bonded between two glass wafers



a smooth surface and carefully cleaned prior to the bonding step. Standard polished silicon wafers pose no problem for anodic bonding.

Foreign particles on the mating surfaces produce nonbonded regions around them, whose dimensions depend on the height of the particles, the bonding temperature, and the applied voltage [34]. The required cleaning procedure depends on the wafers' history. Both hydrophilic and hydrophobic (Si–OH and Si–H terminated chains, respectively) surfaces can bond. Special cleaning procedures, including a 'local clean room', have been developed and are now available with some commercial bonding equipment.

The quality of the bond can be evaluated by three main criteria: void density, bond strength, and residual thermal stress. However, their simultaneous optimization is a rather difficult task. For instance, a higher bonding temperature results in a very low void density and strong bonds, but increases the residual stress. In conclusion, depending on the eventual application, different bonding recipes need to be used.

A good design of the structures to be bonded, together with state-of-the-art sealing equipment, now enable fast, reliable bonding over a whole wafer, for multiple wafer stacks, and for many wafers simultaneously.

5.1.5.2 Fusion Bonding

Silicon-to-silicon fusion bonding involves a high-temperature (1000°C) treatment to strengthen the weak bond formed after hydrophilization and contact at room temperature of two polished silicon wafers [37, 38]. The high-temperature step rules out the use of metal layers. Moreover, bonding failure may occur around cavities due to expansion of trapped air, in addition to the fact that the bond quality decreases drastically with every surface imperfection.

Silicon fusion bonding has been extensively studied during the past few years not only for sensor applications, but also because it represents a good alternative for the fabrication of high quality silicon-on-insulator (SOI) wafers [39].

5.1.5.3 Other Wafer-Bonding Techniques

Silicon-to-silicon bonding with an intermediate thin glass layer produced by sputtering, evaporation, or spin-on of silicate solutions, avoids the thermal effects, but suffers from low deposition rates of the thin glass film and bubble formation within it [40, 41].

Adhesive bonding uses a two-component epoxy mixture as an intermediate adhesive layer [42]. Although the epoxy layer ensures good coverage even over high-topography steps, allows very good bonding, and is performed at relatively low temperatures (65–150 °C), it cannot be used on wafers that contain cavities.

Eutectic bonding is a process adopted from IC packaging and uses an eutectic gold-based alloy preform, such as 80% Au/20% Sn [43]. Slow creeping of the eutectic alloy is the main drawback of this technique.

Thermocompression bonding is a process in which soft metals such as Au are squeezed into each other with high force at elevated temperature. Thin films of metals can be deposited on the surface of wafers by evaporation, sputtering, or plating.

5.1.5.4 SOI Wafers

Figure 5.1.11 shows a general cross-section through an SOI wafer. The wafer consists of 2 silicon layers, a thick substrate (400 μm), and a thinner top layer (0.2 to 60 μm), separated by a thin insulating oxide layer (0.4 to 1.2 μm).

Two main ways of fabricating an SOI wafer [44] are

- BESOI wafers are made by bonding 2 oxidized silicon wafers and thinning one of them until the required thickness is reached.
- SIMOX wafers are made by implanting oxygen at a certain depth into a conventional silicon wafer.

Both kinds of SOI wafers have the same properties with respect to the substrate. The differences appear in the top silicon and the insulating oxide layers.

The *smart cut*, a recently developed technique at Leti, Grenoble, France [45], allows even better control of the top layer's thickness while reclaiming the top silicon wafer used in BESOI wafers. It makes use of implanted hydrogen at the separation depth. The smart cut technique is used in Unibond wafers.

Two major distributors of SOI wafers are Soitec SA (Grenoble, France) and Shin-Etsu Handotai (Tokyo, Japan).

SIMOX and Unibond wafers require epitaxy to increase the thickness of the top layer. An epitaxial layer up to 60 μm thick can be grown on such wafers; the resulting thickness tolerance depends on the epitaxy process tolerances and is typically in the range of 5% of the final thickness (4σ value).



Fig. 5.1.11 Cross section through an SOI wafer

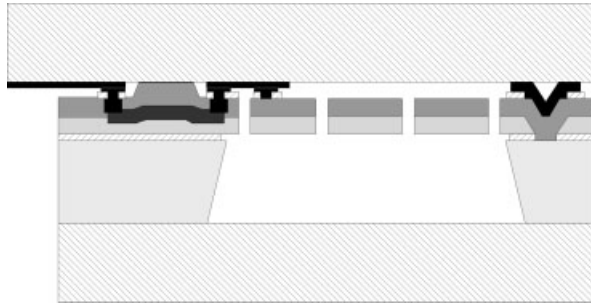


Fig. 5.1.12 Representative bulk-micromachined structure made from SOI wafers by anodic bonding

Figure 5.1.12 shows a typical bulk-micromachined structure fabricated on an SOI wafer by use of the previously described etching and bonding techniques.

5.1.5.5 Sealed Cavities and Press Contacts by Anodic Bonding

Most mechanical sensors have to operate in a stable, well controlled environment, thus requiring perfectly sealed cavities. One of the most important achievements of the anodic bonding technique is the fabrication of such hermetically sealed cavities. Anodic bonding can be performed in different atmospheres and at different pressures, which can both be preserved within a sealed cavity. The creation of reliable reference chambers for pressure sensors or for good control of squeeze-film damping [46] are just some of the opportunities offered by this technique. Nevertheless, this technique also faces some challenges, such as the feed-through of the conductors to and from the sealed cavity. Several solutions have been proposed:

- *Direct crossing* with metal conductors through the bonded area. However, this approach destroys the hermeticity of the cavity and cannot be used for long-term applications.
- *Two-layer polysilicon with backlap/polish* is a reliable technique when used in combination with surface micromachining.
- *Diffused conductors* maintain the hermeticity of the sealed cavity, but, due to the large sodium contamination of the silicon-glass interface, large leakage currents destroy the electrical insulation between the different conductors.
- *Diffused and buried feed-throughs* preserve the advantage of the previous approach while avoiding the negative effects at the silicon-glass interface [47].

Figure 5.1.13 shows in detail the special feed-through developed at and patented by SensoNor asa: within an n -well, several p -type diffused islands are created for each required crossing. An n -epi layer is grown on top, to bury the p -conductors. Highly boron-doped layers are implanted and diffused through the epitaxial layer to contact the buried conductors. A shallow n^+ layer is also implanted immedi-

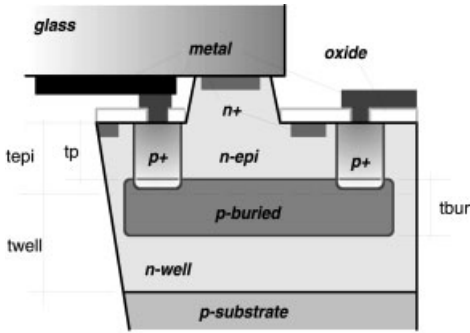


Fig. 5.1.13 Cross section of a diffused, buried feed-through

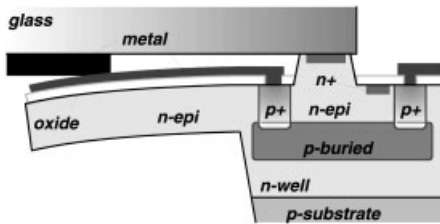


Fig. 5.1.14 Microspring conductor transfer

ately under the anodically bonded areas, to minimize the electrical effects of the silicon-glass interface on the junctions beneath.

The conductors can be transferred to and from the silicon and glass wafers by so-called *press contacts* – the name comes from the fact that the two metallic layers are pressed against each other during the anodic bonding step.

The standard solution, involving a fixed recess formed either in the silicon or in the glass wafer, is shown in Figure 5.1.13. The total thickness of the two metallic films, allowing for their thickness tolerances as well, must exceed the recess depth by at least 100 nm, and at least one of the conductors must be a soft metal, such as aluminum or gold. During bonding, the soft metal is easily plastically deformed and, for metals like aluminum, its native oxides are broken apart, resulting in good ohmic contact.

However, in certain applications, hard metals are required on both bonded wafers. In these cases, the so-called *microspring* solution (Fig. 5.1.14) is a good alternative.

To avoid trapping gas at the bonding interface, the glass wafers receive some cuttings prior to bonding. However, once bonding is accomplished, the gas retained in the sealed cavities can no longer escape. It turns out that trapped gaseous reaction products originating at the bonding interface can alter the final gas pressure within the cavity. A more detailed presentation of these phenomena can be found in [34] and [48].

5.1.6

Examples

To show the strength and varied possibilities of bulk micromachining, some examples used in automotive products are presented in this section.

The first example is a technology originally patented by S. Suzuki et al., Hitachi Ltd., Japan [49]. Their solution for absolute pressure sensors, as well as for relative pressure sensors, is shown in a cross-sectional view in Figure 5.1.15. The piezoresistive silicon sensor element is anodically bonded to a thick glass part that constitutes the vacuum reference volume in the absolute pressure sensor or that contains a hole as a pressure inlet port for the relative pressure sensor. The pressure sensor die is typically housed in a cavity package with pressure inlet ports as part of the body. The surface of the sensor element is usually protected with a gel or other flexible material for corrosion resistance.

Figure 5.1.16 shows a design in which fusion bonding is used to obtain a small die size, as first presented by J. Bryzek et al., Novasensor, California [50]. A wafer that was selectively etched and a plane wafer are fused at high temperature; followed by thinning the nonetched wafer to form the silicon diaphragm. After completing the processing of the silicon part by making passivated piezoresistors and metal interconnects, the fusion-bonded silicon structure is anodically bonded to a glass substrate to form the vacuum reference volume under the diaphragm. The sensor element is housed in a cavity package, with corrosion protectants for the metallic components and the diaphragm as in the first example.

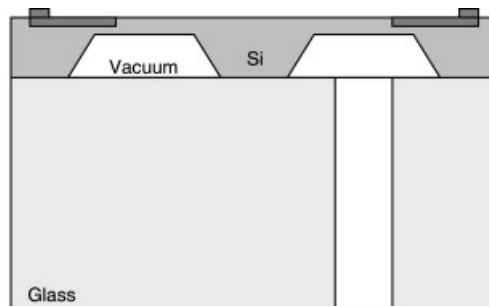


Fig. 5.1.15 Anodic-bonded silicon sensor with sensor for absolute pressure (left) and relative pressure (right) on the same chip

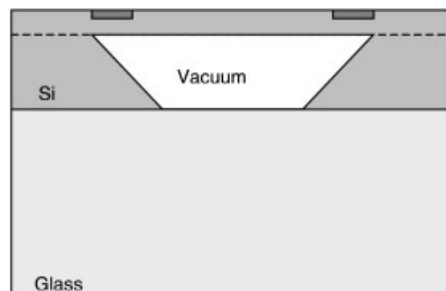


Fig. 5.1.16 Piezoresistive pressure sensor made by silicon fusion-bonding (SFB) to obtain small chip size and anodic bonding to form a sealed vacuum cavity

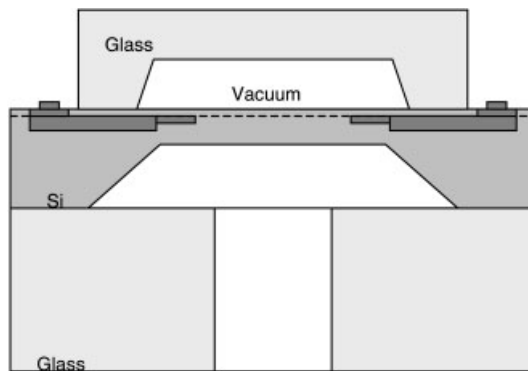


Fig. 5.1.17 Absolute pressure sensor for good media compatibility with pressure inlet port on the bottom of the silicon diaphragm and an on-chip vacuum reference volume, made possible by triple-stack anodic bonding

The piezoresistive absolute pressure sensor patented by H. Jakobsen et al., SensoNor, Norway [51] (cross-sectional view in Fig. 5.1.17), can be packaged in a low-cost epoxy molding as used in the IC industry. The device has a pressure inlet port on the bottom of the diaphragm through an etched hole in the glass anodically bonded to the silicon. The vacuum reference volume is formed by anodic bonding of a glass wafer to the metallized top of the silicon wafer. Diffused buried conductors form electrical connections under the sealed area between the buried piezoresistors in the diaphragm and wire bond pads outside the sealed cavity. The result is a robust and reliable design, with no need for additional corrosion protection, useful for measurements of gas pressures such as in monitoring tire pressure.

Many companies offer bulk micromachined capacitor sensors. A capacitive accelerometer from VTI-Hamlin (Fig. 5.1.18) [52] is built as a die sandwich with three silicon parts put together on the wafer level using glass films to form the seals. A mass-spring system is etched in the middle silicon part, with the springs in the center plane for good symmetry and low cross-axis sensitivity. Counter electrodes on the other two silicon parts face the movable mass and give a differential solution by means of two matched capacitors. Metal bonding pads are formed on one side of the die sandwich, with one pad for each of the three silicon parts. This technology is typically used for low-g accelerometers.

Development of standardized processes with high design flexibility to enable forming a large variety of different sensors on the same wafer or different sensors on the same die is made possible by bulk micromachining. The device (Fig. 5.1.19) has a pressure sensor and an accelerometer on the same chip, made possible by SensoNor's standard process SW4. The die sandwich includes on the left: an absolute pressure sensor, with a pressure inlet hole through the bottom glass, a silicon diaphragm with high-stability piezoresistors in the diaphragm, and a vacuum reference volume on the same side as the wire bonding pads; and on the right: a piezoresistive accelerometer, with its mass-spring system sealed in a sealed cavity. The process technology is available for custom design of microsystems through the European manufacturing service network Europractice.

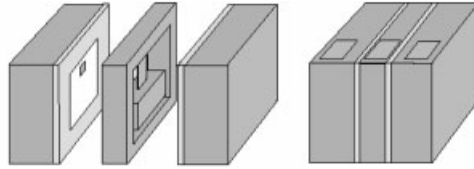


Fig. 5.1.18 Bulk micromachined capacitive accelerometer chip from VTI Hamlin OY showing a mass-spring system in the center chip and counterelectrodes on both side chips before and after sealing

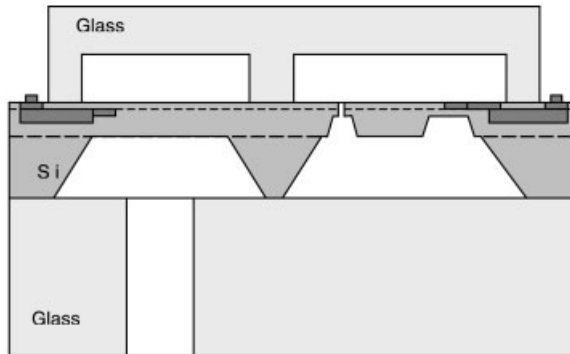


Fig. 5.1.19 Silicon pressure sensor (left) and accelerometer (right) on the same chip, made possible by one of SensoNor's standardized sealed-cavity processes

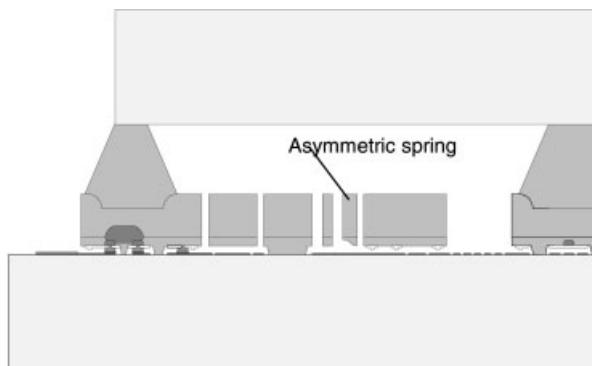


Fig. 5.1.20 Bulk-micromachined angular rate sensor from SensoNor with asymmetric spring for setting up an in-plane excitation movement by applying a vertically oriented electrostatic force

A more advanced process that can include capacitive sensors, electrostatic actuators, and piezoresistive sensors on the same chip is also made possible by bulk micromachining.

Figure 5.1.20 shows a cross section of a bulk-micromachined angular rate sensor from SensoNor asa [53] that uses electrostatic actuation to set up vibrations and capacitance to detect the Coriolis force. The sensor is a glass-silicon-glass die sandwich put together on the wafer level to achieve low production cost. The sealed cavity houses a mass-spring system fixed to the glass by anodic bonding. Near in-plane vibrational motion is excited by applying electrostatic force from electrodes on the glass to springs having asymmetric cross sections. Electrodes on the glass detect the capacitance change that results when an angular velocity is applied perpendicular to the excited vibration in the mass-spring system, which results in the appearance of Coriolis forces. This device is fabricated by combining ion implantation and high-temperature diffusion of silicon, selective etching that stops against a *pn*-junction to form the thickness of the structure, and dry etching to sculpt the mass-spring system. Triple-stack anodic bonding is performed with fabricated metal interconnects on the glass, to form press contacts between the metal tracks on the glass and the metal areas on the silicon.

5.1.7

References

- 1 G. T. A. KOVACS, N. I. MALUF, K. E. PETERSEN, Bulk micromachining of silicon, *Proc. IEEE* **1998**, *86*, 1536–1551.
- 2 K. E. PETERSEN, Silicon as a mechanical material, *Proc. IEEE* **1982**, *70*, 420–457.
- 3 J. C. GREENWOOD, Silicon in mechanical sensors, *J. Phys.: Electr.* **1988**, *21*, 1114–1128.
- 4 R. M. LANGDON, Micromachining of silicon for sensors, *Proc. NATO ASI on Novel Si-Based Technologies* **1989**, 143–172.
- 5 S. LEE, S. PARK, D. CHO, The surface/bulk micromachining (SBM) process: a new method for fabricating released MEMS in single crystal silicon, *IEEE/ASME J. Microelectromech. Syst.*, **1999**, *8*, 409–416.
- 6 Y. KANDA, What kind of SOI wafers are suitable for what micromachining purposes, *Proc. Transducers*, **1991**, *91*, 452–455.
- 7 H. SEIDEL, L. CSEPREGI, A. HEUBERGER, H. BAUMGÄRTEL, Anisotropic etching of crystalline silicon in alkaline solutions, *J. Electrochem. Soc.* **1990**, *137*, 3612–3632.
- 8 B. SCHWARTZ, H. ROBBINS, Chemical etching of silicon, IV. etching technology, *J. Electrochem. Soc.* **1976**, *123*, 1903–1909.
- 9 A. REISMAN, M. BERKENBLIT, S. A. CHAN, F. B. KAUFMAN, D. C. GREEN, The controlled etching of silicon in catalyzed ethylenediamine-pyrocatechol-water solutions, *J. Electrochem. Soc.* **1979**, *126*, 1406–1415.
- 10 O. TABATA, pH-controlled TMAH etchants for silicon micromachining, *Sens. Actuators* **1996**, *A53*, 335–339.
- 11 P. ALLONGUE, V. COSTA-KIELING, H. GERISCHER, Etching of silicon in NaOH solutions, *J. Electrochem. Soc.* **1993**, *140*, 1009–1026.
- 12 V. M. MCNEIL, S. S. WANG, K.-Y. NG, M. A. SCHMIDT, An investigation of the electrochemical etching of (100) silicon in CsOH and KOH, *Proc. Solid State Sensors and Actuators Conf. Hilton Head Isl*, **1990**, 92–97.
- 13 A. MERLOS, M. ACERO, M. H. BAO, J. BAUSSELLS, J. ESTEVE, TMAH/IPA anisotropic etching characteristics, *Sens. Actuators* **1993**, *A37/38*, 737–743.

- 14 K. R. WILLIAMS, R. S. MULLER, Etch rates for micromachining processing, *J. Microelectromech. Systems* **1996**, 5-4, 256–269.
- 15 E. D. PALIK, O. J. GLEMBOCKI, I. HEARD Jr., Study of bias-dependent etching of Si in aqueous KOH, *J. Electrochem. Soc.* **1987**, 134, 404–409.
- 16 N. F. RALEY, Y. SUGIYAMA, T. VAN DUZER, (100) Silicon etch-rate dependence on boron concentration in ethylenediamine-pyrocatechol-water solutions, *J. Electrochem. Soc.* **1984**, 131, 161–171.
- 17 E. PEETERS, D. LAPADATU, W. SANSEN, R. PUERS, PHET – an electrodeless photo-voltaic electrochemical etchstop technique, *J. Microelectromechanical Systems* **1993**, 3, 113–123.
- 18 S. D. COLLINS, Etch stop techniques for micromachining, *J. Electrochem. Soc.* **1997**, 144, 2242–2262.
- 19 E. STEINSLAND, M. NESE, A. HANNEBORG, R. W. BERNSTEIN, H. SANDMO, G. KIT-TILSLAND, Boron etch-stop in TMAH solutions, *Sens. Actuators* **1996**, A54, 728–732.
- 20 B. KLOECK, S. D. COLLINS, N. F. DE ROOIJ, R. L. SMITH, Study of electrochemical etchstop for high-precision thickness control of silicon membranes, *IEEE Trans. Electron Devices* **1989**, ED-36, 663–669.
- 21 E. HERR, H. BALTES, KOH etching of high-index crystal planes in silicon, *Sens. Actuators* **1992**, A31, 283–287.
- 22 R. PUERS, W. SANSEN, Compensation structures for convex corner micromachining in silicon, *Sens. Actuators* **1990**, A21–23, 1036–1041.
- 23 M. BAO, C. BURRER, J. ESTEVE, J. BAUSELLS, S. MARCO, Etching front control of <110> strips for corner compensation, *Sens. Actuators* **1993**, A37/38, 727–732.
- 24 R. HUSTER, A. SOFFEL, Vertically structured silicon membranes by electrochemical etching, *Sens. Actuators* **1990**, A21–23, 899–903.
- 25 D. EICHNER, W. VON MÜNCH, A two-step electrochemical etch-stop to produce free-standing bulk-micromachined structures, *Sens. Actuators* **1997**, A60, 103–107.
- 26 V. MCNEIL, A thin-film silicon micro-accelerometer fabricated using electrochemical etch-stop and wafer bonding technology, *Ph.D. Thesis*, Massachusetts Inst. of Technology, Boston, MA, USA, **1994**.
- 27 M. K. ANDREWS, G. C. TURNER, Bipolar effects in the fabrication of silicon membranes by the anodic etch stop, *Sens. Actuators* **1991**, A29, 49–57.
- 28 A. GÖTZ, J. ESTEVE, J. BAUSELLS, S. MARCO, J. SAMITIER, J. R. MORANTE, Passivation analysis of micromechanical silicon structures obtained by electrochemical etch stop, *Sens. Actuators* **1993**, A37/38, 744–750.
- 29 D. LAPADATU, G. KIT-TILSLAND, M. NESE, S. M. NILSEN, H. JAKOBSEN, A model for the etch stop location on reverse biased pn junctions, *Sens. Actuators* **1998**, A66, 259–267.
- 30 D. LAPADATU, Limits in miniaturisation of bulk micromachined capacitive accelerometers, *Ph.D. Thesis*, K. Univ. Leuven, Belgium, **1996**.
- 31 R. J. SCHUTZ, in *VLSI Technology* (ed. S. M. SZE), McGraw-Hill, New York, NY, USA **1988**, p. 185.
- 32 H. JANSEN, M. DE BOER, R. LENGTENBERGH, M. ELWENSPOEK, The black silicon method: a universal method for determining the parameter setting of a fluorine-based reactive ion etcher in deep silicon trench etching with profile control, *Proc. Micromechanics Europe* **1994**, 60–64.
- 33 W. H. KO, J. T. SUMITO, G. H. YEH, *Bonding Techniques for Microsensors, Micromachining and Micropackaging of Transducers*, Elsevier, Amsterdam, Netherlands **1985**, p. 41.
- 34 A. COZMA, Development of wafer bonding techniques with application to self-testable silicon pressure sensors, unpublished, *Ph.D. Thesis*, K. Univ. Leuven, Belgium, **1998**.
- 35 G. WALLIS, D. I. POMERANTZ, Field-assisted glass-metal sealing, *J. Appl. Phys.* **1969**, 40, 3946–3949.
- 36 K. B. ALBAUGH, Electrode phenomena during anodic bonding of silicon to sodium borosilicate glass, *J. Electrochem. Soc.* **1991**, 138, 3089–3094.
- 37 R. STENGL, T. TAN, U. GOSELE, A model for the silicon wafer bonding process, *Jpn. J. Appl. Phys.* **1989**, 28, 1735.

- 38 C. HARENDT, B. HÖFFLINGER, H. G. GRAF, E. PENTEKER, Silicon direct bonding for sensor applications: characterisation of the bond quality, *Sens. Actuators* **1991**, A25–27, 87–92.
- 39 W.P. MASZARA, Silicon-on-insulator by wafer bonding: a review, *J. Electrochem. Soc.* **1991**, 138, 341–347.
- 40 A. HANNEBORG, M. NESE, H. JAKOBSEN, R. HOLM, Silicon-to-thin film anodic bonding, *J. Micromech. Microeng.* **1992**, 2/3, 117–121.
- 41 H. J. QUENZER, W. BENECKE, Low-temperature silicon wafer bonding, *Sens. Actuators* **1992**, A32, 340–344.
- 42 S. VAN DER GROEN, M. ROSMEULEN, P. JANSEN, K. BAERT, L. DEFERM, CMOS compatible wafer scale adhesive bonding for circuit transfer, *Proc. Transducers* **1997**, 97, 629–632.
- 43 W.H. KO, J. HYNCEK, S. F. BOETTCHER, Development of a miniature pressure transducer for biomedical applications, *IEEE Trans. Electron Devices* **1979**, ED-26, 12, 1896–1905.
- 44 J.-P. COLINGE, *Silicon-on-Insulator Technology. Materials to VLSI*, Kluwer, **1991**.
- 45 M. BRUEL, Silicon on insulator material technology, *Electron. Lett.* **1995**, 31, 1201.
- 46 SensoNor asa, *Data Sheet: SP13 Absolute Pressure Sensor*, SensoNor, Horten, Norway, **1997**.
- 47 A. COZMA, H. JAKOBSEN, R. PUERS, Electrical characterisation of the anodically bonded wafers, *J. Micromech. Microeng.* **1998**, 8, 69–73.
- 48 S. MACK, H. BAUMANN, U. GÖSELE, H. WERNER, R. SCHÖGL, Analysis of bonding-related gas enclosure in micromachined cavities sealed by silicon wafer bonding, *J. Electrochem. Soc.* **1997**, 144, 1106–1111.
- 49 S. SUZUKI et al., Hitachi, Tokyo, Japan, US patent 4322980.
- 50 J. BRYZEK et al., *Silicon Sensors and Microstructures*, Novasensor, Fremont, CA, USA.
- 51 H. JAKOBSEN et al., SensoNor asa, Horten, Norway; US patent 5591679.
- 52 Drawing from HANNU LAATIKAINEN, VTI Hamlin OY, Vantaa, Finland.
- 53 T. KVISTEROY et al., SensoNor asa, Horten, Norway; US patent 6319729 B1.

5.2

Integrated Surface Micromachining Technology

BOB SULOUFF

5.2.1

Introduction

Constructing mechanical structures in or on the surface of a silicon substrate is commonly referred to as surface micromachining. The surface micromachining can be combined with integrated circuit (IC) processes to create a single-chip or integrated microelectromechanical system (IMEMS) [1]. The most common mechanical material used for surface micromachining has been polycrystalline silicon; however, other materials applied by electroplating [2] or vapor depositions have also been used [3]. The defining characteristic for this approach is the temperature profiles that occur during processing. To achieve the desired mechanical properties of polysilicon as well as the flatness on which to construct useful structures, high temperatures, in excess of 900°C for several hours, are required. The thermal budget for IC processing must also be compatible with the surface-micromachined devices. Submicron complimentary metal oxide semiconductor (CMOS)

Fig. 5.2.1 Integrated surface micro-machined accelerometer

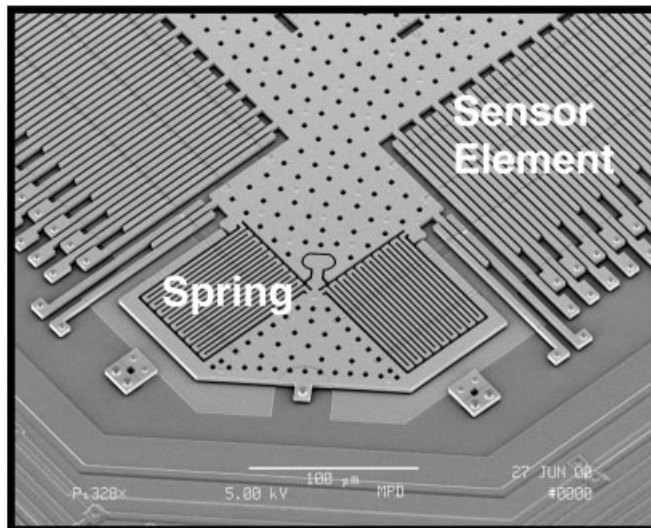
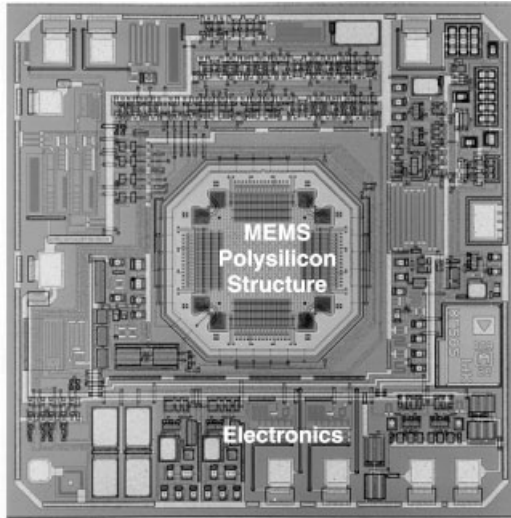


Fig. 5.2.2 Surface micromachined polysilicon structure

structures that have become widely used in digital devices degrade at high temperature and must be processed after mechanical structures that require high-temperature processes. Thus, not only the polysilicon but the electronics process must be thermally compatible. When structural sensor materials are made by plating or deposition, subsequent high-temperature processing could degrade the sensor elements.

Figure 5.2.1 shows a photograph of an integrated surface-micromachined accelerometer, and Figure 5.2.2 is a close-up of the surface-micromachined polysilicon MEMS structure near the center of Figure 5.2.1. The structural material is 3 μm thick polysilicon and the IC process is bipolar and CMOS (BiCMOS) with thin film resistors. This structure thus combines bipolar transistors, CMOS, precision laser-trimmed resistors, and mechanical polysilicon [4].

The integration of sensors and electronics makes it possible to measure very low-level signals and minimizes the number of interconnections at the package level. The single-chip configuration eliminates capacitive parasitics that occur when a sensor element is interconnected with an electronics interface chip. The detectable signal level is improved by eliminating noise due to such interconnections. The sensor can thus be made smaller for a given signal strength, or lower level signals can be measured when the sensor and electronics are integrated. Additional sources of error, from thermal variations and mechanical stress gradients, can be minimized with integrated surface micromachining. However, the additional time needed to process a more complex design and the costs associated with IC processing are factors that must be considered.

5.2.2

Structures of Surface Micromachined Devices

Deposition and etching of surface films and removal of a sacrificial layer under the structural film can produce many different shapes. The most common shape is a capacitive structure that is attached to the single-crystal silicon substrate at an anchor and is free to move above the surface (Fig. 5.2.3).

Integration of the sensor structure with silicon microelectronic fabrication is accomplished by progressing in stages. First, the IC components are created by high-temperature processes. Then the mechanical polysilicon sensor structure is created. Finally, lower-temperature metal and passivation are applied to complete the structure. A topographical area is reserved for the sensor during IC processing. Silicon nitride deposited by low pressure chemical deposition (LPCVD) is then deposited to protect the circuitry as well as to act as an etch-stop layer when the sacrificial layer under the polysilicon mechanical structure is etched. To connect the mechanical structure to the circuitry, polysilicon is used along with n+ doped regions (n+ runners). This allows signals to be transferred from the perimeter of the sensor to mechanical polysilicon anchors within the boundaries of the mechanical structure.

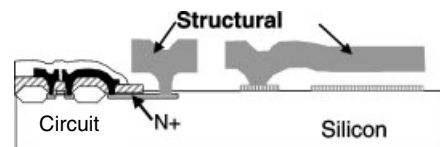


Fig. 5.2.3 Cross-section of integrated surface micromachined device

Structures fabricated from polysilicon are defined not only by the mechanical structure but by variations in the film's properties and internal stresses. Stress gradients through the film can result in curvature, which limits the structure's travel and overall functionality. Polysilicon films must be processed so as to avoid buckling when the sacrificial layer is removed. The thickness of the sacrificial layer defines the gap between the mechanical structure and the underlying substrate. With polysilicon films as long as 500 μm located only 2 μm above the substrate, flatness is critical. In addition, the sidewalls of the polysilicon mechanical layer define both springs and electrical sensing elements. The spacing between these sidewalls therefore is important.

Optimized designs select feature sizes, sidewall angles, and manufacturing tolerances that compensate for variations in process parameters. Design and process trade-offs are illustrated by the inverse relationship between spring width and capacitive sensor gap. As the spring becomes thinner due to variations in etching, the gap increases. Therefore, spring compliance reducing spring stiffness also creates a larger travel allowance due to the increased gap. Structural designs that result in gaps rather than open fields are often used to improve the mechanical structure's uniformity. This helps to minimize feature variations and facilitates in-process measurements. The microenvironment around a feature contributes to the etch rates and the sidewall structures. By keeping gaps the same and minimizing open areas, variation across a mechanical structure is balanced and consistent. This results in better matching of springs and sensor elements [5]. The vibrational modes acting in these structures can also create a nonlinear response and must be carefully designed so that the structure will operate properly with the electronic circuitry used for both read-out and any type of self-test or actuation.

Integrated MEMS devices place a large premium on the silicon chip area, due to the cost of processing the electronics with the sensor element. Protective masking, alignment tolerances, and buffer regions between the sensor element and the circuit can add significantly to the silicon chip area. Structures that include high sidewalls and tall features require more tolerance. Optimization of processes that allow for interconnection to the circuit thus achieving a high density of interconnect is often more cost effective than minimizing the feature sizes of the lithographic processing in the circuit area. Close placement of sensor elements and electronics also makes it possible to minimize the use of bonding pads and compensation circuitry.

Another type of integrated micromachined structure different from the anchored polysilicon surface films occurs by adding the micromachined structure to the wafer after the IC is created by plating or film deposition. Texas Instruments' DLP technology with aluminum metal is the most well known example [3]. However, these types of structures are not used in large-scale automotive production. Another class of micromachined devices is made from silicon-on-insulator (SOI) components and could easily be described as surface micromachined, since they have mechanical structures on the surface. These devices use the buried oxide of the bonded wafer as the sacrificial layer [6]. A similar structure created by epitaxial deposition of silicon over oxide produces a polysilicon structure [7]. Figure

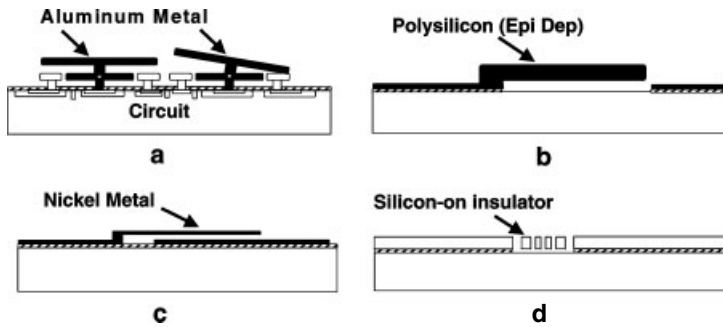


Fig. 5.2.4 Examples of types of surface micromachined structures: **a** aluminum multi-metal mirrors; **b** polysilicon epi-over-oxide; **c** nickel metal; **d** silicon-on-insulator

5.2.4 illustrates these various devices. Several of these structures are used in automotive production in nonintegrated form and doubtless in the future will become integrated as the economics of larger volumes justify their development.

5.2.3

Processing Integrated Surface Micromachined Polysilicon Structures

Surface-micromachined structures are most efficiently integrated with microelectronics, with minimum disruption to the integrated circuit process flow. Using polysilicon as a mechanical material requires different processing than CMOS gate fabrication, but the equipment can be similar. Due to thermal budget limitations when the polysilicon is annealed, the electronics fabrication process must be carefully selected. In the integrated surface-micromachining processes that I have been involved with for large-scale production of automotive accelerometers, the first phase of fabrication is a BiMOS process that uses large geometry devices ($3\text{-}\mu\text{m}$ gates) and dopants (As) that have low diffusion coefficients. This BiMOS process is capable of good performance after the 4 h of polysilicon annealing at 950°C .

Preceding an in-depth explanation, a general overview is given to help understand the process sequence better. We can describe the integrated process as an insertion of the mechanical surface-micromachined process into the middle of the BiMOS flow, before the contact and metallization stages. The processing sequence is then completed by contacting the mechanical structures, depositing and forming the thin film resistors, metallizing the aluminum, and performing passivation. The remaining steps include protecting the circuit structure with a photoresist, etching away the sacrificial layer under the mechanical polysilicon, and then removing the photoresist with a suitable dry etch. This ensures that, in the final steps, when the surface micromachined structures are released, they are not exposed to liquids and to surface tension, which can cause the polysilicon elements to adhere to one another or the substrate.

With a general overview of the process sequence complete, it is now appropriate to examine the details of the MEMS-specific processing considerations. Perform-

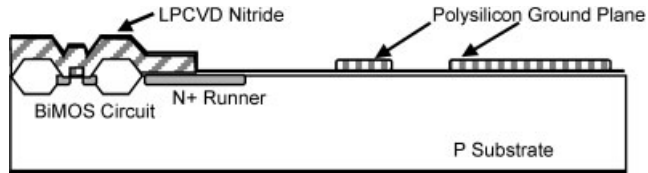


Fig. 5.2.5 Process cross section after deposition of a polysilicon ground plane

ing the BiMOS process before the mechanical polysilicon results in an area in the center of the chip that is free of any circuitry and can be interconnected to the circuit with N+ runners. The following is a detailed description of the MEMS process:

- *Protection layer.* An LPCVD nitride (Si_3N_4) layer is deposited to protect the BiMOS devices, and this $0.12\ \mu\text{m}$ thick layer becomes an etch stop and also protects the silicon substrate from pitting and etch attack.
- *Ground plane polysilicon.* To provide efficient interconnections, a polysilicon layer is deposited over the nitride, before the sacrificial layer is deposited. This provides a ground plane under the mechanical polysilicon or a lower electrode and therefore is referred to as ground-plane poly. Figure 5.2.5 shows that the mechanical polysilicon can also be used as an additional interconnect from either the N+ or the ground-plane poly. The $0.25\ \mu\text{m}$ thick ground-plane polysilicon layer has a sheet resistance of 120 ohms/square after arsenic implantation and annealing. In situ doped material is also used as a polysilicon layer.
- *Sacrificial oxide.* The sacrificial oxide layer is deposited by low-pressure chemical vapor deposition (LPCVD) or plasma-enhanced chemical vapor deposition (PECVD). The layer is made of either undoped glass or phosphorous-silicon glass (PSG)-doped resulting in a final thickness of $1.5\text{--}2.5\ \mu\text{m}$. Low defect density, good etch rate control, uniformity and stress control have been accomplished for undoped oxide with a Novellus Concept One tool.
- *Anchor.* The next step is to open contact holes through the sacrificial oxide to the ground-plane polysilicon layer and/or the n+ runners. These contacts become the electrical and mechanical interconnections or anchors for the mechanical polysilicon structure (Fig. 5.2.6). A wet-dry etch combination provides a

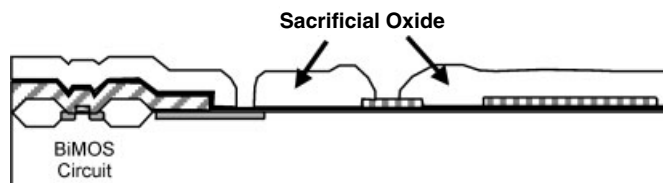


Fig. 5.2.6 Wafer cross section after deposition of sacrificial oxide and contact via's

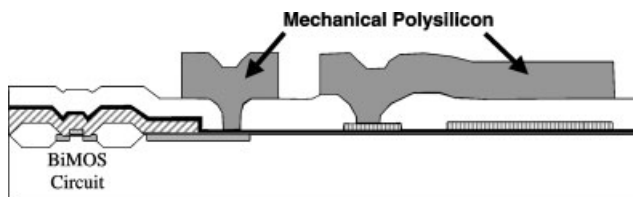


Fig. 5.2.7 Wafer cross section after mechanical polysilicon deposition and etching

shape that improves the strength and reliability of the layer by reducing stress in the mechanical polysilicon at the anchors.

- *Mechanical polysilicon element formation.* The deposited 2–4 μm thick mechanical or sensor polysilicon film is partially amorphous, and may have residual tensile stress. If the deposited film is completely amorphous or completely crystalline, it can buckle when the sacrificial film is removed. Annealing the deposited film typically results in recrystallization, which significantly reduces stress. Section 5.2.4 discusses polysilicon sensor processing in more detail, because this is considered the most important and critical element of integrated surface micromachining. Two ways of reducing the resistivity of the polysilicon material are phosphorus doping after the film is deposited and doping in situ during deposition. Both methods, along with residual oxygen in the film, significantly affect the magnitude of stress in the final film and its flatness after annealing. Although both methods of doping are commonly used, the manufacturing control possible with vertical tube deposition and in situ doping with phosphorus improves product consistency. After the annealing stage, the thick sacrificial oxide layer over the circuitry is removed by wet etching, and the polysilicon is dry etched to create the structure that will become the sensor elements. This is illustrated in Figure 5.2.7.
- *LPCVD or PECVD silicon dioxide protection.* An oxide layer is deposited to a thickness of 0.2 μm to protect the sensor and circuitry during the remaining manufacturing steps.
- *Oxide and nitride removal.* The 0.12 μm thick oxide and nitride layers originally deposited over the circuit area to protect it during MEMS processing are removed. This makes it possible to continue processing the standard microelectronic devices through contact etching, metal deposition, passivation of the circuit areas, and creation of bond pad vias or openings as shown in Figure 5.2.8.
- *Microstructure stabilization.* Before complete removal of the sacrificial oxide, small cavities are etched around and under the mechanical polysilicon and down to the ground-plane polysilicon below the sacrificial oxide layer. These cavities are then backfilled with photoresist, resulting in pillars that support the polysilicon. In a subsequent masking operation, strips of photoresist are placed across the micromachined elements. This results in a web of photoresist material that holds the polysilicon elements in place after complete removal of the sacrificial oxide layer. Subsequent etching of the sacrificial oxide with a buffered

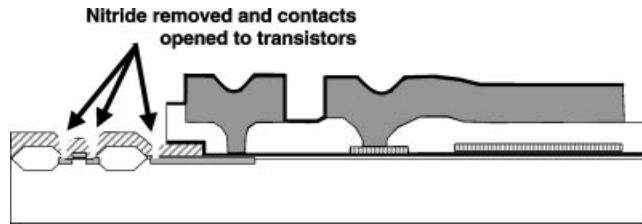


Fig. 5.2.8 Wafer cross section before resuming circuit processing

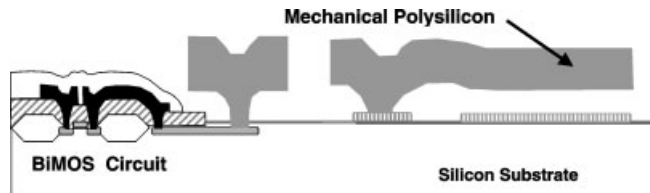


Fig. 5.2.9 Cross section of completed wafer

hydrofluoric acid (HF) solution therefore removes the oxide but does not pull the fragile polysilicon structure together or down to the substrate.

- *Oxygen plasma cleaning.* The integrated surface-micromachined wafer is exposed to an oxygen plasma to remove the photoresist and thus create free 3D structures as shown in Figure 5.2.9.
- *LPCVD formed anti-stiction coating.* The final step in processing the wafer by integrated surface micromachining is to passivate the mechanical structures with a material that reduces the surface energy and prevents sticking due to friction, commonly referred to as stiction. Controlled deposition of films less than 1 nm thick, which can be carefully monitored and measured, is important. A low-temperature LPCVD furnace is used to coat the wafers held in a quartz boat. The deposition process has been designed to be self-limiting so that the film does not build to a thickness of more than a few monolayers of the film. This is a more practical process than vapor deposition from liquid or organic residues left from the dry-etching process. Films like this must be carefully evaluated for long-term failure potential. The interaction of this type of film with ambient gases and their tendency to coalesce into agglomerated masses must be evaluated for long-term stability. Figure 5.2.10 depicts a polysilicon surface oxidized with native oxide (left), as well as an absorbed coat of anti-stiction material (~ 0.8 nm thick) on the surface of the mechanical polysilicon (right). Depositing a thin organic coating reduces the surface energy and the tendency of layers to stick.
- *Electrical evaluation.* Electrical process-control monitors that are included at critical steps in the manufacturing as a means of evaluating the process steps are an important component of surface micromachining. The diffusion profiles and semiconductor junctions, as well as the quality of the oxides and contacts, can all be

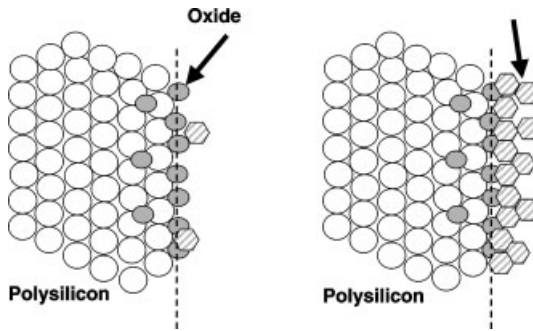


Fig. 5.2.10 Polysilicon with native surface oxide and anti-stiction coating

evaluated with test structures. It is also important to evaluate the MEMS's mechanical properties. Stress rings and various cantilever beams make it possible to evaluate the released film's mechanical stress and stress gradients. These structures, along with in-line inspection tools like the Zygo/Wyco tool and film stress measurements based on wafer bowing, make it possible to monitor the integrated surface micromachining process. One of the most effective methods is a careful electrical probing of the actual device to determine its mechanical characteristics such as damping, capacitance, and natural frequency, as well as its response to the on-chip self-test. The electrical and mechanical parameters can then be examined as a function of location on the wafer. The statistical process-control variations from batch to batch can be monitored to ensure a high yield.

5.2.4

Polysilicon Deposition and Processing

A large body of research exists on the deposition and processing of structural polysilicon [8–10]. The key parameter for films made of structural polysilicon is residual stress and therefore the flatness of the films. The deposition of polysilicon from silane (SiH_4) gas in a horizontal furnace has been extensively used. The addition of phosphorus as a dopant for the polysilicon has been used to control resistivity and stress. Including phosphorus dopant in the sacrificial oxide layer under the structural polysilicon layer and/or as ions implanted after deposition of the polysilicon is the most common methods used.

The temperature and rate of deposition have a significant effect on the polysilicon as does the residual oxygen content. The objective is to achieve a slightly tensile film by depositing a mixed phase of amorphous and smaller (111) oriented grains. The mixed amorphous and crystalline structure occurs in a transition region between mixed-phase and fully crystalline phase, where the deposition conditions encourage increased nucleation and small grains with (111) orientation. To reduce stress, the films are then annealed at temperatures above 900°C . This annealing results in a predominately small-crystalline film with a thin amorphous region at the top. The grain size distribution can be changed by regulating the oxygen concentration in the deposition tube during film growth. This oxygen can

come from leaks in the system or out-gassing. Oxygen increases the nucleation rate and interrupts the formation of large conical grains, thus resulting in finer-grained and lower-stress films. The minimum amount of oxygen needed is typically 2×10^{19} atoms cm^{-3} . Reaction with the silane will diminish this oxygen, thus the deposition rate and oxygen content are mutually related. The distribution of temperature, gas concentration and oxygen along the tube as a function of silane flow can also create significant variations in stress within the polysilicon layer. These parameters are difficult to monitor and maintain within acceptable limits during production. Because stress variations in the film can be detected only after deposition and annealing are performed, the risk to product yield is significant. To control these parameters of gas concentration, temperature and oxygen, a vertical LPCVD furnace with nitrogen (N_2) load lock and multi-hole gas injectors has been applied. With the vertical furnace, the background level of oxygen can be maintained at less than 5×10^{18} , and controlled amounts of oxygen and nitrogen are mixed and added using the injectors that supply phosphorus in the hot zone. This prevents interaction of oxygen with the silane, as well as particles of silicon that can be a byproduct of an uncontrolled oxygen atmosphere. Distribution and consistency are maintained in wafer batches of 100, resulting in cost-effective manufacture of integrated surface-micromachined automotive sensors.

To ensure consistent resistivity throughout the polysilicon structure, an alternative to the implantation and doping process has been applied to films greater than $2 \mu\text{m}$ thick. By adding phosphorus during the deposition step, control of the stress effects in a manner similar to small amounts of oxygen occur. A minimum level of phosphorus is needed for low stress and flat beams, but excessive

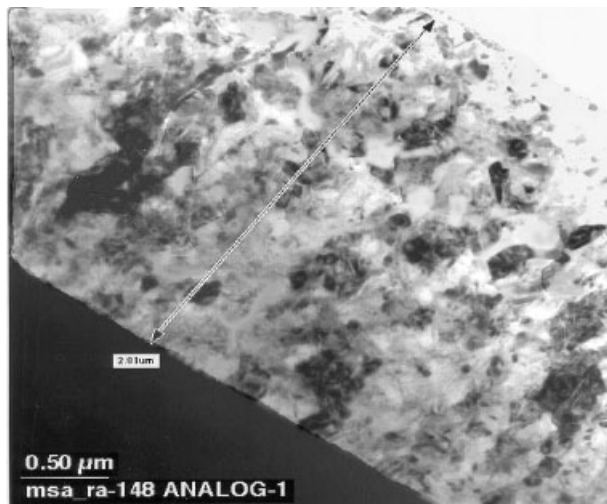


Fig. 5.2.11 Transmission electron micrograph of a cross section through structural polysilicon showing midsize uniform grain structure throughout

amounts result in compressive residual stress. The parameters of deposition rate, oxygen concentration and pressure, phosphorous concentration and pressure, time, temperature, silane flow, and annealing temperature are evaluated by designing experiments with the tools used for production. Optimization, films with low tensile stress and controlled stress gradients through the thickness of the film can be produced at deposition rates of 7 nm min^{-1} . In addition, appropriate processing parameters enable films with very smooth surfaces to be manufactured. The challenge of course is to maintain process control and consistent results. Shown in Figure 5.2.11 is a transmission electron micrograph of a cross section through a deposited and annealed structural polysilicon film. The grain structure is consistent throughout the $2.8 \mu\text{m}$ thickness therefore resulting in a flat film with a minimal stress gradient through the film.

5.2.5

Summary

Integration of surface-micromachined structures and electronics on the same substrate can be achieved at automotive production volumes and quality levels under carefully controlled conditions. Although many other materials can be used for the mechanical structure, such as nickel or aluminum, the properties of polycrystalline silicon have led to its dominating the automotive sensor market.

This section discussed the overall sequence for constructing BiMOS silicon features and for reserving an area on the surface of the micromachined structure. Next comes the deposition of a sacrificial layer of oxide and careful deposition of a film of amorphous, fine-grained silicon from silane gas, under conditions of controlled oxygen and phosphorus concentrations. After annealing the polycrystalline film, the final stage is etching and deposition of metal films and protective coatings and removal of the sacrificial oxide.

This process can be used for automotive devices, such as accelerometers and gyroscopes, that include both circuit and sensor on a single wafer.

5.2.6

References

- 1 K. CHAU, S. LEWIS, Y. ZHAO, R. HOWE, S. BART, R. MARCHESELLI, An integrated force-balanced capacitive accelerometer for low-g applications, in *Tech. Dig. 8th Int. Conf. Solid-State Sensor and Actuators (Transducers '95)*, Stockholm, Sweden, June 1995, 593–596.
- 2 J. C. COLE, A new sense element technology for accelerometer subsystems, in *Tech. Dig. 6th Int. Conf. Solid-State Sensors and Actuators (Transducers '91)*, San Francisco, CA, June 1991, 93–96.
- 3 L. J. HORNBECK, Current status of the digital micromirror device (DMD) for projection television applications, *IEEE International Electron Devices Meeting, Tech. Digest* 1993, 381–384.
- 4 T. A. CORE, W. K. TSANG, S. J. SHERMAN, Fabrication technology for an integrated surface-micromachined sensor, *Solid State Technology*, 1993, 39–44.
- 5 S. BART, H. SAMUELS, Design techniques for minimizing manufacturing variations in surface micromachined accelerome-

ters, *Microelectromechanical Systems*, ASME DSC-Vol. 59, #G01036, 1996, 427–433.

- 6 R. SOULOUFF, B. RIEDEL, Micromachined sensors for advanced microsystems, *Advanced Microsystems for Automotive Applications* (S. KRUEGER, W. GESSNER, eds.), Springer, Berlin, Germany 2001, 201–208.
- 7 M. OFFENBERG, F. LARMER, B. ELSNER, H. MUNZEL, W. RIETHMULLER, Novel process for a monolithic integrated accelerometer, *Tech. Dig. 8th Int. Conf. Solid-State Sensors and Actuators (Transducers '95)*, Stockholm, Sweden, June 1995, Vol. 1, 589–592.
- 8 K. NUNAN et al., Developing a manufacturable process for the deposition of thick polysilicon films for micro machined devices, *Proc. 11th Annual IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop (ASMC)*, September 2000.
- 9 A. VOUTSAS, Experimental and theoretical study of the crystallization of chemical-vapor-deposited mixed-phase silicon films, *J. Appl. Phys.* 76(2), 15 July, 1994, American Institute of Physics, 1994.
- 10 T.I. KAMINS, J.E. TURNER, *Solid State Technology*, Hewlett-Packard, Palo Alto, CA, 1990.

5.3

Surface Micromachining – Discrete

JAN PETER STADLER, MICHAEL OFFENBERG, and FRANZ LÄRMER

5.3.1

Introduction

Although microelectromechanical systems (MEMS) can be created in many different ways, silicon-based micromachining technologies have achieved widest distribution. Initially, silicon micromachining was used to form movable, or at least deformable, microstructures by etching into the bulk of a silicon wafer. Therefore this approach is called bulk micromachining and can deliver real 3D silicon structures formed from the wafer itself. More similar to IC processing technology, surface micromachining consists in building microstructures on top of a silicon wafer using thin films that have been deposited and patterned on the wafer surface. Despite the limitations of the rather simple techniques of deposition and patterning thin films, surface micromachining can produce complex microstructures, which can be used for a large variety of applications. This is true even for rather simple film structures involving only a few different layers. With increasing numbers of layers, the possible complexity, and hence the manufacturing difficulty can be increased almost without limit.

The general advantages and disadvantages are rather obvious. By limiting processes to the wafer surface, the degrees of freedom in which to create real 3D microstructures are much fewer than when bulk micromachining is used. In addition, the dimensions (at least in the z axis) are limited, and the achievable seismic masses are much smaller, but this may be compensated for by proper design and multiple-layer structures. On the other hand, arbitrary shapes can be formed, whereas classical bulk micromachining using anisotropic wet etching suffers from severe design limitations.

The convincing advantage of surface micromachining is its similarity to classical IC fabrication, which is limited to the wafer surface as well. Both fabrication techniques and structures are therefore similar and well understood. Only a few specific processes had to be developed to supplement the established standard semiconductor processing steps.

5.3.2

General Process Flow

The basic concept of surface micromachining is rather simple. In general, a functional material (usually silicon or polycrystalline silicon) is used to form the moving or deformable parts. To free this functional layer from the underlying substrate wafer, a ‘sacrificial’ layer (usually silicon dioxide) is formed prior to deposition of the functional layer. The functional layer is patterned while it is still fixed to and supported by the underlying rigid sacrificial layer. Subsequently, the intentionally movable structures are released by removing the supporting material, usually by etching away in wet or vapor phase etches. The supporting layer is called a sacrificial layer because it is ‘sacrificed’ to release the functional structures above it. This simple technique was first introduced to form metal cantilever beams [1] and later was transferred to silicon micromachining [2].

In the following example a typical, but rather general silicon-based process flow is described in the sequence of the processing steps (Fig. 5.3.1 a–d). The result is a simple micromachined cantilever beam, which – as will be described – can be anchored to the surface in two different ways (Fig. 5.3.1 e–f).

The sacrificial oxide layer is initially formed on top of the silicon substrate by thermal or plasma processing (Fig. 5.3.1 a). If insulation of the functional layer from the silicon substrate is required, the sacrificial layer may be preceded by deposition of an insulating film, for example, silicon nitride. The thickness of the sacrificial oxide layer determines the height of the released cantilever above the silicon substrate. It is typically a few micrometers thick, depending on the specific requirements and application.

Next, the sacrificial layer is patterned and holes are etched into the oxide using established lithography and etching processes. These holes will be filled and thus act as anchor points on the left end of the two cantilevers formed later (Fig. 5.3.1 e). In the next step, the functional polysilicon layer is deposited (Fig. 5.3.1 b). The thickness of this layer determines the mechanical properties of the movable beam. The thicker it is, the stiffer the beam will be in the z axis, which is desirable for structures intended to move only in the xy direction. But its thickness is limited by the capabilities of the deposition process used. The functional layer is next patterned and etched (Fig. 5.3.1 c). Depending on the thickness of the polysilicon layer, specific trench etch processes (as described later on) may be required, especially when this layer is rather thick. Finally, the sacrificial layer is removed (Fig. 5.3.1 d). This is typically done with wet or vapor phase etches to dissolve the silicon dioxide and leave parts of the functional structures free-standing and movable. When using wet etching, special care has to be taken to prevent stic-

tion of the released beams caused by capillary forces during drying. This is a lesser problem with vapor etching, but careful processing is still required. Because the sacrificial layer is removed by underetching the polysilicon from the sides, it is not removed everywhere. This fact can be used to anchor structural components without using dedicated anchor holes, as shown for the right cantilever beam (Fig. 5.3.1e).

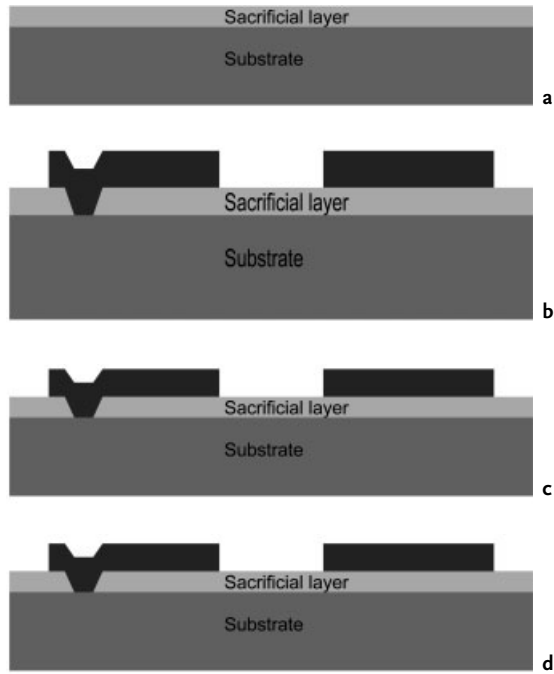


Fig. 5.3.1 a–d Cross sections of a silicon-based MEMS device at several stages during surface micromachining process flow

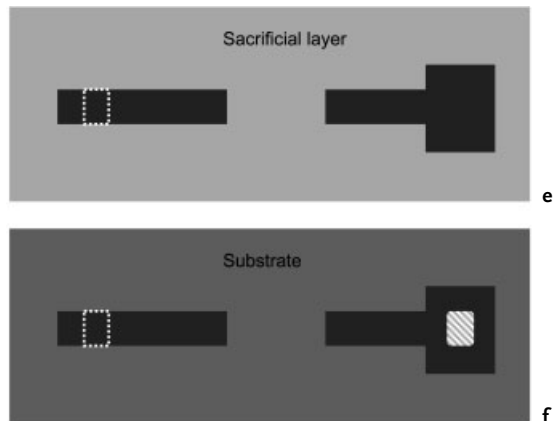
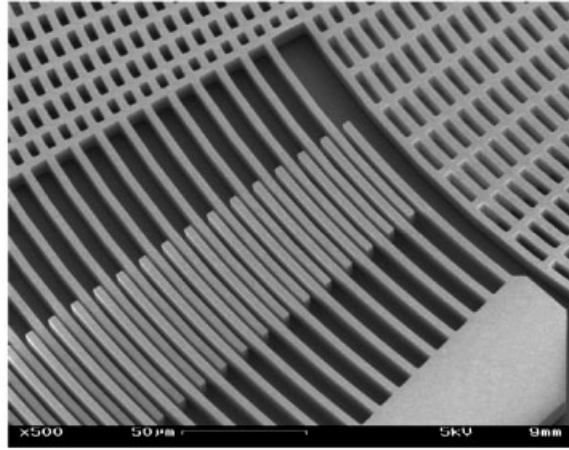


Fig. 5.3.1 e–f Top view of surface-micromachined cantilever beams anchored in two different ways; left: anchor hole; right: unre-moved sacrificial layer

Fig. 5.3.2 SEM (scanning electron microscope) detailed view of a surface-micromachined device



However, the limited underetching imposes serious restrictions on micromachining designers when they want to design structures that require larger areas of the functional layer to be released. To achieve complete release of large areas, they have to be perforated with holes whose sole purpose is enabling the surface to be freed from the underlying sacrificial layer. Although it is not their real purpose, these holes can give rise to beautiful patterns and the characteristic look of surface-micromachined devices (Fig. 5.3.1 e–f). However, recent progress in sacrificial layer etching methods may eliminate this limitation and open up even more degrees of freedom to MEMS designers [3].

5.3.3

Discrete vs. Integrated Approach

When designing sensors, an initial decision has to be made between the discrete approach, with separated sensing element and electronic signal evaluation circuit,

Tab. 5.3.1 Sensor concepts: discrete vs. integrated approach

	Arguments Discrete approach	Integrated approach
Pro	<ul style="list-style-type: none"> Flexibility Time to market/development time Individual yield optimization ASIC can follow S/C generation cycle 	<ul style="list-style-type: none"> Small size Fewer interfaces Multiple area usage On-chip signal evaluation
Con	<ul style="list-style-type: none"> Size 2-chip packaging Minimum signal requirements Larger parasitics (capacitances, resistors) 	<ul style="list-style-type: none"> Complexity Time to market Yield Inflexibility to follow S/C generation cycle

or the monolithic integration of both elements on the same die. There are numerous arguments pro and con for either of these options (Table 3.1).

Surface micromachining can be used for both discrete and integrated concepts, but has to be optimized according to the chosen approach. If discrete sensing elements are to be fabricated, special care has to be taken to maximize the available signals that have to be transmitted to the separate evaluation circuit via chip-to-chip bonds. When using capacitive sensing principles, this means maximizing the areas of the typical capacitive finger structures. In addition to the lateral dimensions, number of fingers, etc., the functional layer thickness is another important parameter. Using a rather thick functional layer is important for achieving working capacitances large enough (up to 1 pF) to enable off-chip signal evaluation.

5.3.4

Processes for Surface Micromachining

5.3.4.1 Epitaxial Polysilicon

When surface micromachining was first proposed it was touted as a technique that can be easily integrated into an existing planar process for integrated circuits. However, the monolithic integration of a microstructure into an electronic circuit is complex and requires numerous protective means and compromises, both on the sensor and on the evaluation circuit. On the other hand, MEMS devices benefit from a host of techniques established for microelectronics. Thus, an obvious choice of material for the free movable beam structure is LPCVD polysilicon (*low pressure chemical vapor deposition – polycrystalline silicon*) used for conductive layers in electronic circuits and deposited around 650°C. However, using a mechanical-grade polysilicon requires certain deviations from the standard processing scheme.

First, the typical thickness of a polysilicon layer in electronic circuits amounts to several hundred nanometers. MEMS applications generally require thicker layers, to reduce Brownian noise in the devices caused by impinging air molecules, for mechanical strength in the out-of-plane direction, and to increase capacitances and seismic mass. In particular, gyroscopes require large proof mass to generate a sufficiently strong sensor signal. However, the deposition rate of LPCVD polysilicon is limited by the pyrolysis of silane (SiH_4) to a couple of hundred nanometers per hour. The common thicknesses of MEMS devices, 2–3 μm , are a compromise between noise performance and feasible processing times.

Second, a mechanical-grade polysilicon requires a low-stress material that is homogenous in the cross section, across a wafer, from wafer to wafer, and from batch to batch. Large compressive stress results in buckling of double-clamped beams. A high stress gradient along the vertical cross section leads to curling of the released structures. A common approach to reduce mechanical stress in LPCVD polysilicon layers is to lower the deposition temperature to below 600°C, where the material is deposited in an amorphous state. However, this approach decreases the thermally activated deposition rate even more.

The deposition rate may be raised by using higher deposition temperatures. A standard process in the IC world is silicon epitaxy, which deposits tens of micro-

meters of monocrystalline silicon within minutes at temperatures of 1180°C. In contrast to the standard approach of single-crystalline epitaxy, the resulting layer is polycrystalline if the substrate is covered with a sacrificial layer and/or a polycrystalline seed layer. This material is dubbed 'epi-polysilicon', first because of the epitaxial reactor used, and second because the process on a polycrystalline seed layer is epitaxial with respect to each individual crystalline grain.

A thick polysilicon layer of about 10 μm can be deposited in about two minutes. In addition, the high temperature allows high surface diffusion of reactive species. In consequence, the as-deposited material is almost ideal in terms of stress and stress gradient. Even very sensitive interferometric tools cannot detect any curling of the released structures, and the internal mechanical stress is negligible. On the other hand, the high surface diffusivity favors the growth of very large grains at the expense of smaller grains. Thus, crystalline analysis shows columnar growth (Fig. 5.3.3), with smaller grains at the bottom and very large grains at the top.

For further processing, the surface is polished by CMP (chemomechanical polishing). Stepper lithography with very high resolution can be performed on the resulting mirror-like surface.

The mechanical properties of the film are modified by doping. For instance, phosphorus atoms are introduced via conventional PSG (Phospho-silicate glass) sources or ion implantation. In-depth studies [4] revealed that incorporation of

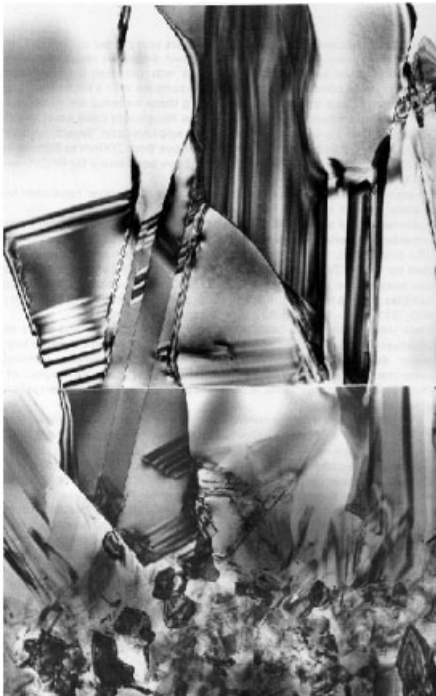
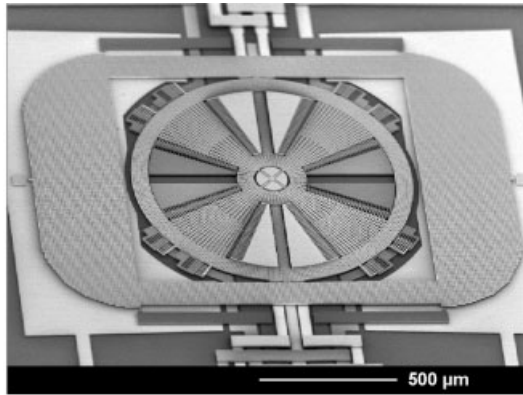


Fig. 5.3.3 TEM (transmission electron microscope) view of epi-polysilicon showing columnar growth of crystals

Fig. 5.3.4 Gyroscope created of epipolysilicon [5]



phosphorus in grain boundaries and the annealing atmosphere affect the stress gradient. Doping concentration and annealing atmosphere are varied to fine-tune the mechanical properties of the released MEMS structures. With an optimized process, extremely large structures, up to 2 mm in diameter, that are anchored only in the center have been fabricated [5] (Fig. 5.3.4).

Process monitors allow determination of stress and ensure proper process control of the mechanical properties. Clamped-clamped beams beyond a certain width start to buckle under compressive stress beyond a certain length, according to Euler buckling criteria. An array of so-called Euler buckling beams of different lengths allows determination of the mechanical stress according to the shortest beam that shows buckling (Fig. 5.3.5, left). Similarly, the stress gradient can be determined with a single-clamped beam by using interferometric methods to determine the curvature of the beam (Fig. 5.3.5, right). Both stress and stress gradient must have excellent uniformity across a wafer and repeatability from wafer to wafer and batch to batch.

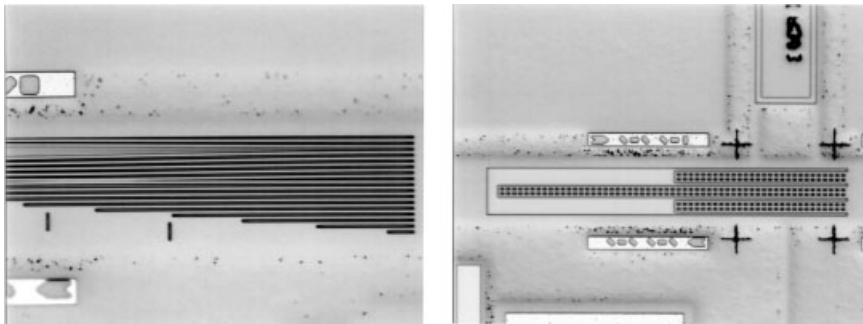


Fig. 5.3.5 Examples of process monitors for stress and stress gradients

5.3.4.2 Trench Etching

A key process in silicon micromachining is deep reactive ion etching (DRIE), or trench etching technology. DRIE has emerged as a standard MEMS technology in recent years, replacing other microstructuring technologies such as wet etch or LIGA (Lithographic by Synchrotron Radiation). In a plasma discharge, reactive species are created together with ions of more-or-less inactive gases supplied to the processing chamber. This generic technology allows the fabrication of nearly arbitrarily shaped geometries in silicon with high accuracy and vertical sidewalls. The directionality of the process is, irrespective of processing details, dependent on the ions' properties: due to their charge, ions can be accelerated towards the substrate electrode by an attractive bias potential and can impinge nearly perpendicularly on the wafer surface. Where they hit the silicon surface, they provide activation energy to the local wafer area and enhance the etching reaction between silicon atoms and chemically active species. Because impinging ions hit the bottom much more often than the sidewalls of a structure, anisotropic etching is possible. In contrast to the ions, uncharged chemically active species reach the wafer surface with no or only little directionality. Vertical ion impact onto the wafer surface is the dominant source of anisotropy, independent of crystal orientation.

In contrast to DRIE, anisotropic wet etching, from the early days of MEMS, is based on the crystal orientation-dependent etching behavior of alkaline solutions, such as KOH, EDP (Ethylendiamin-Pyrocatechol), or TMAH, with fast and slow etching behaviour along different crystal directions in a wafer. For a long time, wet etching was the only available microstructuring technology. For the fabrication of relatively simple geometries, such as piezoresistive pressure and acceleration sensors, this technology proved successful. However, manufacturable geometries were limited by crystal orientation, and extensive corner compensation was needed to protect convex corners from etch attack. In addition, hard coatings like SiO_2 or Si_3N_4 are required to protect masked silicon areas from the aggressive etchant. No satisfactory solution, except etch-boxes to protect the wafer's front during wet etching of its back, were ever found for passivation of metals like Al during the etch, which often meant that wet-etching was not compatible with IC production. These limitations and drawbacks of the early technology led people to look into plasma etching as a more flexible alternative.

Plasma etching has been established in semiconductor technology for a long time. Typical semiconductor etch processes are based on chlorine or bromine gases and include many potential additives to fine-tune desired properties and suppress undesired ones. These processes are, however, limited as to etch rate (typically below $1 \mu\text{m min}^{-1}$), achievable etch depth (a few micrometers), and mask selectivities (20–30:1 for a SiO_2 -hard mask). Photoresist stability is insufficient for deeper etches in chlorine and bromine-based process chemistries [6]. Fluorine-based plasma processes are superior in rate, depth, and mask selectivity and, in particular, enable photoresist masking, but the etch behavior is intrinsically isotropic, due to the nature of the spontaneously etching fluorine radicals. Since fluorine radicals do not need ion enhancement to etch silicon rapidly, anisotropy can be achieved only by adding sidewall passivation to the process. Sidewall

passivation is possible by oxidizing or nitrifying the sidewalls by adding oxygen and/or nitrogen, sometimes in combination with cryogenic temperatures (typically around 170 K) to freeze desorption of etch products from the surface [7], or by depositing sidewall films such as Teflon-type layers. The presence of both fluorine radicals (as etching species) and Teflon-forming monomers (as depositing, film-forming species) in the plasma at the same time leads to recombination and mutual extinction of both species, which makes a mixed process hard to control and tends to lower processing performance. These drawbacks appear more relaxed for plasma sources of very high excitation density [8].

The recombination problem was overcome by the Bosch approach [9], which is a variation of the Teflon sidewall passivation technique. Figure 5.3.6 illustrates the mechanism of the so-called Bosch DRIE process: passivation and etching gases are alternately introduced separately into the process chamber and exposed to a high-density plasma generated by microwave or inductive RF excitation, during passivation and etch cycles. During each passivation cycle, a thin Teflon-like film is deposited on the sidewalls of etched structures from gases like C_4F_8 or C_3F_6 (or other hydrofluorocarbons of low fluorine-to-carbon-ratio, 2:1 or less). During the subsequent etching cycle, part of this film is removed by off-vertical ion impact and is driven deeper into the trench. At the same time the trench bottom is etched by fluorine radicals released from SF_6 or other efficient fluorine radical sources. This local transport of the sidewall film yields local anisotropy at and in the vicinity of the sidewall, but does not negatively affect the main etching reaction away from the sidewall during the etching steps, which would otherwise be fully isotropic. This local anisotropy effect reduces sidewall roughness despite the discontinuous process and leads to a characteristically rounded trench bottom that is beneficial in many applications, for example pressure sensors.

A high-density decoupled plasma is most important to make this approach work, first to achieve sufficient sidewall passivation in a comparatively low-pressure regime, second to obtain a high fluorine concentration and thus high etch rates, and third to achieve a low plasma potential, close to ground potential. With a plasma potential close to ground, ion energy flow to the wafer surface is con-

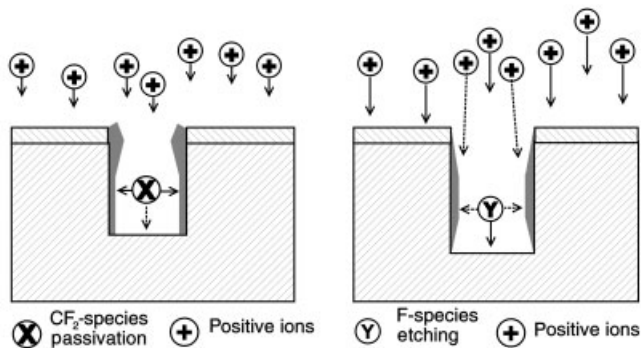


Fig. 5.3.6 Mechanism of the Bosch DRIE etching process

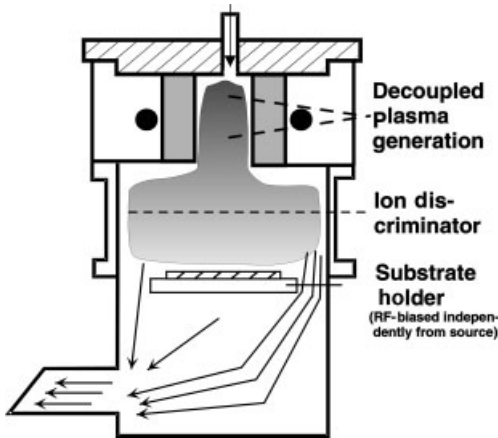


Fig. 5.3.7 High density plasma tool for DRIE

trolled exclusively through substrate electrode biasing by an independent RF source. Decoupling of plasma density and ion energy flow to the substrate wafer, making them independently adjustable, is a must for a successful use of this process.

Figure 5.3.7 shows a diagram of an advanced remote plasma tool configuration for illustration of the process requirements. Traditional plasma sources like RIE (reactive Ion Etching) and triode reactors fail in this context, as do ECR (Electron Cyclotron Resonance) sources in the very low pressure range (<0.1 Pa) tolerated by the cyclotron resonance mode. Today's microwave surfatron sources, inductively coupled plasma sources, and helicon sources expand the pressure window to 10 Pa and even higher and are most appropriate for this kind of process. Optimization of processing parameters and plasma excitation geometry has improved etch rates to well above $10 \mu\text{m min}^{-1}$, with selectivity to standard photoresists in excess of 150:1 [10].

High-power conditions during high-rate processing are responsible for several detrimental electrical disturbances in the plasma. Capacitive coupling of RF currents from the inductive coil into the plasma within the source region leads to variations in plasma potential. Given the comparatively low energy flow from the plasma to the wafer, controlled by moderate RF bias power levels supplied to the substrate electrode, the pronounced sensitivity of DRIE processing to any deformation in the spatial distribution of the plasma potential is obvious. For uniform, high-quality etch results over large wafer areas, plasma disturbances must be suppressed. As a first step, a balanced power supply to the inductive excitation configuration (coil) neutralizes capacitive coupling of currents into the plasma. In addition, careful spatial ion filtering and ion flow homogenization through mechanical or magnetic ion discriminators help to improve uniformity and profile quality over the wafer [11].

Etch-stop at dielectric interfaces is another common occurrence in surface micromachining, when the etch terminates on the sacrificial or isolation oxide beneath the active silicon layer. Notching effects appear in high-aspect ratio (narrow)

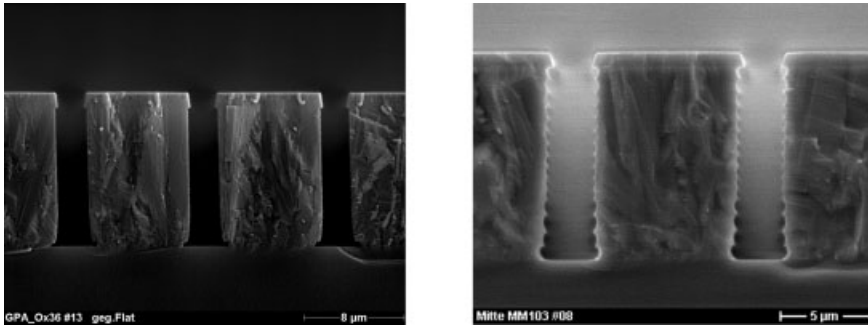


Fig. 5.3.8 Notching in narrow trenches (left); straight sidewalls down to the bottom obtained with pulsed bias (right)

trenches (Fig. 5.3.8, left), as a result of charging at the dielectric interface in a high-density plasma. Sophisticated pulsing schemes have been developed for substrate biasing, to suppress notching without harming other important processing features (Fig. 5.3.8, right). Pulsed biasing enables the dielectrics in high-aspect ratio trenches to be discharged during the off-periods of substrate potential, when carriers of negative charge can reach the trench bottom and neutralize the positively charged ions trapped there.

Fabrication of high-aspect ratio trenches while maintaining vertical sidewalls is another challenge, since profiles tend to develop positive slopes with increasing aspect ratio, mainly because of a limited species' transport to the etch floor. One approach to addressing this problem is to continuously adjust processing parameters with increasing etch depth during processing, for example by ramping down the passivation cycle time or the passivation gas flow. Trenches with an aspect ratio up to around 50:1 can be achieved in this manner, while keeping the sidewalls straight and vertical [12]. A different approach by Denso Corp. to achieving high-aspect ratio trenches is to include oxygen plasma steps in the process flow after every 20–30 'Bosch cycles'. This periodically removes accumulated sidewall polymer material and builds up a silicon oxide type of sidewall passivation, in addition to the Teflon films [13].

5.3.4.3 Sacrificial Oxide Etch

Another challenge in surface micromachining is to remove the sacrificial layer underneath the functional silicon layer without creating in-process stiction. Various proposals have been made [14–19], most based on etching with HF-containing liquids. Capillary forces during removal of the wafer from the etch bath and drying of the etchant must be avoided. This is generally done by replacing the liquid-gaseous transition by the introduction of a solid or supercritical phase or by mechanically stabilizing the structure from underneath.

Some examples are:

- Etching in buffered oxide etchant (BOE) and in-situ replacement of the etchant with water and then with a mixture of photoresist and acetone. The wafer is

then heated on a hot plate, and the photoresist stabilizing the structure from underneath is removed in an oxygen plasma. Critical factors include turbulence in handling the various liquid bathes and the processing conditions during plasma stripping [16].

- BOE followed by in-situ replacement with water and then with various intermediate liquids, ending with cyclohexane [17]. This chemical has a melting point around -5°C and its vapor has a high partial pressure. The wafer immersed in liquid is transferred to a vacuum chamber. Upon evacuation of the chamber, the liquid solidifies and is subsequently sublimated. Processing issues are wafer handling and throughput.
- Another approach is adopted from biological methods: After BOE, the liquid is replaced with liquid CO_2 and dried at the super-critical PT state [18]. Systems are available commercially but suffer from low throughput.
- One approach suitable for high-volume production is to stabilize the movable beam structure with pedestals of photoresist [19]. Rectangular openings are created in the sacrificial oxide layer and are isotropically underetched to create circular holes. These holes are filled with photoresist that is patterned to allow the etching liquid to enter the entire area under the movable beam structure. Subsequently the pedestals are removed in an oxygen plasma. Temperature control during this step is essential to avoid capillary forces due to low viscosity in the resist.
- Finally, stiction during processing can be avoided by using a vapor-phase instead of a liquid-phase etchant [20]. In contrast to the other approaches, this does not require additional masking steps or use processes that are critical with respect to manufacturing throughput. Because vapor-phase etching has proved to be well suited for high-volume manufacturing, we describe this approach in more detail.

Typically, the vapor phase of hydrofluoric acid is used instead of HF liquid. HF vapor is known to strip native oxides during cleaning of silicon wafers [21] and was suggested for sacrificial oxide etching [22]. However, even using the vapor phase can be critical in terms of stiction, because a byproduct of the chemical reaction is water:



This process-induced water can accumulate underneath the released microstructures and can lead to in-process stiction. To prevent this, the wafer is heated so that the excess water evaporates [23]. But selection of the temperature is critical, because water is not only a product of the reaction but also an initiator for hydrolyzing the SiO_2 network. The reaction does not start or is slowed down if water is depleted too much. Thus, there is a processing window limited on the low-temperature end by the onset of stiction and on the high temperature end by reaction kinetics (Fig. 5.3.9).

Fig. 5.3.9 Wafer-temperature window for vapor HF etching between stiction and too-slow etching [20] for constant temperature of the HF liquid (26 °C)

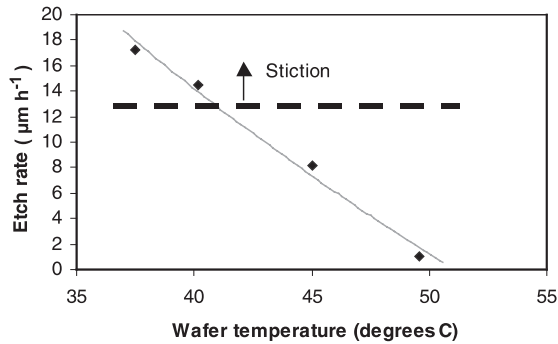
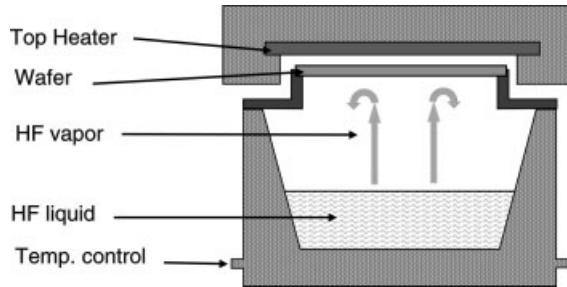


Fig. 5.3.10 Apparatus for vapor HF etching of sacrificial oxide



In practice, a difference of about 10 K between the temperature of the wafer and that of the concentrated HF liquid is recommended. To accelerate the reaction, the HF can be heated to around 50 °C. The wafer is placed a certain distance above the concentrated HF liquid and is exposed to the vapor phase formed (Fig. 5.3.10).

A semiautomated system including a wafer handler and a precleaning station is used for volume production at Bosch. Euler buckling beams are used for process control. The beams are longer than the critical length and thus buckle towards the sides after release. The beams are arranged in an array in the scribe line between the dies. Broad areas of increasing thickness from beam to beam are placed in the center of each beam. As soon as a broad area is completely under-etched the beam buckles sideways. This displacement can be easily observed in an optical microscope and serves as a reliable measure for processing control.

As explained, HF vapor has major benefits for the micromachining process: it is easy to use, is inexpensive, has good process control and high throughput, and requires no photolithographic steps. The latter feature allows the sacrificial etching of structures that are not compatible with photolithography. For instance wafers with aggressive topography or with holes etched all the way through [24]. However, distinct limitations of HF vapor etching need to be considered in the design of a MEMS process. HF vapor is fairly aggressive and easily penetrates common passivation layers. A widely used masking layer is LPCVD silicon nitride (Si_3N_4), which etches at a very low rate in concentrated liquid HF. In contrast,

LPCVD silicon nitride decomposes into a whitish powder within a few minutes when exposed to vapor HF. Organic photoresists are another masking material used for etching in liquid HF-containing solutions. Again, all common resists are penetrated in a short time in HF vapor and do not have any protective effect. As a consequence, when using HF vapor-etching MEMS processes must be designed so that the devices do not require any protective hard masks during production. Alternatively, less common hard passivation materials can be used, as was done for a monolithic integrated demonstrator [25]. In addition, the integrity of any exposed metal layers has to be considered. The process has to be designed carefully, including pre- and post-etch cleaning steps to avoid corrosion of aluminum wire bonding pads.

5.3.5

Applied Process Flow for Surface Micromachining

In practice, when actually manufacturing acceleration sensors by surface micromachining, the overall process flow outline must be filled in with the detailed specific processes – as discussed in the previous sections. This is described below, based on an actual production process used for acceleration sensors [26] and an angular rate sensor made solely by surface micromachining [27, 28].

First, a silicon dioxide layer is thermally grown to a thickness of about 2 μm . Although this oxide will also act as a sacrificial layer, its main purpose is to act as insulation for the next layer added. This next layer will be used to form wiring and counterelectrodes beneath the functional polysilicon layer. It is deposited in a standard LPCVD furnace and doped, for example, by diffusion doping with POCl_3 . The buried polysilicon layer is patterned by standard photolithography and etched by reactive ion etching (Fig. 5.3.11 a). Since it will be covered by subsequent layers it is called buried polysilicon.

Next, the sacrificial layer is deposited as a TEOS (Tetraethylorthosilane) oxide about 1.5 μm thick and is coated with photoresist, patterned and etched to form contact openings to the buried polysilicon layer (Fig. 5.3.11 b).

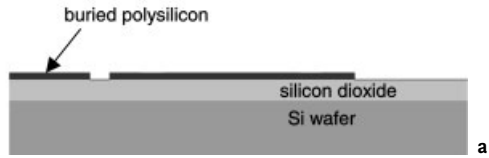
After stripping the photoresist from patterning, a seed layer of polysilicon is grown, again using a conventional LPCVD reactor at temperatures of approximately 600–650 °C.

A functional layer about 10 μm thick is grown in an epi reactor, as described earlier in Section 5.3.4.1. Since the starting material for the epi-polysilicon is polycrystalline, the epi layer is polycrystalline also. Therefore it is called ‘epi-polysilicon’ (Fig. 5.3.11 c).

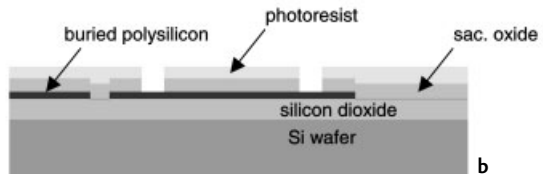
A CMP process is used to smooth the surface of the epi-polysilicon, which is rough both because of the polycrystalline epi growth and topographically, due to the patterning of the underlying layers. Removing as much as several micrometers smooths the surface enough for subsequent processing. To adjust the electrical and mechanical properties, the epi-polysilicon has to be properly doped and annealed. Next, aluminum metallization, as required for wiring and bond pads has to be deposited. It is sputtered conventionally with AlSiCu, patterned,

Fig. 5.3.11 a–f Detailed process flow for surface micromachining

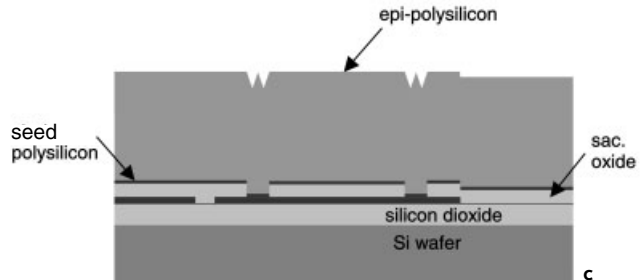
- Thermal oxidation
- LPCVD deposition of buried polysilicon
- Doping of buried polysilicon
- Lithography/etch



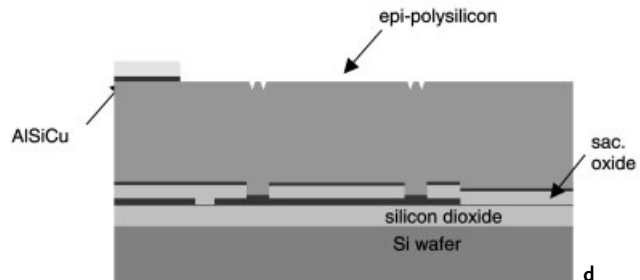
- Deposition of sacrificial oxide (TEOS)
- Lithography/etch



- Deposition of seed layer
- Epitaxial growth of epi-polysilicon



- CMP, doping and annealing of epi-polysilicon
- metal sputtering/lithography/etch



- Trench etch

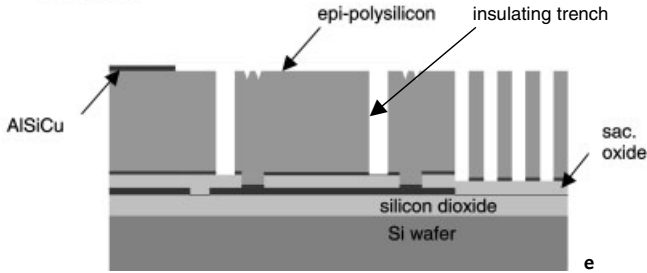
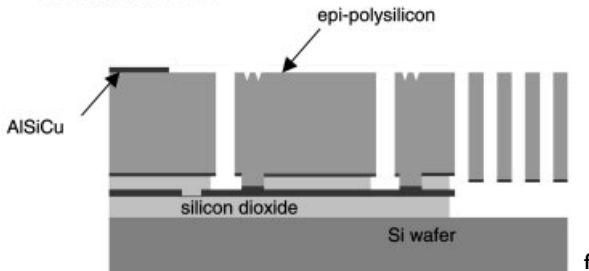


Fig. 5.3.11 e, f

- Sacrificial oxide etch



and etched (Fig. 5.3.11d). Due to the relaxed feature sizes of a few microns, this can still be done using wet/spray etch and does not require dry metal etch.

The epi-polysilicon functional layer is patterned and then etched by the trench etch process described in Section 5.3.4.2. The trench etching forms cantilever beams that act as comb fingers in the acceleration sensor, as well as the insulating trench required for electrical separation of the epi-polysilicon wiring and bond pad structures (Fig. 5.3.11e). Straight, unnotched sidewalls are strictly required, especially for all design elements that form springs or capacitive comb structures.

Finally, the functional layer is released by removing the sacrificial oxide layer by, for example, the vapor etch process described in Section 5.3.4.3. Depending on the actual layout, the sacrificial etch either removes only the sacrificial TEOS layer or removes both the initial oxide layer and the sacrificial oxide layer (Fig. 5.3.11 f). The sacrificial etch not only releases structures that are intended to be freed, but also undercuts anchored layout elements. This tendency has to be taken into account within the layout rules and during subsequent processing, since it can expose the edges of the buried polysilicon layer and remove the adjacent oxide layers that support it. After releasing the delicate movable microstructures need to be protected. Ideally, this protection is implemented on the wafer level immediately after the release process. A bulk-micromachined capping wafer is bonded to the sensor wafer with a low-melting glass seal. Such wafer-level packaging hermetically encapsulates the sensor structure and thereby protects it

against moisture and particles, etc. and from handling and subsequent processing steps.

5.3.6

Surface Micromachined Automotive Sensors

5.3.6.1 Acceleration Sensor

Acceleration sensors for automotive applications are manufactured in high volume using the described process flow [26]. The basic working design is illustrated in Figure 5.3.12.

A seismic mass is suspended on two springs, which allow it to move in the x direction. When acceleration is applied, the seismic mass lags behind, causing the springs to bend. This changes the gap between the moving fingers of the inner electrode and the fixed (immovable) fingers of the outer electrodes. The result is a change in the capacitances between the inner electrode and electrodes 1 and 2, which can be used to determine the applied acceleration a :

$$a \sim \frac{C_1 - C_2}{C_1 + C_2}$$

As can be seen, due to the thick functional epi-polysilicon layer, the comb fingers have a rather high aspect ratio. This enables the springs to deform in the xy -direction while maintaining stiffness to minimize unwanted deformation in the z direction. In addition, the thick layer increases the working capacitances by increasing the capacitance areas of the adjacent fingers.

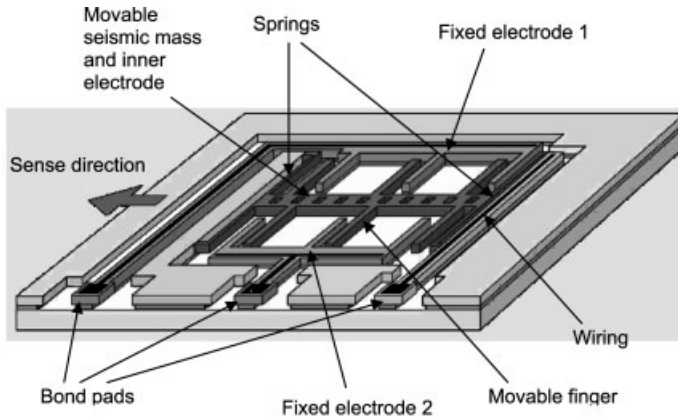


Fig. 5.3.12 Working design of a discrete accelerometer made by silicon surface micromachining: the seismic mass forms the movable inner electrode and is suspended on two springs; the movable capacitive fingers are interdigitated between the outer electrodes 1 and 2, which have fixed (immovable) capacitive fingers

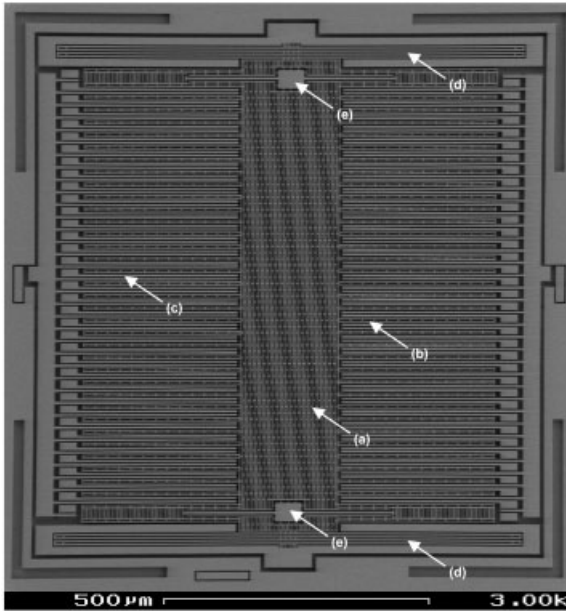


Fig. 5.3.13 Micromachined acceleration sensor with seismic mass (a), movable comb fingers (b), two fixed-finger outer electrodes (c), two supporting springs (d), and two over-range stops (e)

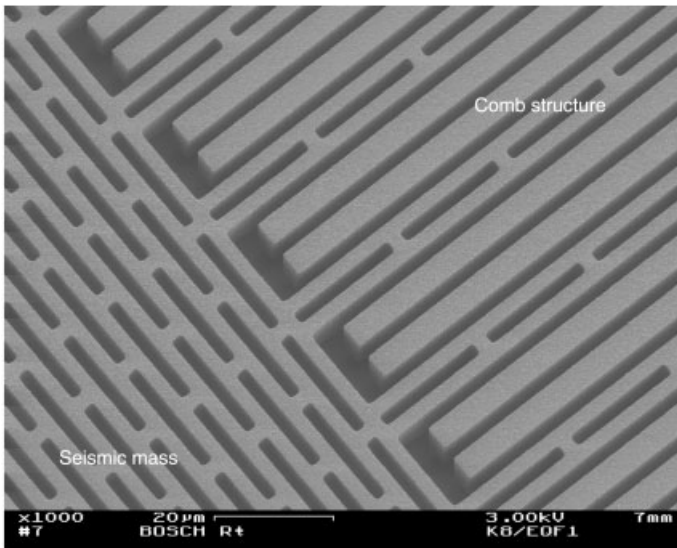


Fig. 5.3.14 Detailed view of seismic mass (left) and comb finger structure (right)

Figure 5.3.13 shows a SEM view of an actual acceleration sensor fabricated at Bosch, which has interdigitated electrodes for in-plane detection of xy accelerations. The movable comb fingers, the central seismic mass, both fixed finger electrodes, and the supporting springs can be identified easily. Mechanical overload protection is achieved by two over-range stops, which are located close to the supporting springs. In case of excessive acceleration, the over-range stops limit the movement of the seismic mass and thereby prevent the capacitive fingers from touching in an uncontrolled manner.

The comb structures is shown at higher magnification in Figure 5.3.14. As can be seen, not only is the seismic mass (left) perforated with holes necessary for its release, but the movable comb fingers also consist of double-beam structures to enhance their mechanical stability. Both positive and negative fixed electrodes, interdigitated with the moving comb fingers, are made as single beams.

5.3.6.2 Angular Rate Sensor

An angular rate sensor can be manufactured with an almost identical process flow as described in Section 5.3.5 and as used for acceleration sensors [27, 28]. The working principle is described in Chapter 7.2. An overall SEM view is shown in Figure 5.3.15. Like the acceleration sensor just described, this angular rate sensor is made solely by surface micromachining. But in contrast to acceleration sensors, the angular rate sensor uses both in-plane and out-of-plane movements, for drive and detection, respectively. The whole circular microstructure, with a rotor diameter of about 1.6 mm, is supported by the central x-shaped springs. Since out-of-plane detection requires underlying counterelectrodes made from the bur-

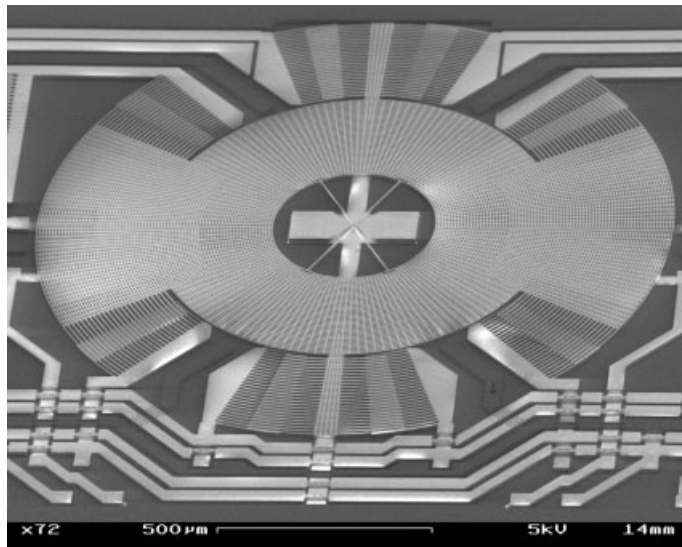


Fig. 5.3.15 Surface micromachined angular rate sensor

ied polysilicon, there is a released gap of only 1.5 μm across the whole diameter. This necessitates even higher requirements on the in-plane planarity of the epitaxial polysilicon functional layer.

5.3.7

References

- 1 H. C. NATHANSON, W. E. NEWELL, R. A. WICKSTROM, J. R. DAVIS, The resonant gate transistor, *IEEE Trans. Electron Devices ED-14*, 117–133, 1967.
- 2 R. T. HOWE, R. S. MÜLLER, Polycrystalline silicon micromechanical beams, *J. Electrochem. Soc.* 130, 1420–1423, 1983.
- 3 L. METZGER, F. FISCHER, W. MOKWA, Polysilicon sacrificial layer etching using XEF2 for silicon acceleration sensors with high aspect ratio, *16th European Conference on Solid-State Transducers*, Prague, Czech Republic, September 15–18, 2002.
- 4 M. FÜRTSCH, M. OFFENBERG, A. VILA, A. CORNET, J. R. MORANTE, Texture and stress profile in thick polysilicon films suitable for fabrication of microstructures, *Thin Solid Films* 296, 177, 1997.
- 5 W. GEIGER, B. FOLKMER, H. SANDMAIER, W. LANG, Hahn Schickart IMIT 12.10.98 New designs, readout concept and simulation approach of micromachined rate gyroscope ESSDERC97.
- 6 K. T. SUNG, S. W. PANG, *J. Vac. Sci. Technol. A11*, 1206–1210, 1993.
- 7 J. W. BARTHA, J. GRESCHNER, M. PUECH, P. MAUQUIN, *Journal of Microelectronic Engineering* 27, 453–456, 1995.
- 8 F. LÄRMER, A. SCHILP, US Pat. 5 498 312.
- 9 F. LÄRMER, A. SCHILP, US Pat. 5 501 893.
- 10 A. SCHILP, M. HAUSNER, M. PUECH, N. LAUNAY, H. KARAGOEZOGLU, F. LÄRMER, *Proc. Microsystem Technologies 2001*, Düsseldorf, Germany, 425–430, March 27–29, 2001.
- 11 F. LÄRMER, A. SCHILP, K. FUNK, M. OFFENBERG, Bosch deep silicon etching: improving uniformity and etch rate for advanced MEMS applications, *Proc. MEMS '99*, Orlando, FL, USA, 211–216, Jan. 17–21, 1999.
- 12 F. LÄRMER, A. SCHILP, US Pat. 6 284 148.
- 13 J. OHARA, K. KANO, Y. TAKEUCHI, N. OHYA, Y. OTSUKA, S. AHITA, A new DRIE process by dual sidewall protection layer, *Proc. MEMS '00*, Miyazaki, Japan, 277–282, Jan. 23–27, 2000.
- 14 C. H. MASTRANGELO, G. S. SALOKA, *Proceedings IEEE-MEMS 1993*, 77, 1993.
- 15 D. KABAYASHI, C. KIM, H. FUJITA, Late News, *International Conference on Solid-State Sensors and Actuators: Transducers 93*, 14, 1993.
- 16 A. KOVACS, A. STOFFEL, *J. Micromech. Microeng.* 2, 190, 1992.
- 17 H. GUCKEL, J. J. SNIEGOWSKI, T. R. CHRISTENSON, F. RAISSI, *Sensors and Actuators A21*, 346, 1990.
- 18 G. T. MULHERN, D. S. SOANE, R. T. HOWE, *Proc. 7th Int. Conf. on Solid-State Sensors and Actuators (Transducers '93)*, IEEE Electron Devices Society, Yokohama, Japan 296, 1993.
- 19 T. CORE, R. HOWE, Analog devices WO 93/21536.
- 20 M. OFFENBERG, B. ELSNER, F. LÄRMER, Vapor HF etching for sacrificial oxide removal in surface micro machining, *Electrochem. Soc. Fall Meeting*, Miami Beach, FL USA, Abstr. No. 671, p. 1056, 1994.
- 21 G. W. RUBLOFF, M. OFFENBERG, M. LIEHR, Integrated processing of MOS gate dielectric structures, *IEEE Transaction Semiconductor Manufacturing* 7(1), 96, 1994.
- 22 T. LOBER, J. HUANG, M. A. SCHMIDT, S. D. SENTURIA, Characterization of the mechanisms producing bending moments in polysilicon micro-cantilever beams by interferometric deflection measurements.
- 23 M. OFFENBERG, US Pat. 5 683 591A
- 24 M. LUTZ, W. GOLDBERGER, J. GERSTENMEIER, J. MAREK, B. MAIHÖFER, S. MAHLER, H. MÜNDEL, U. BISCHOF, A precision yaw rate sensor in silicon surface micromachining, *Proceedings of Transducers '97*
- 25 M. OFFENBERG, F. LÄRMER, B. ELSNER, H. MÜNDEL, W. RIETHMÜLLER, Novel process for a monolithic integrated accele-

- rometer, *Transducers 1995 Eurosensors IX*, 148-C4, 589–592, 1995.
- 26 M. OFFENBERG, H. MÜNZEL, D. SCHUBERT, O. SCHATZ, F. LÄRMER, E. MÜLLER, B. MAIHÖFER, J. MAREK, *Acceleration sensor in surface micromachining for airbag application with high signal/noise ratio*, Robert Bosch GmbH, SAE 960758, 1996.
- 27 K. FUNK, A. SCHILP, M. OFFENBERG, B. ELSNER, F. LÄRMER, Surface-micromachining of resonant silicon structures, *8th International Conference on Solid-State Sensors and Actuators Transducers '95 Eurosensors IX*, Stockholm, Sweden, Late News, 50–52, 25–29 June, 1995.
- 28 R. SCHELLIN, A. THOMAE, M. LANG, W. BAUER, J. MOHAUPT, G. BISCHOPINK, L. TANTEN, H. BAUMANN, H. EMMERICH, S. PINTER, J. MAREK, K. FUNK, G. LORENZ, R. NEUL, A low cost angular rate sensor for automotive applications in surface micromachining technology, *Advanced Microsystems for Automotive Applications 99*, Berlin, Germany, 239–250, 18–19 March 1999.

5.4

Thin Films on Steel

RALF HENN, TOSHIO HONMA, and STEFAN FINKBEINER

5.4.1

Introduction

This section describes thin-film-on-steel technology for automotive applications in detail. Steel substrates are typically used to transform mechanical parameters such as pressure, weight, force, or momentum into displacements and, in consequence, strain. Strain gauges convert the strain into electrical parameters, for example, voltage. In the following, we focus on strain gauges fabricated by thin-film technology on steel substrates. The most important type of sensors using this technology are high pressure sensors (Chapter 7.4) covering a range of about 10–200 MPa [1–3].

Stainless steel substrates are used because of their suitable mechanical properties and stability against corrosion. Thin-film technology on steel substrates achieves high accuracy and long-term stability. A schematic cross section of a thin-film system for the creation of strain gauges on stainless steel is shown in Figure 5.4.1. The thin-film system contains four functional layers: the isolation layer, which also forms the interface to the steel substrate; the sensing layer; the contact layer; and finally, the passivation layer.

An overview of the techniques for fabricating the functional layers is given, with the focus on special aspects resulting from the use of steel substrates. Poly-

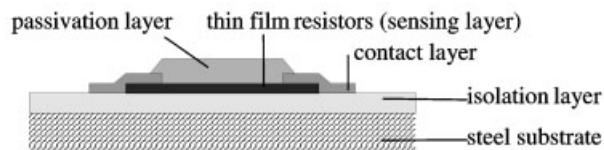


Fig. 5.4.1 Schematic cross section of a strain gauge using thin film technology on a steel substrate

silicon and NiCr are commonly used for the sensing layer in automotive applications and are also discussed in detail. The manufacturing approach contributes significantly to the productivity of a production line and strongly affects the details of the thin film process. Therefore it is discussed first.

5.4.2

Manufacturing Approaches

The manufacturing approach strongly determines the productivity of a high-volume production line. Batch processing minimizes production costs. The individual processing steps are similar to those used in manufacturing integrated circuits (ICs). Hence, to benefit from established semiconductor technology, it is advantageous to create an equivalent to a silicon wafer. Basically two strategies can be followed:

1. *Stainless steel wafer.* A diagram of a stainless steel wafer is shown in Figure 5.4.2 (left). Depending on the geometry of the sensor element and the wafer, many more than 100 elements can be formed on one wafer. For many applications, the back of the wafer undergoes a machining process before thin film processing, for example, to create the diaphragms of high-pressure sensors. After thin film processing of the wafer, it is cut to separate the sensor elements. This step is critical, since damage to the thin film system has to be prevented. Sensor separation can be achieved by, for example, laser cutting techniques. Erosive techniques may also be used, even though the cutting speed is significantly lower.
2. *Batch processing of individual elements in a jig* (Fig. 5.4.2, right). This approach avoids the critical separation step. Typically, a tray with fittings for the individual elements is used. During each processing step, additional components may be used to, for example, clamp the elements. The elements should remain in the same tray throughout the whole processing. Such an approach allows single elements to be treated mechanically with high precision before they are placed

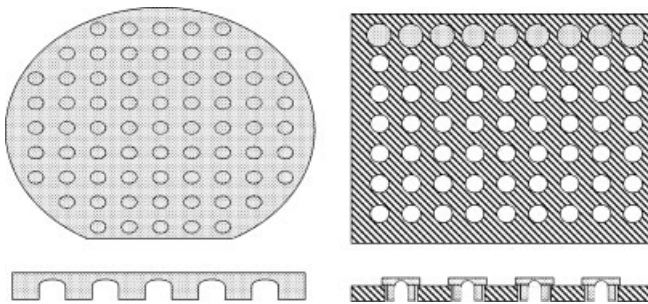


Fig. 5.4.2 Schematic cross sections and top views of a stainless steel wafer (left) and a jig for batch processing separated elements (right – elements are indicated only in the top row)

in the jig. However, the jig has to meet many different requirements, including vacuum compatibility, resistance against the various chemicals used during the thin film process, high temperature stability, and exact positioning of the elements for the different patterning steps. Furthermore, the jig should be compatible with automation.

Both approaches represent a compromise among different requirements and also affect the degrees of freedom allowed for the design of the sensor element. The decision between the approaches depends on the requirements of the application and the production line.

5.4.3

Substrate

The substrate acts as a transducer of the mechanical parameter to be detected into strain. Hence, it contributes significantly to the characteristics of the sensor element. Stainless steel is used as a substrate material for applications involving high mechanical stress, for example, high-pressure sensors, which have to withstand extreme pressure and temperature cycles. Stainless steel is stable in a wide temperature range. It offers high resistance against corrosive media. Its tolerance to overload is favorable because of its low brittleness. Furthermore, its welding performance is good. This is important for mounting the sensor elements in systems.

The material commonly used is stainless steel 17-4 PH (~17% chromium and ~4% nickel) [2, 3, 6], which offers a high strength in combination with high corrosion stability. To obtain optimum performance, the material is solution annealed

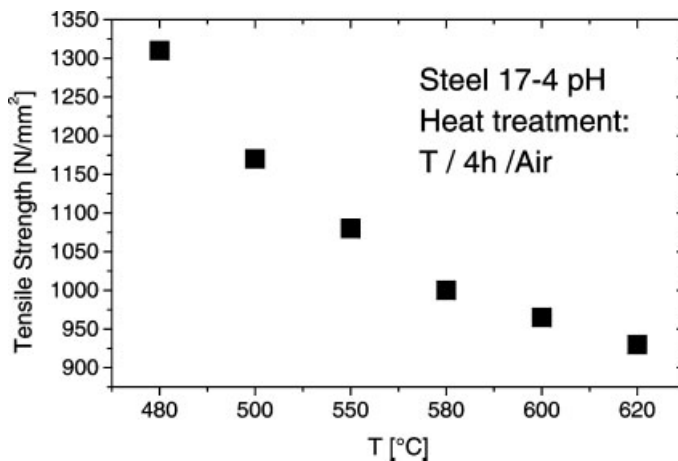


Fig. 5.4.3 Tensile strength of solution-annealed 17-4 PH stainless steel as a function of the temperature of the precipitation-hardening step (data taken from [4])

and then precipitation hardened. The solution-annealed material contains martensite, austenite, and ferrite. The precipitation-hardened material in addition contains intermetallic phases which enhance its hardness and strength [4]. The microstructure depends on the temperature during the precipitation-hardening process. Therefore, mechanical properties such as tensile strength are determined by the temperature of the precipitation-hardening step (Fig. 5.4.3) [4]. The highest values for hardness and tensile strength are obtained at temperatures of about 480 °C. After treatment at higher temperatures, the hardness and strength are lower. To obtain the desired mechanical properties, the temperature applied to the material after solution annealing has to be kept below about 600 °C. In consequence, this is also valid during thin-film processing.

To design a suitable geometry of the substrate, it is common to determine strain profiles by finite element method FEM [5]. It is important to have good control over critical geometrical parameters, because the substrate strongly affects the sensor's characteristics. For high-pressure sensors, for example, the diaphragm (with a typical thickness of a few 100 μm) has to be machined very precisely. Because machining is easier in the solution-annealed state, machining is usually done before the precipitation-hardening step. In doing so, however, one has to take into account a small amount of shrinkage during the precipitation hardening.

Most thin-film processes require a steel surface of high quality. Oil- or water-based polishing processes have been developed to obtain high quality surface finishes. Roughness values R_a below 10 nm can be achieved.

Because thin film processing requires clean-room conditions, cleaning the substrates after manufacturing is very important. Clean surfaces are crucial for achieving a low defect density in the thin films and for good film adhesion. Typically, the cleaning procedure is split into a cleaning step immediately after manufacture of the substrates and a second cleaning step when the process flow enters the clean room. Immediately after manufacturing, the organic and metallic residues from machining and polishing have to be removed. Usually, several cleaning stations with different cleaning solutions and additional mechanical support such as brushing or ultrasonic cleaning are needed to obtain substrate surfaces suited for a thin film process [2, 3]. The details depend strongly on the machining and polishing processes.

5.4.4

Isolation Layer

An isolation layer is required for electrical insulation between the strain gauges and the steel surface. The isolation layer also acts as a dielectric the capacitor formed between the steel substrate and the strain gauges. This strongly determines their electromagnetic compatibility. Another important function is to transmit elastic deformations from the steel substrate to the strain gauge. The material has to exhibit high thermal stability and should be stable against cracking. The main requirements for the insulating layer can be summarized as follows:

- high resistivity
- high breakdown voltage
- good dielectric properties
- good adhesion to steel
- good transmission of elastic deformation
- high thermal stability

Thin films of SiO_x [1–3] or Al_2O_3 [6, 7] are suitable and well established as insulating layers. SiO_x is a well known dielectric from Si technology. It is preferred in most applications, because its insulating behavior is superior to that of Al_2O_3 [8]. However, for the deposition on steel substrates, several aspects have to be considered in addition to Si-technology requirements.

One crucial aspect is the large difference in the thermal expansion coefficients of stainless steel and SiO_x (more than 10-fold, Table 5.4.1). As a result, temperature changes create significant stress at the interface between the steel and the SiO_x . During operation, mechanical stress is introduced in addition to the thermal stress. To prevent the formation of cracks, the internal stress of the SiO_x layer has to be chosen so that the SiO_x remains under compressive stress under all operating conditions. Typically, the internal stress needs to be compressive and on the order of 100 MPa. Stability against cracks is favored by a high density of the deposited layer [9].

Steel substrates, in comparison to silicon wafers, have surfaces with a higher degree of imperfection. As a result, the defect density and breakdown voltage are compromised. The thickness of the SiO_x is usually chosen in the range 5–10 μm , which is required to guarantee a breakthrough voltage greater than 300 V DC.

The steel substrate limits the maximum processing temperature to about 600 °C (see Section 5.4.3). Since high-quality SiO_x is typically grown at higher temperatures, special efforts have to be made to develop a suitable low temperature deposition process. RF sputtering and plasma-enhanced chemical vapor deposition (PECVD) [10] are most frequently used. PECVD techniques using large-area reactors or furnaces allowing batch processing are preferred for high volume production. Besides silane (SiH_4), the organic compound tetraethylorthosilicate (TEOS) is used as a Si source. For the silane process [10, 11], as well as for the TEOS process [12, 13], satisfactory results can be obtained at temperatures below 400 °C. SiO_x layers deposited at lower temperatures tend to have higher silanol (SiOH) concentrations. The properties of the deposited SiO_x layers, such as refrac-

Tab. 5.4.1 Thermal expansion coefficients and densities of stainless steel, SiO_x , and Si_xN_y

<i>Material</i>	<i>Thermal expansion coefficient</i>	<i>Density</i>	<i>Reference</i>
Stainless steel (17-4 PH)	$10.8 \cdot 10^{-6} \text{ K}^{-1}$	7.78 g cm^{-3}	4
SiO_x	$0.5 \cdot 10^{-6} \text{ K}^{-1}$	2.2 g cm^{-3}	8
Si_xN_y	$2.44 \cdot 10^{-6} \text{ K}^{-1}$	3.1 g cm^{-3}	9

tive index, density, strain, and stoichiometry, strongly depend on deposition parameters such as gas flow, pressure, temperature, and RF power. The higher surface mobility of TEOS-deposited layers is the reason that the surface is typically smoother for TEOS SiO_x than for silane SiO_x . However, TEOS processes incorporate more impurities, like C and H.

5.4.5

Sensing Layer

The sensing layer transforms mechanical strain into an electrical parameter. Most often, the dependence of electrical resistance on strain is exploited in strain gauges. The requirements for a suitable strain gauge material can be summarized as follows:

- strong dependence of the resistance on the applied strain
- low temperature dependence of gauge factor
- high resistivity
- low temperature dependence of the resistivity
- high chemical and structural stability

A compromise usually has to be found among these requirements. The priorities strongly depend on the target application of the strain gauge.

Polysilicon and NiCr are currently used for high volume production of high pressure sensors. Because the origin of the change in resistance as a function of the applied strain is different for the two materials, we will analyze the physical contributions to the gauge factor and then discuss these materials in more detail. The Wheatstone bridge, which is commonly used to detect small changes in resistance, is also discussed.

5.4.5.1 Gauge Factor and Piezoresistivity

An important parameter for characterizing strain gauges is the gauge factor K_{GF} , which describes the strain sensitivity and is (within linear approximation) given by the relation between the fractional change of the electrical resistance ($\Delta R/R$) and the mechanical strain (ε_m) as

$$\frac{\Delta R}{R} = K_{GF} \cdot \varepsilon_m \quad (1)$$

To illustrate the physical contributions to the gauge factor, a relation for the gauge factor is derived. The starting point is the resistance R of a conducting wire having resistivity ρ , length l , and cross section A :

$$R = \rho \frac{l}{A}$$

When strain is applied, the relative change in resistance is given by (first order approximation)

$$\frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} + \frac{\Delta l}{l} - \frac{\Delta A}{A}.$$

Having in mind the relations $\frac{\Delta l}{l} = \varepsilon_m$, where ε_m describes the mechanical strain and $\frac{\Delta A}{A} = -2\nu \frac{\Delta l}{l}$, where ν is the Poisson ratio, the relative variation of the resistance can be written as

$$\frac{\Delta R}{R} = \left(1 + 2\nu + \frac{\Delta \rho}{\rho \cdot \varepsilon_m} \right) \cdot \varepsilon_m \quad (2)$$

By equating Eq. (1) and Eq. (2), the gauge factor is derived as

$$K_{GF} = 1 + 2\nu + \frac{\Delta \rho}{\rho \cdot \varepsilon_m} \quad (3)$$

The first two terms in Eq. (3) represent the change in resistance caused by dimensional changes. The last term represents the change in resistivity. The material property by which the bulk resistivity is influenced by mechanical stress applied to the material is called piezoresistivity [14]. Therefore, the last term essentially represents the piezoresistive effect. The piezoresistive effect is larger in semiconductors than in metals. Therefore, geometrical contributions dominate in metal strain gauges, and piezoresistivity dominates in semiconductor gauges, leading to significantly higher gauge factors for the latter (Table 5.4.2) [15].

For most semiconductors, the geometrical contribution to the gauge factor is negligible [15]. From a quantum physical point of view, the piezoresistivity is tightly coupled to the band structure of these materials [16]. Distortions in the crystal lattice change the effective mass (and also the mobility) of the charge carriers and modify the occupation of the bands (changing the effective number of charge carriers) [17]. Often, distortion reduces the crystal symmetry. Thus, existing degeneracies of conduction bands can be lifted [17]. This mechanism enhances the piezoresistive effect, because it further contributes to the variation of the mobility and the effective number of charge carriers.

Tab. 5.4.2 Properties of various strain gauge materials

<i>Material</i>	K_{GF}	ρ ($\mu\Omega$ cm)	TCR^+ (ppm K^{-1})
Pt	6.0	10	1300
Pt ₈₀ Ir ₂₀	5.7	25	3800
Cu ₆₀ Ni ₄₀	2.1	49	-30
Ni ₈₀ Cr ₂₀	2.2	110	85
Ni ₅₀ Cr ₅₀	2.3	200	-5
Si*	≤ 200	-	-
SmS	≤ 260	≈ 1000	≈ -100

* The properties of Si depend strongly on the dopant concentration (Figs. 5.4.4–5.4.7)

+ Temperature coefficient of resistivity

The various mechanisms contributing to piezoresistivity mentioned above are closely related to the symmetry of the crystal and to the distortion by the applied strain. One consequence is that the change in the resistance is different when the strain is applied parallel (longitudinal) and perpendicular (transverse) to the electrical current. In analogy to Eq. (1), this can be accounted for by longitudinal ($K_{GF,l}$) and transverse ($K_{GF,t}$) gauge factors:

$$\frac{\Delta R}{R} = K_{GF,l} \cdot \varepsilon_{m,l} + K_{GF,t} \cdot \varepsilon_{m,t} \quad (4)$$

$\varepsilon_{m,l}$ and $\varepsilon_{m,t}$ correspond to longitudinal and transverse strains, respectively. Furthermore, the orientation of the applied strain with respect to the crystal axes strongly influences its effect. In p-doped Si, for example, the longitudinal gauge factor varies by a factor of 100, depending on the orientation of the applied strain with respect to the crystal axes. The highest values (up to about 200) are observed along the $\langle 111 \rangle$ direction [18].

5.4.5.2 Polysilicon

Si is the semiconductor material that is most widely used for strain gauge applications. Usually p-doped Si is preferred, because it shows higher gauge factors than n-doped Si. The higher gauge factor has its origin in the degeneracy of the heavy-hole and light-hole bands, which is lifted when strain is applied [17]. Mostly boron (B) is used as a dopant for p-Si. Because single crystalline Si can not be grown on stainless steel substrates coated with an insulator, polysilicon is used for such strain gauge applications. However, for a given carrier concentration the gauge factor is considerably smaller than in single-crystalline Si, because of averaging among different crystal orientations in the polysilicon grains [19, 20]. However, the gauge factor is still about 10-fold higher in polysilicon than in metals.

Properties of polysilicon such as the gauge factor, the resistivity, and its temperature coefficient are considerably affected by the polysilicon's morphology – the grain size and the preferred orientation of the grains are especially important. The highest longitudinal gauge factors are obtained for a morphology in which the $\langle 111 \rangle$ orientation makes the largest contribution. Such a constellation can be obtained by preferentially (110) oriented growth in which the [111] axis is aligned within the plane of the polysilicon film [20]. The resistivity of B-doped polysilicon decreases with increasing dopant concentration (Fig. 5.4.4). Although the resistivity of polysilicon is higher than that of crystalline silicon at low dopant concentrations (below about 10^{18} cm^{-3}), the difference becomes smaller with increasing dopant concentration [20]. This observation is consistent with a significant contribution of grain boundaries to the electrical properties at low dopant concentrations. The grain contribution decreases when the dopant concentration is increased. For high dopant concentrations (above about 10^{19} cm^{-3}) the properties of the grains dominate [19–21].

Fig. 5.4.4 Sheet resistance R_s and resistivity ρ as a function of the boron doping concentration N_A in polysilicon (solid line) and crystalline silicon (dashed line) [20].

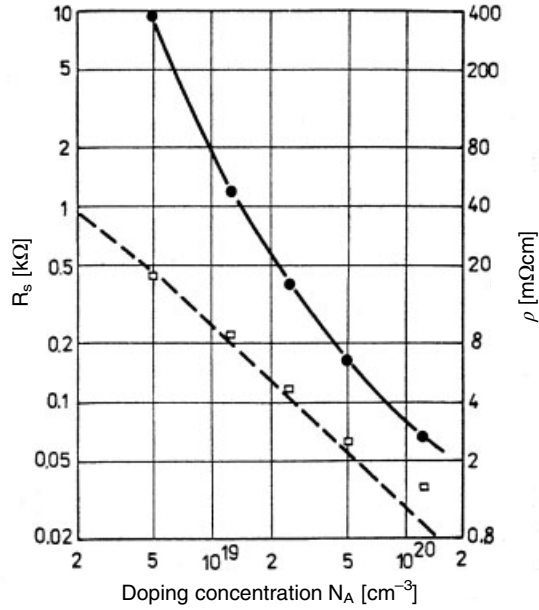
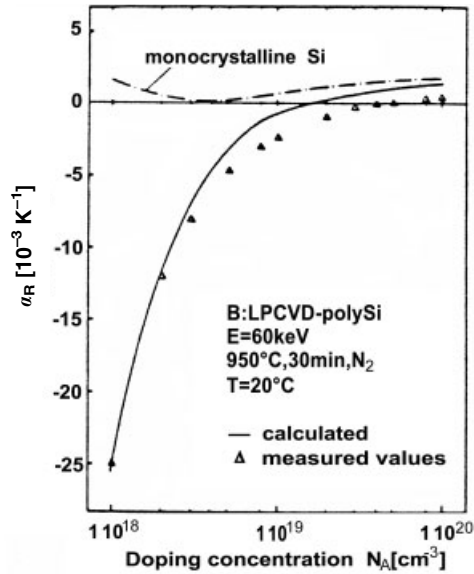


Fig. 5.4.5 Temperature coefficient of resistivity (α_R) as a function of boron-doping concentration in LPCVD polysilicon resistors at 20°C [22].



The temperature dependence of the resistivity α_R of boron-doped polysilicon increases monotonically with increasing dopant concentration (Fig. 5.4.5). Although the temperature dependence is negative at low dopant concentrations, it becomes positive at high concentrations. The negative temperature dependence for low dopant concentration is attributed to trapping of carriers at the grain boundaries,

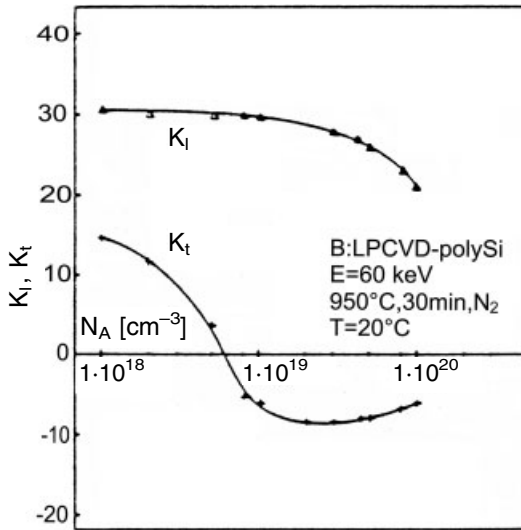


Fig. 5.4.6 Longitudinal (K_l) and transverse (K_t) gauge factors as a function of boron-doping concentration N_A in LPCVD polysilicon resistors [22]

leading to thermally activated conductivity [20–22]. At high dopant concentrations, the contributions of the grains themselves dominate and the grain boundaries become negligible.

The transverse gauge factor shows a stronger dependence on the dopant concentration than the longitudinal gauge factor [20, 22] (Fig. 5.4.6). At high dopant concentrations the longitudinal gauge factor decreases with increasing concentration. Although the transverse gauge factor is positive at low dopant concentrations, it becomes negative at high concentrations.

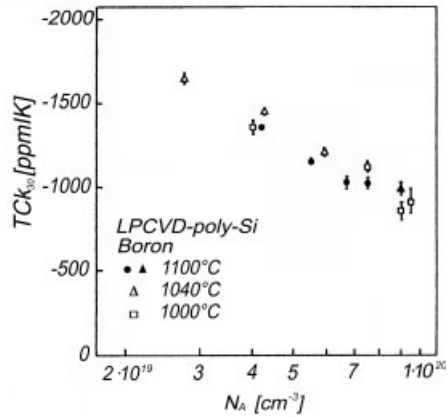
At lower doping concentrations the gauge factor also depends on the grain size: it increases with increasing grain size [21, 23, 24]. As mentioned above, at higher dopant concentrations the influence of the grain boundaries becomes negligible. Correspondingly, no dependence of gauge factor on grain size is observed at high dopant concentrations.

With increasing temperature the gauge factor decreases, independent of the dopant concentration [23, 24]. However, with increasing dopant concentration the temperature dependence of the gauge factor is significantly reduced [23, 24] (Fig. 5.4.7).

In summary, p-doped (boron) polysilicon is frequently used as a material for strain gauges on steel substrates. Even though the gauge factor is smaller than at lower dopant concentrations, highly doped polysilicon is usually preferred because of the lower temperature dependence of its resistivity and gauge factor. A preferential $\langle 110 \rangle$ oriented growth favors high gauge factors.

For the deposition of boron-doped polysilicon the maximum processing temperature of 600°C (due to the stainless steel substrate) constitutes a severe limitation. Two main strategies are followed for the deposition of polysilicon. One technique is based on laser (re-)crystallization [25, 26] in at least two steps. First, the material is deposited in an amorphous state and then it is crystallized in a second

Fig. 5.4.7 Differential temperature coefficient of the gauge factor of boron-doped polysilicon films as a function of dopant concentration N_A for different annealing temperatures [24]



step. However, deposition of highly boron-doped polysilicon in one step is preferred for high-volume production. Thus, the second technique consists in plasma-enhanced chemical vapor deposition (PECVD) [1] using silane (SiH_4) and diborane (B_2H_6). If the ratio of B_2H_6 to SiH_4 is high enough, polysilicon can be grown with a grain size of about 60 nm at a substrate temperature of 580 °C [1]. The of B_2H_6 supports an $\langle 110 \rangle$ oriented growth.

5.4.5.3 NiCr

The term $\Delta\rho/(\rho\varepsilon)$ in Eq. (3) can be neglected for most metal strain gauges, that is, the geometrical contribution dominates [15]. Assuming a material with a constant volume when strain is applied (i.e., $\nu=0.5$), the gauge factor is expected to be $K_{GF} \approx 2$. In addition to most pure metals, many alloys fall into this category of strain properties. Compared with pure metals, alloys have the advantage of higher resistivity and a lower temperature dependence of the resistivity. Examples include NiCr, CuNi, and PtIr (Table 5.4.2).

NiCr alloys are widely used because of their high resistivity, the low temperature coefficient of the resistivity, their good adhesive properties, and their high long-term stability [27]. The ratios of Ni:Cr commonly used vary between 80:20 and 50:50 [2, 28]. The electrical properties depend strongly on the Ni:Cr ratio. Starting from pure Ni, the resistivity increases with increasing Cr concentration [2, 3, 29]. While the resistivity increases, its temperature coefficient TCR decreases from positive to negative values with increasing Cr concentration [28, 29] (see Fig. 5.4.8 for $\text{Ni}_{100-x}\text{Cr}_x$ and Fig. 5.4.9 for $\text{Ni}_{90-x}\text{Cr}_x\text{Si}_{10}$). A decrease in TCR with increasing resistivity is generally observed for alloys as first discovered by Mooij in 1973 [29]. A theoretical explanation does not yet exist.

The temperature dependence of $\text{Ni}_{50}\text{Cr}_{50}$ is near zero (Fig. 5.4.8), which is why $\text{Ni}_{50}\text{Cr}_{50}$ is often chosen for metal strain gauge applications. The important material parameters are a resistivity of 200 $\mu\Omega\text{cm}$ and a TCR of -5 ppm K^{-1} (after heat treatment for 24 h at 150 °C) [3], the corresponding gauge factors are $K_{GF,1}=2.3$ and $K_{GF,t}=-0.83$ [3]. In some cases, small amounts of Si are added (up to 10%) to

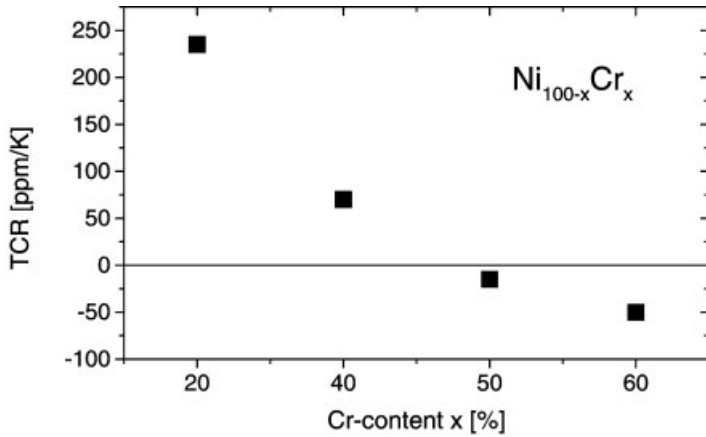


Fig. 5.4.8 Temperature coefficient of the resistance of $\text{Ni}_{100-x}\text{Cr}_x$ alloys as a function of the Cr content x . The data are taken from [3]

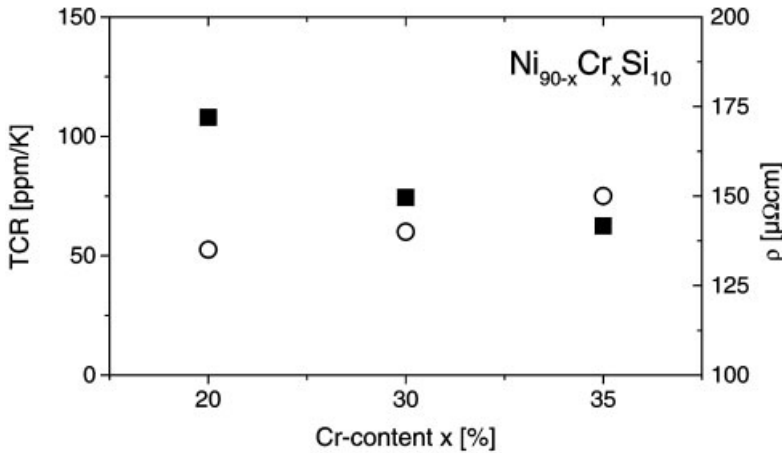


Fig. 5.4.9 Temperature coefficient of the resistance (squares) and resistivity (circles) of $\text{Ni}_{90-x}\text{Cr}_x\text{Si}_{10}$ alloys as a function of the Cr content x . The data are taken from [28].

further reduce the temperature coefficient of $\text{Ni}_{50}\text{Cr}_{50}$ [29, 30]. Small amounts of Si can also improve the thermal stability of the resistive films [30].

To obtain a sufficiently high sheet resistance for strain gauges, the thickness of the NiCr layers have to be in the range of 50–100 nm. If the thickness is significantly below 100 nm, surface effects become important for the electrical behavior. In consequence, the electrical properties deviate from the bulk properties [31]. Typically, the resistivity increases with decreasing thickness while the temperature coefficient of the resistance decreases. Another consequence is that the surface

Tab. 5.4.3 Assessment of NiCr and polysilicon with respect to the selection criteria for a strain gauge material (+: very good, 0: sufficient)

Criterion	NiCr	Polysilicon
Gauge factor GF	0	+
Temperature dependence GF	+	0
Resistivity	0	+
Temperature dependence of resistance	+	0
Stability	+	0
Reproducibility of deposition process	+	0

roughness at the interface to the underlying layer has to be controlled carefully to obtain reproducible electrical properties.

The gauge factor of NiCr alloys has low temperature dependence, decreasing only slightly with increasing temperature. With respect to the sensitivity of strain gauges on steel substrates, this decrease in the gauge factor is partially compensated for by the decreasing Young's modulus of the steel substrate.

Magnetron sputtering [32] from a sintered target or evaporation [32, 33] can be used to deposit NiCr(Si) layers. Magnetron sputtering is usually preferred, and the deposition typically occurs without heating of the substrates. The plasma does heat the substrates during deposition; however, the substrate temperature usually stays below 100 °C, resulting in amorphous films. Incorporating oxygen into resistive films enhances their thermal stability [30, 34]. Oxygen can be incorporated by reactive sputtering in an Ar/O₂ atmosphere or by thermal treatment in an atmosphere containing O₂ (e. g., air) after deposition [2, 3, 30].

A thermal treatment is important for stabilizing the sheet resistance and the TCR [2, 30, 35], suitable temperatures being approximately 300 °C [2, 35]. Treatment in air leads to an increase in the resistance and a shift of the TCR in the positive direction [35]. A comparison of polysilicon and NiCr as strain gauge materials is shown in Table 5.4.3. Although polysilicon offers significantly higher gauge factors and higher resistivity, NiCr has the advantage of a lower temperature dependence, higher stability, and better reproducibility of the deposition process. The decision between the different materials depends strongly on the target application.

5.4.5.4 Wheatstone Bridge

Arranging four strain gauges in a Wheatstone bridge configuration is an established design for detecting small fractional resistance changes (Fig. 5.4.10).

The output voltage ΔV is given by the fractional resistance changes in the single resistors $\Delta R_i/R_i$ and the applied voltage V [36]:

$$\frac{\Delta V}{V} = \frac{1}{4} \left(\frac{\Delta R_1}{R_1} - \frac{\Delta R_2}{R_2} + \frac{\Delta R_3}{R_3} - \frac{\Delta R_4}{R_4} \right) \quad (5)$$

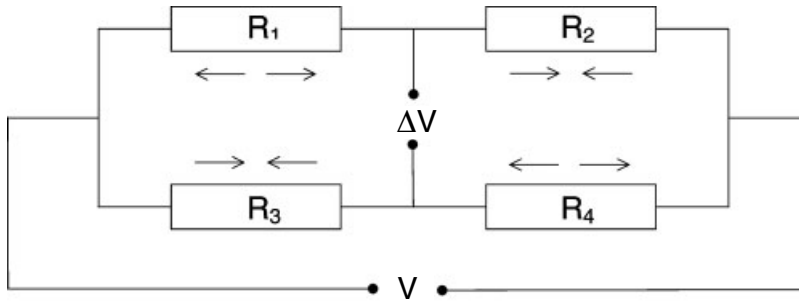


Fig. 5.4.10 Schematic view of four resistors in a Wheatstone bridge configuration; V is the applied voltage and ΔV is the output signal V

Because of the bridge configuration, the variation in resistance is transformed into a voltage signal. Temperature compensation is efficient if all resistors in the bridge are exposed to the same temperature and show have the same temperature dependence. To maximize the output signal, neighboring resistors have to experience resistance changes that differ in sign. Therefore, neighboring resistors are either located in areas with opposite signs of strain (e. g., for pressure sensors – at the center and edge of the membrane) or, if the longitudinal and transverse gauge factors differ in sign, the gauges are oriented in alternating transverse and longitudinal directions with respect to the strain.

The output signal can be estimated as follows. The sensing unit is usually operated in an elastic regime. Therefore, the maximum strain for steel substrates is limited to a value of about 0.1%. Considering a gauge factor of 2, the fractional resistance variation $\Delta R_i/R_i$ is expected to be $2 \cdot 10^{-3}$ (Eq. (1)). If we further assume the same fractional resistance variation for all four strain gauges (but not the same signs), the output signal $\Delta U/U$ is expected to be 2 mV/V (Eq. (5)).

In addition to being used in strain gauges, resistors are used for temperature detection. Temperature detection is performed by metals whose resistivity shows a significant, almost-linear temperature dependence. Examples are Ni, Ti, and Pt.

5.4.6

Metal Contacts

In addition to the strain gauge material itself, a second conducting layer is usually required to electrically connect the strain gauges with the contact pads. The requirements for a good contact layer can be summarized as follows:

- good electrical contact with the sensing layer
- low resistivity
- compatibility with downstream processing steps (e. g., bonding, soldering)
- stability to environmental conditions
- thermal stability
- good adhesion to adjacent layers

- stable interfaces to adjacent layers
- temperature coefficient of resistivity matching that of the sensing layer
(In some designs, the sign of the temperature coefficient of resistivity in the contact material is opposite to that in the sensing layer, for temperature compensation.)

Depending on the application, it may be difficult to meet all requirements with a single layer. So a sandwich of layers consisting of, for example, an adhesion layer, a diffusion barrier, and the contacting layer, may be used.

Applications in which the electrical connections of the sensing element are made by soldering are rare, Ni can be used as contact material in such applications [6]. To protect Ni against oxidation, it is often covered by a very thin layer of Au [6]. Most applications make use of bond contacts of Al or Au wires. Frequently used contact materials include Ni (Al wire), Al, or Au. NiCr is a good adhesion layer for Ni. Therefore, Ni is a popular contact material for NiCr layers, which can then function simultaneously as both a sensing and an adhesion layer. However, Au is used for most automotive applications, because it is the most stable against corrosion. Ti and NiCr are suitable adhesion layers for Au.

For Al bonds on Au or vice versa, the Kirkendall effect leads to a well known failure mechanism [37]. The Kirkendall effect results from the different diffusion coefficients of Au and Al in different phases of the system AlAu [38]. The Al₂Au phase ('purple phase'), which is always present, plays a crucial role. The diffusion coefficient of Al through Al₂Au is much higher than that of Au. As a result, voids are created as soon as the temperature reaches the diffusion temperature, which reduces the stability of the bond contact.

If the Au contains some Pd impurities, the behavior of Al–Au bond contacts is much more stable [39]. The creation of AlAu phases is believed to be impeded by the Pd.

Standard deposition techniques for the above materials are evaporation and sputtering. Sometimes, Ar sputtering is done before deposition to remove thin oxide layers or contamination, to ensure better electrical contact and to improve adhesion.

5.4.7

Passivation

For most applications, the sensor area is exposed to the environment. Therefore, a passivation layer is required. Important qualities for the passivation layer are:

- high electrical resistance
- high resistance to moisture penetration
- high resistance to ion penetration
- good adhesion
- high temperature stability
- low material stress

Si_xN_y and SiO_x are commonly used as passivation layers [40]. It is known from experience with integrated circuit devices that alkali cations (e.g., sodium) diffuse into SiO_x [41], which may be the origin of reliability problems that occur during operation at high temperatures and high voltages. In addition, Si_xN_y shows higher resistance to humidity. Therefore, Si_xN_y is superior to SiO_x as a passivation layer. The thickness is typically 300–600 nm.

Similar to SiO_x the thermal expansion coefficient of Si_xN_y is significantly lower than that of the steel substrate (Table 5.4.1). Consequently, the internal stress in the Si_xN_y layer has to be chosen accordingly (see Section 5.4.4), to prevent the development of cracks during operation. A compressive stress of 150 MPa is typical. High density also leads to higher stability against cracks [9] and improve resistance to humidity.

Si_xN_y is usually deposited by PECVD [40]. Due to temperature restrictions imposed by the steel substrate, the maximum deposition temperature is limited. A further limitation can result from layers that have been deposited before the Si_xN_y . When Au is used as a contact material, the deposition temperature is limited to $<280^\circ\text{C}$ because Au diffuses at higher temperatures. PECVD processes for silicon nitride deposition at temperatures below 280°C are well known and commercially available. However, the low deposition temperature leads to a higher hydrogen content, making the films mechanically and chemically less stable [40].

5.4.8

Patterning

Three layers have to be patterned to fabricate strain gauges: the sensing layer, the contact layer, and the passivation layer. Usually, the highest precision is required when patterning the sensing layer. Minimum line widths are about $30\ \mu\text{m}$. Examples of layouts for high-pressure sensor elements are shown in Fig. 5.4.11.

In response to requirements and manufacturing approaches, different patterning techniques have been developed. Photolithographic techniques similar to those used in Si technology [42, 43] are used. Shadow mask techniques are well

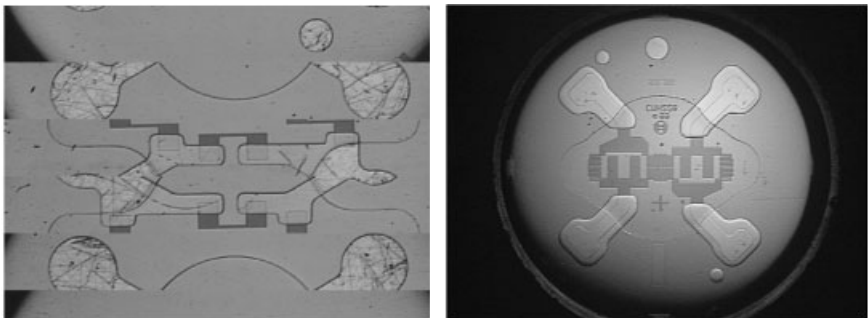


Fig. 5.4.11 View of thin film system in high-pressure sensor elements of Nagano Keiki Co. (left) and Robert Bosch GmbH (right)

established for fabricating patterns whose sizes are considerably greater than 100 μm . Laser patterning techniques have been considered, but are not widespread today. In the following we look at the established techniques.

Photolithography consists of several steps: resist coating, exposure of the resist, development of the resist, etching of the layer, and stripping [42]. The design of each step is strongly affected by the manufacturing approach. To illustrate, spin coating is typically used to deposit a resist. Although this technique can be used to process a stainless steel wafer, topography considerations make it not feasible for batch processing on a jig. Alternative techniques, like roller coating or spin coating, were developed for jig processing. Important parameters for obtaining sufficiently homogeneous resist layers are the conditioning of the surface, the composition of the resist, and the treatment while drying the resist. In particular, the solvent concentration strongly affects the homogeneity of each element.

For exposure, techniques with a larger depth of focus, such as proximity or projection exposure, are used. Such techniques can handle the reduced planarity of steel substrates compared to semiconductor wafers.

For etching, depending on the layer to be patterned, wet or dry processes can be used. For wet processing steps such as developing, etching, and stripping, special attention has to be paid to removing all residues from the jig. Often, special cleaning steps have to be introduced for the jig. Of course, the media used depend strongly on the resist and the layers to be patterned.

A shadow mask technique can be used if the required resolution is sufficiently low, as is true for the contact and the passivation layers. During deposition of the layer to be patterned, a metal sheet (mask) is typically used, which has openings only in those areas where material is to be deposited. With this technique, deposition and patterning are combined into a single processing step. Special fixtures are required to mount the masks, and special attention has to be paid to preventing generation of particles. Masks perform best in directed deposition techniques like evaporation; however, they can also be used in sputtering and PECVD processes.

5.4.9

Trimming

The manufacturing tolerances of the single strain gauges result in a distribution of the offsets of the Wheatstone bridges. Depending on the signal processing concept used in the sensor, the distribution may be too broad. If so, it can be reduced significantly by trimming.

Analog trimming concepts, in which the offset is measured online while resistors are adjusted by laser trimming, are widely used. This type of trimming has to be done before deposition of the passivation layer, to ensure that the laser cut will be covered by the passivation layer. The laser has to be chosen and adjusted carefully to avoid damage to the insulation layer, which could reduce its long-term stability.

Other parameters such as the temperature coefficient or the output signal can also be trimmed and are discussed elsewhere [15].

5.4.10

Summary and Outlook

Thin film technology on steel is well suited for high precision sensors in automotive applications. Well established applications include high pressure sensors in the range of 140 to 200 MPa. For these applications, a manufacturing volume of several million sensors per year has been reached.

Challenges for the future include improving the accuracy beyond the high level that has already been reached and extending the variety of applications. Future applications can be force or momentum measurement for sensing weight, braking force measurement, and momentum measurement on the wheel bearing.

5.4.11

Acknowledgements

We thank M. Metz, J. Muchow, and C. Schelling for critical reading of the manuscript.

5.4.12

References

- 1 K. KAMIMURA, Y. ONUMA, T. HOMMA, *Sens. Actuators* **1988**, *13*, 71–77.
- 2 A. GARCIA-ALONSO, J. GARCIA, E. CASTANO, I. OBIETA, F. J. GARCIA, *Sens. Actuators* **1993**, *A37–38*, 784–789.
- 3 A. GARCIA-ALONSO, E. CASTANO, I. OBIETA, J. GARCIA, F. J. GARCIA, *Vacuum* **1994**, *45*, 1103–1105.
- 4 *Data sheet N700*, Böhler Edelstahl GmbH, Kapfenberg, Austria.
- 5 J. MUCHOW, A. KRETSCHMANN, R. HENN, K. SKROBANEK, S. FINKBEINER, H. R. KRAUSS, *Proc. SPIE* **2002**, *4755*, 143–148.
- 6 G. SCHULTES, *wägen + dosieren* **1996**, *3*, 3–10.
- 7 O. DÖSSEL in *Sensoren – Technologie und Anwendung*, NTG-Fachberichte, VDE-Verlag, Berlin, **1986**, *93*, 143–147.
- 8 S. M. SZE, *Physics of Semiconductor Devices*, 2nd ed., John Wiley & Sons, New York, NY, USA, **1981**, p. 405.
- 9 A. K. SINHA, H. J. LEVINSTEIN, T. E. SMITH, *J. Appl. Phys.* **1978**, *49*, 2423–2426.
- 10 G. LUCOVSKY, *Adv. Mater. Opt. Electron.* **1996**, *6*, 55–72.
- 11 M. F. CEILER, P. A. KOHL, S. A. BIDSTRUP, *J. Electrochem. Soc.* **1995**, *142*, 2067–2071.
- 12 B. L. CHIN, E. P. VAN DER VEN, *Solid-State Technology* **1988**, *4*, 119–122.
- 13 W. KULISCH, T. LIPPMANN, R. KASSING, *Thin Solid Films* **1989**, *174*, 57–61.
- 14 B. KLOECK, N. F. DE ROOIJ in *Semiconductor Sensors*, chap. 4 (S. M. SZE, ed.), John Wiley & Sons, New York, NY, USA, **1994**.
- 15 H. S. SCHAUMBURG, *Sensoren*, B. G. Teubner, Stuttgart, Germany, **1992**, p. 141.
- 16 Y. KANDA, *Sens. Actuators* **1991**, *A28*, 83–91.
- 17 P. Y. YU, M. CARDONA, *Fundamentals of Semiconductors*, Springer, Berlin, Germany **1996**, p. 63 ff.
- 18 O. JÄNTSCH in *Sensorik* (W. HEYWANG, ed.), Springer, Berlin, Germany **1984**, 25.
- 19 P. J. FRENCH, A. G. R. EVANS, *J. Phys. E: Sci. Instrum.* **1986**, *19*, 1055–1058.
- 20 D. SCHUBERT, W. JENSCHKE, T. UHLIG, F. M. SCHMIDT, *Sens. Actuators* **1987**, *11*, 145–155.

- 21 P. J. FRENCH, A. G. R. EVANS, *Solid-State Electron.* **1989**, *32*, 1–10.
- 22 P. KOPYSTYNSKI, E. OBERMEIER, *Sens. Actuators* **1989**, *17*, 69–73.
- 23 J. Y. W. SETO, *J. Appl. Phys.* **1976**, *47*, 4780–4783.
- 24 H. SCHÄFER, V. GRAEGER, R. KOBBS, *Sens. Actuators* **1989**, *17*, 521–527.
- 25 E. LÜDER, *Sens. Actuators* **1986**, *10*, 9–23.
- 26 J. BINDER, W. HENNING, E. OBERMEIER, H. SCHABER, D. CUTTER, *Sens. Actuators* **1983**, *4*, 527–536.
- 27 J. RÖLKE, *Electrocomponent Sci. Technol.* **1981**, *9*, 51–57.
- 28 T. NAOHARA, K. SHINOHARA, A. INOUE, T. MASUMOTO, H. TOMIOKA, *Rapidly Solidified Mater., Proc. Int. Conf.* (P. W. LEE and R. S. CARBONARA, eds.) San Diego, CA, USA **1986**, 249–253.
- 29 J. H. MOOIJ, *Phys. Status Solid. A* **1973**, *17*, 521–529.
- 30 W. GAWALEK, *Thin Solid Films* **1984**, *116*, 205–210.
- 31 K. NITZSCHE, H.-J. ULLRICH, *Funktionswerkstoffe der Elektrotechnik und Elektronik*, Alfred Hüthig, Heidelberg, Germany **1986**.
- 32 F. D. BARLOW, A. ELSHABINI, R. BROWN in *Thin Film Technology Handbook, chap. 1* (A. ELSHABINI-RIAD and F. D. BARLOW, eds.), McGraw-Hill, New York, NY, USA, **1998**.
- 33 L. LASSAK, K. HIEBER, *Thin Solid Films* **1973**, *17*, 105–111.
- 34 H. DINTNER, R. MATTHEIS, G. VOGLER, *Thin Solid Films* **1989**, *182*, 237–246.
- 35 J. H. MOOIJ, M. DE JONG, *J. Vac. Sci. Technol.* **1971**, *9*, 446–449.
- 36 H. PAUL, *Druckaufnehmer mit metallischen Dehnungsmessstreifen*, Vortragsreihe in der Technischen Akademie Wuppertal, **1990**.
- 37 T. RAYMOND, *Semicond. Int.* **1989**, *9*, 152–158.
- 38 E. PHILOFSKY, *Solid-State Electron.* **1970**, *13*, 1391–1399.
- 39 J. F. HAAG, *DVS-Berichte* **1990**, *129*, 131–136.
- 40 P. SCHMID, M. ORFERT, M. VOGT, *Surf. Coat. Technol.* **1998**, *98*, 1510–1517.
- 41 S. M. SZE in *Physics of Semiconductor Devices, 2nd ed.* John Wiley & Sons, New York, NY, USA, **1981**, p. 392.
- 42 L. K. WHITE, *RCA Review* **1986**, *47*, 345–379.
- 43 C. H. MASTRANGELO, W. C. TANG, in *Semiconductor Sensors, chap. 2* (S. M. SZE, ed.), John Wiley & Sons, New York, NY, USA **1994**.

5.5

Thin Films on Silicon

MATTHIAS FÜRTSCH and STEFAN FINKBEINER

5.5.1

Introduction

Twenty years after the first major review paper on silicon as a material for mechanical applications was published by K.E. Petersen in 1982 [1], devices produced by silicon micromachining technology play a dominant role in many kinds of sensor applications, especially for inkjet heads and for sensors in the automotive industry. A major success factor in silicon micromachining technology is its close link to the well-established field of silicon microelectronics. In production schemes for silicon microstructures, most of the processing steps, equipment, and materials come from the microelectronics industry. The switch from microelectronics to micromachining can be performed by adding a limited number of sensor-specific processing steps. These include deposition and etching of selected thin film materials.

In this contribution, we concentrate on the sensor-relevant applications of thin films. These are used, for example, for:

- membrane layers (e.g., silicon, polysilicon, silicon oxide, silicon nitride)
- electrical connections and resistors (e.g., aluminum, platinum, polysilicon)
- sacrificial layers (e.g., silicon oxide)
- functional layers (e.g., polysilicon)
- passivation (e.g., silicon oxide, silicon nitride)

For automotive applications, the various functions require several special material properties. Here, we concentrate on properties of thin films rather than on their production processes. A wide range of publications deal with sensor-specific processes, for example, silicon reactive ion etching (RIE) using the Bosch trench process [2] and sacrificial oxide etching [3, 4]. The details of standard deposition and structuring processes are described in numerous books on semiconductor technology (e.g., [5, 6]), and they are not discussed in depth here.

In Section 5.5.2, we focus in detail on the essential material properties of thin films for sensors in automotive applications. In addition, several examples of measurement structures to determine these properties are given.

The topics of subsequent sections are the different materials used for thin films, specifically, silicon oxide (Section 5.5.3), silicon nitride (Section 5.5.4), polysilicon (Section 5.5.5), and other materials, such as metals and silicon oxynitride, as well as novel thin film materials that may appear in future sensor designs (Section 5.5.6). In these sections, we adhere to the following structure: first, the basic deposition process is briefly described, followed by the common function of the layer in sensor design. Subsequently, we discuss in detail the important material properties.

5.5.2

Material Properties

5.5.2.1 Requirements for Automotive Applications

The properties of thin films in automotive sensors often have to meet different requirements than those in microelectronics. In general, the electrical properties are of less importance than the mechanical properties. The special requirements and constraints for the materials are detailed below.

High volume applications. The production output in the automotive industry is very high and typically well above 1 million sensors per year and application. Therefore, the properties of the materials must be extremely reproducible, and the production processes need to exhibit very good long term stability, and the products must be insensitive to small variations in processing.

Robust sensor materials. Due to the harsh environmental conditions in an automobile, for example, in the engine compartment, and the necessary long life ex-

pectancy of the sensors, the membranes, cantilevers, etc., must be temperature stable and possess a very good long-term durability.

In consequence, the following material properties are of paramount importance:

- The intrinsic stress of the thin films. High stresses lead to mechanical instability from, for example, buckling of microstructures to fracturing of membranes. In addition, high wafer bowing due to high stress can lead to handling problems.
- The stress gradient within the thin films. High stress gradients lead to extreme curvatures of the released structures, thereby modifying the functionality of, for example, comb-structure acceleration sensors produced by Bosch or Analog Devices.
- The fracture strength of the thin films. Under working conditions, external mechanical stresses act on the sensors; the respective materials need to withstand these loads.

5.5.2.2 Measurement Methods

5.5.2.2.1 Stress and Stress Gradient

Thin films can exhibit intrinsic compressive or tensile stress. A thin film under compressive stress tries to expand; a thin film under tensile stress tries to contract. Once the optimum values for stress and stress gradients are found, it is crucial to maintain a stable process. Several possible techniques for measurement and in-process control of stress and stress gradients are discussed below.

The easiest way to measure stress in thin films after deposition is to analyze the change in the radius of curvature of the wafers before and after film deposition on one side, as first proposed by Stoney [7]. However, this technique usually requires the use of test wafers. After complete processing of the wafers, the stress can be obtained by measuring the deflection of membranes or indicator structures [8]. To measure compressive stress, the buckling technique on double-side supported bridges [9] and harp-like structures [10] can be applied.

A stress gradient is caused by an inhomogeneous stress profile within the thin film and results in an internal bending moment, causing vertical deflection of released microstructures. Possible causes of stress gradients include an inhomogeneous grain or dopant distribution in the thin film (Section 5.5.5). A thin film has a negative stress gradient when the top part of the structure is more compressive than the bottom part; upon release, the thin film exhibits concave curvature; for a positive stress gradient, the situation is reversed. The direction in which a structure bends depends on the stress gradient and the type of suspension. For example, the deflections of two structures made from a material with a positive stress gradient are shown in Figure 5.5.1. Beams supported on one side bend upwards (Fig. 5.5.1a), whereas beams suspended on both sides bend downwards (Fig. 5.5.1b).

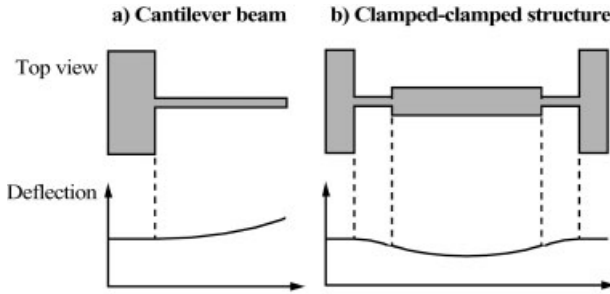


Fig. 5.5.1 Dependence of the deflection of structures on the suspension type for a material with a positive stress gradient, for **a** a cantilever beam and **b** a clamped-clamped structure

An example of the practical consequences of vertical deflection is shown in Figure 5.5.2 for a comb-type accelerometer fabricated with polysilicon. In Fig. 5.5.2 a, a positive stress gradient leads to an upward deflection of the fixed outer beams and a downward deflection of the proof mass. In Fig. 5.5.2b, the situation is reversed due to a negative stress gradient in the material [11].

The measurement of a stress gradient Γ is fairly simple. The linear stress gradient Γ , which in this definition is a mathematically calculated average value, can be derived from the deflection z and the length x of cantilever beams suspended on one side in relation to Young's modulus E_m , as

$$\Gamma = 2 \cdot E_m \cdot \frac{z}{x^2} \quad (1)$$

An additional contribution to the deflection can be introduced by an in-plane deformation of the supporting layer, which modifies the calculated stress gradient [12].

5.5.2.2.2 Mechanical Strength

A comprehensive review of the measurement of mechanical properties such as fracture strength or adhesion strength can be found in the literature [13]. In general, the best way to determine mechanical properties is to test micromachined structures that are as close to the actual design as possible using, for example, the micromanipulator developed at the University of Uppsala [14].

The fracture strength is commonly measured by applying a static load to membranes or cantilever beams. Due to the statistical nature of fracture processes, many tests have to be performed to obtain a statistically relevant result. For membrane structures, knowing the burst pressure is essential. Section 5.5.5 presents several tensile-fracture testing methods for surface micromachined polysilicon cantilever beams.

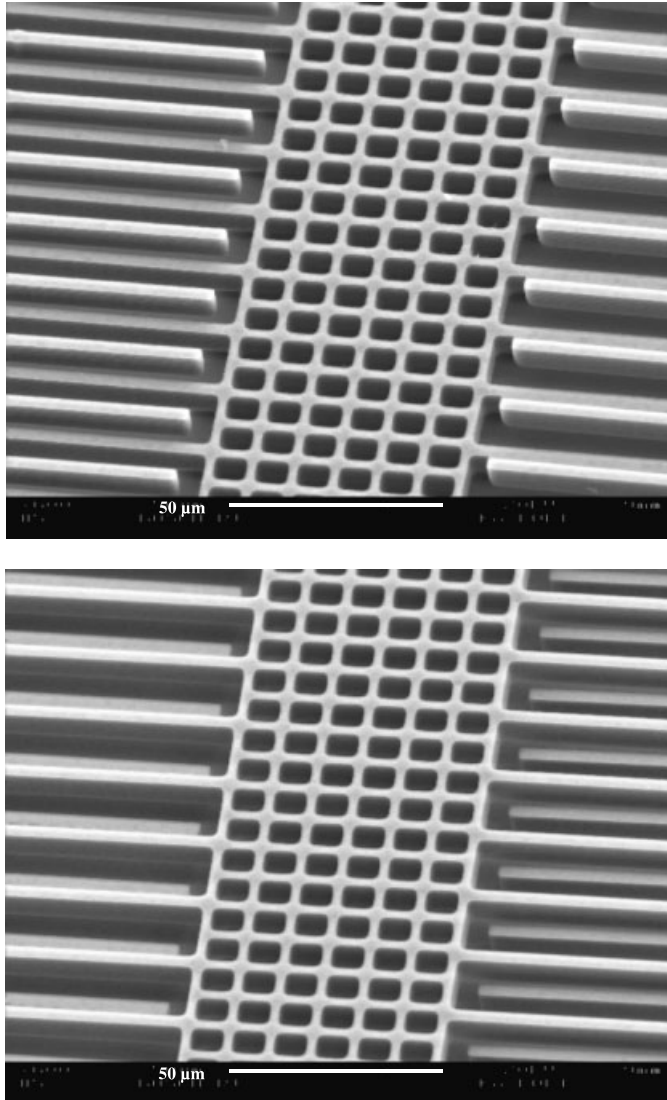


Fig. 5.5.2 Close-up of the proof mass of an acceleration sensor made of a material with **a** positive and **b** negative stress gradient

The adhesive properties of thin films are a crucial material property, because micro-machined structures generally consist of several layers deposited on top of one another. Insufficient adhesion inevitably leads to device failure. Only a few methods for measuring adhesion strength are mentioned in the literature. A widespread qualitative test of film adhesion is the tape test, in which a piece of adhesive tape is pressed on the surface and then pulled off. If the film adhesion is good, no material

is peeled off with the tape. If the tape test fails, the adhesion is not sufficient for using such films in sensors designed for automotive environments. Other adhesion tests include scratch testing, pull testing, and shear testing [11].

5.5.3

Silicon Oxide

In this and subsequent sections, the thin films commonly used in silicon micro-machined structures are discussed in detail, with particular focus on the important film properties mentioned in Section 5.5.2.

5.5.3.1 Processing and General Function

Silicon oxides (SiO_x) are the most widely used thin films in silicon microelectronic and micromechanical devices. Similar to silicon nitride (Section 5.5.4), these amorphous films exhibit dielectric properties. Silicon oxide is often utilized as part of a dielectric membrane, as a passivation or insulating layer, or as a sacrificial layer, which can be etched with hydrofluoric acid (HF)-containing etchants. Two different approaches to forming a silicon oxide thin film are

- *Thermal oxide* is grown directly on silicon by exposing the wafers to an oxidizing gas (O_2 or H_2O) at an elevated temperature, usually above 900°C . Usually, a large number of wafers are treated in parallel in a furnace. During thermal oxidation, the oxidizing species diffuses through the oxide layer to the Si-SiO₂ interface, with two consequences: first, the growth of an oxide of thickness x consumes a layer of silicon of thickness $0.44 x$. Second, due to the diffusion mechanism, the growth rate decreases significantly with increasing thickness, limiting the practical maximum thickness of these layers to about $3\ \mu\text{m}$. Thermal oxides have a high density and are the best oxides in terms of structural quality and electrical properties. The HF etching rate is generally slower than for CVD oxides (next paragraph).
- *Chemical vapor-deposited oxides* (CVD) are deposited on the wafers at temperatures usually well below those used in thermal oxidation. The thin films are deposited from the vapor phase directly on top of the wafer surface, resulting in some differences from thermal oxides: CVD oxides can be deposited on any type of substrate (including metals) and the growth rate is linear; thus, in principle very thick layers of silicon oxide can be produced with this technique. Depending on the deposition temperature, different reactions are utilized. The most important types of oxide are *PECVD oxide* (plasma-enhanced CVD), deposited at a temperature between 250°C and 400°C by means of a reaction of silane (SiH_4) and oxygen (O_2) or nitrous oxide (N_2O); and *TEOS oxide*, deposited at a temperature between 650°C and 800°C by means of the decomposition of tetraethyl orthosilicate (TEOS, $\text{Si}(\text{OC}_2\text{H}_5)_4$). PECVD oxides usually incorporate a significant amount of hydrogen or nitrogen, up to 2%. Upon thermal annealing, hydrogen diffuses out of the PECVD oxide, increasing the density of the film. PECVD oxides are usually used as passivation layers, because of their relatively low deposition temperature. This allows their deposition after metallization with aluminum, after which processing temperatures cannot exceed 450°C .

5.5.3.2 Intrinsic Stress in Silicon Oxide Layers

Of the layers mentioned above, the thin films showing the widest range of stress are the PECVD oxides. The stress state of the oxide layers becomes important when such films are used as part of a membrane or as a passivation layer. Membranes that are under considerable compressive stress tend to buckle, severely changing their mechanical properties. In contrast, considerable tensile stress in membranes can lead to crack formation and fracture. The passivation properties of PECVD oxide layers, for example, towards humidity, depend sensitively on their composition, which, in turn, has a great influence on the stress.

Table 5.5.1 shows typical stress data for selected types of oxide. Thermal oxide layers are very stable in terms of stress, similar to TEOS-deposited oxides, and their stress is nearly uninfluenced by deposition parameters. However, depending on the deposition parameters, the stress in PECVD oxide can vary significantly as shown below. The reasons for the large variations in stress in PECVD oxides are manifold.

- PECVD oxides can contain hydrogen or nitrogen impurities on a percentage level; the incorporation is greatly influenced by the deposition process.

Tab. 5.5.1 Typical stress values for different oxide thin films

Oxide type	Stress (MPa)
Thermal oxide	About -300 to -400 (compressive)
TEOS	About -100 to -200 (compressive)
PECVD oxide, as deposited	From -300 (compressive) to + 200 (tensile), depending on deposition parameters
PECVD oxide, annealed at $T > T_{\text{dep}}$	Shift towards tensile stress due to structural changes and stress relaxation at higher annealing temperatures, typically $>750^\circ\text{C}$ [41]

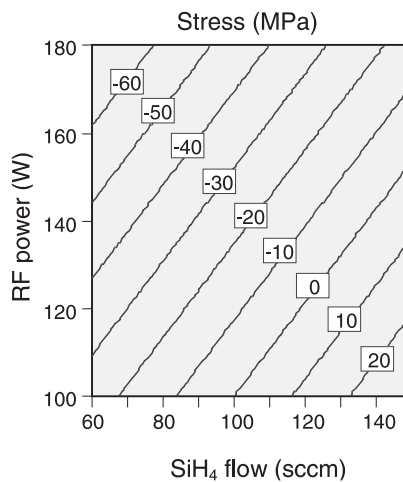


Fig. 5.5.3 Dependence of stress in a PECVD-oxide thin film on selected deposition parameters like gas flow and plasma power

- The silicon-to-oxide ratio usually differs from the optimum 1:2 ratio ($\Rightarrow \text{SiO}_x$). Stoichiometric films ($x=2$) are usually less tensile than silicon-rich films ($x < 2$).

During deposition, many processing parameters affect the stress of the deposited layer, including the gas flow rates, the pressure during deposition, the plasma RF power, and the deposition temperature. For example, data on the stress of the oxide as affected by selected deposition parameters are shown for an Applied Materials AMAT 5000 PECVD machine in Figure 5.5.3. The large influence of gas flow rate is clearly visible. The dependence of stress in PECVD oxides on these parameters is different for every piece of manufacturing equipment and can even differ for two similar models from the same manufacturer.

5.5.4

Silicon Nitride

5.5.4.1 Processing and General Function

Silicon nitride (Si_3N_4) is commonly used in silicon microelectronics and micromechanics as a passivation layer and, because of its dielectric properties, it is also widely used in capacitors. The dielectric constant of silicon nitride is twice as high as the dielectric constant of silicon oxide, thus allowing higher capacitances for identical capacitor geometries.

Silicon nitride as a passivation layer on top of an electronic circuit or a metal structure is an excellent diffusion barrier against water and protects the electronic device from organic and metallic (e.g., Na, K) contaminants. Silicon nitride is also used as a masking layer for wet anisotropic etching of silicon (in KOH), as part of a dielectric membrane, and for mechanical protection in micromechanics: during face-down handling of the front of electronics while processing the back of the wafer, a silicon nitride passivation layer prevents defects and scratches on the sensitive front side.

Silicon nitride is deposited by chemical vapor deposition (CVD). Two methods are commonly used:

- Low-pressure (LP) CVD deposition of silicon nitride from SiCl_2H_2 (dichlorosilane, DCS) and NH_3 at temperatures between 700°C and 900°C at a pressure of < 100 Pa to form Si_3N_4 on top of a wafer. Due to the high processing temperatures, the hydrogen content of *LPCVD-silicon nitride* is only about 3%, and most of the hydrogen resides in NH groups [15]. The deposition rate is typically around 4 nm min^{-1} [16]. Multiwafer deposition furnaces are used to obtain higher throughput. Because of the required high deposition temperature, LPCVD-silicon nitride is not well suited as a backend passivation for aluminum-metallized wafers.
- Plasma-enhanced (PE) CVD deposition of silicon nitride is typically performed at temperatures between 250°C and 400°C from silane (SiH_4) and ammonia (NH_3) gases. The plasma enhances the deposition rates. The deposition rate strongly depends on the deposition temperature and RF power, typical rates being around 400 nm min^{-1} [17]. Because of the low deposition temperature,

the hydrogen content of *PECVD-silicon nitride* is much higher than that of *LPCVD-silicon nitride*. Depending on the deposition temperature, the hydrogen content is about 10%–20% [17] but values as high as 39% have been reported [15]. The hydrogen enhances the etch rate of silicon nitride in (dilute) HF acid [17].

5.5.4.2 Intrinsic Stress in Silicon Nitride Layers

The intrinsic stress in silicon nitride layers is important especially in micromechanical devices: the characteristic features of membranes based on silicon nitride or thin silicon structures covered with silicon nitride are very sensitive to stress variations in these layers. Similar to *PECVD-oxide*, the intrinsic stress of silicon nitride is influenced by several deposition parameters such as gas flow, deposition temperature, and radiofrequency (RF) power.

The stress of *LPCVD-silicon nitride* can be varied from tensile stresses of 1 GPa to compressive stresses of –200 MPa [16, 18, 19] depending on the process temperature and especially on the composition of the process gases. Changing the ratio of gaseous dichlorosilane (DCS) to ammonia (NH₃) from (for example) 2:1 to 8:1 results in a stress change from 1 GPa tensile to a small compressive stress. The stress change can be explained by differing volume fractions of silicon within the resulting silicon nitride layer [19].

Despite the fact that the RF power is an additional parameter influencing the intrinsic stress of *PECVD-silicon nitride*, their stress range is less than that of *LPCVD* layers. By varying the deposition parameters for *PECVD-silicon nitride*, stress values between –300 MPa and +300 MPa can be obtained [17]. Similar to *PECVD-oxides*, subsequent thermal annealing pronouncedly shifts the stress towards more tensile values.

5.5.5

Polysilicon

5.5.5.1 Processing and General Function

The term ‘polysilicon’ arises from the structure of these silicon layers, which are essentially polycrystalline in contrast to single-crystalline silicon substrates, due to the growth of these films on amorphous starting layers (in silicon surface micromachining, a silicon oxide layer is usually used as both a seed layer and a sacrificial layer). Polysilicon is widely used in sensor technology: it can be used as part of a membrane layer, as an electrical connector, or as a part of a thermopile structure. In this contribution, we focus on its most important function – as the functional layer in surface-micromachined structures. In surface micromachining basically two approaches for producing polysilicon films are used:

- Deposition in a conventional *LPCVD* reactor by thermal decomposition of silane (SiH₄) at temperatures between 550 °C and 630 °C. The growth rate at this temperature is around 5 nm min^{–1}. The typical thickness of such thin films is about 2 μm, as used, for example, by Analog Devices for their surface-micromachined

accelerometer [20]. The film thickness is limited in particular by deposition of polysilicon on the walls of the furnace and the necessary maintenance runs to clean the walls. In this contribution, we refer to these films as *LPCVD-poly*.

- Deposition of a thin LPCVD polysilicon starting layer at around 590°C, followed by epitaxial growth of a thick polysilicon film. Epitaxy of silicon by means of thermal decomposition of silane (SiH_4) or silicon tetrachloride (SiCl_4) is commonly used for growing lightly doped, high-quality, single-crystalline silicon layers on heavily doped silicon substrates in the construction of bipolar devices. However, if the seed layer is polycrystalline, the epitaxial layer is also polycrystalline. The growth rate at a temperature of 1180°C is about 3000 nm min⁻¹. Using this approach, films with a thickness of 10 µm and more can be deposited, as used, for example, by Bosch for their surface-micromachined devices [11]. In this contribution, we refer to these films as *epipoly*.

Electrical conductivity, which is necessary for using polysilicon layers as capacitive elements, is assured by doping the polysilicon with dopants such as boron (B) for p-doped, or arsenic (As) or phosphorous (P) for n-doped layers. This doping can be performed both in-situ during the deposition process and ex-situ by diffusion or ion implantation. The appropriate doping technique has to be chosen according to the specific requirements of the application with respect to doping level and distribution. As shown below, the doping distribution can have a significant influence on the mechanical properties of polysilicon films.

5.5.5.2 Intrinsic Stress and Stress Gradients of Polysilicon Layers

The most significant mechanical properties for surface-micromachined structures are the intrinsic stress and the stress gradient in the functional layers. Of the two properties, the stress gradient is more important. Compressive stress might lead to buckling of sensor structures, but buckling can usually be prevented by designing appropriate compensation structures [11]; a high stress gradient, however, cannot be compensated for. Both properties are linked to the stress profile in the layers. The intrinsic stress is the average of the stress profile throughout the layer, the stress gradient reflects the shape of the stress profile.

For mass production as necessary for the automotive industry, it is advantageous to have stable processing conditions and to know exactly the key process parameters influencing the stress. In this respect, according to the literature, the two functional layers, LPCVD-poly and epipoly, differ as detailed in the following sections.

5.5.5.2.1 LPCVD-poly

Many publications describe the stress behavior of LPCVD-poly films. However, to our knowledge, no comprehensive and complete publication on the processing parameters that affect the stress profile within these layers is available.

The main influence on the stress in LPCVD-poly stems from their structure, which depends very much on the deposition temperature and the subsequent annealing cycle [21].

At processing temperatures below about 580 °C, the films are essentially amorphous after deposition; at a temperature of 630 °C or higher, the films are polycrystalline, exhibiting a columnar grain structure with a strong preferential {110} orientation. A transition regime is located in between. After deposition, the films are usually under compressive stress [22]. The structure of the polycrystalline LPCVD-poly (deposited at 630 °C) is very stable to annealing temperatures up to 1100 °C, and the influence on the stress is correspondingly low. Production of films with low compressive or tensile intrinsic stress is performed by optimizing the annealing procedures for LPCVD-poly initially having an amorphous or randomly oriented polycrystalline structure [23]. Doping of LPCVD-poly, especially with high concentrations of phosphorus, leads to large increases in the grain size, which strongly affects the stress behavior of LPCVD-poly films.

Information on parameters that influence stress gradients in LPCVD-poly is less readily available. However, one study showed that the grain growth that occurs during annealing of heavily doped films reduces the stress gradient in LPCVD-poly films [24].

The stability of the properties of polysilicon films to processing is not ideal. The structure and intrinsic stress of LPCVD-poly films created by different organizations but with the same deposition recipe were different [21]. The properties of LPCVD-poly can differ even on wafers from the same deposition run, due to small temperature fluctuations in the deposition furnace [25]. However, since several LPCVD-poly surface-micromachined sensors are in fact produced in high volumes for the automotive industry (e.g., by Analog Devices, Infineon, and Motorola), these literature data may not represent the current state of the art.

5.5.5.2.2 Epipoly

The dependence of stress and stress gradient in epipoly films on processing parameters has been well studied [10]. In contrast to LPCVD-poly, the grain structure of thicker films has virtually no impact on the stress behavior. The three major influences, in order of increasing importance, are: the layer thickness, the doping process, and the annealing scheme. The stress dependence is fully explained by a theoretical model based on the diffusion of oxygen atoms into and out of the epipoly layers during thermal processing [26]. The striking effect of oxygen on the mechanical properties of epipoly was illustrated by the following experiment.

Epipoly samples were post-annealed after standard processing, under alternating oxygen (1200 °C, 2 h) and nitrogen (1100 °C, 7 h) atmospheres, and the intrinsic stress and stress gradient were measured (Fig. 5.5.4). The final annealing atmosphere is important: oxygen results in highly compressively stressed layers having a high negative stress gradient, and nitrogen leads to layers having significantly smaller compressive stress and stress gradients. The effect of the annealing atmosphere is reversible, which can be explained by diffusion of oxygen atoms into and out of the epipoly. During annealing in an oxygen atmosphere, oxygen atoms diffuse into the epipoly layer, causing a marked increase in compressive stress close to the surface and, consequently, a shift of the stress gradient to

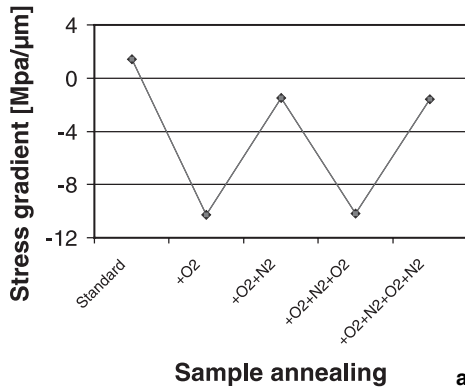
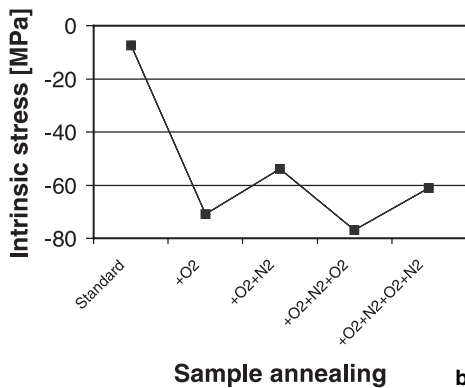


Fig. 5.5.4 Reversibility of **a** stress gradient and **b** stress in epipoly by alternately annealing in different atmospheres: nitrogen, 1100°C, 7 h; oxygen, 1200°C, 2 h



negative values. Upon annealing in a nitrogen atmosphere, the oxygen atoms diffuse out of the surface of the epipoly due to the decreased solid solubility of oxygen in inert atmospheres [27], thus reducing the compressive stress and stress gradients.

Epipoly is an ideal candidate for high-volume production. Its stress properties are virtually insensitive to many processing parameters, especially during the growth process, when the deposition temperature may be decreased from 1180°C to 1150°C. The obvious dependence of the properties on the annealing procedure enables the stress properties to be tailored to the specific needs of the application. Moreover, the annealing process can be closely monitored during production, assuring reproducible layers in mass production.

The surface-micromachined sensors produced by Bosch utilize the epipoly process. In-process data for the past 5 years clearly show the excellent stability and reproducibility of these films.

5.5.5.3 Fracture Strength of Polysilicon Layers

Silicon surface micromachining requires a material with a very high degree of elasticity and an absence of plastic deformation. Figure 5.5.5a shows a typical U-beam suspension structure with a beam width of $3\ \mu\text{m}$ and a folded beam length of $320\ \mu\text{m}$. A ring attached to the beam is subjected to a tensile load by pulling with a needle. The ring can be pulled more than $200\ \mu\text{m}$ without fracturing (Fig. 5.5.5a). When the ring is released, the structure returns to its starting position. No residual damage is observed even after repeated pull testing. Figure 5.5.5b shows a single-beam test structure subjected to a bending load. The bending force is so high that a neighboring beam is bent as well. Again, after release of the beam, the structure returns to its original position without any sign of damage.

The fracture strength of polysilicon layers can be measured by burst pressure testing of membranes or by using sophisticated surface-micromachined tensile test structures based on, for example, selective ion implantation [28], an electrostatic force grip [29], or a micromanipulator [14]. The fracture strength determined by the different techniques varies significantly with the different measurement principles. However, several conclusions can be drawn from published tensile testing experiments:

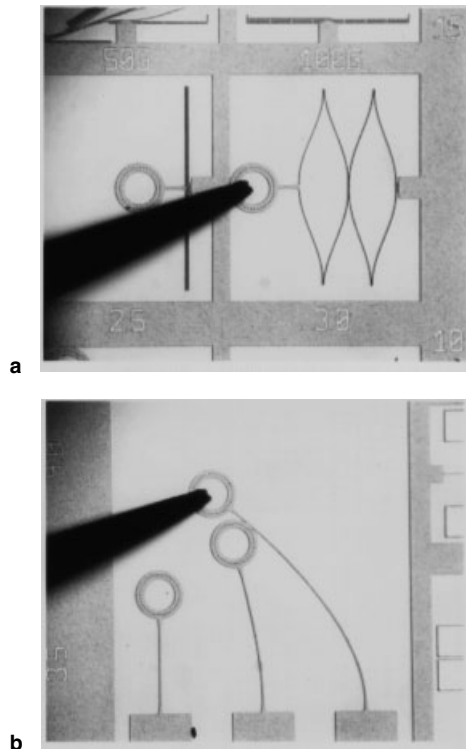


Fig. 5.5.5 Micrographs showing the high degree of elasticity of epipoly silicon structures with respect to **a** tensile and **b** bending stress. Epipoly structures were used for this experiment.

- The fracture behavior of polysilicon layers is governed by the sudden growth of a crack or cracks through the structure. These cracks are initiated at defects or imperfections in the surface.
- In general, the surface roughness is highest on the dry-etched sidewalls of structures, so this roughness is the decisive parameter.
- For LPCVD-poly, phosphorous doping leads to a higher surface roughness and hence decreased fracture strength [28].
- For epipoly, doping and stress level have no significant effect on the fracture strength [30].
- A study comparing epipoly and single-crystalline silicon SOI layers found no significant difference in fracture strength. Both layers were dry etched using the same process and therefore exhibited similar sidewall roughness, which suggests that grain boundaries in polysilicon layers do not constitute weak spots for fracture under tensile loads [11].

5.5.6

Other Materials

In addition to the thin films discussed above, several additional materials are used in silicon-based sensors for automotive applications. These include metals as well as other, more exotic, materials.

5.5.6.1 Aluminum

The dominant metal layer used in conventional silicon microelectronics technology is aluminum (Al), usually doped with silicon to reduce spiking and with copper to increase resistance to electromigration. Aluminum is used for electrical connections, for contacting the sensing elements, and as the standard material for bonding pads, which are needed to assure electrical connection to the 'outside world'.

Most commonly, aluminum layers between 0.5 μm and 2 μm thick are deposited by a sputtering process and then annealed at a temperature of about 400 °C to achieve the desired electrical properties.

In automotive sensors, aluminum bonding pads are the obvious choice: they can be used for bonding both aluminum and gold wires, and the bonding quality is sufficiently good for the lifetime of the device. However, for surface-micromachined structures, special care has to be taken regarding the bonding pads to prevent corrosion when the sacrificial oxide is etched.

In terms of mechanical stress, aluminum layers on silicon have to be handled carefully. Due to plastic deformation, aluminum can induce thermal stress in silicon devices [31], severely affecting the mechanical characteristics of structures as a function of temperature and causing hysteresis. Consequently, aluminum thin films are rarely used as part of a membrane.

5.5.6.2 Platinum

The metal platinum (Pt) is not used in modern microelectronic CMOS applications. As an impurity in silicon, platinum possesses electronic states located close to the middle of the bandgap. It exhibits a high diffusion coefficient and large capture cross sections for minority carriers. Therefore, the presence of platinum would severely change the CMOS transistor characteristics and is not allowed inside a silicon microelectronics fab.

For discrete micromechanical devices that do not include integrated electronics, the situation is different. Platinum is an excellent choice for sensors based on a thermal measurement principle and is used in very high volumes, for example, in the air mass flow sensor chip produced at Bosch. Here, a platinum thin film serves as both a heater and a temperature sensor on a thin dielectric membrane consisting of silicon oxide and silicon nitride. The advantages of platinum as a thin film for thermal sensors compared to, for example, polysilicon are as follows:

- High temperature coefficient of the resistivity (TCR), close to the bulk value of 3850 ppm K^{-1} .
- Almost linear dependence of TCR on temperature from room temperature to more than 400°C .
- Very high chemical stability and inertness.
- High specific electrical resistivity, ensuring high measurement resolution with small devices.
- Negligible plastic deformation in the working temperature range.

Thicker platinum layers are used extensively for temperature sensors, for example, deposited on ceramic substrates. Compared to other thin film metals, such as gold, silver, or nickel, platinum has a low thermal expansion coefficient, thus minimizing stress problems due to the differential thermal expansions of the various membrane layers.

Most commonly, platinum is sputter-deposited on the wafer and annealed at temperatures above 600°C to promote recrystallization and to adjust the TCR to the desired value.

In sensors for automotive applications, special care has to be taken to assure the adhesion of platinum thin films to adjacent layers over the lifetime of the devices.

5.5.6.3 Additional Materials

Another thin film material is *single-crystalline silicon*, which is used in, for example, Bosch's bulk-micromachined pressure sensors to assure sufficient membrane thickness (more than $10 \mu\text{m}$). These silicon films are usually produced by epitaxial processing of the silicon substrate with appropriate dopant concentrations and using KOH etching with a pn-etch stop. Another form of single-crystalline silicon layers is the so-called *SOI layer*, in which a single-crystalline silicon layer of variable thickness is attached on top of a silicon oxide layer. SOI wafers can be used to produce surface-micromachined devices by selectively removing the buried oxide film; however, due to the comparably high costs of these wafers, no high-volume applications for automotive sensors are on the market now.

Silicon oxynitrides (SiO_xN_y) are thin films that are basically a mixture of silicon oxide and silicon nitride, produced in a CVD-process by adding nitrous oxide (N_2O) to the gases used for silicon nitride deposition. By changing the oxide-to-nitride ratio, the properties of these films can be modified towards improved thermal and moisture stability and lower stress compared to pure silicon oxide or silicon nitride thin films [32]. At low oxygen concentrations ($\text{O}/(\text{O} + \text{N}) < 0.3$), oxynitride layers have good diffusion barrier characteristics [33] and oxidation resistance.

Besides platinum, several other metal thin films can be used to produce thermal sensors. Examples include *gold* (Au) and *nickel* (Ni).

Additional novel materials, which may find their way into commercial products include *silicon carbide* (SiC), *diamond* (C), *germanium* (Ge), and *silicon germanium* (SiGe). However, no high-volume automotive applications of these materials have been reported so far.

SiC is an excellent material suitable for high temperature, high frequency, and high power electronics [34, 35]. Therefore, it is well suited for applications close to engines. Besides, because of its chemical inertness to many media even at elevated temperatures, SiC is a well-known passivation layer. Because of the high price of monocrystalline SiC substrates, only CVD-deposited layers are attractive for implementation in high-volume applications. Until now, the low deposition and etching rates of SiC and the lack of appropriate production equipment have been the most important handicaps in integrating SiC layers in high-volume MEMS applications.

PECVD-deposited amorphous diamond layers (on silicon) are attractive, due to their mechanical strength and outstanding chemical stability [36, 37]. Until now, however, silicon structures meet the mechanical requirements described in Section 5.5.5, and diamond is not needed in automotive applications now.

Ge and SiGe are under discussion for future automotive sensor applications: Ge can be selectively etched to Si and thus may be used as a sacrificial layer [38]. The deposition temperature needed for CVD-SiGe is lower than that for silicon, so SiGe layers can be deposited even on top of metallized wafers. SiGe is CMOS-compatible [39, 40]. Sensor structures integrated in or located on top of a CMOS wafer are feasible because of the processing temperature and compatible chemistry. SiGe/Ge is a very promising material for future automotive sensors. However, as with SiC and diamond layers, the necessity still has to be proven.

5.5.7

Conclusions

In this contribution, we discussed the materials used as thin films on silicon for sensors in automotive technology. Most of these materials are also used in silicon microelectronics; however, due to the different applications, other material properties are relevant for micromechanics. Moreover, some special CMOS-incompatible metals such as platinum and gold are used in, for example, thermal sensors, due to the excellent thermal behavior of their electrical characteristics.

The mechanical properties of thin films used as membranes or moving structures and produced either by bulk or surface silicon micromachining are of crucial importance. In either type of application, the mechanical stress and the fracture properties of the applied thin films determine the behavior and long-term reliability of the device. In automotive applications, the environmental constraints (e.g., high ambient temperatures and temperature variations) and the required long life expectancy of the sensors require the use of high-quality thin films with well-controlled properties.

In the previous sections, we summarized the current understanding of mechanical stresses in several individual thin films. One more significant fact has to be mentioned: the stress behavior of thin film multilayer stacks is not simply the sum of the stresses of the individual layers. Instead, stacked thin films can pronouncedly affect each other's properties [41]. Consequently, this influence has to be considered when selecting materials for an application.

Bulk-micromachined membranes are usually formed from dielectric materials like silicon oxide or silicon nitride combined with additional materials; for example, in pressure sensors, silicon is used to increase the membrane thickness to the required values; and in thermal sensors, platinum or other metals are needed for the sensing elements. The overall stress state of the membranes has to be controlled well to prevent buckling (under high compressive stress) or fracture (under high tensile stress). With proper processing control, silicon oxide and silicon nitride thin films meet this requirement, making them ideal candidates for membrane-type devices.

Surface-micromachined moving structures are usually made up of polysilicon layers and are basically produced by two different techniques, to produce thinner or thicker films. In the meantime, sensors based on polysilicon are widely used and represent state-of-the-art technologies. The fracture strength of polysilicon films is generally sufficient for the loads that appear in an automotive environment.

Silicon micromachining is a widely accepted technology for sensors for automotive applications today, and chances are very good that it will remain so in the future.

5.5.8

Acknowledgements

We would like to thank our colleagues at Robert Bosch GmbH Sensor Development, especially, Dipl.-Ing. Doris Schielein and Dr. Christoph Schelling for their valuable comments and suggestions.

5.5.9

References

- 1 K. E. PETERSEN, *Proc. IEEE* **1982**, *70*, 420–457.
- 2 M. OFFENBERG, F. LÄRMER, B. ELSNER, H. MÜNZEL, W. RIETHMÜLLER, *Proc. 8th Int. Conf. Solid-state Sensors and Actuators (Transducers '95 – Eurosensors IX)*, Stockholm, Sweden, pp. 589–592.
- 3 A. KOVACS, A. STOFFEL, *J. Micromech. Microeng.* **1992**, *2*, 190–192.
- 4 M. OFFENBERG, B. ELSNER, F. LÄRMER, *Extended Abstracts Electrochem. Soc. Fall Meeting*, Miami Beach, USA, **1994**, *94-2*, 1056–1057.
- 5 S. WOLF, R. TAUBER, *Silicon Processing for the VLSI Era, Vol. 1 Process Technology*, 2nd ed., Lattice Press, Sunset Beach, CA, USA, **2000**.
- 6 S. M. SZE, *Semiconductor Devices, Physics and Technology*, 2nd Ed., John Wiley & Sons, New York, NY, USA, **2001**.
- 7 G. G. STONEY, *Proc. R. Soc., London*, **1909**, *A82*, 172–175.
- 8 F. ERICSON, S. GREEK, J. SÖDERKVIST, J. Å. SCHWEITZ, *Proc. 8th Int. Conf. Solid-state Sensors and Actuators (Transducers '95 – Eurosensors IX)*, Stockholm, Sweden, **1995**, pp. 84–87.
- 9 H. GUCKEL, T. RANDAZZO, D. W. BURNS, *J. Appl. Phys.* **1985**, *57*, 1671–1675.
- 10 M. FÜRTSCH, M. OFFENBERG, H. MÜNZEL, J. R. MORANTE, *Proc. SPIE Micromachining and Microfabrication Process Technology III*, **1997**, 130–141.
- 11 M. FÜRTSCH, *Dissertation*, Shaker Verlag, Aachen, **1999**.
- 12 W. FANG, J. A. WICKERT, *J. Micromech. Microeng.* **1996**, *6*, 301–309.
- 13 J. Å. SCHWEITZ, *MRS Bull.* **1992**, *XVII*, 34–45.
- 14 S. JOHANSSON, *Proc. 1st IARP Workshop on Microrobotics and Systems*, Karlsruhe, Germany, **1993**, pp. 72–83.
- 15 F. H. P. M. HABRAKEN, A. E. T. KUIPER, *Mater. Sci. Eng.* **1994**, *R12*, 123–175.
- 16 J. G. E. GARDENIERS, H. A. C. TILMANS, C. C. VISSER, *J. Vac. Sci. Technol.* **1996**, *A14*, 2879–2892.
- 17 T. J. COTLER, J. CHAPPLE-SOKOL, *J. Electrochem. Soc.* **1993**, *140*, 2071–2075.
- 18 G. M. WELLS, M. REILLY, R. NACHMAN, F. CERRINA, M. A. EL KHAKANI, M. CHAKER, *MRS Symposium Proc.* **1993**, *306*, 81–89.
- 19 S. HABERMEHL, *J. Appl. Phys.* **1998**, *8*, 4672–4677.
- 20 T. A. CORE, W. K. TSANG, S. J. SHERMAN, *Solid State Technol.* **1993**, *10*, 39–47.
- 21 T. I. KAMINS, *Sens. Actuators* **1990**, *A21-23*, 817–824.
- 22 J. ADAMCZEWSKA, T. BUDZYNSKI, *Thin Solid Films* **1984**, *113*, 271–285.
- 23 H. GUCKEL, W. BURNS, C. C. G. VISSER, H. A. C. TILMANS, D. DEROO, *IEEE Trans. Electron Devices* **1988**, *35*, 800–801.
- 24 T. A. LOBER, J. H. HUANG, M. A. SCHMIDT, S. D. SENTURIA, *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA **1988**, pp. 92–95.
- 25 P. KRULEVITCH, R. T. HOWE, G. C. JOHNSON, J. HUANG, *Dig. Tech. Papers, Transducers '91*, San Francisco, USA, **1991**, pp. 949–952.
- 26 M. FÜRTSCH, M. OFFENBERG, H. MÜNZEL, J. R. MORANTE, *Proc. Eurosensors XII*, Southampton, England, **1998**, pp. 59–62.
- 27 J. C. MIKKELSEN JR., *Proc. MRS Fall Meeting*, Boston, MA, USA, **1985**, *59*, 19–30.
- 28 M. BIEBL, H. VON PHILIPSBORN, *Proc. 8th Int. Conf. Solid-state Sensors and Actuators (Transducers '95 – Eurosensors IX)*, Stockholm, Sweden, **1995**, pp. 72–75.
- 29 T. TSUCHIYA, O. TABATA, J. SAKATA, Y. TAGA, *Trans. IEEE Sensors and Micro-machines Soc.*, **1996**, *116-E(10)*, 441–446.
- 30 S. GREEK, F. ERICSON, S. JOHANSSON, M. FÜRTSCH, A. RUMP, *J. Micromech. Microeng.* **1999**, *9*, 245–251.
- 31 D. S. GARDNER, *IEEE Trans. Electron Devices*, **1988**, *35*, 2160–2169.
- 32 K. TAKASAKI et al., *Electrochem. Soc. Ext. Abs.*, **1980**, *80-2*, 260.
- 33 A. HASHIMOTO, M. KOBAYASHI, T. KAMIJOH, H. TAKANO, M. SAKUTA, *J. Electrochem. Soc.*, **1986**, *133*, 1466.
- 34 M. MEHREGANY, C. A. ZORMAN, N. RAJAN, G. H. WU, *Proc. IEEE* **1998**, *86*, 1594–1609.
- 35 P. G. NEUDECK, *SiC Technology, in VLSI Handbook, The Electrical Engineering Handbook Series* (W. K. CHEN, ed.), CRC

- Press and IEEE Press, Boca Raton, FL, USA, 1998, pp. 6.1–6.24.
- 36 J. WANG, J. E. BUTLER, D. S. Y. HSU, C. T.-C. NGUYEN, *Proc. MEMS 2002*, Las Vegas, NV, USA, 2002, pp. 657–660.
- 37 E. KOHN, W. EBERT, M. ADAMSCHIK, P. SCHMID, A. DENISENKO, *New Diamond and Frontier Carbon Technology* 2001, 11(2), 81–100.
- 38 B. LI, B. XIONG, L. JIANG, Y. ZOHAR, M. WONG, *J. Microelectromechanical Systems*, 1999, 8, 366–372.
- 39 A. E. FRANKE, Y. JIAO, M. T. WU, T.-J. KING, R. T. HOWE, Hilton Head 2000, 2000, pp. 18–21.
- 40 S. SEDKY, A. WITVROUW, H. BENDER, K. BAERT, *IEEE Trans. Electron Devices* 2001, 48, 377–385.
- 41 C. M. A. ASHRUF, P. J. FRENCH, C. DE BOER, P. M. SARRO, *Proc. SPIE Micro-machining and Microfabrication Process Technology III*, 1997, pp. 149–159.

5.6

Ceramic Materials and Technologies for Exhaust Gas Sensors

HARALD NEUMANN, HANS-MARTIN WIEDENMANN, JOHANN RIEGEL

5.6.1

Introduction

This section gives an overview of the materials and technologies used in current and future exhaust-gas sensors. Because of the dominant role of ZrO₂-based oxygen sensors, this sensor type is used as the main example for introducing the various material and technology aspects in more detail. However, we want to cover briefly most other sensor types and alternative materials as well.

5.6.2

Overview of Exhaust-Gas Sensors

To lead into this broad subject of what materials and technologies are suitable for exhaust-gas sensors, two main questions should be considered first: what is valuable to be measured or what sensor type is needed in an exhaust gas and what are the environmental conditions?

For almost 20 years the focus was on air/fuel ratio sensors, also known as oxygen sensors or lambda sensors [1]. They are standard in almost all gasoline engines and are now becoming standard in diesel engines [2]. In addition, the exhaust-gas temperature should be of interest. And the upcoming NO_x reduction catalyst and current activities in diesel technologies have intensified investigations of specific NO_x, HC, and NH₃ sensors. Exhaust-gas temperatures up to 1100°C combined with mechanical vibration at levels of 130 g limit the choice of sensor materials to ceramics. Ceramics are considered robust enough to last for 15 years and 150 000 miles without any deterioration being a requirement included in the latest California automotive emission standards.

Beyond the technical requirements, cost is also a major driver in the continuous search for low cost materials, technologies, and miniaturization.

A ceramic sensor element is typically composed of a substrate, electrodes, catalytic layers, and functional layers, based on various physical and chemical effects. In some cases, for example, ZrO_2 , the substrate and functional materials are identical.

With multilayer technology being used in exhaust-gas sensors, the complexity is increased by using the third dimension.

5.6.3

Material Classes used in Exhaust-Gas Sensors

5.6.3.1 Ion Conductors

5.6.3.1.1 Zirconia

The best known principle in the field of exhaust-gas sensors is the ZrO_2 -based oxygen sensor (lambda sensor) used to determine the air/fuel ratio (see also Section 7.15). Walther Nernst derived the measuring principle equation that bears his name for galvanic cells in 1889 (Figs. 5.6.1 and 5.6.2). He also showed that the ionic conductivity of ZrO_2 could be significantly improved when the zirconium ion is partially replaced by bivalent calcium or trivalent yttrium, creating additional oxygen vacancies in the crystalline lattice.

The electrode reaction:

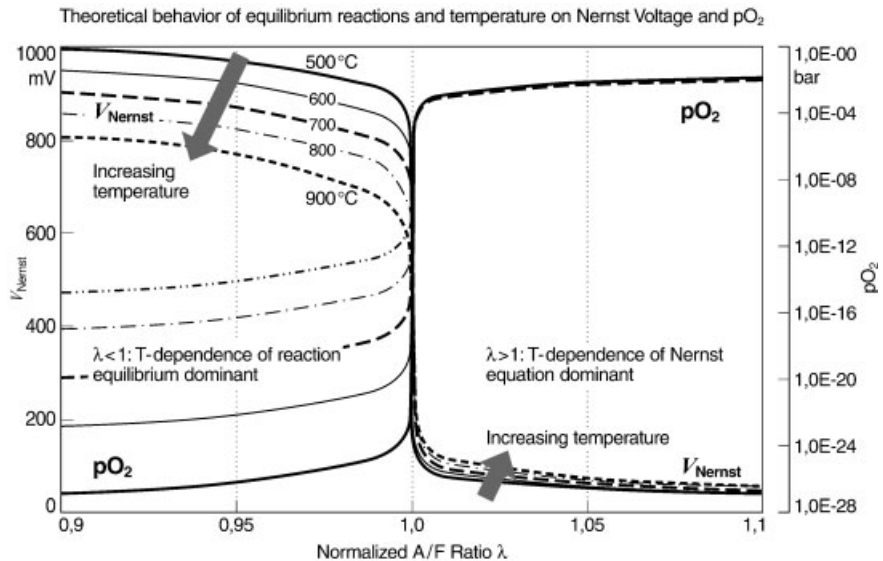
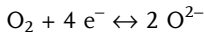


Fig. 5.6.1 Calculated Nernst voltage and pO_2 vs. air/fuel ratio

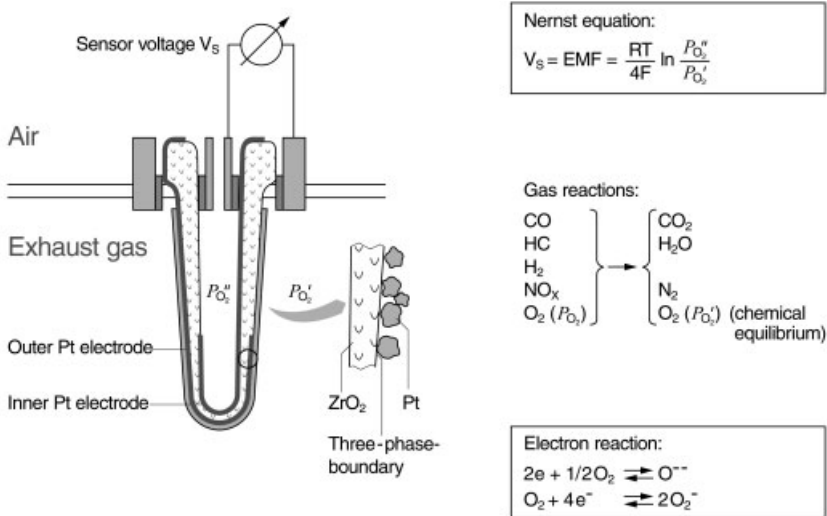


Fig. 5.6.2 ZrO_2 -based oxygen sensor principle and three-phase boundary between zirconia, platinum and exhaust gas (triple points)

on both the exhaust ($'$) and reference ($''$) sides generates an electromotive force obeying the following law:

$$V_S = \text{EMF} = \frac{RT}{4F} \ln \frac{pO_2''}{pO_2'}$$

where R is the universal gas constant; T is the absolute temperature; F is the Faraday's constant; pO_2'' is partial pressure of oxygen at the reference electrode; and pO_2' is the partial pressure of oxygen at the measuring electrode. To measure the correct pO_2' first chemical equilibrium of the raw exhaust gas must be adjusted at the outer electrode by completing the indicated reactions (Fig. 5.6.2). As a result, Fig. 5.6.1 shows the Nernst voltage where above the stoichiometric ratio the pre-factory of the equation and below the stoichiometric ratio the kinetics of the chemistry determine the temperature dependence of the signal.

Partially stabilized zirconia containing around 4 mol% of yttria (≈ 8 wt%) provides good ionic conductivity combined with sufficient mechanical stability (Fig. 5.6.3). The ZrO_2 phase composition varies, depending on the Y_2O_3 concentration [3]. Although pure ZrO_2 is monoclinic at room temperature, it becomes tetragonal at 1200°C and cubic above 2370°C . The tetragonal and cubic phases can be partially stabilized (PSZ, partially stabilized ZrO_2 ; TSZ, tetragonal stabilized zirconia), resulting in less phase transformation and improved durability in high-temperature operation [4].

At higher Y levels (e.g., 9–10 mol%) the material becomes completely cubic (FSZ, fully stabilized zirconia), which gives the highest ionic conductivity but weak mechanical properties.

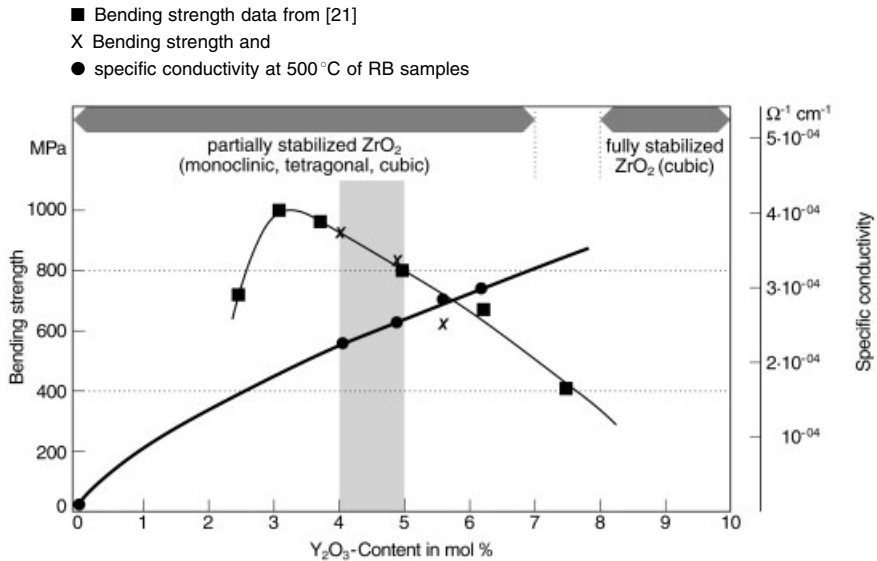


Fig. 5.6.3 Ionic conductivity and bending strength of ZrO_2 as a function of Y_2O_3 content

Tab. 5.6.1 Phase and material composition of commercial ZrO_2 -based oxygen sensors

	Type A	Type B planar	Type C	Type D	Type E
% monoclinic ¹⁾	25	0	1	20	23
% tetragonal ¹⁾	<10	60	55	<10	<10
% cubic ¹⁾	main	40	46	main	main
wt% Y_2O_3 ²⁾	8.8	8	7	10	10
wt% Al_2O_3 ²⁾	2.5	0.9	8.5	not determined	2.5
wt% SiO_2 ²⁾	0.5	1.1	0.2	not determined	0.1

1) x-ray diffraction; 2) chemical analysis

Another way to control the mechanical properties is to reinforce ZrO_2 with small amounts of Al_2O_3 or silicates. Typical compositions of various ZrO_2 -electrolyte materials are shown in Table 5.6.1.

5.6.3.1.2 Nasion

Electrochemical type-III sensors have been constructed for measuring NO_x , CO_2 , SO_2 , and many other environmentally harmful gases in very low concentration ranges [5]. In these multilayer type-III gas sensors, no direct relation between a solid electrolyte and the target gas exists without the mediation of an attached *auxiliary phase*.

A sodium-ion conductor (Nasion) is used as the solid electrolyte. Various oxides have been proposed for the auxiliary phase, depending on the target test gas, includ-

ing NaNO_3 , $\text{Li}_2\text{SO}_4 \cdot \text{CaSO}_4$, and BaCO_3 . For automotive use, NO_x sensors based on Nasicon with a rare earth perovskite auxiliary phase look promising [6, 7].

5.6.3.2 Semiconducting Ceramics

Instead of ion conductivity, another class of oxides providing electronic conductivity can also be used for gas detection. Promoted by various dopants, the adsorption of the measured gas components or the partial pressure of oxygen modifies the defect chemistry in the material, leading to a variation of bulk or surface conductivity.

5.6.3.2.1 Bulk Conductors

The best known ceramic is TiO_2 which is used in oxygen sensors of this conductor type for various vehicle applications [8]. The operating principle is based on changes in the bulk conductivity due to the concentration of oxygen vacancies, indicated by the formulation TiO_{2-x} .

At high ceramic temperatures the oxygen vacancy equilibrium (V_{O}) of the TiO_2 lattice responds quickly to the change of the partial pressure of oxygen p_{O_2} in the exhaust gas (Fig. 5.6.4).

Because the oxygen ion is negatively charged, the ionized-oxygen vacancies act as donors and the electrical conductivity obeys following equation:

$$\sigma_T = A \cdot p_{\text{O}_2}^m \exp(-E/kT)$$

where A is a constant; m is a constant that depends on the type of bulk defect; E is the activation energy; k is the Boltzmann constant; and T is the absolute temperature.

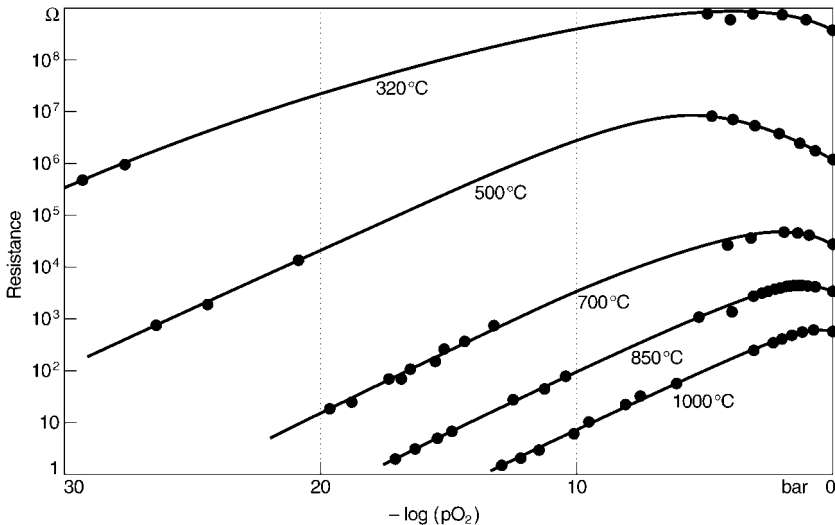


Fig. 5.6.4 Dependence of the electrical resistance of TiO_2 on temperature and p_{O_2} [22]

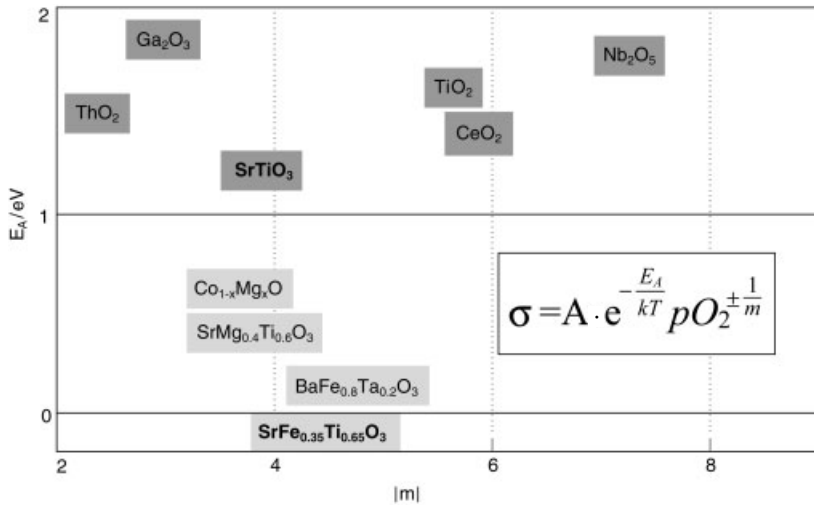


Fig. 5.6.5 Properties of semi-conducting oxide materials [23]

For good response time and accuracy, the sensitive material should be porous and as thin as possible, and the ceramic temperature needs to be in the range of 600–1000 °C. These requirements are met by applying the semiconducting ceramics as porous thick-film layers on a heated Al₂O₃ substrate.

Another promising material is Ga₂O₃. Recent investigations on 1-μm-thick sputtered layers showed sufficient thermal and chemical stability for automotive applications [9]. At temperatures above 800 °C Ga₂O₃ is a potential O₂ sensor, and at lower temperatures it also shows good sensitivity to other reducing gas species such as H₂, HC, and CO [10, 11]. The exploitation of these properties is very dependent on precise control of the ceramic temperature, which needs to be kept within 20 K.

Additional known examples are various oxides having a perovskite structure, for example, SrTiO₃ [12] (Fig. 5.6.5). Although the defect chemistry of the perovskites is very well studied, neither a Ga₂O₃-based nor a SrTiO₃-based commercial application is known.

5.6.3.2.2 Surface Conductors

A representative surface-conducting material is SnO₂ which is already used for detecting CO [13]. The working principle is to increase the dwelling time of a CO or HC molecule over that of a competing species on the surface of the SnO₂ microstructure. This leads to local reduction and formation of a thin conducting SnO_{2-x} layer around the individual grains, initiating an electron-hopping process. The resulting electrical resistance is a measure of the CO or HC concentration (Fig. 5.6.6). Because this sensor shows a permanently high signal in a reducing atmosphere, it is necessary to provide a lean base gas; otherwise the sensitive material cannot distinguish between a rich gas composition and a high CO concentration.

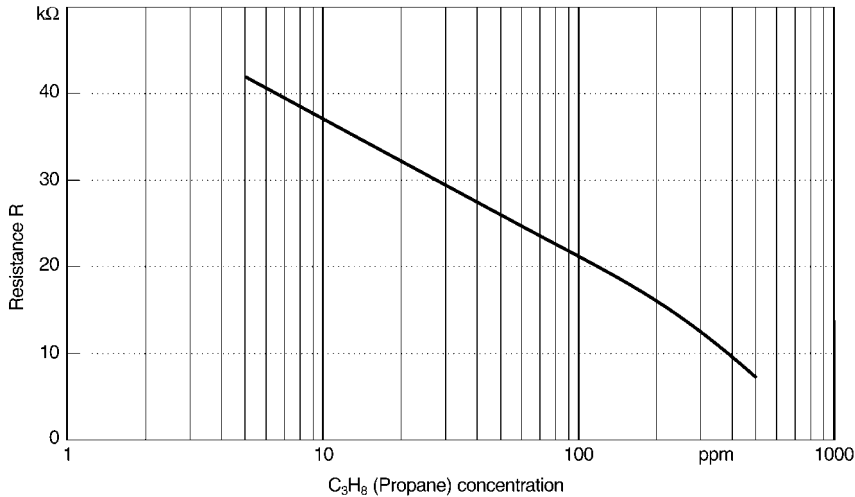


Fig. 5.6.6 Electrical resistance of SnO₂ as a function of propane concentration

5.6.3.3 Electrodes and Other Metal and Ceramic Catalysts

Electrodes and catalytic materials in front of the electrode (pre-catalyst) play an important role in the design of specific functions or characteristics of exhaust gas sensors. For excellent O₂ detection in nonequilibrium raw emissions, platinum has become established as the dominant electrode material, because it efficiently supports the ionic-atomic transfer reaction of oxygen. Additionally, it ensures that the raw gas quickly achieves thermodynamic equilibrium, to enable an accurate lambda measurement. Sometimes it can be useful to add additional metals such as Rh, Pd, or Au to improve or prevent NO_x reduction or to add ceramic additives such as zeolites or CeO₂ to adsorb oxygen or other species [14–16].

5.6.3.4 Structural Materials and Substrates

The exhaust gas sensor business is dominated by two substrate materials: ZrO₂, with its excellent thermal and mechanical properties, and Al₂O₃. Although the thermal and mechanical behavior of Al₂O₃ is slightly inferior to that of zirconia

Tab. 5.6.2 Comparison of different substrate material properties

	Y-ZrO ₂	Al ₂ O ₃
Strength [MPa]	typical 700	300–400
Fracture toughness K _{1c} (MPa m ^{1/2})	4–7	3–3.5
Thermal expansion coefficient (10 ⁻⁶ K ⁻¹)	9–10	7–8
Young's modulus (GPa)	220–240	350
Heat conduction (W/mK)	5	30
Phase transformation	yes ¹⁾	no

1) toughening and crack avoiding due to the tetragonal to monoclinic phase transformation

(Table 5.6.2), the great advantage of alumina is its perfect electrical insulation up to very high temperatures, as long as the levels of alkaline impurities can be kept low.

To further optimize properties, the two materials are sometimes mixed or joined together.

5.6.4

Ceramic Technologies and Manufacturing Methods for Exhaust-Gas Sensors

5.6.4.1 Processing of Yttria-Stabilized Zirconia Powder

The ceramic industry has developed various highly sophisticated manufacturing methods to meet the material requirements of homogeneous microstructure and phase distribution of the sintered ceramic. These are the key factors for predictable mechanical and functional properties.

The classical mixed oxide method is still a standard way to obtain yttria-stabilized zirconia. Fine-grained chemically processed ZrO_2 and Y_2O_3 powders plus flux additives such as silica and alumina are mixed homogeneously in a sequential dry- and wet-milling processes. During sintering, the yttrium ion diffuses into the zirconia lattice. Because grain growth is affected by the local yttria concentration, grain size distribution is broad. Larger grains tend to have lower yttria stabilization and become monoclinic at room temperature, and smaller grains tend to have higher yttria stabilization and remain cubic or tetragonal. This phase mix is not necessarily a disadvantage, but experience is required to control it properly. For a more homogeneous distribution and better phase control, for example, tetragonal stabilized zirconia, a prereacted yttria and zirconia powder should be used. A well known prereaction method is to coprecipitate, calcine, and re-mill the yttria/zirconia material to a submicron powder. Various plasma fusion processes can be used to establish the same homogeneous yttria distribution by melting the ceramic powders simultaneously. Extensive milling is needed after recrystallization to turn the large crystals back into a submicron powder.

5.6.4.2 Standard Manufacturing Route for Thimble-Type Sensors

5.6.4.2.1 Shaping Thimble-Type Sensors

Sensor elements in thimble-type O_2 sensors represent about 75% of the oxygen sensor market today. These elements are mainly shaped by dry pressing and grinding of granulated ZrO_2 .

The elements can also be made by casting aqueous slurries. These two methods require prefiring at low temperatures so that they can be handled in an automated large scale process. Elements can also be injection molded similar to plastic parts, in which the ceramic powder is mixed into an organic matrix. This third method leads to a very robust green (unfired) stone that can be easily processed further without prefiring.

5.6.4.2.2 Application of Electrodes, Layers, and Coatings

To build up functional or protective layers, various methods from other technical industries have been adapted to the specific needs of ceramic sensor materials. Important criteria for the selection of the various technologies are the shape of the ceramic substrate (e.g., thimble or planar design), if it is co- or post-fired, if the layer material is expensive (e.g., noble metals) or inexpensive (e.g., alumina, magnesia spinel), and if the geometry and thickness need to be controlled precisely.

5.6.4.2.3 Layers in Thimble-Type Sensors

Two major classes of technology are used for electrode formation.

In *thin-film technology* (layer thickness $< 1 \mu\text{m}$), a microporous platinum layer is deposited on the already fired ceramic by thermal evaporation, sputtering, chemical vapor deposition, or electrolytic or electroless deposition. The microporosity of the thin electrode provides sufficient access of the exhaust gas to the three-phase boundary.

In *thick-film technology* (layer thickness $> 1 \mu\text{m}$), roll or tampon printing methods well known in the porcelain industry are suitable for applying layers of defined geometry on the green ceramic. The thickness is controlled by precise adjustment of the paste viscosity and solid content. To provide an adequate three-phase boundary, zirconia is added to the platinum to form a cermet (ceramic-metal compound).

For thicker ($100 \mu\text{m}$) functional or protective layers, various dipping or spraying techniques are used. An important method is plasma spraying, in which the ceramic powder (e.g., magnesium spinel) is fed into a 20000°C argon plasma and is sprayed as liquid droplets on the fired ceramic. By controlling plasma energy, spraying parameters, and base stone temperature, the droplets recrystallize at the

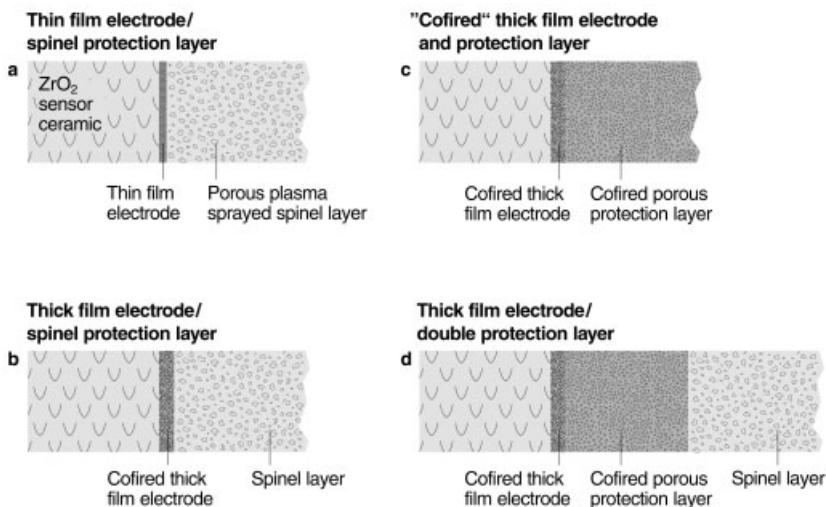


Fig. 5.6.7 Designs of an electrode/protection layer system [24]

surface and form a defined porous structure providing good adhesion to the substrate. The various design options based on co-firing and post-firing or a combination of both are shown in Fig. 5.6.7.

5.6.4.3 The Multilayer Processing of Planar Sensors

5.6.4.3.1 Ceramic Tape Casting

For planar sensors manufactured by multilayer technology, the recently developed ceramic tape casting process was improved further [17, 18]. ZrO_2 powder is mixed into an organic matrix of binders, plasticizers, and solvents and is tape cast onto a Mylar supporting tape or a steel belt. After drying, the ceramic tape is separated from the support tape and the flexible green tape can be punched to form individual substrates.

5.6.4.3.2 Application of Layers on Planar Sensors

Because planar sensors can have more complex monolithic structures, new techniques have been developed and some of the well established thimble techniques have been replaced. Today, modern planar sensors are completely co-fired. All electrodes, as well as functional and protective layers, are applied prior to sintering. Borrowed from the electronic industry (where it is used to create circuit boards and hybrid circuits) the screen printing process is ideal for applying layers of precise thickness (dried, typically 10–20 μm thick) and geometry. The planar substrates and the simultaneous printing of many sensors at the same time result in a very cost-effective production method. Continual improvements (e.g., automated visual alignment, new screens) allow for fine line structures in the 100 μm range. If higher resolution is required, this process can be combined with photolithographic techniques known from the semiconductor industry. A scheme of the whole process flow is shown in Fig. 5.6.8.

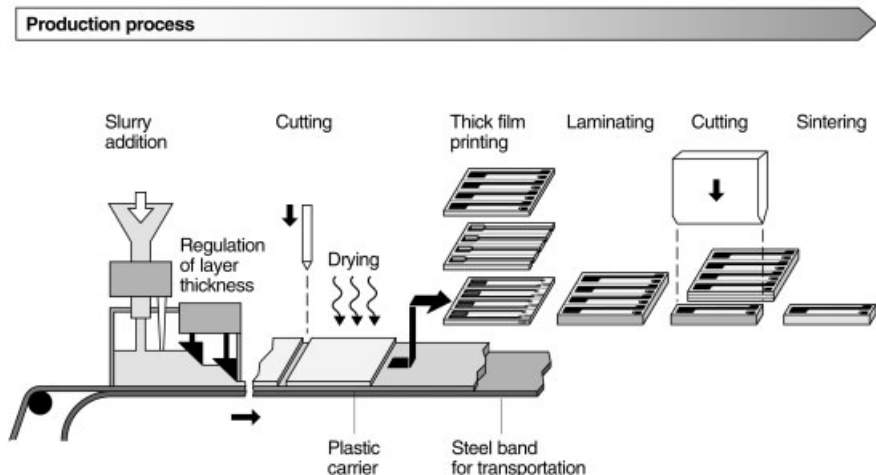


Fig. 5.6.8 Manufacturing flow for creating planar ZrO_2 -based oxygen sensors [24]

5.6.4.3.3 Special Ceramic Multilayer Techniques

To complete a multilayer planar sensor, various additional processing steps are necessary. Connections between the different planes or sheets of the sensor are made by punched or drilled holes that are filled or wall coated with a conducting material. Punching or stamping can be used to create buried cavities or registration holes for alignment. Lamination under pressure at a temperature of about 80°C fuses the different tapes together. Standard cutting techniques with ceramic or tungsten carbide blades are used to separate the laminate into individual pieces. In a hybrid process, the pieces are separated after firing, along scratched lines. It is important to handle the unsintered flexible pieces carefully, to prevent defects such as warping or delamination. If necessary, after sintering sharp edges can be ground to prevent additional thermomechanical stresses upon rapid heating during operation.

5.6.4.4 Printed Heaters in Exhaust-Gas Sensors

Electrical heating of the sensor plays an important role in exhaust-gas sensors because almost all sensor principles work only at high temperatures, and the efficiency of the heater determines the time from engine start to control readiness (light-off time). Two types of heaters are used in oxygen sensors.

5.6.4.4.1 Rod-Shaped Heaters for Thimble-Type Sensors

The manufacturing process for fabricating rod-shaped heaters is similar to those used in planar technology, because the active heater pattern is typically printed on an alumina tape that is wound around a pre-fired alumina rod or tube. The alumina allows sintering at low O₂ pressure and therefore inexpensive tungsten is the typical heater material. However, leaks and the residual porosity of the materials lead to a continuous increase in electrical resistance, caused by slow tungsten oxidation.

5.6.4.4.2 Planar Heaters for Co-fired Zirconia-Based Sensors

Platinum is the standard heater material for zirconia-based sensors because an oxidizing atmosphere is mandatory for preventing zirconia decomposition. The material cost disadvantage is balanced by the low amount of platinum that is needed for these low-mass sensor elements and the stability of the heater throughout its lifetime.

To adjust heater resistance, the heater composition is typically modified by adding an insulating ceramic phase (e.g., alumina) to the platinum powder. The more insulating phase is mixed into the platinum the higher the resistance. This cermet technology leads to the added benefit of improved stability, because the ceramic phase prevents grain growth or formation of platinum droplets at high temperatures (Fig. 5.6.9).

Investigations show that a 5 vol% insulating phase significantly improves heater durability. However, good control of grain size distributions and paste preparation allows even an excess of insulating phase to form a lead which can carry currents of several amperes, according to percolation theory. But even with this knowledge,

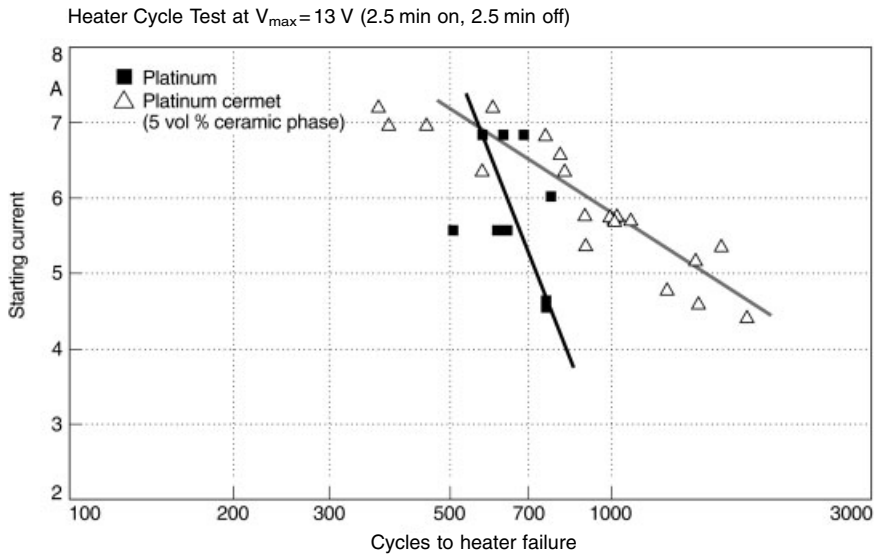


Fig. 5.6.9 Effect of ceramic phase in heater material on heater failure behavior

it is still a challenge to meet the very narrow heater requirements, because geometry and sintering shrinkage also affect the resistance.

5.6.4.5 Firing Techniques

ZrO_2 requires sintering temperatures about 1400°C and Al_2O_3 even higher, up to 1600°C . If catalytic layers such as electrodes are involved, the sintering temperature is kept as low as possible to maintain optimized catalytic capabilities. Precisely matched temperature profiles (which depend on the organic formulation) allow all organic constituents to burn off under a defined oxidizing atmosphere prior to densification and without any damage to the internal structure or delamination. A typical sintering shrinkage of 15–20% requires good process and material control at all levels to keep the final sensor element geometry within allowed tolerances.

5.6.4.6 Activation and Conditioning for Specific Sensor Characteristics

A very important aspect of gas sensors in automotive exhaust-gas environments is aftertreatment of the electrodes to control a specific sensor behavior. For example, to measure nonequilibrium raw emissions, the sensor needs excellent catalytic ability. Various methods are known to improve electrodes in ZrO_2 -based sensors. One well known method is to increase the active platinum surface area and the three-phase boundary area by partial reduction of zirconia close to the electrode. This occurs when the ceramic is exposed to a reducing atmosphere at high temperatures or when an electrical cathodic current is applied through the electrode and electrolyte. A similar effect can be achieved by chemical etching of the elec-

trode through the porous protective layer. Another method involves incorporating a pre-catalyst in front of the measuring electrode before or after firing. Very fine particles of Pt, Rh, or Pd can be deposited inside the pores of the protective layer. If there is a need to inhibit catalytic reactions, for sensitivity to NO_x or HC, the same methods can be used with different materials, for example deposited Au.

Another requirement is robustness of the sensor characteristics to various poisons that come from the engine, such as phosphorus, silicon, lead and sulfur originating from the fuel; from oil additives; and from seals and coatings burned off during driving [19, 20]. Even if the electrodes have some degree robustness, it can be improved by introducing various getter materials, which can be deposited on top of the layer or inside the pores binding the poisoning materials prior to reading the electrode.

5.6.5

References

- 1 H.-M. WIEDENMANN, G. HOETZEL, H. NEUMANN, J. RIEGEL, F. STANGLMEIER, H. WEYL, *Exhaust Gas Sensors in Automotive Electronics Handbook, chap.6, 2nd ed* (R. K. JURGEN, ed.) McGraw-Hill, New York, NY, USA 1999.
- 2 M. BIRK, G. ENGEL, R. LEONHARD, *Symposium Steuerungssysteme für den Antriebsstrang von Kraftfahrzeugen*, Berlin, Germany, 1997.
- 3 H. TSUBAKINO, M. HAMAMOTO, R. NOZATO, *J. Mater. Sci.* **1991**, 26, 5521–5526.
- 4 S.-Y. LIU, I.-W. CHEN, *J. Am. Ceram. Soc.* **1991**, 47, 1197–1216.
- 5 N. YAMAZOE, N. MIURA, *MRS Bull.*, June 1999, 37–43.
- 6 T. KIDA, K. SHIMANOE, N. YAMAZOE, *Sens. Actuators*, Chemical vol. 75, B3, **2001**, 179–187.
- 7 N. MIURA, S. YAO, Y. SHIMIZU, N. YAMAZOE, *Solid State Ionics* **1994**, 70/71, 572–577.
- 8 A. TAKAMI, *Ceram. Bull.* **1988**, 67 (12).
- 9 M. FLEISCHER, H. MEIXNER, *J. Mater. Sci. Lett.* **1992**, 11, 1728–1731.
- 10 U. LAMPE, M. FLEISCHER, H. MEIXNER, *Sens. Actuators*, **1994**, B17, 187–196.
- 11 Siemens Matsushita Components GmbH +Co KG, *Development Kit Manual, Gas Sensors* April 1996.
- 12 W. MENESKLOU, H.-J. SCHREINER, R. MOOS, K. H. HAERDTL, E. IVERS-TIFFÉE, *Materials for smart systems III, MRS Fall Meeting Proc.* **1999**, 604, 305–310.
- 13 E. W. WILLIAMS, N. TOMLINSON, M. T. CHENEY, A. G. KEELING, *J. Mater. Sci. Mater. Electron.* **2000**, 11, 369–372.
- 14 T. OGASAWARA, H. KURACHI, *SAE-Paper 880557*, SAE-Symposium, Detroit, MI, USA, 1988.
- 15 T. M. GUER, R. A. HUGGINS, *J. Appl. Electrochem.* **1987**, 17, 800–806.
- 16 T. MOSER, F. STANGLMEIER, B. SCHUMANN, S. THIEMANN-HANDLER, *Sensoren und Messsysteme, Tagung Ludwigsburg, March 13–14, 2000*, VDI-Bericht 1530, VDI-Verlag, **2000**, 159–167.
- 17 G. N. HOWATT, U.S. Patent 2582993, **1952**.
- 18 R. E. MISTLER, D. J. SHANEFIELD, R. B. RUNK, *Foil Casting of Ceramics, in Ceramic Processing before Firing*, (G. Y. ONADA and L. L. HENCH), John Wiley and Sons, New York, NY, USA, **1978**, 411–448.
- 19 P. S. BRETT, A. NEVILLE, W. H. PRESTON, J. WILLIAMSON, *SAE-paper 890490*, SAE-Symposium, Detroit, MI, USA, **1989**.
- 20 F. UEDA, S. SUGIYAMA, K. ARIMURA, S. HAMAGUCHE, K. AKIYAMA, *SAE-paper 940746*, Detroit, **1994**.
- 21 M. MATSUI, T. SOMA, I. ODA, *Advances in Ceramics, Vol. 12, Science and Technology of Zirconia II* (N. CLAUSSEN, M. RUEHLE, A. H. HEUER, eds.), The American Ceramic Society, Columbus, OH, USA, **1984**, 371–381.

- 22 M. J. ESPER, E. M. LOGOTHETIS, J. C. CHU, *SAE-paper 790140*, SAE-Symposium, Detroit, MI, USA, 1979.
- 23 E. IVERS-TIFFÉE, W. MENESKLOU, K. H. HAERDTL, J. RIEGEL, *7th Int. Symp. Ceramic Materials and Components for Engines*, DKG, Goslar, June 19–21, 2000, Wiley-VCH, Weinheim, Germany 2001.
- 24 J. RIEGEL, H. NEUMANN, H.-M. WIEDENMANN, *Solid State Ionics Conference, Cairns, June 2001*, Elsevier, Amsterdam, Solid State Ionics, 2001, 152/153, 783–800.

5.7

Magnetic-Field Sensor Technologies

HENRIK SIEGLE

5.7.1

Introduction

Magnetic-field sensors are widely used in automotive applications because of their robust design, their non-contacting and thus wear-free operation and their low manufacturing costs, in particular for solid-state sensors, which can be fabricated using batch processes. Their main fields of application in automobiles are speed and position sensing. In total, about one third of today's automotive sensors are based on the magnetic principle [1].

One high-volume application is the wheel speed detection for anti-lock braking (ABS) systems. Wheel speed information is also needed in modern vehicle dynamics control (VDC) and navigation systems. Both require, in addition to the wheel speed, the steering angle as an input value, which is also often provided by magnetic sensors. A classic field of application is the power train, in which magnetic sensors deliver information about the cam and crank shaft positions as well as the transmission speed.

Besides these applications of very high volume, more and more position sensors, in particular in safety-relevant and drive-by-wire systems, are replaced by magnetic sensors, because these work without contact and thus without wear. Examples are pedal-position sensors for electronic gas and electrohydraulic braking. Additional quantities which are already sensed magnetically today or which will

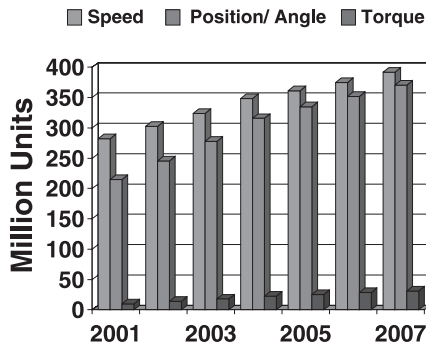


Fig. 5.7.1 Market development for selected magnetic sensor applications

become of interest in the future are the position of electric motors, steering torque and electrical current. Figure 5.7.1 outlines the development of the magnetic sensor market for some example automotive applications.

The applications described above are mainly implemented with inductive sensors or with sensors based on Hall and magnetoresistive effects (AMR, GMR). In the following sections I focus on the latter technologies. Both of these sensor technologies allow batch processing and thus low-cost devices. At the end I give a short overview of other magnetic sensor technologies that have found their way into some automotive applications and some that are in the process of penetrating this market.

As for all automotive sensors the specific application and the place in the car where the sensor is installed defines the requirements. Sensors in the interior for example have to withstand temperatures only up to $+85^{\circ}\text{C}$, whereas near the engine block or near the brakes, temperatures can reach $+150^{\circ}\text{C}$. In the future, car manufacturers will demand an even higher temperature resistance. One very specific requirement for magnetic sensors is electromagnetic compatibility. Due to the fact that more and more electronic systems and electric actuators are being installed in automobiles, engineers have to deal with higher magnetic stray fields, which may affect the sensor properties and falsify their output signal. Therefore, tomorrow's magnetic sensor needs not only to have a better performance (higher accuracy, higher sensitivity, longer lifetime) but must be even more robust. Finally, the sensor costs decide whether the developed solution will be successful in the market.

5.7.2

AMR Sensors

AMR sensors are based on the anisotropic magneto-resistive effect (AMR effect), which occurs in ferromagnetic transition metals. In these materials, electrical resistance is a function of the angle between the electrical current and the direction of magnetization. An external magnetic field can change the direction of the magnetization and thus the electrical resistance, allowing the AMR to be used as the transducer effect in magnetic-field sensors.

Although discovered as early as 1857 by William Thomson [2], only in the late 20th century has progress in thin-film deposition techniques allowed systematic studies of the AMR effect, resulting in its first applications in the read heads of hard-disk drives [3, 4] and in magnetic sensors for industrial purposes in the 1980s and 1990s [5–7]. The breakthrough of AMR sensors for automotive applications occurred in the mid 1990's with their use as speed sensors in anti-lock braking systems. Since the AMR material is rather optimized, the focus of R&D today is on application-specific sensor designs and on new or improved evaluation electronics.

5.7.2.1 Fundamentals

The AMR is a quantum mechanical effect. In the presence of an external magnetic field electrons with different spin orientations (spin-up, spin-down) possess

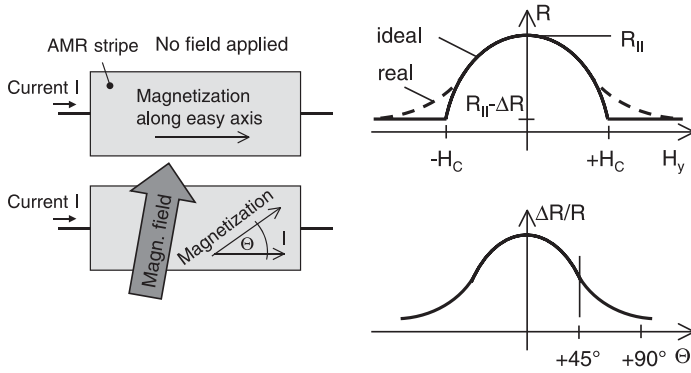


Fig. 5.7.2 Measuring geometry and AMR transfer curve

different energy levels, resulting in a magnetic field-dependent conductivity. Below, I give only a phenomenological description (for details please see [5–7]).

Usually, AMR sensors are made of narrow stripes of $\text{Ni}_{81}\text{Fe}_{19}$, so-called permalloy, sputtered as thin films with a thickness of about 40 nm and several micrometers wide. The electrical resistance in these thin films is

$$R(\Theta) = R_{||} - (R_{||} - R_{\perp}) \cdot \sin^2(\Theta) = R_{||} - \Delta R \cdot \sin^2(\Theta) \quad (1)$$

with Θ being the angle between the directions of the current and the magnetization, $R_{||}$ the resistance component parallel and R_{\perp} perpendicular to the magnetization (Fig. 5.7.2). In permalloy, the size of the effect $\Delta R/R$ is in the range of 2 to 4% at room temperature.

AMR sensors utilize the effect, that the internal magnetization vector \mathbf{M} can be rotated by an external magnetic field. In the simplified one-domain model of Stoner and Wolfarth [8] a magnetoresistive layer is always completely magnetized up to its saturation magnetization \mathbf{M}_s . An external field can change only the direction of the magnetization, not its magnitude. In the ideal case, the \mathbf{M} vector would directly follow the external field, independent of its strength. In real sensor structures however, with their rather long, very thin, narrow strips (length $l \gg$ width $w \ll$ thickness t), the shape anisotropy generates stray fields which affect the rotational behavior of the magnetization resulting in anisotropy (Figure 5.7.2). Without an external field, the preferred direction of the magnetization, the so-called easy axis, is parallel to the length of the stripe. The axis perpendicular to the easy axis is called the hard axis.

By minimizing the overall energy the correlation between the rotation of the magnetization and an external field is:

$$\sin(\Theta) = \frac{H_y}{H_c}, H_y < H_c \quad (2)$$

with H_y being the component of the external field perpendicular to the easy axis (parallel to the hard axis). H_c is the total anisotropy field strength, consisting of an internal material constant and of a geometric part according to:

$$H_c = H_{\text{int}} + \frac{t}{w} M_S \quad (3)$$

From Eq. 3 it becomes clear that the sensitivity of an AMR sensor can be adjusted over a large range by changing the thickness-to-width-ratio t/w . Inserting Eq. 3 into Eq. 1 leads to the quadratic resistance relation of AMR sensors:

$$R(H_y) = R_{\text{II}} - \Delta R_{\text{max}} \left(\frac{H_y}{H_c} \right)^2, H_y < H_c \quad (4)$$

For fields larger than H_c , the AMR stripe is saturated. The magnetization is aligned completely parallel to the external field and no field-dependent change in resistance is observed. Of course, this describes only the ideal case within the limits of the assumed single-domain model. In real sensor structures we always find misaligned domains, which deviate from quadratic behavior, because the domains interact and behave differently according to their position in the strip. Boundary domains, for example, are more fixed than domains in the center of the stripe. Therefore, the real MR curve shows a smoother transition into the saturated state (Fig. 5.7.2). Unfortunately, the magnetic behavior in real sensor structures cannot be described analytically; micromagnetic modeling is necessary [9].

The component of the external field parallel to the easy axis does not affect the resistance. But as soon as it exceeds H_c , the magnetization can flip. This has important implications for the linearized AMR curve that is described below.

For low-field sensor applications, the quadratic change in resistance with θ is not suitable. The desired transfer curve should be linear and well-defined. As can be seen in Figure 5.7.2, the curve is linear in the region of 45° . The method used to cause the current to flow at a 45° angle in the AMR film is called barber-pole biasing and is accomplished by a layout technique that places low-resistance shorting bars across the AMR strips. The current, preferring to take the shortest path through the AMR film, thus flows from one bar to the next at a 45° angle. By this means the transfer curve is symmetric around zero field with maximum sensitivity and linear for small external fields.

In practice, two problems arise. First, the above described flipping of the magnetization vector along the easy axis has to be suppressed. Second, real sensor structures exhibit more than one domain. Their alignment along the easy axis is essential for repeatable, low-noise, and low-hysteresis output signals. Therefore, a supporting field along the easy axis is usually applied by placing a magnet behind the sensor element. This additional field affects the sensor sensitivity, since it acts as a repelling force similar to the shape anisotropy.

5.7.2.2 Sensor Design, Performance, and the Manufacturing Process

AMR resistors made of permalloy have a strong temperature coefficient of 0.3–0.4% K⁻¹, as is true for all metals. To minimize drifts due to temperature changes, four AMR resistors are usually arranged in a Wheatstone bridge configuration, which also leads to an output signal twice as large as for a single resistor when constant voltage is applied. AMR resistors are meander-shaped to increase their base resistance, which is typically in the kΩ range. The sensitivity around the operating point of such device is defined as

$$S = \frac{\Delta U/U_0}{B} = \text{mV}/(\text{VmT}) \quad (5)$$

and can be tuned over a large range depending on the application. Today, AMR sensors in automobiles are used to detect magnetic fields in the range from a few μT (e.g. in compass applications) up to some 10 mT in position and angle sensors. Typical AMR sensitivity is in the range of 3–4 mV V⁻¹mT⁻¹ [10, 11] and thus about 2–4 times higher than Hall effect sensors. The sensitivity can be enlarged to above 10 mV V⁻¹mT⁻¹ by adjusting the design and can be improved even more by integrating flux gates. The temperature-sensitivity coefficient is –0.2% K⁻¹ for a single resistor and –0.4% K⁻¹ for a differentially working Wheatstone bridge.

Thin-film magnetoresistive sensors are prepared by processes similar to those used in microelectronics, so that it is reasonable to integrate AMR sensors with the evaluation circuit usually used in the bipolar CMOS technique. However, AMR sensors are usually offered in a simpler hybrid technology, in which both chips – sensor and electronics – can be fabricated by means of their optimized technology. Contamination risks and the large sensor size of about 1–2 mm² normally make the production costs of one-chip solutions more costly.

The sensor production process starts with the deposition of permalloy of several 10 nm thickness on a smooth oxidized silicon wafer by sputtering. Other compounds, such as NiCo, exhibit higher MR effects than permalloy but suffer from larger coercivity and some magnetostriction. The first leads to hysteresis, the second to stress sensitivity and consequently to temperature drifts. Care has to be taken that the substrate is smooth. Roughness increases the coercivity and supports the formation of random magnetic domains. As described above, sensor properties are best for a few large domains. To optimize the domain structure, AMR sensors are annealed in vacuum at temperatures between 250 °C and 400 °C with a magnetic field applied along the easy axis. A similar effect can be obtained by applying a magnetic field during the deposition process. For angle sensors, isotropic behavior is desired, so the annealing as well as the permalloy deposition is carried out without applying a magnetic field.

The metallization for contacts and circuitry is accomplished with aluminum, followed by a protection layer made, for example, as a stress-free combination of SiO₂ and Si_xN_y. The structuring and lithography are mostly standard micro-electronic process steps, except for the etching of permalloy, which needs a specific type of wet etch.

Some sensors include a second metallization layer to integrate a planar coil. Its additional field can be used as a flipping field for offset reduction or as a closed-loop compensation to keep the transfer curve at maximum sensitivity, what is especially necessary for low-field applications. In a third application, angle sensors, planar coils generate an auxiliary field to expand the measuring range.

The AMR sensor elements fit into standard mold packages with iron-free lead-frames. Often, the application necessitates a specific package design, as for speed sensors, in which the reading configuration does not allow much space. Today, the assembly and packaging technology is the limiting factor for the AMR temperature range. Although an AMR sensor by itself can operate at temperatures well above 200°C, the standard sensor packages are specified for temperatures below +150°C.

Sensor errors arise from scatter in the processing parameters during manufacture. Variations in sensitivity, offset, and their temperature coefficients are important. An AMR-specific problem for a long time but meanwhile mostly overcome was the bridge offset, with its strong and often unpredictable temperature drift, so that an automatic offset calibration had to be integrated into the evaluation circuit. This can be based, for example, on the flipping method mentioned above; or if the sensed magnetic field alternates, the offset can be derived and subtracted by a minimum-maximum method. The manual trimming of the offset with a trim resistor within one leg of the bridge to balance both outputs helps only if the sensor operates in a narrow temperature band or if the required accuracy is not very high.

The main suppliers for AMR elements suited for automotive applications are Honeywell (HMC series) and Philips (KMI and KMZ series). Smaller companies, such as HL Planartechnik and Sensitec, both in Germany, also deliver application-specific AMR sensors.

5.7.2.3 AMR Angle Sensors

Magnetoresistive sensors can cover the whole spectrum of magnetic-field sensor applications, because they can detect both, the strength and the angle of the sensing field. In this section I explain their application as angle sensors.

AMR angle sensors are operated in the saturating mode at magnetic fields typically above 50 mT. As described above, for large fields the AMR effect turns into a pure angle dependency. Variations in the applied field do not affect the output signal. This measuring principle is fairly independent from assembly and magnet tolerances, as well as from aging- and temperature-induced changes in the magnetic field strength.

Figure 5.7.3 shows the principle setup and layout of a conventional AMR 180° sensor consisting of 8 AMR resistors that are combined into 2 Wheatstone bridges. One of these bridges is turned by 45°. When an in-plane sensing field is rotated by, for example, simply rotating a permanent magnet above the sensor, the first bridge delivers a sine output signal, the other bridge a cosine signal, according to:

$$V_1 = A(T) \sin(2\alpha), V_2 = A(T) \cos(2\alpha) \quad (6)$$

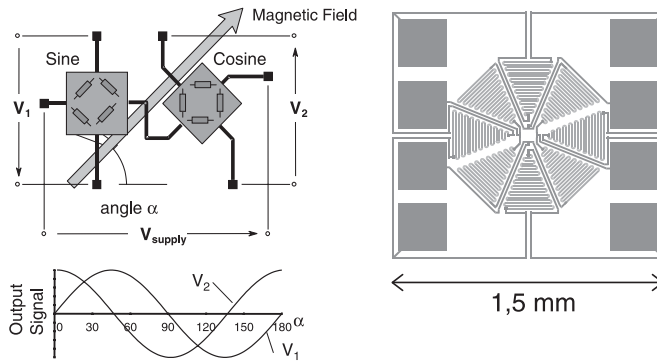


Fig. 5.7.3 Principle and layout of an AMR angle sensor for 180° measuring range

Using the arctangent function, the angle of the external magnetic field can be calculated as:

$$\alpha = \frac{1}{2} \arctan\left(\frac{V_1}{V_2}\right) \quad (7)$$

In an evaluation circuit this calculation is usually done with the so-called CORDIC algorithm. The amplitudes of both bridges are almost the same, because the bridges are placed into one another and thus have experienced the same manufacturing processes and will be operated at the same temperature. Therefore, in this operation the temperature-dependent signal amplitudes $A(T)$ are removed, resulting in delivery of pure angle information largely without any temperature drifts. Of course, small asymmetries in the bridges always remain leading to offset voltages with their own temperature dependencies.

As can be seen from Eqs. (6) and (7), the AMR output signal depends on twice the external magnetic field angle, restricting the absolute measuring range to 180°. Because many applications (e.g. in steering systems) need absolute angle information over a full rotation, we developed at Bosch an AMR sensor with an extended measuring range of 360° [12]. This AMR360 element consists of an additional planar coil, which is placed above the AMR resistors (Fig. 5.7.4). An electrical current generates an auxiliary field, resulting in small changes in the output signal of both AMR bridges, making it possible to perform range discrimination. A signal change, approximately only 1% of the signal amplitude, which requires a current of only a few milliamps, is sufficient to ensure a faultless range discrimination.

Today's angle sensors based on the AMR effect allow an accuracy of better than 1° to be obtained over the full measuring range and for temperatures up to +150°C. The main causes of errors are offset drifts ($< 2 \mu\text{V V}^{-1}\text{K}^{-1}$ per bridge). Inaccuracies due to form and uniaxial or crystal anisotropy are only in the range of

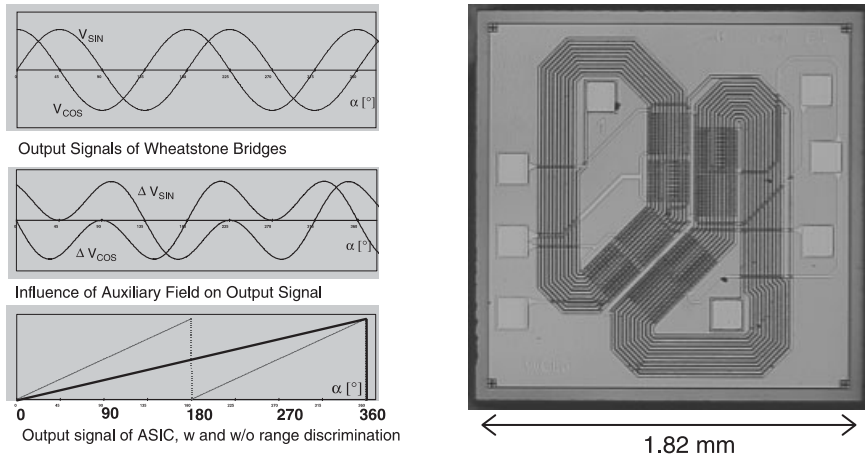


Fig. 5.7.4 Principle and layout of the AMR360 sensor

0.1° – 0.2° and 0.02° , respectively. Their contributions become important only when using weaker magnetic fields. In the application of course, the engineer has to pay care to correctly design the magnetic circuit. Non-centric alignment of the external magnet, whose angle is to be measured, can easily produce inaccuracies on the order of 0.1° or more.

5.7.3

GMR Sensors

In the late 1980's scientists in France and Germany observed high magnetoresistance in multilayers of alternating ferromagnetic and non-ferromagnetic materials [13, 14]. Since the changes in resistance were as high as 70%, this effect was named giant magnetoresistance or GMR. Since then, intensive R&D has led to further improvements in materials and by 1995 the first devices were available on the market [15]. At the end of the 1990s GMR-based sensors were installed as read heads in hard disk drives [4]. In this field of application, GMR has today replaced the formerly used AMR technology due to its better sensitivity, allowing much higher data-storage densities. Recent prototypes achieve storage densities of 100 Gb in^{-2} [16]. GMR angle sensors were first reported and introduced on the market by Siemens/ Infineon in 1998 [17, 18]. It is only a question of time as when GMR sensors will also find their way into the automotive business: MELCO recently announced the production of GMR position sensors for cam and crank shaft applications that will be delivered to car manufacturers in 2003 [19, 20]. Intensive R&D by other suppliers of automotive sensors, for example, Philips [21] and Bosch [22], has led to the development and presentation of prototypes for various magnetic sensor applications based on the GMR effect.

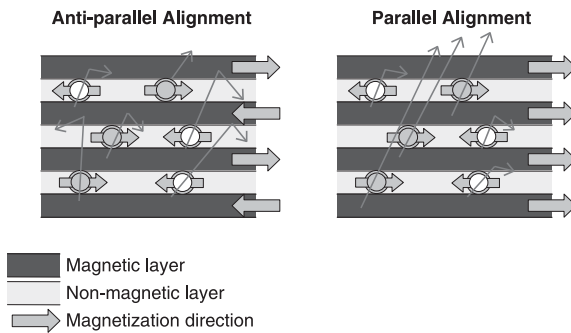


Fig. 5.7.5 Spin-dependent scattering

5.7.3.1 Spin-Dependent Scattering

The GMR effect occurs in multilayers of alternating ferromagnetic and non-ferromagnetic materials. Due to spin-dependent scattering of the current-carrying electrons, the resistance of these sandwiches can be altered by changing whether the magnetization of the ferromagnetic layers is parallel or anti-parallel (Fig. 5.7.5). Electrons in layers with parallel magnetization undergo less scattering at the interfaces. They have longer mean free paths, resulting in lower electrical resistance. In the anti-parallel state, the scattering rate is increased and also the resistance. For spin-dependent scattering to be a significant part of the total resistance, the layers must be thinner than the mean free path of electrons in the bulk material. Typical mean free paths in ferromagnets are tens of nanometers, so the layers themselves must each be several nanometers thick.

Various approaches can be used to obtain a controlled alignment of the magnetization in magnetic multilayers. The most important material systems for devices are anti-ferromagnetic coupled multilayers and spin valves, which is discussed below.

5.7.3.2 GMR Multilayers

The anti-parallel alignment in GMR multilayers is achieved by an indirect exchange coupling between the magnetic layers, for which the thickness and material properties of the non-magnetic layer are critical. The sign of the coupling oscillates as a function of this non-magnetic spacer layer and the coupling strength decreases with increasing spacer thickness [23, 24]. With the right choice of spacer thickness, the magnetic layers are aligned anti-parallel in zero field (anti-ferromagnetic state) and the electrical resistance is maximum. Applying an external magnetic field forces the magnetization directions of the magnetic domains to be more and more parallel with respect to the applied field and the electrical resistance decreases. At saturation, the external field exceeds the anti-ferromagnetic coupling, and all domains are oriented parallel to the external field. In this parallel state, the electrical resistance reaches its minimum.

Figure 5.7.6 displays an MR curve of a Bosch-GMR multilayer consisting of CoFe and Cu. Increasing the number of alternating layers increases the GMR

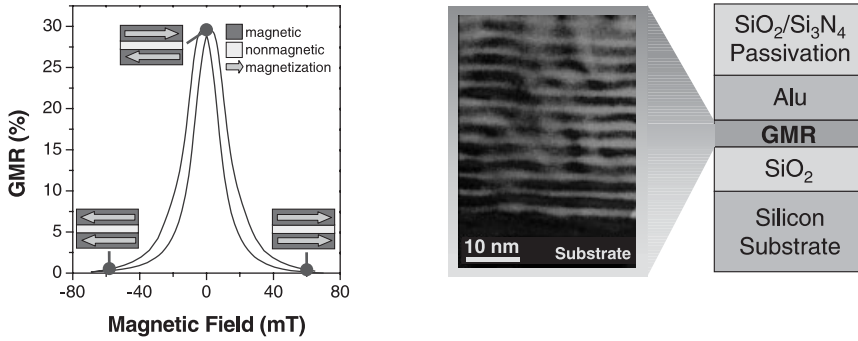


Fig. 5.7.6 MR curve and TEM picture of coupled multilayers

effect. In the situation depicted here, 20 bilayers results in a relative effect of 25–30% at room temperature. The GMR effect is defined as

$$\text{GMR} = \frac{R_{\min} - R_{\max}}{R_{\min}} = \frac{R_{\text{saturated}} - R_0}{R_{\text{saturated}}} \quad (7)Q11$$

The thickness of the Cu spacer layer (Fig. 5.7.6) was chosen to fit the second anti-ferromagnetic coupling maximum ($t_{\text{Cu}} = 2.15$ nm). Reducing the thickness to the first maximum ($t_{\text{Cu}} = 1.1$ nm) increases the effect even more and also results in a minimum hysteresis, but the slope is much flatter and thus the device is less sensitive, because, due to the stronger coupling, the saturated state is reached at higher external fields. In addition, thinner layers mean less temperature stability. The example stack shown here withstands ambient temperatures up to 270 °C for one hour.

5.7.3.3 Spin Valves

Spin valves were introduced shortly after the discovery of the GMR effect in anti-ferromagnetic coupled multilayers in 1991 [25]. In this material system, controlled alignment of the magnetization directions is obtained by ‘pinning’ one of the magnetic layers. The second magnetic layer, separated by the nonmagnetic spacer, is called the free layer, and its magnetization follows an applied external field. To minimize coupling between the magnetic layers, the spacer in spin valves is thicker than in GMR multilayers, but a residual coupling always remains, because the thickness cannot be made too thick, otherwise the GMR effect diminishes. In addition, roughness of the interface between the layers results in the so-called orange-peel coupling. The residual coupling plays an important role in the transfer curve of sensor devices, because it shifts the MR curve away from zero field and gives rise to significant anisotropy not suitable for angle applications.

Pinning of a magnetic layer is accomplished by exchange biasing with an additional antiferromagnetic layer. Many different materials have been studied, the most important being FeMn, NiO, IrMn, and PtMn [26–30]. The materials differ

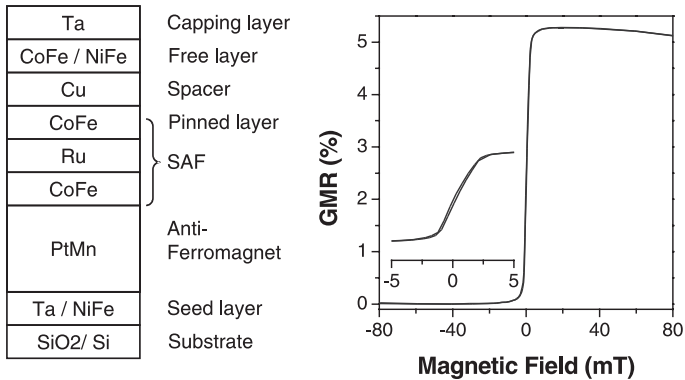


Fig. 5.7.7 Spin valve setup and MR curve

in their pinning strength. The blocking temperature of FeMn, NiO, and IrMn, that means the temperature at which coupling is lost, which corresponds to the Neel temperature of bulk antiferromagnets, is too low for automotive applications. Therefore, at Bosch we developed a material system based on PtMn. To further increase its pinning strength, we added a synthetic antiferromagnet (SAF) between PtMn and the pinned layer. The SAF's setup is similar to that of GMR multilayers, but the layers are coupled in their first antiferromagnetic maximum and include ruthenium as a spacer material.

Figure 5.7.7 shows a typical MR curve for our spin-valve material. Because it contains only two active magnetic layers, the effect is much smaller than that in GMR multilayers, but the slope at zero field is much steeper and so is the sensitivity of the device. To increase the magnitude of the effect, the whole stack can be doubled or mirrored (dual spin valve) and reflecting nano-oxide layers (NOL) can be included. These act as mirrors for the conduction electrons, which are specularly reflected (spin conservative) back into the magnetic layer. With these measures, effects of greater than 20% can be obtained [30].

5.7.3.4 Manufacturing Process

The manufacturing process of GMR sensors is similar to that of AMR sensors. Smooth oxidized silicon wafers are used as substrates. The metallization, protection, and structuring processes are comparable. The greatest difference is the deposition of the active sensor layers, since these contain up to 8 or more materials with thicknesses in the subnanometer range. Only the tremendous recent progress in thin-film deposition techniques allows the controlled and repeatable manufacturing of these metallic films. Although GMR stacks were first deposited by MBE (molecular beam epitaxy), ultra high-vacuum DC-magnetron and ion-beam sputtering are today the established techniques. To ensure the best homogeneity over the wafer, the layers are sputtered with an oscillating technique, by which the silicon wafer is rotated underneath the operating cathodes. Shaper or special cathode geometries further improve the homogeneity.

All GMR stacks start with a specific seed layer, which defines the growth mode of the subsequent films, and end with a capping layer to avoid unwanted oxidation of the stack. Further processing is similar to that for AMR sensors, but care has to be taken that processing temperatures are well below the degradation limit, which is in the range of 250–350°C. For the magnetic materials CoFe and NiFe, magnetostriction-free compositions are chosen. PtMn as the pinning material in spin valves needs an additional field-anneal step during its formation. Further annealing steps are performed to increase the lifetime.

The GMR-sensing material can be directly integrated with semiconductor processes. Because of their higher sensitivity than AMR, the sensor area can be made much smaller, fitting well with other semiconductor structures. The sensing elements are fabricated after most of the semiconductor fabrication operations are complete. Due to the topography introduced by the many layers of bipolar-CMOS electronics, areas with no underlying transistors or connections are usually reserved for the GMR resistors. Otherwise, additional layers (e.g., spin-on glass) have to be added to smooth out the topography. Integrated sensors have been demonstrated by NVE and MELCO [15, 19, 20]. So far, only a few companies sell GMR sensors for applications outside the hard disk drive business [15, 17–20]: NVE put GMR sensors on the market based on coupled multilayers for speed, position, and current sensors. Infineon provides a 360° angle sensor, and MELCO announced a position sensor for cam and crank shaft applications that will enter the market in 2003.

5.7.3.5 GMR Speed Sensor

Figure 5.7.8 shows the layout of a rotational speed sensor that is designed as a gradiometer. The Wheatstone bridge is split into two half bridges, which are sepa-

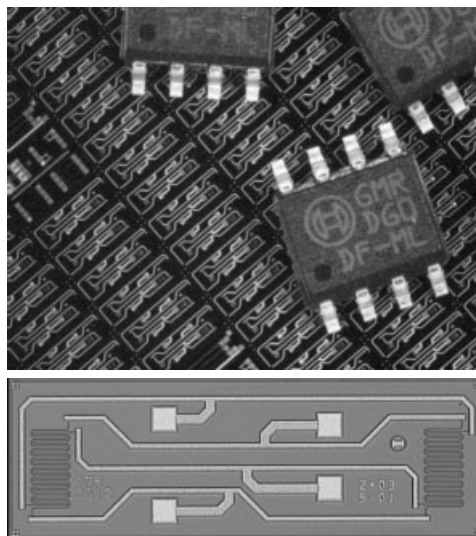


Fig. 5.7.8 GMR speed sensor (ca. $1 \times 3 \text{ mm}^2$)

rated from each other by a distance of 2.5 mm. The gradiometer is sensitive only to field differences or gradients; homogeneous interfering stray fields shift the operating point of the sensor but do not generate any signal. Both GMR multilayers and spin valves can be used as the sensing material.

With GMR multilayers (see transfer curve in Fig. 5.7.6), the operating point should be shifted into the steep linear region with high sensitivity. This can be achieved by placing a supporting magnet behind the sensor, similar to AMR sensors, or by integrating this magnet into the GMR stack as an additional layer. It is crucial that the integrated layer possess very good and stable coercivity, otherwise the speed sensor would react differently depending on the distance to the target wheel. In our example, we integrated a CoCrPt layer approximately 10 nm thick [31]. To reduce through this bias film, an additional high-resistance layer is placed between the integrated bias and the sensing layers.

When using spin valves as a sensing material, the transfer curve *must* be symmetric around zero with minimum hysteresis, and the linear working range must be sufficient, otherwise even small interfering fields saturate the sensor. The transfer curve can be tailored by optimizing the spin-valve stack and the sensor design. For example, the transfer curve can be tilted to expand the working range by utilizing the form anisotropy of narrow GMR strips.

The GMR speed sensor shown in Fig. 5.7.8 has excellent sensitivity – more than $10 \text{ mV V}^{-1} \text{ mT}^{-1}$ – and very good immunity against interfering fields. It can be operated at temperatures up to $+190^\circ\text{C}$. The offset stability, which is important for signal quality, is similar to that of AMR sensors and well below $2 \mu\text{V V}^{-1} \text{K}^{-1}$.

Due to the higher signal and the better sensitivity compared to the established Hall and AMR technologies, the GMR speed sensor allows larger air gaps between it and the target wheel. With the first prototypes we achieved air gaps of about 7 mm, compared to the 2–3 mm that is standard today. This is even more impressive when considering the fact that the strength of the magnetic field of the target wheel decays nearly exponentially with increasing distance. On the other hand, keeping the air gap constant, the higher sensitivity of GMR makes it possible to reduce the diameter of the target wheels, to increase the number of poles on these wheels, to enhance the accuracy in timing applications, or to use weaker and consequently cheaper magnets.

5.7.3.6 GMR Angle Sensor

In GMR angle sensors, only spin valves can be used as the sensing material, because the GMR effect in coupled multilayers is isotropic. Analogous to AMR-based sensors, GMR resistors are arranged in two Wheatstone bridges. In a rotating external field, one of these bridges generates a sine signal, the other a cosine signal. The angle of the external field is derived by applying the arctangent function to both signals. The main advantage of GMR is that it is a uniaxial effect, that is, the resistance is proportional to the cosine of the angle θ of the external field with respect to the pinning or reference direction:

$$R = R_0 - \Delta R \cdot \cos(\theta) \quad (9)$$

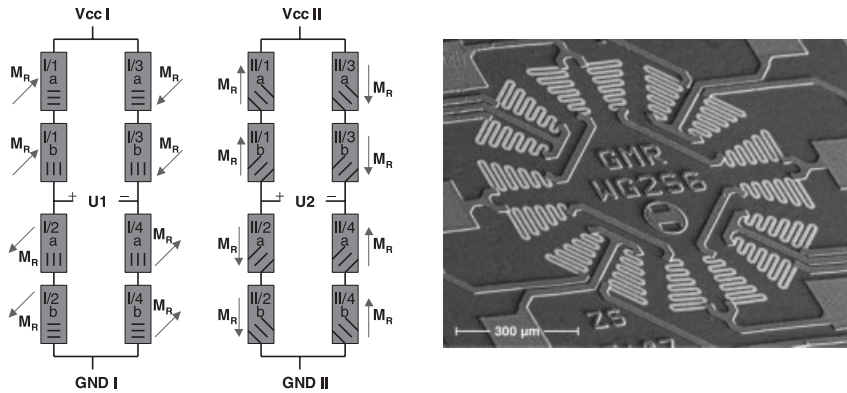


Fig. 5.7.9 GMR angle sensor

Therefore, GMR angle sensors have a ‘natural’ 360° measuring range, in contrast to the 180° covered by AMR sensors. The angle signal calculated from the arctangent does not repeat after 180° as occurs with AMR sensors, so that no additional sensor or planar coils are necessary to sense a complete rotation of an external field.

To create a sine and a cosine GMR bridge, the single resistors need at least two pinning directions with a 90° phase difference and an additional two directions to double the output signal (Fig. 5.7.9). After deposition, the pinning in spin valves containing antiferromagnets usually has a preferred direction. To change it, the material has to be heated above the Néel temperature (blocking temperature) of the antiferromagnet while applying an external field in the desired pinning direction. During cooling, the pinning direction is fixed. The challenge is to do this imprinting of different pinning directions on the micrometer scale of the devices. Two different methods were developed. In the first, current flowing through the GMR resistors or through an additional circuit provides local heating. The second method utilizes short laser pulses, which are focussed on the selected sensor region [32, 33].

Figure 5.7.9 shows an example layout of a GMR angle sensor. The full bridges are split into 16 single elements that are arranged in a circle. The rotational symmetry further minimizes angle errors. We achieve an accuracy better than 1.5° over the complete temperature range from -40 to $+150^\circ\text{C}$. Since the spin valve material is extremely robust, another feature of the sensor is its immunity against interfering fields, exceeding 100 mT at 200°C .

Compared to AMR, GMR technology allows absolute angle measurement over one complete rotation without any additional measures. The inherent 360° measuring range is also advantageous for small-angle applications, due to the larger linear response range. In addition, similar to the speed sensor, weaker and thus more cost-efficient magnets can be used.

5.7.3.7 Related XMR Effects

Besides the above multilayers and spin valves, other GMR-related material systems have been developed or are presently under investigation.

Granular systems consist of ferromagnetic particles embedded in a nonferromagnetic, conducting matrix, for example, NiFe particles in an Ag matrix. If the distance between the ferromagnetic particles is about the same as the mean free path of the electrons, spin-dependent scattering can be observed. The granular materials are very stable over temperature but suffer from very low sensitivity, making their application unlikely.

Tunneling magnetoresistance (TMR) material has a similar stack as do spin valves, but (instead of a conducting spacer) an isolating barrier, usually Al_2O_3 , is placed between the two ferromagnetic layers. The electrons tunnel perpendicularly through the TMR stack, which is why it is called tunneling magnetoresistance. The tunneling current is largest when the magnetization directions in both magnetic layers are parallel and smallest in the anti-parallel state. The magnitude of the effect ranges from 30% to more than 50%. Presently, such companies as IBM, Infineon, Motorola, and others are working intensively on TMR elements for the use in new nonvolatile memories, called *magnetic RAMs (MRAM)*.

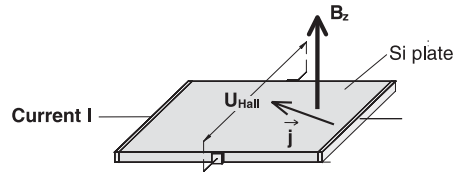
Colossal magnetoresistance (CMR) was discovered in 1993 and is, in contrast to GMR and TMR, an intrinsic effect [34]. The extreme change in resistance, on the order of 100%, can be explained by the suppression of a metal-to-insulator transition at the Curie temperature T_c due to an external magnetic field. Below T_c the spins of electrons in ferromagnetic materials are aligned parallel (corresponding to the Néel temperature of antiferromagnets). At temperatures above T_c CMR materials are insulators or semiconductors, and below T_c , ferromagnetic metals. The CMR effect was observed for the first time in manganese oxides (e.g., LaSrMnO_3). The electrical current through this material is carried by electrons hopping between the Mn sites. This hopping occurs only if the magnetic moments of the corresponding Mn atoms are aligned parallel. By applying an external magnetic field, this ferromagnetic state can be reached at temperatures below T_c and the electrical resistance can be decreased. There are two reasons why the CMR is not technically used so far. First, most CMR materials have a T_c close to room temperature, and second, colossal resistance changes are obtained only at very high magnetic fields, on the order of several Tesla. Therefore, CMR materials have a very restricted temperature range and very low sensitivity.

5.7.4

Hall Sensors

Hall sensors are based on the Hall effect, which was discovered by E.H. Hall in 1879 (Fig. 5.7.10). If a long, flat, current-carrying conductor is placed in a magnetic field, the moving charges experience a net force mutually perpendicular to the direction of the current flow and the magnetic field. Under the influence of this Lorentz force, the electrons pile up on one edge of the conductor and the pos-

Fig. 5.7.10 Hall effect



itive charges on the other edge. This Hall voltage is directly proportional to the applied field according to

$$V_H = S \cdot I \cdot B \quad (10)$$

where S is a constant material-dependent sensitivity. By holding the control current constant, the Hall voltage can be used to measure the magnetic field.

Typical sensor materials are n-type silicon, GaAs, InSb, and InAs. Because of their larger carrier mobility, III–V semiconductors allow much higher sensitivities and also better dynamic behavior. Their frequency-response capabilities are well above the 10–20 kHz typical of silicon Hall sensors. Moreover, GaAs in particular has, due to its larger bandgap, also a lower temperature coefficient for its resistance and sensitivity. Nevertheless, most Hall sensors are based on silicon, because it is much cheaper and the sensor can easily be integrated with an evaluation circuit. Because all sensors need some kind of electronics, the integration of silicon Hall elements can be done at nearly zero cost. Two other advantages of monolithic integration are, first, better immunity to electromagnetic interference, and second, very compact size, allowing standard and single-mold packages.

The sensitivity of silicon Hall elements ranges from some mV T^{-1} to 100 mV T^{-1} , which is about 2–4 times more sensitive than AMR sensors and 10 times more sensitive than GMR sensors.

Severe disadvantages of silicon Hall sensors are the huge offsets, the massive variations in sensitivity ($\pm 20\%$), and the large temperature drifts. That is why they are not often used for precise measurements of magnetic fields. But, because of their compatibility with standard semiconductor processes, they are much cheaper than all other solid-state magnetic field sensor technologies, which is the deciding advantage for many cost-driven automotive applications. Nevertheless, the scattering in the manufacturing processes and stress due to packaging and assembly can lead to offset signal values that are far above the desired sensor signals ($\pm 200 \text{ mV}$), so that calibrating steps in the sensor ASIC are inevitable.

In one method, the sensor is calibrated after installation. The offset is subtracted and the amplification is adjusted to the altered sensitivity. In addition, the temperature drift of the sensor and the drift of the permanent magnet used in the specific application can be programmed into the ASIC [35]. In a second method, the offset is actively compensated for by chopping [36]. Here, the applied current and the taps of the Hall voltage are continuously commutated. An advanced version of this technique is the spinning-current method, in which 8 or even 12 different current directions are used [37, 38]. The current direction is spun

around the central axis of the sensor element step-by-step. Meanwhile, the Hall voltage is measured at the contacts that are perpendicular to the current flow. Offset compensation takes place in time by summation of the measurement results. This method is more complex but is capable of reducing the offset to $10\ \mu\text{T}$, so that even compass applications are possible [37]. For a modulated signal, as in speed detection, the minimum and maximum signals can be detected and the offset can easily be subtracted.

In contrast to AMR and GMR sensors, Hall sensors are typically used to measure the strength of an applied external magnetic field, because only the normal field component is detected. To also measure the field direction, supporting measures (e.g., a special magnetic circuit) have to be added. However, most Hall applications in cars are restricted to simple switching or linear field measurements. Hall switches have an integrated comparator with predefined switching points and a digital output that can be adapted to various logic systems. A standard Hall switch has only a single Hall plate; differential Hall sensors contain two Hall plates that are some millimeters apart from each other. The output transistor is switched according to the difference of the magnetic field between the plates. Similar to the above described GMR gradiometer (Fig. 5.7.8), this arrangement has the advantage of being more immune to interfering fields and of giving twice the output signal.

Such a differential Hall sensor for wheel speed applications resulted from a joint development project by Infineon and Bosch [39]. In addition to the standard speed signal, this sensor also detects the direction of rotation and provides further information on the amplitudes and the quality of the magnetic signal at the wheel, allowing the driver to be warned when reduced functionality occurs. The sensor has an extremely low cutoff frequency of approximately 1 Hz, corresponding to an automobile velocity of $2\ \text{cm s}^{-1}$. In addition to Infineon, Honeywell, Allegro, Melexis, and Micronas are manufacturing Hall sensors for automotive applications.

5.7.5

Other Magnetic Sensor Technologies

The Hall and AMR technologies, and probably in the near future also the GMR technologies, are the solid-state magnetic sensor technologies that cover most of today's automotive applications together with inductively working sensors. A variety of other methods can be used to sense magnetic fields, but they are not very relevant to the automotive sector: SQUIDS (Superconducting Quantum Interference Device), search coil magnetometers, fluxgates, magneto-inductive sensors, magnetodiodes, fiberoptic magnetometers, and so on. An overview is given in [40]. Below, two other principles that are also popular in automotive applications are briefly outlined.

Magnetoresistors, similar to Hall sensors, utilize the deflection of carriers due to the Lorentz force when an external magnetic field is applied. The deflection increases the path of the carriers through the device and thus increase the electrical resis-

tance. Usually the sensor is made of InSb or InAs; both compounds are III–V semiconductors with high carrier mobility. To enhance the sensor effect, conducting needles of NiSb are embedded into the material. They lead to a further increase in the path the carriers have to take through the sensor. As is usually done in producing magnetoresistive sensors, the sensor resistors are formed as meanders and connected in a Wheatstone bridge configuration to optimize the output signal and minimize temperature effects. The sensors are very robust but of low sensitivity. Applications in the automobile are speed and position sensing.

Low cost, simplicity, reliability, and zero power consumption make *reed switches* popular in many automotive applications [41]. They consist of a pair of flexible ferromagnetic contacts hermetically sealed in a container, often glass, filled with an inert gas. The magnetic field along the long axis of the contacts magnetizes the contacts, causing them to attract one another and close the circuit. A reed switch, together with a separate magnet, form a simple proximity switch. The magnet, fixed to the moveable part, activates the reed switch when it comes close enough.

The author thanks BMBF for the support within the “Leitprojekt Magneto-Elektronik”, FKZ 13N7379/4.

5.7.6

References

- 1 Strategy Analytics, *Active Sensor Market Trends*, Part I, Market Study, March 2001.
- 2 W. THOMSON, *Proc. R. Soc. London*, A8 (1857) 546–550.
- 3 M.H. KRYDER, An introduction to magnetic recording heads, *Proceedings of the NATO Advanced Study Institute*, ed. by G. C. HADJIPANAYIS, Dordrecht, Netherlands. Kluwer (2001) pp. 449–466.
- 4 C. H. TSANG, R. E. FONTANA JR., T. LIN, D. E. HEIM, B. A. GURNEY, M. L. WILLIAMS, *IBM J. Res. Dev.* 42 (1998) 103.
- 5 U. DIBBERN, Magnetoresistive sensors in: *Sensors, Vol. 5: Magnetic Sensors*, eds. W. GÖPEL, J. HESSE, J. N. ZEMEL, K. J. OVEERSHOFF. VCH, Weinheim (1989).
- 6 K. DIETMAYER, *Technisches Messen*, 68 (2001) 269.
- 7 S. TUMANSKI, *Thin Film Magnetoresistive Sensors*, Institute of Physics Publishing, Bristol, 2001.
- 8 E. C. STONER, E. P. WOLFARTH, *Philos. Trans. R. Soc. London A, Math. Phys. Sci.* 240A (1948) 599.
- 9 J. FIDLER, T. SCHREFL, *J. Phys. D: Appl. Phys.* 33 (2000) R135.
- 10 Compare application notes from Philips: www.semiconductors.philips.com/acrobat/various/SC17_GENERAL_APP_1996_1.pdf, p. 5 application notes at this URL, please supply complete URL(s) and page title(s)
- 11 Compare application notes from Honeywell: *The World of Magnetic Sensors*, www.ssec.honeywell.com/magnetic
- 12 U. CADUFF, H. SCHWEREN, H. KITTEL, Low cost angle sensors for multi-purpose applications, *SAE Conference*, Detroit, Michigan, March 6–9, 2000; 2000-01-0090, patents DE 19839446 A1, US 6433535 B1
- 13 M. BAIBICH, J. BROT, A. FERT, F. NGUYEN VAN DAU, F. PETROFF, P. ETIENNE, G. CREUZET, A. FRIEDERICH, J. CHAZELAS, *Phys. Rev. Lett.* 61 (1988) 2472.
- 14 G. BINASCH, P. GRÜNBERG, F. SAURENBACH, W. ZINN, *Phys. Rev. B* 39 (1989) 4828–4830
- 15 J. DAUGHTON, A. V. POHM, R. T. FAYFIELD, C. H. SMITH, *J. Phys. D: Appl. Phys.* 32 (1999) R169.
- 16 H. KANAI, K. NOMA, J. HONG, *Fujitsu Sci. Tech. J.* 37 (2001) 174.
- 17 H. A. M. VAN DEN BERG, W. CLEMENS, G. GIERES, G. RUPP, M. VIETH, J. WECKER,

- S. ZOLL, *J. Magn. Magn. Mater.* 165 (1997) 524.
- 18 Infineon Technologies, *Semiconductor Data Book 04-99*, 175–196.
- 19 K. JOST, SAE Automotive Engineering 2002, September issue
- 20 Mitsubishi Electric Global website: <http://global.mitsubishielectric.com>
- 21 K.-M.H. LENSSON, D.J. ADELERHOF, H.J. GASSEN, A.E.T. KUIPER, G.H.J. SOMERS, J.B.A.D. VAN ZON, *Sens. Actuators A85* (2000) 1; C. GIEBELER; D.J. ADELERHOF, A.E.T. KUIPER, J.B.A. VAN ZON, D. OELGESCHLAGER, G. SCHULZ, *Sens. Actuators, A*, 91 (2001) 16.
- 22 H. SIEGLE, *Autotechnik* 9 (2001) 14; C.P.O. TREUTLER, H. SIEGLE, *Technisches Messen*, 68 (2001) 280.
- 23 P. GRÜNBERG, *Phys. Rev. Lett.* 57 (1986) 2442.
- 24 S. S.P. PARKIN, *Phys. Rev. Lett.* 64 (1990) 2304.
- 25 B. DIENY, V.S. SPERIOSU, S. METIN, S.S. PARKIN, B.A. GURNEY, P. BAUMGART, D.R. WILHOIT, *J. Appl. Phys.* 69 (1991) 4774.
- 26 B. DIENY, *J. Magn. Magn. Mater.* 136 (1994) 335.
- 27 J. NOGUÉS, I.K. SCHULLER, *J. Magn. Magn. Mater.* 192 (1999) 203.
- 28 M. LEDERMAN, *IEEE Trans. Magn.* 35 (1999) 794.
- 29 G.W. ANDERSON, Y. HUAI, M. PAKALA, *J. Appl. Phys.* 87 (2000) 5726.
- 30 S. ARAKI, M. SANO, S. LI, Y. TSUCHIYA, O. REDON, T. SASAKI, N. ITO, K. TERUNUMA, H. MORITA, M. MATSUZAKI, *J. Appl. Phys.* 87 (2000) 5377.
- 31 M. RABE, U. MAY, H. SIEGLE, REISS, HÜTTEN, *J. Appl. Phys.*, submitted
- 32 J.K. SPONG, V.S. SPERIOSU, R.E. FONTANA, Jr., M.M. DOVEK., T.L. HYLTON, *IEEE Trans. Magn.* 32 (1996) 366–371.
- 33 A. JOHNSON, G. MÖRSCH, H. GUNTHER, WO 02082111.
- 34 R. VON HELMOLD, J. WECKER, B. HOLZAPFEL, L. SCHULTZ, K. SAMWER, *Phys. Rev. Lett.* 71 (1993) 2331.
- 35 V. HILIGSMANN, B. MARSHALL, C. PEPIN, P. RIENDEAU, *Proceedings SENSORS EXPO Baltimore*. Helmer, Peterborough, NH, USA, 1999, 417–425.
- 36 Micronas, *Sensors Overview* at www.micronas.com/products/overview/sensors/index.php.
- 37 P. DE JONG, F.R. RIEDIJK, J. VAN DER MEER, *Sensors Proc. IEEE* 2 (2002) 1440–1444.
- 38 H. LEMME, *Elektronik* 24 (1994) 57; H. LEMME, *Elektronik* 3 (1998) 40.
- 39 V. GUSSMANN, D. DRAXELMAYR, J. REITER, T. SCHNEIDER, R. RETTIG, *SAE Conference 200*, SAE 2000-01-C058, (2000) 389.
- 40 J.E. LENZ, *Proc. IEEE* 78 (1990) 973.
- 41 M.J. CARUSO, T. BRATLAND, C.H. SCHMIDT, R. SCHNEIDER, *Sensors, USA*: Helmers Publishing, vol. 15, no. 12, Dec. 1998, p. 34–46.

5.8

Assembly, Packaging

KURT WEIBLEN and JAN PETER STADLER

5.8.1

Introduction

This section describes a standardized approach to assembly, packaging, and housing of sensors for automotive applications. It is based on general considerations driven by the stringent requirements of the harsh environments within automotive applications and the high volume manufacturing of these sensors. This approach can be applied to various sensor types (acceleration, pressure, yaw rate, chemical, and others) typically made by silicon micromachining technology. The main driver is always cost. Often, the cost for assembly and packaging can be more than half the total cost of a complete sensor product, either directly or indirectly via the underlying sensing element. Therefore, only an overall approach taking into account the various, often conflicting, considerations can yield optimum product solutions. In general, high volume manufacturing is required, driven by automotive volumes of at least a few ten thousand up to several million pieces.

5.8.2

General Considerations

From the packaging point of view new products have to be flexible to be applicable to varying customer requirements. Every automotive manufacturer will insist on his individual connector, or more general, his mechanical interface. The sensor packaging has to be able to meet these specific requirements in a flexible way, within the basic concept and without additional cost.

The package has to protect the sensing element from the harsh environment (i.e., the engine compartment) over the whole lifetime of the automobile. It has to support zero failures during ECU (electronic control module) manufacturing and car assembly and during the full lifetime of the automobile. This requires robust designs for high volume manufacturing at minimum cost yielding highly reliable products. This greatly extends the demands on automotive-grade assembly and packaging, especially as compared with direct consumer products.

As described within the different application section of this book, an automotive sensor has to be protected against various influences and loads. In addition to the extreme temperature range and temperature shock are two important requirements, which are specific requirements for micromechanical sensors. These are protection against mechanical overload, that is, e.g. from falling down, and protection against electromagnetic interference (EMI), for example, from cell phones. Dropping a sensor from only 1.5 m high onto a concrete surface results in acceleration up to 10 000 times the specified range of a typical accelerometer. The high radiation power of modern cell phones can produce EMI strong enough to cause severe problems for sensitive electronic circuits, as used for signal evaluation of

automotive sensors. Both problems have to be considered during the early concept phase to avoid late, and thereby costly, changes.

High volume manufacturing at low cost requires fully automated manufacturing lines. This has to be taken into account both for the packaging design and the manufacturing processes and requires robust, well established processes with wide tolerances. Typical high volume assembly and packaging is based on pick-and-place, with cycle times of a few seconds and ideally under clean, dry (no glue, wet chemicals, etc.), and cold (no annealing, curing) conditions.

5.8.3

Standardization

5.8.3.1 Drivers for Standardization

To be successful within the automotive field, sensor suppliers have to meet many challenges. One important issue is standardization. Worldwide competition requires competitive product cost despite individual specifications. This can be achieved only with standardized concepts and processes, forming building blocks that can be used in multiple ways. The main drivers for standardization are

- **Development cost, time and risk**
Due to their high complexity, an automotive microsystem cannot always be developed from scratch but has to be based – at least partly – on established solutions and building blocks.
- **Time to market**
Time to market is continually increasing in importance and can be improved only through evolutionary development and improvement of existing products; otherwise, the always decreasing time goal between project start and volume production cannot be met.
- **Competition of automotive suppliers**
The automotive market is based on worldwide competition. This forces suppliers to serve customers on a worldwide basis, thereby increasing pressure on local suppliers and niche players. Manpower and materials are expensive resources. Therefore cost-competitive products can only be formed when manufacturing is simple and reliable.
- **Manufacturing volume ramp-up and utilization**
A completely new product requires new manufacturing equipment and production lines, which are not fully utilized during initial ramp up to the final manufacturing volume. A reasonable return on investment (ROI) can be achieved only when initial parts can be manufactured with existing manufacturing infrastructures. This requires reuse of concepts, in other words, standardization. Once the production line is established, expansion of production capacity is straightforward and can be done incrementally.
- **Product quality and maturity**
Every new development, every new process, and every single modification bears risks. The risk increases with increasing system complexity, especially due to the fact that typically not every potential interaction is known or understood.

Product maturity, and hence high product quality within the initial manufacturing volumes, can be achieved only when established elements are used where ever possible. New elements and new processes, as required by new product features, inevitably add a lot of innovation. Therefore an unnecessarily 'all new' approach has to be avoided, and changes have to be limited as much as possible.

- Well known = proven?

Automotive manufacturers have experienced many problems and thus tend to be rather conservative. They expect as much innovation and improvement as possible with as little change as necessary. The best known approach to these conflicting targets is a standardized, modular concept, with changes to only one element at a time.

5.8.3.2 Package Requirements

Automotive applications impose numerous requirements on the packaging and assembly of sensors and microsystems. To deliver competitive products, these requirements have to be met with cost-effective solutions, standardized but inherently flexible.

5.8.3.2.1 Flexibility to Meet Customer Requirements

As described, standardization is an essential way to achieve high volume, high reliability products. But it always bears the problem of resolving conflicting standards. Therefore there is no sense in trying to convince customers of what is good for them. The inherent problem of differing interfaces (e.g., different connectors for every automotive manufacturer) cannot be solved in general; it is a prominent example of these conflicting standards. Consequently, the corresponding customer requirements have to be accepted and considered in the basic concept. Many other requirements have to be taken into account too, like acceptable weight, available volume, and desired appearance.

The ideal packaging and assembly concept has to be flexible to meet these varying customer requirements while keeping the necessary standardized concept so as to achieve the overall cost and quality targets.

5.8.3.2.2 Manufacturability

Manufacturing has to be fully automated, using simple, robust processes. Minimizing cycle time is a must. Thermal processes, like curing of glue, should be avoided or must be highly parallelised to achieve the typical short cycle time of pick-and-place operations.

The final test and calibrating requirements are often underestimated. Both can contribute to major parts of the total cost and therefore have to be an integral part of the manufacturing concept. A typical example is testing at different temperatures (low, room, and high temperatures). According to the high thermal masses, testing at different temperatures requires long heating and cooling times and can increase costs for testing and trimming considerably. Specific measures to simpli-

fy testing and trimming processes can pay off quite easily in system cost and performance.

5.8.4

Cost

5.8.4.1 General Considerations

Significant parts of the system and product cost are defined during the initial concept phase. This requires a high-level consideration of all system components and their effect on costs. This necessarily has to consider both direct and indirect implications. Basic options for packaging to be considered comprise

- **Packaging levels**
Although wafer-level packaging is increasingly used for microsystems, the next higher level of packaging is still important. Various approaches are possible: *chip on board* solutions promise a small footprint, attractive cost, and reduced interconnects by avoiding a complete level of packaging. But keeping one more packaging level and using e.g. mold packages has important advantages too. Obviously, testability, standardized interconnects, and standardized handling during subsequent manufacturing and processing are much easier with packaged chips.
- **Hermeticity?**
Hermetic packages are expensive, but the total costs associated with passivation of sensitive microsystems within nonhermetic packages may be even more expensive. The same applies to EMI problems, which are prevented by hermetic metal packages.
- **When to test & trim?**
'As early as possible' is not always the best solution. When the manufacturing process is robust and the yields are high, test-and-trim operations can be effectively combined later in the manufacturing sequence. Otherwise, testing has to be done early to avoid adding value to nonfunctional parts. The best approach is to combine test-and-trim steps with final testing (end-of-line trimming).
- **Batch processing?**
Parallel processing in larger batches seems to have an advantage over serial processing of single parts. But there is no general answer to this question – the best processing method depends so strongly on volume, batch size, yield, cycle time, and many other factors, that it has to be decided for each case.

5.8.4.2 Cost Drivers

In addition to the general considerations, which have to be assessed for each application separately, are several ever-occurring problems:

- Yield losses and failures caused by manufacturing processes that are not fully developed or not yet robust enough.
- Additional testing and inspections to assure quality requirements.

- EMI is always a problem in automotive applications, but especially for automotive sensors dealing with small signals. In trying to minimize costs, the expensive protective measures needed against EMI are often underestimated. Because overall EMI performance can be judged only very late in the development process, this can cause very late, and hence expensive, changes.
- Specifications
During early development the detailed specifications are not yet known. To minimize risk, this may lead to unnecessarily stringent electrical and mechanical specifications, which drive up both development and product costs.

5.8.5

Sensor Packaging Concepts

5.8.5.1 Modules

The first-level packaging has to be defined carefully and in consideration of the full system, including test-and-trim requirements. Figure 5.8.1 illustrates the principle relations between standardization, cost, and flexibility.

There are many reasons to use packaged chips, that is, modules, instead of a chip on a board:

- Module packages are robust.
- Handling is simple.
- The packages protect against the environment.
- The modules can be individually labeled and traced.
- Modularization enables standardization on a lower level of added value.

5.8.5.1.1 Plastic Mold Packages

Standardized plastic mold packages are the most cost-effective packages available. Packaging services are offered by large subcontractors, mainly located in South East Asia, which have established a worldwide standard for IC packaging. Therefore, flexibility for sensor-specific requirements is limited or even impossible, because existing manufacturing lines have to be used. Automotive quality is available, but not standard.

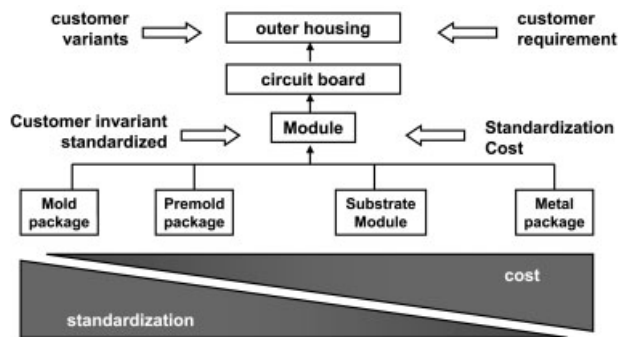


Fig. 5.8.1 General relationships between standardization, costs, and flexibility

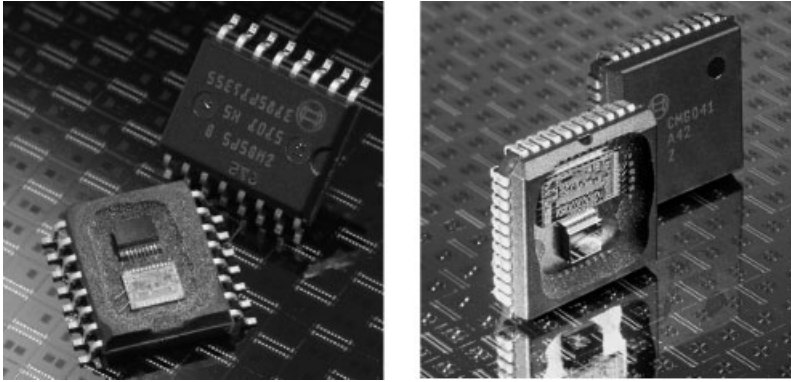


Fig. 5.8.2 Typical plastic mold packages as used for automotive sensors (left: accelerometer; right: angular rate sensor)

Mold packages are well established, easy to handle, and can be used on every printed circuit board assembly and placement line. If mold packages can be used (e.g., in the devices shown in Fig. 5.8.2), they offer the best potential for minimizing packaging cost. But plastic mold packages are designed for integrated circuits having purely electrical functionality. Therefore the micromechanical elements have to be protected against damage during mold packaging, for example, by using wafer-level packaging. When using standard mold packages for sensors and micro systems, a careful analysis of the potential effects and interactions of package and the device(s) to be packaged is essential. Well known problems are (1) package-induced stress causing parasitic behavior and thermal drifts, and (2) overstress due to excessive pressure during molding.

5.8.5.1.2 Premold Packages

Due to the nature of sensing intensive physical parameters (i.e., pressure, flow, etc.) the sensor has to have access to the environment to be sensed. Thus mold packages are inappropriate, but premolded packages may be an alternative (Fig. 5.8.3). Although more expensive than the standard mold package, the pre-molded version offers the required accessibility and retains many of the typical mold package advantages, especially when the external dimensions are kept standard.

5.8.5.1.3 Ceramic Modules

When molded or premolded packages cannot be used for whatever reason, for example, a need for hermetic sealing, modules using ceramic substrates (or PCB-printed circuit board) offer almost unlimited flexibility. They can be used for simple multichip modules containing two or more elements (e.g., a sensing element and an evaluation circuit), may include external components, or may be used to construct complete systems. Stress-optimized package configurations for sensitive microsystems and sensors are no problem (Fig. 5.8.4).

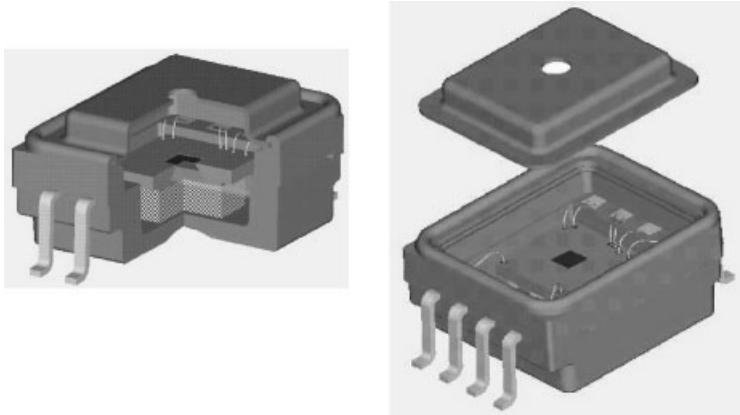


Fig. 5.8.3 Automotive pressure sensor using premold package

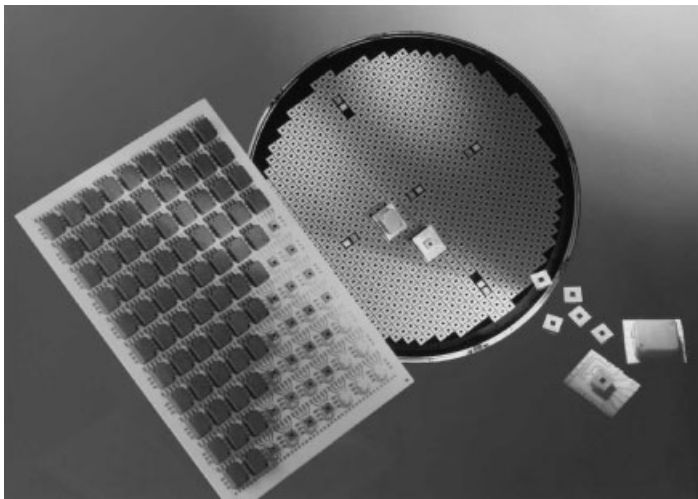


Fig. 5.8.4 Micromechanical pressure sensor with stress-optimized package on ceramic substrate

5.8.5.1.4 Metal Packages

For addressing special needs such as hermetic and vacuum enclosures, metal packages are an established alternative (Fig. 5.8.5). But from a cost perspective this can be justified only if other options are unacceptable.

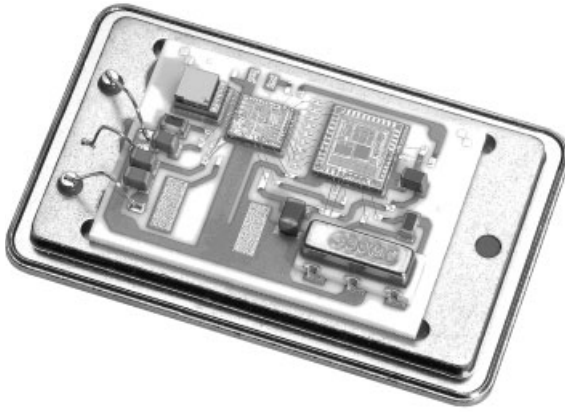


Fig. 5.8.5 Accelerometer within metal can package

5.8.6

Sensor Housing

5.8.6.1 Housing Requirements

Many automotive sensors, when packaged as described in Chapter 5.8.5, are incorporated into larger units, namely electronic control units (ECU) comprising electronic devices together with the necessary sensors. A good example is the central restraint system ECU, which contains the main accelerometer as well as all other electronics. But for the same restraint system, external sensors are used as well. For side-impact detection, additional accelerometers are placed at the bottom of the sidedoor pillar. These additional accelerometers are called peripheral acceleration sensors (PAS). In addition, up-front sensors (UFS) are mounted at the front of a car for early impact detection. In both applications, as in many others, due to the mounting requirements separate sensor housings are required.

A sensor housing contains a sensor module and a printed circuit board, which has established as worldwide standard for electronic circuits. The external inter-

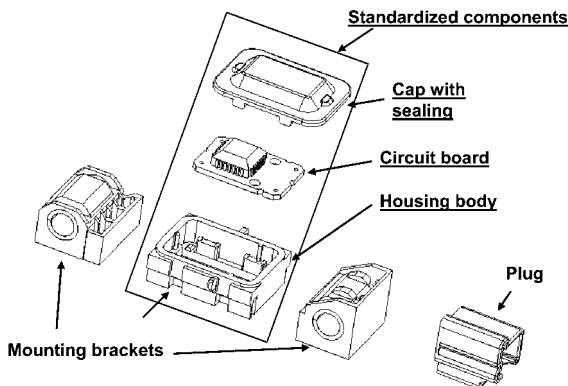


Fig. 5.8.6 Modular approach to constructing a sensor housing from standardized components and a custom-specific parts

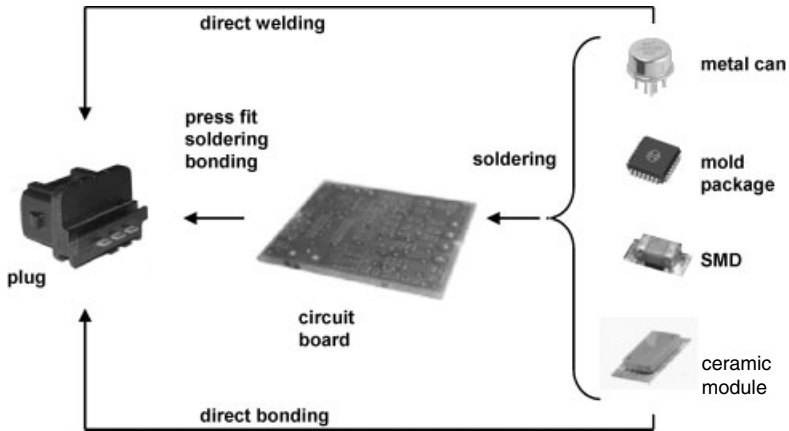


Fig. 5.8.7 General approach to the connection technology when constructing sensor housings

facing is made through a connector plug, which is almost always built as part of the housing itself. Figure 5.8.6 illustrates the modular approach to constructing an accelerometer housing with standardized components and customer-specific parts.

The standardized housing body contains the electronic module and is sealed with a cap. It has to be flexible enough to accommodate all sensor modules, including any additional functions, and should not need to be changed to accommodate individual parts, like the mounting brackets or the electrical interface, specified by the individual automotive customer.

A very general question within this modular concept is the choice of circuit board and connection technology. Each pathway in Figure 5.8.7 corresponds to a possible way of building a sensor. Metal packages can be welded directly onto metal inlay parts, and ceramic modules can be directly bonded. All the illustrated modules contain a printed circuit board (PCB).

5.8.6.2 Circuit Board

Printed circuit boards follow worldwide standards that are well established and cost effective. They offer unsurpassed flexibility, especially when considering circuit size and complexity as well as contacting concepts. PCB technologies and applications have been described in many textbooks.

For sensor applications, printed circuit boards are manufactured and assembled in larger sized batches. If possible, testing and trimming are also done in the same way. Finally, the PCB is divided into individual sensor modules. This approach is key to minimizing cost.

5.8.6.3 Housing Body

A well established material for electronic housings is polybutylene terephthalate (PBT) strengthened with glass fibers. It is a cost-effective plastic compound with

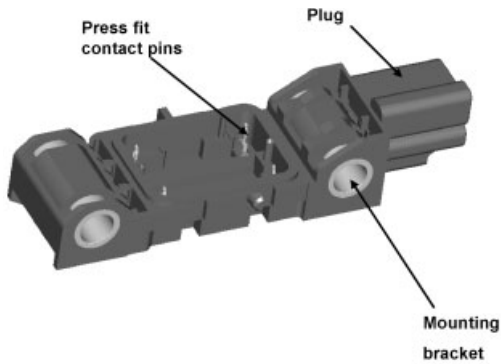


Fig. 5.8.8 Sensor housing made of PBT

good environmental resistance and can be used even within the engine compartment. The housings are manufactured by injection molding, which enables high flexibility in terms of forms and size (e.g., Fig. 5.8.8). Metal parts for the inner and outer connections can be incorporated into the housing. However, the metal inlays cannot be tightly sealed physically for media separation because of the metal/plastic interface. The manufacturing process is well known and housings are offered by numerous suppliers worldwide.

5.8.6.4 Cap and Sealing

To protect the electronic components, the housing has to be tightly sealed, especially when the sensor will be used in the engine compartment or anywhere outside the passenger compartment. Metal and plastic covers glued to the housing body are widely used. Another approach is to completely fill the housing, including the circuit board, with casting compounds. A new approach uses transmission laser welding (see Section 5.8.7.3 of this book). It can connect a housing body and cap made of the same plastic material and therefore can avoid potential problems caused by different thermal expansion coefficients. In addition, it eliminates any limitations due to the use of metal caps, that is, limited freedom for forming and corrosion.

5.8.7

Processes

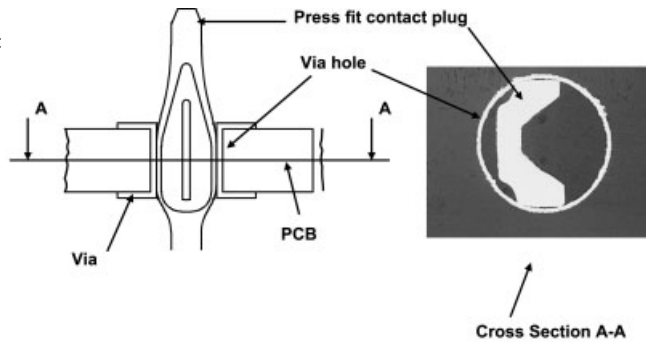
5.8.7.1 Placement and Soldering

Placement and soldering of components on a PCB are well established and are not explained here.

5.8.7.2 Press Fit Technology

A key element in cost-effective sensor housings is the electrical connection between the circuit board and the outer connector. Good connections can be made very easily using press fit technology, which can simultaneously make the electrical connection and position the circuit board inside the housing (Fig. 5.8.9). This is a dry cold process adding only minimal costs. When used properly, press fit plugs can be used as the mechanical fixture and not only the electrical connec-

Fig. 5.8.9 Cross section and top view of press fit plugs



tion. If required, additional electrically inert press fit plugs can be incorporated for the sole purpose of adding mechanical anchor points.

The electrical properties of press fit connections are at least equivalent to, if not superior to, soldered connections. Process quality can be monitored by creating a force/travel diagram during the press fit.

To limit the forces needed to make the connection it is necessary to apply a tin surface to the PCB or the plug surface, which can result in creation of tin slivers. This is a general quality concern, although it has not been observed to cause electrical shorts. Using combinations of tin on the PCB surface and nickel or nickel/gold on the plugs prevents this problem. Although the reason is not fully understood, slivers do not form when using these combinations of surfaces.

5.8.7.3 Transmission Laser Welding

A new process for tightly sealing housings is transmission laser welding. This process takes advantage of the differential absorption of laser radiation by the two pieces to be joined. The absorption coefficients can be adjusted by using dyes of various colors and concentrations in the materials used. Usually the sensor cap is transparent, and the housing absorbs the laser beam, thus bringing the radiation energy to the intended interface. The housing is directly heated, and the cap indirectly. Mirrors guide the laser beam so that it is scanned along the circular welding path (Fig. 5.8.10).

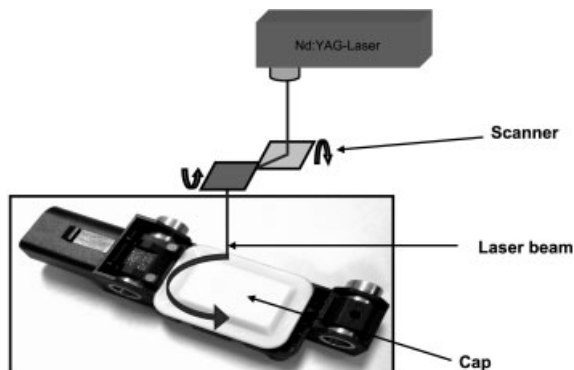


Fig. 5.8.10 Overview of transmission laser welding

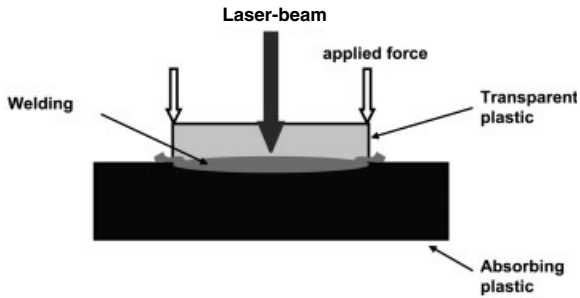


Fig. 5.8.11 Detailed diagram of transmission laser welding

This method is called quasisimultaneous, because the laser beam scans at frequencies of several tens of Hertz, which is fast enough to almost-simultaneously melt the material all along the laser path. The pieces being joined can also be pressed together (Fig. 5.8.11), thereby squeezing out excess material. This eliminates the problems of gaps and nonuniform start/end points associated with classical laser welding.

The whole process can be done within seconds and with freely programmable laser scanning patterns. This makes this process ideally suited for automotive applications.

5.8.7.4 Sensor-Specific Processes

Besides the standardized approach described above, other processes for packaging electronics are available. When using different processes, because of special needs of the sensor or the application, care has to be taken to avoid unwanted interactions with the sensor itself. Existing processes usually have to be modified. In general, using nonstandardized approaches to sensor-specific processes should be restricted whenever possible.

5.8.8

Summary

A modular approach to packaging automotive sensors has been described. It uses standardized processes and components, enabling high flexibility and quality at minimum cost. In general, automotive sensors made according to this approach

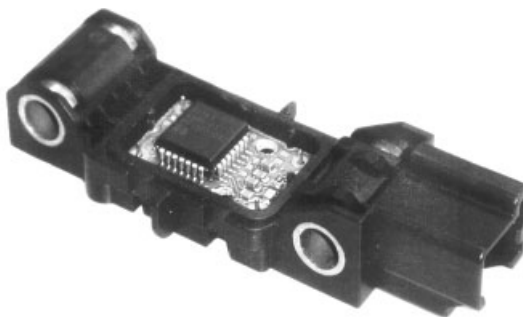


Fig. 5.8.12 Acceleration sensor

- are sealed and resistant to external conditions
- can be used within the engine compartment
- can be manufactured in a highly automated process at cycle times of only seconds
- are flexible, robust, and cost-effective

Figures 5.8.12–5.8.15 show several examples of automotive sensors constructed according to modular concepts. Often, different sensors can even be manufactured in the same manufacturing line.

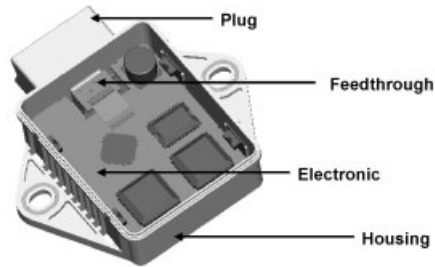


Fig. 5.8.13 Angular rate sensor

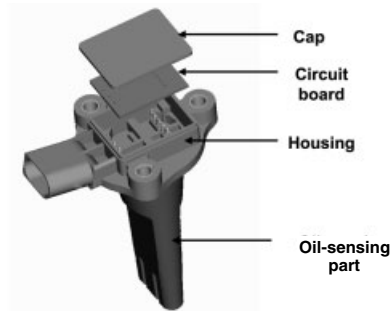


Fig. 5.8.14 Oil condition sensor

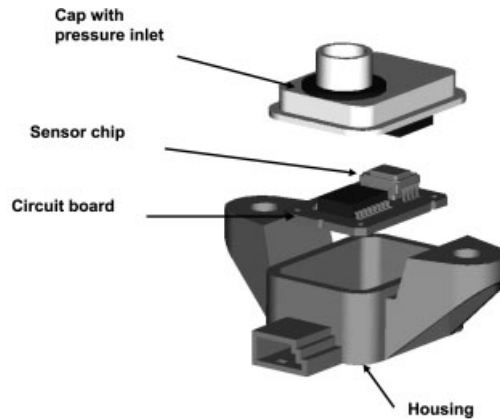


Fig. 5.8.15 Peripheral pressure sensor

5.9

Reliability

ROLAND MÜLLER-FIEDLER

5.9.1

Introduction

In modern cars, microsensors play an essential and increasingly important role as the interface between the vehicle, with its complex functions of motor management, chassis systems, safety, as well as comfort and convenience on the one hand, and the respective electronic control units on the other hand. Microsensors need to function in a harsh environment over the entire lifetime of an automobile. Therefore reliability aspects have come more and more into the focus of engineering and research activities in microsystem technology.

By definition, reliability is the sum of all characteristics of a device relating to its ability to meet specified requirements under well-defined conditions over a given period of time [1]. Although the major concern of quality is to assure a high yield and proper functioning of the product with respect to customer specifications at the time of product release, reliability is a constituent of quality and describes changes in quality over time. The major parameter to qualify and quantify the reliability of a device to give a numerical definition is the mean time to failure (MTTF). In consequence, the main objective of all efforts in microsystem technologies to improve reliability is to increase the lifetime of MEMS.

This section describes a methodical procedure that allows reliability issues to be approached efficiently. MEMS reveal specific reliability aspects, which differ considerably from the reliability issues of integrated circuits and macroscopic devices. A classification of typical MEMS-failure modes is given, as well as an overview of lifetime distribution models. The extraction of reliability parameters is a 'lack of failures' situation using accelerated aging and suitable models. In a case study, the implementation of the methodology is illustrated with a real-life example of dynamic mechanical stress on a thin membrane in a hot-film mass-airflow sensor.

5.9.2

Reliability as a Key Parameter for MEMS Development and Market Success

The implementation of microsystems in automotive applications is certainly one of the driving forces for the breakthrough of MEMS as an industrial technology on a mass-production level. During the early development of microsystem technologies, the emphasis was mainly on processes, components, and sensor functions. However, many of their applications in the automobile constitute distinct safety features. Lives of passengers and pedestrians depend on the proper functioning of steering, brakes, and airbags. Consequently, it is crucial that microsensors assure accurate, reliable, failsafe operation during an automobile's lifetime, which may be 15 or even 20 years.

The environmental conditions for microsystems in automotive applications are harsh. The sensors have to endure temperatures from -40°C to $+85^{\circ}\text{C}$ in the pas-

senger compartment, to +125 °C under the hood, and to +140 °C if mounted on the engine. During storage and transportation, temperatures can reach -60 °C. In addition to static long-term low and high temperatures, automotive equipment has to withstand thermal cycling under shock-type conditions and relative humidities of up to 85% at 85 °C.

Therefore, for commercial suppliers, reliability issues have come more and more into the focus of engineering and research activities. Reliability has become a key parameter for the success of MEMS in a market with distinct safety implications [2–7].

5.9.3

Reliability Methodology

5.9.3.1 Implementation of Reliability in the Development Process

In addition to safety aspects, cost considerations are a driving force behind the goal of achieving reliability. Similar to the situation in the beginning of microtechnology development, the early phases of product development are often dominated by considerations of design, functionality, producibility, and costs. Reliability issues are often postponed to later stages. But the expenses of correcting problems overlooked in an early phase increase drastically as product development moves onward. Feedback about indispensable corrective measures concerning de-

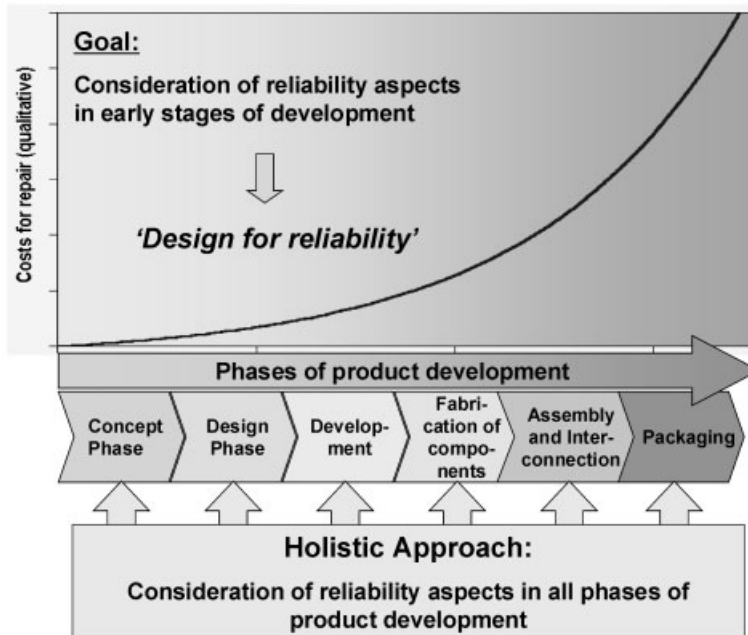


Fig. 5.9.1 Holistic approach to considering reliability in all phases of development

sign, materials, and processes is more valuable the earlier it is communicated to the engineering team. In the very early stages, supplementary improvements can still be initiated and incorporated into the product design at relatively low costs.

An essential goal is therefore to consider reliability aspects from the very beginning of a development project and to include it already in the concept and design phases ('design for reliability'). This holistic approach to increasing sensor lifetime is based on the idea of considering reliability aspects in all phases of product development (Fig. 5.9.1).

This approach requires an integrated process comprising networked, interdependent collaboration among engineers with broad areas of background knowledge. A high level of broad expertise covering diverse scopes from modeling and simulation, design and layout, process technology, high-volume manufacturing, assembly, and packaging through characterization and stress testing is necessary to accomplish the 'design-in' of reliability into the product. The various processing steps in the different phases have to be checked with respect to their mutual interaction to achieve concurrent, comprehensive optimization.

In the various phases, distinct reliability aspects must be taken into account. In the *concept phase*, for example, the basic behavior of movable structures like membranes, suspended beams, and springs must be carefully taken into consideration. Initial decisions on the feasibility of candidate approaches to assembly and interconnection technologies are also needed in this early stage of development. In the *design phase*, basic material properties such as Young's modulus, electrical and thermal conductivity, fracture toughness, breakdown voltage, coefficient of thermal expansion (CTE), and residual stress, as well as microstructure-related geometry parameters such as layer thickness, width of structures, gap size, and curvature, have to be assessed not only with regard to the function of the component but also their influence on long-term stability. The design 'space' may be reduced considerably by implementing reliability aspects. During *development and fabrication* of components, tests have to be performed to obtain initial experience and data about the impact of such environmental parameters as temperature, humidity, and mechanical shock on critical structures. Characterization procedures have to be designed and implemented to facilitate the study of distinct features in this early stage of product life cycle. The *assembly and interconnection* phase has to take into account that combining various materials can cause additional problems due to mismatches between coefficients of thermal expansion (CTE), which in turn may induce thermomechanical stress and thereby limit both functionality and durability. Thorough data collection and documentation is necessary to provide a complete database of the results of tests and experiments in all development phases.

The key to success is a methodical approach, which basically comprises the following steps:

- Determination of the significant material and design parameters affecting the lifetime of a microsystem by:
 - experimental verification and
 - simulation (finite element method [FEM], analytical).

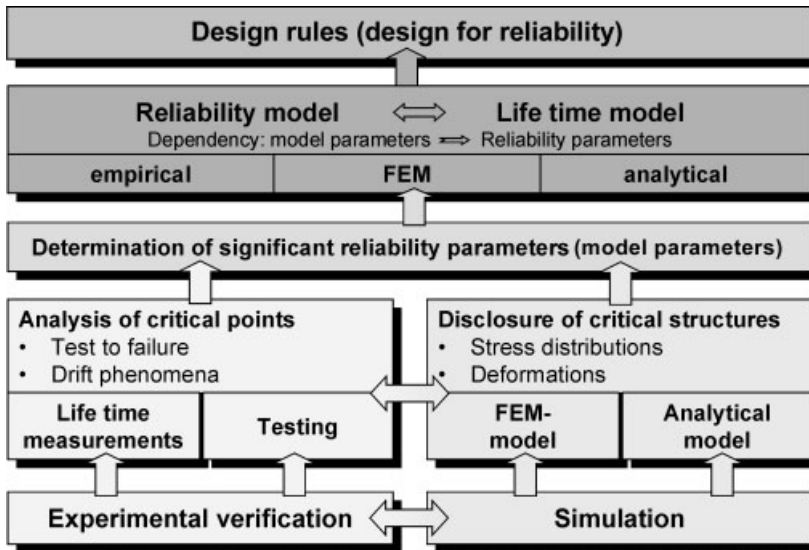


Fig. 5.9.2 Reliability methodology

- Understanding the physical mechanisms behind failures.
- Set up of reliability models.
- Deduction of design rules for increased reliability in current product development as well as in future generations of the product.

The methodology for designing in reliability is shown in Figure 5.9.2.

5.9.3.2 Specific Reliability Aspects of MEMS

The fabrication processes of microsystems are to some extent based on processes established for high-volume production of integrated circuits. Therefore, many reliability issues concerning MEMS are already well known from IC manufacturing. But the scope of MEMS reaches far beyond the utilization of IC processes. The whole range of micromachined mechanical elements such as membranes, movable beams, and comb structures poses new challenges on the assurance of reliability. New concepts are required to investigate and eliminate MEMS-specific failure mechanisms, which have not yet been subjected to testing and qualification procedures in IC production.

The reliability aspects of MEMS also reveal some basic differences in comparison to lifetime considerations for macroscopic devices. The ratio of surface to volume increases linearly with additional miniaturization, resulting in forces that eventually are no longer dominated by mass but by surface effects. Friction and sticking are therefore typical problems of microstructures, which have to be considered in the design.

Lifetime issues of microsystems can be roughly divided in two major fields: issues of the microstructure itself and issues related to assembly and packaging. Re-

liability aspects concerning microstructure include, for instance, inherent mechanical and electrical material properties, but also include effects caused by the combination of different materials in a common device. For example, mismatches between coefficients of thermal expansion can result in mechanical stress, and chemical reactions at interfaces can lead to oxidation of metal layers. The design of movable structures such as membranes and springs has to take into account the maximum deflection and the possibility of surface contacts, which can result in sticking, jamming, or fracture [8].

Assembly and packaging have to protect the microstructure from environmental influences such as humidity and corrosive atmospheres. Often, a very low pressure has to be maintained within an enclosed space in the package and must be maintained below a certain level over the entire life of the system. In addition, the development of assembly and packaging approaches has to consider that these are the final steps in the production process, which can influence the already completed microstructure in an unpredictable way. Temperature range, media pressure, and mechanical load during the packaging process have to be selected carefully to avoid micro damage, which can gradually grow over time and cause subsequent failure of the complete system.

The manufacturing processes represent the link between the reliability issues of the microstructure on one hand and those of assembly and packaging on the other. The process flow must take into account, for example, the impact of the applied temperature range, pressure, and media on the sensor structure, which at the time of packaging is already completed. Dicing of sensor structures that include delicate elements, such as freestanding cantilevers or thin membranes, is usually very critical and requires some kind of protection which can be effectively provided by zero-level packaging on the wafer scale.

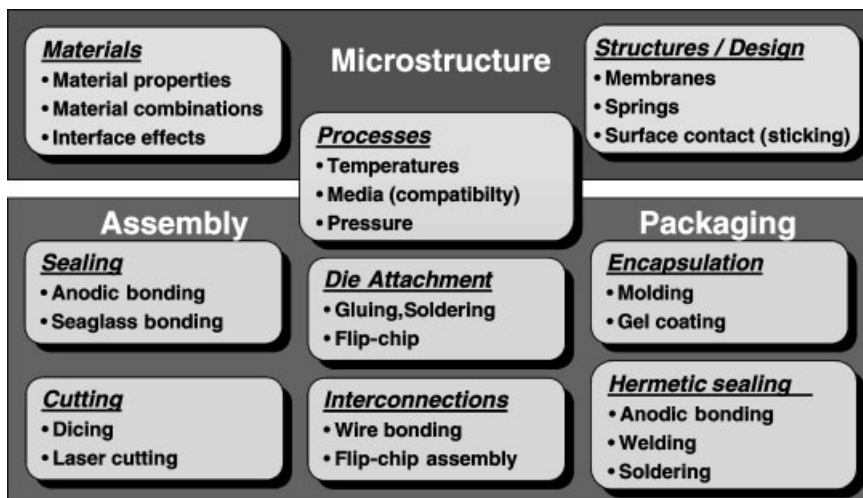


Fig. 5.9.3 Reliability aspects of microsystems

Fig. 5.9.4 Schematic illustration of the degree to which the relations between functional parameters and model parameters are understood

		Functional parameters			
		Sensor characteristic		Reliability parameters	
Model parameters	Geometry parameters	+	+	+	-
	Material parameters	+		-	-

An overview of the reliability aspects of microsystems is given in Figure 5.9.3.

In Section 4.1 an approach for integrated design flow was introduced, which not only focuses on the required functionality of the system, but also takes into account the demands of testability, reliability, and the entire process flow in an early stage of system design. The basic idea behind this approach is to establish a model-based relationship between process monitors (model parameters) and functional parameters, which comprise the characteristic curve, reliability parameters, and cost aspects. The benefit of this design method relies highly on a thorough understanding of the dependency of the functional parameters on the model parameters. Although the influence of geometry and material parameters on the sensor's characteristic curve is well understood in most cases and can be described with simulation tools, much less is known about the dependency of lifetime on the same parameters (Fig. 5.9.4).

In particular, the question of how material parameters affect the reliability of micro components or systems poses a serious challenge to the design and process flow. Extensive work is needed in the near future to gain a deeper understanding of basic material properties. Characterization of the creeping behavior, fracture strength, Young's modulus, the coefficient of thermal expansion, etc. is needed to build up a broad knowledge base of important material parameters, which will enable the system designer to extract the necessary data for a dependable estimation of reliability parameters.

5.9.3.3 Classification of Failures

For sensors in the automobile environment, mechanical shock load, humidity, elevated temperature, and temperature shock are the main causes of failure. Table 5.9.1 gives an overview of typical failure issues that are most common in micro-sensors and which need to be considered in reliability studies.

It is worth mentioning that 'failure' does not necessarily mean the total breakdown of a device, as occurs with mechanical fracture or an electrical short circuit. Failure comprises all mechanisms that result in a state of operation in which the system no longer meets the required specifications.

To gain insight into the physical, chemical, and electrical factors that can cause a component or system to fail, we need to take a closer look at some characteristic properties relating to the occurrence of failures. Asking 'when, how, and why'

Tab. 5.9.1 Classification of failures

<i>Mechanical</i>	<i>Electrical</i>	<i>Electro-mechanical</i>	<i>Thermal</i>	<i>Thermo-mechanical</i>	<i>Chemical</i>
<ul style="list-style-type: none"> • Shock type loads • Resonant vibration • Membrane fracture • Spring fracture 	<ul style="list-style-type: none"> • Charging • Dielectric breakdown • Short circuit • Open circuit 	<ul style="list-style-type: none"> • Resistor drift • Offset drift • Electro-migration • Stiction 	<ul style="list-style-type: none"> • Material softening • Material relaxation • Creeping 	<ul style="list-style-type: none"> • Mismatch in coefficients of thermal expansion • Build up of mechanical stress 	<ul style="list-style-type: none"> • Corrosion • Oxidation due to humidity

yields valuable information about some ‘fingerprints’ that are diagnostic of categories of failures. The answers can be roughly classified as follows:

- When?
 - Early failures or infant mortality: manufacturing defects, foreign particles, material defects
 - Midlife or premature failures: roughly constant failure rate due to random factors
 - Wearout period: increasing failure rate in the last stages of life due to growing defects and/or material fatigue
- How?
 - Sudden or catastrophic failures
 - Gradually growing failures
- Why?
 - Internal mechanisms resulting in predetermined failures
 - External causes leading to random failures

Some combinations of answers to these three questions rarely occur in the life of a product, but others are specific for certain failure modes and can therefore give valuable hints as to the underlying mechanism. A good visualization of this approach is the ‘reliability triangle’ [1], in which solid lines indicate typical combinations of when, how, and why (Fig. 5.9.5).

A closer look at the three stages of product life when failures occur, and how often failures occur during these stages, reveals an interesting feature of the dependency of failure rate on time, which is typical of most mass-produced items. In the first stage, defects, foreign particles, and manufacturing or handling failures result in an ‘infant mortality’ behavior, with a relatively high failure rate at time zero that decreases rapidly with age of the product. This region lies in the domain of quality assurance, and defective parts are detected and eliminated with suitable screening procedures, for example, burn-in. More details on the field of test procedures are described in Section 5.8.

In contrast, the midlife stage has a constant failure rate, because failures mainly occur randomly due to external circumstances. This period is often re-

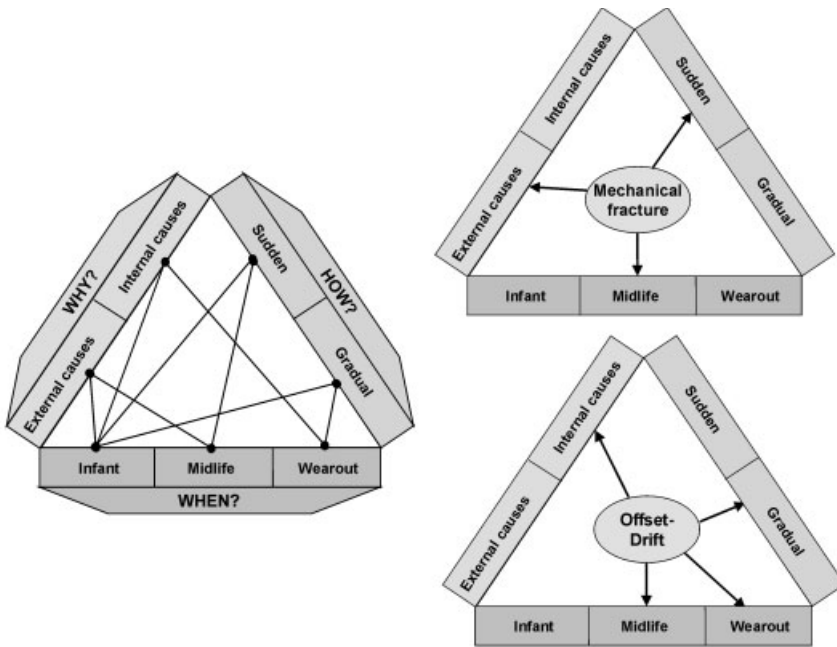


Fig. 5.9.5 Reliability triangle [1] showing two specific examples: sudden failure and offset drift

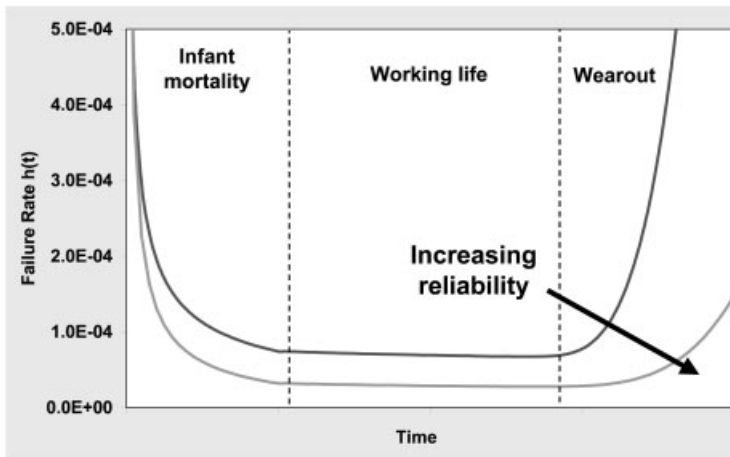


Fig. 5.9.6 Schematic bathtub curve of failure rates over time; with increasing reliability, the curves are lower and less steep at the end of a product's life

ferred to as the working life of a system or component, since the most systems spend most of their lifetimes in this stage. Therefore, the main focus of reliability activities is to further decrease the failure rate during this time period.

In the last stage of life, failures are dominated by wearing out of materials, which results in a continuously increasing failure rate. This cannot in principle be avoided, so the reliability issue is to shift the beginning of this stage to a later time by using proper materials and suitable material combinations and by choosing geometrical design parameters, which take into account, for example, the velocity of crack propagation or the kinetics of chemical reactions. All together, the behavior of the failure rate over the entire lifetime of a device, which can be observed by investigating a large number of identical devices, shows a typical shape resembling a bathtub (Fig. 5.9.6).

5.9.4

Quantitative Characterization: Lifetime Distribution Models

Quantitative descriptions of reliability are useful only in the context of these boundary conditions (Fig. 5.9.7):

- What is the purpose of the system?
- What is the required duration of the purpose?
- What are the typical conditions under which the system must function?

The goals of a quantitative description of reliability are *determination of the failure rate* during the working lifetime and *prediction of the characteristic lifetime* of a system or component.

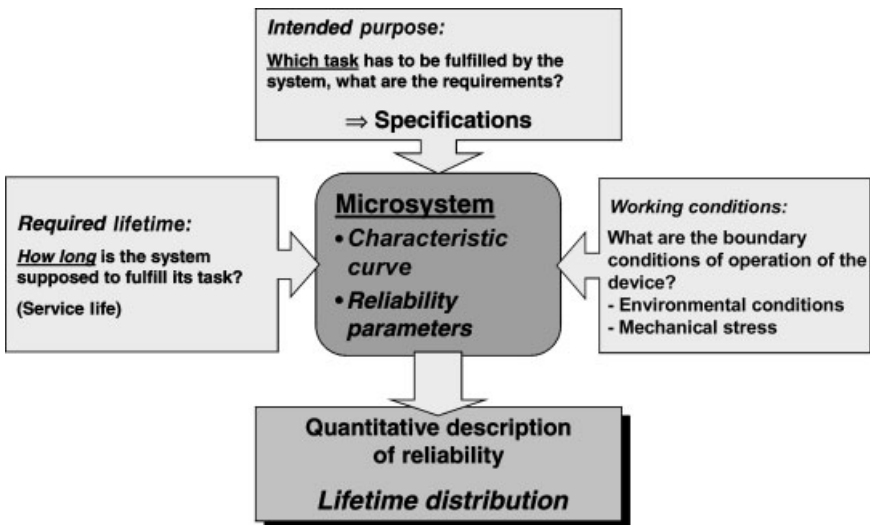


Fig. 5.9.7 Basic considerations for a quantitative description of system reliability

Tab. 5.9.2 Probability functions for reliability calculations

Reliability function $R(t)$ (survival function)	Cumulative distribution function $F(t)$ (DCF)
$R(t)$	$F(t) = 1 - R(t)$
<ul style="list-style-type: none"> • Fractin of units that survived to time t • Probability for a unit to survive to time t 	<ul style="list-style-type: none"> • Fraction of units that failed up to time t • Probability for a unit to fail up to time t
$R(t) = 1 \rightarrow 0$	$F(t) = 0 \rightarrow 1$
for $t = 0 \rightarrow \infty$	for $t = 0 \rightarrow \infty$
Probability density function $f(t)$ (PDF)	Failure rate $h(t)$ (hazard function)
$f(t) = \frac{dF(t)}{dt}$	$h(t) = \frac{f(t)}{R(t)}$
Fraction of original population that failed between time t and $t+dT$	Fraction of remaining population at time t that failed between time t and $t+dT$

Therefore, probability functions describing the lifetime distribution have to be defined; they allow the future behavior of devices to be calculated based on data from experiments performed on a limited population. Table 5.9.2 lists the most frequently used probability functions for reliability calculations and gives brief definitions, as well as the range of values and their principle behavior over time.

When the probability density function $f(t)$ is known (or $R(t)$ or $F(t)$), the mean time to failure MTTF, which is the characteristic lifetime, can easily be calculated as

$$\text{MTTF} = \bar{T} = \frac{\int_0^{\infty} t \cdot f(t) dt}{\int_0^{\infty} f(t) dt}$$

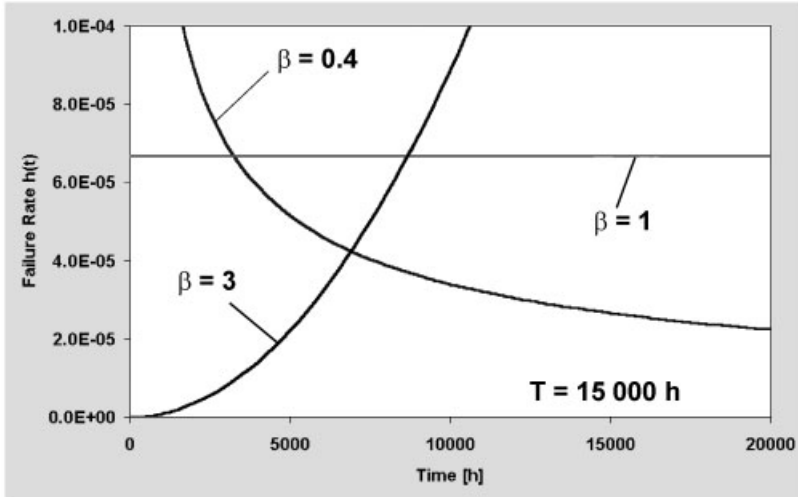
Therefore it is a basic concern of reliability to determine the mathematical form of $f(t)$ or $F(t)$. A few fundamental functions are available in the literature which serve as lifetime distribution models and which in many cases can describe the time behavior of failures by adjusting one or more parameters. An overview of these distribution models is given in more detail in [9, 10]. As one of these models, the Weibull distribution is a powerful tool that often allows a very good description of the actual reliability functions by fitting two parameters. Table 5.9.3 lists the various functions $R(t)$, $F(t)$, $f(t)$, and $h(t)$ for the Weibull distribution.

The Weibull distribution is completely described by the shape parameter β and the characteristic time T . We should mention that T is related to but not identical to the mean time to failure (MTTF). T is actually the time by which 63% of the original population fails. Figure 5.9.8 shows how the time dependence of the failure rate changes with the shape parameter at constant T .

With shape parameter $\beta < 1$ a decreasing failure rate can be simulated, insuch as on the left side of the bathtub curve (Fig. 5.9.6). For $\beta = 1$ the Weibull distribution reduces to a pure exponential with a constant failure rate that is given by the reciprocal of the characteristic time T .

Tab. 5.9.3 Probability functions of the Weibull distribution

Reliability function $R(t)$	Cumulative distribution function $F(t)$
$R(t) = e^{-(t/T)^\beta}$	$F(t) = 1 - e^{-(t/T)^\beta}$
Probability density function $f(t)$	Failure rate $h(t)$
$f(t) = \frac{\beta}{T} \cdot \left(\frac{t}{T}\right)^{\beta-1} \cdot e^{-(t/T)^\beta}$	$h(t) = \frac{\beta}{T} \cdot \left(\frac{t}{T}\right)^{\beta-1}$

Fig. 5.9.8 Weibull distribution for three different shape parameters β

$$h(t) = \text{const} = \frac{1}{T}$$

For an exponential distribution function, the mean time to failure and the characteristic time T are identical. When $\beta=1$ the Weibull distribution represents the region of the working life. For $\beta > 1$ the failure rate increases with time, as when the product is wearing out.

The common way to plot a Weibull distribution is based on the fact that the $\ln(\ln(F(t)))$ is linearly dependent on the logarithm of time t :

$$F(t) = 1 - \exp\left(-\left(\frac{t}{T}\right)^\beta\right) \rightarrow \ln(\ln(F(t))) = \beta \cdot \ln(t) - \beta \cdot \ln(T)$$

β is the slope of the curve, and T is the time where the straight line cuts the 63% level (Fig. 5.9.9).

The Weibull distribution is a pure mathematical function, but in many applications two parameters are sufficient to describe the actual time dependency of a

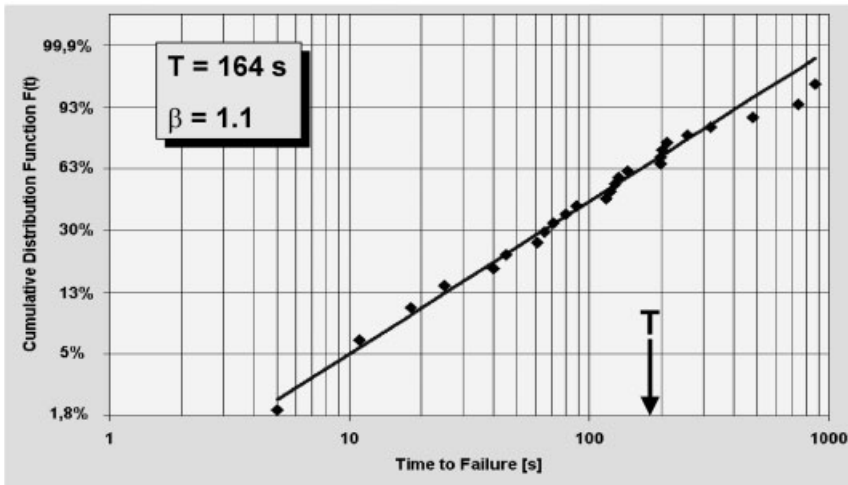


Fig. 5.9.9 Typical Weibull distribution for $\beta=1.1$ and $T=164$ s (25 samples)

system with this empirical model. The shape parameter and the characteristic time depend on the failure mode. The goal of improving reliability is to increase T in the region where $\beta \approx 1$.

5.9.5

System Reliability

The description of failure modes with reliability functions and their respective parameters is, in the first place, suitable for individual microstructures which are the primary source of failures and which fail in total. The corresponding reliability features in a complete complex system, which consists of many subsystems, components, and structures, can be deduced in several steps by breaking the system down to its basic building blocks as diagrammed in Figure 5.9.10.

Then, the functionality of the system is described using series and parallel arrangements of the underlying building blocks. Employing a bottom-up approach, in each of these models the reliability functions on a certain level of complexity are mathematically linked to the components in the next sublevel as outlined in Figure 5.9.11.

In this way, system reliability can be described and assessed by the successive buildup of basic building blocks in series or parallel arrangements. The desirability of implementing parallel designs is largely a matter of cost. It is obvious that in the series model the overall reliability is dominated by the weakest link of the system. On the other hand, redundancy is a distinguished feature of the parallel model, generally leading to higher reliability and lower failure rates.

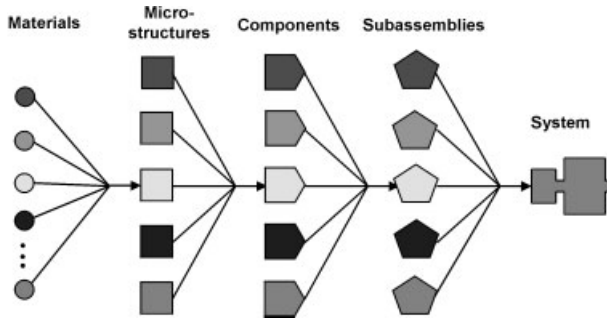


Fig. 5.9.10 Breakdown of a system to its basic building blocks

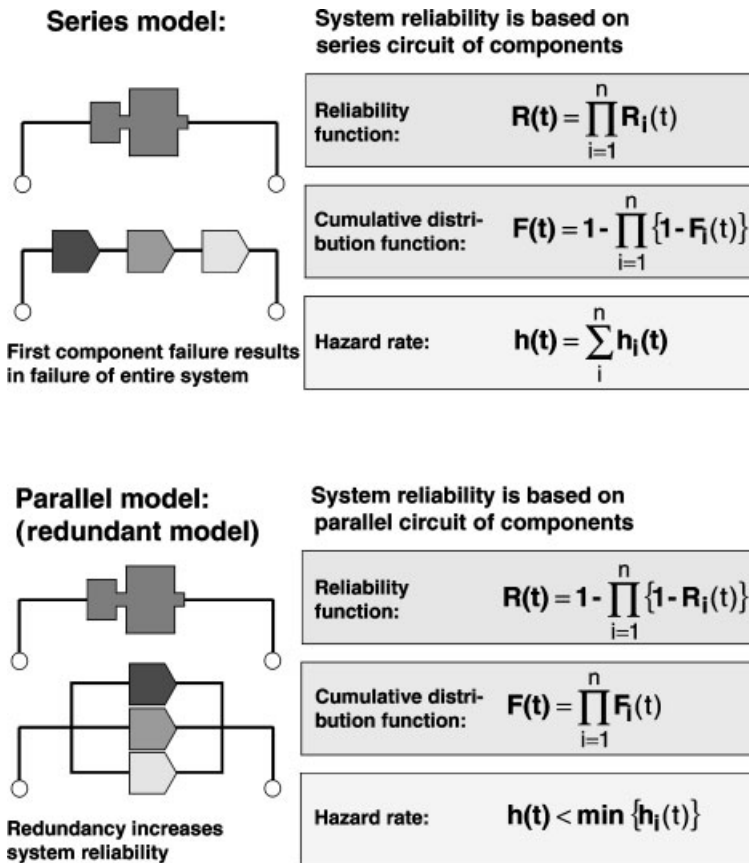


Fig. 5.9.11 Deduction of system reliability functions from the underlying component functions

5.9.6

Extraction of Reliability Parameters

The reliability parameters, such as the mean time to failure, have to be determined in experiments under well defined conditions. Failure rates of microsystems for automotive applications are typically in the range of a few ppm (parts per million). This may sound negligible, but due to the large number of sensors sold every year and their increasing numbers in each car, even this failure rate must be decreased further. However, the engineer who tries to investigate failure mechanisms is confronted with the problem of 'lack of failures' in the sense that he finds too few defective samples for a thorough failure analysis. Thus, due to the lack of a statistical basis, the quality of lifetime predictions under normal in-use conditions would be poor.

Therefore, failures have to be provoked artificially by exposing the sensor to increased stress. Methods for accelerated aging include application of these types of stresses:

- Mechanical
 - vibration
 - shock type load
- Thermal
 - constant high temperature
 - thermal cycling
 - thermal shock

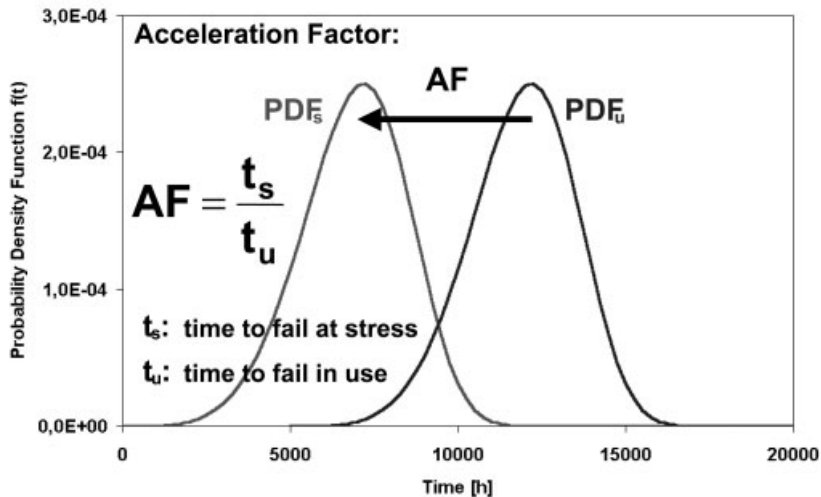


Fig. 5.9.12 Effect of accelerated aging and definition of acceleration factor AF

- Electrical
 - voltage
 - current
 - electromagnetic interference
- Chemical
 - humidity
 - corrosive atmosphere

Experiments for accelerated aging have to be conducted under well defined conditions. The age-acceleration factor is defined as the lifetime in use (t_u) divided by the lifetime under stress (t_s) (Fig. 5.9.12).

Proper design, performance, and analysis of the experiment should facilitate the investigation of a single MEMS failure mode by provoking a specific underlying failure mechanism without interference from other sources.

Suitable models have to be developed so as to extrapolate from the experimental results to nominal real-life conditions by calculating the age-acceleration factor that relates the lifetime under experimental stress to the in-use lifetime. As with reliability functions, various models that apply to the different age-acceleration methods are described in the literature: [1, 9]

- Thermal stress at constant temperature, Arrhenius model:

$$AF = \frac{t_1}{t_2} = \exp \left[\frac{E_A}{k} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \right]$$

- Thermal cycling between temperatures T_1 and $T_2 = T_1 + \Delta T$, Coffin-Manson model:

$$N_f = A \cdot \frac{1}{f^\alpha} \cdot \Delta T^\beta$$

Here f is the cycle frequency, N_f the number of cycles to failure, and ΔT the temperature difference. The Coffin-Manson model is widely accepted for describing material degradation (e.g., solder materials), fatigue fracture, and crack propagation.

- Thermal, electrical, chemical, and other stresses, Eyring model:

$$T = A \cdot T^a \cdot \exp \left(\frac{E_A}{kT} \right) \cdot \exp \left[\left(B + \frac{C}{T} \right) S_1 \right] \cdot \exp \left[\left(D + \frac{E}{T} \right) S_2 \right] \cdot \dots$$

The first factor holds for thermal stress similar to the Arrhenius model, and the following factors are suitable for describing any kind of stress when appropriate values are chosen for the nonthermal parameters B and D and for the parameters C and E , which describe temperature interaction effects. a is a constant, characteristic for the material.

It is often impossible to determine the appropriate age-acceleration factor to use for quantitative extrapolation from the experimental results to real life conditions.

In these cases the experiments must be conducted under well defined, reproducible conditions so that relative comparisons are possible and to enable a sound prediction of improved reliability.

5.9.7

Real-life Example

The approach diagrammed in Figure 5.9.2 has been used in a real-life reliability problem concerning the dynamic mechanical stability of a thin-film membrane. The membrane is part of the sensing element of a hot-film mass-airflow sensor. It consists of several dielectric layers with an overall thickness of about $1.5\ \mu\text{m}$ and a thin-film metallization that forms heating elements and thermosensors on a silicon chip. The schematic layout of the sensor is shown in Figure 5.9.13; its principle of operation is described in more detail in Section 7.6.

The sensor is installed in the air intake manifold of the car, measuring the aspirated air mass. The measuring principle requires the membrane to be exposed to the airflow and thereby also to dust and other particles, which are either not removed by the air filter or are inherent in the intake pipe. The impact of such particles on the membrane imposes a shock-type mechanical load, which can cause field failures as well as 0-km failures due to membrane fracture. The task was to identify the relevant geometry and material parameters limiting the lifetime of the sensor element and to deduce a model for their effects on membrane stability [11].

A full-scale model of the air intake manifold, where the complete sensor (with signal processing ICs, housing, cable fittings, etc.) can be tested under real-life conditions, was available. However, failure rates of components in automotive applications under normal conditions are on the order of a few ppm, which corre-

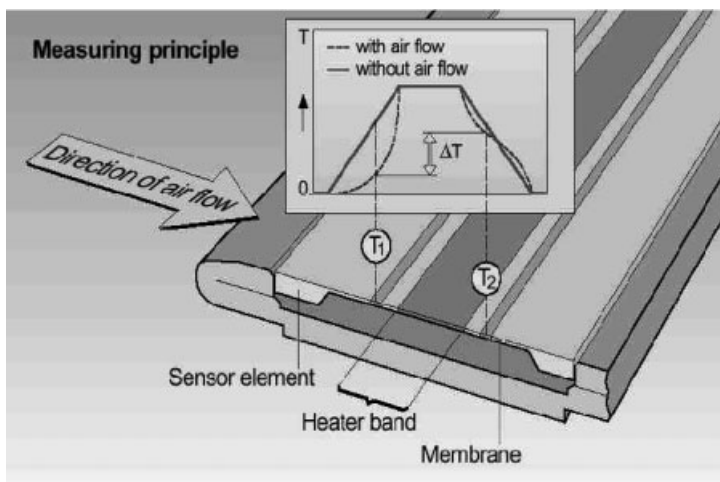


Fig. 5.9.13 Schematic layout of mass airflow sensor: Schematic layout

sponds to a useful life of close to 10^9 hours. Therefore accelerated aging is mandatory for obtaining enough data for statistically significant conclusions within moderate times at reasonable cost. To study the cause for failure due to particle impact, the aging process can easily be accelerated by exposing the sensor to a defined air stream and introducing dust particles into the air at a constant rate by means of a dispersion unit.

Designing the experimental setup requires some understanding of the failure mechanism, because the acceleration must affect the failure mechanism in a well defined way and should not be affected by, for example, the general material behavior. The experimental setup must allow for individual adjustment of the relevant experimental parameter as well.

Ideally, one parameter at a time is changed in both experiment and simulation, facilitating conclusions as to that parameter's influence. However, in this example such an ideal experiment is impossible because the shape and mass of the dust particles vary over a wide range, and the damaging particle cannot be back traced. To overcome this inherent variability, a huge statistical basis is needed, which directly translates to experiments that are not only expensive but also time-consuming. In addition, exact simulations by means of FEM are virtually impossible. An alternative is to use uniform particles. Since the aim of the study is to understand the effects on reliability and not to faithfully reproduce an in-use failure, this is not a serious restriction. Therefore 'standardized' particles with ideal spherical shape and a very narrow size distribution were used.

Another drawback of a realistic mockup is that preparing complete sensors is extremely costly. Since only the behavior of the MEMS component is of interest, we decided to build a test bed that enabled unpackaged membrane chips to be tested. For the severest possible shock load, the membrane was positioned vertical to the air and thus to the particle stream. The intensity, velocity, and size of the particles could be adjusted individually to separate the effects of energy and momentum.

A series of fully functional sensors were tested until failure (defined for this application as membrane fracture). A sample of the recorded survival times is shown in Figure 5.9.14.

A huge scatter in survival times is apparent. Thus, cumulative damage seems an unlikely mechanism, and statistical methods are needed to evaluate the nature of the failure. For this, the data were plotted in a Weibull diagram (Fig. 5.9.15). Since the data points can be approximated by a straight line, a Weibull model can be used to describe the reliability parameters. Because $\beta \cong 1$, the failure rate is constant, and total failure is a random event. With aging effects thus ruled out, the event causing the damage must be singular.

Although the Weibull model allows qualitative description of the sensor reliability and thus the comparison#Q15# to other variants, more specific experiments were needed to identify and understand the parameters of the failure mechanism itself. Thus, dedicated experiments, accompanied by simulations, were designed. With this setup, it became feasible to test several membranes differing in thickness, layer composition, and length/width ratio in a reasonable amount of time. The effects of particle size, velocity, and intensity#Q16# on lifetime were also investigated.

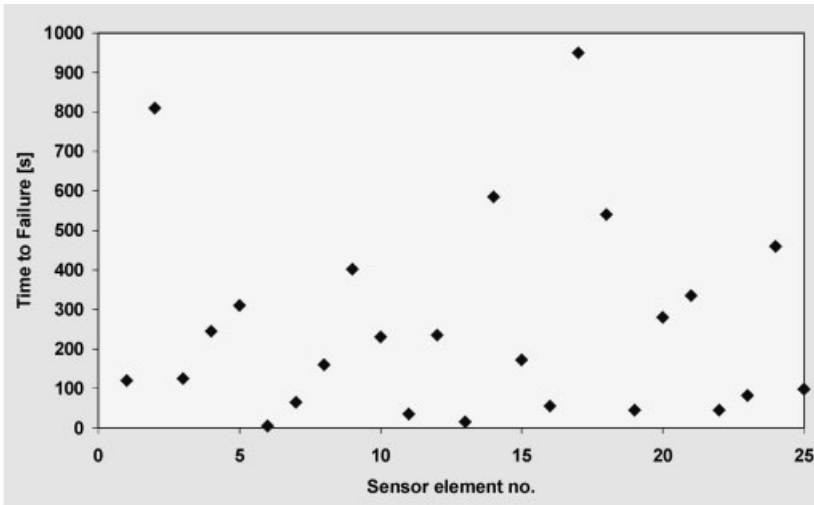


Fig. 5.9.14 Lifetimes of full-scale models of air intake manifold subjected to dust-particle stream

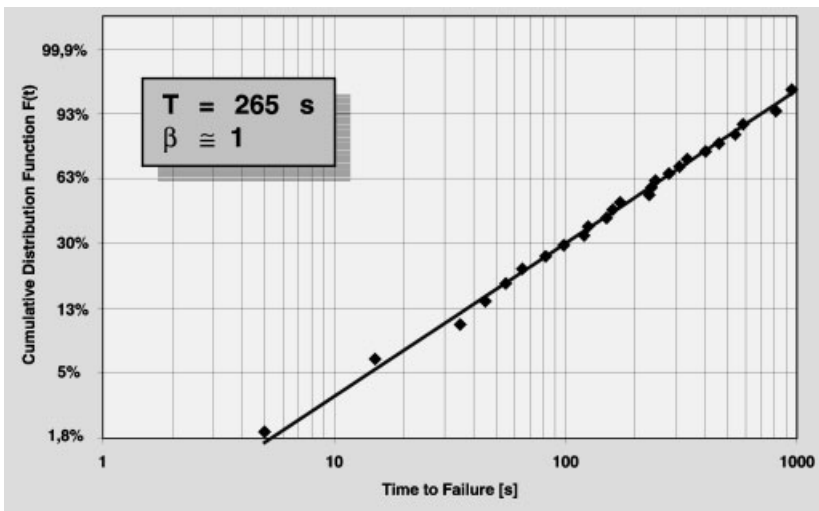


Fig. 5.9.15 Weibull representation of lifetimes of full-scale models of air intake manifold subjected to dust-particle stream

To gain further insight into the failure process, the experiments were accompanied by simulations using FEM. This method has gained wide acceptance in the field of MEMS for analyzing material properties [12], stress distribution [13, 14], and resonance effects, contributing to understanding in fields such as packaging and integration. Because previous experiments revealed that resistance under

static loads could not be translated to dynamic loads, it was necessary to carry out explicit FEM simulations, in contrast to static simulations. The entire impact event was simulated, and load histories were obtained. The criterion used to compare the severity of the impacts was the maximum principal stress component, which is customary for brittle materials, combined with qualitative judgements of the time history of the load, because recent research showed that maximum values alone are not sufficient to judge the resistance of a MEMS structure to shock load [8].

By means of the combined approach of experiment and simulation, the influence of several parameters, such as membrane layer composition, residual stress, thickness, and particle-related properties, were quantified. An empirical model was deduced which allows design rules to be developed for membranes with increased time to failure under dynamic mechanical loads.

Besides intrinsic stress and membrane composition, the location of the particle impact considerably affected the lifetime. This decisive reliability parameter was discovered with this general methodology: starting from knowing only that the failure event is singular, first the effect of particle size was examined in experiment and simulation. In the next step, the impact location was varied in the simulation, and in the experiment certain areas of the membrane were covered so that only the same regions as in the simulation were exposed to particle impact. The FEM simulations revealed that impacts induce high stress not only at the impact point, but also at adjacent edges (Fig. 5.9.16).

In the experiments, membranes with covered edges showed significant increases in lifetime, to the point that sensor elements were virtually indestructible with the standard experimental conditions. If tested under real-life conditions, the sensors would still fail, but would have substantially longer lifetimes. Thus, the region of impact is an important factor determining a sensor's resistance to particle damage.

More detailed studies involving simulation of the entire membrane and different membrane variations were pursued with the aim of optimizing the edge re-

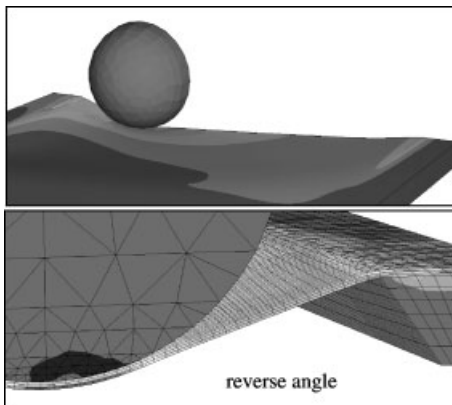


Fig. 5.9.16 Stress distribution in sensor membrane immediately after particle impact

gions. Since this design goal was previously unknown, significant improvements could be realized by means of a redesign of structural elements in the vicinity of the edges. Because the generalized nature of the experiments allows more general application of the experience obtained, a knowledge base is available from which future designs involving membranes subjected to shock loads will also benefit.

5.9.8

References

- 1 M. OHRING, *Reliability and Failure of Electronic Materials and Devices*, Academic Press, San Diego (1998).
- 2 S. TADIGADAPA, N. NAJAFI, Reliability of microelectromechanical systems (MEMS), in *Reliability, Testing, and Characterization of MEMS/MOEMS* (R. RAMESHAM, ed.), *Proc. SPIE 4558* (2001) 197–205.
- 3 A. HARTZELL, D. WOODILLA, MEMS reliability, characterization and test, in *Reliability, Testing, and Characterization of MEMS/MOEMS* (R. RAMESHAM, ed.), *Proc. SPIE 4558* (2001) 1–5.
- 4 S. ARNEY, V.A. AKSYUK, D.J. BISHOP, C.A. BOLLE, R.E. FRAHM, A. GASPARYAN, C.R. GILES, S. GOYAL, F. PARDO, H.R. SHEA, M.T. LIN, C.D. WHITE, Design for reliability of MEMS/MOEMS for light-wave telecommunications, in *Reliability, Testing, and Characterization of MEMS/MOEMS*, R. Ramesham (ed.), *Proc. SPIE 4558* (2001) 6–10.
- 5 S. ARNEY, Design for reliability of MEMS/MOEMS for lightwave telecommunications, in *Proc. 2nd International Conference COTS MEMS 2001 (Advances in Application of Integrated Commercial Off-The-Shelf Microelectromechanical Systems)*, Quincy, MA, USA (2000).
- 6 D.M. TANNER, N.F. SMITH, L.W. IRWIN, W.P. EATON, K.S. HELGESEN, J.J. CLEMENT, W.M. MILLER, J.A. WALRAVEN, K.A. PETERSON, P. TANGYUNYONG, M.T. DUGGER, S.L. MILLER, *MEMS Reliability: Infrastructure, Test Structures, Experiments, and Failure Modes, SAND2000-0091*, Sandia National Laboratory, Albuquerque, NM (2000).
- 7 S. ARNEY, Designing for MEMS reliability, *MRS Bulletin* 26(4) (2001) 296–299.
- 8 U. WAGNER, J. FRANZ, M. SCHWEIKER, W. BERNHARD, R. MÜLLER-FIEDLER, B. MICHEL, O. PAUL, Mechanical reliability of MEMS structures under shock load, *Proc. 11th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis* (2001) 657–662.
- 9 NIST/SEMATECH *Engineering Statistics*, Internet Handbook, <http://www.itl.nist.gov/div898/handbook> (Accessed Oct 2002).
- 10 E. SUHIR, *Applied Probability for Engineers and Scientists*, McGraw-Hill, New York (1997).
- 11 H.-P. TRAH, J. FRANZ, J. MAREK, Physics of semiconductor sensors, *Adv. Solid State Phys.* 39 (1999) 25–36.
- 12 T. KRAMER, O. PAUL, Surface micromachined ring test structures to determine mechanical properties of compressive thin films, *Sensors Actuators A* 92 (2001) 292–298.
- 13 J. BAGDAHN, D. KATZER, M. PETZOLD, Strength and reliability of silicon based MEMS for automotive and medical applications, *Materialsweek 2000, Proceedings* (2000).
- 14 J.Y. PAN, P. LIN, F. MASEEH, S.D. SENTURIA, Verification of FEM analysis of load-deflection methods for measuring mechanical properties of thin films, *1990 Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, 1990.

5.10

Test of Automotive Sensors

WINFRIED SCHÖNENBORN, FRANK SCHATZ and GOTTFRIED FLIK

Testing is a very essential part of the manufacturing process of automotive sensors. Since these sensors are intended for use in safety systems (airbags, ESP (Electronic Stability Program), etc.), their functionality has to be guaranteed under all circumstances. At the same time, automotive sensors have to be produced in extremely high volumes at competitive cost levels. Therefore, batch techniques (micromachining, thin-film technology, thick-film technology, etc.), similar to those used in IC fabrication or – more generally – electronic device fabrication, have become widely accepted in the last two decades for fabricating automotive sensing element.

Often, most of the functionality of modern automotive sensors is dominated by the sensing elements, which are typically fabricated by one of the above processes. In contrast to their enormous impact on the function of the sensor, the cost added by these sensing elements is normally rather low – approximately 10–25% of the overall sensor cost. Unfortunately, state-of-the-art test methods, which have been developed for testing conventional electronic devices, turn out to be inappropriate or insufficient for testing sensing elements.

In this section we therefore introduce the special challenges of automotive sensor testing. First, some specific characteristics of sensing devices with respect to testability for functional parameters are pointed out. Then, conclusions are deduced leading to an optimized approach to test methods suitable for batch fabrication processes (i.e., model-based tests, test functions, test-pattern based tests, parameter extraction methods based on on-wafer tests). Practical examples show the potential of the methodical approach.

5.10.1

Introduction to Microsensor Metrology and Testing

The accuracy and reliability of microsensors under severe environmental conditions are of prime importance in automotive applications. To meet the required quality at competitive cost levels, high yields within the product specifications have to be ensured under conditions of high-volume production. In addition to the continuous development of refined sensor designs and processing technologies, appropriate testing concepts and methods are indispensable to guaranteeing specified quality and sustained economic success.

Obviously, a sensor's function can be tested at the end of the fabrication process after the sensor is completely assembled and fully packaged. At this time a primary input signal (pressure, acceleration, yaw rate, mass flow, etc.) can be applied directly, and a comprehensive test of the specified performance is possible. Unfortunately, by the time a defective sensor component is packaged, loss of time and capital is maximum, since a fully packaged sensor needs to be discarded.

A more cost-efficient approach is to test the individual sensor components prior to assembly, that is, at a time when the level of value added is still low. For ASICs

(application specific ICs), appropriate methods and standardized equipment enabling high throughput, high precision testing at the wafer level are readily available based on highly automated probe stations and purely electrical stimuli and response signals.

However, by far most sensors convert nonelectrical (primary) input signals into electrical output. At first sight, the adoption of automated probe stations seems to be limited to purely electrical networks on sensor elements. However, electrical stimuli provided by the probe station may replace the nonelectrical primary input signals to the sensor transducer element, if a built-in self-test function (BITE) can be incorporated into the sensor design, allowing transducer-related (i.e., sensitivity-related) fundamental functional parameters to be tested.

When a built-in self test is not feasible, we encounter difficulties because primary input signals cannot be easily provided within a standardized probe setup. To our knowledge, no commercial testing equipment is available that allows a set of primary stimuli like acceleration, pressure, torque, or mass flow to be applied to the transducer elements of sensor chips on the wafer level, with the required speed and precision. At the moment we are therefore left with purely electrical stimuli for testing microsensor devices.

In this situation a holistic sensor design approach may prove valuable [1]. The approach may include these steps:

1. Design and model the sensor function on the basis of a comprehensive set of model parameters (i.e., geometrical dimensions and material properties) and calculate tolerance bands for all layout-specific model parameters (including those for wafer-level packages) so that functional specs are safely met (see Section 4.1).
2. Identify functionally critical model parameters for which the tolerance bands of available batch fabrication processes do not comply with those calculated above. Consider alternative processing technologies and redesign the sensor so that the number of critical parameters is minimized (see Section 4.1).
3. Concurrently design test structures and develop test models dedicated solely to extracting critical model parameters from electrical precision measurements on the wafer level. Improve fabrication processes using test structures early in the sensor development phase and well in advance of starting production. Eventually, reduce the number of test structures and the wafer space they occupy, optimize testing time and equipment, and then commit a test setup to production.

In contrast to IC tests [2], model parameters for microsensors may differ significantly depending on the specifics of the transducer principle, the fabrication sequence used, and the intended sensor application. In this context a 'library' of test models and testing methods is a valuable tool not only to set quality gates in a production line but also to accelerate the qualification of newly developed manufacturing processes and newly developed functional designs. In this sense, high throughput precision testing methods capable of extracting material properties (e.g., stress gradients, shear moduli, thermal conductivity, heat capacity) and 3D geometries (film thicknesses and widths, edge losses, side wall inclination angles,

etc.) from directly accessible electrical quantities appear to be a prerequisite for microsensor product commercialization.

Below, we indicate the status and trends in microsensor testing and metrology.

5.10.2

Microsensor Wafer-Level Testing

Holistic design and test approaches are only at their beginnings in the world of MEMS-based mass production. Corresponding tools are still in development and show a relatively low degree of standardization. Therefore the situation may be characterized as a set of coexisting isolated partial solutions.

5.10.2.1 Standard Electrical Tests of Functional Parameters

As mentioned above, some functional parameters of microsensors are readily accessible to high-speed electric testing on the wafer level. Often such tests are limited to functional substructures such as heaters (for microanemometers) or resistor bridges (for pressure sensors), where testing of input or output resistance, bridge offsets, and thermal coefficients of offsets with respect to specified tolerance bands is in principle straightforward and is widely used in monitoring quality within a production line. Together with data obtained by in-process monitoring (e.g., lateral CD (Critical Dimension) measurements), valuable information on thin-film material parameters can be deduced. In general, the database is still too small to allow tests regarding conformity with transducer-related specifications, that is, sensitivity, linearity, and cross-sensitivity specifications.

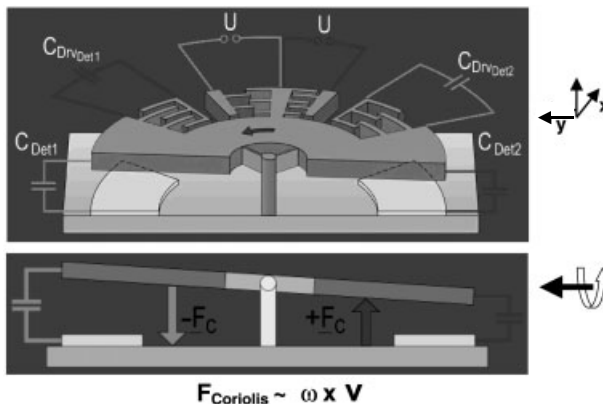


Fig. 5.10.1 Schematic drawing of a rotational disc gyro. The rotor (in blue) is oscillating around the z axis, driven by fixed comb structures. A yaw rate around the y axis generates a Coriolis force in the z direction. The resulting seesaw-like out-of-plane motion of the rotor is detected by measuring the difference in capacities between the rotor and detecting electrodes beneath it

5.10.2.2 Gyro: Functional Parameter Tests (BITE)

Testing concepts for a surface-micromachined rotating disc gyro represent an interesting example of a direct functional parameter test based on electrical probe data sets. The basic functional design is shown in Figure 5.10.1. A surface-micromachined rotor is driven to near resonance by an AC voltage applied to the lateral comb drive electrodes. Ideally, the oscillation of the rotor should be restricted to an in-plane motion around the z axis. If an external stimulus rotates the oscillator around the in-plane axis (y), the Coriolis force causes a seesaw-like out-of-plane motion (around the x axis). A capacitive yaw rate-related sensor signal is picked up with the help of buried electrodes beneath the rotor.

Obviously, this gyro design requires tightly controlled surface micromachining processes with respect to the thicknesses of the polysilicon and sacrificial layers and to stress gradients within the polysilicon rotor. However, the tolerance band widths and process capability indices available for these model parameters were not indicative of significant yield losses in a factorial analysis.

When designs for gyro-based sensors first began to be developed, the occurrence and effects of processing-induced asymmetries in the 3D geometry of the comb drive electrodes were underestimated. Trench etch-related side wall inclination angles, that is, deviations from an exactly rectangular parallelepiped in the shaped polysilicon, may give rise to significant mechanical coupling between the in-plane driving motion and the out-of-plane sensing motion. This ‘parasitic’ quadrature effect – although phase shifted – may exceed the compensation range of the signal-evaluation circuit, thus necessitating wafer-level testing of quadrature-induced offset signals.

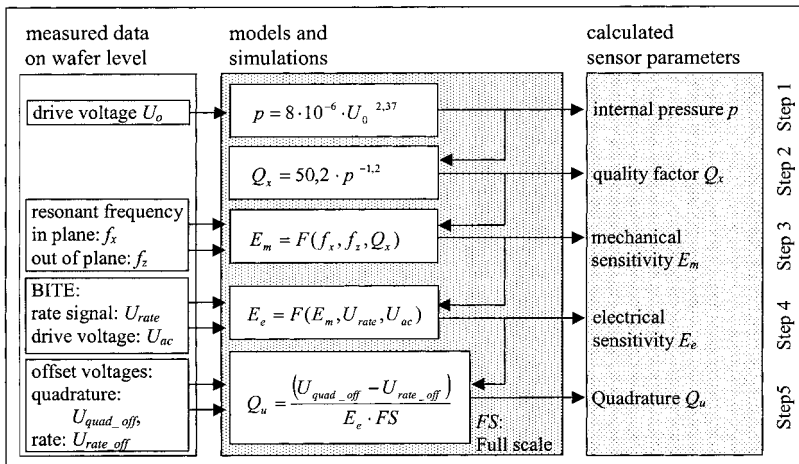


Fig. 5.10.2 Test concept for determination of the quadrature number of a rotating disc gyro. The section on the left lists all electrically measured properties. Sections to the right show the sequence of models and simulations employed to extract sensor parameters, including the quadrature number

The corresponding test concept is shown schematically in Figure 5.10.2. In a simplified form the quadrature number

$$Qu = (U_{\text{quad_off}} - U_{\text{rate_off}}) / (E_{\text{electr}} \times FS)$$

is calculated from the difference between the quadrature offset ($U_{\text{quad_off}}$) and the rate signal offset ($U_{\text{rate_off}}$), normalized to the full-scale sensor response signal, that is, normalized to the product of the electrical sensitivity (E_{electr}) and the working range (FS).

First, an in-plane oscillation with an amplitude of 4° is stimulated and controlled by a feedback loop. Due to their 90° phase difference, rate signal offsets (due to phase demodulation errors) and quadrature offsets can be demodulated separately and measured.

Second, we use an additional set of provided buried electrodes to stimulate an out-of-plane (see-saw) motion on the device level. From the drive voltage and the corresponding signal response of this BITE, we could calculate the electrical sensitivity if the mechanical sensitivity were known. However, the mechanical sensitivity is not directly accessible and is therefore calculated from in-plane and out-of-plane resonance frequencies and the in-plane quality factor.

Both resonance frequencies are therefore measured in a third step. Because the quality factor depends on the pressure inside the sealed pre-package and may therefore be process-dependent, a fourth and fifth step are necessary: the encapsulated pressure is deduced from an empirically determined relation between pressure and the in-plane drive voltage needed to stimulate a 4° amplitude oscillation.

This approach to a quadrature test comprises acquisition of 7 directly accessible electrical measurands and a 5-step modeling sequence. As a multiple functional parameter test of considerable breadth and expense, it delivers control of resonant quality factors, rate offsets, and mechanical and electrical sensitivities, in addition to quadrature-induced offsets.

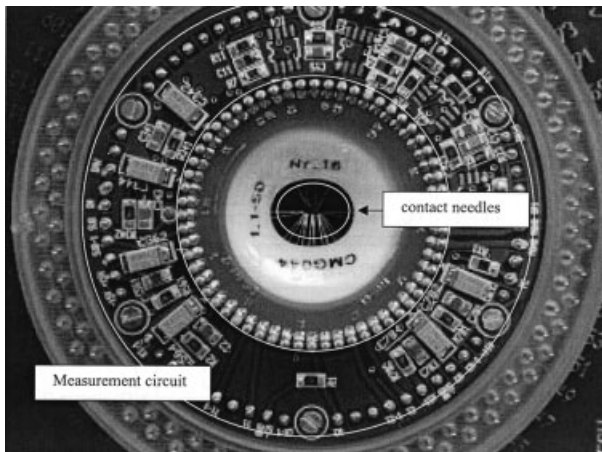


Fig. 5.10.3 Probe card with integrated capacity-to-voltage converter. Compact construction leads to small parasitic capacities

In performing these electrical measurements we have to be aware that signal levels are often very small. After all, the change in capacity of the rotating disc gyro is on the order of 40 fF at 100° s^{-1} with a resolution specification of $<1\%$ (3 s). Consequently, the distance between the electrical contacts on the sensor chip and the capacity to voltage converters in the signal-evaluation circuit has to be as short as possible. Parasitic capacities must be avoided. In addition, the signal-evaluation system should be symmetric and shielded. Figure 5.10.3 shows a probe card with contact needles including the measurement circuit. Digital signal processing (DSP) is used for noise reduction, additionally improving the testing time and reproducibility.

5.10.2.3 Methods of Extracting Model Parameters

As outlined Section 5.10.1 knowledge of model parameters in relation to their tolerance bands enables model-based predictions of functional parameters and may offer a more fundamental and efficient alternative way of monitoring a manufacturing process. Extracting model parameters early in the sensor development phase enables one to identify and focus on 'critical' processes that need improvement respect to their tolerance bands, reducing waste of time and money due to unfocused troubleshooting or design recursion cycles. This enables an engineering process that may indeed make it first and make it right.

5.10.2.3.1 Gyro: Model Parameters Accessible from Functional Test

To illustrate the extraction of model parameters from functional parameter tests, we return to the example of the rotating disc gyro and the 5-step model outlined in Section 5.10.2.2.

It is often valuable to visualize the distribution of model parameters as a 2D wafer map. For example, from the empirical test model (step 5 in Fig. 5.10.2), the pressure inside the vacuum-sealed pre-packages can be presented straightforwardly as a function of wafer position (see left part of Fig. 5.10.4) giving hints on how to improve on the homogeneity of the sealing process.

Similarly, the thickness of the polysilicon layer that forms the resonating disc can be monitored and visualized as shown in right part of Fig. 5.10.4. Film thickness and edge-loss effects observed in reactive ion etching are extracted from in-plane and out-of-plane resonance frequencies.

5.10.2.3.2 Accelerometers: Extraction of Geometry Parameters

Another approach to extracting geometrical parameters for acceleration sensors is based on a spring-mass system which is capacitively read out and fabricated by surface micromachining. Based purely on static capacitance-voltage measurements, characteristic model parameters related to polysilicon thickness, side wall overetching, and differences in side wall inclination angles can be determined on the wafer level [3]. This approach offers a good compromise between depth and breadth of testing, that is, the significance of the results and the cost of testing.

First, a bias voltage applied between the fixed comb drive electrodes and the movable comb drive electrodes connected to a seismic mass suspended by a

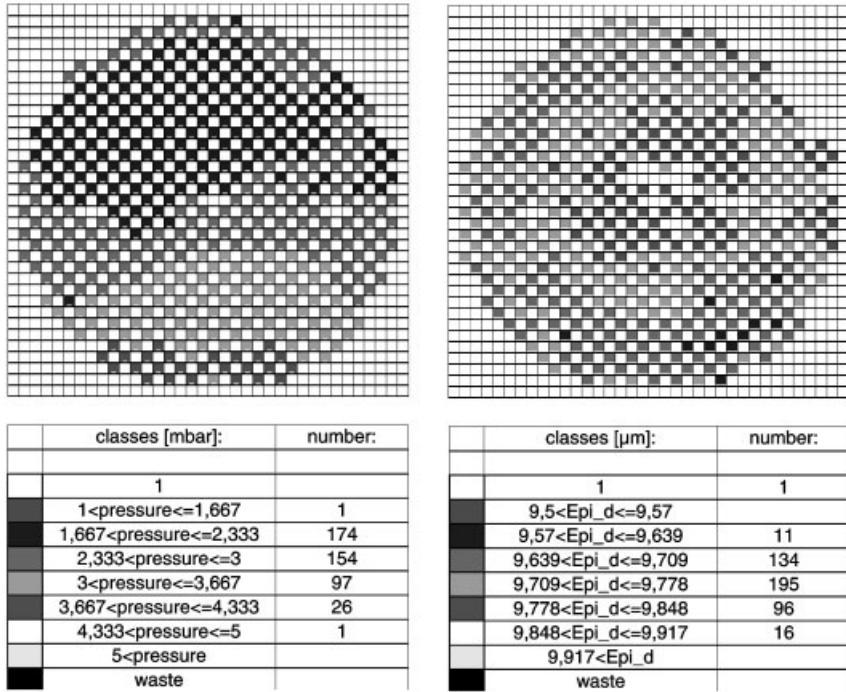


Fig. 5.10.4 Wafer maps of model parameters: encapsulated pressure inside vacuum-sealed pre-packages of rotating disc gyros (left); thickness (Epi_d) of the polysilicon rotor (right)

spring results in an attractive electrostatic force. The resulting lateral (i.e., in-plane) displacement of the polysilicon proof mass can be calculated from the equilibrium condition of the elastic and electrostatic forces. As the bias voltage increases, the distance between the fixed and movable comb electrodes decreases. From the resulting capacitance-voltage characteristic, a test function can be established correlating the thickness of the functional polysilicon layer with the side wall overetch using linear regression techniques.

In a second step, a bias voltage is applied between the substrate and the movable proof mass resulting in vertical (i.e., out-of-plane) displacement. As the bias voltage increases, the distance between the proof mass and the substrate decreases, due to the attractive electrostatic force. Hence, the capacitance between the proof mass and substrate increases. Until recently, the polysilicon trench-etch process led to slightly sloping side wall profiles. Again, as a consequence of an asymmetric beam cross section, a vertical (out-of-plane) actuation of the proof mass implies a horizontal (in-plane) motion, which can be calculated from the measured in-plane capacitance between the fixed and movable comb drive electrodes. Assuming a trapezoidal cross section of the beams, the difference in the side wall inclination angles can be calculated using the orthogonal set of test functions derived from in- and out-of-plane capacitance-voltage characteristics.

The testing concepts described so far make explicit use of the fact that additional drive electrodes, together with those used to capacitively sense a primary sensor input signal, may be used for built-in tests. ‘Design for test’ concepts therefore appear to be limited to the class of sensors employing reversible transducer mechanisms (i.e., capacitance or piezoelectricity) and to their processing sequences.

5.10.2.3.3 Mass Flow Sensors: Extraction of Thermal Material Properties

As opposed to inertial sensors, micromachined anemometers are often based on micro hot plates, that is, combinations of heaters and thermometers on thermally insulating membranes or multilayered thin films [4]. Functional sensor parameters are obviously determined by thermal conductivities and heat capacities, which can be monitored to reject faulty chips. Until recently, suitable wafer-level testing methods have not been available, and most sensor designs were based on literature values, despite the fact that these values depend strongly on fabrication processing parameters.

The approach of Raudzis et al. [5, 6] is based on the so-called 3-omega technique [7] and is easily adapted to thin- or thick-film technology: only a heater element with a high thermal coefficient of resistance (TCR) having 4-point probe geometry is required on the material system being tested.

A periodic heating current with angular frequency ω induces temperature oscillation within a thermoresistor. In the example below, this leads to a diffusive thermal wave, whose amplitude ΔT and penetration depth depend on the thermal properties of the substrate and on the frequency of temperature oscillation. The electrical resistance of the heater itself depends on the temperature and therefore oscillates with angular frequency 2ω . The measured voltage drop over the thermoresistor thus contains a small component $V_{3\omega}$ with angular frequency 3ω which is detected with lock-in techniques. The thermal wave amplitude ΔT is then calculated from $V_{1\omega}$ and $V_{3\omega}$ and the TCR heater values. Information on the thermal conductivity λ is obtained from the slope $\partial(\Delta T)/\partial(\log(f))$. Figure 5.10.5 shows the calculated and normalized slope $\partial(\Delta T)/\partial(\log(f))$ over a wide range of frequencies ($f = \omega/2\pi$) for a SiO_2 film on a silicon substrate.

In the low-frequency region of Figure 5.10.5 (zone I), the thermal changes penetrate so deeply that thermal coupling effects to the back of the wafer dominate (adiabatic or isothermal coupling to the prober chuck). At somewhat higher frequencies (zone II), the thermal properties of the substrate are relevant. At even higher frequencies (zone III), the depth of thermal penetration is reduced to the scale of the thickness of the SiO_2 film, and the lower thermal conductivity of SiO_2 compared to silicon causes the curve to rise. Zone IV is dominated by effects due to the non-zero lateral extension of the heater potentially allowing the study of anisotropy effects in thermal conductivity.

This testing method enables us to simultaneously monitor three different model parameters (i.e., thermal conductivity, heat capacity, and film thickness) – not only for a single thin film system but also for multilayer systems and buried structures.

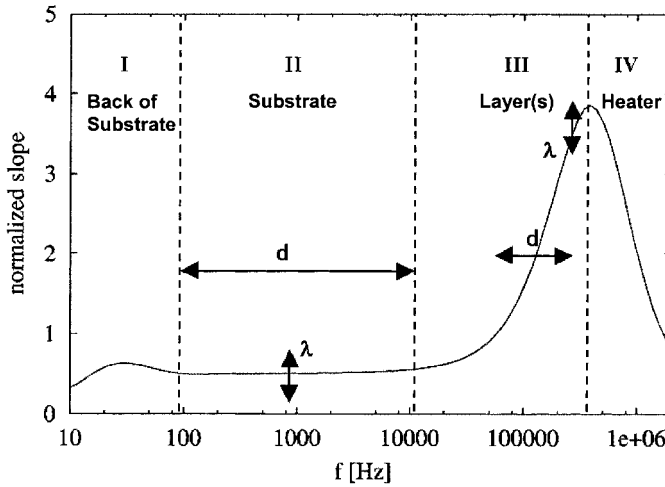


Fig. 5.10.5 Three-omega technique: the calculated normalized slope ($\partial(\Delta T)/\partial(\log(f))$) of the thermal oscillation amplitude ΔT as a function of frequency f for a 100 nm thick aluminum heater element on a 500 nm thick SiO_2 layer and a 1 mm thick silicon substrate is shown as a function of thermal oscillation frequency f . With increasing frequency, the thermal wave penetrates deeper, allowing different sections of a (multilayered) sample to be scanned

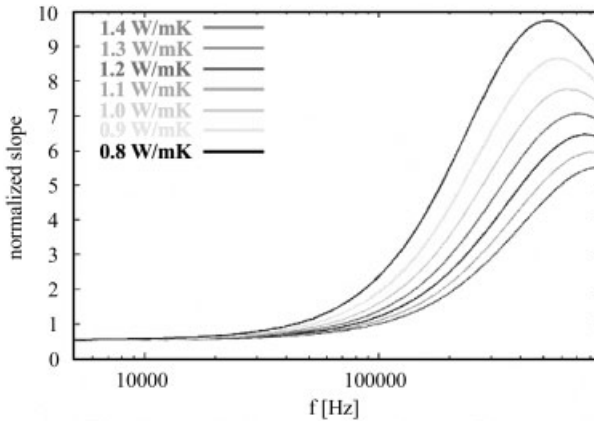


Fig. 5.10.6 Three-omega-technique: influence of different thermal conductivities on the calculated normalized slopes ($\partial(\Delta T)/\partial(\log(f))$) for 500 nm thick SiO_2 films (substrate: 0.5 mm silicon, heater: 100 nm aluminum)

The sensitivity of the method to variations in a single parameter is impressive. Figure 5.10.6 shows the calculated frequency dependence of the normalized slope in tests of SiO₂ films of different thermal conductivities on a silicon substrate. Variations in thermal conductivities of 0.1 W mK⁻¹ should be easily resolvable at frequencies above 100 kHz.

Through sophisticated testing models, it appears feasible to achieve significant depth and breadth of model parameter tests simultaneously, even with a single, simple test structure. It is obvious that a library of similar test methods for all the potentially critical geometric and material parameters must be readily available and verified at least for those functional sensor designs that are being considered for high-yield mass production. Moreover, these methods represent a powerful tool for quality assessment of newly developed micro system technologies and processing modules.

5.10.2.4 Identification of Defects and Anomalies

So far, we have assumed that a change in an equipment-related MEMS-processing parameter set leads to a defined statistical distribution of geometric and material parameters allowing yield predictions based on functional modeling. However, in addition, we need to take stochastic effects into account and develop testing methods to detect defects and anomalies that detract from product quality and which cannot now be traced to their root causes or controlled sufficiently during production.

Defects include (dust) particles on or within suspended interdigitated seismic masses, leading to electrical shorts and to deviations in moments of inertia. Another example relates to the appearance of crystallographic defects such as dislocations or stacking faults in epitaxial superstructures, causing leakage currents that might impair sensor reliability in the field. Defect densities may of course vary within a wafer, from wafer to wafer, and from lot to lot. Unfortunately, at present not many models are available that link tolerance bands of functional parameters to specific defects and their tolerable density distributions.

Whereas phenomena like sticking of surface-micromachined structures appear early in the testing sequence described above for rotating disc gyros, the detection of stochastically appearing static surface charges on the detection electrodes, leading to out-of-spec offset potentials, requires extensions to this testing routine. Again, the provision of a built-in test structure proves valuable.

A positive DC bias voltage is superimposed on the AC test signal on the BITE electrode, driving the see-saw into a oscillating motion around an inclined equilibrium position. The test signal response from the signal detection electrodes (yaw rate out) is recorded. Then a negative DC bias voltage is applied to the BITE electrode such that the absolute values of the respective test-signal responses are identical. As illustrated in Figure 5.10.7, the presence of surface charges is detected on the wafer level from asymmetries in the sensor characteristics. Faulty devices are recognized and marked with an ink dot to prevent them from being packaged.

Often, the presence of a defect does not change model parameters strongly enough to drive them out of their specification range. Nevertheless, the sensor chip must not be delivered to a customer, because the risk of it failing during its life-

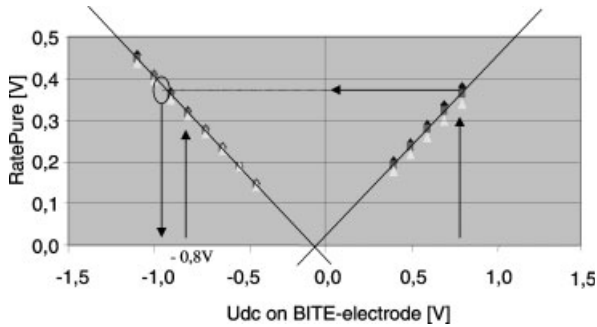


Fig. 5.10.7 Rate signal of a gyro for different BITE bias voltages. Due to the potential of static surface charges, an asymmetry in the rate signal for positive and negative offset voltages is generated and detected

time is too high. For instance, a small part of a vibrating comb structure might have come off and might have escaped visual inspection. The resulting change in resonance frequency can be detected, but may be just within specifications. However, the debris could cause an electrical short or blockage at any time in the life cycle of an automobile, with potentially significant damage and costs.

To identify such defects, the behavior of the sensor chip being tested is compared with that of its neighboring chips on the wafer and checked for consistency. Assuming only small variations in other process-related factors, defect-induced singularities (anomalies) may be identified with sufficient statistical significance.

5.10.2.5 Modification of Testing Equipment and Facilities

As pointed out in Section 5.10.1, conventional IC testing equipment does not permit a primary input signal (e.g., pressure or acceleration) to be applied to a sensor being tested at the wafer level. Efforts are being made to modify IC testing facilities accordingly.

For instance, to test air quality sensors a standard wafer prober has been upgraded with a gas feeding system that allows each sensor element to be exposed to a well-defined mixture of NO_x and CO_x gases (cooperation of Suess GmbH and Robert Bosch GmbH). The sensitivity to various gas compositions can now be measured on the wafer level. To achieve good consistency with test results obtained by exposing individual sensor chips in a gas chamber, the gas feeding system had to be optimized with respect to gas flow velocity and relative incidence angles of the gas stream. Cycle times (i.e., relaxation times needed to achieve equilibrium conditions) need further improvement.

Bay and Branebjerg [8] presented 3 different approaches to testing membrane-based pressure sensors employing

- a local pressure chamber around the sensor element, to apply a real pressure
- a needle to apply mechanical force
- an air stream to deflect the membrane

As mentioned, various approaches to modifying existing testing equipment are conceivable. However, to assess these ideas, the main criterion is the ratio of benefit to cost. If there is any possibility to test without introducing additional equip-

ment complexity, that is, if reliable functional models are available to forecast yield based on electrically determined model parameters, standard test equipment is used. Especially for a market like automotive sensors, with a wide variety of sensor functions and considerable cost pressure, the penetration of new dedicated (i.e., sensor function specific) testing equipment appears fairly difficult in near future.

5.10.3

Concluding Remarks

Model parameter tests on the wafer level represent a basic building block within the framework of a holistic sensor design approach.

First, in an early phase of product development, testing of material and geometric parameters is of prime importance for revealing and ameliorating potential weaknesses in sensor design and layout hand, or, for improving existing processing technologies, and for calibrating and releasing new equipment. Test results in the development phase determine which model parameters remain critical with respect to yield and thus need to be tested in the ramp-up phase to real production using time- and cost-optimized test methods and algorithms to guarantee the quality of delivered sensors. Based on wafer-level test data, factors that detract from yield and quality can be identified reliably and troubleshooting may be focused on them. To this end, libraries of proven model parameter testing methods, test models, and test structures must be available for each sensor function and processing technology under consideration.

A prerequisite for efficient use of model parameter test data is the availability of flexible, comprehensive data storage and data handling systems that allow a full data set to be traced and analyzed for each sensor chip, including all relevant processing parameters and in-process monitors (i.e., for the complete manufacturing line). Standardized data formats and a toolbox of data analysis software for creating trend plots, histograms, and wafer maps of critical model parameters are essential for assessing drifts and anomalies (preferentially by generating automatic alarm flags). Data mining can reveal valuable, previously unknown correlations of model parameters and details of the correlations of processing and model parameters with good statistical significance provided by large databases [2]. Refinement of functional modeling algorithms is possible if chip-specific data from the final sensor test of fully packaged sensors can be consistently included in the database, thus enabling verification or falsification of model-inherent approximations and hypotheses in addition to elucidating packaging-induced effects on sensor performance.

In the future, new additions to the existing portfolio of processes and of sensor and actuator functions will inevitably call for extensions of wafer (substrate) level testing concepts and methods for obtaining both geometric and material parameters. In our opinion, wafer level testing is essential for the sustained economic success of microsystem technologies well beyond the scope of automotive applications.

5.10.4

References

- 1 H.-P. TRAH, J. FRANZ, J. MAREK, Physics of semiconductor sensors, *Adv. Solid State Phys.* 39, 25, 1999.
- 2 A.N. SRIVATAVA, Data mining for semiconductor yield forecasting, *Future FAB International*, p. 67, 2001.
- 3 M. MAUTE, S. KIMMERLE, J. FRANZ, J. HAUER, D. SCHUBERT, H.-R. KRAUSS, D.P. KERN, Parameter extraction for surface micromachined devices using electrical characterization of sensors, *Technical Proceedings of the Fourth International Conference on Modeling and Simulation of Microsystems*, Hilton Head Island, South Carolina, Computational Publications, Cambridge, MA, p. 406, 2001.
- 4 J. MAREK, M. ILLING, *Microsystems for the automotive industry*, *IEDM Technical Digest*, 2000.
- 5 C. RAUDZIS, F. SCHATZ, D. WHARAM, Adaption of the 3ω -Method for testing of MEMS, *Nanotech Proceedings 2003*, in press.
- 6 G. FLIK, H. EISENSCHMID, C. RAUDZIS, F. SCHATZ, W. SCHÖNENBORN, H. P. TRAH, Microsensors for automotive applications: metrology and test, *MRS Proceedings*, 687, p. 1, Materials Research Society, Warrendale, 2001.
- 7 D.-G. CAHILL, Thermal conductivity measurement from 30 to 750 K: the 3ω method, *Rev. Sci. Instrum.* 61(2), 802, 1990.
- 8 J. BAY, J. BRANEBJERG, Functional testing of micromachined mechanical sensors on wafer level, *Euroensors XIV*, p. 317, Microelectronic Centret, Lyngby, 2000.

6 Evaluation Circuits

6.1 Capacitive Evaluation Circuits

BERNHARD BOSER

6.1.1 Introduction

Many sensors use capacitors for interfacing to electronic circuits. The ability of the electronic circuit to resolve minute changes in capacitance translates into excellent sensor resolution. When used as displacement sensors, capacitive interfaces achieve resolutions down to the sub-atomic level. For example, the gyroscope described in [1] resolves displacements as small as 10^{-14} m. For comparison, the classical radius of an electron is 2.8×10^{-15} m and the distance between atoms in the silicon crystal is orders of magnitude larger, 2.5×10^{-10} m. The capacitance resolved by the circuit is 10^{-20} F, six orders-of-magnitude less than the gate capacitance of a transistor with sub-micron dimensions.

Comparisons of published results suggest that capacitive sensing is as good or better than alternative techniques, including tunnelling, piezoresistive, and piezoelectric interfaces [2]. Other attractive features of capacitive interfaces include compatibility with many sensor fabrication processes, low temperature coefficient, and low power dissipation.

Most capacitive evaluation circuits do not achieve the maximum possible resolution but are limited by the electromechanical interface, shortcomings in the electronic circuits, or stray signals coupling into the detector and corrupting the output. Section 6.1.2 below illustrates approaches to maximize the sensitivity of capacitive sensor interfaces, potential error sources, and approaches to minimize them. Electronic circuit options are discussed in Section 6.1.3.

6.1.2

Transducer Interface6.1.2.1 **Introduction**

Although not strictly part of the evaluation circuit, the interface between the transducer and the electronics has a significant impact on overall sensor performance and must therefore be carefully designed. The performance difference between an optimally designed interface and a seemingly very similar solution is often an order-of-magnitude or more. In addition to the design of the sensing capacitor, attention must be paid to parasitic capacitors and resistors, the effect of electrical signals on the transducer that can result in resonant frequency shifts or actuation, and stray electrical signals that might corrupt the measurement result.

The capacitance between conductors is a function of conductor size A , spacing z , and the permeability ϵ of the material between the conductors. For parallel-plate capacitors with dimensions much larger than the gap the capacitance is

$$C = \epsilon \frac{A}{z} \quad (1)$$

For reference, the capacitance of $100 \mu\text{m} \times 100 \mu\text{m}$ plates with $1 \mu\text{m}$ air gap is 88 fF.

Many different arrangements are used for capacitive sensing. Parallel-plate type interfaces sense variations of the plate spacing, z . Conversely, overlap type interfaces such as combs detect variations of the effective capacitor area. The two approaches and combinations thereof differ in sensitivity, displacement range, electrostatic force, and electrical spring constant generated.

Capacitive interfaces can be further divided into single-ended and complementary structures. The latter sense variations of opposite sign of two nominally equal capacitors. Compared to single-ended solutions, complementary designs usually exhibit significantly smaller offset and improved sensitivity.

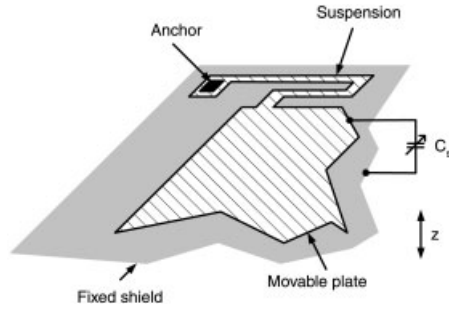
The sections below describe three representative capacitive interfaces. The single-ended parallel-plate style is used in inertial and pressure sensors where fabrication constraints preclude a complementary design. Many microfluidic applications, such as particle sorting or fluid level detection, and proximity sensors fall into this category also. Section 6.1.2.3 describes the advantages of complementary interfaces. The last section is devoted to comb style sensors.

6.1.2.2 **Single-Ended Parallel Plate Interface**

Fig. 6.1.1 shows a simple parallel-plate capacitor suspended with flexures permitting the structure to move in the z direction. Applications include acceleration [3, 4] and pressure sensors [5, 6]. Similar structures are also used for sensing fluid levels or biological assays. In this case both electrodes are fixed and instead the relative permittivity ϵ_r of the medium between changes.

If the plate dimension W is much larger than the spacing z , the capacitance of the structure is

Fig. 6.1.1 Parallel-plate capacitor with suspension



$$C(z) = \epsilon \frac{W^2}{z} \tag{2}$$

The sensitivity to translation in the z direction is

$$\frac{C(z)}{dz} = \frac{C(z)}{z} \tag{3}$$

For small displacements the sense capacitor can be represented by a fixed capacitor C_0 and a change ΔC ,

$$C(z) = C_0 + \Delta C \approx C(z_0) + z \left. \frac{dC(z)}{dz} \right|_{z=z_0} \tag{4}$$

Fig. 6.1.2 shows a simple electrical model in which the sense capacitor is placed in series with a reference capacitor C_{ref} . Also shown are parasitic capacitors C_{Pi} .

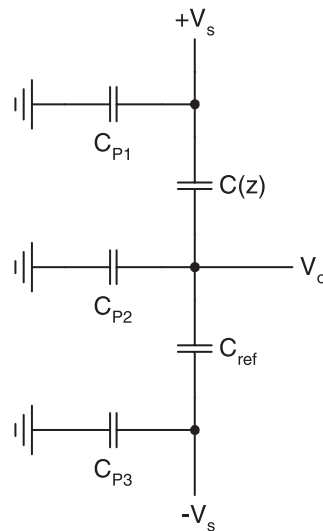


Fig. 6.1.2 Electrical equivalent circuit for a single-ended parallel-plate capacitive transducer interface

Especially in micromachined sensors, these parasitics are often comparable to or larger than the sense capacitance and can significantly reduce the sensor performance. The relatively small variation of the relative permeability compared with the range of conductance of different materials is the reason why parasitics tend to be much more important in sensors with capacitive rather than (piezo)-resistive interfaces.

The output voltage V_0 of the circuit is

$$V_0 = V_s \frac{\Delta C + C_0 - C_{\text{ref}}}{C_0 + C_{\text{ref}} + C_{P2} + \Delta C} \approx V_s \frac{\Delta C + \overbrace{C_0 - C_{\text{ref}}}^{\text{offset}}}{C_0 + C_{\text{ref}} + C_{P2}} \quad (5)$$

In addition to the desired dependence on ΔC , it has a matching-dependent offset and gain that depends on parasitics. Any deviation of the reference capacitor C_{ref} from the nominal value of the sense capacitance C_s appears as offset. Since in many applications ΔC is much smaller than C_0 , this offset often exceeds the signal. Offset cancellation should therefore occur early to minimize the dynamic range of the readout electronics. Care should also be taken for the trimming not to introduce a poor temperature coefficient. One solution fabricates the reference with the same process and in close proximity to the sense capacitor. The z axis accelerometer shown in Fig. 6.1.3 [7] utilizes two micromachined structures for the sense and reference. The suspension of the reference structure has been made intentionally stiff.

The second problem highlighted by Equation (5) is the dependence of the sensitivity of the capacitive interface to parasitic capacitance, C_{P2} . This capacitance significantly degrades the sensor output as it approaches or exceeds the sense capacitance, C_0 . In principle, signal attenuation suffered as a result of excessive parasitics can be compensated by subsequent electronic amplification. In most cases, however, electronic noise and other errors are amplified also, negating any potential gain in sensor resolution.

In many implementations the parasitics C_{P1} and C_{P2} at the terminals of the sense capacitor differ. In this case, it is advantageous to connect the output V_0 to the terminal with the smaller parasitic. A good layout also contributes to minimizing parasitics. For example, careful placement and sizing of anchors often results in reduced parasitic capacitance. In micromachined processes, low capacitance in-

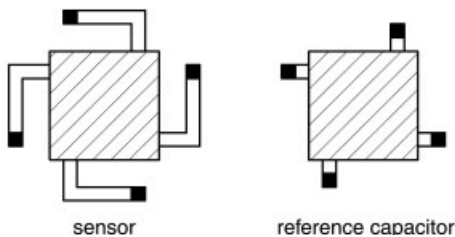


Fig. 6.1.3 Z axis accelerometer with reference capacitor

terconnects over the substrate can often be realized as ‘air-bridges’, structural polysilicon separated from the substrate by an air-gap.

It is interesting to note that, in the absence of parasitics, the sensitivity of the parallel-plate interface is independent of capacitor size

$$\frac{dV_0}{dz} = \frac{1}{2} \frac{V_s}{z} \quad (6)$$

This expression suggests that to maximize the sensitivity only the gap must be minimized, but the absolute value of the capacitor (i.e., its area) is irrelevant. When parasitics are included in the analysis, their value dictates the minimum sense capacitance below which sensitivity suffers.

Parasitics are an important factor in the decision for monolithic integration versus two-chip solutions. Owing to their small dimensions, the practically achievable sense capacitance in many surface micromachined devices with typical 2 μm gaps and film thickness is often 100 fF or less. Since chip-to-chip interconnections, even using solder bumps, add at least 1 pF parasitic capacitance, on-chip electronics is usually required to achieve the highest possible sensitivity. Bulk-micromachined devices have much higher sense capacitance; in this case a two-chip solution may not incur a resolution penalty.

In addition to reducing sensitivity, parasitic capacitors can also couple unwanted signals from the substrate or adjacent interconnections into the sensor output. Because of the high sensitivity of capacitive sensing circuits, sources of errors that are irrelevant for other circuits are often significant in capacitive sensors. For example, the capacitance between a surface micromachined sensor and a pad is on the order of 1 fF, creating a potentially significant coupling path. Shields, conservative spacing, and complementary electronic signals all help to reduce errors.

According to Equation (5), the sensitivity of capacitive interfaces can be maximized also by increasing the voltage V_s applied to the sensor. Aside from supply voltage limitations, several other considerations set an upper limit on the value of V_s , including electrostatic force and undesired actuation. The so-called pull-in effect sets an absolute limit. The magnitude of the negative electrostatic spring mentioned earlier increases with the applied voltage. When the sum of mechanical and electrical spring constant becomes zero the capacitor plates collapse. For parallel-plate capacitors the pull-in voltage V_{pi} is [8]

$$V_{\text{pi}} = \frac{2}{3} z \omega_r \sqrt{\frac{m}{1.5C_0}} \quad (7)$$

For $z = 1 \mu\text{m}$, $\omega_r = 2\pi \times 10 \text{ kHz}$, $m = 10^{-9} \text{ kg}$, and $C_0 = 100 \text{ fF}$, pull-in occurs at $V_{\text{pi}} = 3.4 \text{ V}$.

Electrostatic charging is another potential source of errors in the capacitive interface. Exposed insulating surfaces such as nitride or other oxides are prone to electrostatic charging and discharging. To the electronic circuits, these charges appear like signals, resulting in offset drift. Counter measures include strapping insulated sur-

faces with grounded conductors or packaging the sensor in a low-humidity environment to provide a leakage path for trapped carriers. A dc component in V_s should be avoided as the resulting field gives a preferred direction to carriers (e.g., ionized water) leading to a noisy potential gradient that can be picked up by the electronics.

6.1.2.3 Complementary Parallel Plate Interface

Complementary structures exhibit substantially lower offset and reduced susceptibility to errors by using two nominally equal sense elements that change in opposite direction. This can be achieved for example by adding a second fixed electrode above the movable plate [3]. Most planar processes have no provision for stacking a third electrode above the movable structure. In this case, complementary parallel-plate sensing can still be achieved with the planar comb structure shown schematically in Fig. 6.1.4. The sense capacitance consists of many individual capacitances between movable fingers and fixed electrodes connected in parallel to detect in-plane displacements.

Fig. 6.1.6 shows the equivalent electrical circuit diagram. The individual finger capacitances are shown as single lumped devices. The sample component values given are for an accelerometer built in a typical surface micromachining process with $2\ \mu\text{m}$ thick structural polysilicon and $2\ \mu\text{m}$ gaps. For such low aspect ratio, fringing fields add significantly to the sense capacitance compared with the value predicted by the parallel-plate capacitor Equation (1). Fig. 6.1.5 compares FEM simulation [10, 11] results [9] with parallel capacitor estimates as a function of gap x over film thickness t . For $x/t = 1$, fringing fields increase the sense capacitance by 41% and the change $\Delta C = dC/dx$ by 27%. Compared to a large parallel-plate capacitor, the sensitivity $\Delta C/C$ is 10% lower.

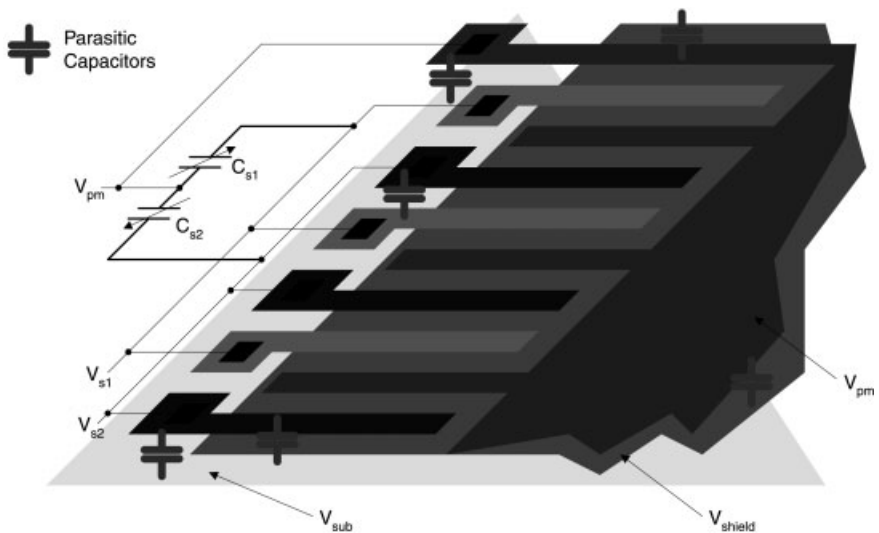


Fig. 6.1.4 Parallel-plate style comb interface

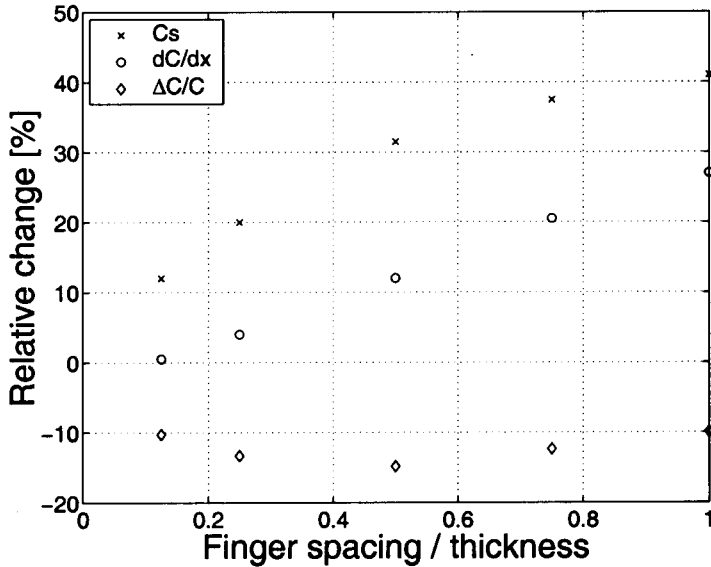


Fig. 6.1.5 FEM simulation of the capacitance of parallel plate style combs

The low ratio of layer-thickness to gap is also the reason for the relatively large parasitics, which in many designs exceed the value of the actual sense capacitance, negating some of the gain in sensitivity of the complementary over the single-ended design. Increasing the length of the fingers will increase the sense capacitance, but also decreases the resonant frequency of the individual fingers, and can lead to instability when feedback is applied [12].

The relatively large area of the sensor contributes substantial parasitic to the substrate, C_{Psub} , or shield, $C_{Pshield}$. The shield is used to attenuate vertical actuation of the structure by electrically shorting V_{pm} to V_{shield} . The shield also isolates the structure from noise coupling through the substrate. The capacitance from the shield to the substrate is often very large and care must be taken to provide a low resistance connection that does not interfere with the rate at which the voltage on the shield is varied by the sense electronics.

Also shown explicitly in Fig. 6.1.6 is the interconnection resistance R_{pm} of the movable structure. In sensors with long and skinny mechanical suspensions it is often large and can contribute significant electronic noise. Highly doped substrates and connecting all suspensions electrically in parallel reduce this source of error. Other interconnection resistances are often negligible and omitted from the circuit diagram. Silicon resistivity should be kept sufficiently low also to prevent depletion of the capacitor electrodes.

If the voltages used for sensing are balanced, complementary parallel-plate capacitor interfaces exert no net force in the sensitive direction, as long as the sense finger is exactly centered between the fixed electrodes. The structure is still sus-

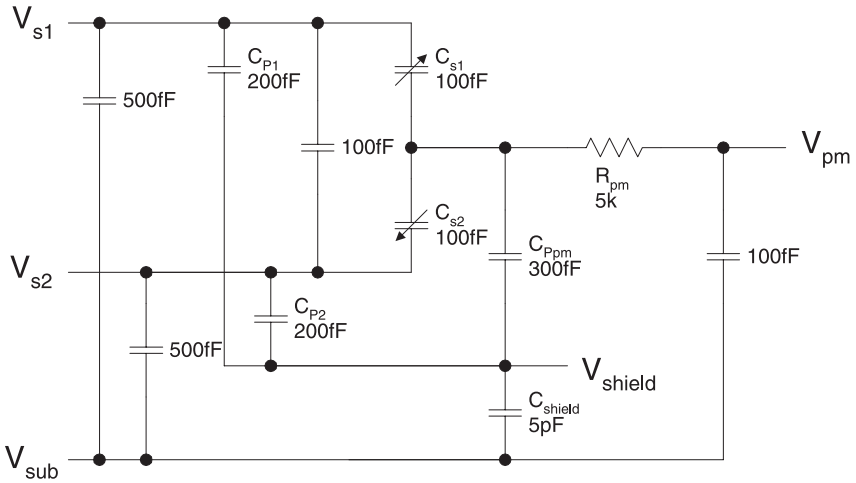


Fig. 6.1.6 Equivalent circuit for parallel-plate style comb interface

ceptible to the negative electrical spring effect and pull-in. Lateral combs reject these phenomena to first order.

6.1.2.4 Lateral Comb

Lateral combs exploit variation in electrode overlap rather than spacing. Fig. 6.1.7 shows a typical arrangement. A differential interface can be realized with combs on either side of a mechanical structure. Since the structure moves in the direction of the fingers, displacements in lateral combs are not limited by the gap between electrodes.

To first order, the capacitance of each finger is

$$C(x) = \epsilon \frac{t(x + x_0)}{d} \tag{8}$$

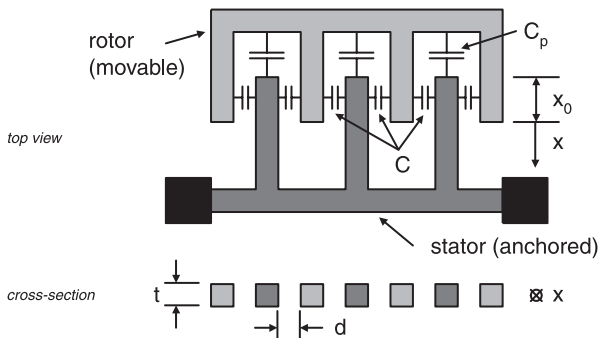


Fig. 6.1.7 Lateral comb interface

where x_0 is the nominal finger overlap and x is the displacement. Since the sensitivity dC/dx is independent of x the interface creates no electrical spring. Aside from edge effects, the structure is also not susceptible to pull-in in the x direction. Instead, sensors with lateral comb structures usually pull in by rotating about the z axis.

The penalty for using lateral combs is reduced sensitivity compared to parallel-plate type interfaces for a given spacing x

$$\left. \frac{\Delta C}{C} \right|_{\text{lateral comb}} = \frac{1}{x_0 + x}$$

$$\left. \frac{\Delta C}{C} \right|_{\text{parallel plate}} = \frac{1}{x}$$

Since, in most cases, the nominal overlap x_0 is much larger than the displacement x , parallel-plate interfaces have considerably larger sensitivity. Lateral combs are most often used for actuation where large range is required or in electromechanical oscillators, where the position-dependence of the negative spring of parallel-plate interfaces introduces nonlinearity [13]. Parallel-plate structures are usually preferred for maximum displacement resolution.

Lateral combs also exert a vertical uplift on the structure, an effect that can be used in situations where such a force is needed but electrodes above the structure are not available [4]. An appropriately engineered ground plane underneath the comb reduces the effect when it is not desired [14].

6.1.3

Circuits

6.1.3.1 Introduction

Broadly, two approaches are used for measuring capacitance. The first is analogous to an Ohm meter and realized, for example, by applying a voltage to the capacitor and measuring the resulting current. Since capacitors do not pass dc signals, an ac excitation is required. Alternatively the variable capacitor is used as a frequency-determining element in an oscillator circuit [5]. The discussion here focusses on the first technique, which is most widely used in automotive sensors.

6.1.3.2 Circuit Topologies

Fig. 6.1.8 shows conceptually the two most common topologies for sensing small capacitance changes. For single-ended interfaces, a reference capacitor takes the place of one of the sense capacitors. For practical reasons, an ac waveform is used as excitation V_s , resulting in an amplitude-modulated output V_0 that must be demodulated. Circuits for demodulation and biasing with the resistors R_{bias} will be discussed in Section 6.1.3.3. To first order they have no impact on circuit operation and are ignored in the conceptual analysis below.

In Fig. 6.1.8a the variable capacitors are configured as a capacitive voltage divider. Conceptually, the center electrode ‘samples’ the electrostatic field set up by the sense

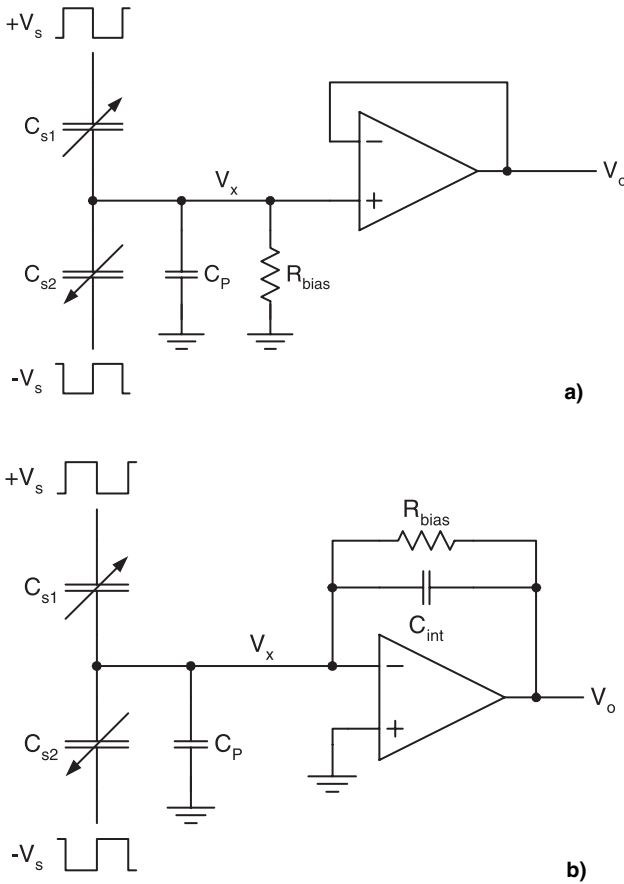


Fig. 6.1.8 Capacitive sensing with a) high, b) low impedance interfaces

voltage V_s . The second approach, depicted in Fig. 6.1.8b, uses an amplifier to provide a low impedance at the common node V_x . A difference between C_{s1} and C_{s2} results in a proportional current flowing into the integrating capacitance, C_{int} .

The voltages at the outputs of the two circuits are

$$V_{oa} = V_s \frac{\Delta C}{C_s} \frac{1}{1 + \frac{C_P}{2C_s}} \quad (9)$$

$$V_{ob} = V_s \frac{\Delta C}{C_s} \frac{2C_s}{C_{int}} \quad (10)$$

In both cases the output is proportional to $\Delta C/C_s$ and the sense voltage V_s . Increasing the sense voltage results in higher sensitivity but is limited by supply voltage, electrostatic force exerted on the variable capacitor, and pull-in.

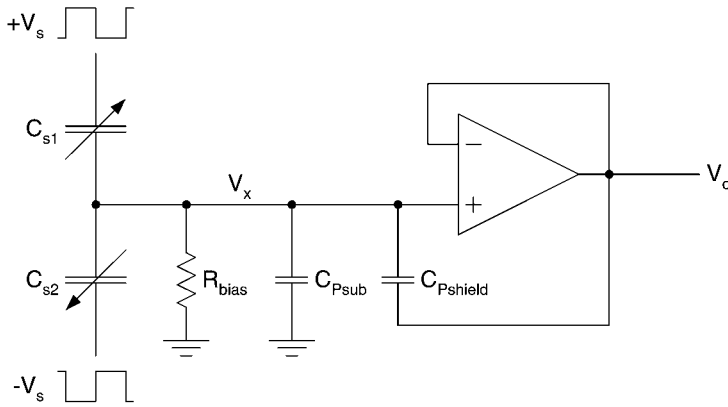


Fig. 6.1.9 Capacitive readout with driven shield

In the voltage divider case, parasitics significantly reduce the sensitivity when they become comparable to or larger than the sense capacitance. In Fig. 6.1.9 the parasitic is broken up into shielded and unshielded components, with the shield driven by the output of the unity-gain buffer amplifier. In this case, only the unshielded portion of C_P appears in Equation (9). In fact, the sensitivity of the circuit can be increased further by increasing the gain A_v of the amplifier slightly beyond unity to simulate a negative capacitance $C_{Pshield}(1 - A_v)$ at node V_x . This approach is not advisable, however, for several reasons. First, if the overall capacitance at node V_x approaches zero, the circuit will oscillate. Since the value of the parasitics is often poorly controlled, large margins are needed to avoid this possibility. Additionally, while the negative capacitance approach increases the output signal, it also increases the electronic noise, usually negating any potential benefit.

The output voltage of the circuit in Fig. 6.1.8b is independent of C_P . This result assumes ‘infinite’ amplifier gain, a condition that is usually met with sufficient accuracy. Equation (10) suggests that arbitrarily high sensitivity can be achieved by simply lowering the integrating capacitance C_{int} . Since, however, the closed-loop bandwidth of the amplifier is proportional to $\frac{C_{int}}{2C_s + C_P + C_{int}}$, small values of C_{int} usually result in unacceptable speed. A practical value is $C_{int} \approx 2C_s + C_P$ [16]. Note also that C_P includes the input capacitance of the amplifier. Decreasing C_{int} substantially below the suggested value in turn requires a faster amplifier with consequently larger input capacitance.

Fig. 6.1.10 shows a variation of the current sense approach where the amplifier is connected differentially to the sense capacitors. Compared to the single-ended version in Fig. 6.1.8b, the differential circuit has improved power supply rejection ratio (PSRR) and ability to reject errors such as substrate noise or coupling from other signals. This is because any error that couples into both inputs of the amplifier equally will be rejected by the amplifier’s common-mode rejection ratio (CMRR).

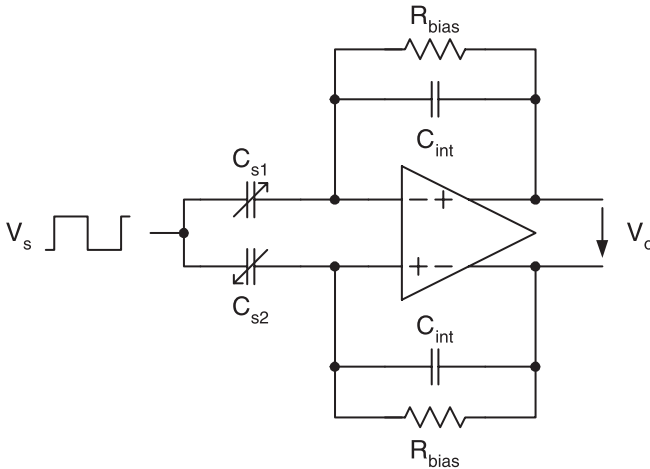


Fig. 6.1.10 Pseudo-differential capacitive readout

Whether or not this rejection, and hence a differential connection, are needed depends on the application. Chips containing multiple sensors or two-chip solutions in which the interface between the sensor and the electronics passes over bond-wires are obvious candidates to gain from this arrangement. The relative size of parasitics in the differential versus the single-ended arrangement should also be considered when deciding which topology to use.

The differential solution adds some additional circuit complexity in the form of auxiliary feedback loops to set the common-mode voltage at the amplifier input and output [17, 18].

6.1.3.3 Modulation and Biasing

Since capacitors do not pass dc current, special steps must be taken to set the charge stored on each capacitor and protect the circuit from drift due to leakage currents, carrier generation, etc. The resistors in Fig. 6.1.8 labelled R_{bias} serve this purpose. Additionally, measurements are usually not performed at dc, but rather at some carrier frequency f_M , typically a few 100 kHz or MHz. The higher frequencies are needed especially with CMOS circuits to reject the electronics' own low frequency flicker noise which often extends to the low MHz. The only exception are applications in which the signal itself is modulated or free from dc content, such as resonant gyroscopes [1] or microphones. Even in those cases modulation is often used.

Two different schemes are used for modulation: chopper stabilization, which works similarly to an AM radio, and correlated double sampling (CDS), which performs two measurements of the noise only and the noise plus signal [19]. Fig. 6.1.11 shows conceptual diagrams of the two techniques. Although the additional circuitry looks complex, actual implementations are surprisingly simple [20, 21].

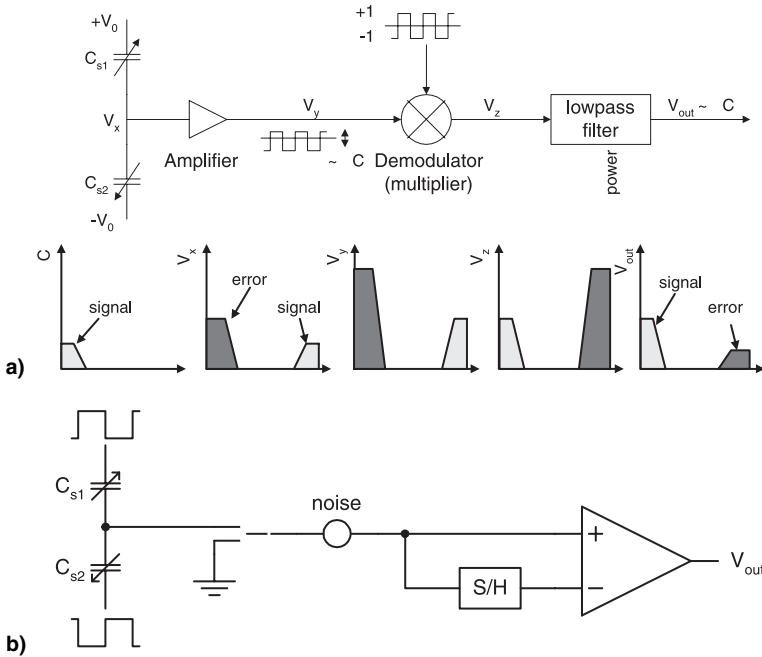


Fig. 6.1.11 Modulation techniques: a) chopper stabilization, b) correlated double sampling

CDS is somewhat simpler since no demodulator and filter are needed. However, since the signal is sampled, CDS suffers from noise aliasing. Unless eliminated with appropriate circuit techniques [22, 23], this will result in an elevated noise floor. Chopper stabilization is preferred in printed circuit board implementations since all necessary components are readily available.

The biasing resistors R_{bias} shown in Fig. 6.1.8 set the charge on the sense and integrating capacitors. Without them, the node connected to the amplifier input would drift, causing the amplifier output to rail. While the value of the resistor is not critical, it usually must be several $M\Omega$ or $G\Omega$ so that its thermal noise will not reduce the sensor resolution. Such large resistors are difficult to integrate in standard CMOS processes and require large area. In one commercial acceleration sensor [24], the area of the biasing resistor is larger than that of the sensor.

Various circuits are used to simulate a large resistor with active components that take up significantly less chip area. A particularly simple solution consists of a switch that periodically charges the sense capacitors to a known value. This solution can be conveniently implemented in combination with correlated double sampling. Since the sensor capacitance is often small, the noise aliased by this re-setting can be significant [22].

MOSFETs operating in the triode region can also be used as resistors. Fig. 6.1.12 shows a simplified circuit diagram of a solution used in a surface-microma-

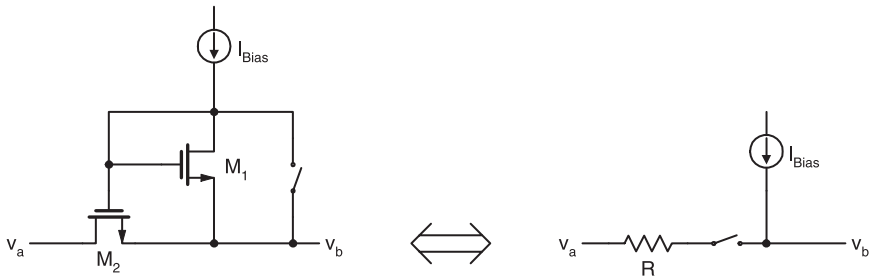


Fig. 6.1.12 MOSFET simulating a high valued resistor

chined gyroscope [1]. The current I_{Bias} is set by a reference such that the transconductance of M_1 is $g_{m1} = 1/NR_{\text{ref}}$, where N is a device ratio and R_{ref} a physical resistor in the bias circuit. Device M_2 has a much smaller aspect ratio than M_1 . Since the sensor output is small, M_2 is in the triode region with resistance

$$R = \frac{(W/L)_2}{(W/L)_1} N R_{\text{ref}} \quad (11)$$

For typical values, $\frac{(W/L)_2}{(W/L)_1} = 100$, $N = 10$, and $R_{\text{ref}} = 20 \text{ k}\Omega$, $R = 20 \text{ M}\Omega$. Much higher values than this are limited in practice by matching. Especially for processes with thin oxide, the overdrive voltage of M_1 approaches typical threshold voltage mismatch if R_{ref} is made much larger. A better solution for achieving even higher resistance is duty-cycling M_2 with the switch indicated in Fig. 6.1.12, which in practice would also be implemented with a MOSFET. For an on/off ratio of 1/50, the equivalent resistance R increases to $1 \text{ G}\Omega$. The bias current flows into terminal v_b , which is usually connected to the output of an amplifier where it is absorbed without corrupting the signal.

6.1.3.4 Noise

Most capacitive readout circuits are extremely sensitive. For example, a 1 aF change of a capacitor biased at 1 V corresponds to a charge difference of only six electrons, but can readily be detected by many readout circuits. The high sensitivity translates into an equally high susceptibility to errors and noise.

Since stray capacitors to substrate, the supplies, or other conductors are usually significantly larger than the detection limit of the sensor, an unusually high degree of care must be taken to minimize capacitive coupling. In addition to minimizing stray capacitors, measures include electrostatic shields, differential structures, and correlated double sampling (CDS) or chopper stabilization. After successful mitigation of these errors, circuit performance is limited ultimately by electronic noise.

Electronic noise manifests itself as random fluctuations of voltage or current. Two kinds of electronic noise are relevant to capacitive readout circuits: flicker or $1/f$ noise, and thermal noise. The former dominates at low frequencies and can be

thought of as a slowly varying offset voltage. The magnitude of flicker noise is highly device specific. JFETs exhibit orders-of-magnitude lower flicker noise than MOSFETs. Unfortunately JFETs are not available in most technologies, and even the selection of stand-alone amplifiers with JFET inputs is limited. Fortunately, chopper stabilization and CDS, introduced in the previous section, successfully reject flicker noise to the point where it is negligible compared to other noise sources.

Thermal noise, on the other hand, is a fundamental property of all dissipative elements such as resistors or transistors. For example, a thermal noise current source appears in parallel with all physical resistors. While the instantaneous value of the noise current cannot be predicted, physics accurately predicts its statistics. The mean of the noise current is zero, and the variance is given by

$$\overline{i_n^2} = \frac{4k_B T}{R} \quad (12)$$

In this equation, $k_B T \approx 4 \times 10^{-21}$ J are the Boltzmann constant and absolute temperature (290 K) and R is the value of the resistor. For $R = 1 \text{ M}\Omega$, $\overline{i_n^2} = 1.6 \times 10^{-26} \text{ A}^2/\text{Hz}$. The unit 'A²/Hz' reflects the fact that noise is distributed evenly up to very high frequencies. Only the fraction that is within the bandwidth of the sensor and its electronics actually corrupts the signal.

For optimum noise performance the bandwidth of a sensor and its evaluation electronics should be limited to a value no larger than that required by the application. For example, an accelerometer with a noise floor of 1 mG ($1 \text{ G} = 9.8 \text{ m/s}^2$) in a 1 kHz bandwidth will resolve 500 μG when its bandwidth is reduced to 250 Hz. Further bandwidth reduction results in additional improvement up to the point where a different noise source dominates (e.g., flicker noise).

For illustration, the circuit from Fig. 6.1.10 is redrawn in Fig. 6.1.13 with the noise sources v_n and i_n from the amplifier and bias resistor shown explicitly. The variances are determined from Equation (12) for the resistor, and the datasheet of the amplifier. The noise specification of amplifiers includes also an input referred current noise, which (for amplifiers with CMOS inputs) is usually negligible, and for amplifiers with bipolar input is so large that BJT input stages simply cannot be used in most capacitive readout circuits.

At the frequencies of interest, the variance of the noise at the output of the amplifier is

$$\overline{v_{on}^2} = \underbrace{\left(1 + \frac{2C_s + C_P}{C_{int}}\right)^2}_{\text{amplifier}} \overline{v_n^2} B + \underbrace{\left(\frac{1}{2\pi f_M C_{int}}\right)^2}_{\text{resistor}} \overline{i_n^2} B \quad (15)$$

The terms in parentheses are the noise transfer functions from the respective source to the output V_0 of the circuit and B is the bandwidth of the filter. Since the noise is proportional to the bandwidth, B should be chosen as small as permitted by the application.

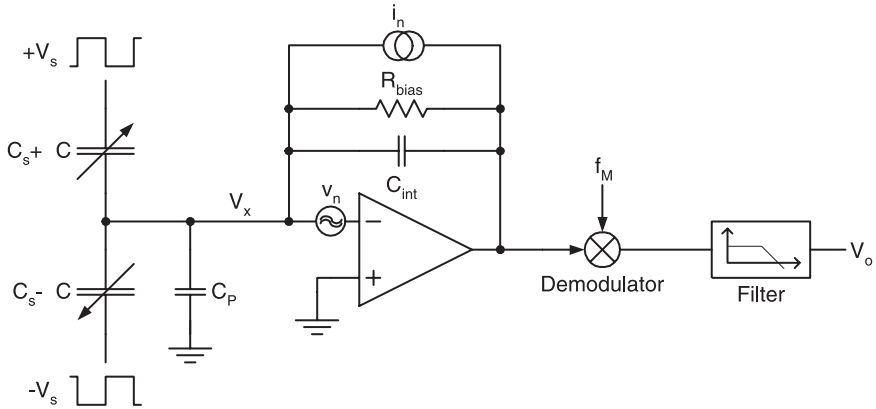


Fig. 6.1.13 Electronic noise sources in capacitive readout circuit

Assuming $R_{bias} = 20 \text{ M}\Omega$, all capacitor values equal to 2 pF , $\overline{v_n^2} = (5 \text{ nV}/\sqrt{\text{Hz}})^2$ (a typical value for a good low noise amplifier), and $f_M = 1 \text{ MHz}$, the noise contributions from the resistor and amplifier are $2.3 \text{ nV}/\sqrt{\text{Hz}}$ and $20 \text{ nV}/\sqrt{\text{Hz}}$, respectively. In this example the noise from the biasing resistor is negligible compared to the contribution from the amplifier, thanks to the high value of R_{bias} .

Dividing the noise at the output by the gain in Equation (10) results in the ability of the circuit to resolve change of capacitance

$$\sqrt{\Delta C^2} = \frac{C_{int}}{2V_s} \left(1 + \frac{2C_s + C_p}{C_{int}} \right) \sqrt{B} \sqrt{\overline{v_n^2}}. \quad (14)$$

The noise from R_{bias} has been omitted for simplicity. Using the same parameters as above and $V_s = 1 \text{ V}$ and $B = 1 \text{ kHz}$, the standard deviation of the equivalent input noise of the circuit is 0.63 aF . Since noise follows a Gaussian distribution, there is a 32 percent probability that a particular error sample is larger than this value. Datasheets usually specify a three times larger value, 2 aF , to reduce the probability for larger errors to 0.3%. Note that thermal noise from the amplifier is the only error source included in this analysis. In practice, other error sources are often relevant also.

Equation (14) confirms the earlier observations that to maximize resolution, parasitic capacitance must be kept small. In fact, to first order the resolution is proportional to $2C_s + C_p$. Apparently, thermal noise considerations impose no penalty on a fully integrated sensor with small sense and parasitic capacitance, since resolution improves by the same factor. Only the combination of small sense capacitance and large parasitic (e.g., due to connection to off-chip electronics) suffers from reduced sensitivity.

6.1.3.5 Examples

Capacitive readout circuits, such as that shown in Fig. 6.1.13 can be readily implemented with off-the-shelf components using operational amplifiers for single ended [25, 26] or differential designs [27, 28]. Special purpose chips for capacitive sensing [29, 30] are also available.

A monolithic three-axis accelerometer with three independent capacitive readout circuits on a single chip is described elsewhere [7] (Fig. 6.1.14). The circuit is similar to Fig. 6.1.10 and achieves $0.085 \text{ aF}/\sqrt{\text{Hz}}$ resolution with 100 fF sense capacitors. The noise is actually dominated by Brownian noise in the sensor itself, as tests in vacuum demonstrate. The actual capacitance resolution is therefore somewhat better than stated. This circuit uses correlated double sampling (CDS) for biasing and to reject flicker noise.

Fig. 6.1.15 shows the operation of the circuit in three phases. For clarity, devices that are inactive in a particular phase are omitted from the diagram. The first phase resets the sense capacitors by shorting the sense capacitors to ground. This solution circumvents the need for a bias resistor. When the switches are opened, a random voltage with standard deviation $\sqrt{k_B T/C_s}$ is stored on each sense capacitor. Given a sense capacitance of only 100 fF, this standard deviation is 200 μV at room temperature and hence much larger than the minimal signal. Geometric constraints on the sense capacitor interface preclude enlarging C_s to attenuate this error. Instead, care is taken in subsequent phases to prevent the error voltage from corrupting the signal.

Fig. 6.1.15 b shows the circuit topology in the offset cancellation phase. The proof-mass is still tied to ground and ideally (assuming the value of C_s has not changed) the differential output voltage of the integrator is zero. In practice it is corrupted by offset and flicker noise from the amplifier, switch charge injection errors, and the $k_B T/C$ noise stored on the sense and integrating capacitors. The resulting error voltage is stored on capacitors C_H . When these capacitors are disconnected at the end of the offset cancellation phase, all errors are cancelled to first order with the exception

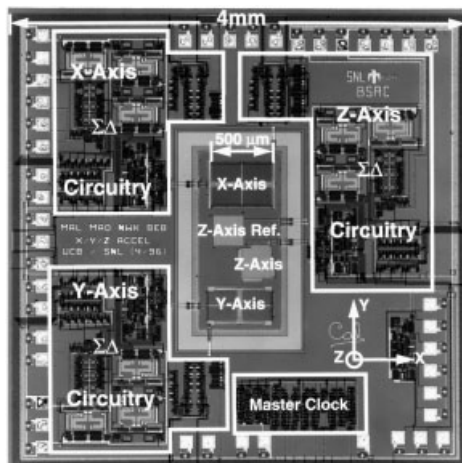


Fig. 6.1.14 Die photograph of a monolithic three-axis accelerometer [7]

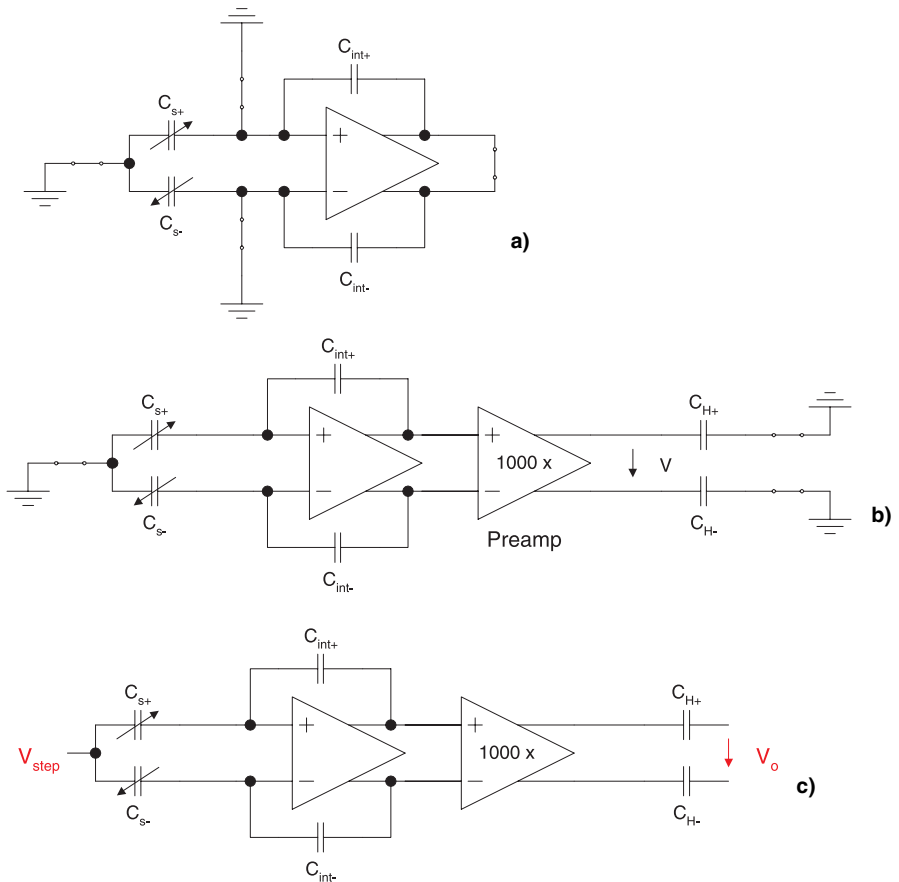


Fig. 6.1.15 Timing of correlated double sampling (CDS): a) circuit reset, b) sampling noise cancellation, c) sensing

of the $k_B T/C_H$ noise sampled onto C_H . This residual error can be made small by appropriately sizing C_H . The signal-to-noise ratio is boosted further by the gain of the preamplifier. Additional offset cancellation circuitry not shown in Fig. 6.1.15 prevents the high-gain preamplifier from railing.

Finally the proof-mass potential is raised to V_{step} (Fig. 6.1.15c). Only a mismatch of the sense capacitors (or other mismatch in the circuit, e.g., between the integrating capacitors) results in a non-zero differential output voltage ΔV_0 . Note, however, that for identical sense capacitors the output is zero, as is required. Thermal noise sampled onto C_S and C_{int} is effectively cancelled.

Many other integrated solutions are documented in the literature. These include single-ended and differential solutions employing chopper stabilization [1, 20, 24, 31], circuits using correlated double sampling [2, 3, 32] and applications for capacitive readout circuits and corresponding implementations [15].

6.1.4

References

- 1 J. F. CHANG, J. A. GEEN, S. J. SHERMAN, S. R. LEWIS, in *ISSCC Dig. Tech. Papers*, San Francisco, **2002**, pp. 426–427.
- 2 X. JIANG, F. WANG, M. KRAFT, B. E. BOSER, in *Proc. IEEE Solid-State and Actuator Workshop*, Hilton Head Island, SC, **2002**, pp. 202–205.
- 3 T. SMITH, O. NYS, M. CHEVROULET, Y. DECOULON, M. DEGRAUWE, in *ISSCC Dig. Tech. Papers*, J. WUORINEN (ed.), San Francisco, **1994**, pp. 160–161, IEEE.
- 4 C. LU, M. LEMKIN, B. E. BOSER, *IEEE J. Solid-State Circuits* **1995**, 30 (12), 1367–1373.
- 5 F. RUDOLF, S. SOUILLAC, A. HOOGWERF, T. SMITH, M. CHEVROULET, B. DE GEETER, in *Proc. Sensor Expo.*, Detroit, **1997**, pp. 15–20.
- 6 ZHANG YAFAN, S. MASSOUD-ANSARI, MENG GUANGQING, KIM WOOJIN, N. NAJAFI, in *Technical Digest. MEMS 2001*, Interlaken, Switzerland, **2001**, pp. 166–169.
- 7 M. LEMKIN, B. E. BOSER, *IEEE J. Solid-State Circuits* **1999**, 34(4), 456–468.
- 8 O. BOCHOBZA-DENGANI, Y. NEMIROVSKY, *J. MEMS* **2001**, 10(4), 601–615.
- 9 M. A. LEMKIN, PhD dissertation, UC Berkeley, **1997**.
- 10 K. NABORS, J. WHITE, *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems* **1991**, 10(11), 1447–1459.
- 11 Ansoft Corporation, *Maxwell Solver, Electrostatic Package*, **1996**.
- 12 J. I. SEEGER, X. JIANG, M. KRAFT, B. E. BOSER, in *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, **2000**, pp. 296–299.
- 13 C. T. C. NGUYEN, in *Proc. 1995 IEEE Ultrasonic Symposium. An International Symposium*, vol. 1, M. LEVY, S. C. SCHNEIDER, B. R. McAVOY (eds.), Seattle, WA, USA, **1995**, pp. 489–499.
- 14 W. C. TANG, M. G. LIM, R. T. HOWE, *J. MEMS* **1992**, 1(4), 170–178.
- 15 L. K. BAXTER, *Capacitive Sensors – Design and Applications*, Electronics Technology, IEEE, New York, **1997**.
- 16 W. M. C. SANSEN, Z. Y. CHANG, *IEEE Trans. Circuits Systems* **1990**, 37(11), 1375–1382.
- 17 P. R. GRAY, P. J. HURST, S. H. LEWIS, R. G. MEYER, *Analysis and Design of Analog Integrated Circuits*, 4th edn. Wiley, New York, **2001**.
- 18 B. RAZAVI, *Design of Analog CMOS Integrated Circuits*, McGraw-Hill, Boston, **2001**.
- 19 C. C. ENZ, G. C. TEMES, *Proc. IEEE* **1996**, 84(11), 1584–1614.
- 20 J. WU, G. K. FEDDER, L. R. CARLEY, in *ISSCC Dig. Tech. Papers*, IEEE, San Francisco, **2002**, pp. 428–429, IEEE.
- 21 B. E. BOSER, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, Chicago, **1997**, pp. 1169–1172.
- 22 B. E. BOSER, in *RF Analog-to-Digital Converters; Sensor and Actuator Interfaces; Low-Noise Oscillators, PLLs and Synthesizers*, W. M. C. SANSEN, R. J. VAN DE PLASSCHE, J. H. HUIJSING (eds.), Kluwer, Dordrecht, **1997**.
- 23 X. JIANG, J. I. SEEGER, M. KRAFT, B. E. BOSER, in *Proc. 2000 Symposium on VLSI Circuits*, Hawaii, **2000**, pp. 16–19.
- 24 S. J. SHERMAN, W. K. TSANG, T. A. CORE, D. E. QUINN, in *Digest of Technical Papers, 1992 Symposium on VLSI Circuits*, **1992**, pp. 34–35, IEEE.
- 25 Texas Instruments, *TLC226x advanced lincmos rail-to-rail operational amplifiers*, Datasheet, Texas Instruments, **2001**.
- 26 Maxim, *MAX4475 low-noise, low-distortion, wide-band, rail-to-rail op amps*, Datasheet, Texas Instruments, Maxim, **2001**.
- 27 Texas Instruments, *THSA4120, THSA4121 high-speed fully differential I/O amplifiers*, Datasheet, Texas Instruments, **2001**.
- 28 Analog Devices, *AD8132, low-cost, high-speed differential amplifier*, Datasheet, **2002**.
- 29 Analog Microelectronics, *CAV414 converter ic for capacitive signals*, Datasheet, **2001**.
- 30 Micro Sensors, *MS3110 universal capacitive readout IC*, Datasheet, **2001**.
- 31 S. J. SHERMAN, W. K. TSANG, T. A. CORE, R. S. PAYNE, D. E. QUINN, K. H. L. CHAU, J. A. FARASH, S. K. BAUM, in *Technical Digest Int. Electron Devices Meeting 1992*, San Francisco, CA, USA, **1992**, pp. 501–504, IEEE.
- 32 N. YAZDI, A. MASON, K. NAJAFI, K. D. WISE, *Sens. Actuators A* **2002**, 84(3), 351–361.

6.2

Resistive Evaluation Circuits

RAINER KIENZLER and OLIVER SCHATZ

6.2.1

Introduction

One of the oldest evaluation techniques for detecting small resistive changes is the ‘Wheatstone bridge’ method. The sensor designer seeks to arrange selected piezoresistive measurement elements in a Wheatstone bridge configuration. This makes it possible for imperfections and disturbances to be cancelled out by matching the resistors, and gives high sensitivity to the differential signal.

The differential signal must be further amplified, adjusted, and filtered to make it into the desired waveform and signal level. This can be done with the well-known analogue integrated circuit design technique, in which the equality of active and passive components placed near each other on the chip is used to amplify only the differential signal. This is called the matching technique. Trimming techniques are also used to compensate for unwanted influences.

Alternatively, the analogue signal can be converted immediately to a digital bit stream. Subsequently mathematical cancellation algorithms are used to correct non-ideal effects. This is best done with $\Delta\Sigma$ A/D-techniques in which a high sampling rate gives accurate signal detection.

This can further be enhanced by the so-called ‘lock-in technique’, in which modulated signals are used, which after demodulation separate the desired signal from the noise. Both techniques are described here by examples: one is a monolithic pressure sensor with moderate piezoresistive bridge signals and the other is a signal evaluation circuit for a high-pressure sensor with very small sensor signal levels.

6.2.2

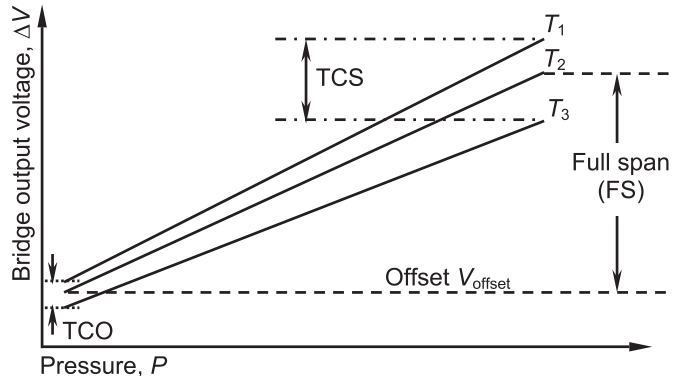
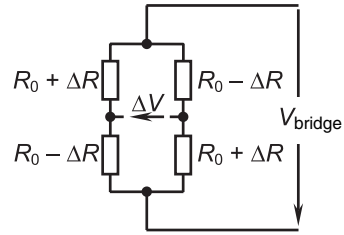
Requirements of a Piezoresistive Wheatstone Bridge

Before describing the signal evaluation principles, the general observed imperfections of the piezoresistive bridge, which must be taken care of in the evaluation circuit, will be discussed.

Four diffused piezoresistors located on a diaphragm are connected in a Wheatstone bridge configuration (Fig. 6.2.1). An applied pressure causes a resistance change in each resistor due to the piezoresistive effect. The output of the Wheatstone bridge (ΔV) supplied by voltage V_{bridge} is processed by an integrated circuit.

Both the sensitivity and the offset of the bridge are functions of temperature (Fig. 6.2.2). The temperature coefficient of sensitivity (TCS) is always negative and relatively constant (typically $-0.2\%/K$). The bridge offset creates the largest part of the temperature coefficient of the bridge offset (TCO) and is caused by the TCS compensation technique used.

Fig. 6.2.1 Equivalent circuit of the Wheatstone bridge

Fig. 6.2.2 Characteristics of an uncalibrated Wheatstone bridge supplied with constant voltage V_{bridge} (temperature $T_1 < T_2 < T_3$)

The sensitivity (slope of the characteristic in Fig. 6.2.3) is temperature-independent after calibration of TCS. The parallel shift is called TCO and is caused by the electrical offset of the bridge. This offset is proportional to the bridge supply voltage (V_{bridge}). Because the bridge supply voltage is varied with temperature to com-

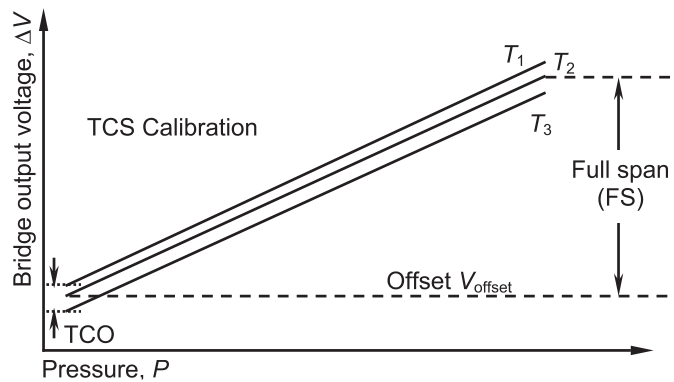


Fig. 6.2.3 Characteristics of a Wheatstone bridge after calibration of temperature coefficient of sensitivity

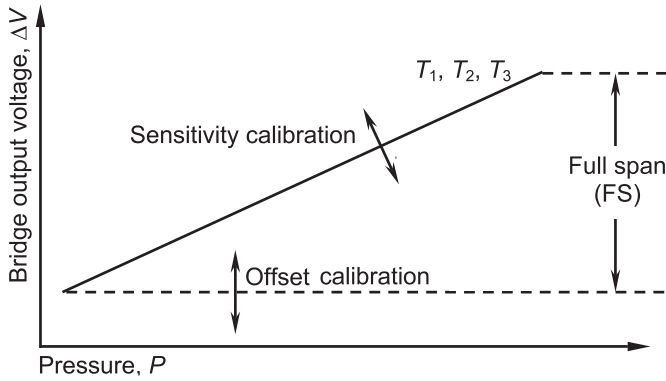


Fig. 6.2.4 Characteristics of a Wheatstone bridge after calibration of temperature coefficient of bridge offset

compensate for TCS, TCO is increased accordingly. The value and sign of TCO compensation is hence dependent on the electrical bridge offset.

During TCO calibration the characteristics at different temperatures are shifted onto a single trace, which is temperature-independent (Fig. 6.2.4). The slope (sensitivity) and the offset will be calibrated for the desired application.

The major parameters are defined as follows.

Sensitivity

$$S = \frac{\Delta V(P) - \Delta V(P_0)}{(P - P_0) \cdot V_{\text{bridge}}} \left[\frac{\text{mV}}{\text{V} \cdot \text{Pa}} \right]$$

Full Span

$$\text{FS} = S \cdot (P - P_0) \cdot V_{\text{bridge}} \text{ [mV]}$$

Temperature Coefficient of Sensitivity

$$\text{TCS} = \frac{\text{FS}(T) - \text{FS}(T_0)}{\text{FS}(T_0)} \frac{1}{T - T_0} \left[\frac{1}{\text{K}} \right]$$

Temperature Coefficient of Offset (Pressure, $P = 0$)

$$\text{TCO} = \frac{\Delta V(T) - \Delta V(T_0)}{\text{FS}(T_0)} \left[\frac{1}{\text{K}} \right]$$

where

T = temperature

T_0 = reference temperature

P = pressure

P_0 = reference pressure

An evaluation circuit must fulfill several requirements:

- compensation of TCS (increasing of V_{bridge} over temperature, +0.2%/K)
- compensation of offset
- compensation of TCO
- voltage amplifying (sensitivity)
- delivery of output current (source, load).

Such evaluation circuits can be analog or digital. One example of each type is described here.

6.2.3

Example of Analog Evaluation Circuit

6.2.3.1 General Description

Four piezoresistors connected to a Wheatstone bridge are supplied with an adjustable voltage by the bridge voltage supply, OP1. The resulting bridge output voltage is proportional to the applied pressure. A two-stage analog amplifier consisting of the V/I signal amplifier and the final amplifier, OP2, amplifies the sensor signal. The output signal is proportional to the VCC voltage (Fig. 6.2.5).

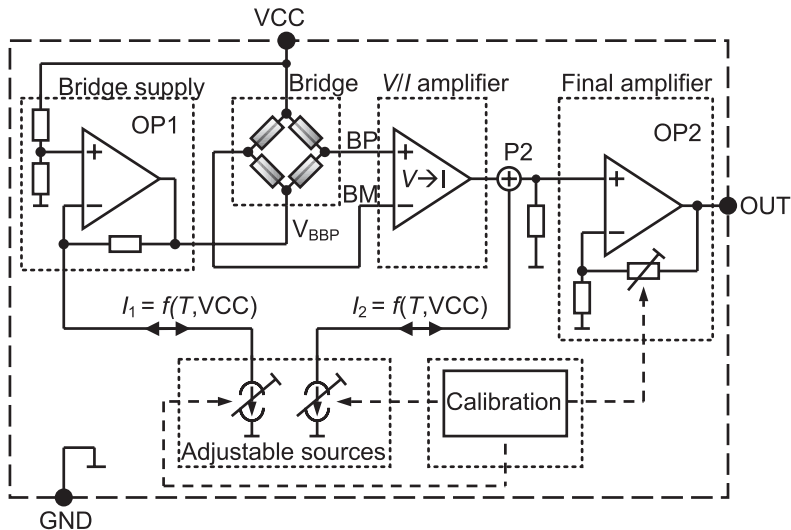


Fig. 6.2.5 Block diagram of analog signal processing

The bridge and the evaluation circuit parameters are not stable over wafer-lot variations. For this reason the desired pressure versus output signal relation is adjusted by an electrical calibration. The TCS of the bridge is adjusted by variation of the bridge supply voltage via OP1. Therefore, a temperature-proportional and adjustable current, I_1 , is fed into the negative input of OP1. This current flows through the feedback loop resistor to the output terminal. The positive temperature coefficient of the current I_1 causes a decrease in the bridge base potential (V_{BBP}) with increasing temperature. The output voltage of the bridge at the same pressure is now constant over temperature. Offset voltage and TCO are adjusted by a variable current I_2 supplied to OP2. The sensitivity (total gain) is adjusted by the final amplifier OP2. The calibration procedure of the sensor is controlled by a logic circuit. The control logic selects the desired parameter and adjusts the calibration currents and the gain.

6.2.3.2 Function of Main Circuit Blocks

Bridge Supply Voltage

The non-inverting input of the operational amplifier (Fig. 6.2.6) is connected with a resistor divider to the supply voltage. Any current flow into the inverting input causes a change in output voltage V_{BBP} , which is proportional to the current. To achieve a supply voltage and temperature-proportional current I_1 an analog current multiplier is used.

V/I Signal Amplifier

The bridge output voltage (BP-BM) is buffered by two voltage followers (Fig. 6.2.7) and is converted into a current by a resistor. The resulting current $I_{V/I}$ is proportional to the bridge output voltage.

Sumpoint P2

The V/I signal amplifier current output $I_{V/I}$ and the TCO and offset compensation current I_2 are added at the sumpoint P2 (Fig. 6.2.8). The offset compensation current is a function of VCC; the TCO compensation current is a function of temperature and VCC. The circuit for TCO compensation is equivalent to the circuit of the TCS compensation. Additionally a sign bit for positive or negative gradient

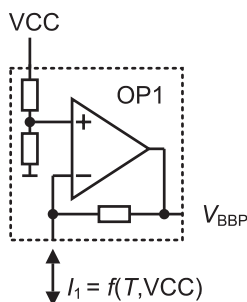


Fig. 6.2.6 Bridge supply

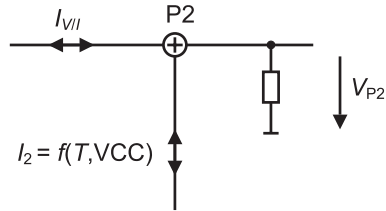
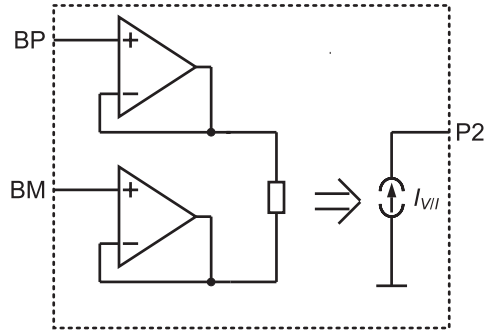
Fig. 6.2.7 V/I signal amplifier

Fig. 6.2.8 Sumpoint P2

is implemented. The added current is converted back into a proportional voltage V_{P2} in the resistor at P2.

Final Amplifier

The gain of the non-inverting final amplifier (Fig. 6.2.9) is adjusted with an trimmable resistor network in the feedback loop. To detect short circuits between the output and ground or supply voltage two active output limiters are included in the final stage.

Adjustable Sources for TCS, Offset, and TCO

Each weight current is active by default and can be deactivated during sensor calibration by zapping of the specific thyristor (Fig. 6.2.10). The remaining active cur-

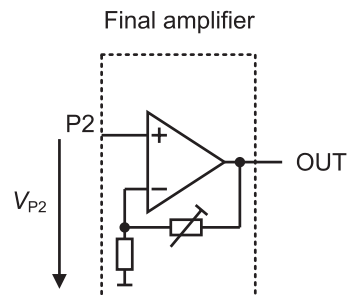


Fig. 6.2.9 Final amplifier

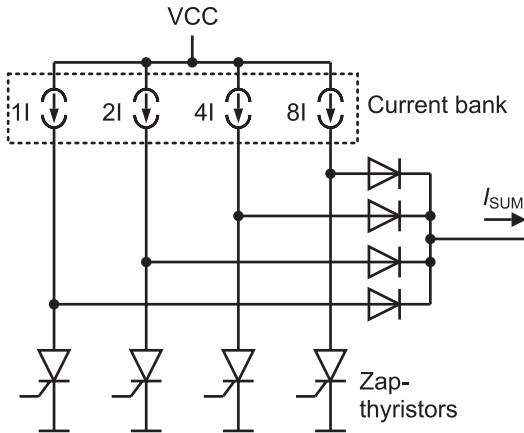


Fig. 6.2.10 Adjustable current source

rents are summarized in one common node before being fed into the next circuit stage. The thyristor is an open circuit when not zapped and is permanently shortened via a zap pulse. After zapping of the thyristor the output of the current source is shortened to ground. Thyristor zapping is done during sensor calibration. Therefore, an additional pin for zap current input is necessary (not in the drawings).

The calibration technique used makes trimming of the sensor possible at the end of the fabrication process (including housing of the sensor). Furthermore only electrical signals are required (no laser cutting is needed). The calibration data are stored permanently on the chip.

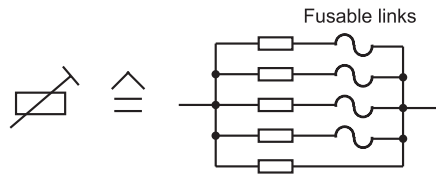
During the calibration process thyristors are triggered prior to zapping to test their influence on the output signal. There are other possibilities for calibration trimming.

- Resistor trimming with laser cutting (continuously adjusting resistor values).
- Diode zapping instead of thyristor zapping (additional circuit necessary).
- Oxide zapping (short circuited small on-chip capacitors).
- EPROM (storing electrical charge on floating MOS gates).
- Fusible links (Zener zap, metal bridge blowing, polyresistor zap).

Calibration Circuit of the Final Amplifier

A trimmable resistor network in the feedback loop of the operational amplifier OP2 is aligned during the calibration process. The total resistance of the resistor network is increased by opening fusible links (Fig. 6.2.11). A test calibration is not possible. At the end of the hole trimming a special link is fused to deactivate the complete control logic for safety reasons. Consequently, accidental trimming is prevented.

Fig. 6.2.11 Final amplifier calibration circuit



6.2.3.3 Summary

This example shows an analog solution with a V/I converter. Another possibility is an instrumental amplifier instead of the V/I converter. The analog evaluation circuit is economical relating to wafer-area, but there are a lot of design features necessary for a high accurate signal evaluation. All circuit blocks must be designed to be insensitive to temperature and mechanical stress.

Fig. 6.2.12 shows an example of a monolithic analog evaluation circuit with integrated diaphragm. The circuit blocks are made as described above. Typical parameters of the sensor are as follows.

- Sensitivity of the bridge: $20 \text{ mV/V} \cdot \text{bar}$.
- Voltage gain: 38 dB.
- Accuracy: 1%.
- Temperature range: -40 to 125°C .

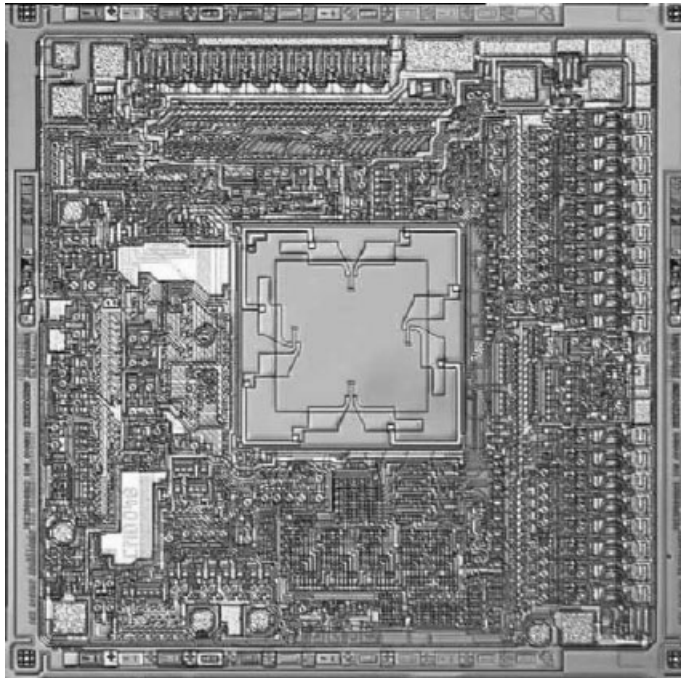


Fig. 6.2.12 Analog integrated pressure sensor

6.2.4

Example of Digital Evaluation Circuit

The application-specific integrated circuit establishes a single-chip sensor evaluation for high-pressure gauges. On the electrical side these gauges consist of a resistive Wheatstone bridge delivering a very tiny signal (typically 2 mV/V supply voltage). For high-precision applications an evaluation scheme with high offset and gain stability is necessary (Fig. 6.2.13).

For this reason the ASIC concept is based on the modulation of the pressure measurement signal to a carrier frequency f_m of approximately 5 kHz by square wave voltage supply of the gauge (H-bridge circuit). The signal is preamplified and then converted into a single-bit data stream with a data rate of $f_{\Delta\Sigma} = 256 f_m$ (about 1.28 Mbps) by a second-order $\Delta\Sigma$ modulator. The analog front-end (preamp and $\Delta\Sigma$) is produced in fully differential SC circuitry.

In the digital domain the signal is decimated to a rate of $4 f_m$ (at 13 bit width) by a sinc^3 decimation filter and then demodulated by multiplication with a sine of frequency f_m . Careful adjustment of the phases results in a sampled sine with sample values 1, 1, -1, -1, 1, 1, -1, -1,

The signal band (formerly centered at f_m) is thus shifted to dc. The following digital FIR filter with transmission zeros at f_m , $2 f_m$, and $3 f_m$ cancels out the dc offsets and $1/f$ noise of the analog front-end (now centered at f_m) as well as some overtones of the original square wave aliased into the Nyquist band of the signal sampled at $4 f_m$.

Signal conditioning is done by addition of a 13 bit offset and multiplication of a 13 bit factor. Thus, production tolerances of the gauge can be accounted for. The

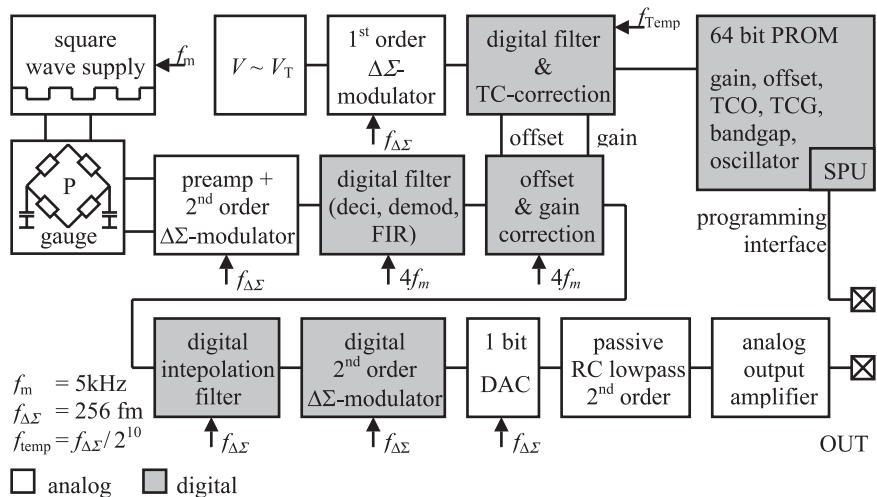


Fig. 6.2.13 Concept of a high-pressure sensor

corrections reduce the dynamic signal range. So the word length can losslessly be reduced from 13 to 11 bit.

In order to regain an analog signal, first the sample rate is increased again to $f_{\Delta\Sigma} = 256 f_m$ in a digital interpolation filter, which interpolates piecewise linearly between the 4 f_m samples. Hereafter it is converted to a single bit data stream by a digital second-order $\Delta\Sigma$ modulator. The 1 bit digital-to-analog conversion is done by a simple CMOS inverter switching between ground and the ratiometric supply voltage V_S . Most of the quantization noise (located at higher frequencies due to the $\Delta\Sigma$ modulation) is cut off by a second-order passive RC lowpass. The analog signal is amplified by 1.68 and driven at the output at low impedance. But now back to the signal conditioning. The offset factor (O) and gain factor (G) mentioned above are calculated in the following manner.

$$O = O_0 + \text{TCO} \cdot (T - T_0)$$

$$G = G_0 + \text{TCG} \cdot (T - T_0)$$

with

O_0 = offset factor at constant temperature

G_0 = gain factor at constant temperature

T_0 = constant temperature

T = actual temperature

TCO = temperature coefficient offset

TCG = temperature coefficient gain

Consequently not only sensor sensitivity and offset but also the linear temperature coefficients of sensitivity and offset can be trimmed. The trimming values are gain (13 bit), TCG (11 bit), offset (13 bit), and TCO (11 bit).

The temperature measurement is done by a first-order $\Delta\Sigma$ modulator produced in single-ended SC circuitry. It is fed by a temperature-proportional voltage. A bandgap voltage serves as the reference. The single-bit data stream from this converter is passed to a periodically ($f_{\text{temp}} = f_m/4$) resetted up-down counter, which may be seen as an accumulate-and-dump style first-order decimator. The counter is read out at variable time points after the reset time. The read-out times are proportional to the TC trimming values. So the read out values are the product of temperature and TC. After addition with the G_0 or O_0 data they are used as offset and gain factor in the correction stage.

The trimming values are held in an 8×8 bit oxide zapping PROM array. The array is programmed during the sensor end of line trimming procedure. It also holds 4 bit for trimming of the bandgap voltage reference and 2 bit for adjustment of the main clock oscillator. Last but not least, a special bit inhibits further programming for safety reasons. Each 8-bit data word in the PROM contains a parity bit. The data is repeatedly downloaded into a latch array, which holds the valid trimming information. If a parity error occurs, the analog output is set to about 0 V; outside the plausible operation range. The end-of-line programming is done via a separate programming interface pin.

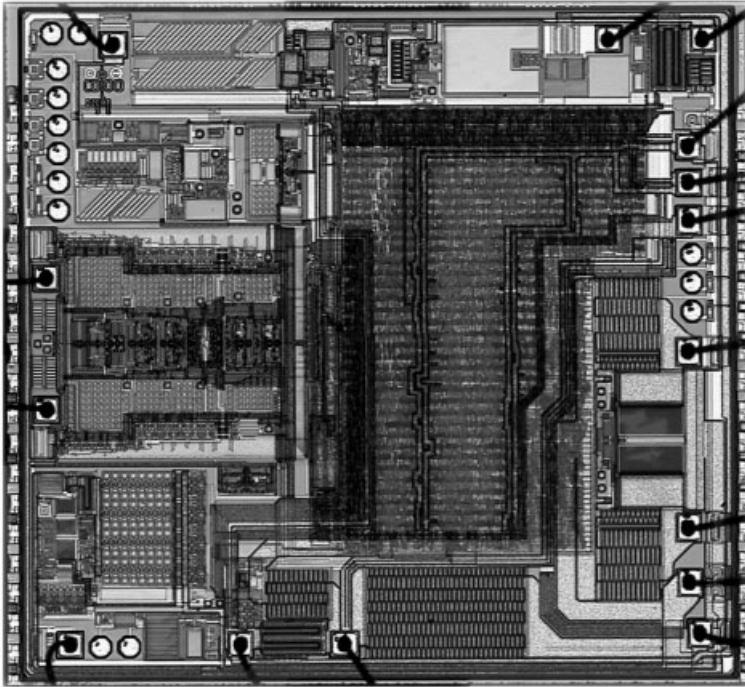


Fig. 6.2.14 Digital signal processing application-specific integrated circuit

Fig. 6.2.14 shows the chip photo of the realized signal processing ASIC for high-pressure sensors. The signal conditioning operates in the above described manner. The parameters of the sensor are as follows:

- Sensitivity bridge: 10 mV at high pressure (140 bar, 1800 bar).
- Voltage gain: 52 dB.
- Accuracy: 0.5%.
- Temperature range: -40 to 140°C .

6.2.5

Communication Interfaces

After signal conditioning, the information must be transferred to the system. Fig. 6.2.15 shows different possible solutions that are used today in the automotive world.

Dependent on the system architecture you can use a classical analog interface, which can be realized as a continuous voltage signal between certain levels (e.g., 0–5 V). Additionally there will often be a signal clipping incorporated and an out-of-range region defined to identify fatal failures of the sensor or the connections. Alternatively you can use a pulse-width signal, where the information is continuously encoded in the pulse width of a constant clock signal. This can easily be processed by standard microcontrollers.

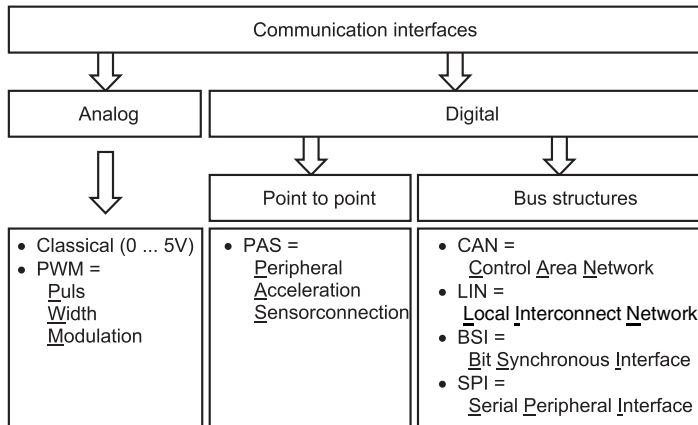


Fig. 6.2.15 Communication interfaces

If high safety requirements and/or high signal accuracy are needed digital interfaces will be more advantageous. The digital solution can be designed as a point-to-point connection or a bus structure. Today a current interface like the well-known automotive PAS (peripheral acceleration sensor connection) interface, as used for side airbag systems, is very popular for connecting the sensor and the electronic control unit. The advantage of this low-impedance connection is a high robustness against electromagnetic interference and the possibility of transferring signal and supply over the same twinpair cable.

For bus systems there are several standard or quasi-standard bus structures available. For high-speed connection you can use the popular automotive CAN bus (control area network) or the SPI interface (serial peripheral interface) often used to connect the sensor to the microcontroller inside the control unit. For lower speeds the LIN bus (local interconnect network) is available. In both cases a microcontroller inside the sensor unit is necessary, which works with a stable quartz clock frequency.

For low cost application the BSI bus (bit synchronous interface) can be used where the data rate is synchronized bitwise and not framewise. Therefore the clock can be regenerated out of the data stream without the need for a quartz-based clock.

7

Applications

7.1

Accelerometers for Automotive Applications

DAVID MONK, DRAGAN MLADENORIC and MARK SKAW

7.1.1

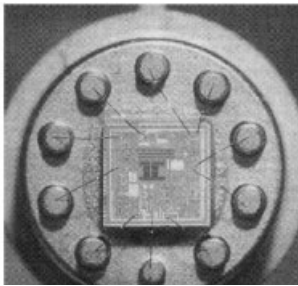
Crash-Detection Accelerometers: System Partitioning

The first crash-detection accelerometers appeared in Europe in the early 1980s. Mercedes started production of an airbag system with a seismic mass and a CuBe spring with thin-film pick up in the early 1980s. In 1984, a piezoelectric sensor was released for production. The production of automotive airbag systems increased significantly when the Intermodal Surface Transportation Efficiency Act became law in 1991 in the USA. This law required that all vehicles produced in the US would be equipped with an airbag by 1998. By 1998, airbag systems had achieved a 90% penetration rate into the US market compared with only a 5% penetration rate in 1993 [1]. Initially, the systems were enabled by electromechanical, 'ball-in-tube' type sensors produced by Breed and TRW (Thomson Ramo-Wooldridge) [2, 3]. With the early systems, three sensors were used typically. Two of the sensors would be placed in the crush zone near the front bumper. A third sensor was placed in the passenger compartment. This 'safing' sensor was used to prevent inadvertent deployment [2]. Microelectrical mechanical systems (MEMS) for airbags still have a safing sensor for this purpose.

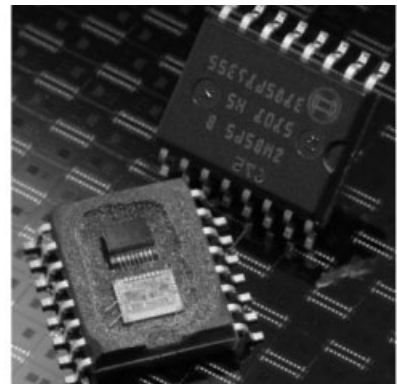
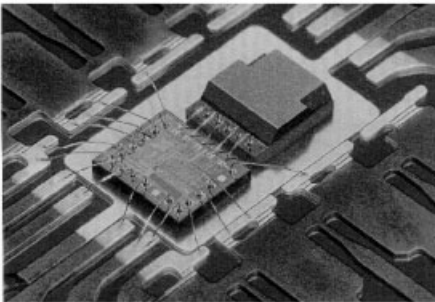
One significant advantage afforded by MEMS accelerometers is the possibility of single-point sensing. This gave a reduction in system cost by removing some of the wiring and harnesses. In the late 1980s, several companies, including Analog Devices, Motorola, Bosch, Denso, Ford Microelectronics, SensoNor, Temic, IC Sensors, and others, began work on micromachined accelerometers for the airbag deployment market. A well-documented example of a MEMS accelerometer is the Analog Devices ADXL50, which began as an exploratory market study [1], but became the first automotive production CMOS-integrated crash-detection accelerometer in the world. Technical details of the ADXL50 development can be found as a case study [4].

Motorola and Bosch, on the other hand, developed an accelerometer as a new market offering for already well-established automotive semiconductor businesses. In fact, automotive pressure sensors using bulk micromachining had been in production for a decade at Motorola when the surface micromachined accelerometer development started in 1991. Published descriptions of the first Z-axis device [5–7] and prototypes emerged in 1994, followed by full-scale, high-volume production in 1996. Since that time, Motorola has developed laterally-based sensing (both X- and XY-lateral accelerometers) for 20–250 g ranges. Moreover, a folded-beam Z-axis device has taken the place of the original ‘trampoline-style’ Z-axis accelerometer for the 5–250 g range. Finally, Motorola Semiconductor Products Sector has begun a joint development project with CEA-LETI on high aspect ratio SOI-based (silicon on insulator) MEMS. Bosch focused their accelerometer work on epi-polysilicon, deep RIE (reactive ion etch) based linear accelerometers. Like Motorola, these sensors are two chip: a sensing element and a control integrated circuit (IC).

These examples are highlighted because they represent two distinct engineering approaches to the accelerometer market. Fig. 7.1.1 shows this contrast in system partitioning, integration, and packaging technologies that were used by these two approaches in the automotive accelerometer market. Fig. 7.1.1a is the original ADXL50 device [8, 9]. This device is an integrated surface micromachined device



a)



b)

Fig. 7.1.1 Automotive crash-detection accelerometers: **a)** Analog Devices ADXL50 [8, 9] and **b)** Motorola's Z-axis accelerometer [5, 6,

21] on the left and the Bosch XY-lateral accelerometer (courtesy Dr. Jiri Marek) on the right

in a BiCMOS process. The capacitive transducer is highlighted in the picture. It is a lateral accelerometer device (i.e., parallel to the plane of the silicon surface), meaning that the plane of sensitivity to acceleration is parallel to the plane of the surface of the die. Because the die and micromachined device are exposed during packaging, a special sawing process is used to eliminate the possibility of the wafer sawing process causing damage to the mechanical device. This integrated approach offers a technology platform with lower parasitic capacitance and good signal-to-noise ratio.

Fig. 7.1.1b shows the Motorola two-chip accelerometer and the Bosch two-chip accelerometer. The accelerometer, or ‘g-cell’, is surface micromachined in a non-CMOS-integrated process. The original Motorola accelerometer is a Z-axis device, meaning that the plane of acceleration sensitivity is perpendicular to the plane of the chip. The Bosch accelerometer example in Fig. 7.1.1b is a lateral accelerometer, with acceleration sensitivity parallel to the plane of the chip [10, 11]. These were the first X- and XY-lateral accelerometers that came into production in 1997. In both cases, the g-cell is bonded at wafer-level, using glass frit bonding technology, to provide a sealed environment for the accelerometer during packaging and in operation. SensoNor, Ford Microelectronics, and Denso also followed the multi-chip approach to MEMS-based accelerometers. This ‘combinational technology’ provides the ability to mix and match MEMS with a variety of circuit technologies without lengthy integration process development time.

Crash detection for airbag deployment is presently the largest automotive opportunity for inertial sensors. The initial market was for front-impact crash detection, but side-impact crash detection has also become prevalent. Side impacts typically require sensors with a higher accelerometer range (100–250 g range versus the 20–100 g range for front impact). This application can typically be served with product modifications on the existing, predominantly surface micromachined-based accelerometer technology. However, there are several other opportunities in which micromachined inertial sensors are being used, including: vehicle dynamics systems, suspension and ride control, navigation, and roll over [12–15]. These require higher sensitivities (better signal-to-noise ratios for the sensor), so high aspect ratio surface micromachining or bulk micromachining are often used in these applications. This chapter surveys the technology required for automotive inertial sensor applications, including: transducer design concepts, MEMS fabrication, signal conditioning circuit design, package and test development, board mounting, reliability, typical specifications, and the future trend of networked satellite accelerometers.

7.1.2

Sensing Fundamentals

7.1.2.1 Transducer Design

Accelerometers can be described using the lumped-parameter mechanical model shown in Fig. 7.1.2.

Typically, an accelerometer (or g-cell) consists of a proof mass that is suspended with a spring, or compliant beam, in the presence of some damping, and anchored to a fixed reference. Coupling this mass-spring-dashpot model (Fig. 7.1.2)

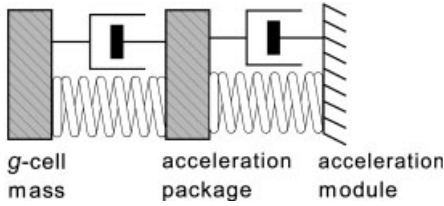


Fig. 7.1.2 A simplified accelerometer mechanical model

with Newton's second law of motion yields the mechanical transfer function for the accelerometer [16].

$$a(t) = x(t)t^2 + \frac{D}{M}x(t)t + \frac{K}{M}x(t) \quad (1)$$

where $a(t)$ is the acceleration, $x(t)$ is the displacement of the g-cell (proof) mass, M is the mass, K is the spring constant, and D is the damping coefficient.

This model produces a quality factor of

$$Q = \frac{\sqrt{KM}}{D} \quad (2)$$

a natural resonance of

$$\omega_r = \sqrt{\frac{K}{M}} \quad (3)$$

and a static sensitivity of

$$S = \frac{1}{\omega_r^2} \quad (4)$$

This simplistic model allows some of the properties of the system to be deduced.

- Resonant frequency: will increase with increasing spring constant and/or decreasing mass.
- Quality factor: will increase with increasing spring constant, mass, and/or decreased damping.
- Static sensitivity: will increase with decreasing resonant frequency.

Finally, mechanical noise is the result of Brownian motion of gases around the proof mass. This noise component can be described by

$$n_B = \sqrt{\frac{4K_B T \omega_r}{QM} \left[\frac{m \sqrt{\text{Hz}}}{\text{s}^2} \right]} \quad (5)$$

where K is Boltzmann's constant and T is temperature [K].

7.1.2.2 Self-Test

One requirement of crash-detection accelerometers is to provide a self-test feature. In one manifestation of this feature, a portion of the accelerometer is used to actuate the structure on demand. Fig. 7.1.12g shows an example of a lateral accelerometer from Motorola and the associated self-test actuation section of the design. In this case, the self-test actuates the structure to an equivalent 20 g (versus the 40 g range for the crash-detection application). Self-test is performed upon initial vehicle ignition turn-on to determine the health of the sensor before operation.

7.1.2.3 Overtravel Stops

Mechanical shock is one of the automotive reliability tests that an accelerometer must survive. Often, these excursions are specified beyond the maximum g-range of the device. Mechanically, this shock magnitude creates the possibility of substantial proof mass movement. To maintain integrity of the mechanical structure in this environment, many accelerometers use another structural material layer to provide a stop to the proof mass that inhibits the movement. An example of an overtravel stop on a lateral accelerometer is shown in Fig. 7.1.12g.

7.1.2.4 Stiction

Stiction has become one of the most pervasive quality issues with surface micromachined devices. Stiction can be described as the phenomenon in which microstructures become stuck to their substrates. Several authors have provided reviews of the stiction phenomenon and literature [17–21] and comparative studies of various stiction reduction techniques [22]. Typically, stiction occurs at one of two times (particularly for surface micromachined structures): post-sacrificial-layer-etch and during usage. The former is called ‘release stiction’ and the later is called ‘in-use stiction’. Packaging engineers are most often concerned with in-use stiction because of the potential reliability problems that it can cause; however, a phenomenon similar to release stiction can be a problem if devices are exposed to liquids during assembly: particularly during the wafer sawing operation. Therefore, both types of stiction and techniques for improving yield and reliability are described here. Analog Devices uses a controlled amount of moisture and an organic coating that is deposited before sealing the package to ‘seal’ the microstructure from moisture during operation [23].

Lastly, it should be noted that there have been techniques published to remove the effect of stiction after the event. Gogoi and Mastrangelo have published a technique for using Lorentz forces to lift the microstructure that is stuck [24].

Design Techniques

Release stiction has been recognized as a problem since the late 1980s, especially for surface micromachined structures. During drying, surface tension from the liquid-vapor interface causes a downward force on the structural layer. If the layer touches the substrate, it is prone to stick onto the surface. It is hypothesized that etch products and/or contaminants in the rinse water can then precipitate out of solution during drying and cause a bond that is stronger (e.g., a chemical bond)

than the electrostatic bonding between the two semiconductors [25, 26]. Mastrangelo and Hsu were among the first to provide an analytical model of this effect [27–29]. Others have followed their lead [30, 31]. At the University of California, Berkeley, dimples [32–34] or meniscus-shaped microstructures [35] were designed into the thin films constituting the structural layers of the microstructure to minimize the surface area for stiction. Fedder and Howe used breakable tethers to hold the device in place temporarily during sacrificial layer etching and drying. These tethers could be broken or melted after release, and release stiction was never encountered [36]. Alternatively, roughening the surface of the structural layer (often polysilicon) was attempted through modifications of the sacrificial layer etch process [37, 38].

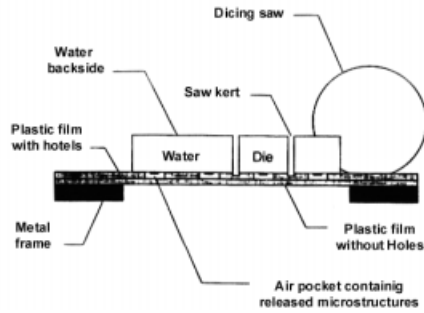
Rinse/Dry Processes

However, the repeatability of these types of solution is suspect; therefore, the general consensus is that surface tension forces must be overcome (i.e., avoided or eliminated) to minimize release stiction. Additional effort that has been placed on modifications to the etch process include the use of gas phase HF totally to eliminate liquid from the sacrificial etch process [39–41] and the use of high-temperature sacrificial layer etch processes. The simplest approach is to use a rinse solution that exhibits a lower surface tension force during drying. Several groups have used IPA (isopropyl alcohol) drying or other low surface tension solutions, like *n*-hexane [42]. But it is not feasible to eliminate this problem with this approach because the surface tension force induced by even low surface tension solutions, like alcohols, is still high enough to cause release stiction [43]. Therefore, a group at the University of Wisconsin has experimented with freeze-drying and sublimation to eliminate the surface tension forces [26, 44–49]. Mastrangelo et al. have presented a process for using polymer (parlylene) studs [50, 51]. This is similar to the process that has been developed for the production of accelerometers using photoresist [52], and a slightly different process that uses divinylbenzene [53]. Finally, and probably most successful on the research-scale, is the use of supercritical carbon dioxide drying [54, 55]. Of course, all of this assumes that the structural materials exhibit tensile stress so that they do not inherently bend down onto the substrate [56].

Special Packaging Techniques

Micromachined devices are more difficult to saw because of induced stress that affects the device performance, delicate microstructures that can be broken during the sawing operation, and/or stiction resulting from the water used to cool the wafer saw during operation. For example, in the case of the Motorola production accelerometer, glass frit wafer bonding is used to seal the microstructure hermetically within the wafer-level package before sawing (Fig. 7.1.13b) [57, 58]. This eliminates the potential problems listed above. In the case of the Analog Devices production accelerometer, a special upside-down sawing process is used (Fig. 7.1.3). In this case, the wafer is placed on a special sticky-tape that includes a

Fig. 7.1.3 Upside-down wafer sawing process used to protect the Analog Devices accelerometer during die separation [89]



recess to accommodate the microstructure. By flipping the wafer over, the microstructure can be isolated from the water and from any flying debris during the sawing process. In addition, because of the recess in the sticky-tape, the microstructure is protected from damage. Finally, a specially designed pick and place tool is then used to remove the dice and turn them right-side-up for subsequent die attachment.

In the case of in-use stiction, it is hypothesized that moisture from the environment (relative humidity) comes in contact with the MEMS structural surfaces. If, during operation, these structures come in contact, the moisture can cause a temporary bond that, like release stiction, can then become permanent with time. To reduce in-use stiction, three basic techniques have been attempted. The first is to use a hermetic seal around the microstructure to eliminate the possibility of moisture encountering the structure. Secondly, the use of techniques to minimize the work of adhesion has been employed. Specifically, Houston et al. have used ammonium fluoride to reduce the work of adhesion on surface micromachined structures [59, 60]. Lastly, various coatings and/or surface treatments have been used on the microstructure to eliminate the chance of contact between two surfaces that have the prevalence to stick (e.g., polysilicon and silicon, each material with a native oxide). The University of California, Berkeley has pioneered techniques of using self-assembled layer monolayer coatings to minimize in-use stiction [18, 25, 59, 61]. Also, other researchers have used fluorocarbon coatings to minimize the in-use stiction [62–64].

7.1.3

Signal Conditioning

Fig. 7.1.1 shows an example of variations in system partitioning and integration on very similar products. The Analog Devices single-chip accelerometer is in Fig. 7.1.1a and the Motorola or Bosch two-chip accelerometer is in Fig. 7.1.1b. Each has particular advantages. For example, the single chip solution may improve the signal-to-noise ratio and reduce the number of interconnects. Whereas the multiple chip system may have a better time-to-market as a result of the flexibility afforded by the separation of the MEMS processing from the circuit fabrica-

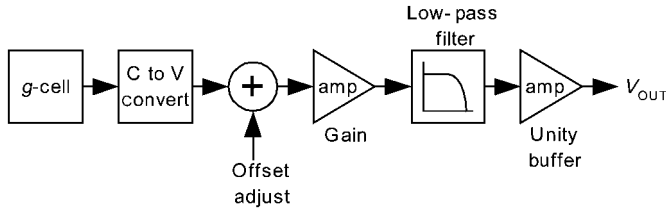


Fig. 7.1.4 Ideal accelerometer architecture

tion. Both of these perform very similar functions: they provide an analog 0–5 V output based on acceleration input, yet monolithic integration is not required to achieve this function in a single package.

Most silicon accelerometers are based on a micromachined variable capacitance element (g-cell) that is converted to a voltage using a C–V converter and then amplified, filtered, and buffered to provide an analog output as shown in Fig. 7.1.4. To date, open-loop implementations for capacitive read-out circuits are more widely employed than closed-loop systems, primarily as a result of the stability of such systems [16]. Interface electronics for micromachined sensors depend not only upon the transduction technique (input specification) and the product requirements (output specification) but also on the packaging approach, as parasitics are introduced when a multiple-die packaging technique is used.

7.1.3.1 Capacitance to Voltage Conversion

Fig. 7.1.5 shows an example of interface electronics for capacitive devices. Capacitive transduction has several advantages. Capacitance is very sensitive as a transduction method with good dc response, noise performance, and low power dissipation, and it is relatively temperature insensitive. It also is compatible with surface micromachining technology. Sensor interface circuits with capacitance transducers require careful design, especially to provide shielding from unwanted electrostatic forces, electromagnetic interference (EMI), and to minimize the impact of parasitic capacitances [65]. Fig. 7.1.5 a is an implementation of a capacitance position measurement circuit for inertial sensors. As is mentioned above, parasitic capacitance, and the resulting noise, can be complicating factors for the design of C–V converters. Fig. 7.1.5 b represents a differential capacitance sense interface that is meant to overcome some of these parasitic capacitance and noise concerns [66].

7.1.3.2 Filtering

The history of airbag development began with mechanical ‘ball-in-tube’ or ‘roll’ sensors that made an electrical contact when sufficient energy over a sufficient time was encountered. Because these sensors were essentially digital ‘on-off’ devices, more complex detection algorithms were difficult to implement; and they were not fast enough to be used for the evolving side-impact detection systems. The development of small silicon micromachined accelerometers for less than \$5

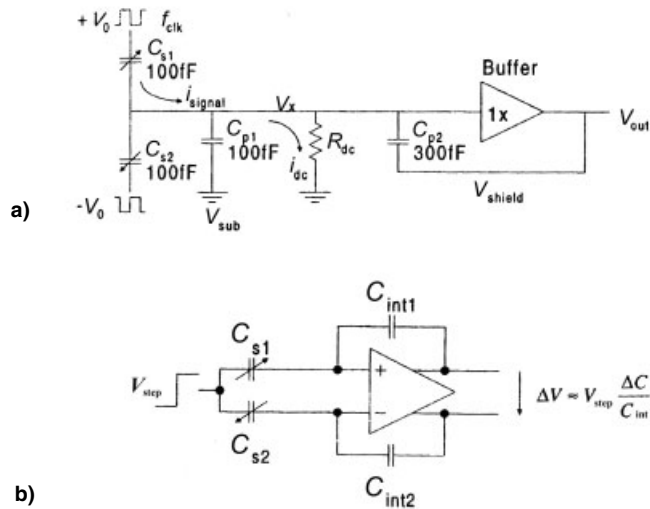


Fig. 7.1.5 Capacitive interface electronics for MEMS: **a)** inertial sensors [65], and **b)** a differential position sense interface [65]

shifted airbag development towards a single centrally-mounted sensor in which the actual acceleration of the vehicle (referenced to the vehicle frame) was integrated to determine a velocity-against-time profile in order to define the trigger point at which the airbag must be deployed. The time for such a decision is usually less than 10–50 ms depending on the speed at impact and the type of object impacted. So accurate determination of the trigger threshold is important.

Since the detection methodology was based on the deceleration of a large mass, early in the development it was typically defined that the frequency bandwidth of interest should be less than 400 Hz. This low bandwidth would reduce the need to perform extremely fast A/D (analogue to digital) conversions and mathematical calculations that would be needed to follow the higher frequency components of a crash signal.

As airbags gained more field experience it became apparent that the force and timing of deployments needed to be controlled for a wider range of crash types in order to avoid possible injury to passengers. This requirement increased the need to monitor the accelerometer signal for longer periods of time than previously required.

In developing these new algorithms, the requirements for sensors added more of them outside of the central control module for two reasons. First, some types of crash were difficult to detect within the short time frame required for deployment (such as offset pole crashes) and the solution was to provide a high-g ‘pre-crash’ sensor within the front of the vehicle. And second, the need for side-impact detection required deployment decisions even faster than the frontal crashes (within approximately 5 ms). Both of these requirements moved accelerometers to ‘satellite’ positions closer to the perimeter of the vehicle where they are more likely to pick up impacts from small objects.

Sources of Distortion

With the advent of acceleration detection over longer timeframes or satellite sensors near the perimeter of the vehicle, problems began to surface with signal distortion from the accelerometers. These occurred due to several specific sources.

The first was from the 'safing' sensor that was used as a redundant firing feature, and was usually placed within the central module. The most common safing sensors are small ball or roller mechanical devices that make an electrical contact when they experience decelerations around 5 g. Because of their mechanical mass these safing sensors can create a shock pulse up to 1000 g at very high frequencies. Sometimes the shock occurs when the ball or roller rebounds after its initial closure. Some of this shock can be transferred directly to an accelerometer mounted within the same module. In early systems this shock would usually occur after the deployment decision had been made, so whether it corrupted the accelerometer signal was not a concern. But this is not the case for newer systems with staged airbags and longer times for the deployment decision.

The second was the impact of rocks near sensors that may be placed on the floor structure for side-impact detection. In these locations rocks or other debris hitting the bottom of the vehicle can cause high-g shocks greater than 300 g at frequencies above the 400 Hz bandwidth of the accelerometer. The last category involves various small objects such as birds, rocks, tree limbs, balls, or other road debris that can impact the side of the vehicle and thus affect sensors mounted in the 'B' pillar or within the skin of the doors. These can also create high-g forces with high-frequency content.

When subjected to such high-g, high-frequency shocks the output of the accelerometer would be driven to full scale (or saturated) if the 400 Hz bandpass filter was not present. In theory the 400 Hz bandpass filter would suppress these signals so that they would contribute little to the output signal. But, as will be discussed, such ideal accelerometers are not easily produced.

Practical Accelerometer Analog Interface Circuit Design

In the ideal architecture, shown in Fig. 7.1.4, all of the signal amplification occurs before the bandpass filter so that any noise, which is also amplified, can be suppressed along with any high-frequency acceleration signals. Sources of noise are the Brownian noise within the g-cell due to molecules hitting the sensing plates, noise within the signal path, and noise within the circuit's transistors themselves. Because the noise is passed through the filter, the signal-to-noise ratio of the output is improved. This low-pass filter is usually a 2- or 4-pole Bessel function, which reduces the output by 40 or 80 dB/decade with a unity gain for the filter. The output stage would also be a unity gain buffer solely for driving an external load. The overall gain in the system is chosen to give an output calibration such that the most commonly expected signals will use the complete dynamic range of the output. If a wider range were provided then the common level of signals of interest would be too small to be digitized with sufficient resolution to perform an accurate integration to calculate the velocity. Distortion of the output can occur from five design related issues: gain distribution, input stage saturation, g-cell saturation, g-cell resonance, and g-cell damping.

In the ideal architecture all forces applied to the g-cell will be amplified the same regardless of their frequency. Therefore the gain would be distributed so that the input to the bandpass filter is the same as the output of the sensor. What occurs when the g-cell is subjected to a force greater than the output will follow (i.e., a 400 g shock with a 40 g output range) and the input is at a frequency less than the bandpass filter (400 Hz) is considered an ‘in-band overload.’ The representative waveform is shown in Fig. 7.1.6.

As shown in Fig. 7.1.6 the output signal hits the upper limit and then recovers when the signal falls back within the range. The same signal saturation is also occurring at the input to the low-pass filter. In the application this would be the expected result since the pulse is within the frequency band of interest. If such pulses are occurring in the application, then the accelerometer’s output range calibration would be changed to analyze fully such higher level signals.

Now consider the same situation in which the high-g pulse occurs at a frequency greater than the pass band (i.e., 4000 Hz with a 400 Hz pass band). This would be an ‘out of band overload’ as represented in Fig. 7.1.7. It would be expected that the pulse would be attenuated 10000:1 and yield less than a 0.4 g output; but instead there may be a noticeable shift in the zero level. The signal at the

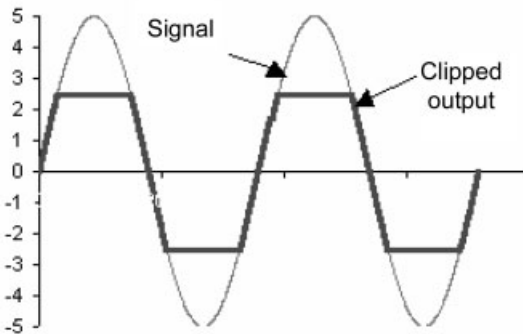


Fig. 7.1.6 Saturation of output within pass band

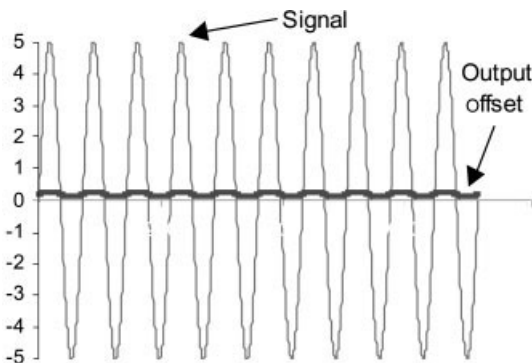


Fig. 7.1.7 Saturation of output above pass band

input of the low-pass filter has saturated at the 40 g level and stayed flat until the pulse energy falls below 40 g. The Fourier series for the input source would have a single frequency term at 4000 Hz; but the distorted waveform has added components at odd harmonics above 4000 Hz and more importantly it now has a dc or steady state component, if the signal is not centered perfectly at the input of the low-pass filter. The low-pass filter attenuates the fundamental and its entire harmonics, but cannot suppress the new dc component, which now shows up in the output.

Such distortion can be a serious consideration in crash-detection algorithms as any unwanted shocks will be integrated into the crash pulse. Such distortions can either add to the velocity integration causing a deployment when not desired; or subtracting from the velocity integration causing non-deployment when required. This is shown graphically in the examples in Fig. 7.1.8.

The key criterion is that the input to the low-pass filter must not be distorted or the signal that is passed to the output will be a false representation of the original signal. Therefore, the ideal architecture must be revised to move some of the signal gain after the low-pass filter, as shown in Fig. 7.1.9.

Making this trade-off in gain distribution has the primary effect of decreasing the signal-to-noise ratio in the output. So the amount of gain moved after the filter must be carefully chosen to handle only the expected high-g signals. For the example shown in Fig. 7.1.9 a gain of 7.1.10 would be placed after the filter and a

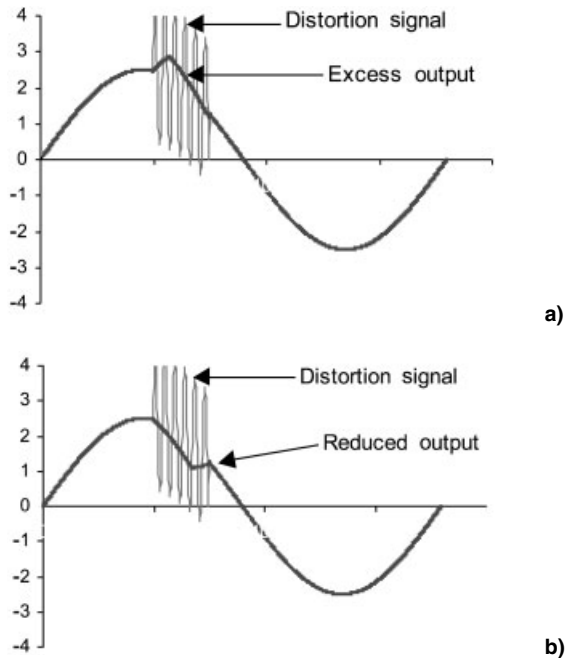


Fig. 7.1.8 Crash signal distortion causing non-deployment

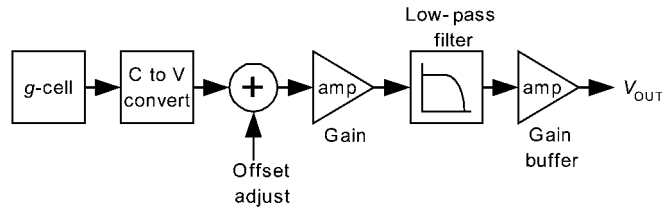


Fig. 7.1.9 Practical accelerometer architecture

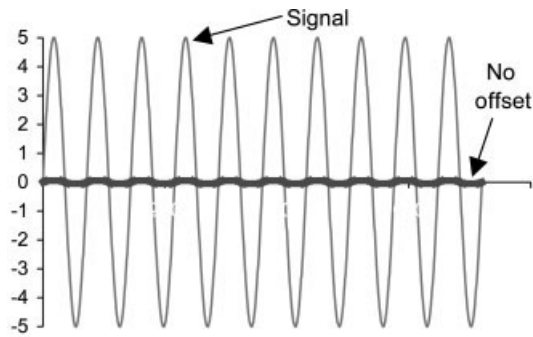


Fig. 7.1.10 Saturation prevented above the pass band

representation of the output is shown in Fig. 7.1.10. As can be seen this redistribution of gain solves the issue, but requires knowledge of the worst-case signals to be applied and an acceptance of higher noise in the output signal. The input stage converts the variable capacitance of the g-cell to an analog voltage.

There is also a 'gain' factor in the conversion from capacitance to voltage that is independent of the electrical gain in the amplifier. This conversion is usually a switched capacitor design that has the signal captured during one or more phases of a clock. The architectures vary here, but all have similar outputs. The components affecting the gain include the size of the capacitor used in the C-V converter versus the capacitance of the g-cell, the applied voltages within the C-V conversion, and the sensitivity of the g-cell (percentage change in capacitance per unit of acceleration). If these parameters are not matched for a given g-cell design then the C-V converter's output could saturate even though the amplified signal is within the input range of the low-pass filter. It is therefore possible to saturate the C-V converter at 300 g and the input to the filter at 400 g, but the net effect is that the design starts to distort the output at 300 g (the weakest link).

Saturation can also occur mechanically within the g-cell if the moving element strikes a mechanical overtravel stop and can no longer move. While this is mechanically possible with all silicon micromachined g-cells, it is usually the least likely source of saturation. In order to provide good linearity the range of movement of the g-cell sensing elements is usually less than 2% of the total range and mechanical stops may be used to limit the movement to less than 2/3 of the total

range. So any mechanical saturation of the g-cell will occur at least 30 times higher than the range for which the accelerometer was designed.

Effects of Board Mounting and Safing Sensor on Accelerometer Performance

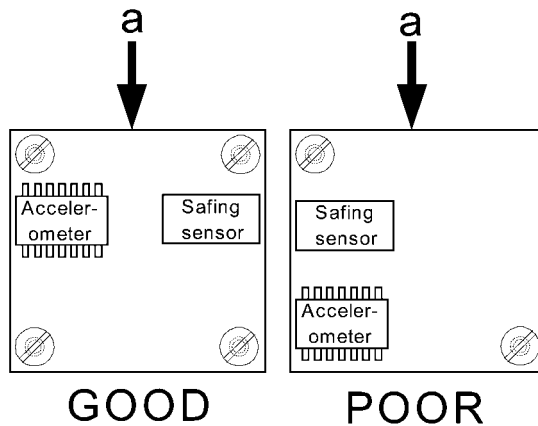
The preceding discussion has centered about the electrical gain in the accelerometer. But the applied stimulus is mechanical and, therefore, any mechanical gain in the system must be considered. The accelerometer looks like a combined spring mass system with the g-cell itself forming one spring mass system and the package forming a second spring mass system as shown in Fig. 7.1.2 and equation 1 (this model can also be extended to the sensor housing and to the car itself). The resonant frequency of the g-cell is based on its mass and spring constant and is quite high due to the small mass (equation 3). Typical resonant frequencies are 7–35 kHz. At resonance there is some mechanical gain in the absence of any damping. Therefore, if the applied shock pulse happens to have frequency components at this resonant frequency those components will be amplified and the resulting signal given to the C–V converter will be higher than the actual applied level.

Similarly, the package can have a resonant frequency due to the mounting leads and mass of the package body. This resonant frequency is lower at about 6–8 kHz; and its damping will depend on the mounting methods. It is critical that the resonant frequencies of either the g-cell or the package not match any expected applied pulses.

The g-cell's spring mass system can also have damping that results from the moving plates squeezing any gas encapsulated within the g-cell as the plates move closer. Due to the extremely small dimensions the amount of damping can be limited. In a vertical or 'Z' axis g-cell the plate area is much larger compared with the gap so these structures are typically over-damped and show no mechanical gain at their resonant frequency. A photograph of a typical Z-axis g-cell is shown in Fig. 7.10.12 f. In lateral or 'X' axis g-cell designs the plate area is distributed among a number of small fingers that do not have much area versus their gap, so these structures are typically under-damped and may have gain at their resonant frequency. A photograph of a typical X-axis g-cell is shown in Fig. 7.1.12g. Both g-cell designs can have their damping increased by changing the type of encapsulated gas or increasing its pressure. But these effects are more pronounced on the Z-axis designs; and within normal manufacturing limits the X-axis design can not be easily made over-damped without very high aspect ratio etching techniques (like those shown in Fig. 7.1.12c and i). Therefore, the resonant frequency of an X-axis design is much more likely to enhance distortions than a Z-axis design.

Mounting an accelerometer onto a printed wiring board (PWB) requires attention to both additional sources of resonance and transmission of shocks from other components. The PWB can become resonant either as a total unit or have resonance about the mounting points. Therefore, it is important that the accelerometer be mounted as close as possible to a rigid mounting point as shown in Fig. 7.1.11. Resonance can also be reduced by using a conformal coating covering

Fig. 7.1.11 Accelerometer and safing sensor mounting axis recommendations



either all of the PWB or just the accelerometer packaging itself. For central modules with safing sensors it is also important not to align the axis of the safing sensor and the accelerometer as shown in Fig. 7.1.11. If the safing sensor is in line with the accelerometer along the crash axis the shocks generated by the safing sensor will be fed directly into the accelerometer. On the other hand, if the safing sensor is mounted parallel to the crash axis, then the shocks from the safing sensor will radiate more towards the less sensitive cross axis of the accelerometer. It is also helpful if the safing sensor is mounted next to a mounting point.

Finally, other considerations for the sensing module are to avoid hardened metal cases or brackets as these do not have inherent damping and will have higher resonant peaks than soft metals.

7.1.3.3 Specifications

Accelerometers generally fall into two application categories of impact or motion measurements. The specifications for each type of application vary with regard to sensitivity, zero-g offset, frequency range and noise. In all cases there is a desire to have the smallest possible package size and lowest power consumption.

Sensitivity and Accuracy

The most significant requirement is the sensitivity of the device to acceleration. The low-range full scale outputs of ± 1.5 , ± 5 , and ± 10 g are common for rollover detection, tilt detection, anti-lock braking and safing for medium range accelerometers. The medium range full scale outputs of ± 40 , ± 50 , ± 70 , and ± 100 g are common for central crash sensing and safing of high-range accelerometers. The high-range full-scale output of ± 250 g is common for impact sensing on the side and front extremities of the vehicle [66]. In all cases the desired accuracy of the output sensitivity in mV/g is better than 5% for limited temperature and voltage ranges and 7% over the extremes found in automotive applications. Tilt and rollover detection can benefit from higher accuracy. Linearity to better than 1% is common with little or no measurable hysteresis.

Offset

All accelerometers are bi-directional and, therefore, possess a zero-g output level at middle scale, which is commonly called the 'offset' of the device. Accuracy in the offset is most desired for the low-g applications of tilt and rollover detection as any offset drift can be convolved with angular change. In the medium and high-g impact applications the offset is less critical or may even be ignored by AC coupling the output. Offset variations are usually an artifact of stresses induced within the package over temperature and during the assembly process. Offset errors less than 5% of the full scale range are achievable.

Bandwidth

The application of accelerators to date is usually associated with monitoring a large mass device. Therefore, the required bandwidth of the signal does not have to be much greater than 400 Hz for impact applications or greater than 50 Hz for other applications. The inherent noise sources within the sensor are reduced by a filter supplied in the accelerometer signal conditioning with either a 2-pole or 4-pole roll off rate. As pointed out above (Section 7.3.2), the presence of such filters requires attention to possible distortion issues when the accelerometer is subjected to high frequency, high-level impacts. It is important that the bandwidth of the impact application be well behaved as the filter roll off will affect any analysis of impact pulses. Filter roll off frequency (-3 dB point) accuracy is usually within 10%.

Noise

Noise within an accelerometer is created from several sources. The noise comes from the switched capacitor design, inherent thermal noise within any devices, the 'flicker' noise of transistors, and the Brownian noise in the g-cell transducer due to random motion of atoms (see equation 5). Signal-to-noise ratio of 60 dB is common, but lower g devices with higher gains may be worse.

Temperature and Voltage

Micromachined accelerometer technologies and their associated ASICs are generally capable of operating at temperatures from -40 °C to +125 °C and voltage ranges from 2.7-6.0 VDC. But there is a trade-off in accuracy over the range of voltage and temperature. Generally, the tightest accuracy is specified for -40 to +85 °C and 4.75-5.25 V dc.

Mechanical

In addition to the maximum acceleration that might cause the signal conditioning to saturate, as described in Section 7.3.2, there is a maximum acceleration at which the g-cell will impact its stops and/or be damaged by severe shocks. In order to achieve good linearity the g-cell will normally only use 2-3% of its maximum range of travel for its calibrated range. Mechanical overtravel stops are pro-

vided to prevent full travel in which the capacitance plates might touch in order to prevent stiction. The stops are typically at about $2/3$ of the maximum travel. Therefore the device can be driven at 20–30 times its calibrated range before hitting the stops. Merely touching the stops will not damage the g-cell as silicon is an elastic material that does not yield. On the other hand, if the shock to the stop is too high, the g-cell could break. Typical shock limits are up to 2000 g for less than 0.5 s. Drop shocks onto a concrete surface of 1.2 m are also permissible.

The other mechanical specifications of interest are the resonant frequency and damping of the g-cell and the resonant frequency of the package when properly mounted on a printed wiring board. The lateral g-cell designs may have resonant frequencies in the range 15–40 kHz depending on the design range and are usually underdamped with up to 15 dB of gain at resonance. The vertical g-cell designs may have resonant frequencies of 10–30 kHz and are usually overdamped with up to –15 dB of attenuation at their resonant frequency.

7.1.4

Fabrication

Automotive MEMS fabrication technology is not yet as standardized as circuit technology. Among the various technologies are bulk micromachining, surface micromachining, bonded wafer, SOI, epi-polysilicon, and combinations of these. Examples of each of these technologies are shown in Fig. 7.1.12. Of course, this brief section cannot possibly do justice to a topic that has been the genesis of several conferences (Transducers, Hilton Head, MEMS, ECS-Sensors, SPIE Microfabrication, ASME-MEMS, etc.). Rather, it is meant to show the variety of micromachining technologies that are the basis for actual automotive products today.

7.1.4.1 Bulk Micromachined/Piezoresistive or Capacitive

Bulk micromachined processes use the silicon wafer itself as the structural material. Typically, an alkaline solution (usually KOH or TMAH) is used to etch into the silicon substrate to create a 3D structure: such as a spring-proof mass [16, 67]. This approach has been used for pressure sensors in the automotive industry since the late 1970s/early 1980s. For inertial sensors, bulk micromachining is usually employed when high sensitivity (very low-g range/high resolution) is required. An example of a capacitive version of this technology for accelerometers is the VTI low-g sensor (Fig. 7.1.12b) [12].

7.1.4.2 Surface Micromachined/Capacitive

Surface micromachining does not usually require two-sided processing of the silicon, rendering it often more CMOS-compatible. A structural material is patterned over the top of a sacrificial material. Subsequently, the sacrificial material is etched, leaving the anchored structural material free to move. Capacitance is used almost exclusively as the transduction technique with surface micromachined devices. An example of a three-layer polysilicon surface micromachined accelerometer is shown in Fig. 7.1.12f.

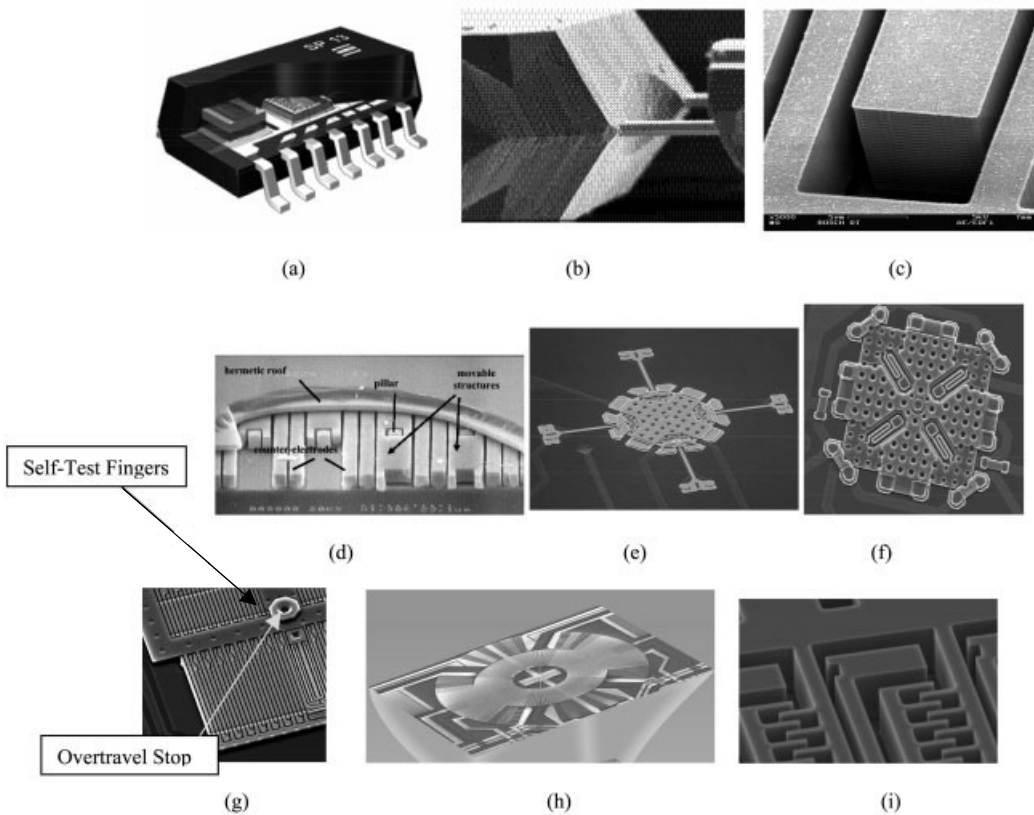


Fig. 7.1.12 Examples of various automotive-qualified MEMS inertial sensor technologies: bulk micromachining (a–c), **a**) SensoNor Tire Pressure Monitor Sensor that includes an inertial roll switch [90], **b**) VTI low-g accelerometer [12]; surface micromachining (c–g), **c**) Bosch thick polysilicon lateral accelerometer [91], **d**) Infineon surface-micromachined pressure sensor [92], **e**) initial ‘trampoline-style’

Motorola Z-axis accelerometer [21], **f**) Motorola ‘Folded-beam’ Z-axis accelerometer, **g**) Motorola X-lateral accelerometer (an Analog Devices surface micromachined lateral accelerometer is shown in Fig. 7.1.1 a); **h**) epi-polysilicon surface micromachined lateral accelerometer [93]; **i**) silicon-on-insulator surface micromachining [94]; **j**) bonded wafer [69, 95]; **k**) nickel electroplated [96]

7.1.4.3 Other Transducers and Associated Processing

A new transduction technique that is being employed for inertial sensing is based on convective heat transfer. Temperature sensors around a central heater track convection as a function of acceleration input [68]. While this is not yet qualified for automotive production, it is a candidate for future devices.

7.1.4.4 Wafer Bonding

Wafer bonding or cavity sealing is present on many MEMS products. Wafer-level packaging was first used in pressure sensors (Fig. 7.1.13 a) to create an absolute

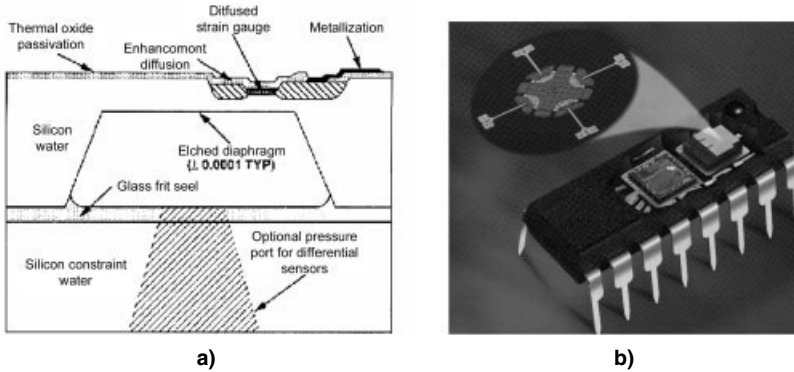


Fig. 7.1.13 a) Wafer bonding to create an absolute vacuum cavity for an absolute pressure sensor device [21] was adapted to several automotive accelerometers, including the Motorola version (b or Fig. 7.1.1b)

vacuum reference for absolute pressure sensors but has also been expanded to accelerometers to create a controlled atmosphere for the device (Fig. 7.1.13b) and to protect the device from the assembly environment. Complete reviews of wafer-to-wafer bonding for MEMS are given by Schmidt [69] and Ko et al. [70].

Fig. 7.1.13 shows examples of an absolute pressure sensor and an accelerometer that employ wafer bonding through the use of an adhesive layer bonding, using low-temperature (450–500 °C) glass frit. In general, wafer bonding can be categorized based on the following processes: direct wafer bonding, anodic bonding, and intermediate layer bonding. A detailed review of these techniques can be found elsewhere [23]. In general, accelerometer devices that use wafer bonding have been used in production with an intermediate bonding layer: glass frit in the case of the Motorola and Bosch accelerometers, and a reflowed glass in the case of the VTI low-g accelerometer. SensoNor has used anodic bonding for their production accelerometer.

Motorola has been producing pressure sensors since 1979 (Fig. 7.1.13a), and accelerometers since 1996 (Fig. 7.1.13b) using a glass frit intermediate bonding material to form wafer bonds [71]. In this process, glass paste is screen printed onto a ‘constraint’ substrate. Glass pastes are powdered glass with a solvent and a binder. These are usually proprietary compositions [72]. This material is then glazed at 400 °C before placing the constraint wafer in contact with the ‘active’ wafer. Once these wafers are joined, they are fired at approximately 550 °C to form the bond. The advantage of glass frit bonding over direct wafer bonding and anodic bonding is that it can be performed at low temperature without any large electrical fields. Therefore, with non-CMOS-integrated or bipolar-integrated devices, the glass frit bonding process can be performed at the end of the fabrication and just before wafer sawing and assembly begins. This allows flexibility in the fabrication process and protects the device from damage during the assembly process. Also, glass frit bonding processing can occur at different pressures to create absolute pressure sensors (with vacuum reference cavities) and/or specified damping-level accelerometers.

Finally, several groups have attempted monolithic sealing of microstructures [73]. This was first attempted with the sealing of a light source using surface/bulk micromachining [74, 75]. After this effort, McNair et al. used a thick phosphosilicate glass mesa above a microstructure sealed with low-stress silicon nitride or polysilicon using surface micromachining processing [76]. Leboutitz et al. [77] are using permeable polysilicon [43, 78, 79] as a shell for later sealing. In this process, thin (1500 Å or less) polysilicon is permeable to HF. The sacrificial layer etching, which does not require etch access holes, etches the underlying glass (particularly phosphorus-doped glass), while leaving the thin polysilicon intact. The permeable polysilicon can, then, be sealed with a subsequent low-pressure chemical vapor deposition LPCVD. Aigner et al. have also presented this as an alternative bonding technique [80].

7.1.5

Backend Processing

Packaging, assembly, trim/calibration, and test of MEMS/MST devices represents a considerable portion of the cost of MEMS/MST products. Swift and Schuster [81, 82] both showed that the package and test can represent over 50% of the product cost for a microsystem. Discussion during each of the last two ‘Commercialization of Microsystems’ conferences has confirmed this point. In fact, often industrial MEMS engineers and managers suggest that a typical cost breakdown is roughly 33% silicon content, 33% package, and 33% test. It has been reported, in some cases, as greater than 70% packaging cost [3, 83–87]. Furthermore, of this silicon content, often less than half of it is the microsensor or actuators. The rest of the silicon content is the control or interface electronics cost.

7.1.5.1 Assembly Process

A complete overview of the backend assembly and test technologies can be found elsewhere [23]. This section represents a summary of that review. Packaging of MEMS/MST devices poses unique challenges when compared with more conventional electronic packaging because physical signals induced by the package can be misinterpreted by the MEM device as signals from the environment [57]. This was observed very early in the development of sensors and actuators. Senturia and Smith noted in 1988, ‘it is necessary to design the microsensor and the package *at the same time*’ [88]. They made two very important observations in this work: ‘packaging people’ and ‘sensor people’ are usually not the same people, and they do not always work well together; and consideration of the package can result in elimination of possible sensor designs because they would not be feasible in a particular package.

7.1.5.2 Package Design

Fig. 7.1.14 shows just a glimpse of the myriad of sensor packages. Package design standardization has been a problem with MEMS device commercialization. For accelerometers, standards are evolving more quickly because the devices can be de-

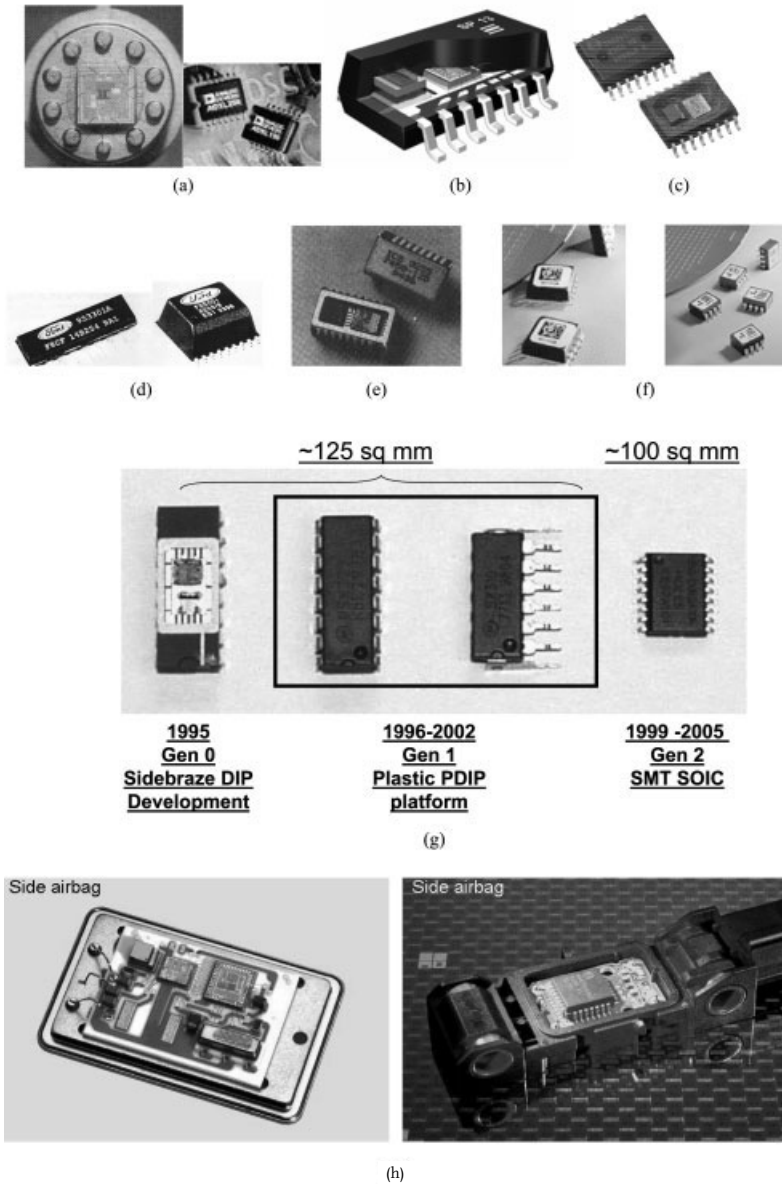


Fig. 7.1.14 Examples of automotive inertial sensor packages: a) ADI's accelerometer in a header-style package and a SOIC-style package [8, 89, 97]; b) an overmolded combination pressure and inertial sensor for Tire Pressure applications [90]; c) the Bosch 2-chip PLCC 28-pin automotive accelerometer; d) the

Ford Microelectronics accelerometer in a plastic package [98]; e) the EG&G IC Sensor accelerometer ceramic package [23]; and f) the VTI; g) the progression of automotive accelerometer packages produced by Motorola; and h) inertial module-level packages from Bosch (courtesy Dr. Jiri Marek)

signed into standard CMOS packages. Lack of standards increases cost, provides the need for unique assembly techniques, process tools, and test equipment, and increases time-to-market.

7.1.5.3 Semi-Custom Trim/Test Systems

Very little has been published about the testing procedures for micromachined devices. In addition to the standard analog and digital test equipment, a physical interface is also required (e.g., pressure for pressure sensors, shaker tables for accelerometers). These physical interface modules cause the test systems to be custom, which invariably increases the test cost because of high depreciation per unit. A typical process flow includes four major types of testing: ‘class probe’ for monitoring process control parameters in the front-end, ‘unit probe’ for monitoring circuit functionality at the wafer-level, ‘trim’ for calibrating the sensors after packaging, and ‘final test’ for monitoring functionality of the device after completion of the assembly process. Most MEMS test processes only include the full physical stimuli after packaging: at trim and final test. As a result, these process steps are critical to insure proper accuracy and functionality of the outgoing product.

7.1.5.4 Reliability

Reliability of automotive devices is a prerequisite, much like quality. Underhood requirements are stringent. Tab. 7.1.1 lists some typical requirements. Significant detail can be found by reviewing the AEC Q-100 specification.

7.1.6

Future Trends: Networked Satellite Sensors and Side-Impact Sensors

Government safety regulations have been the compelling market driver for crash-detection accelerometers. Initially, mechanical sensors were used at several locations within the car to obtain the crash signal. The MEMS accelerometer enabled

Tab. 7.1.1 Some typical automotive underhood reliability requirements

Operating temperature range	−40 to 125 °C
Storage temperature range	−40 to 150 °C
Thermal shock/temperature cycling	−40 to 125 °C, multiple cycles (×10–1000)
Humidity exposure	85 °C/85% RH for 1000 h 60 °C/90% RH for 1000 h Autoclave (121 °C/100% RH) for 96–168 h
Drop shock	Multiple times, 100–200 g
Mechanical shock	Multiple times, 1500–4000 g sometimes over temperature and with power at the device
Fluid compatibility	Multiple automotive fluids in direct exposure
EMI/RFI exposure	Various conditions

a system that could employ single-point sensors. However, airbag systems are becoming more sophisticated than the original implementation because of the addition of side airbags, seatbelt pre-tensioners, and occupant sensing systems that allow smarter deployment algorithms. As more sensing requirements are added to the airbag systems, bus architectures become much more attractive to handle the additional required sensor inputs. Future airbag system will also require improved reliability and the crash-detection response time. Fig. 7.1.15 shows an example of an airbag deployment system that has these features.

Smarter sensors will be required at each of the nodes on the automotive bus to achieve this goal. This requirement is being driven by: 1) the need for improved accuracy (to <5%) through temperature compensation and normalization, 2) remote communications and addressability (with multiple sensors on a single bus), 3) more integrated testability (e.g., autodiagnosis), 4) minimizing the effect of signal processing distortion that results from a larger excitation force than nominal application (oversaturation), and 5) occupant detection for more appropriate airbag deployment force determination. To implement these new devices, substantial system cost must be eliminated to provide incentive for the high cost of new automotive product introduction. Multiple, distributed sensors will communicate with one central processing unit that will make the final deployment decision. These sensor functional blocks will include calibration, A/D, a state machine, and a physical interface to the automotive bus: functions that have been included in other system components in previous architectures. Feature integration, whether on-chip or in-package, provides a means for system cost reduction. An example of

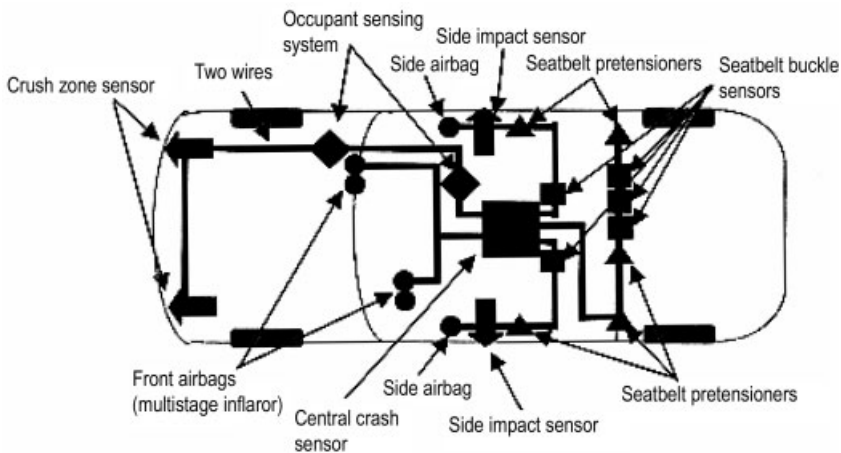


Fig. 7.1.15 An example of an airbag system [99]. To provide more reliable, faster, and more fail-safe sensing, crash sensors are being moved to the external part of the vehicle. This requires some signal conditioning at

the point of use before sending signals through a bus network to a central processing unit. A physical interface to the bus is also required as a part of the 'satellite' sensor

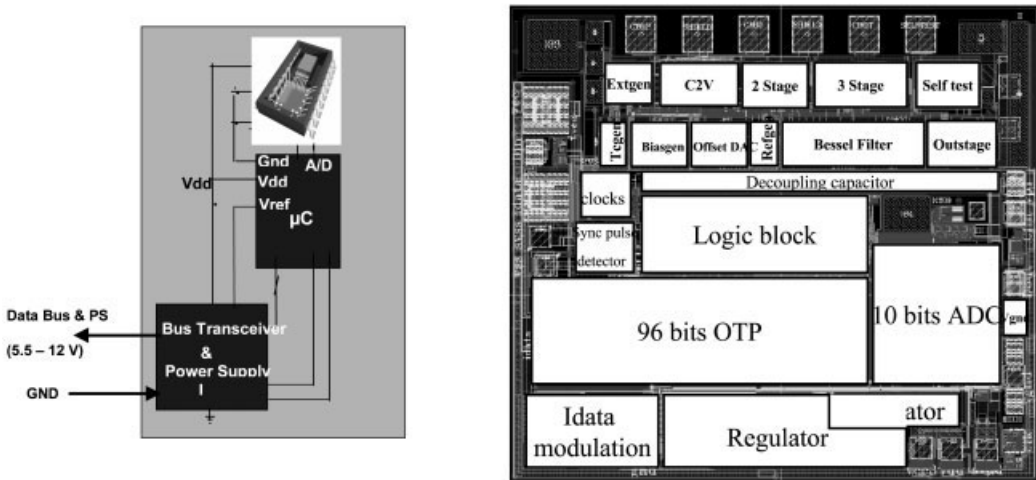


Fig. 7.1.16 a) An application diagram and b) a layout and block diagram of the Motorola SmartMOS7lv ASIC for satellite accelerometer applications

a sensor interface circuit to be used with the Motorola 2-chip accelerometer architecture is shown in Fig. 7.1.16.

This sensor will improve on the functionality of existing accelerometers with the incorporation of an A/D and integrated state machine. The silicon technology has 0.4 μm drawn gate length. It has voltage scaleable analog CMOS devices with maximum VGS of 9 V, and breakdown voltages in the range 7.5–55 V. This 22-layer, triple metal layer process combines existing digital logic with analog circuitry and the ability for integrated power devices. It is ideal for an automotive environment, mixed-signal chip, like that needed to do signal conditioning for a networked accelerometer.

As these networked sensors become more prevalent, clusters of sensors at a given node are also likely to be developed, which will introduce additional functions to the sensor node. For example, in the vehicle dynamics application, both inertial and angular rate sensors will be required. It is likely that these will be clustered into a single module that communicates to the ECU via the sensor network.

7.1.7

Conclusions

Automotive accelerometers have been in production for over a decade, driven by the US Intermodal Surface Transportation Efficiency Act of 1991. MEMS-based accelerometers have won acceptance because they provide a means of single-point sensing for airbag crash detection, which has been the driving application. In the late 1980s, the choice of micromachining technology was unclear. Through the past 10 years of production, surface micromachining, capacitive sensing technol-

ogy has become the technology of choice for crash-detection accelerometers in the 20–250 g sensing range. Low-g sensors, for vehicle dynamics, have generally been bulk micromachined, capacitive sensing technology, for the 1–2 g range, although surface micromachined products are gaining acceptance in this acceleration-range. Stiction remains the most pervasive process quality issue, although design, process rinsing and drying techniques, and special packaging techniques have provided acceptable engineering solutions. Many automotive suppliers today have chosen a ‘System in Package’ approach: a MEMS transducer chip and CMOS-based ASIC built together in a single package. The notable exception to this is the ADI integrated BiCMOS, surface micromachined accelerometer. Most production ASIC technology has been open-loop, switch-capacitive C–V converter architecture. Package technology has been driven down in size from through-hole DIP-style (dual in-time package) packages to surface mount SOIC-style packages to the development of surface mount QFN packages. Test technology is generally partially custom because of the need for physical stimulus (shaking the part). The general trend in the market is to demand lower cost, higher reliability/quality/yield, and higher signal-to-noise (demand for low-g sensors). Finally, with the proliferation of safety features within the car, networked inertial sensors, with some logic at the point-of-use will become much more prominent in the future.

7.1.8

Acknowledgments

Throughout the decade that the author has been in the MEMS industry, there have been several influences and mentors. While much of this work cites specific publications (see references), some of the underlying ideas come from past conversations with Joe Giachino, Bob Sulouff, Richie Payne, Christopher Hierold, Jiri Marek, Ray Roop, Lj Ristic, Mark Shaw, Steve Hendry, Raymond Wiley, John Schuster, Aristide Tintikakis, Demetre Kondylis, Tom Hatfield, Bishnu Gogoi, Dan Koury, Andrew McNeil, Rick August, Walter Czarnocki, Sung Jin Jo, Mike Chapman, Kelvin Blair, Dragan Mladenovic, Yoshio Sekiguchi, Hervé Mureau, Marc Osajda, Kais Mnif, Brett Richmond, and others. Specific thanks goes to Ray Roop for reviewing the manuscript.

The ideas expressed herein are those of the author and do not necessarily reflect the views or positions of Motorola Inc.

7.1.9

References

- 1 A. HUSEIN and R.S. PAYNE, in *Proc. 11th Int. Conf. Solid-State Sensors and Actuators (Transducers '01/Eurosensors XV)*, Munich, Germany, 2001, pp. 1284–1287.
- 2 M. MADOU, *Fundamentals of Microfabrication*, 2nd edn, CRC, Boca Raton, FL, 2002.
- 3 D.S. EDDY and D.R. SPARKS, *Proc. IEEE* 1998, 86, 1747–1755.
- 4 S.D. SENTURIA, *Microsystem Design*, Kluwer, Boston, MA, 2001.
- 5 L. RISTIC, R. GUTTERIDGE, B. DUNN, D. MIETUS, and P. BENNETT, in *Proc. IEEE*

- Solid State Sensor and Actuator Workshop*, Hilton Head, SC, 1992, pp. 118–121.
- 6 L. RISTIC, R. GUTTERIDGE, J. KUNG, D. KOURY, B. DUNN, and H. ZUNINO, in *Proc. Transducers '93*, Yokohama, Japan, 1993, pp. 810–813.
 - 7 L. RISTIC, D. KOURY, E. JOSEPH, F. SHEMANSKY, and M. KNIFFIN, in *Proc. Micro System Technologies '94*, Berlin, Germany, 1994, pp. 77–84.
 - 8 T.A. CORE, W.K. TSANG, and S.J. SHERMAN, *Solid State Technol.* 1993, 39–47.
 - 9 R.S. MULLER, in *Proc. 11th International Conference on Solid-State Sensors and Actuators (Transducers '01/Eurosensor XV)*, Munich, Germany, 2001, pp. 20–23.
 - 10 M. OFFENBERG, H. MUNZEL, D. SCHUBERT, B. MAIHOFER, F. LARMER, E. MULLER, O. SCHATZ, and J. MAREK, SAE Technical Paper 960758, presented at *The Engineering Society for Advancing Mobility Land Sea Air and Space*, 1996, reprinted in *Sensors and Actuators 1996*, SP-1133, 35.
 - 11 M. OFFENBERG, M. ILLING, M. SCHÖFTHALER, R. SCHELLIN, and D. ULLMANN, in *Proc. Sensor '99*, Nürnberg, Germany, 1999, pp. 293–298.
 - 12 H. KUISMA, in *Proc. 11th International Conference on Solid-State Sensors and Actuators (Transducers '01/Eurosensor XV)*, Munich, Germany, 2001, pp. 430–433.
 - 13 J. SULOUFF, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, San Francisco, CA, 1991, pp. 170–176.
 - 14 R. SCHELLIN, A. THOMAE, M. LANG, W. BAUER, J. MOHAUPT, G. BISCHOPINK, L. TANTEN, H. BAUMANN, H. EMMERICH, S. PINTER, J. MAREK, K. FUNK, G. LORENZ, and R. NEUL, in *Proc. Advanced Microsystems for Automotive Applications '99*, Berlin, Germany, 1999, pp. 239–250.
 - 15 M. LUTZ, W. GOLDERER, J. GERSTENMEIER, J. MAREK, B. MAIHÖFER, S. MAHLER, H. MÜNDEL, and U. BISCHOF, in *Proc. Int. Conf. Solid-State Sensors and Actuators: Transducers '97*, Chicago, IL, 1997, pp. 847–850.
 - 16 N. YAZDI, F. AYAZI, and K. NAJAFI, *Proc. IEEE* 1998, 86, 1640–1659.
 - 17 R. MABOUDIAN and R.T. HOWE, *Tribol. Lett.* 1997, 215–221.
 - 18 R. MABOUDIAN and R.T. HOWE, *J. Vac. Sci. Technol.* 1997, 15 (1), pp. 1–20
 - 19 R. MABOUDIAN, *Surf. Sci. Rep.* 1998, 30, 207–269.
 - 20 N. TAS, T. SONNENBERG, H. JANSEN, R. LEGTENBERG, and M. ELWENSPOEK, *J. Micromech. Microeng.* 1996, 6, 385–397.
 - 21 L. RISTIC, *Sensor Technology and Devices*, Artech, Boston, MA, 1994, p. 524.
 - 22 J.Y. KIM and C.-J. KIM, *Proc. IEEE*, The Teuth Annual Workshop on Micro Electro Mechanical Systems. 1997, 442–447.
 - 23 P. RAI-CHOUDHURY, *MEMS and MOEMS: Technology and Applications*, SPIE Press Monograph, SPIE Press, Bellingham, WA, USA, 2000.
 - 24 B.P. GOGOI and C.H. MASTRANGELO, *J. Microelectromechanical Systems* 1995, 4, 185–192.
 - 25 R.L. ALLEY, G.J. CUAN, R.T. HOSE, and K. KOMVOPOULOS, in *Proc. IEEE Solid-State Sensors and Actuators Workshop*, Hilton Head Island, SC, 1992, pp. 202–207.
 - 26 H. GUCKEL and D.W. BURNS, *Sensors and Actuators* 1989, A20, 117–122.
 - 27 C.H. HSU and M.A. SCHMIDT, in *Proc. Solid-State Sensor and Actuators Workshop*, Hilton Head Island, SC, 1994, pp. 151–155.
 - 28 C.H. MASTRANGELO and C.H. HSU, *J. Microelectromechanical Systems* 1993, 2, 33–43.
 - 29 C.H. MASTRANGELO and C.H. HSU, *J. Microelectromechanical Systems* 1993, 2, 44–55.
 - 30 Y. YEE, M. PARK, and K. CHUN, *J. Microelectromechanical Systems* 1998, 7, 339–344.
 - 31 F.M. SERRY, D. WALLISER, and G.J. MACCLAY, *J. Appl. Phys.* 1998, 84, 2501–2506.
 - 32 W. TANG, PhD Dissertation, University of California, Berkeley, 1990.
 - 33 L.S. FAN, PhD Dissertation, University of California, Berkeley, 1989.
 - 34 Y.-C. TAI and R.S. MULLER, in *Proc. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, 1998, p. 88.
 - 35 T. ABE, W.C. MESSNER, and M.L. REED, in *Proc. MEMS '95*, Amsterdam, The Netherlands, 1995, pp. 94–99.
 - 36 G.K. FEDDER and R.T. HOWE, in *Proc. IEEE, MEMS*, 1989, pp. 63–68.

- 37 Y. YEE, K. CHUN, and J. D. LEE, in *Proc. 8th Int. Conf. Solid-State Sensors and Actuators, and Eurosensors IX*, Stockholm, Sweden, 1995, pp. 206–209.
- 38 R. L. ALLEY, P. MAI, K. KOMVOPOULOS, and R. T. HOWE, in *Proc. 7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, 1993, pp. 288–291.
- 39 T. A. LOBER and R. T. HOWE, in *Proc. Solid-State Sensors and Actuator Workshop*, Hilton Head Island, SC, 1988, pp. 59–62.
- 40 J. ANGUITA and F. BRIONES, *Sensors and Actuators* 1998, A64, 247–251.
- 41 M. OFFENBERG, B. ELSNER, and F. LÄRMER, in *Proc. Electrochemical Society Fall Meeting*, Miami Beach, FL, 1994, p. 1056.
- 42 P. R. SCHEEPER, J. A. VOORTHUYZEN, W. OLTHIUS, and P. BERGVELD, *Sensors and Actuators* 1992, A30, 231.
- 43 D. J. MONK, PhD Dissertation, University of California, Berkeley, 1993.
- 44 H. GUCKEL and J. J. SNIEGOWSKI, PhD Dissertation, University of Wisconsin, 1990.
- 45 H. GUCKEL, J. J. SNIEGOWSKI, T. R. CHRISTENSON, and F. RAISSI, *Sensors and Actuators* 1990, A21–A23, 346.
- 46 J. J. SNIEGOWSKI, PhD Dissertation, University of Wisconsin, 1991.
- 47 R. LEGTENBERG and H. A. C. TILMANS, *Sensors and Actuators* 1994, A45, 57–66.
- 48 N. TAKESHIMA, K. J. GABRIEL, M. OZAKI, J. TAKAHASHI, H. HORIGUCHI, and H. FUJITA, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, San Francisco, CA, 1991, pp. 63–66.
- 49 D. KOBAYASHI, T. HIRANO, T. FURUHATA, and H. FUJITA, in *Proc. IEEE MEMS '92*, 1992, pp. 214–219.
- 50 C. H. MASTRANGELO and G. S. SALOKA, in *Proc. IEEE MEMS '93*, 1993, pp. 77–81.
- 51 C. H. MASTRANGELO, X. ZHANG, and W. C. TANG, in *Proc. Transducers '95: 8th Int. Conf. Solid-State Sensors and Actuators, and Eurosensors IX*, Stockholm, Sweden, 1995, pp. 612–619.
- 52 M. ORPANA and A. O. KORHONEN, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, San Francisco, CA, 1991, pp. 266–269.
- 53 F. KOZLOWSKI, N. LINDMAIR, T. SCHEITER, C. HIEROLD, and W. LANG, in *Proc. 8th Int. Conf. Solid-State Sensors and Actuators, and Eurosensors IX*, Stockholm, Sweden, 1995, pp. 220–223.
- 54 G. T. MULHERN, D. S. SOANE, and R. T. HOWE, in *Proc. 7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, 1993, pp. 296–299.
- 55 C. W. DYCK, J. H. SMITH, S. L. MILLER, E. M. RUSSICK, and C. L. J. ADKINS, in *Proc. SPIE Symp. Micromachining and Microfabrication*, Austin, TX, 1996, pp. 225–235.
- 56 M. BIEBL, G. T. MULHERN, and R. T. HOWE, in *Proc. 8th Int. Conf. Solid-State Sensors and Actuators, and Eurosensors IX*, Stockholm, Sweden, 1995, p. 198–201.
- 57 L. RISTIC and M. SHAH, in *Proc. WESCON '96*, Anaheim, CA, 1996, pp. 64–72.
- 58 M. L. KNIFFIN and M. SHAH, *Int. J. Microcircuits and Electronic Packaging* 1996, 19, 75–86.
- 59 M. R. HOUSTON, R. T. HOWE, and R. MABOUDIAN, *J. Appl. Phys.* 1997, 81 (8), pp. 3474–3483.
- 60 M. R. HOUSTON, R. MABOUDIAN, and R. T. HOWE, in *Proc. 8th Int. Conf. Solid-State Sensors and Actuators, and Eurosensors IX*, Stockholm, Sweden, 1995, pp. 210–213.
- 61 U. SRINIVASAN, M. R. HOUSTON, R. T. HOWE, and R. MABOUDIAN, *J. Microelectromechanical Systems* 1998, 7, 252–260.
- 62 P. F. MAN, B. P. GOGOI, and C. H. MASTRANGELO, *J. Microelectromechanical Systems* 1996, 6, 25–34.
- 63 M. NISHIMURA, Y. MATSUMOTO, and M. ISHIDA, in *Technical Digest of the 15th Sensor Symposium*, Kawasaki, Japan, 1997, pp. 205–208.
- 64 Y. MATSUMOTO, K. YOSHIDA, and M. ISHIDA, in *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers '97)*, Chicago, IL, 1997, pp. 695–698.
- 65 B. E. BOSER, in *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers '97)*, Chicago, IL, 1997, pp. 1169–1172.
- 66 D. ULLMANN, G. BRISCHOPINK, M. SCHÖFTHALER, R. SCHELLIN, B. MAIHÖFER, J. SEIBOLD, and J. MAREK, SAE Technical Paper 1999-1901-0757.

- 67 K. E. PETERSEN, *Proc. IEEE*, **1982**, *70*, 420–457.
- 68 MEMSIC, *MEMSIC: The Measurement and Control of Motion through MEMS Technology*, **2002**, *4*, <http://www.memsic.com>.
- 69 M. A. SCHMIDT, *Proc. IEEE* **1998**, *86*, 1575–1585.
- 70 W. H. KO, J. T. SUMINTO, and G. J. YEH, in *Micromachining and Micropackaging of Transducers* (C. D. FUNG, P. W. CHEUNG, W. H. KO, and D. G. FLEMING), Elsevier, Amsterdam, **1985**, pp. 41–61.
- 71 S. A. AUDET and K. M. EDENFELD, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, Chicago, IL, **1997**, pp. 287–289.
- 72 T. A. KNECHT, in *Proc. Transducers '87*, **1987**, pp. 95–98.
- 73 J. M. BUSTILLO, R. T. HOWE, and R. S. MULLER, *Proc. IEEE* **1998**, *86*, 1552–1574.
- 74 C. MASTRANGELO, PhD Dissertation, University of California, Berkeley, **1991**.
- 75 C. H. MASTRANGELO and R. S. MULLER, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, San Francisco, CA, **1991**, pp. 245–248.
- 76 L. LIN, K. M. MCNAIR, R. T. HOWE, and A. P. PISANO, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, **1993**, pp. 270–273.
- 77 K. S. LEBOUITZ, R. T. HOWE, and A. P. PISANO, in *Proc. 8th Int. Conf. Solid-State Sensors and Actuators, and Eurosensors IX*, Stockholm, Sweden, **1995**, pp. 224–227.
- 78 M. W. JUDY and R. T. HOWE, in *Proc. IEEE Micro Electro Mechanical System Workshop*, Ft. Lauderdale, FL, **1993**, pp. 265–271.
- 79 D. J. MONK, P. KRULEVITCH, R. T. HOWE, and G. C. JOHNSON, in *Proc. MRS Spring Annual Meeting*, San Francisco, CA, **1993**, pp. 641–646.
- 80 AIGNER et al. *Circuits Assembly*, **1995**, *March*, 70–72.
- 81 J. SCHUSTER, presented at *Solid-State Sensors and Actuators Workshop*, Hilton Head, SC, **1998**, unpublished (personal communication).
- 82 J. SWIFT, presented at *Solid-State Sensors and Actuators Workshop*, Hilton Head, SC, **1998**, unpublished (personal communication).
- 83 M. MADOU, *Fundamentals of Microfabrication*, CRC Press, Boca Raton, FL, **1997**.
- 84 C. SONG, in *Proc. Int. Conf. Solid-State Sensors and Actuators*, Chicago, IL, **1997**, pp. 839–842.
- 85 H. REICHL, *Sensors and Actuators* **1991**, *A25–A27*, 63–71.
- 86 R. ALLAN, *Electronic Design* **1997**, *20 Jan*, 75–88.
- 87 G. BEARDMORE, in *Proc. IEEE Colloquium on Assembly and Connection in Microsystems*, **1997**, pp. 2/1–2/8.
- 88 S. D. SENTURIA and R. L. SMITH, *Sensors and Actuators* **1988**, *15*, 221–234.
- 89 K. H.-L. CHAU and J. SULOUFF, R. E., *Microelectronics J.* **1998**, *29*, 579–586.
- 90 SensorNor, 6 January 2002. <http://www.sensonor.com/sp13.html>.
- 91 C. GAHN, D. DROSTE, D. SCHUBERT, and M. OFFENBERG, in *Proc. 16th Eur. Conf. Solid-State Transducers*, Prague, Czech Republic, **2002**, to be published.
- 92 H. KAPELS, R. AIGNER, and C. KOLLE, in *Proc. 11th Int. Conf. Solid-State Sensors and Actuators (Transducers '01/Eurosensors XV)*, Munich, Germany, **2001**, pp. 56–59.
- 93 K. FUNK, H. EMMERICH, A. SCHILP, M. OFFENBERG, R. NEUL, and F. LÄRMER, in *Proc. 12th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '99)*, Orlando, FL, **1999**, pp. 57–60.
- 94 H. GRANGE, P. REY, F. MICHEL, M. BORDEL, B. DIEM, and T. LEGAY, in *Proc. 28th Eur. Solid-State Device Research Conf. (ESSDERC '98)*, Bordeaux, France, **1998**, pp. 200–203.
- 95 P. W. BARTH, *Sensors and Actuators*, **1990**, *A21–A23*, 919–926.
- 96 M. W. PUTTY, PhD Dissertation, University of Michigan, **1995**.
- 97 A. Devices, 15 May 1999, <http://www.analog.com/industry/iMEMS/>.
- 98 W. M. STALNAKER, L. S. SPANGLER, G. FEHR, and G. FUJIMOTO, in *Proc. Int. Symp. Microelectronics*, **1997**, pp. 197–202.
- 99 R. FRANK, *Understanding Smart Sensors*, 2nd edn, Artech House, Boston, MA, USA, **2000**.

7.2

Yaw-Rate Sensors

OLIVER SCHATZ

7.2.1

Applications and Requirements

Yaw-rate sensors measure the rotation of a body along a selected axis or the change in angle per unit of time, independent of the reference system. The unit of measurement is degrees per second ($^{\circ}/s$). ‘Angular rate sensor’ and ‘gyroscope’ are synonyms for ‘yaw-rate sensor’. Typical applications in the automotive sector include the following.

- **Electronic stability program (ESP) or vehicle dynamics control (VDC)**
The turning movement of a vehicle around the vertical axis is permanently monitored, for example, changes in direction due to driving through curves or skid movements. The actual value is compared with that expected from the angle of the steering wheel and the vehicle’s speed. If there is a difference, a computer-controlled intervention re-stabilizes the vehicle by braking one or more wheels individually. In future, more sensitive systems will even be able to correct automatically for side winds (enhanced VDC), and for such purposes yaw-rate sensors with a particularly high resolution are required.
- **Roll-over protection**
Yaw-rate sensors recognize roll-over accidents and activate safety equipment, such as head airbags, to protect the passengers.
- **Navigation systems**
Navigation systems record the position of a vehicle using satellite-based location systems (GPS) and match the determined position with the navigation system’s digital street map. Yaw-rate sensors, in combination with wheel-speed sensors, permit interpolation in situations where no satellite reception is possible, such as when driving through tunnels or if the GPS signals cannot be clearly interpreted due to multiple reception as a result of reflections from houses in urban situations, for example.
- **Automatic distance control**
A distance sensor (e.g., a radar sensor) records the distance to vehicles ahead and automatically reduces the speed if the distance to the vehicle in front is not sufficient. Using a yaw-rate sensor, it is possible to pivot the recording area of the sensor when driving through curves. Knowing the steering angle alone is not sufficient because of the flotation angle of the tires. At present automatic distance control systems use the yaw-rate sensor of the VDC via a bus system. Tab. 7.2.1 shows typical specifications for different applications.

Tab. 7.2.1 Typical specifications of yaw-rate sensors

Application	Range ($^{\circ}/s$)	Resolution ($^{\circ}/s$)	Bandwidth (Hz)	Tolerance (%FS)
VDC	± 100	0.1	50	± 1
VDC enhanced	± 100	< 0.02	50	± 2
ROSE ¹⁾	± 250	2	15	± 5
Navigation	± 70	< 0.5	15	± 5

1) Roll over sensing

7.2.2

Functional Principles

Measurement principles based on mechanical gyroscopes or fiber-optic gyroscopes have been known for a long time. These measurement principles provide large signal values. Sensors of this kind are used, for example, in military and space travel applications. However, their use in cars is impractical for reasons of cost. As the accuracy requirements in cars are somewhat lower, vibration gyroscopes are used in automotive applications. These sensors measure the Coriolis acceleration occurring during a rotation in connection with an oscillation movement. The oscillation movement must be actively produced by the sensor. These sensors therefore have two components: the drive unit and the detection unit.

The excitation can be realized by electrostatic, electromagnetic, or piezoelectric methods. The detection is carried out by capacitive, piezoelectric, or piezoresistive techniques. The principle of vibrating gyroscopes can be traced back to Foucault's pendulum in 1851. There are numerous design possibilities for the construction of vibrating gyroscopes [1–5, 9]: Fig. 7.2.1 shows one principle.

A mass M_1 oscillates with a frequency ω_D radial from pivot S. By a turning movement around pivot S, M_1 moves around this pivot with a yaw rate ω_z .

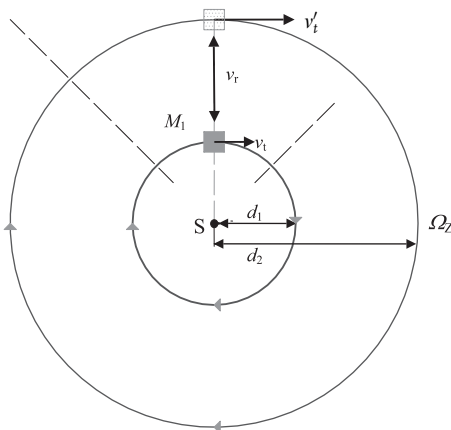


Fig. 7.2.1 A mass M_1 oscillates radial to the pivot S and is accelerated tangentially by Coriolis forces when yaw rate occurs: d_1 , minimum distance; d_2 , maximum distance; v_t , tangential speed of the mass due to yaw rate; v_r , radial speed of the mass due to the oscillation; Ω_z , yaw rate

As the distance of the mass from pivot S is now varied continuously, M_1 runs between a large trajectory around S (distance d_2 =maximum) and a small trajectory around S (distance d_1 =minimum). As a result of this, the trajectory velocity v_t also changes (Fig. 7.2.2). The change of the velocity correlates directly with the yaw rate, as explained in the mathematical description below.

Generally, the position of mass M_1 in Fig. 7.2.2 can be expressed as

$$\vec{r}(t) = (R_1 + R_2 \sin(\omega_D t)) \begin{bmatrix} \sin(\Omega_z t) \\ \cos(\Omega_z t) \end{bmatrix} = r(t) \vec{e}_r(t)$$

where $e_r(t)$ is the radial unit vector, $R_1 = (d_1 + d_2)/2$, $R_2 = (d_2 - d_1)/2$.

The velocity is the first derivative of the position

$$\begin{aligned} \vec{v}(t) &= \frac{d}{dt} \vec{r}(t) \\ &= (R_2 \omega_D \cos(\omega_D t)) \begin{bmatrix} \sin(\Omega_z t) \\ \cos(\Omega_z t) \end{bmatrix} + (\Omega_z)(R_1 + R_2 \sin(\omega_D t)) \begin{bmatrix} \cos(\Omega_z t) \\ -\sin(\Omega_z t) \end{bmatrix} \\ &= v_r(t) \vec{e}_r(t) + v_t(t) \vec{e}_t(t) \end{aligned}$$

$e_t(t)$ is the tangential unit vector. The velocity $v(t)$ has a radial component $v_r(t)$ and a tangential component $v_t(t)$.

The acceleration of M_1 is the first derivative of the velocity and can be expressed as

$$\begin{aligned} \vec{a}(t) &= \frac{d}{dt} \vec{v}(t) \\ &= -(R_2 \omega_D^2 \sin(\omega_D t) + \Omega_z^2 (R_1 + R_2 \sin(\omega_D t))) \begin{bmatrix} \sin(\Omega_z t) \\ \cos(\Omega_z t) \end{bmatrix} \\ &\quad + 2(\Omega_z)(R_2 \omega_D \cos(\omega_D t)) \begin{bmatrix} \cos(\Omega_z t) \\ -\sin(\Omega_z t) \end{bmatrix} \\ &= a_r(t) \vec{e}_r(t) + a_t(t) \vec{e}_t(t) \end{aligned}$$

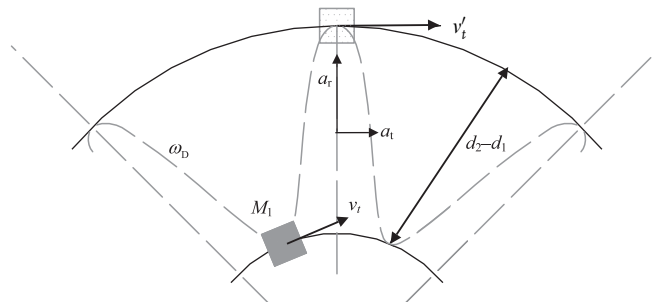


Fig. 7.2.2 Segment of Fig. 7.2.1: the mass M_1 oscillates radial to the pivot S with the frequency $\omega_D \gg$ yaw rate Ω_z ; $d_2 - d_1$ = working stroke; a_r , Coriolis acceleration; a_t , radial acceleration; v_t , tangential speed

The tangential component $a_t(t)$ is the Coriolis acceleration and (by comparison with $a(t)$ and $v(t)$) can be written as

$$a_t = 2\Omega_z R_2 \omega_D \cos(\omega_D t) = 2\Omega_z v_r$$

More generally, the expression for the Coriolis acceleration is

$$\vec{a}_{\text{coriolis}} = 2\vec{\Omega} \times \vec{v}$$

By measuring this acceleration, a signal proportional to the yaw rate is available. At a constant yaw rate, this signal is sinusoidal and the amplitude of the signal increases linearly with frequency ω_D and working stroke R_2 .

The radial component $a_r(t)$ is nearly independent of Ω_z due to the fact that with typical values for R_1 , R_2 , ω_D , and Ω_z the Ω_z term of the radial components $a_r(t)$ is negligibly small

$$a_r(t) \approx -R_2 \omega_D^2 \sin(\omega_D t)$$

The formula shows that the radial acceleration is shifted by 90° out of phase to the desired signal $a_t(t)$ and that it increases quadratically with ω_D .

For a system with a working amplitude $(d_2 - d_1)/2$ of $25 \mu\text{m}$ and a drive frequency of 10 kHz, the following values are obtained at a turning rate of $0.1^\circ/\text{s}$

$$a_t = 5.6 \times 10^{-3} \frac{\text{m}}{\text{s}^2} \quad a_r = 98.5 \times 10^3 \frac{\text{m}}{\text{s}^2}$$

Compared with the acceleration due to gravity at $1g = 9.8 \text{ m/s}^2$, in this example the signal-to-noise ratio is $560 \mu\text{g}$ to $10000g$, which results in approximately -145 dB .

From the example given above it can be seen that much effort is needed to detect a small Coriolis signal in the presence of a disturbing acceleration that is many orders of magnitude larger. These two signals can be distinguished by their direction and phase. Therefore, a high directional selectivity between the drive direction and the detection direction is required. Otherwise the drive movement is transferred to the detection path.

Unfortunately, with increasing drive frequency, the relation a_t/a_r becomes worse. Nevertheless, one tries to realize drive frequencies as high as possible, in order to keep the influence of external stimulation, such as bodywork resonance, as low as possible (Section 7.2.4).

The suppression of the radial acceleration is realized mechanically by using acceleration sensors that are not sensitive to lateral accelerations, and by keeping an orientation of 90° between the drive and detection direction as precisely as possible.

This precise perpendicular alignment can never be perfect. Production tolerances and out of plane vibrations of the drive because of limited stiffness in tangential and vertical directions cause a remaining disturbance signal, which is called the 'quadrature signal'. This signal has to be suppressed by the electronic evaluation circuit by making use of the phase difference between the Coriolis signal and the disturbing radial signal.

For reasons of symmetry, the sensors are normally also formed in mirror image, so there are two bodies oscillating opposite to each other. In consequence the center of gravity of the system remains at rest. For the evaluation, the Coriolis signals of the individual signals are subtracted. Owing to the phase relationship, this doubles the desired signal, and the disturbing lateral acceleration (from driving through curves for example) is eliminated. Fig. 7.2.3 shows an example of a capacitive type of a yaw-rate sensor system.

The sensing element (1) with two acceleration sensors AC_1 , AC_2 feeds the Coriolis signals via two capacitance-to-voltage converters (3, 3') to a summing node. The different signs double the Coriolis signal and eliminate common mode acceleration signals.

The synchronous demodulator (4) eliminates the quadrature signal. This demodulator multiplies the input signal with a reference signal, which is generated by the sensor drive velocity. The signal conditioner (7) corrects any time delay and phase shift, because the reference signal has to relate in the phase exactly to the velocity of the sensor drive.

According to the following equation, the demodulated signal V after the demodulation is

$$\begin{aligned} V &= (C_1 \cos \omega_D t \times \Omega_z + C_2 \sin \omega_D t) \times \cos \omega_D t \\ &= C_1 \cos^2 \omega_D t \times \Omega_z + C_2 \sin \omega_D t \cos \omega_D t \\ &= \frac{C_1}{2} \times \Omega_z \times (1 + \cos 2\omega_D t) + \frac{C_2}{2} \sin 2\omega_D t \end{aligned}$$

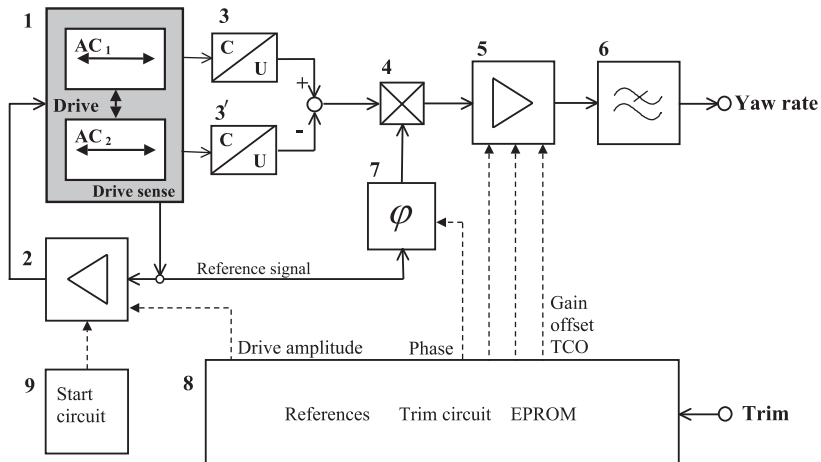


Fig. 7.2.3 Block schematic of a capacitive yaw-rate sensor system: 1, sensor element; 2, drive control unit; 3, capacitance to voltage converter; 4, synchronous demodulator; 5, amplifier; 6, low pass filter; 7, signal conditioner for the reference phase; 8, control unit for trimming; 9, start circuit

By filtering the signal with a low pass filter, the AC components proportional to $2\omega_D$ are suppressed and only the yaw rate ω_z proportional term remains. C_1 and C_2 are circuit-dependent constants.

The amplifier (5) boosts the signal, eliminates temperature coefficients, and adjusts the offset. The low pass filter (6) limits the bandwidth. The drive circuit is a closed loop system to achieve a stable drive oscillation. It consists of a structure (not drawn) that detects the oscillation movement of the drive of the yaw-rate sensor element, a control unit (2), and an actuator (not drawn). The start circuit (9) initiates the drive oscillation at power on. The block (8) generates all necessary adjustment signals and includes the logic circuit for the trim and an EPROM (erasable programmable read-only memory) for the storage of the trim data.

The stability of the drive oscillation and of the exact demodulation substantially determine the achievable tolerances of the sensor. This requires complex and demanding electronic circuitry.

For sensors based on quartz technology some blocks are unnecessary (e.g., capacitance to voltage converters) and other blocks can be designed more simply owing to a higher level of the signals. This advantage, however, is generally reduced by the complexity of the manufacturing of the sensor element itself.

Because of the small signal of vibrating gyroscopes, the sensitivity can be increased by a resonant evaluation. In this case, the drive frequency is set at a value close to the resonance frequency of the acceleration sensors. As a result, the Coriolis signal is increased by a factor K , which can be as large as the quality (Q) factor of the acceleration sensors

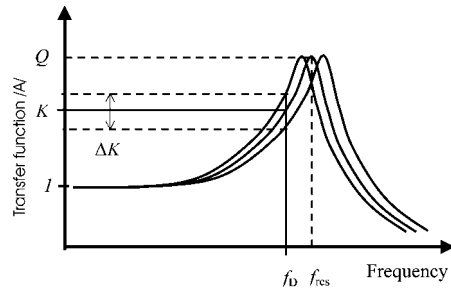
$$a_t = 2 \times v_r \times \Omega_z \times K$$

As it is difficult to measure the resonance frequency of the acceleration sensors in operation, the position of drive and resonance frequency must be defined by the design. For some constructions, the frequencies can be adjusted in the volume production (e.g., laser trim of the sensor elements).

However, the higher the Q -factor, the more complex is the realization, as the resonance frequency also varies due to production tolerances. The closer the drive frequency to the resonance frequency, the stronger the influence at the turning rate signal by frequency drifts, due to lifespan for example, and temperature effects. For this reason, it is more practical to place the drive frequency in the range of the lower gradient on the resonance curve of the acceleration sensors, and thus to exchange the gain in sensitivity with stability (Fig. 7.2.4).

The drive path is normally operated resonantly, to reduce the driving force to a minimum. Depending on the mechanical layout of the sensor elements, Q -factors of 10 to more than 1000 can be achieved. This normally requires the operation of the sensor elements in a vacuum in order to eliminate air damping.

Fig. 7.2.4 Frequency shifts of the resonant frequency f_{res} of the transfer function $|A|$ of the detection path cause changes of the factor K , which increases the sensitivity in a resonant system. The same occurs if the drive frequency f_D varies. Q , Q -factor of the detection system



7.2.3

Examples

7.2.3.1 Linear Vibrating System Based on Quartz Technology

A tuning fork made of quartz is stimulated in such a way that the ends oscillate in opposite directions to each other, as shown in Fig. 7.2.5. Viewing the arrangement at the ends of the tuning fork, it is easy to recognize the principle of Fig. 7.2.1.

When a yaw rate occurs, the Coriolis force leads to a deflection of the ends of the tuning fork out of the oscillating plane. This sensor element consists of a metal tuning fork, on which four piezoelements are located: two for the drive in the lower area and two for detection in the upper area.

By making the sensor element completely out of quartz, the tuning fork body itself is used both for the drive and for detection because of the piezoelectrical properties of the quartz material. Suitable electrode structures supply the drive voltages and pick up the acceleration voltages.

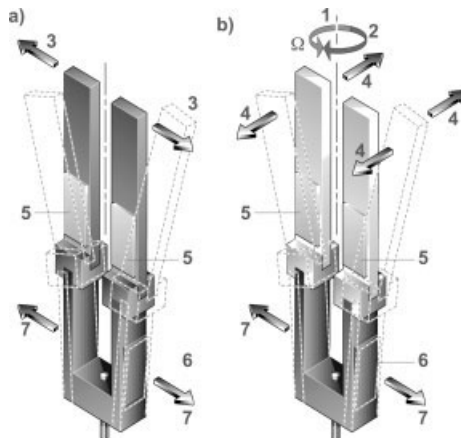


Fig. 7.2.5 Yaw rate on basis of a piezo tuning fork: a) without, b) with yaw rate. 1, 2, yaw-rate directions; 3, oscillation without yaw rate; 4, Coriolis forces; 5, piezo elements for detection; 6, piezo elements for the drive; 7, drive oscillation

Sensors such as this show very high resolution and are therefore theoretically suitable for all applications. The favorable properties of the sensor element keep the evaluation electronics simple.

However, obstacles to this are the cost for the production and assembly of the sensor element. The aging effect of the quartz material requires a pre-aging process, depending on the application.

The fundamental advantages of the tuning fork structure are as follows.

- Low driving voltages.
- The Coriolis acceleration is directly available as a voltage signal.
- Inherent frequency is very stable with a high Q -factor.
- Inherent frequencies in the range 10–40 kHz can be achieved.
- High level of sensitivity, resolution $<0.01^\circ/s$ possible.
- Simple electronic evaluation circuit.

The fundamental disadvantages are the following.

- Costly mechanical manufacturing process, in some cases individual processes.
- Manufacturing tolerances of the sensor element often require a mechanical alignment, by laser trimming of the element for example, in order to minimize the quadrature signal.
- Complicated mounting technique.

One way of producing a tuning fork in silicon by micromachining has been described [5]. This design avoids some of the disadvantages mentioned above (Fig. 7.2.6, Fig. 7.2.7).

Viewing the structure in the direction of the sensitive axis, one can recognize a principle similar to Fig. 7.2.1. Owing to the paddle structure, the Coriolis forces

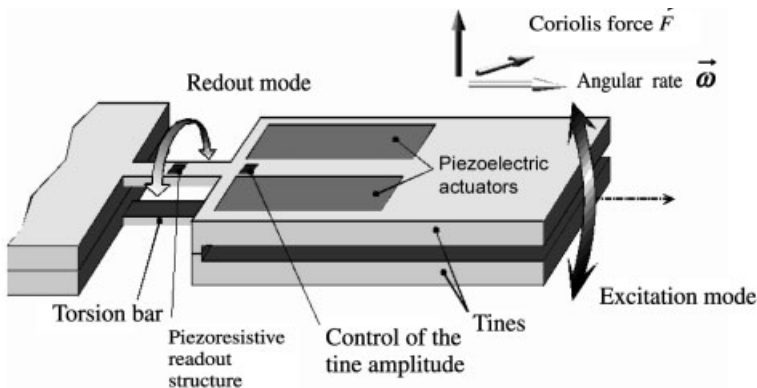


Fig. 7.2.6 Functional principle of a silicon micromachined tuning fork, the tines are excited by piezoelectric actuators, the Coriolis forces bend the suspensions and are detected by piezoresistive resistors in the suspension bars. Source: Conti Temic, Produktlinie Sensorsysteme

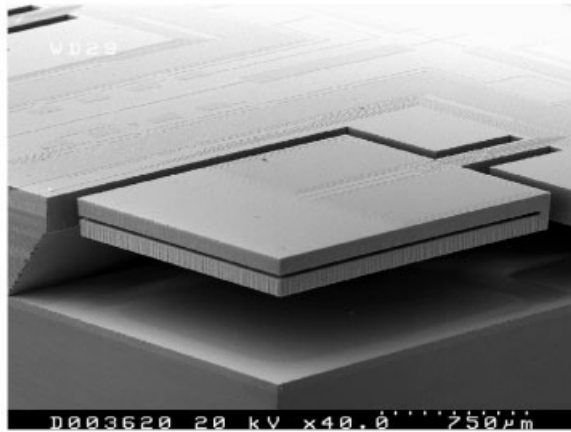


Fig. 7.2.7 Scanning electron micrograph of a silicon micromachined tuning fork. Source: Conti Temic, Produktlinie Sensorsysteme

cause a torsion in the suspension, which is detected by implanted resistors. These resistors change their value with mechanical stress, due to the piezoresistive effect. The drive is produced by deposited and structured piezolayers. This principle effectively decouples the drive and detection. The divided piezoactuators allow a test signal to be applied, which causes a torsion of the paddles and can be detected as yaw rate to check the functionality, even in operation.

7.2.3.2 Linear Vibrating System in Silicon Bulk- and Surface Micromachining

This sensor is also based on the principle shown in Fig. 7.2.1. The oscillating masses are manufactured from a wafer bulk material by an etching process (bulk micromachining). The masses are suspended from the frame by U-shaped structures, representing springs, which serve also as coupling of the masses to each other.

The mechanical resonance frequency is mainly determined by the masses and the coupling stiffness between the masses, here 6 kHz. For the detection an acceleration sensor is placed on each of the two masses, arranged in a comb structure by surface micromachining. The drive works electro-dynamically. A permanent magnet causes a magnetic field vertical to the surface of the sensor element. Conductors on the masses, stimulated by an electrical current, experience a Lorentz force, which drives the masses.

In an other conductor loop placed on the masses, a voltage is induced by the magnetic field, proportional to the oscillating velocity. This voltage signal is the input for the drive control circuit. Its phase is used as reference signal for the demodulation [6] (Fig. 7.2.8, Fig. 7.2.9).

The advantages of this sensor principle are as follows.

- Micromachined techniques enable batch production (silicon wafer processing).
- Simple construction and connection techniques.

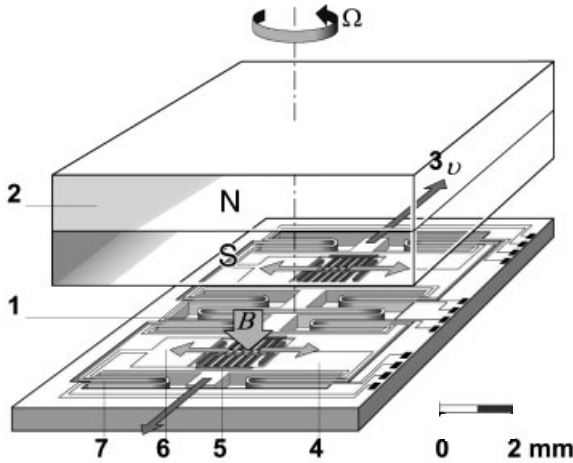


Fig. 7.2.8 Silicon micromachined yaw-rate sensor with electromagnetic drive: 1, coupling spring; 2, magnet; 3, oscillating direction; 4, oscillating mass; 5, acceleration sensor; 6, direction of Coriolis acceleration; 7, spring to

the frame; Ω , yaw rate; ν , oscillation velocity; B , magnetic field. The dimensions of the active sensor surface are about $6\text{ mm} \times 8\text{ mm}$. Source: Robert Bosch GmbH

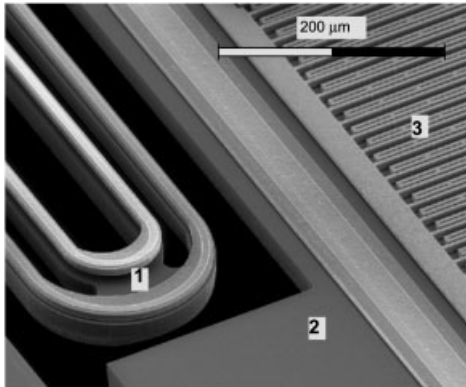


Fig. 7.2.9 Details of the sensor element: 1, spring; 2, mass; 3, Coriolis acceleration sensor. Source: Robert Bosch GmbH

- Electrical decoupling of detection and drive.
The disadvantages are the following.
- Complex electronic circuitry for the evaluation.
- Inherent frequency limited by increasing quadrature signal.
- Temperature drift of the magnet must be taken into consideration.

As a result of micromachining, the sensor element can be produced relatively cheaply. For the assembly of the sensor element, standard mounting processes can be used. Compared to quartz sensors these advantages are compensated in a certain way by a more complex electronics. However, due to modern semiconduc-

Fig. 7.2.10 Sensor module: 1, yaw-rate sensor element; 2, evaluation circuit; 3, acceleration sensor; mounted on a ceramic substrate. Source: Robert Bosch GmbH

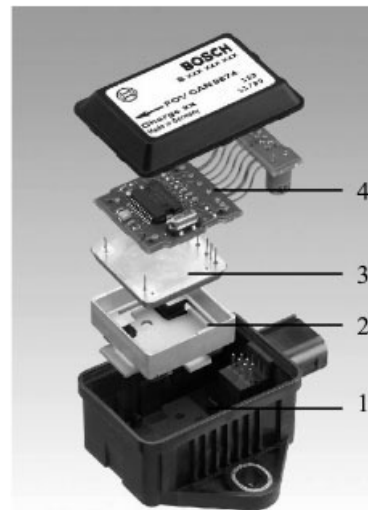
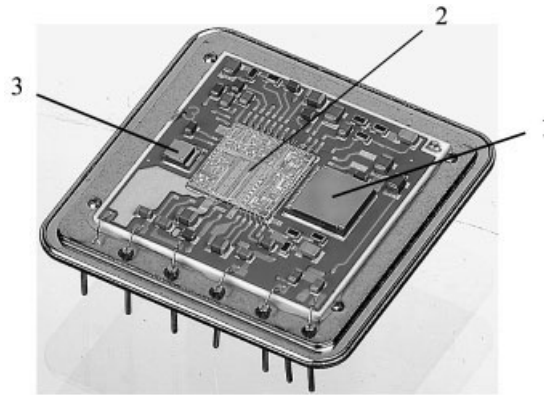


Fig. 7.2.11 Exploded view of the complete yaw-rate sensor system. 1, plastic housing; 2, damper; 3, sensor module; 4, PCB with microcontroller and CAN interface. Source: Robert Bosch GmbH

tor processes, the electronic costs are not the dominant factor any more and are decreasing rapidly at the same time. The relatively low resonance frequency caused by the geometric conditions requires additional measures in vehicle applications, in order to avoid vibrations, for example from the bodywork (see 7.2.4 Application). This sensor is used for vehicle dynamics control.

The following figures show the sensor module with sensor element and evaluation electronics, as well as the complete sensor in a plastic housing with additional microcontroller and CAN (control area network) interface (Fig. 7.2.10, Fig. 7.2.11).

7.2.3.3 Linear Vibrating System in Silicon Surface Micromachining

The functional principle is again in line with Fig. 7.2.1. Oscillating masses and acceleration sensors are produced in the same plane and consist (depending on the micromachining technique) for example of 2–10 μm thick epitaxially deposited polycrystalline silicon.

The drive is produced electrostatically: comb structures, consisting of stationary fingers on the ‘mainland’ and of fingers on the masses to be moved, interdigitated with each other. By applying an oscillating voltage, the masses start oscillating. One example of a sensor element, produced in a 10 μm surface micromachining process, is illustrated in Fig. 7.2.12.

Advantages of this design are as follows.

- Manufacture of drive and detection together, thus very favorable manufacturing costs.
 - One micromachining process for the complete sensor element.
- The disadvantages are the following.
- Disturbances between drive and detection are favored by the same capacitive principle.
 - Capacitive drive enables only low levels of energy coupling, this requires either an extensive drive structure or a high drive voltage and/or the operation of the sensor element in a vacuum.
 - The unfavorable geometrical relations from area to thickness limit the inherent frequency.

The major advantage relates to the production costs of the sensor element. Owing to the high ratio of area to thickness of the active sensor structure, undesired oscillating modes must also be taken into account when designing the element. This limits the inherent resonance frequency in the range 10–20 kHz, depending on the process.

Depending on the chosen system (resonant/not resonant, operation in a vacuum or at ambient pressure) typical resolutions from $0.5^\circ/\text{s}$ to below $0.1^\circ/\text{s}$ in measure-

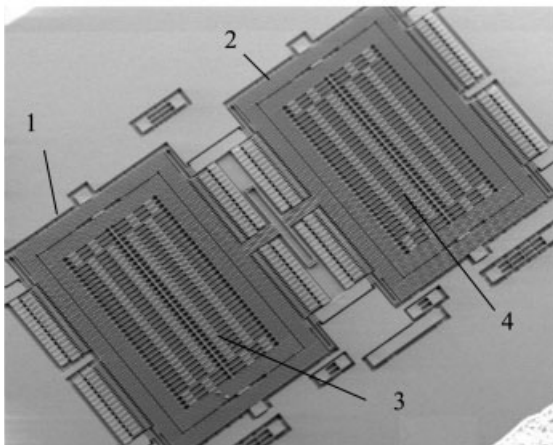
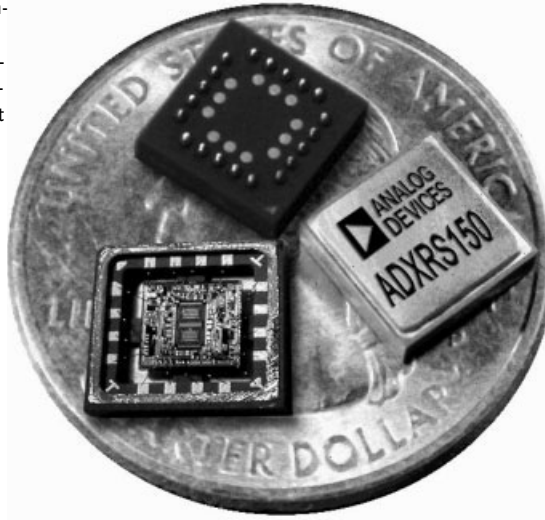


Fig. 7.2.12 Sensor element produced by a micromachining process in 10 μm epi-poly-silicon: 1,2, oscillating masses; 3,4, acceleration sensors.

Source: Robert Bosch GmbH

Fig. 7.2.13 Single chip yaw-rate sensor with a range of $\pm 150^\circ/\text{s}$ and a resolution of $0.5^\circ/\text{s}$. The sensor element works under ambient air pressure conditions with a non-resonant evaluation and is packaged in a $7\text{ mm} \times 7\text{ mm}$ ceramic housing.

Source: Analog Devices Inc



ment ranges of $\pm 100^\circ/\text{s}$ can be achieved. In most sensor systems the sensing function and the evaluation are realized on separate chips. This allows the optimization of the single processes and leads to high yield in the manufacturing. To reduce parasitic effects on the electrical connection between the chips, the minimum capacitance of the element is limited to about 0.5 pF with a resolution of a few aF (10^{-18}).

There exists as well a single chip approach, in which the sensing element and most parts of the electronic circuit are integrated on the same die (Fig. 7.2.13). This approach needs a more complex semiconductor process that influences the yield. But due to much smaller parasitic effects, the sensing element can be reduced and the electronics can be less complex. So the sensor can be very small.

7.2.3.4 Rotational Vibrating System in Silicon Micromachining

For the following example, the function is not directly obvious from Fig. 7.2.1. A flat disc-shaped rotor, suspended by flexible beams at a central point, is set into horizontal, rotational vibrations by comb drive structures. If a turning movement occurs in the plane of the sensor chip, the rotor responds with a perpendicular tilt due to the conservation of angular momentum by Coriolis forces. The distance of the rotor disc to the substrate is detected capacitively and provides a signal proportional to the yaw rate. As the rotary oscillation periodically changes its direction, the whole structure executes tilt oscillations towards the substrate with the frequency of the drive (Fig. 7.2.14, Fig. 7.2.15).

The structure is made of epi-poly-silicon with a diameter of 1.6 mm , a thickness of $10\text{ }\mu\text{m}$, and a distance of $1.5\text{ }\mu\text{m}$ between the detection electrodes and the substrate. A description of the sensor system can be found in Sections 4.1.4.1 and 4.1.4.2 [7]. In principle, the same advantages and disadvantages apply as listed under 7.2.3.3.

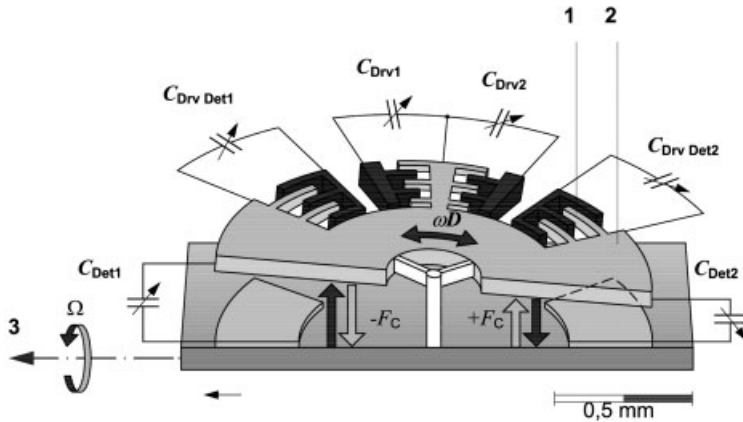


Fig. 7.2.14 Principle of the micromachined rotational yaw-rate sensor: 1, comb structure for drive and drive detection; 2, rotating mass; 3, sensitive direction; F_C , Coriolis force;

ω_D , drive oscillation. The yaw rate Ω causes Coriolis forces F_C , which change the direction periodically with the drive oscillation ω_D . This results in a tilt motion towards the substrate

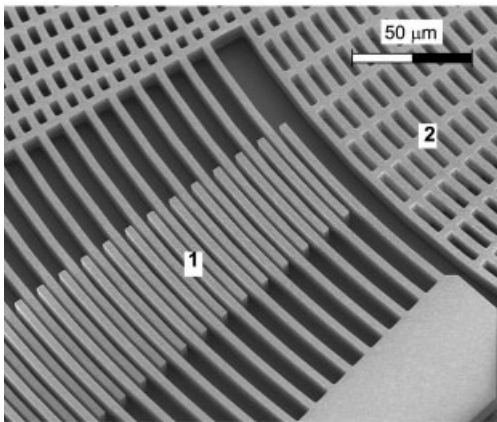


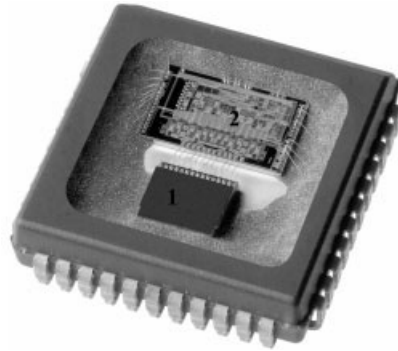
Fig. 7.2.15 Details of the sensor element: 1, comb drive; 2, rotational oscillator. Source: Robert Bosch GmbH

In addition, the sensor principle permits a complete self-test of the sensor, including the rotary oscillator (BITE=built in test). For the self-test, the rotary oscillator is deflected electrostatically via additional electrodes between the oscillator and the substrate. The sensor structure reacts to this in precisely the same way as if a turning movement has occurred: the sensor detects a turning-rate signal. Testing is also possible during operation and thus permits a permanent function monitoring.

Owing to the sensing axis 'in plane', the sensor is used in applications in which an assembly on a horizontal printed circuit board is desirable, for example in air-bag control units for roll-over detection.

The relatively small measurement effect limits the achievable resolution to about $0.5^\circ/\text{s}$ and requires the operation of the sensor in a vacuum. The sensor system is packaged in a PLCC(Plastic Leaded Chip Carrier)-housing (Fig. 7.2.16).

Fig. 7.2.16 Sensor module in an SMD mountable PLCC housing for roll-over detection ($\pm 250^\circ/\text{s}$) and navigation ($\pm 70^\circ/\text{s}$): 1, sensor element with protective cap; 2, evaluation circuit. Source: Robert Bosch GmbH



7.2.3.5 Vibrating Shell Yaw-Rate Sensor

The functional principle of these kind of sensors is based on the effect that was discovered by G. H. Bryan in 1890 [8]. The sensor element of this gyroscope consists of a hollow, axial symmetrical resonator, such as a ring or a cylinder. It is possible to stimulate radial oscillations, which form a kind of standing wave on the wall of the cylinder. If the cylinder rotates around its axis, the nodes of the oscillation waves remain at rest, similar to Foucault's pendulum.

In reality, the nodes move as well, but with a distance factor to the turning of the cylinder. This distance factor is called the Bryan factor. The moving of the nodes can be detected and is a measure for the yaw rate. This principle, when produced by precision mechanics, gives an excellent resolution (about $0.01^\circ/\text{s}$) and is used in the aircraft industry. The following yaw-rate sensor for automotive application works on this basis (Fig. 7.2.17).

A hollow cylinder is stimulated to radial resonance oscillations by piezoelements located on it. As the cylinder is turned around its axis, the nodes of the

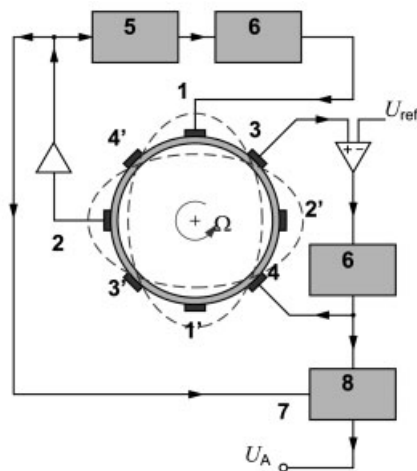


Fig. 7.2.17 1–4, Piezo element pairs; 5, electronics; 6, band-pass filter; 7, reference phase; 8, synchronous demodulator; U_A , output; Ω yaw rate; U_{ref} , self-test input. Piezo pair 1 drives the oscillation, piezo pair 2 controls the amplitude, piezo pair 3 detects movement of the nodes due to yaw rate, piezo pair 4 adjusts the nodes to a minimum at the pair 3



Fig. 7.2.18 Yaw-rate sensor on the principle of the resonant cylinder: 1, cylinder with piezo elements.
Source: Robert Bosch GmbH

standing waves start to move on the cylinder wall. This movement is detected by piezoelements and a closed loop system forces the nodes back, so that the waves stay constant relative to the cylinder. The regulation voltage of the closed loop is a measure of the yaw rate. All together eight piezoceramic elements are attached to the cylinder symmetrically and equidistant.

They operate as: drive, amplitude control of the drive, detectors for the oscillation nodes, and as actuators to keep the nodes in place. The principle enables a complete self-test by simulating a yaw rate via the piezoactuators of the closed loop. A yaw-rate sensor based on the principle of a vibrating cylinder is used in VDC systems (Fig. 7.2.18). This principle has been used in a similar way in silicon micromachining and is used in a yaw-rate sensor for VDC [9].

7.2.4

Application in the Vehicle

The application of yaw-rate sensors in vehicles requires particular attention. The dynamic requirements for the sensors are high: Coriolis accelerations in the range of milli-g must be detected correctly, and at the same time the accelerations occurring in the range of several g must not interfere with the sensor function (such as when driving over potholes). In the case of the roll-over sensor this even applies for accelerations $> 30\text{ g}$, such as in a crash.

External accelerations superimpose the sensor element's acceleration due to the Coriolis force, that is, the desired signal. By means of a symmetrical construction, it is possible to eliminate errors by subtracting the Coriolis signals of the oscillating masses; the external acceleration forces affect these masses in the same direction, the Coriolis forces, in contrast, act out-of-phase.

Measurement errors occur if the linear range of the Coriolis detectors is exceeded. It is therefore recommended that yaw-rate sensors, especially when permanent availability is required as in the case of vehicle dynamics control, are placed as centrally as possible inside the vehicle to reduce the influence of disturbing accelerations.

Yaw-rate sensors are spring/mass systems that can be stimulated by external influences in an undesired way on one of the operating frequencies, that is on the drive frequency, the detection resonance, and (depending on the sensor system) also on the difference and summation frequencies.

Impacts (stones hitting against the base of the vehicle when driving on unpaved roads, potholes, accidents) cause mechanical frequencies with a broad spectrum. Resonant properties of the mounting location in the car amplify these disturbances. If the frequencies match with the sensors frequencies, this can lead to malfunction of the sensor, even at nominally slight external accelerations.

The remedy to this is to place the resonance frequencies of the bodywork and the sensor as far apart as possible, and to make sensors with high working frequencies. The mounting of mechanical dampers to prevent transfer of the bodywork resonance frequencies to the sensor is common practice. Mechanical dampers within the sensor itself are also used. Typical working frequencies for yaw-rate sensors are 2–40 kHz, bodywork and printed circuit board resonance frequencies are typically below 5 kHz, or for very stiff light metal bodywork, up to 20 kHz.

The sensitivity of the sensors should also be taken into consideration in the car assembly. The shock resistance level for typical sensors allows shock acceleration in the range of 5000 g, even for the typical fragile structures of yaw-rate sensors. Undamped impacts, from assembly tooling for example, however, reach accelerations of over 10000 g and thus can destroy the sensors or cause preliminary damage. A more robust layout of the sensor is basically in conflict with the desired high sensitivity.

7.2.5

References

- 1 N. HEDENSTIERNA, S. HABIBI, S. M. NILSEN, T. KVISTERØY, Bulk Micromachined Angular Rate Sensor based on the 'Butterfly-Gyro Structure', *Proc. MEMS 2001*, Interlaken, Switzerland, *Technical Digest, IEEE Robotics and Automation Society*, Interlaken, Switzerland, **2001**.
- 2 K. Y. PARK, H. S. JEONG, S. AN, S. H. SHIN, C. W. LEE, Lateral Gyroscope Suspended By Two Gimbals Through High Aspect Ratio ICP Etching, in *Proc. Conf. Transducers*, Technical Papers, Institute of Electrical Engineers of Japan, Sendai, Japan, **1999**.
- 3 T. TSUCHIYA, Y. KAGEYAMA, H. FUNABASHI, J. SAKATA, Vibrating Gyroscope Consisting Of Three Layers Of Polysilicon Thin Films, in *Proc. Conf. Transducers*, Technical Papers, Institute of Electrical Engineers of Japan, Sendai, Japan, **1999**.
- 4 T. FUJITA, K. HATANO, K. MAENAKA, T. MIZUNO, T. MATSUOKA, T. KOJIMA, T. OSHIMA, M. MAEDA, Vacuum sealed Silicon Bulk Micromachined Gyroscope, in *Proc. Conf. Transducers*, Technical Papers, Institute of Electrical Engineers of Japan, Sendai, Japan, **1999**.
- 5 S. SASSEN, R. VOSS, J. SCHALK, E. STENZEL, T. GLEISSNER, R. GRÜENBERGER, F. NEUBAUER, W. FICKER, W. KUPKE, K. BAUER, M. ROSE, Silicon Angular Rate Sensor for Automotive Applications with Piezoelectric Drive and Piezoresistive Read-out, in *Proc. Conf. Transducers*, Technical Papers, Institute of Electrical Engineers of Japan, Sendai, Japan, **1999**.
- 6 M. LUTZ, W. GOLDERER, J. GERSTENMEIER, J. MAREK, B. MAIHÖFER, S. MAHLER, H. MÜNZEL, U. BISCHOF, A Precision Yaw-rate Sensor In Silicon Surface Machining, in *Proc. Conf. Transducers*, Digest of Technical Papers, Transducers Research Foundation Staff, Chicago, IL, USA, **1997**.
- 7 R. SCHELLIN, A. THOMAE, M. LANG, W. BAUER, J. MOHAUPT, G. BISCHOPINK, L. TANTEN, H. BAUMANN, H. EMMERICH, S. PINTER, J. MAREK, K. FUNK, G. LORENZ, R. NEUL, A Low Cost Angular Rate Sensor For Automotive Applications In Surface

- Micromachining Technology, in *Advanced Microsystems for Automotive Applications '99*, DETLEF E. RICKEN, WOLFGANG GESSNER, Berlin, Germany, 1999.
- 8 G. H. BRYAN, *Proc. Camb. Phil. Soc.* **1890**, VII (24 Nov), 101–111.
- 9 A. J. HARRIS, J. S. BURDESS, D. WOOD, R. LANGFORD, G. WILLIAMS, M. C. L. WARD, M. E. MCNIE, *J. Micromech. Microeng.* **1998**, 8, 284–292.

7.3

Pressure Sensors

IWAO YOKOMORI and YASUTOSHI SUZUKI

7.3.1

Introduction

Nowadays, pressure sensors are typical of automotive sensors and are widely used in automotive applications. The history of the automotive pressure sensor began with the manifold absolute pressure sensor used for engine combustion control in 1979 [1]. Then it found wider applications in the high-pressure field, such as suspension pressure detection and air-conditioner refrigerant pressure detection. After the introduction of OBD, a US on-board diagnosis standard, their application also extended to low-pressure fields, such as vaporized gasoline leak detection. The application of pressure sensors is now extending further to the very high-pressure field such as gasoline fuel injection and diesel common-rail fuel injection systems [2].

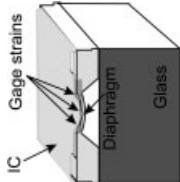
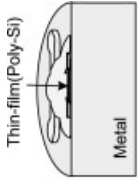
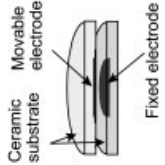
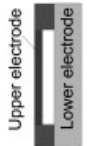
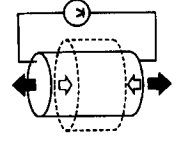
7.3.1.1 Principles and Characteristics of Pressure Sensors

Tab. 7.3.1 provides a summary of the principles and features of the pressure sensors. The pressure sensors are broadly divided into three groups according to which principle they employ: piezoresistive sensors, capacitive sensors, and piezoelectric sensors.

Fig. 7.3.1 shows the principle of the piezoresistive sensor. Diffused resistors (gages) are formed on the thin-walled section called the diaphragm. An applied pressure is detected via the piezoresistive effect, which is the change in electrical resistance when a stress is applied to the diaphragm. The sensitivity is determined by the material, diameter, and thickness of the diaphragm. The thin-film piezoresistive sensor offers low sensitivity because the piezoresistive coefficient of thin-film silicon is less than one-third of that of single-crystal silicon.

Fig. 7.3.2 shows the principle of the capacitive sensor. The electrode consists of a fixed electrode and a movable electrode formed in the thin-walled diaphragm. When a pressure is applied, the diaphragm is distorted and the width of the gap between the above two electrodes is changed. The sensor retrieves the change in the gap width as the change in capacitance. The change in capacitance ΔC is expressed by the following equation.

Tab. 7.3.1 Principles and features of pressure sensors

Principle	Piezoresistance		Capacitance		Piezoelectricity
	Silicon piezoresistance	Thin-film piezoresistance	Ceramic capacitance	Thin-film capacitance	
Structure	 <p>IC Gage strains Diaphragm Glass</p>	 <p>Thin-film (Poly-Si) Metal</p>	 <p>Ceramic substrate Movable electrode Fixed electrode</p>	 <p>Upper electrode Lower electrode</p>	
Sensitivity	Middle $\Delta R = 1/2R\Delta\sigma\pi_{44}$	Low $\pi_{44} = \text{One sixth of single-crystal silicon}$	High	High	Low
Integration	Easy	Difficult	Difficult	Easy	Difficult
LSI process	Easy	Difficult	Difficult	Easy	Difficult

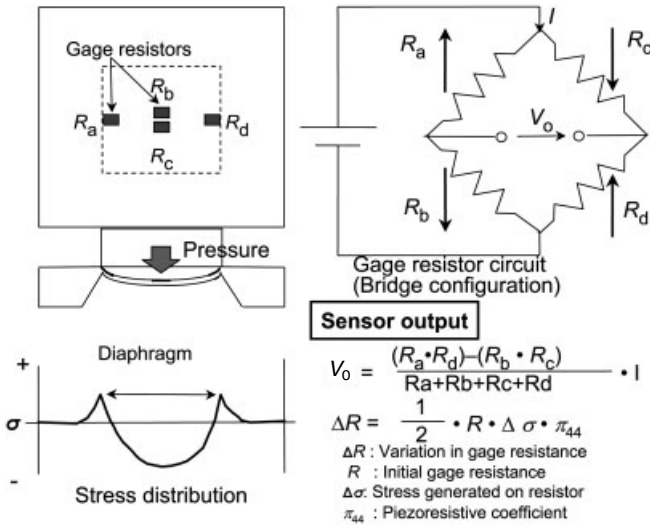


Fig. 7.3.1 Principle of piezoresistive sensor

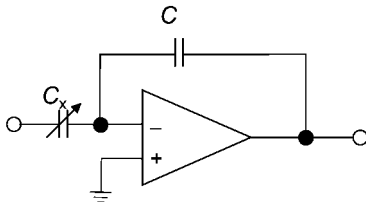


Fig. 7.3.2 Principle of capacitive sensor

$$\Delta C \propto \frac{W^4}{\Delta t^3} \times \Delta P \tag{1}$$

The sensitivity of the capacitive sensor is inversely proportional to the cube of the gap width, which means that the technology of narrowing the gap evenly is important for the sensor.

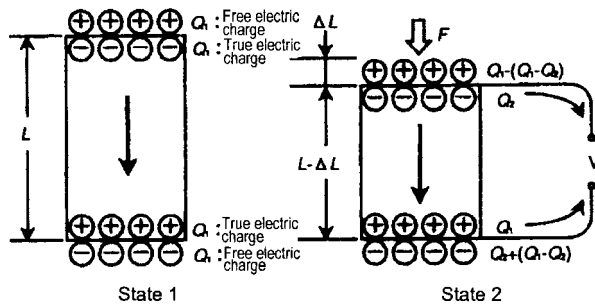
Fig. 7.3.3 shows the principle of the piezoelectric sensor [3]. The sensor utilizes the voltage generated when a force or a strain is applied to a piezoelectric material.

7.3.1.2 Application of Automotive Pressure Sensors

Fig. 7.3.4 gives a summary of the main automotive pressure sensors now commercially available, and their applications, according to the pressure range.

Our automotive pressure sensor line offers a wide range of choices. At one end of the range is the gasoline vapor pressure sensor for 5 kPa used to detect leaks from the gasoline pipe. At the other end is the common-rail pressure sensor for

Fig. 7.3.3 Principle of piezo-electric sensor




	Low	Medium	High		
Pressure range	5 k Pa	100 k Pa	1~10 MPa	20 M Pa	200 M Pa
Product	Gasoline vapor 	Intake pressure 	Air-conditioning Suspension 	Fuel pressure 	Common rail 

Fig. 7.3.4 Automotive applications of DENSO pressure sensors (Si piezoresistive sensors)

Tab. 7.3.2 Operating temperature and accuracy requirements for automotive pressure sensors

Example for accuracy requirements

<i>Application</i>	<i>F.S. pressure</i>	<i>Accuracy</i>	<i>Temperature range [°C]</i>
Evaporation monitoring	5 KPa	2%FS	-30 to 120
Speed density EFI	100 KPa	1%FS	-30 to 120
Suspension	2 MPa	2%FS	-30 to 120
Air-conditioning	3.5 MPa	2%FS	-30 to 135
Gasoline diesel injection	20 MPa	2%FS	-30 to 120
Common-rail	200 MPa	1%FS	-30 to 120

200 MPa used in the common-rail system: in other words 40 000 times the pressure at the low end of the range.

The automotive pressure sensor is characterized by its strict requirements. Tab. 7.3.2 gives sample accuracy requirements for automotive pressure sensors [2]. The pressure sensor used in the engine compartment is required to operate over a temperature range from -30 to 120 °C, and the pressure sensor for air-condi-

tioners from -30 to 135°C . Accuracy of 1–2% full-scale deflection (FSD) is required of both, in spite of their wide operating temperature ranges.

7.3.2

Technologies

To satisfy the demand for pressure sensors with a wide operating temperature range and high accuracy, three critical technologies, relating to the pressure-sensing element, the signal-processing circuit, and the package, are required. To detect medium pressure, the pressure-sensing element for the absolute pressure sensor has a reference cavity (a vacuum cavity in most cases) while that for the differential pressure sensor uses barometric pressure as the reference pressure. The signal-processing circuit amplifies and conditions the output signal from the pressure-sensing element, and then produces an electric signal necessary for the vehicle-mounted electronic control unit (ECU). The package is required to have a structure to support the sensor, derive signals, and protect the internal components. The following sections, taking the intake pressure sensor with a silicon piezoresistive element as an example, give detailed explanations of the sensing element, the signal-processing circuit, and the package.

7.3.2.1 Pressure-Sensing Element

The silicon piezoresistive sensor detects distortion. Therefore diffused resistor (gage) layout and diaphragm geometry design have a profound effect on the sensor characteristics. We utilize simulation technology for pressure-sensing element design.

Gage Layout

The sensitivity of a piezoresistive pressure sensor depends on the piezoresistive coefficient. Silicon crystal face selection and gage layout on the crystal face are important because of the anisotropy of the piezoresistive effect. Silicon (100) and (110) are often used with P-type diffused resistors to achieve a desired sensitivity. The next consideration is the thermal stress effect originating from the silicon crystal face. Fig. 7.3.5 shows the stress-distribution maps for silicon (100) and silicon (110) by the finite element method (FEM).

The output voltage generated by thermal stress results from the difference in gage resistance change, $R_a - R_b$, due to gage positions. For silicon (110), the output voltage is low because only the stress in the current direction (the A–A' direction) must be taken into account. For silicon (100), stress components both in the current direction (the A–A' direction) and its orthogonal direction (the B–B' direction) must be taken into account. R_a and R_b , the gage resistance changes for R_a and R_b , are of opposite signs, resulting in a high output voltage. In one specific example, silicon (100) is used with a unique gage layout on the thin-walled diaphragm.

Fig. 7.3.6 shows its gage layout called a picture-frame design [4]. This picture-frame design with a bridge configuration in a tiny area provides both better linearity and a higher signal-to-noise ratio with the small gage area.

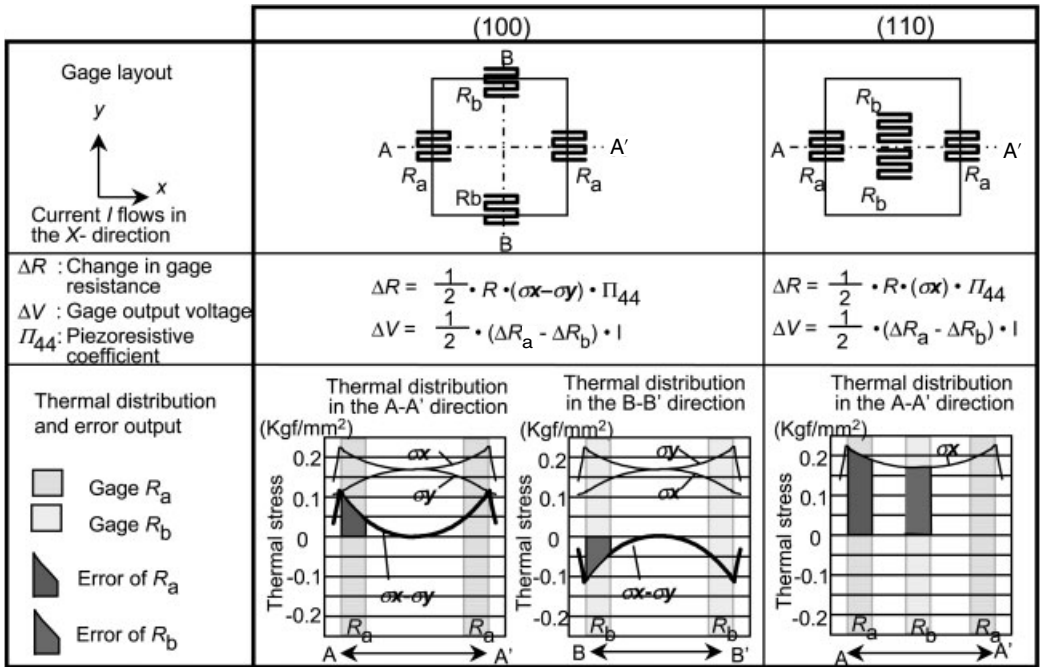


Fig. 7.3.5 Stress distribution maps for silicon (100) and silicon (110)

Diaphragm Geometry

In conjunction with gage layout, the geometry of the thin-walled diaphragm (thickness and diameter) determines the sensor characteristics. The sensor sensitivity is proportional to the diaphragm diameter and inversely proportional to the second power of the diaphragm thickness.

$$S \propto \sigma \propto A/t^2 \tag{2}$$

The diaphragm is formed by anisotropic etching using an etching solution such as KOH. Fig. 7.3.7 shows the anisotropically-etched diaphragm geometries for silicon (100) and silicon (110) designed via an etching simulator [5].

The thin-walled area for the anisotropically-etched silicon (100) diaphragm is made in a square. On the other hand, the thin-walled area for the silicon (110) diaphragm can be made in a square or in another polygon. Since the shape of the diaphragm thin-walled area has a great effect on the sensor characteristics, the diaphragm geometry is designed so as to optimize the sensor characteristics.

7.3.2.2 Signal Processing Circuit

The signal from the pressure-sensing element must be processed in some way into a desired electric output in order to be transferred to the vehicle ECU. The signal-processing circuit design can be divided into the following three parts.

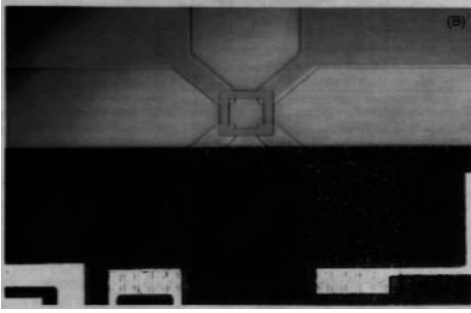


Fig. 7.3.6 Picture-frame photograph

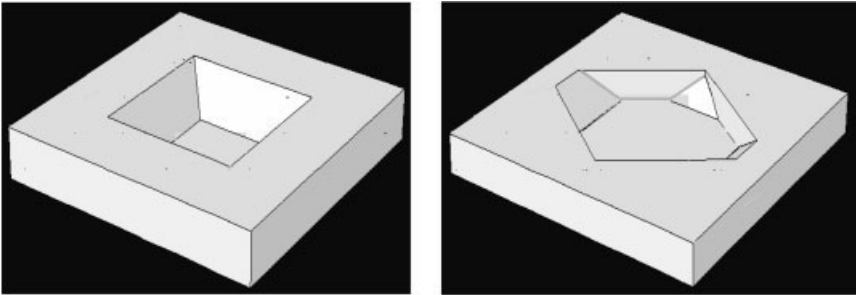


Fig. 7.3.7 Diaphragm geometries for silicon (100) and silicon (110)

Temperature-Compensation Circuit

P-type silicon piezoresistive elements in high concentration of 10^{20} cm^{-3} are used to make a bridge configuration. The resistance of those elements increases with temperature (positive temperature characteristic). On the other hand, the piezoresistive coefficient π_{44} determines the sensibility decrease with temperature (negative temperature characteristic). The temperature-compensation circuit is made by making use of these characteristics. Fig. 7.3.8 shows the principle of the temperature-compensation circuit and the concept of temperature compensation.

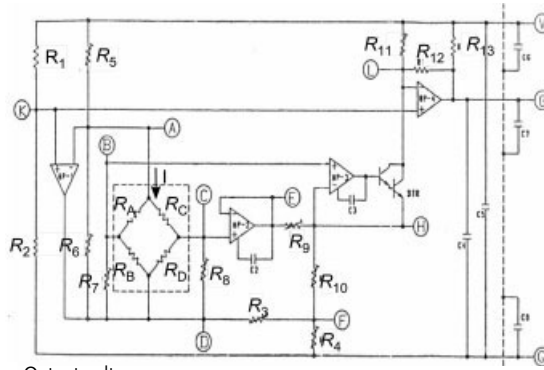
Signal Conditioner

The sensor output, a continuous analog function, is expressed by the following Taylor series [2].

$$V(P) = A_{000} + A_{100}P + A_{010}t + A_{001}\sigma + A_{200}P^2 + A_{020}t^2 + A_{002}\sigma^2 + A_{110}Pt \\ + A_{101}Pa + A_{011}\sigma + A_{300}P^3 + A_{030}T^3 + A_{003}\sigma^3 + \dots \quad (3)$$

where P =pressure, t =temperature, σ =strain of the element, A_{lmn} =partial differential constant of each term. Up to which term in equation 3 to correct must be determined by considering the balance between cost and accuracy. The concept

Fig. 7.3.8 Principle of temperature compensation circuit



Output voltage

$$= \Delta R \cdot I = G \cdot R \cdot \epsilon \cdot I = G_0(1+\beta T)R_0(1+aT)\epsilon I = G_0R_0\{1+(a+\beta)T+a\beta T^2\}\epsilon I$$

R_0 : Resistance at RT, G_0 : Gage ratio at RT, a : Resistor temperature coefficient >0 , β : Gage ratio temperature coefficient <0

From $(a+\beta)T > a\beta T^2$, Output voltage $= G_0R_0\{1+(a+\beta)T\}\epsilon I$

If gage concentration is chosen so as to make $(a+\beta)$ zero, then output voltage becomes $G_0R_0\epsilon I$, independent of temperature T

for the correction is: correct the fewest possible terms by improving the sensor device (gage) accuracy.

With the pressure sensor, linear errors (i.e., offset voltage, offset voltage temperature characteristics, sensitivity, and sensitivity temperature characteristics) can be corrected and reduced with a simple circuit.

On the other hand, correction and reduction of nonlinear errors, such as TNO (non-linearity of offset voltage temperature characteristics), TNS (non-linearity of sensitivity temperature characteristics), and Nlp (non-linearity of pressure characteristics) requires complex circuits and large amounts of memory, with many terms needing to be corrected. That means a larger chip size and a greater number of man-hours for error correction, resulting in greater cost. From the viewpoint of cost reduction, we correct only linear errors by conditioning as few as four characteristics: room-temperature offset, room-temperature sensitivity, high-temperature offset, and high-temperature sensitivity. As for nonlinear errors, we take measures against TNO, TNS, and Nlp by optimizing the pressure sensor device structure. Linear errors are corrected with the circuit. Correction methods include:

- Laser trimming of the thin-film resistor.
- Electric trimming using EPROM.

A pressure sensor employing a DSP (digital signal processor) recently appeared on the market. This is a result of progress in CMOS microprocessing, providing the potential for size and cost reduction of the CMOS digital signal processor that would replace the conventional analog signal processor utilizing. Fig. 7.3.9 shows the digital signal conditioner of the sensor employing a DSP [6].

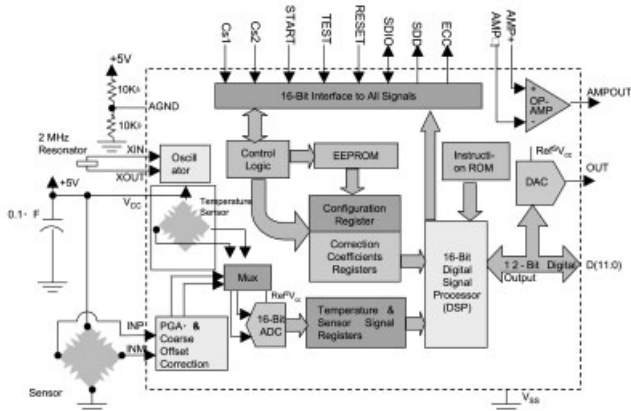


Fig. 7.3.9 Digital signal conditioner with digital signal processor

Electric Noise-Protection Circuit

For the purpose of size and cost reduction, we integrate circuit elements for electric noise protection into one chip by semiconductor circuit integration technology. Fig. 7.3.10 shows the integration of circuit elements for EMC (electromagnetic compatibility) [2].

The first generation of the sensor was realized in a hybrid configuration, comprising a discrete sensor device and circuit elements. For EMC, the entire package was contained in a sealed case to reduce radiated noise. Induced noise from the harness was conducted to the sealed case via a feedthrough capacitor, and then eliminated by eddy current. In the second generation, circuit elements integration and consequent wire length minimization eliminated the need for a sealed case. In the third generation, a low-pass filter, comprised of a capacitor and a resistor, has been additionally integrated to eliminate the need for the feedthrough capacitor. That results in a very simple pressure sensor device with circuit elements for EMC integrated into a single chip.

7.3.2.3 Package

Against the background of installation standardization and for the purpose of cost reduction by eliminating the hose, the intake pressure sensor recently has been directly mounted on the engine. That means a harsher usage environment for the sensor, involving more severe contamination at the pressure detection point and more severe sensor vibration.

Fig. 7.3.11 shows a photograph and Fig. 7.3.12 shows an assembly structure of the surface-detection type intake pressure sensor. The sensor device of the surface-detection type intake pressure sensor is directly exposed to various contaminants. We therefore adopt a structure featuring the following.

- The signal-processing circuit susceptible to contamination is placed in the mold package and only the pressure-sensing element is exposed outside the package.

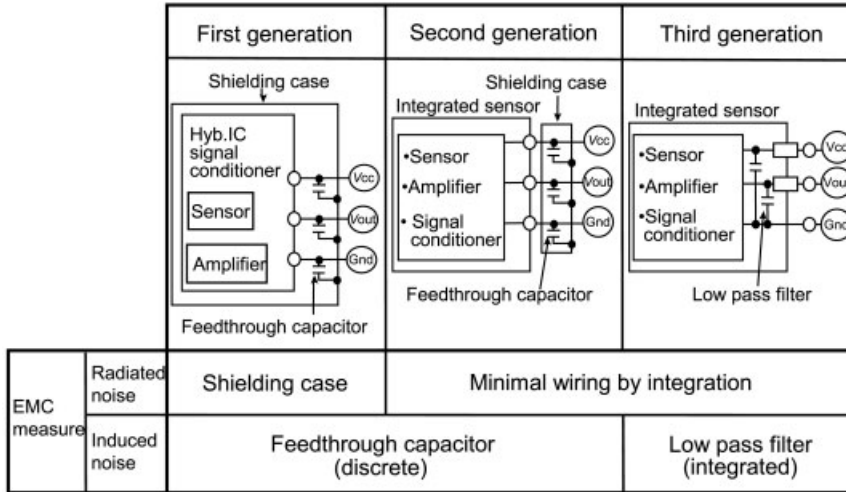


Fig. 7.3.10 Integration of circuit elements for EMC (electromagnetic compatibility)



Fig. 7.3.11 Intake pressure sensor photograph

- The pressure-sensing element is protected by a three-layer protection structure comprising a SiN passivation film with superior moisture resistance, a parylene film for protecting the bonding wire, and gel for protecting the whole element. This structure makes it possible to achieve the creation of a surface-detection intake pressure sensor with high contamination immunity.

7.3.3

Other Examples

7.3.3.1 Fuel Vapor-Pressure Sensor

Since the sensor is intended to detect very low pressures of 5 kPa, reduction in the sensing element error is important to achieve high accuracy. We therefore put some ideas into designing the sensing element glass base and the package [2].

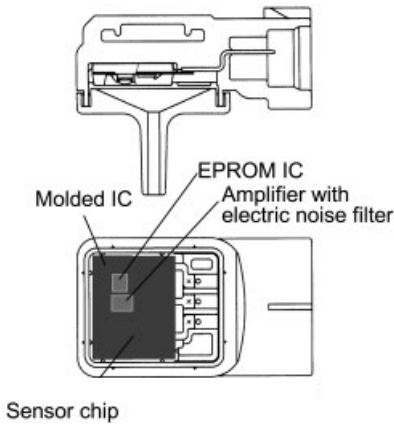


Fig. 7.3.12 Intake pressure sensor assembly structure

Glass Base

The actual pressure sensor is offered in a package made of resin or metal. When the pressure device is directly bonded to the package, the sensor characteristics are affected by the difference in physical properties such as thermal expansion coefficient and Young's modulus between silicon and the package material. The glass base is used for lessening the effect of the difference. We have learnt, however, that the thermal expansion coefficient of the glass base itself also has a profound effect on the temperature characteristics of the sensor.

Silicon and glass are anodically bonded at a high temperature about 400 °C. Therefore, the following distortion occurs at room temperature (25 °C).

$$\int_{25}^{400} [a_G(T) - a_{Si}(T)]dT = -250 \text{ ppm}$$

Likewise, the following distortion occurs at a temperature of 60 °C.

$$\int_{60}^{400} [a_G(T) - a_{Si}(T)]dT = -230 \text{ ppm}$$

As a result, the amount of distortion changes by 20 ppm as the temperature changes from 25 to 60 °C.

We incorporated the above values in the analysis of the sensitivity's temperature characteristics. The analysis results are shown in Fig. 7.3.13. The temperature characteristics of the glass base thermal expansion coefficient from 400 °C (for anodic bonding) to 25 °C (room temperature) are important to reduce TNS. We achieve near-zero TNS by making the above temperature characteristics of glass close to those of silicon.

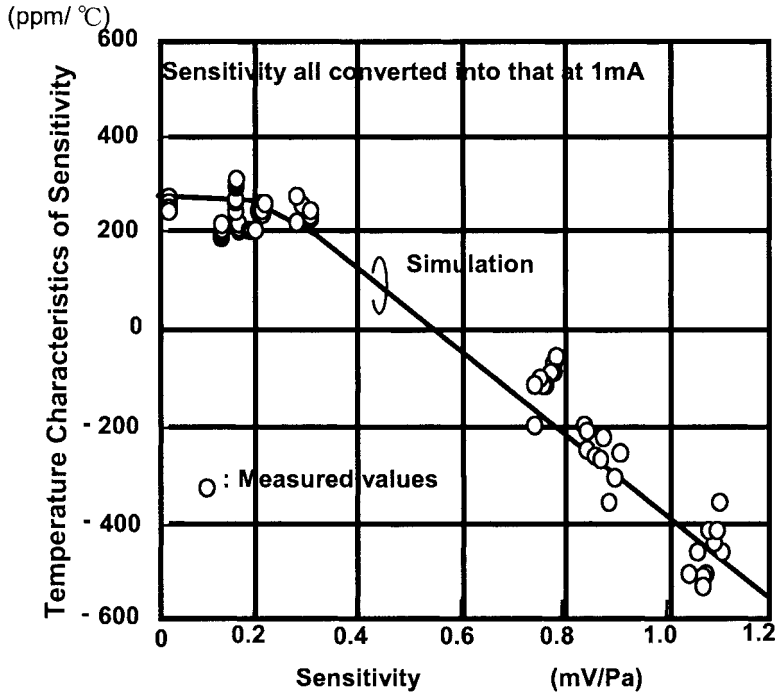


Fig. 7.3.13 Analysis results

Package

To improve the fuel vapor-pressure sensor accuracy, we reduced the mechanical strain conveyed from the package to the sensor device. Fig. 7.3.14 shows the structure of the fuel vapor-pressure sensor for 5 kPa. We chose a relatively thin silicon diaphragm, 14 μm , to achieve the sensitivity required for 5 kPa detection. That makes the sensor device more susceptible to the mechanical strain conveyed from the resin package. To solve that problem, we analyzed the effect of the mechanical strain from the package by FEM.

Fig. 7.3.15 shows the analysis results. The key factors for improving the accuracy of the low-pressure sensor for 5 kPa are as follows.

- A stem that stands midway between the package and the glass in thermal expansion coefficient, to be added between them.
- Fluorine-based adhesive with low Young's modulus and high gasoline resistance to be used to bond the glass base to the case.

These measures allow reduction in the effect of mechanical strain on the sensing element, resulting in a pressure sensor with a high accuracy of 2% FSD at 5 kPa.

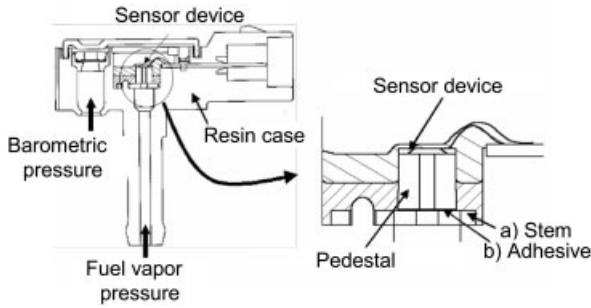


Fig. 7.3.14 Structure of fuel vapor-pressure sensor

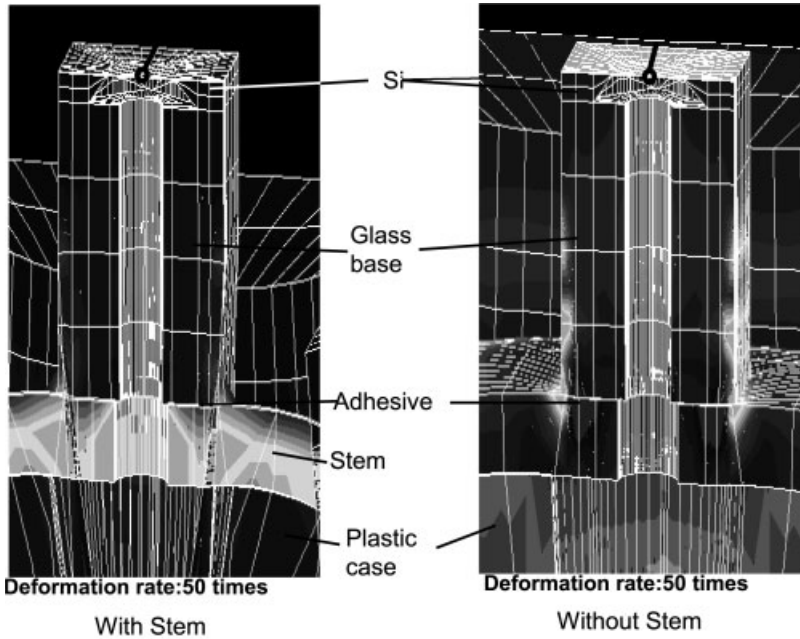


Fig. 7.3.15 Analysis results

7.3.3.2

Fuel Pressure Sensor

Fig. 7.3.16 shows the fuel pressure-sensing principle. Fuel pressure is converted into silicon oil pressure via the diaphragm. The silicon integrated pressure sensor device utilizing piezoresistive effect detects the silicon oil pressure. The conventional high-pressure sensor has a hermetically sealed structure for oil sealing. The large number of parts in this structure makes for high cost. We have developed a new technology to seal oil in a resin case, in which the molecular diameter of the sealed-in oil is larger than that of the resin case [7].

Fig. 7.3.16 Fuel pressure-sensing principle

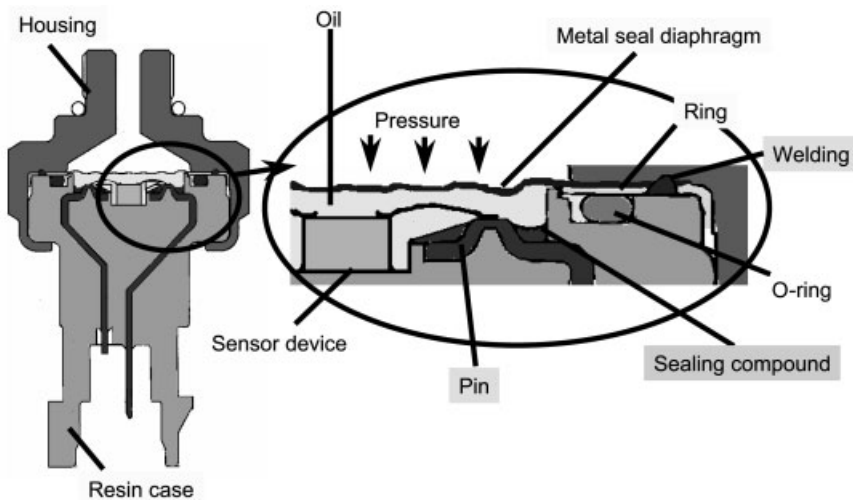
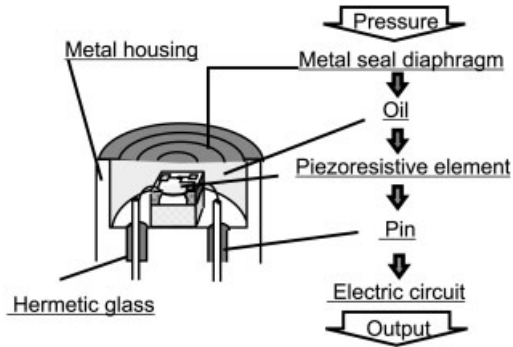


Fig. 7.3.17 Schematic cross section

Fig. 7.3.17 shows a schematic cross section of the fuel pressure sensor. The sensor device integrating piezoresistive gages and circuit elements into one chip is bonded to the center of the sensor. Oil is sealed in the resin package by pin sealing and O-ring sealing of the metal diaphragm. With pressure resistance up to 40 MPa, the sensor is applicable to a wide variety of automotive pressure media including fuel, oil, refrigerant, and air.

7.3.3.3 Common-Rail System Pressure Sensor

The common-rail system pressure sensor is required to have high-pressure resistance at 200 MPa and a high accuracy of 1% FSD as the system covers a higher pressure range and therefore requires higher accuracy. Fig. 7.3.18 shows a schematic cross section of the common-rail system pressure sensor [8]. The metal stem has a diaphragm for receiving very-high-pressure fuel at 200 MPa. We adopt

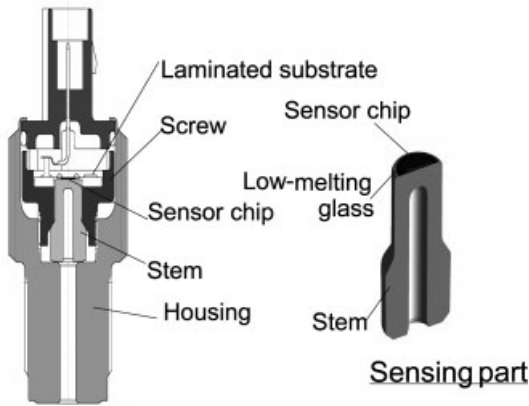


Fig. 7.3.18 Schematic cross section of common-rail system pressure sensor

a metal touch-sealed structure for the sensing part, to ensure a good seal between metals against pressure. A single-crystal silicon piezoresistive element, bonded directly to the diaphragm, detects pressure stress on the diaphragm for fuel pressure detection. To achieve an accuracy of $\pm 1\%$ FSD, the structure must be designed taking the above into consideration. Here we describe in detail the (100) single-crystal silicon substrate and gage layout optimization for high accuracy.

(100) Single-Crystal Silicon Substrate

The Nlp of the silicon (100) piezoresistive element is described by the following equation

$$\text{Nlp} \approx -0.25\eta_{44}/\pi_{44} \times \{(\sigma'_{xx} - \sigma'_{yy}) + (\sigma_{xx} - \sigma_{yy})\} \times 100 (\% \text{FSD}) \quad (4)$$

where η_{44} = quadratic coefficient of π_{44} .

As is evident from equation 4, Nlp can be made zero by making both $(\sigma'_{xx} - \sigma'_{yy})$ and $(\sigma_{xx} - \sigma_{yy})$ zero. Fig. 7.3.19 shows the result of FEM stress analysis by sensor device simulation. With the piezoresistive gages arranged concentrically with the inner circumference of the metal stem, both $(\sigma'_{xx} - \sigma'_{yy})$ and $(\sigma_{xx} - \sigma_{yy})$ can be made equal to zero, resulting in zero Nlp. On the other hand, since TNS is proportional to the temperature characteristics of $(\sigma'_{xx} - \sigma'_{yy}) + (\sigma_{xx} - \sigma_{yy})$, TNS can be made nearly zero by locating the piezoresistive gages so as to make $(\sigma'_{xx} - \sigma'_{yy}) + (\sigma_{xx} - \sigma_{yy})$ zero.

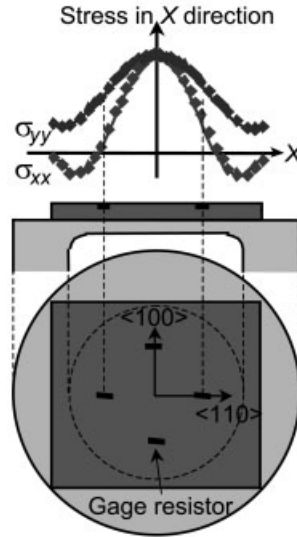
Gage Layout Optimization

The offset voltage temperature characteristics are given by the following equation.

$$e = 1/2 \times i \times R\pi_{44}(\sigma_{xx} - \sigma_{yy}) \quad (5)$$

where i = gage current; $R\pi_{44}$ = coefficient. Equation 5 indicates that in order to make TNO equal to zero, $(\sigma'_{xx} - \sigma'_{yy})$ and $(\sigma_{xx} - \sigma_{yy})$ must be equal independently

Fig. 7.3.19 FEM stress analysis result for (100) silicon



of temperature when the applied pressure is zero. We therefore conducted FEM analysis to determine the optimum gage layout. With this method, the stress value differential between the silicon chip low- and high-temperature sides is obtained in the form of a distribution. If the piezoresistive gages are arranged in the zero-stress area, theoretically TNO becomes zero.

Fig. 7.3.20 shows the results of prototype evaluation for Nlp, TNO, and TNS of the pressure sensor with the optimal gage layout on its (100) silicon element. Nlp, TNO, and TNS are nearly zero, as expected.

7.3.3.4 Media-Compatible Pressure Sensors

Pressure sensors present a potential reliability problem when pressure signals originate from harsh environments, such as that in an engine manifold or in an automotive tire.

Two approaches were investigated to provide selective encapsulation for surface-micromachined pressure sensor devices [9]. The first method uses a glass-frit-bonded bulk micromachined silicon cap wafer, and the second method uses a patterned polymeric gel-like material in the assembly area to create this dam.

Fig. 7.3.21 shows the selective encapsulation processes.

1) Patterned Polymeric Dam Processing (Fig. 7.3.21 a, b)

A polymeric dam material is dispensed into a square shape around the transducer diaphragm and cured. The most critical aspects of this process is the adhesion of the various layers together and the media compatibility of the materials.

2) Silicon Dam Processing: (Fig.7.3.21 c, d, e)

The second approach for selective encapsulation presented here uses a bulk-micromachined bonded wafer with the surface micromachined pressure sensor.

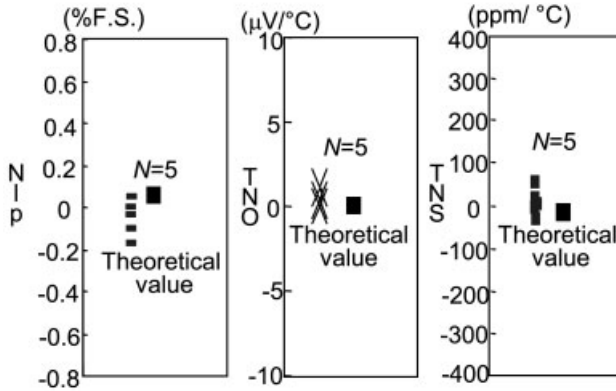


Fig. 7.3.20 Prototype evaluation results

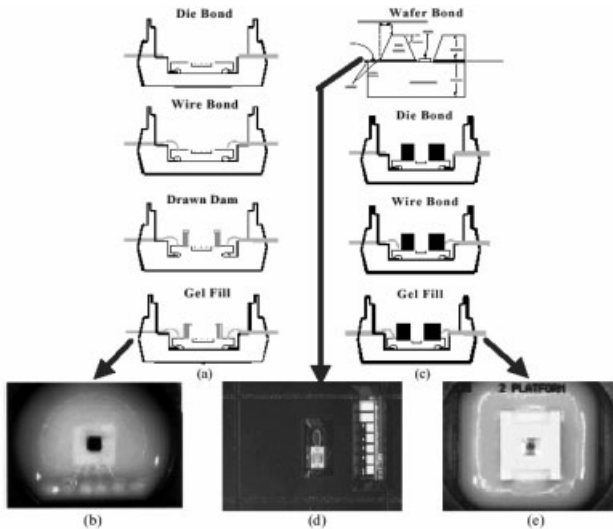


Fig. 7.3.21 The selective encapsulation processes

Following wafer bonding, the silicon dam die has two openings: one in the center for exposure of the P-cell to the pressure environment and the second for the wirebond pads.

A critical point that was discovered during this assembly development was a design rule for the height of the die with respect to the leadframe, the wirebond, near the leadframe, may be higher than the top of the dam. Both methods have passed exposure testing to harsh chemical environments, including nitric acid and salt water, at elevated temperature.

7.3.3.5 A Pressure Sensor for X-by-wire Automotive Applications

Many of the servo-assisted electro-hydraulic applications operate in two pressure ranges: a low range, where high precision is needed, and a high pressure one, where stability is a primary requirement.

Fig. 7.3.22 shows a silicon pressure sensor with a two levels sensitivity[10]. The sensor consists of a membrane at which centre is applied , from top, the membrane will move freely downward. Fig. 7.3.23 shows sensor response extrapolated from FE Simulation.

At 30 bar the cylinder will enter into contact with the silicon bulk plate.

The membrane will not be able to move freely any more and will behave rather like a ring fixed at the two sides. The sensitivity will be roughly one quarter than the one between 0 and 30 bar.

7.3.4

Conclusions and Future Trends

The automotive applications of semiconductor pressure sensors utilizing the large piezoresistive coefficient of silicon have been greatly expanded, owing to the following three advantages.

- Silicon is a superior physical quantity detector.
- Low-cost process employing semiconductor technology in widespread use can be utilized.
- Advanced semiconductor technology allows integrated structure.

This means that the pressure sensor has become a paradigm of the evolution of the semiconductor sensor, and can be said to be a leader among similar sensors.

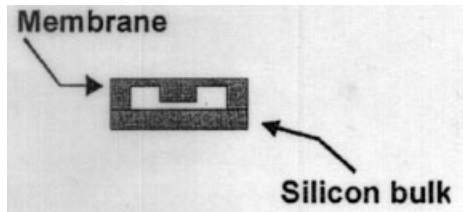


Fig. 7.3.22 Diagram of the sensor chip profile

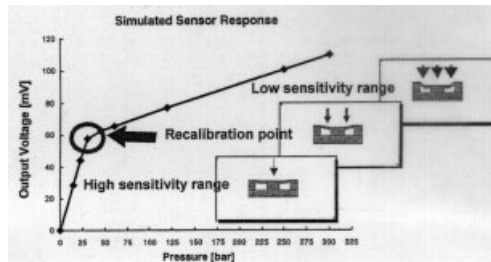


Fig. 7.3.23 Sensor response extrapolated from FE simulation

To develop any sensor with the potential to replace the piezoresistive pressure sensor, it is essential to impart the ability to detect the same physical quantity from very low to high pressure, to employ micromachining technology, and to produce a simple one-chip package [11]. If these goals can be achieved, the prospect of combining reduced cost with a many-fold increase in the number of pressure sensor applications per vehicle may no longer be a dream.

7.3.5

References

- 1 R. H. GRACE, *Proc. Sens. Expo Baltim*, **1999**, 351–358.
- 2 I. YOKOMORI, T. MIZUNO, *CONVERGENCE 2000*, 2000-01-C054.
- 3 K. NISHIO, *Sensors for Engine Control*, San-kaido, Tokyo, Japan, **1999**.
- 4 G. BITKO, A. MCNEIL, *J. Appl. Sens. Technol.* **2000**, 62–67.
- 5 K. ASAUMI, Y. IRIYE, K. SATO, *Proc. IEEE Micro Electro Mechanical Systems (MEMS) Workshop*, **1997**, 412–417.
- 6 A. MAKDESSIAN, M. PARSONS, *J. Appl. Sens. Technol.* January, **2001**, 38–46.
- 7 S. OTAKE, M. ONODA, K. NAGASE, *Spec. Publ. SAE 1998*, SP-1312, 61-68.
- 8 Y. SUZUKI, H. TANAKA, M. IMAI, M. HARRISON, N. OBA, *Spec. Publ. SAE 2002*, SP-1076.
- 9 G. LI, J. SCHMIESING, A. MCNEIL, K. NEUMANN, B. GOGOI, G. BITKO, S. PETROVIC, J. TORRES, M. FUHRMANN, D. J. MONK, *Transducers '01*, **2001**, 1C3-01.
- 10 L. TOMASI, G. KROETZ, E. WIESER, W. SUEDEKAMP, P. THIELE, E. OBERMERIER, *Advanced Microsystems for Automotive Applications Yearbook 2002*, Springer, **2002**, 261–270.
- 11 J. MAREK, in *Proc. 13th Eur. Conf. Solid-State Transducers*, September 12–15, **1999**, The Hague, The Netherlands, 1–8.

7.4

High-Pressure Sensors

JÖRG GEBERS

7.4.1

Introduction

Measurement of high pressures in the range 100–2000 bar is required in new cars. Stringent regulations regarding car safety and pronounced awareness of environmental considerations are behind the severe demands being made on high-pressure sensors. As a result, there is a need for high-pressure sensors that deliver reliable accuracy with lifelong stability using low-cost technologies. Such sensors are used in the following applications (Tab. 7.4.1).

- Electrohydraulic brakes (EHB) [1]
This system is able to generate pressure in the wheel cylinders independently of driver action. Six pressure sensors are necessary to detect the pedal travel ('braking request'), the pressure of the high-pressure accumulator, and the brake pressure in the individual wheels.
- Electronic stability program (ESP) [2]
This system improves driving safety by preventing lateral instability of the vehicle. Besides sensors for wheel speed, steering-wheel movement, and yaw rate, a high-pressure sensor is needed to detect the brake pressure.
- Gasoline direct injection [1]
Compared with conventional manifold injection, a reduction in fuel consumption of up to 20% can be achieved by injecting gasoline directly into the combustion chamber. When this is done, the fuel pressure is increased to 120 bar to ensure fuel atomization. From the pressure detected by the sensor (Fig. 7.4.1) the quantity of injected fuel can be derived from the length of injection period. Fig. 7.4.2 shows the main components of gasoline direct-injection systems and pistons indicating the combustion chambers of the engine.
- Diesel direct injection [1]
The diesel common-rail system allows control of the injection pressure independently of engine speed. Since the quantity of injected diesel is defined by the injector opening time and the system pressure, the latter has to be detected by a

Tab. 7.4.1 High-pressure sensors in automotive applications and typical pressure sensor ranges

Automotive application	Full scale pressure (bar)
Electrohydraulic braking (EHB)	200
Electronic stability program (ESP)	250
Gasoline direct injection	140
Diesel direct injection	2000



Gasoline Direct Injection 14 MPa



Diesel-Common Rail 180 MPa

Fig. 7.4.1 High-pressure sensor for gasoline and diesel direct injection

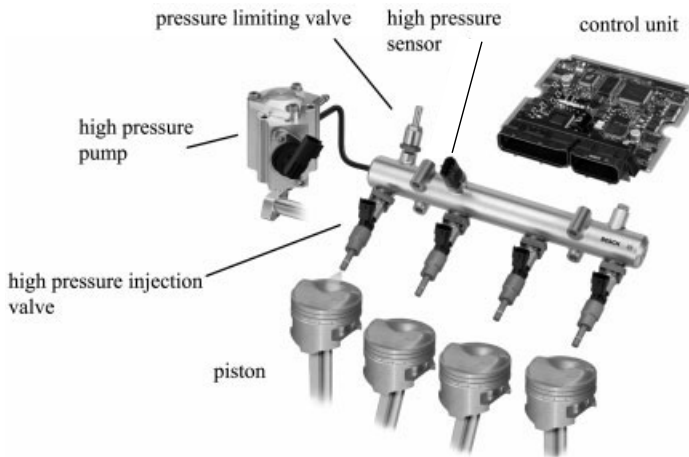


Fig. 7.4.2 Components of gasoline direct-injection system with high-pressure sensor

pressure sensor. Sensors with a nominal pressure up to 2000 bar are necessary to cover pressure pulses generated by the pump.

7.4.2

Requirements

Typical requirements for automotive high-pressure sensors can be defined from the environmental conditions in the engine compartment and the needs of the applications, as follows.

- Hydraulic mounting interface, usually by a threaded connection (Fig. 7.4.1).
- Customer-specific electrical connector.
- Environmental conditions:
 - temperature between -40 and 120°C ,

- humidity and salt-spray,
- industrial climate.
- High burst-pressure and strict separation of hydraulic medium from environment (no leakage).
- Resistance against vibration (strongly dependent on the location in the car).
- Electromagnetic resistance up to 200 V/m.
- Direct electrostatic discharge.
- Characteristic deviation of less than 2.5% of full scale deflection (FSD) in the above mentioned temperature range and less than 0.5% FSD over a lifetime.
- Self-diagnostic functions for safety reasons.
- Low cost.

To fulfill these requirements with one process technology there are only a few appropriate sensor principles. The main challenge is to maintain accuracy despite the environmental conditions and vibrations.

7.4.3

Technology

High-pressure sensors differ mainly in the measuring principle used and the type of internal connection. Most sensor designs use a steel diaphragm to separate the pressure fluid from the sensor signal and environment.

7.4.3.1 Sensor Elements Based on Steel Diaphragms

This section deals with sensors using a steel diaphragm to give a hermetically sealed design. By strictly separating the hydraulic medium from the sensor electronics and the environment, high safety requirements and high burst-pressures can be guaranteed. The measuring principle of these types of sensing elements can be seen in Fig. 7.4.3. When pressure is applied to the diaphragm the stress causes a strain. This elongation can be detected by strain gages that are placed on the opposite side of the medium. Since there is stress in the middle of the membrane and strain in the outer regions, a Wheatstone bridge can be designed with two strain gages in the middle and two others on the edge. To maximize the output signal, the locations of the strain gages on top of the diaphragm are determined by finite-element methods (FEM) (Fig. 7.4.4). The maximum values of strain and stress are used to locate the strain gages. By optimizing the shape of the diaphragm the output signal can be made to have a linear correlation with the applied pressure.

The pressure range can be chosen by changing diaphragm parameters like thickness, diameter, and edge radius between the membrane and the drill-hole. The design of the diaphragm is important to guarantee an appropriate signal range for the amplifying and calibrating function of the ASIC (Application-Specific Integrated Circuit) and sufficient pulsation stability during the lifetime. A stainless steel such as 17-4 PH is usually used for the diaphragm because of its high yield strength and resistance against corrosion.

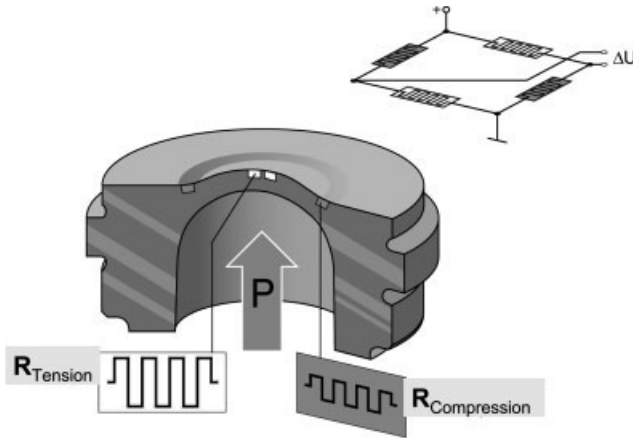


Fig. 7.4.3 Measuring principle of steel diaphragm (in this case with thin-film strain gages)

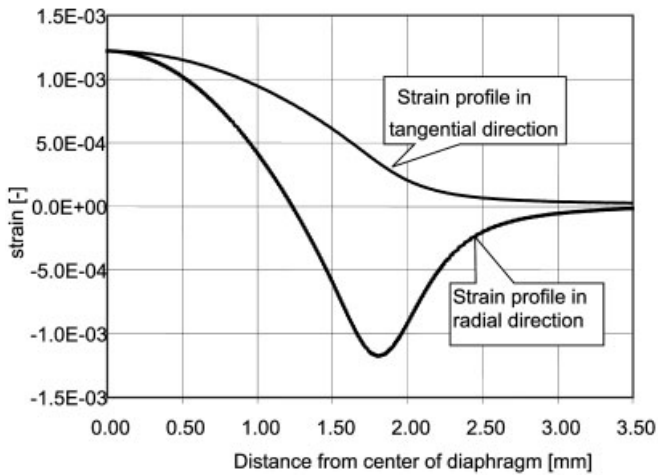


Fig. 7.4.4 Strain profile of steel diaphragm as a result of finite-element simulation

Placing strain gages on steel diaphragms can be done in several ways. In automotive applications up to 2000 bar three different technologies are widely used: thin-film technology, microfused silicon strain gages, and a fused single chip (Fig. 7.4.5).

Thin-Film Strain Gages on Steel Diaphragm

With this sensor technology, good stability against corrosion, high accuracy, and lifetime stability can be achieved. Detailed aspects of thin-film technology are described in Chapter 5.4 (Thin films on steel). To achieve low production costs, vacu-

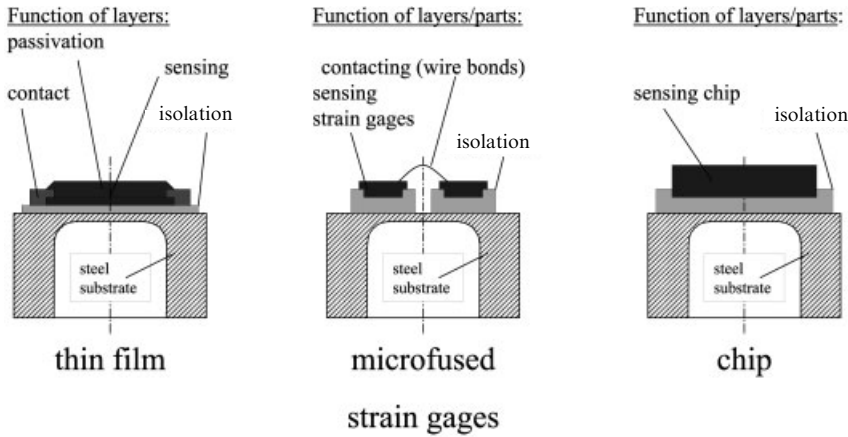


Fig. 7.4.5 Technologies for strain gages on steel substrates (not to scale)

um evaporation and sputtering processes are used, which are quite similar to those used for manufacturing integrated circuits (ICs). Depending on the production strategy more than 100 parts can be handled in one 'steel wafer' or jig. The thickness of the thin-film layers produced on the diaphragm is some micrometers. Since these layers are thin compared with the diaphragm (which is a few 100 μm thick), they do not influence the deformation of the diaphragm. A glass isolation layer is used for electrical insulation between the strain gages and the steel substrate (Fig. 7.4.5). The isolation layer has to allow a strain of about 0.15% and have very good adhesion to the steel and a high resistance (Fig. 7.4.4). A sensing thin film is deposited on the isolation layer. This layer is etched to receive the structure of the strain gages after a photolithographic process. This layer must have a low temperature coefficient because of the high influence on the sensor characteristic. To reach a high sensitivity of the sensing element the strain gages must be small, to detect only the maximums of tensile and compressive strain on the diaphragm surface. From the production point of view the layout of the strain gages should be wide enough for a stable production process. A second conducting layer is required to form the electrical connection of the strain gages and the contact pads. To protect the thin-film layers against environmental and mechanical effects a passivation layer of Si_xN_y or SiO_x is used.

Before depositing the passivation layer, the Wheatstone bridge of high accuracy sensors has to be trimmed to zero by laser cuts in the sensing layer.

The material of the sensing layer strongly influences the characteristics of the pressure sensor and the difficulty of the calibration process. Fig. 7.4.6 shows the change in sensitivity of two different sensing layer materials (poly-Si and thin-film NiCr metal) against temperature. Poly-Si shows a nonlinear dependence on temperature, whereas NiCr is quite linear. Also the variation of the sensitivity is much lower for thin-film NiCr. Both high linearity and low variation of sensitivity contribute to a simple and therefore cost-effective calibration concept for high-pres-

**Poly-Si thin film sensor element
change of output span related
to output span at room temperature**

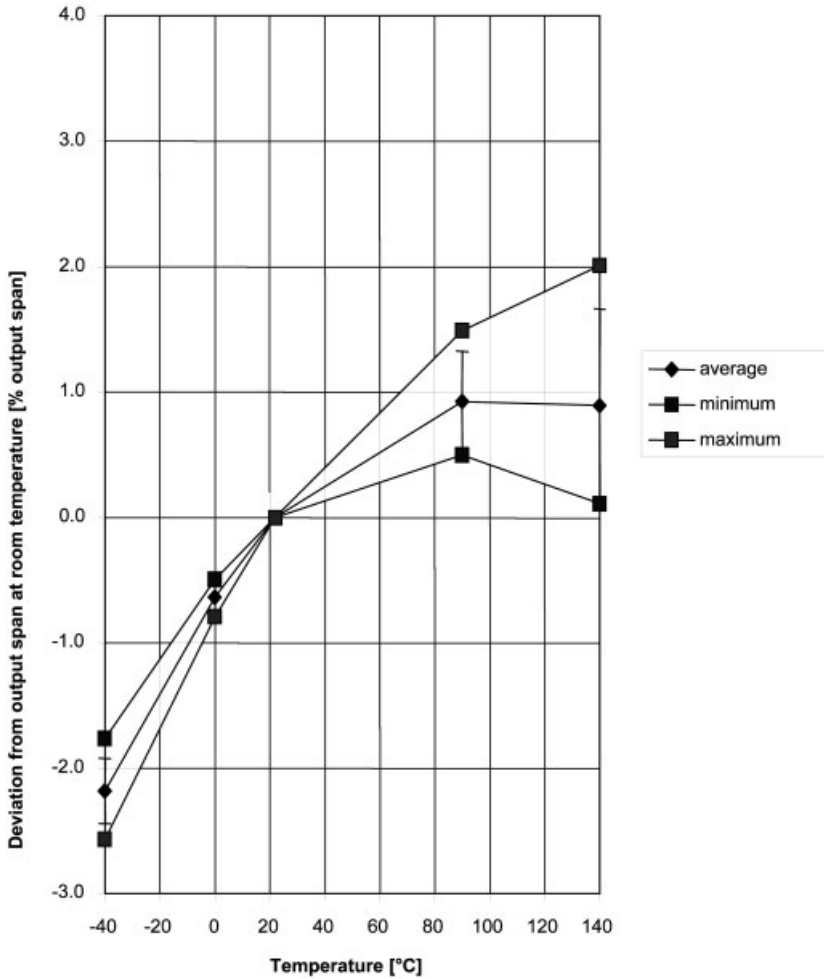


Fig. 7.4.6a

Fig. 7.4.6 a and b Comparison of output span of poly-Si and metal (NiCr) thin-film sensor elements

sure sensors with a metallic thin film such as NiCr. Furthermore, such sensors benefit from the general advantages of thin-film technology [3, 4].

- High accuracy.
- Low temperature dependency of offset and output span.
- Best lifetime stability because of cheap integrated pre-aging during the production process.
- Possibility of miniaturization and cost reduction.

**Metal (NiCr) thin film sensor element
change of output span related
to output span at room temperature**

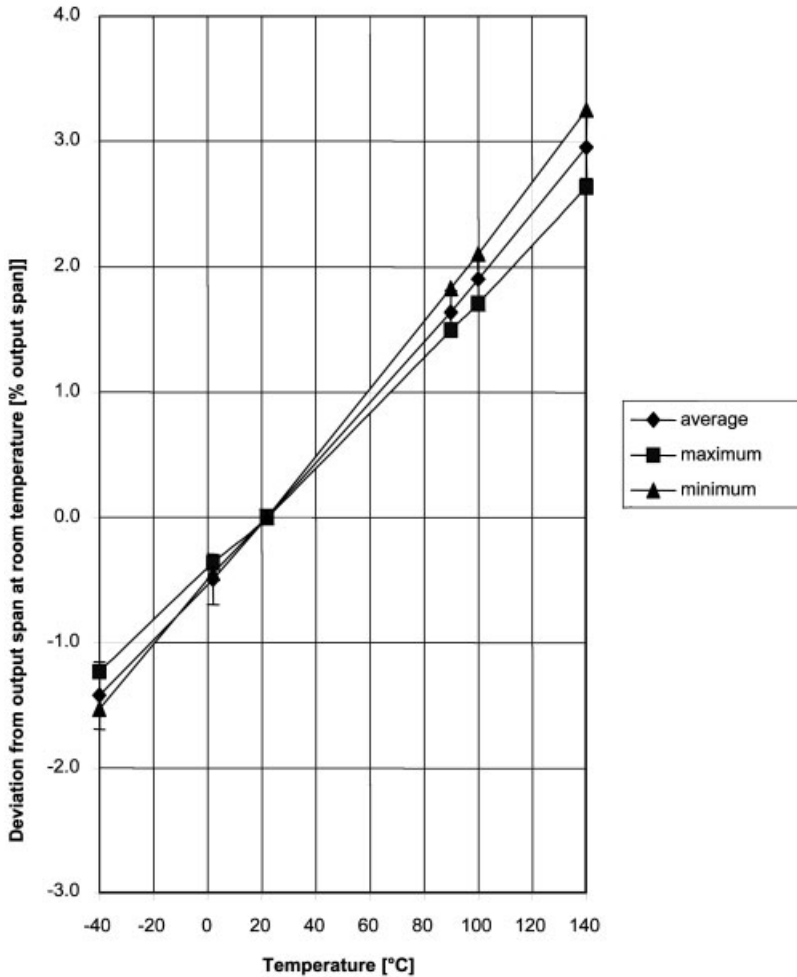


Fig. 7.4.6b

Microfused Strain Gages on Steel Diaphragm

Silicon strain gages can be placed on steel diaphragms by fusing them in low melting point thick-film glass [5]. In this case a higher surface roughness of the diaphragm is allowed, since the glass has a high thickness compared to thin-film isolation layers. Therefore no polishing is necessary and a sandblasted surface is sufficient. Using silicon materials with high temperature coefficients makes calibration difficult. Since the placing of the silicon strain gages is difficult, downsizing of the

sensing element is limited. Furthermore, pre-aging is necessary to get lifetime stability comparable with thin-film sensors.

Single Chip on Steel Diaphragm

In addition to thin-film and microfused strain gage technologies, sensors can be made using a Si chip in which all the strain gages of the Wheatstone bridge are integrated. The thick-film glass bonding process (chip to steel diaphragm) limits the sensor accuracy. An advantage of this measuring principle is the low specification on the surface finish of the diaphragm. On the other hand, the handling of the chip and its high costs are disadvantages of this design. A second chip to do the amplification, calibration, and compensation is still needed. Deformations in the sensing chip because of diaphragm elongation prevent it being used as a sensor ASIC.

7.4.3.2 High-Pressure Sensor Packages

To meet the automotive requirements the design of the high-pressure sensor plays a decisive role with respect to the measurement task. Fig. 7.4.7 shows a cross section of a pressure sensor for gasoline direct injection systems.

The thin-film element is welded to the thread part by electron or laser beam welding. In order to have little influence on the function of the diaphragm, the mechanical and thermal stress during welding has to be minimized. A hexagon is welded to the thread part, to be able to apply the torque needed for tightening the sensor to the fuel rail. The signal of the Wheatstone bridge is conducted to a contact carrier by wire bonds. From there it goes to the hybrid or circuit board, where the ASIC is placed for signal amplification, calibration, and compensation of the characteristic. Contacting from the circuit to the connector terminals is done by springs, which balance the mounting tolerance ranges of all parts. To fulfill the vi-

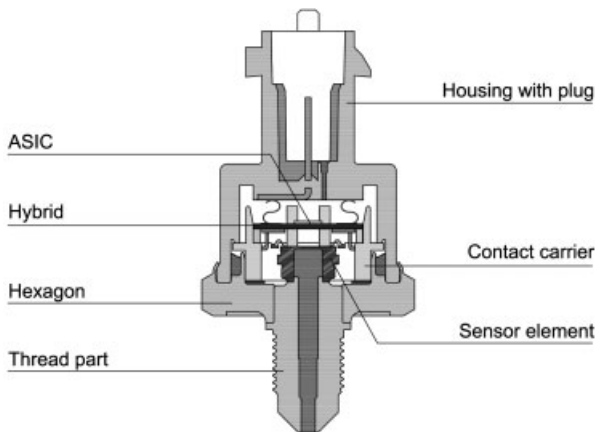


Fig. 7.4.7 Cross section of sensor package for metal thin-film sensor used for gasoline direct injection

bration requirements the housing with its connector are edge bordered to have a positive engagement between housing and hexagon. To protect the sensor from environmental and mechanical effects the housing is adhesive bonded to the hexagon.

Every production process that applies high mechanical stress to the sensor could result in a change of zero-offset. Hence the calibration of the sensor should be done as late as possible to have a chance to compensate those influences. The sensor shown in Fig. 7.4.7 is calibrated when completely mounted by using one of the connector terminals for communication with the ASIC. Doing this enables one to compensate for production influences and to prevent additional conductor tracks and their potential influence on electromagnetic resistance of the sensor.

Sensors used in brake applications are often highly adapted to hydraulic units, so that their package designs depend strongly on their integration environment.

7.4.3.3 Oil-Filled Sensors

Another design used for pressures up to 100 bar is an oil-filled single-chip sensor [6]. Oil transforms the pressure applied from a steel membrane to a single chip immersed in the oil. The advantage of this measuring principle is the low-cost of production. On the other hand, the difficulty of ensuring absolute tightness of the oil filling under automotive environmental conditions makes it hard to achieve high accuracy over the lifetime. Also, the optimization of the electromagnetic resistance is complicated, since the circuit components have to be immersed in the oil.

7.4.3.4 Endurance Testing Results

Fig. 7.4.8 shows the results of endurance testing of a metal thin-film high-pressure sensor with the design shown in Fig. 7.4.7. Typical deviations from the ideal characteristic can be seen. The sensor has been tested in a gasoline direct injection car for 162 000 km. The deviation is shown for pressures up to 140 bar and temperatures between -40 and 140°C . Hysteresis can be seen for increasing and decreasing pressures at each measuring temperature. The maximum deviation of about 0.3% FSD demonstrates the long-term stability of the sensor package design and the thin-film technology.

7.4.4

Summary and Outlook

There are several automotive braking and direct injection applications requiring high-pressure sensors to detect important system parameters. An overview of requirements and measuring principles was given here. In particular, a sensor package design using metal thin-film technology on a steel diaphragm and an optimized calibration process done with a completely mounted sensor has been endurance tested in more than 100 vehicles with very good results.

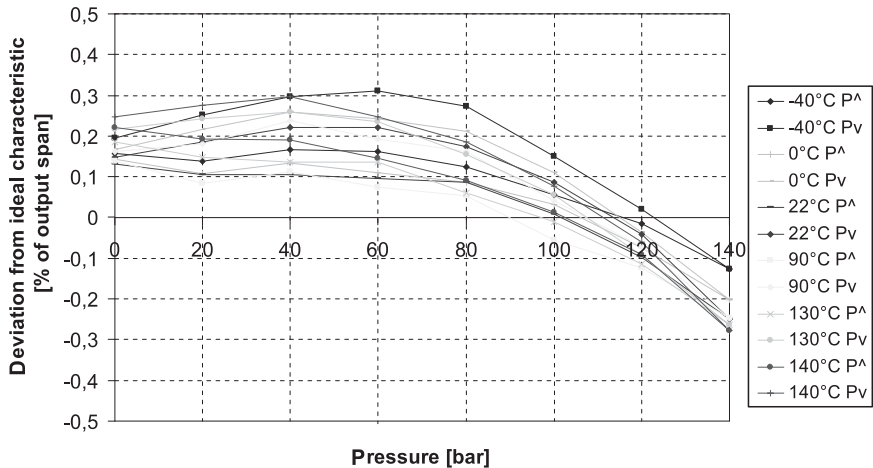


Fig. 7.4.8 Small deviations of sensor endurance tested for 162 000 km

To reach future requirements for brake and direct injection systems requires high-pressure sensors that have the following properties.

- Higher pressures and improved accuracy.
- Better vibration resistance.
- Lower costs.

7.4.5

Acknowledgments

We thank Andre Kretschmann for critical reading of the manuscript.

7.4.6

References

- 1 BAUER, H. (ed.), *Automotive Handbook*, 5th edn, Robert Bosch GmbH, Stuttgart, 2000, p. 507, 673.
- 2 ACHENBACH, W., STOLL, U., *Weltneuheit bei den Bremsen: Die Sensotronic Brake Control (SBC)*, *Automobiltechnische Zeitung (ATZ)/MTZ-Sonderausgabe*, 2001, 10 Der neue Mercedes SL.
- 3 TRÄNKLER, H.-R., OBERMEIER, E. (eds.), *Sensortechnik*, Springer, Heidelberg 1998, p. 343.
- 4 BONFIG, K.W., BARTZ, W.J., WOLF, J. (eds.), *Sensoren*, 2nd edn, Meßaufnehmer, Expert Verlag, Renningen, 1988, p. 86.
- 5 SIDHU, S.K. *Packaging Very High Pressure Transducers from Common Rail Diesel Injection Systems*, SAE Technical Paper 982020.
- 6 OTAKE, S., ONDADA, M., NAGASE, K., *Automotive High Pressure Sensor*, SAE Technical Paper 980271.

7.5

Temperature Sensors

ALDO BOJARSKI and WERNER FICHTE

7.5.1

Introduction

Temperature sensors have a long history in automotive applications. Monitoring the engine temperature was one of the first sensor applications in vehicles. Today, temperature is one of the most measured physical characteristics in modern vehicles. Applications include monitoring and control of engine parameters, coolant return temperature measurement for fan-speed control, gear-box temperature monitoring, fuel temperature measurement, intake-air temperature measurement, outside-air temperature measurement, passenger-compartment temperature measurement, climate control unit parameter measurement, exhaust-gas temperature measurement before and after catalyst, and measurement of compressed-air temperature after turbo charger or compressor. The sensors are required for diagnostics, to control the performance of the powertrain, or to provide data for emission-control systems and convenience equipment. A modern mid-range vehicle today has 5–10 temperature sensors.

7.5.2

Technologies

The temperature range that needs to be measured can be split into two major ranges.

The first, -40 to 170 °C, covers the requirements in the passenger compartment, climate control system, and powertrain applications. The sensors are mostly based on negative thermal coefficient resistors (NTCs). Other technologies are capable of meeting the technical requirements in this temperature range (PTCs, Pt 1000 resistors, integrated circuits with linear outputs (KTY-series) or copper or nickel resistors) but none of these has a major market share so far.

The second range covers exhaust temperature measurements from -40 to 760 °C for diesel engines and even up to 1200 °C for gasoline engines. This market is currently driven by legislative emission requirements worldwide like ULEV (ultra-low emission vehicles), OBD (on-board diagnosis), and EURO4. Improved emission systems need temperature measurements in the exhaust system before and after the catalyst.

7.5.2.1 Sensors for -40 to 170 °C

Sensing Element

For the range up to 170 °C, most sensors are based on NTCs. These resistors are made of powder mixtures of metal oxides such as iron, zinc, cobalt, manganese, copper, and nickel. These powder mixtures are formed into discs or wafers and sintered at high temperatures between 1000 and 1400 °C. The cured product is a ceramic material.

The discs or wafers are contacted by a silver paste on two opposite surfaces and are fitted with leads. While discs are produced to the final size, wafers are cut to a final chip size depending on the resistance of the wafer. Usually a coating and special aging process follows to ensure long-term stability. The innovative waver technology was the breakthrough to cost-effective mass production of accurate NTCs with defined properties.

NTCs are available as SMD resistors (surface mounted devices), as well as in special packages like glass-coated or fitted to lead frames. The size of the NTC can be as small as 1 mm×0.5 mm×0.6 mm (and smaller) or there are discs of 25 mm diameter [1].

The curve of resistance against temperature varies from different NTC manufacturers. Some NTCs are available from several manufacturers. Others are special and can not be matched exactly over the entire temperature range by other manufacturers. This should be considered in advance to avoid excluding suppliers by the specified curve or tolerances. An example of an NTC curve is given in Tab. 7.5.1 and Fig. 7.5.1.

These curves are very nonlinear; note also that the sensitivity is between 2%/K and 6%/K and is much greater than the sensitivity of metals or silicon temperature sensors.

Later we will see that the nonlinearity is not a problem in use. It is not possible here to go into all the details of NTC theory for temperature coefficient alpha, beta value, self-heating effects, thermal time constant, thermal dissipation constant, stability, and reliability. Good information can be found elsewhere [1, 2] and in NTC manufacturers' handbooks, catalogs, or web pages [3–10].

Tab. 7.5.1 Resistor values of a NTC

<i>Temperature</i> (°C)	<i>Resistance</i> (Ω)	<i>Temperature</i> (°C)	<i>Resistance</i> (Ω)
-40	75 501	40	1199
-35	54 495	45	983.4
-30	39 764	50	810.9
-25	29 319	55	672.2
-20	21 833	60	560.1
-15	16 414	65	468.9
-10	12 452	70	394.5
-5	9 529	75	333.4
0	7 353	80	282.9
5	5 719	85	241.2
10	4 482	90	206.4
15	3 539	95	177.3
20	2 814	100	152.9
25	2 252	105	132.4
30	1 814	110	114.9
35	1 471	115	100.2
40	1 199	120	87.6

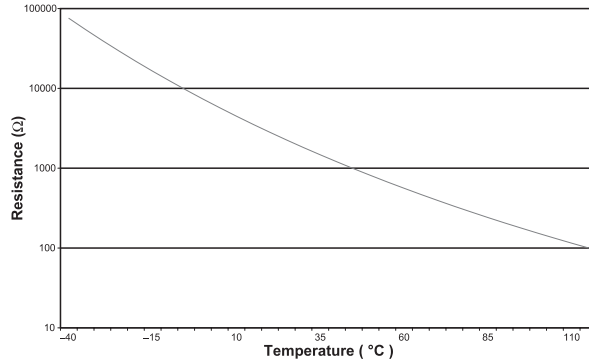


Fig. 7.5.1 Variation of negative thermal coefficient resistance with temperature

The application engineer should notice that NTCs can be point or curve matched. A point-matched NTC has a guaranteed tolerance at one temperature point, which is usually 25°C, but upon agreement other points can be matched by most manufacturers.

Away from this point the tolerance of resistance may increase with distance from the trimmed point. These thermistors are preferred if one point of temperature is of major interest (e.g., 100°C, 0°C). When a temperature range needs to be measured, curve-matched thermistors should be used. They allow a measurement accuracy between 0.2 and 0.5 K over a wide range of temperature [1].

NTCs are characterized by their 25°C resistance in Ω and the B value. The B value is determined by the ceramic material and describes the gradient of the resistive curve over temperature. Therefore, if the resistances R_1 at temperature T_1 and R_2 at temperature T_2 are measured, the B value is

$$B_{T_1/T_2} = \frac{T_2 \cdot T_1}{T_2 - T_1} \cdot \ln \frac{R_1}{R_2} \quad (1)$$

where

T_1 = temperature 1 in Kelvin

T_2 = temperature 2 in Kelvin

R_1 = NTC resistance at temperature T_1 in Ω

R_2 = NTC resistance at temperature T_2 in Ω

A common value for T_1 is 25°C (298.15 K) and T_2 is often 100, 75, or 50°C. For the NTC shown above we calculate a $B_{25/100}$ value of

$$B_{25/100} = \frac{(100 + 273.15) \cdot (25 + 273.15)}{(100 + 273.15) - (25 + 273.15)} \cdot \ln \frac{2252}{152.9} \quad (2)$$

$$B_{25/100} = 1483.4 \cdot 2.69 \quad (3)$$

$$B_{25/100} = 3990 \quad (4)$$

The B value is typically between 2000 and about 5000. Because of nonlinearity, the $B_{25/50}$, $B_{25/75}$, and $B_{25/100}$ values of the same resistor are different. NTCs are available for temperatures up to 350 or even 500°C.

Other sensing elements like PTCs, semiconductors, diodes, or complete integrated circuits are found in a few applications, but NTCs dominate the automotive sector. This is because of their robustness, low cost, high gradient of transferring temperature changes into resistive changes, no electromagnetic induction problems, good ESD (electrostatic discharge) properties, and size that can be adjusted for quick response times or high power-dissipation.

Another advantage is the very simple operating circuit. In the applications below it can be seen that just one or two resistors at the control device site are enough to get a signal with 70% span or more of the supply voltage. This is why the other sensing elements and technologies mentioned are inferior to NTCs in general automotive applications.

Automotive Designs

In most automotive applications the temperature of a gas, oil, water, or coolant is measured. This determines the design of most automotive sensors. An example is shown in Fig. 7.5.2.

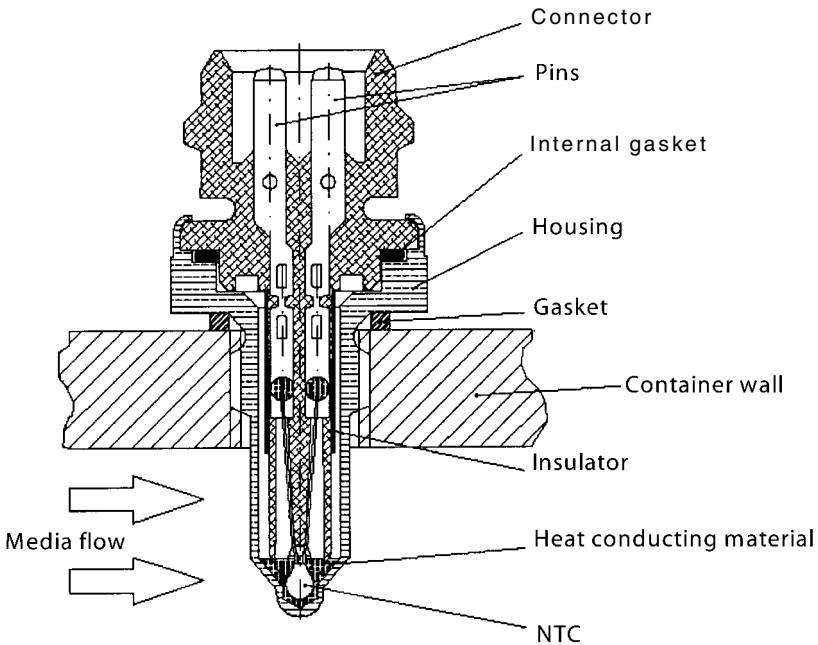


Fig. 7.5.2 Diagram of automotive sensor

Tab. 7.5.2 Trends in NTC-based temperature sensors

Housing	<p>Brass is the dominant material offering good durability for most fluid applications. Plated steel, aluminum alloys, and stainless steel can be found as well.</p> <p>Air applications use preferred plastic covers or just epoxy resin coatings.</p> <p>There are even sensors with no additional NTC coatings, other than the NTC manufacturer's covering.</p> <p>'Snap-in' assembly designs are getting more and more popular for fluid as well as gas sensors.</p>
Connector	<p>Depends on original equipment manufacturer's philosophy.</p> <p>Sensors with thread often specify the hexagonal size upon the size of the connector, to allow for assembly with nut drivers.</p>
Connector pins	<p>Most sensor use two pins. Sensors with ground contact through the housing are found as well. Pins are often tin plated some are gold plated.</p>
NTC	<p>Many sensors use NTCs with leads. The leads are soldered or welded to the connector pins. For 25 °C values see chart.</p> <p>Lead frame NTCs reduce cost, but require high volume production</p>
Heat conducting filler	<p>Traditionally silicon-oil based pastes are used. They tend to 'bleed out' and loose heat conductivity over time.</p> <p>New sensors use filled silicon or epoxies</p>
Gasket	<p>Aluminum or copper are standard materials. The gasket is often fitted to the sensor behind the thread at the sensor manufacturers site. O-rings are common too</p>

The classic sensor consists of a housing, connector, NTC, heat-conducting material, and gasket. So far there is no standard in the automotive industry for temperature sensors. In the market we find an endless variety of threads, NTC curves, connectors or cable exit with connector, gasket materials, and housing materials. There are certain trends though (Tab. 7.5.2). In a market survey 35 temperature sensors used by European manufacturers for underhood applications

Tab. 7.5.3 Specification parameters for temperature sensors

Parameter	Typical values
Surrounding media to be measured	Oil, water, outside air, engine coolant
Temperature range	-40 to 125 °C, -55 to 170 °C
Resistance value at 25 °C/NTC curve	Typical value 1–10 kΩ
Measuring accuracy	Point or curve: 5, 3, 1% or 3, 1, 0.5 K
Dimensions, fit, connector	–
Response time	Depends on media, speed of media, sensor design and NTC properties. For values see application examples below

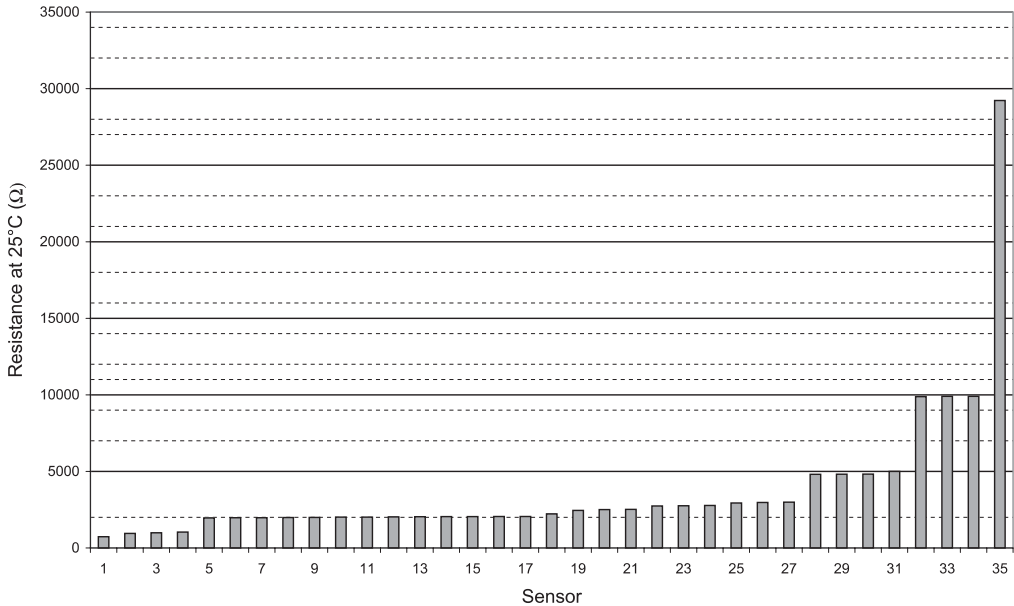


Fig. 7.5.3 Resistances of temperature sensors

were analyzed. All of them used NTCs. The resistance of these sensors at 25 °C is shown in Fig. 7.5.3. To specify an automotive sensor, only a few key parameters are needed (Tab. 7.5.3).

Signal Formation

The final signal needed for further processing in an electronic control unit (ECU) is usually a voltage. This voltage is then digitized by an A/D converter (ADC) for further processing. The voltage span should be as large as possible to utilize the given range and resolution of the ADC. Most ADCs have a operating voltage of 5 V. The digitized signal range should be between 0.5 and 4.5 V.

Using NTC temperature sensors, the circuit to generate a voltage is a simple pull-up resistor or (for further linearization and stretching of the signal range) an additional parallel resistor (Fig. 7.5.4). The resulting signal voltages are shown in Fig. 7.5.5, Fig. 7.5.6, and Fig. 7.5.7.

The plot shows that the signal has a high gradient at 2.5 V of approximately 50 mV/°C and a span of more than 4 V. Nonlinearities are compensated in the subsequent digital processing if required. With a 10-bit ADC resolutions of 1 °C and better are possible.

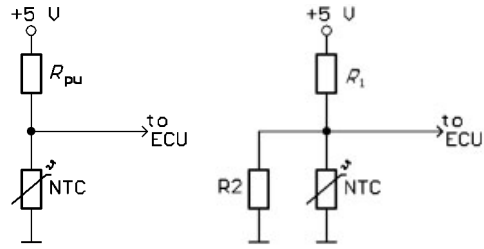


Fig. 7.5.4 Circuits for NTC operation

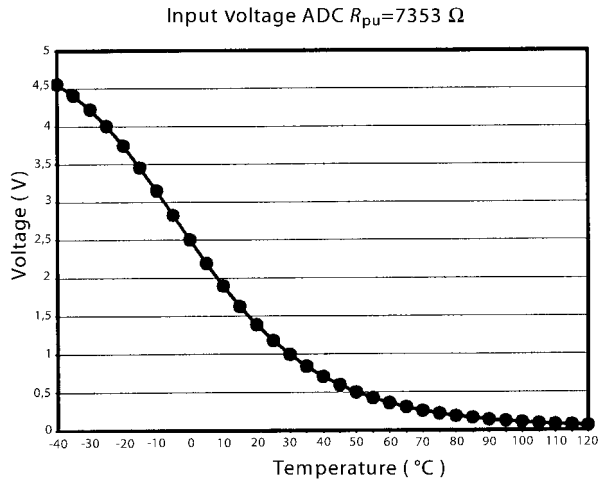


Fig. 7.5.5 Curve input voltage NTC with pull-up resistor

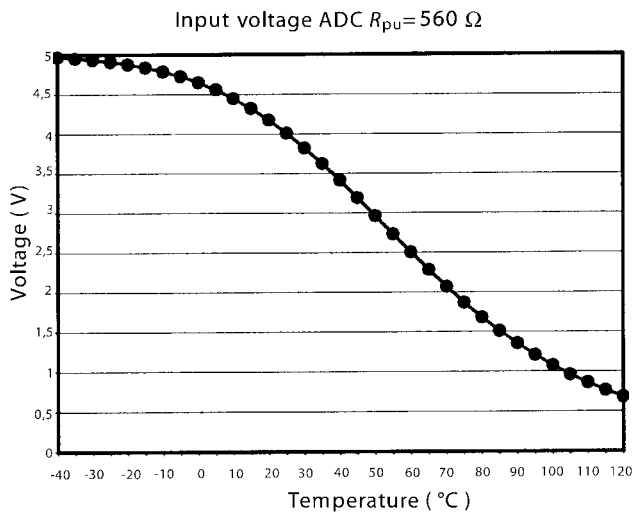


Fig. 7.5.6 Curve input voltage NTC with pull-up resistor

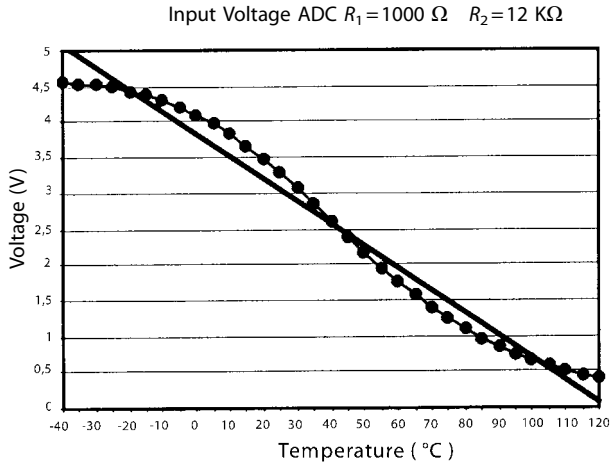


Fig. 7.5.7 Curve input voltage NTC with pull-up and parallel resistor

7.5.2.2 Sensors for Higher Temperatures, up to 1100°C

Sensing Element

For the temperature range up to 800 or even 1100°C sensors are not yet used in high volumes. There are several possible technologies. The first is to qualify special NTCs for these temperatures [3], the second is to use thermocouples, and the third is platinum resistors: there are others. The battle has just started and so far there is no clear favorite in the market.

Thermocouples use the physical effect of a voltage generated between conductors of different materials when their welding point is at a higher temperature than the open ends of the wires. The voltage generated is proportional to the temperature difference between the hot and cold sides. For proper operation these sensors require thermocouple extension wires. To measure the absolute temperature at the hot side, another temperature measurement for the cold side is required. The signal is a voltage of $\mu\text{V}/^\circ\text{C}$. It can be increased by cascading elements. A typical signal is 60 mV/1000°C. To deal with this signal, high-resolution ADCs are required in the ECU. There are costs that need to be considered, but thermocouples are used by some manufacturers.

Some NTCs are rated for temperatures up to 750°C continuously, or up to 1000°C for 50 h [3]. The major advantage is compatibility of the system with existing temperature sensors in low-temperature automotive applications, so the signal is easy to process in the control unit. This is a big plus for these sensing elements. There are good chances that NTCs will be used in this temperature range. Most other things will apply as described above for these high-temperature sensors as well. Several NTC manufacturers are currently developing high-temperature NTCs.

Platinum sensors are currently useful because of their potential accuracy and long-term stability in this temperature range. These sensing elements use the

thermal coefficient of Pt metal (Tab. 7.5.4). According to DIN EN 60751 the curve of Pt100 sensors for temperatures above 0 °C is

$$R(T) = R_0 \cdot (1 + AT + BT^2) \quad (5)$$

where

R_0 = resistance at temperature 0 °C = 100 Ω

$A = 3.90802 \times 10^{-3} \text{ } ^\circ\text{C}^{-1}$

$B = -5.802 \times 10^{-7} \text{ } ^\circ\text{C}^{-2}$

T = temperature in °C

As an example, a Pt100 resistor has 138.5 Ω at 100 °C. Based upon the linear calculated thermal coefficient at the 100 °C and 0 °C points

$$R(T) = R_0 \cdot (1 + aT) \quad (6)$$

the resulting 'linear' factor a is often given as

$$a = 0.00385 \text{ K}^{-1} \quad (7)$$

In the past Pt100 resistors were standard sensing elements. The technology was based on wire-wound resistors. Because Pt is an expensive material and the mass of metal used will affect the response time of a sensor, a Pt100 was the best compromise between resistance and cost. Higher resistances will increase the voltage output and allow lower currents. The goal is to have a high resistance with a low mass to get a maximum voltage signal at a low driving current.

To reduce the amount of Pt used and in order to gain more signal, new technologies are used today to manufacture Pt sensing elements. Most Pt sensing elements are now in the form of thin-film resistors. This technology can produce

Tab. 7.5.4 Resistances for Pt200

Temperature (°C)	Resistance (Ω)	Temperature (°C)	Resistance (Ω)
0	200.0	600	627.2
50	238.8	650	659.0
100	277.0	700	690.3
150	314.6	750	720.9
200	351.7	800	751.0
250	388.1	850	780.5
300	424.0	900	809.5
350	459.3	950	837.8
400	494.1	1000	865.6
450	528.2	1050	892.8
500	561.8	1100	919.4

higher resistances and save Pt material as well. First market applications use Pt200 or Pt1000 sensors [10].

Even Pt sensors are challenged up to the limit at 1000°C. The ceramic chip material becomes conductive itself, bending the curve of the Pt element above 600°C. The Pt itself tends to diffuse away and to be affected by the surrounding materials, so it needs to be covered. Doping effects change the thermal coefficient of the Pt over time.

Different suppliers have different solutions and patents to stabilize their sensing elements. One of them is to use special doped Pt with the pay off of changed A and B coefficients. One manufacturer gives [9,10]

$$A = 3.829 \times 10^{-3} \text{ } ^\circ\text{C}^{-1}$$

$$B = -5.85 \times 10^{-7} \text{ } ^\circ\text{C}^{-2}$$

$$\alpha = 0.00375 \text{ K}^{-1}$$

Other manufacturers publish values of α between 0.00350 and 0.00380. This is not a critical loss of sensitivity, but in terms of interchangeable sensor suppliers this is important. The accuracy of the sensors itself is given as $\pm 2\text{--}3 \text{ K}$ at 0°C and $\pm 10\text{--}15 \text{ K}$ at 1000°C [10].

Automotive Designs

The automotive application of high-temperature sensors is measuring the temperature of exhaust gas. This means an aggressive atmosphere that might be reductive or oxidizing, contain sulfur, nitrogen, hydrogen, oxygen, and all kinds of compounds thereof: CO_x , NO_x , H_2O , SO_x , HC_x . This, in conjunction with the temperature, determines materials for housing and fittings; a design is shown in Fig. 7.5.8.

There are also designs in which the sensing element is not covered by the stainless steel housing, but is in direct contact with the gas flow, to achieve minimal response times.

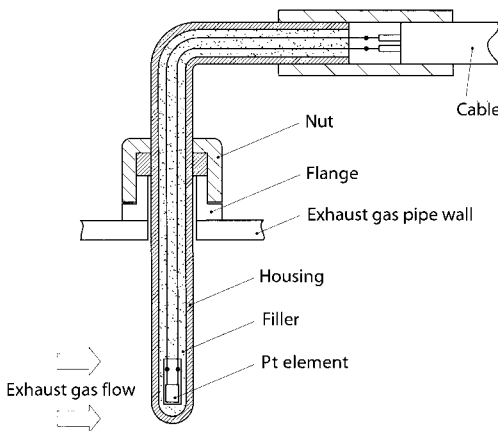


Fig. 7.5.8 Exhaust-gas temperature sensor

Signal Formation

Using Pt200 temperature sensors, the simplest circuit for generating a voltage is again a pull-up resistor. For better linearity and higher signal yield a current source can be used (Fig. 7.5.9).

With a pull-up resistor of 2 kΩ to +5 V the output voltage span between 0 and 1000 °C is 1.05 V. The sensitivity at 0 °C is 1.5 mV/K, decreasing to 0.67 mV/K at 1000 °C (Fig. 7.5.10). The more expensive version with a 2 mA current source generates 1.33 V between 0 and 1000 °C with a sensitivity of 1.5 mV/K at 0 °C and 1.1 mV/K at 1000 °C.

Using a 10-bit ADC (0–5 V) only 20–26% of the possible range is utilized, which compares with an 8-bit converter. The resulting resolution is 4–7 °C per digit. To improve the resolution, additional amplifiers with high requirements for offset and amplification or high-resolution ADCs are required. In these calculations other failures are not included: there may be failures caused by resistance of the connection cable or connectors and their thermal properties.

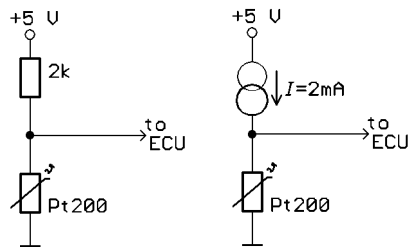


Fig. 7.5.9 Circuit for Pt200 application

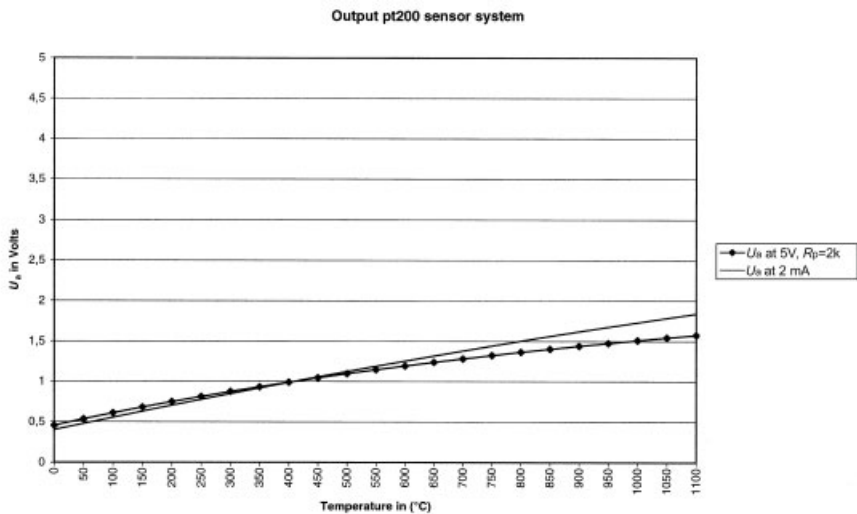


Fig. 7.5.10 Curve input voltage for Pt200

7.5.3

Examples of Applications

The following figures give some examples of applications of temperature sensors from AB Elektronik Sachsen GmbH.



Fig. 7.5.11 Standard sensor with low response time and preassembled aluminum gasket for sensing oil and water temperature up to 170°C in engine and gearbox



Fig. 7.5.12 Sensor with brass housing on media side; two isolated NTC circuits with different curves for different purposes for sensing coolant temperature



Fig. 7.5.13 Sensor with brass housing on media side, front side O-ring for pipe assembly for sensing fuel temperature after fuel pump



Fig. 7.5.14 Sensor with brass housing and front side O-ring; snap-in design for coolant by-pass for exhaust-gas catalyst retrofit control unit



Fig. 7.5.15 Sensor with brass housing for sensing oil, fuel, and coolant temperature in commercial vehicles



Fig. 7.5.16 Sensor with reduced brass housing; front side O-ring for sensing coolant and fuel temperature in passenger cars

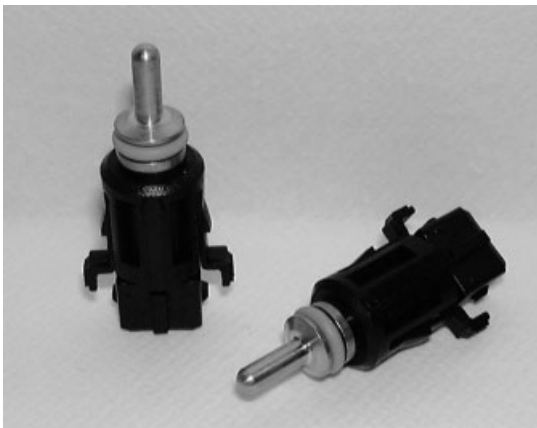


Fig. 7.5.17 Sensor with brass housing, snap-in design, snap-in locked with female connector for sensing coolant return temperature, fan-speed control



Fig. 7.5.18 Sensor with brass housing on media side, front side O-ring, cable exit, flange assembly for sensing fuel and coolant temperature in commercial vehicles

Fig. 7.5.19 Sensor with plastic housing, NTC with direct air contact protected by cage, snap-in design for quick assembly for sensing intake-air temperature in passenger cars



Fig. 7.5.20 Sensor with metal housing, NTC with direct air contact, protected by cage for sensing intake-air temperature in passenger cars



Fig. 7.5.21 Sensor with plastic nozzle, metal housing for sensing intake-air temperature in commercial vehicles





Fig. 7.5.22 Sensor with plastic housing, snap-in design, NTC directly exposed to air for sensing intake-air temperature in passenger cars

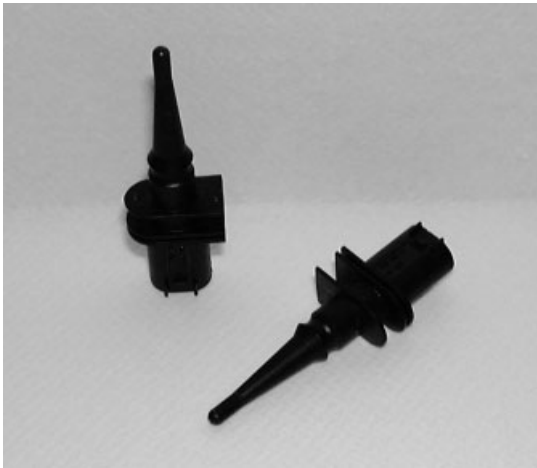


Fig. 7.5.23 Sensor with closed plastic housing for sensing outside-air temperature in passenger cars, ice warning



Fig. 7.5.24 Sensor with plastic housing, snap-in design, NTC directly exposed to air, heat radiation shield, additional port for air flow for sensing passenger-compartment air temperature

Fig. 7.5.25 Two sensors for measurement of metal temperature: upper sensor for battery temperature measurement, lower sensor for cylinder head temperature in air-cooled engines



Fig. 7.5.26 Open NTC, plastic housing for climate control unit and outside temperature for passenger cars



7.5.4

Summary and Outlook

Temperature sensors based on NTCs have established a dominant position in automotive applications up to 170°C for more than 50 years. The production technology has improved over the last decade to increase accuracy and long-term stability. Published accuracy figures of 3–5% due to manufacturing yields are out of date. In the overall mix of cost, robustness, signal yield, accuracy, and stability NTC-based sensors fulfill all automotive requirements and measure temperatures with an accuracy and resolution of less than 1 K. And this at very reasonable cost for sensor, connecting cable and connectors, and signal measuring unit. Even though other technologies have improved as well, none is expected to take the place of the NTC.

The missing ‘standard curve’ for all sensors has not been established in the market yet, but is not a real problem either. Several manufacturers have established internal or ‘house standard’ curves. Standard packages have been established within the manufacturers as well, but different connector systems are still an obstacle to standardization.

One current market trend is towards smaller sensors with less weight and lower cost. The sensor should be fixable as quickly and simply as possible. At the

same time, response-time requirements are steadily increasing. This is possible and supported by smaller NTCs and adjusted designs.

In the field of sensors for temperatures up to 800 or even 1100 °C the market is still diverse, but a strong trend towards Pt sensing elements exists because these elements have advantages over thermocouples and a long history and background in this temperature range. Technology and customer demands are pushing towards Pt200 or even Pt1000 elements. It is expected that one of these elements will have the best chance for high-temperature applications.

7.5.5

References

- | | |
|---|--|
| <ol style="list-style-type: none"> 1 EPCOS AG, <i>NTC Thermistors</i>, data book 2002. 2 www.betatherm.com/therm_theory.html 3 www.thermometrics.com 4 www.epcos.com 5 www.hdk.co.jp 6 www.murata.com/thermis/ntc/faq/ 7 H.-R. TRÄNKLER, E. OBERMEIER, <i>Sensortechnik</i>, Springer, Heidelberg 1998. | <ol style="list-style-type: none"> 8 heraeus sensor nite/CiS Institut für Mikrosensorik gGmbH
www.cismst.de/deutsch/infocenter/report/rep54/platin.htm 9 S. CARSTENS, <i>Auto & Elektronik</i>, 2001, 4, 56–60. 10 www.heraeus-sensor-nite.de |
|---|--|

7.6

Mass-Flow Sensors

UWE KONZELMANN

7.6.1

Introduction

To meet the requirements of fuel consumption and emissions, modern fuel-injection systems require precise metering of the intake air mass. There are two main approaches: direct and indirect measurement. In the indirect method the density of the air upstream of the intake valves is determined by measuring pressure and temperature, volumetric flow is determined from engine speed and displacement, including the volumetric efficiency of the engine. With direct measurement the disadvantages of the indirect method (adulteration due to exhaust gas recirculation and changes in the volumetric efficiency) can be avoided.

Air-flow metering is used in gasoline and diesel engines. In passenger cars more than 50% use air-flow metering; engine displacement has a wide range of 1–6 L.

For metering air mass various physical principles can be used. The first automotive application in 1972 used a vane type air-flow meter [1]. The signal from a vane meter has to be corrected for air density variations, so the thermal measurement principle was chosen for the next generation of air-flow meters.

The first generation of thermal air-flow meters used a thin platinum wire as a sensing element in the main air passage of the metering duct [2]. Alternative developments used a platinum wire wound around a small ceramic or glass cylinder as a sensing element, located in a bypass air passage [3]. For the second generation of thermal air-flow meters film-type sensors were used [4]. Platinum resistors were printed on a glass or ceramic substrate. Sensor elements based on a silicon chip are used for the third generation of thermal air-flow meters [5], and large-scale production of the third generation started in 1996.

7.6.2

Technologies

The present state-of-the-art is to use a thin-film sensor for thermal air-flow meters. The thin film is deposited on a planar or cylindrical substrate. A planar substrate allows for resistor structures, which make complex measurement tasks possible, for example back-flow detection. By exploiting microsilicon technology the resistors can be located on a thin diaphragm, which results in very short response times.

7.6.2.1 Microsilicon Sensor Element

A microsilicon sensor element is produced by deposition of the materials forming the diaphragm and the resistors onto a silicon wafer. By etching the reverse side of the wafer the silicon material is removed to leave the diaphragm. The schematic layout of the Bosch microsilicon sensor element is shown in Fig. 7.6.1 and Fig. 7.6.2. The diaphragm consists of layers of silicon oxide and silicon nitride. The resistors are made of platinum. The layout of the resistors forms a heating zone with a heater and a heater-control resistor. The heater is controlled by a heater-control circuit to maintain a temperature above ambient. The excess temperature depends on the ambient temperature, which is detected by a platinum resistor on the silicon chip outside the diaphragm. At an ambient temperature of 20°C the excess temperature is approximately 180 K. The voltage to drive the heater can be used as a measurement of air-flow rate.

Additional resistors are located upstream and downstream of the heating zone. Without air flow the temperature upstream and downstream of the heating zone

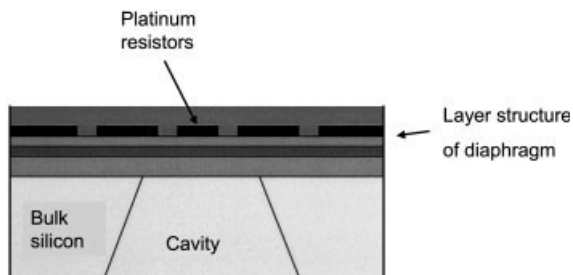


Fig. 7.6.1 Structure of the Bosch microsilicon sensor element for air flow

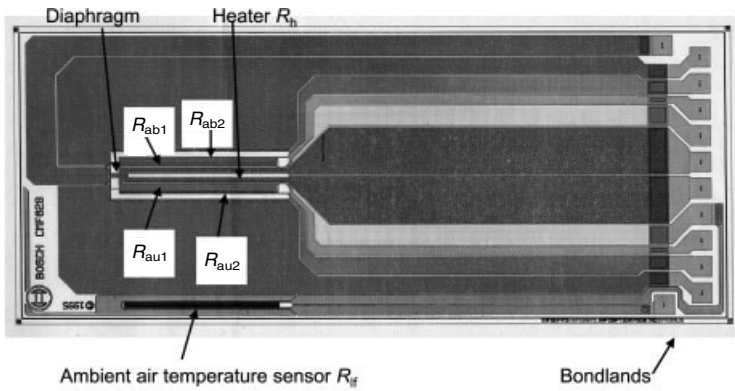


Fig. 7.6.2 Electrical layout of the Bosch microsilicon sensor element for air flow. The different resistors are shown in different colors. The air mass signal is generated using the resistors R_{ab1} , R_{au1} , R_{ab2} , R_{au2} in a Wheatstone bridge

is the same. With air flow the region upstream of the heater is cooled by the air flow. This temperature difference is detected by the resistors and is used as an air-flow signal. Owing to the symmetric layout of the sensor elements, a measurement of the direction of the airflow is possible (Fig. 7.6.3 and Fig. 7.6.4).

A flow measurement that is able to detect a flow direction can also be achieved by using two heating zones, located on the upstream and downstream side of the diaphragm. The upstream heater will need a higher voltage to maintain its temperature. The difference between the voltages to operate the upstream and downstream heater then corresponds to an air-flow signal that is dependent on the direction of the flow.

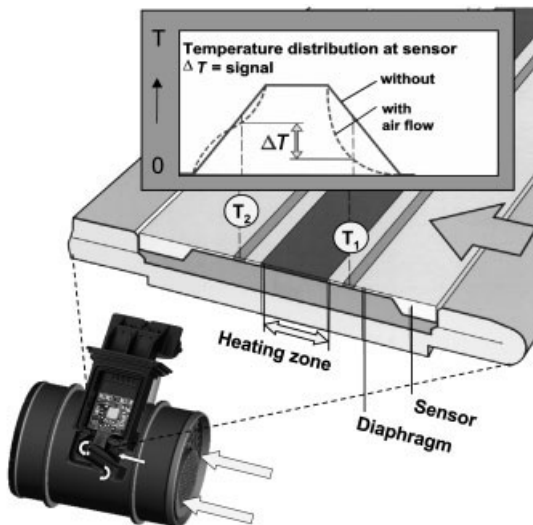
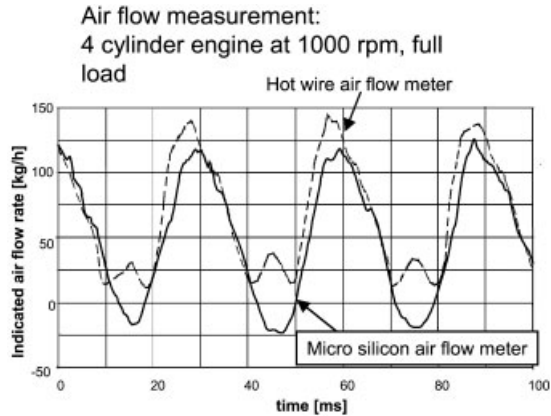


Fig. 7.6.3 Function principle of the Bosch mass air-flow meter.

A cross section of the sensor element shows the heating zone where the temperature is controlled to be about 180 K above ambient temperature. Without air flow the temperature is identical upstream and downstream of the diaphragm. With air flow the upstream side of the diaphragm is cooled. Temperature resistors upstream and downstream of the diaphragm detect the resulting temperature difference, which is used as an air-flow signal

Fig. 7.6.4 Back-flow detection with the microsilicon sensor element. Comparison of the indicated air mass of a microsilicon and of a hot-wire air-flow meter. The microsilicon air-flow meter indicates the back flow that appears periodically



7.6.2.2 Mounting the Sensor Element

To have a steady flow over the diaphragm surface of the microsilicon sensor element, flow separation has to be avoided upstream of the sensor element. Therefore it is necessary to mount it flush to a wall, with an upstream edge shaped with radii to give a steady flow. Another requirement is to induce no mechanical stress on the sensor chip. Therefore it is mounted like a beam on one side of the chip only. A third requirement is that the electrical connections are protected from the environment.

Fig. 7.6.5 shows a design that fulfills the above requirements by gluing the chip into a cavity in sheet metal. The sheet metal has a shape that avoids flow separation at its leading edge. The chip is mounted flush to the wall with only small gaps between chip and sheet metal upstream and downstream of the chip. The protection of the electrical connection is achieved by a partition wall, which separates the flow channel and the electronics chamber.

7.6.2.3 Sensor Package

For an air-flow meter the package plays a decisive role with respect to the measurement task. One aspect is that the package for an air-flow meter is determined

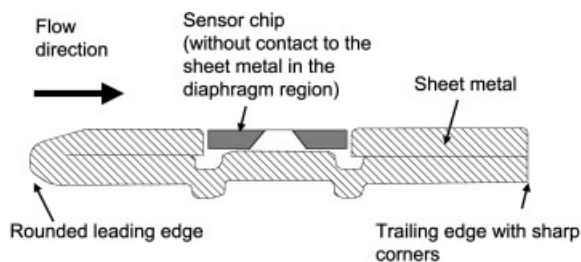


Fig. 7.6.5 Cross section of the mounting of a sensor element in sheet metal

by the need to form part of the intake tract of the engine. The other aspect is the need for a carrier for the sensor element, electronics circuit, flow passage, and connector. For the Bosch microsilicon air-flow meter a modular concept was chosen. The air-flow meter consists of the metering duct, which has flanges to fit to the adjacent parts of the intake tract. There are also flow straighteners, which are necessary to have a stable air flow within the metering duct. Sensor element, electronics circuit, flow passage, and connector are integrated in a plug-in module (Fig. 7.6.6). This module can be inserted in metering ducts of various sizes or directly into an air cleaner as long as there is a duct-like cross section suitable for measurement (Fig. 7.6.7).

As an aerodynamic measurement has to be performed it is necessary that the shape of the plug-in module follows aerodynamic principles. The range of air-flow rates to be measured is roughly 1:70. The flow velocities V are in the range 0.7–50 m/s, which give a Reynolds number in the range 1400–100 000. The non-dimensional Reynolds number, Re , is given by $Re = VL/\eta$ where L is the length of the plug-in measured in the flow direction and η is the dynamic viscosity of air. For Reynolds numbers of around 100 000 the aerodynamic knowledge for airfoils is valid (although it is in the low-velocity or small-scale range, such as for radio-controlled model airplanes or birds). For Reynolds numbers at the lower limit the flow behavior is different. Especially for flow around contours, the flow will separate immediately after the biggest cross section of the contour. The flow separation point is very sensitive to dimensional tolerances, temperature, vibration, and oscillating flow. As flow separation has a big influence on the flow field and therefore on the velocity near the sensing element, separation has to be avoided, or at least fixed. This can be done by using sharp corners and edges (Fig. 7.6.5). Also big regions of separated flow have to be avoided, because they may become un-

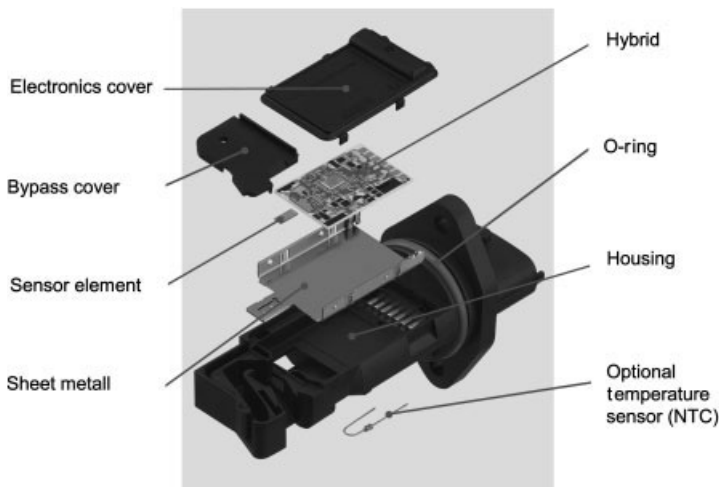


Fig. 7.6.6 Structure of the plug-in module

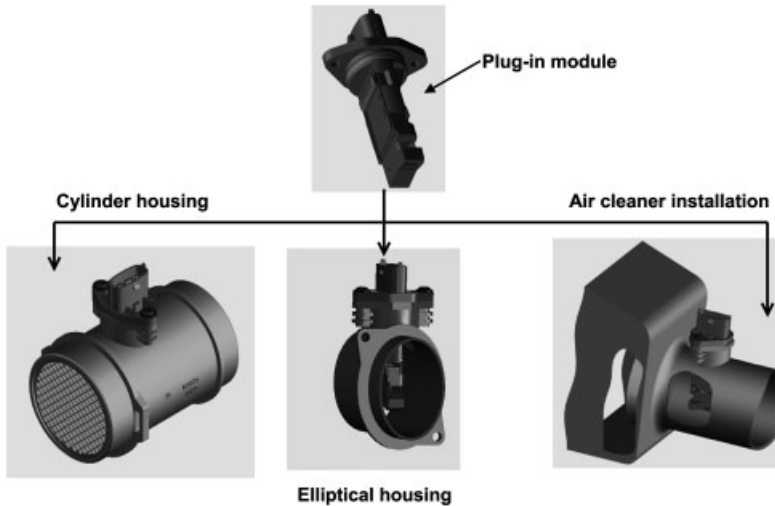


Fig. 7.6.7 Installation possibilities of the plug-in module

stable at certain Reynolds numbers. This can be achieved by generating longitudinal vortices, which reduce the region of separated flow. They can be produced by a three-dimensional structure at an edge or corner. This leads to geometrical design rules for air-flow meters.

- On the upstream side, round airfoil-like shapes are necessary.
- On the downstream side, a reduction in cross section has to be achieved with sharp corners and edges.

The shape of the flow passage that guides the air to the sensor element plays an important role in the function of an air-flow sensor (the bypass, Fig. 7.6.8). Cross section and length are the important factors. As an accelerated flow has the most stable boundary layers and therefore is less vulnerable to flow separation, it would be favorable to have a decreasing cross section (in the flow direction). To avoid a restriction of air flow in the bypass the cross section at the outlet should not be too small in relation to the inlet. Therefore a reduction of cross section for stable accelerated flow is only done in the vicinity of the sensor element and in curved parts of the bypass (Fig. 7.6.8).

The length of the flow path of the bypass from inlet to outlet L_B in relation to the length of the main air passage from bypass inlet to outlet L_M has an influence on the dynamic behavior. With oscillating air flow the mean air-flow rate in the bypass in relation to the mean air-flow rate in the main air passage is dominated by the relation L_B/L_M . If $L_B/L_M \gg 1$ the mean air-flow rate in the bypass is higher than without oscillating air flow [3, 5].

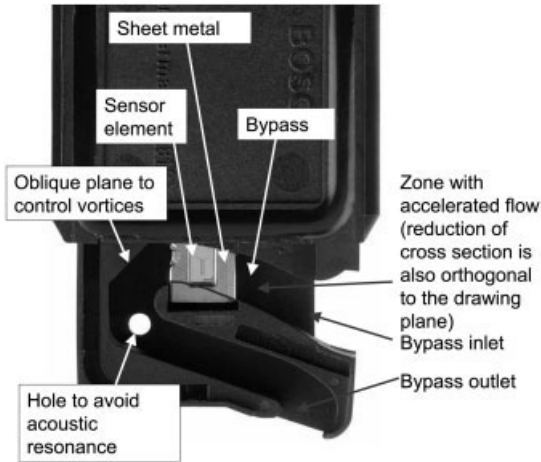


Fig. 7.6.8 Photo of the partially opened plug-in module showing the bypass shape

7.6.3

Signal Formation

The necessary function blocks for an air-flow meter are the following.

- Heater control.
- Signal amplifier.
- Offset correction.
- Decoupling of electric disturbances (overvoltage, Electromagnetic Compatibility (EMC)).

The Bosch microsilicon air-flow meter is available with analog (HFM5) or digital (HFM6) signal evaluation circuits.

7.6.3.1 Analog Output Air Flow Meter

The heating zone of the diaphragm consists of the resistors R_h (heater) and R_{hf} (sensor of heater temperature; Fig. 7.6.9). For heater control the resistors R_1 , R_2 , R_{hf} , R_{lf} (sensor for ambient temperature) are arranged in a Wheatstone bridge, which is adjusted by trimming the resistors R_1 and R_2 . Owing to the temperature dependency of R_{lf} the heater temperature changes dependent on the ambient temperature and, therefore, compensates for temperature dependencies of the air-mass signal. The controller (p-type) feeds the heater R_h until the bridge circuit is in balance. Therefore R_1 and R_2 control the excess temperature of the heater.

The resistors R_{ab1} , R_{au1} , R_{ab2} , R_{au2} located upstream and downstream of the heating zone on the diaphragm of the sensor element and are arranged in a Wheatstone bridge (sensor bridge, Fig. 7.6.9 and Fig. 7.6.10), which gives a basic signal of about 200 mV. The bridge is fed by a reference voltage U_{ref} of 5 V. The bridge voltage is connected to an analog amplifier with digitally adjustable gain. The offset of the bridge output is controlled depending on the temperature of the ambient air (Fig. 7.6.10).

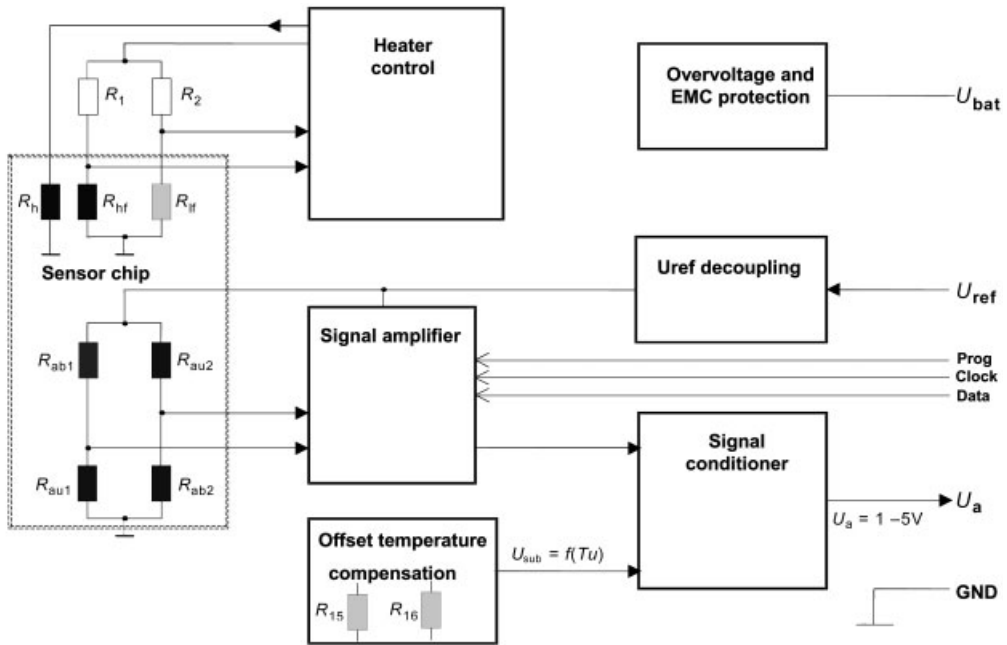


Fig. 7.6.9 Electric function blocks of the Bosch air-flow meter (analog design)

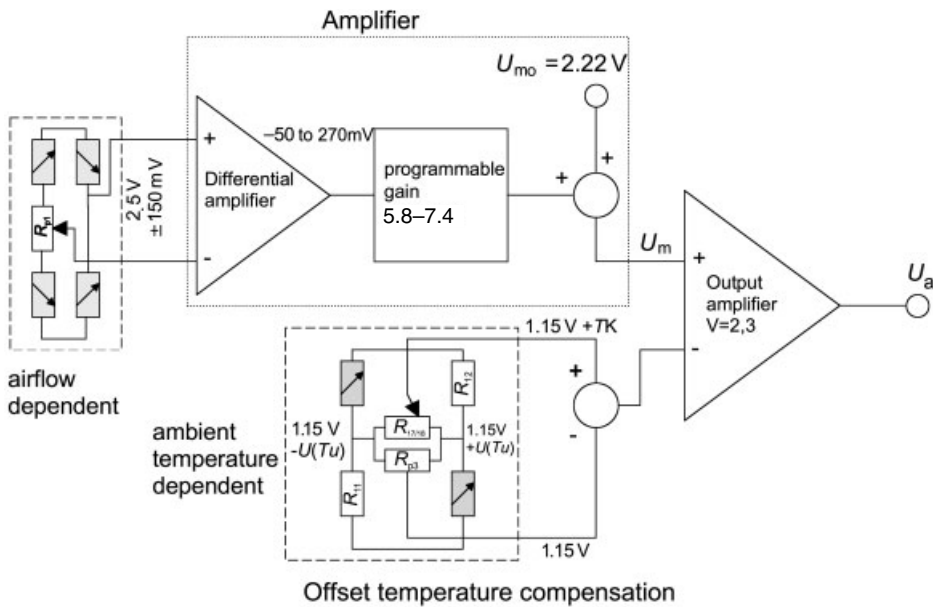


Fig. 7.6.10 Signal path of the analog design air-flow meter

The calibration of the air-flow meter HFM5 is performed with the following steps (Fig. 7.6.10).

- Calibration of heater temperature by laser trimming of resistor R_1 .
- Calibration of output signal without air-flow at 20°C by laser trimming of R_{P1} , R_{11} , and R_{12} .
- Calibration of output signal without air-flow at 80°C by laser trimming of R_{17} and R_{18} .
- Air mass calibration
 - Measurement of output at different air-flow rates.
 - Calculation of possible amplification range for every measured air-flow rate, so that the target of the characteristic tolerance is met.
 - Define optimal amplification, which fulfills characteristic tolerance for all measured air-flow rates.
 - Programming of amplifier. The amplification of the digital part can be adjusted with a 6-bit binary network of zener diodes.

7.6.3.2 Digital Version

The heater control circuit of the digital air-flow meter HFM6 is similar to the HFM5. In addition the voltage at R_{lf} (U_{lf}) is detected and used as an ambient air temperature signal. The sensor bridge is evaluated similarly to the analog circuit. The voltage U_{lf} and the output of the sensor bridge U_{br} are converted to a digital signal S_{lf} and S_{br} in an A/D converter (Fig. 7.6.11). The digital signal is used as

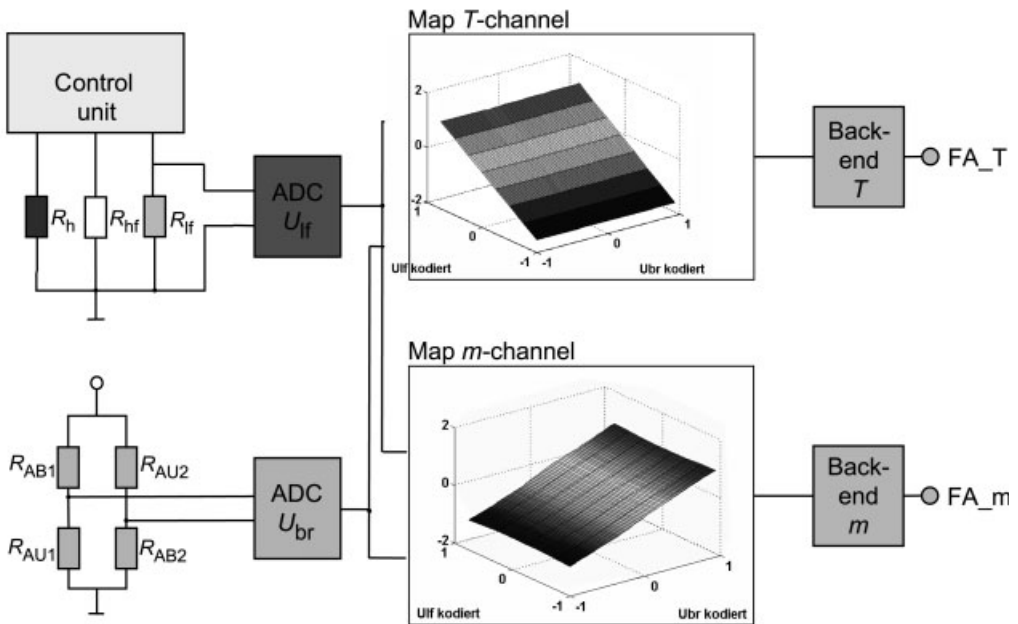


Fig. 7.6.11 Electric function blocks of the Bosch air-flow meter (digital design)

the address for two correction maps; one for temperature, one for air mass. The map stores correction values, which are added to S_{if} and S_{br} . The corrected value is then forwarded to the output drivers (back end) for air mass and temperature, and a frequency signal is generated.

For calibration of the air mass and temperature signal a measurement is performed at zero air-flow at a temperature of 20°C and at 80°C. The output signal of air mass is measured at five different air-flow rates at 20°C. Then a correction map is calculated for an optimal approach to the target characteristic. The digital concept with correction maps enables high precision air-flow measurement. The precision depends on the calibration procedure.

The temperature signal, which is needed for optimum temperature compensation of the air-flow meter, is available as an additional signal and can be used for engine control purposes.

7.6.4

Examples

Examples of the application of the air-flow meter HFM5 and HFM6 are shown in Fig. 7.6.7. Various standard metering ducts are available. Different sizes of the duct are available. Most of the application requirements can be met with these standard ducts. Especially due to spatial restrictions in the engine compartment special solutions are necessary. Therefore ducts with non-standard length and/or elliptical shapes are available. Metering ducts are available with different kinds of flow straighteners. The standard type consists of a combined grid with plastic flow rectifier and a wire mesh (Fig. 7.6.12). For reduced pressure drop, but aerodynamically less effective, metering ducts with only wire mesh are available.

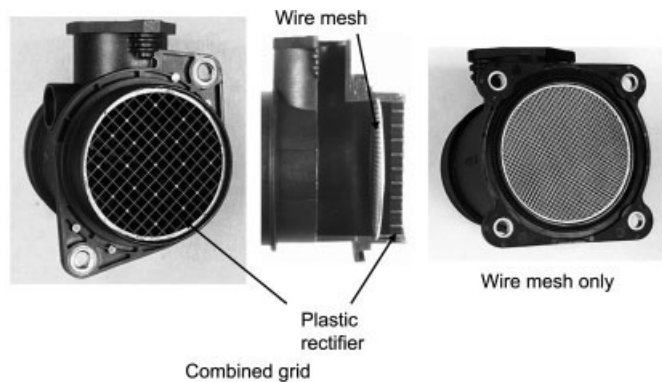


Fig. 7.6.12 Metering duct for air-flow meters showing different types of flow straighteners

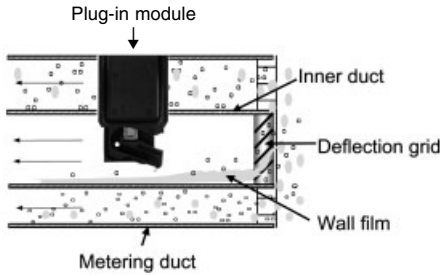


Fig. 7.6.13 Principle of water protection using a deflection grid: the water is deflected to the wall of the inner duct where it remains by adhesion. Direct contact of the water with the sensor element is avoided

Owing to increasing requirements with respect to contamination, a special type of metering duct was developed. The target of this design was to enable air-flow measurement with water droplets in the air flow. Because of their thermal properties, water droplets will lead to useless measurements with thermal air-flow meters. Removal of water was achieved using a special metering duct with deflection grid and inner duct. The principle is shown in Fig. 7.6.13. A deflection grid with angled guiding vanes gives momentum to the air and water droplets in the direction of the wall of an inner duct. Owing to their higher density the water droplets

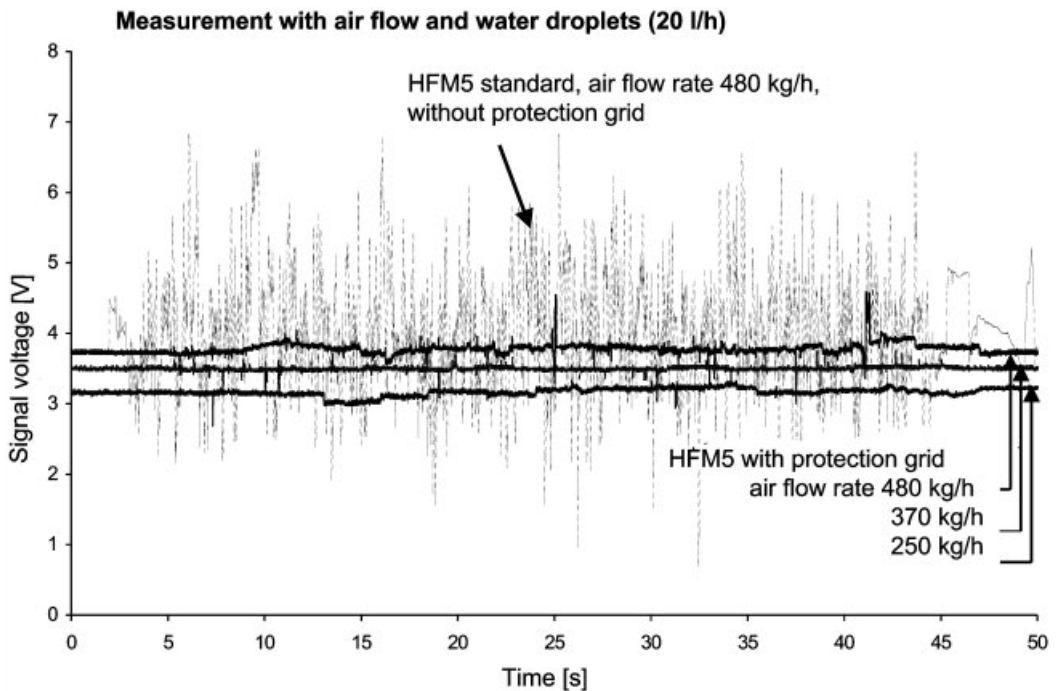


Fig. 7.6.14 Measurement of air flow including water droplets. Without the protection grid big disturbances of the air-flow signal appear,

whereas with the protection grid the disturbances are minimized

Fig. 7.6.15 Bosch air-flow meter with protection grid: HFM5-CI



will hit the wall. The water adheres to the wall, remains as a film, and is transported past the plug-in module without reaching the sensor element. This allows correct metering of the air flow. Also contamination of the sensor element with water is avoided. Contamination of the sensor element with dust is reduced by the deflection grid. A measurement of air flow with water droplets is shown in Fig. 7.6.14. The benefit of this special type of metering duct can be seen very clearly. Without the protection grid big disturbances of the air-flow signal appear, whereas with the protection grid, the disturbances are minimized. Fig. 7.6.15 shows an example of the air-flow meter with the protection grid.

Later the protection grid was included in the plug-in module. The design is shown in Fig. 7.6.16. The air flow is split into two parts within the bypass. The first part that will reach the sensor element has to go around a deflection edge with an angle of 180° . Owing to inertial forces large particles and water droplets cannot enter the first path to the sensor element. Dirt and water therefore remain in the second path and will leave the bypass by the particle/water outlet. Measurements confirmed that the good dynamic properties of the Bosch standard HFM5 prevail, although robustness was increased beyond that of the metering duct with protection grid.

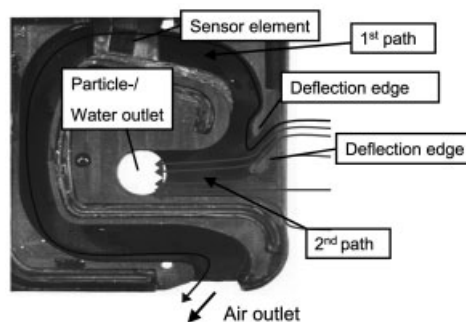


Fig. 7.6.16 Plug-in module with integrated protection for increased robustness. The bypass shape with deflection edge is shown. Dirt and water are separated from the first path of the bypass, which is used for metering the air mass, by inertia effects

7.6.5

Summary and Outlook

An overview of thermal air-flow meters is given here. Examples of microsilicon sensor elements and requirements for the packaging of air-flow meters are presented. Also different possibilities for signal formation are described. Finally it is shown that the good dynamics inherent to microsilicon sensor elements can be combined with robustness increasing measures that allow the air-flow meter to operate under severe environmental conditions.

7.6.6

References

- 1 BAUER, H. (ed.) *Automotive Handbook*, 4th edn, Bosch, Stuttgart, 1996, p. 470.
- 2 SAUER, R., SAE Technical paper 800468.
- 3 NISHIMURA, Y., OHYAMA, Y., SASAYAMA, T., UENO S., OHTANI T., SAE Technical paper 830615.
- 4 SAUER, R., SAE Technical paper 880560.
- 5 KONZELMANN, U., HECHT, H., LEMBKE, M., SAE Technical paper 950433.

7.7**Radar Sensors**

PETER KNOLL

7.7.1

Introduction

A significant number of accidents could be avoided through timely threat recognition and appropriate maneuvers for collision avoidance. This may be achieved either by a suitable warning to the driver or by automatic support to longitudinal or lateral control of the vehicle.

Adaptive cruise control (ACC) is a function for longitudinal control of vehicles, which relieves the driver of a significant part of the task of driving, and in a comfortable manner. ACC can be seen as a further development of cruise control, which maintains constant speed. In regions where traffic situations commonly permit steady speeds (such as in the USA) this function has been widely accepted: cruise control is fitted to more than 90% of US vehicles. In regions characterized by more unsteady traffic conditions with frequent speed changes, cruise control is less often found. In Europe this function is fitted to less than 10% of vehicles.

ACC is a logical derivative of cruise control; it automatically controls the speed to adjust the distance of the car from the vehicle in front and speed relative to preceding vehicles by employing suitable actuator systems [1].

After repeated research attempts in the 1960s and 1970s, the first ACC system was introduced to the Japanese market in 1995 by Mitsubishi. Toyota subsequently also introduced an ACC system. DaimlerChrysler introduced ACC in 2000, followed by BMW. Other European motor vehicle manufacturers will follow soon.

ACC radar technology presently operates in the long-range region (up to 120 m); this will soon be used also in the short-range (up to 20 m) to give additional safety and comfort functions in the automobile. Short-range radar devices are being developed, which will create a 'virtual safety belt' around the car. The signals from these sensors, partly combined with signals from the long-range radar, will make possible various new functions, such as ACC stop & go, pre-crash detection, and parking assistance.

The worldwide demand for long-range radar (LRR) devices for vehicles with ACC functions is estimated to reach 1 million units in 2003, climbing to 2.4 million units by 2005, while the worldwide demand for the short-range radar (SRR) sensor function is expected to reach 5.8 million units in 2005, from over 850 k units in 1999 [2].

7.7.2

Sensor Technologies

Fig. 7.7.1 shows the detection zones of the different surround-sensing technologies being used for driver assistance systems. It can be seen that radar technologies play an important role for long-range applications, and short-range applications as well. A possible substitution technology for radar is Lidar (light detection and ranging).

Radar is the acronym for 'radio detection and ranging'. This means that electromagnetic waves are used to measure the distance between a radar antenna and an obstacle. The simple form of the radar equation expresses the maximum radar range R_{\max} in terms of the key radar parameters and the radar cross section of the target.

$$R_{\max} = 4 \sqrt{\frac{P_t G A_e \sigma}{(4\pi)^2 S_{\min}}} \quad (1)$$

where

P_t = transmitted power [W]

G = antenna gain

A_e = antenna effective aperture [m^2]

σ = radar cross section of the target [m^2]

S_{\min} = minimum detectable signal [W]

Except for the radar cross section of the target, the parameters of this simple radar equation can be influenced by the radar designer. If long ranges are desired, the transmitted power should be large, the radiated energy should be concentrated into a narrow beam (large transmitting gain), the echo energy should be re-

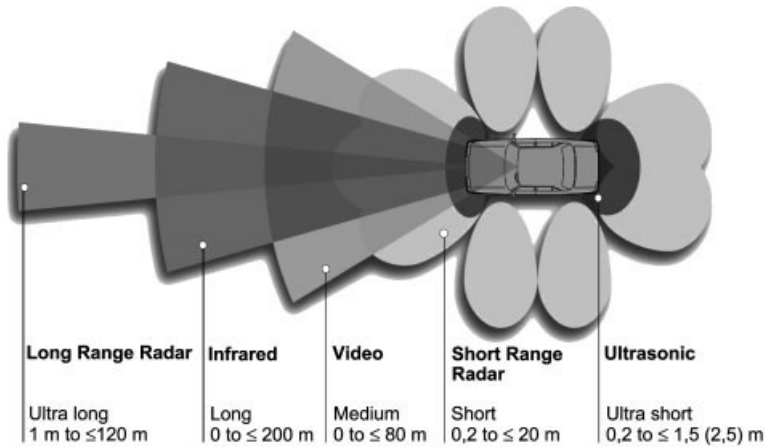


Fig. 7.7.1 Sensors for surround sensing

ceived by a large antenna aperture (also synonymous for large gain), and the receiver should be sensitive to weak signals [3].

To measure a distance with radar a flight-time measurement of the electromagnetic waves is made. The flight time t_{flight} is the duration between the emission of a transmitter signal and the reception of its echo from the target

$$t_{\text{flight}} = 2 \frac{d}{c} \quad (2)$$

where

d = distance between sensor and obstacle

c = velocity of light (300 000 km/s)

At a distance of 100 m the flight time is 0.66 μs . Thus, very short time intervals must be measured. To ensure adequate resolution between multiple obstacles (e.g., 1 m), a time separation of a few nanoseconds is necessary together with an adequate bandwidth of over 300 MHz.

The relative speed between the target and own vehicle is of basic importance for ACC. The maximum allowable error for the relative speed is in the range 0.25–0.5 m/s with the higher value being relevant at greater distances.

The relative speed v_{rel} between the car and an obstacle can be measured with the Doppler effect. A signal transmitted by the radar is subject to a frequency shift f_{Doppler} when it is reflected by an object moving towards or away from the own system.

$$f_{\text{Doppler}} = 2 \frac{f_c v_{\text{rel}}}{c} \quad (3)$$

where

f_c = carrier frequency (76.5 GHz)

For a relative speed of 1 m/s the frequency shift is roughly 510 Hz.

7.7.2.1 Long-Range Radar System

The long-range radar sensor used today for ACC is a frequency modulated continuous wave (FMCW) 76.5 GHz radar with a field of view of $\pm 4^\circ$. With a distance range of 2–150 m (depending on obstacle size and reflectivity) it is able to detect multiple objects and to measure distance, relative speed, and the angle to an object simultaneously. The angle measurement is determined using three beams working in parallel without any switch.

Frequency Modulated Continuous Wave Radar

Long-range radar devices use FMCW with an indirect flight-time measurement. Fig. 7.7.2 shows a very simple block diagram of a FMCW radar that utilizes the Doppler shift to detect moving targets.

In practice CW radar is much more complicated. A CW radar transmits while it receives. It compares the frequency of the transmitted signal with the frequency of the received signal. A precondition for the measurement is that the transmitted frequency changes over time, which is usually done by a ramp-form linear modulation with a slope of $m = df/dt$. When the received signal returns after the flight time $t_{\text{flight}} = 2d/c$, the transmitted frequency has changed by $f_{\text{Doppler}} = t_{\text{flight}}$. Thus, the flight time, and indirectly the distance between transmitter and obstacle, can be calculated by determining the frequency difference between the transmitted and the received signal.

The transmitter generates a continuous (non-modulated) sinusoidal oscillation at the frequency f_c , which is then radiated by the antenna. On reflection by a moving target, the transmitted signal is shifted by the Doppler effect by an amount $\pm f_d$, depending upon the direction of the relative movement. To utilize the Doppler shift, the radar must be able to recognize that the received echo signal has a frequency different from that which was transmitted. The transmitter leakage signal acts as a reference to determine that a frequency shift has occurred. The detector (mixer) multiplies the echo signal at a frequency $f_c \pm f_d$ with the transmitter leakage signal. The Doppler filter allows the difference frequency to pass and rejects the higher frequencies. Stationary clutter echoes are rejected.

All modern ACC radar sensors make use of the Fourier transformation for signal processing. Viewed in a simple manner, the Fourier transformation is a calculation-intensive transformation from the time domain to the frequency domain and reverse. A series of ‘measured values’ defined in frequency steps, the frequency spectrum, is derived from a series of defined time steps. Modern signal

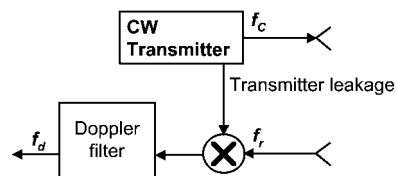


Fig. 7.7.2 Frequency modulated continuous wave radar: block diagram

processors are powerful enough to perform this transformation with a large number of measurement values (about 1000) in a few milliseconds.

Radar Antenna

The radar antenna, in addition to transmitting and receiving the radar waves, is responsible for the focussing of the millimeter waves. In principle, planar antennas, lens antennas, or reflector antennas can be used. Today, monostatic lens antennas are commonly in use.

In lens antennas, the generation of the beam is independent of focussing. Transmission and reception is performed by feed elements, which may be patch elements or the end of waveguides. The wide, exited beam is focussed with a dielectric lens, comparable to optical focussing but not completely the same because of diffraction effects.

Monostatic antennas use the same focussing mechanism for the transmitted and the received signals, with the signal channels being separated electronically. Direction-sensitive decoupling between the transmitted and the received signals is required for the FMCW radar. As indicated in Fig. 7.7.3, this can be achieved with a circulator or with suitable directional coupler structures. Fig. 7.7.4 is a photograph of the Bosch 77 GHz sensor.

The antenna window for the millimeter-waves is a lens of plastic material, which can be heated by wires in order to increase the availability in the winter season. The whole sensor is mounted in air cooling slots of the vehicle front end or behind plastic bumper material by means of a model-specific bracket.

As microwave devices above 50 GHz are rarely used in civil applications, a useful design is necessary for the ACC radar. The development must concentrate on reasonable cost, high environmental robustness, and (very important for the automotive application) small size.

A 76.5 GHz Gunn-oscillator on the radar transceiver delivers the output power for the three beams and the same number of mixers for the demodulation of the received waves. The modulation is controlled by a frequency-locked loop using a

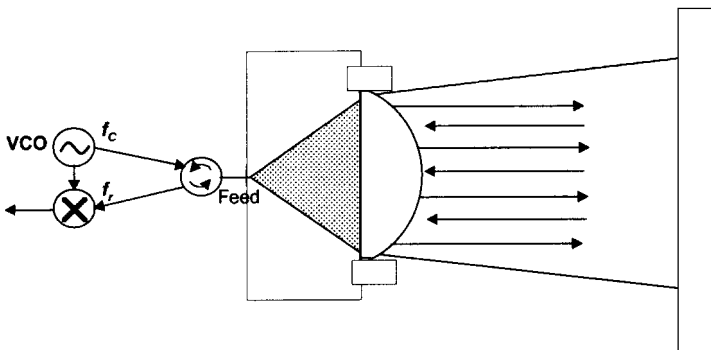


Fig. 7.7.3 Frequency modulated continuous wave radar with monostatic lens antenna and circulator

Fig. 7.7.4 77 GHz sensor



12.65 GHz reference oscillator. This control circuit guarantees linearity of the modulation sweeps and the compliance of the frequency band limits (76 GHz).

The digital signal-processing control unit is the link to the actuator subsystems and to external connections for operation switches. Usually, a 24-bit digital signal processor is used assisted by an ASIC for the fast Fourier transformation. Technical data of the sensor are as follows.

- Temperature range: -40 to 85°C .
- Power consumption: 13 W.
- Package size: $12.4\text{ cm} \times 9.1\text{ cm} \times 9.7\text{ cm}$.
- Weight: $<0.6\text{ kg}$.

7.7.2.2 Short-Range Radar System

As a first step, short-range microwave sensors will be introduced that operate at a frequency of 24 GHz. These sensors can be used to build a ‘virtual safety belt’ around the vehicle covering a variety of functions.

The short-range radar (SRR) uses flight-time measurement, but is adapted to the specific requirements for the short-range (20 cm to 20 m).

Fig. 7.7.5 shows the block diagram of a SRR sensor. A CW oscillator feeds its signals to a power splitter. Its outputs go to two high-speed switches in both channels shown in the diagram. In the upper path the signals of a pulse generator are first shaped and then given to the high-speed switch. From this component the signals go to the transmitter antenna. In the lower parallel path an adjustable delay forms reference signals, which are given to a high speed switch in the receiving path. The received echo signal is mixed with the output of the CW oscillator, which acts as a coherent reference to allow recognition of any change in the received echo-signal frequency. Coherence means that the phase of the transmitted

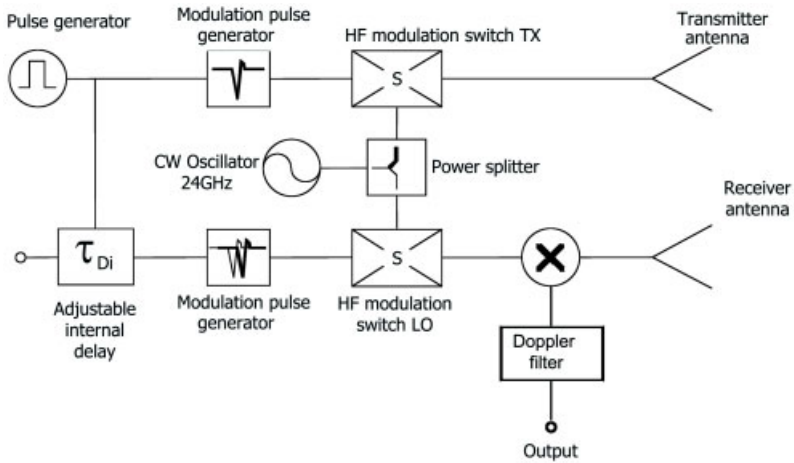


Fig. 7.7.5 Block diagram of a pulse radar

pulse is preserved in the reference signal. The change is detected by the Doppler filter [3]. Fig. 7.7.6 shows a picture of a SRR sensor.

The setup shown uses discrete radio-frequency components. The μC and the signal processing is integrated on the board. The pulse width of the 24 GHz radar is 300 ps. Some technical data of the sensor are as follows.

- Radiated power: 20 dBm (peak) EIRP (effective isotropic radiated power).
- Range: 0.2–20 m.
- Opening angle: horizontal $\pm 40^\circ$ (at -3 dB), vertical $\pm 10^\circ$ (at -3 dB).
- Object resolution: 20 cm.
- Accuracy: ± 3 cm.
- Measurement cycle: minimum 2 ms.
- Temperature range: -40 to 85°C .
- Multi-target capability.



Fig. 7.7.6 24 GHz short-range radar sensor. The high frequency circuit is on the rear side of the substrate. On the right the antenna patches are displayed

The unsymmetrical radiation behavior is achieved by the arrangement of the antenna patches. It can be varied by using dielectric lenses in front of the antenna.

Short-range microwave sensors are being introduced in series in the near future. The major benefit to the customer is the multi-usage capability of this sensor opening up the opportunity to build a sensor platform.

7.7.3

Examples

Radar sensors are used for single functions, where the function is based on one sensor type only, other functions are based on a sensor data fusion, where the objects detected by the different sensors are compared and matched. By doing this, stable, quick, and reliable object information is achieved.

7.7.3.1 Adaptive Cruise Control

ACC differs from other vehicle control functions, especially in that the function is performed by several electronic control units (ECUs). While conventional control systems consist of a sensor and actuator environment around a central ECU, ACC adds functions to existing systems. A truly new component is the sensor for measuring the distance, relative speed, and lateral position of potential target vehicles, using laser optics or millimeter waves. This component often contains the logic for controlling vehicle movement. The latter is affected by commands of ECUs for engine and brake control electronics.

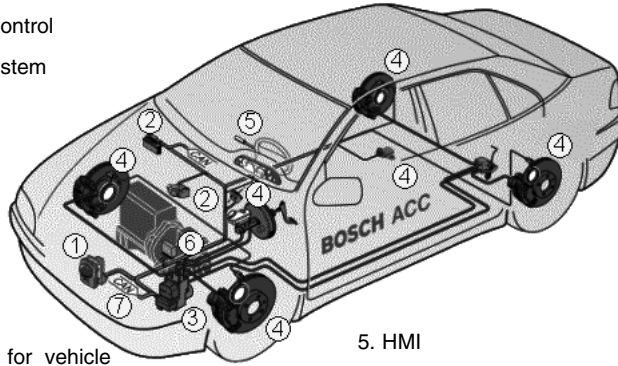
Current ACC designs are intended for use on freeways and other high-speed roads. Extension to urban traffic and to automatic traffic congestion following is desirable, and matter of actual development. In addition to the basic ACC sensors, complicated additional sensors are required to completely recognize all objects and obstructions relevant to driving. To avoid misuse of the standard ACC, a minimum speed is defined. Below this speed (30–50 km/h is typical), the ACC control function is switched off, or at least limited to prevent positive acceleration. Fig. 7.7.7 shows the basic components of an ACC system.

The raw data of the long-range radar are tracked and filtered, so that the ACC control algorithm can select the relevant object. This selection requires the predicted curvature of the subject vehicle course. Basically the yaw rate (coming from ESP, electronic stability program) and radar object information is used for course prediction. With the selected object a control chain starts resulting in commands for engine control and/or the brakes in order to establish a convenient and safe following behind a preceding vehicle. Further system aspects of ACC are described elsewhere [4, 5].

Basic Functions

The minimum function of any ACC is the normal cruise-control function, namely constant speed control at a speed set by the driver (set-speed control). It occurs when no preceding vehicle requires a slower speed. It thus is also used when a preceding vehicle is driving faster than the ACC vehicle. If the ACC car is approaching a slower-travelling vehicle, the car decelerates (brake interference in-

1. Radar sensor with integrated ACC controller
2. Engine control
3. Brake system



4. Sensors for vehicle yaw measurement

5. HMI

6. Automatic transmission

7. Data network (CAN)

Fig. 7.7.7 Adaptive cruise control system from Bosch

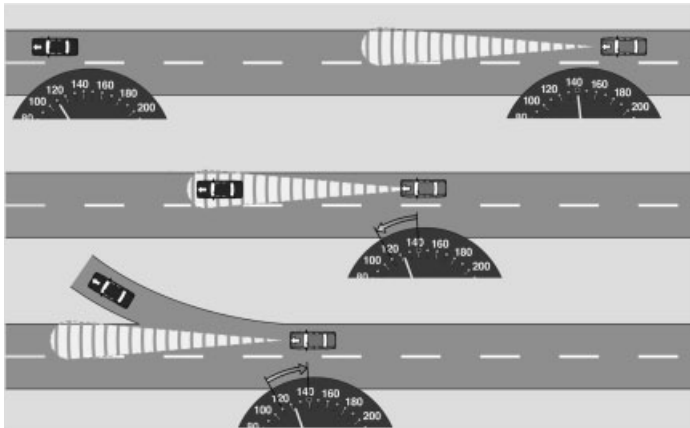


Fig. 7.7.8 Basic function of an adaptive cruise control system

cluded) and the safety margin to the vehicle in front is maintained. If the vehicle in front turns off the highway, the originally chosen speed is resumed. Fig. 7.7.8 shows the different scenarios.

ACC sensors make it possible to follow the car in front at a short distance, but sensor developers have agreed that they will not support using ACC to follow at shorter distances than the driver without ACC would maintain.

Target Selection

Despite sophisticated logic for object selection and control, there always remain a few situations where ACC does not provide the performance desired by the driver. Typical features of such situations are ambiguous vehicle constellations, special

road geometries, and the principal difficulty of predicting lane changes. An especially severe limitation of the ACC function is the discrimination of stationary targets. This result is that ACC only considers objects for which a minimum speed has been measured (typical values 20 km/h or 20% of the speed of the ACC vehicle). This limitation is necessary because of the limited capabilities of the sensor, but also because of the control definition. Since objects other than vehicles are measured with the ACC sensor, there are many things on, next to, and above the road, which the sensor has difficulty distinguishing from vehicles. This leads to a high risk of false reactions, if they are not removed from the control cycle by discrimination as stationary objects. Even if the stationary object is detected as a 'real' vehicle, frequently the desired reaction is not to stop, but to pass it without delay. A vehicle parked halfway on the road and halfway on the shoulder would be typical of this situation. The possible reaction 'stop in front of stationary targets' is not included in the ACC function. The considerations previously listed regarding the uncertainty of predicting the course of the ACC vehicle at larger distances and at large relative speeds also apply to stationary targets.

Topology Limitations

Both vertical and horizontal curves in the road limit ACC function. The sensor range is additionally limited by the geometry of the road, especially at hilltops and sharp curves.

Weather Limitations

Despite various degrees of weather resistance of the technology used, overall robustness against weather cannot be achieved. In extreme weather situations (e.g., strong snowfall, extreme rain, icing of the vehicle), full functionality of the sensor cannot be guaranteed. Large amounts of dirt may also reduce the sensor function. When the sensor range is reduced to less than that of the driver, the sensor logic should give the driver information to this effect.

Display and Control Requirements

The basic display and control requirements of the man-machine interface are the following.

- Activate and deactivate the ACC function (i.e., the transition between the passive waiting status and the active control status).
- Display the ACC status (i.e., passive, active, not available).
- Override ACC function by gas pedal.
- Feedback front driver reaction, although activation of ACC is principally independent of choice.

ACC Stop & X

The next steps in ACC development are the so called ‘stop & X’ functions. Improvement in performance is achieved by combining the signals from other sensors with the signals from the ACC sensor by data fusion. The necessity for this step is shown in Fig. 7.7.9, showing the low-speed following (LSF) function (also often called ACC ‘stop & roll’). While the vehicles in front of the ACC car are detected safely by the radar, objects in the vicinity (like the motorbike to the right of the car) are not detected due to the limited angle resolution of the ACC radar. Only a near-range sensor with a wide angular detection range can detect this situation. Short-range radar devices are suitable for this function.

The customer benefits of such a system are as follows.

- Driver stress is further reduced during traffic jams.
- The ‘stop & go’ control functions up to speeds of 20–30 km/h behind the vehicle in front.
- Deceleration down to zero speed.
- Acceleration with drivers release.

ACC stop & go covers the whole speed range of the vehicle, is also suited for urban traffic, and will combine ACC LSF with video sensors. This combination will eliminate the drawbacks of the current systems with respect to the evaluation of stationary objects [6].

7.7.3.2 Short-Range Radar

The 24 GHz SRR sensor allows general distance and velocity measurement with high resolution. This opens up the possibility of achieving the following functions.

- Parking aid, backing aid.
- Blind spot detection.
- Measurement of parking space (length and depth).
- Parking assistant guiding parking maneuvers.
- Pre-crash detection for front and side.
- Short-range sensor for ACC stop & go function (may require additional sensor).

Although the introduction of this kind of sensor will most probably be driven by one or two single functions, investigations have shown that eight sensors (depending on car size) allow coverage of an almost complete surround view with all its functionality. Multiple sensors are also necessary to reduce the cost of the system. The resulting concept for short-range sensing is depicted in Fig. 7.7.10.

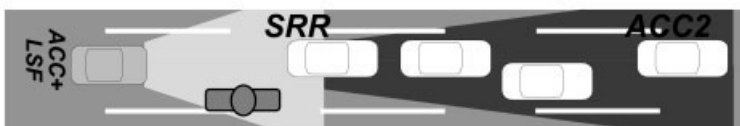


Fig. 7.7.9 Adaptive cruise control low-speed following

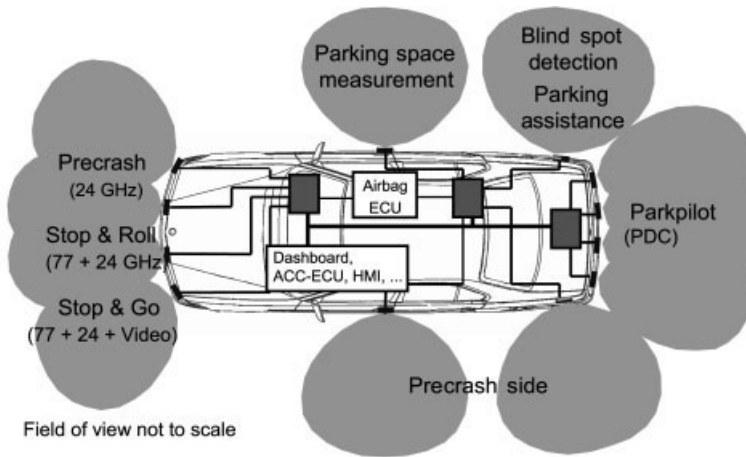


Fig. 7.7.10 Short-range radar platform

Pre-Crash Function

As an example of the functions described above, the pre-crash function is described in more detail here.

Functions of passive safety based on SRR sensors are a major issue for a complete safety concept for the automobile. Crash-optimized vehicle structures and passenger cells are a matter of course for new vehicles just as coordinated restraint systems are, consisting of belt tensioners and airbags. In addition, the use of side and head airbags is increasing rapidly; foot and knee airbags will soon follow in series production. The focal point of airbag development is circumstance- and requirement-appropriate activation with the help of intelligent sensing and triggering electronics. This depends on the severity and progression of the accident as well as on seat occupation.

Airbag control has primarily focused up to now on central control, or rather, acceleration sensors distributed throughout the vehicle, which provide information at the moment of a crash. However, better triggering strategic information from the pre-crash phase (just before collision) is needed. Bosch pre-crash sensing is sub-divided into three different function levels: 'pre-set' (pre-crash setting of algorithm thresholds), 'pre-fire' (pre-crash firing of reversible restraints), and 'pre-act' (pre-crash engagement of active safety devices). The short-range radar platform is especially suited for use with these functions.

Fig. 7.7.11 shows the pre-crash roadmap. In the first stage (pre-set), radar sensors provide information such as relative speed to obstacle and expected moment of collision. The system includes this information in its triggering decision. Because the radar sensors cannot provide any information on mass or condition of the obstacle, the system decides on actual triggering of a restraint medium based on acceleration sensors. For example, this pre-set function allows the belt to be tightened earlier or the triggering point can be optimized for the second airbag

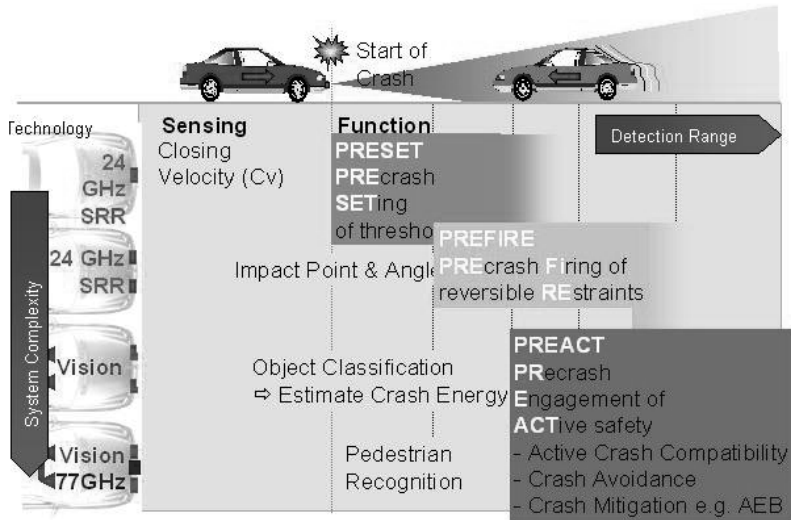


Fig. 7.7.11 Pre-crash sensing roadmap

stage. In addition, this extra information prevents undesired airbag triggering for minor accidents. Overall, the pre-set function improves identification of the crash severity and accident type.

Starting in 2004, the second level (pre-fire) will be feasible. Using pre-crash information, restraint systems can be triggered before the crash according to the requirement. But because information about the mass of the obstacle is not yet available before the crash, new restraint mediums must be developed, for example, the reversible belt tightener, which can function several times.

The third level (pre-act) has the goal of minimizing the severity of the crash or even avoiding the accident through active system intervention, for example through automatically introduced emergency braking. Within the development stages of pre-crash, this is certainly the level with the most ambitious goals. In addition to video sensors and long-range radar, Bosch is also working on tying all prior described sub-systems into a network. This vision will require a completely new overall system architecture and safety strategy [7].

Introduction of Systems for Warning and Vehicle Interference

Fig. 7.7.12 shows a possible introduction scenario for the 24 GHz technology in a sensor platform.

Beyond the year 2003, the large complexity of functions and variety of sensors that is expected by then will be accounted for by employing comprehensive sensor fusion [8].

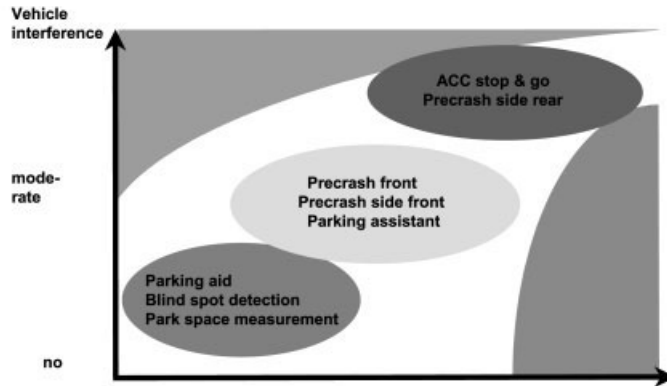


Fig. 7.7.12 Possible introduction scenario for a 24 GHz platform

7.7.5

Summary and Outlook

Radar-based functions for environmental sensing are in rapid progress. Some applications, such as adaptive cruise control, are already on the market, others will follow soon. Meanwhile complexity of the functions will grow steadily. A significant step will be made with the introduction of a 24 GHz platform fulfilling a multi-usage concept for the sensors. In the sequel, a significant extension of functions is expected. Radar technology will be introduced first in up-market cars, followed by a wider distribution as lower price levels are achieved.

Significant changes are expected in the sensors for the applications mentioned here. In order to further reduce the size of the sensors, patch antennae will also be used for long-range radar applications. Scanning antennae will change from a mechanical scan, as it is used today, to an electronic scan making use of phased array technologies. In the field of electronic devices we can expect that MMICs (microwave monolithic integrated circuits) and SiGe technology will be used in the future.

7.7.6

References

- 1 WINNER, H. in *Automotive Electronics Handbook*, 2nd edn. McGraw Hill, New York, NY, USA, 1999, pp. 30.1–30.38.
- 2 *Strategy Analytics*, March 2001.
- 3 SKOLNIK, M.I. *Introduction to Radar Systems*, 3rd edn. McGraw Hill, New York, NY, USA, 2001.
- 4 WINNER, H., WITTE, S., UHLER, W., LICHTENBERG, B., SAE technical paper 961010.
- 5 WATANABE, T., KISHIMOTO, N., HAYAFUNE, K., YAMADA, K., MAEDE, N., in *Proc. 2nd IST World Congress*, Yokohama, Japan, 1995, pp. 1229–1235.
- 6 OLBRICH, H., BEEZ, T., LUCAS, B., MAYER, H., WINTER, K., SAE technical paper 980607.
- 7 ZEHNALL, M., presented at *5th International Automotive Press Conference*, Boxberg, 2001.
- 8 SEGER, U., KNOLL, P.M., STILLER, C., in *Proc. Sensor Vision and Collision Warning Systems, Convergence Conference*, Detroit, MI, USA, 2000.

7.8

Video Sensors

PETER KNOLL

7.8.1

Introduction

Intense research and feasibility studies have demonstrated that surround sensing can provide a broad spectrum of vehicle functions, ranging from driver warning functions through driving support functions and collision mitigation systems to complete autonomous vehicle guidance [1–5]. Up to now, though, few systems have reached the market. New systems are being developed and will be introduced when new sensors become available.

Pictures have the highest information content for humans. People perceive the vehicle surroundings with their eyes, and evaluate traffic situations with their intelligence and experience. Hence, for the development of driver assistance systems it seems obvious to acquire images, extract relevant details, and deduce hazardous situations by means of intelligent image processing.

As a first step, functions for driver warning or information (e.g., lane-departure warning, night-vision improvement, or distance warning) are being introduced. As a second step, functions that intervene with vehicle dynamics through braking, steering, and acceleration, will open up new possibilities for collision mitigation and collision avoidance. Fig. 7.8.1 shows the detection zones of the different surround sensing technologies being used for driver assistance systems, comparable with Fig. 7.7.1 in Section 7.7, but with special emphasis on video technology. The figure shows the high potential of video technology for a wide range of functions.

Adaptive cruise control (ACC), a function for longitudinal control of vehicles, has been introduced first. ACC relieves the driver of a significant part of the task of driving, and in a comfortable manner. But the system still suffers from some drawbacks.

- Radar allows only the measurement of distances and a very limited spatial resolution with scanning systems, so in practice stationary objects, being of relevance in certain situations, are suppressed.
- Radar has a narrow beam. A camera with a wider angle would improve detection.
- Radar cannot detect lane borders and allows no classification of objects; this is only possible with video technology.

Video systems were used first in Japan in the 1990s as maneuvering aids for camper vans, later reversing cameras were developed, showing the image from a camera with wide-angle optics on a display. In 1999 the first ACC stop & go system became available in Japan: this used a stereo camera. In 2000 a lane-departure warning system was introduced in trucks in Europe: it used a mono camera. Also in 2000 a night-vision improvement system based on a far infrared (FIR) sensor came on to the market in Cadillac Seville cars.

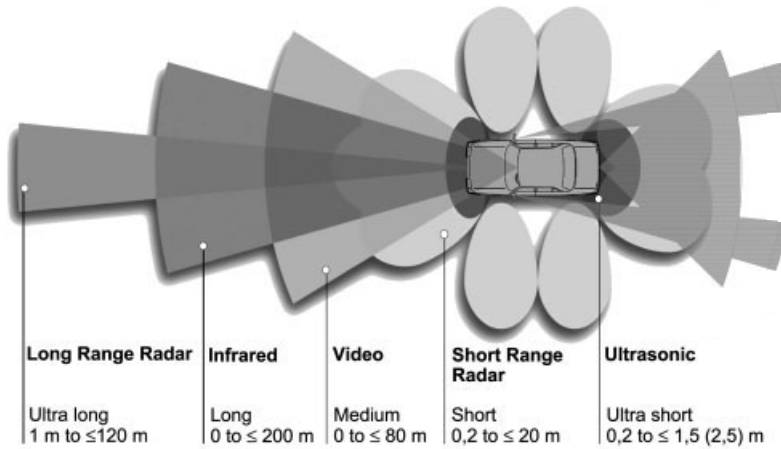


Fig. 7.8.1 Detection zones of different surround sensing devices for driver assistance systems

The worldwide demand for video cameras for vehicles is estimated (optimistically) to reach 500 k units in 2003, and climb to 4.8 million units by 2005. For 2010 a volume of more than 63 million units is foreseen [6].

7.8.2

Components of a Vision System

Fig. 7.8.2 shows the basic components of a camera system, and the signal flow. A vision system or optical measurement system consists of the elements illustrated in the figure. A lens collects radiation from a source, such as a light bulb or an illuminated object. A sensor converts the collected radiation into an electronic charge, which can be preprocessed by electronics. The preprocessed information is con-

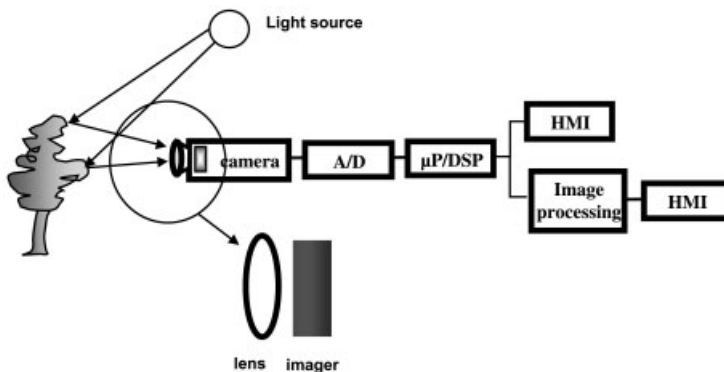


Fig. 7.8.2 Signal flow in a camera system

verted into digital form for treatment in a computer. In the case of automotive vision systems, the purpose of this image-processing step is either to enhance certain details of the image or display the modified image to the driver, or to extract certain types of information contained in the image by using special algorithms. This information can then be used to issue a warning to the driver or, by interacting with suitable actuators in the vehicle, to bring about braking or steering.

7.8.3

Fundamentals of Photosensing

Incident photons pass into a semiconductor device in which most of the photons interact with a target to produce electron-hole pairs. These photocharge pairs need to be separated in an electric field before they recombine again, leading to the flow of a photocurrent, which is proportional to the incident light intensity over many orders of magnitude. Two types of interactions of photons with solid-state materials have to be considered for an understanding of the properties of an image sensor: absorption and reflection.

On top of the image sensor, there is a scratch-resistant transparent covering and protective materials, often in the form of dielectric layers such as silicon dioxide or silicon nitride. In the bulk of the semiconductor there is first a space-charge region. Below this, the field-free bulk of the semiconductor follows, which can be as thin as a few μm or as thick as many $100 \mu\text{m}$ [7].

Because of the sequential process of *photocharge generation*, virtually all photons that are absorbed in the semiconductor material are converted into an electronic charge [8]. There is a strong spectral dependence, however, of the mean absorption depth at which this photoconversion takes place. Short-wavelength light is predominantly absorbed at the surface, while red light penetrates deeply into the bulk of the semiconductor. As a consequence, the achievable spatial resolution degrades significantly with wavelength. Images taken in the red or infrared spectral region (e.g., night-vision improvement systems) show much less contrast than images taken in green or blue light. For this reason, image sensors are often covered with an optical filter, cutting off the infrared portion of the incident light.

In the absorption process, a photon loses its energy by creating one or more charge pairs, consisting of an electron and a hole. The efficiency of this process is usually measured by the *quantum efficiency*, η , describing how many charge pairs are created and electronically detected per incident photon.

Once an electron-hole pair has been created, it must be separated within a certain time before it recombines again and loses all information about the previous presence of the photon. This recombination lifetime depends on the quality and purity of the semiconductor. In high-quality low-doped silicon used in CMOS processes, for example, the lifetime can be as long as several tens of microseconds. This is the time available for separating the photocharge and moving the different charge types to suitable storage areas.

Photogenerated charge carriers moving under the influence of an electric field represent a current, the so-called photocurrent, that is proportional to the incident light

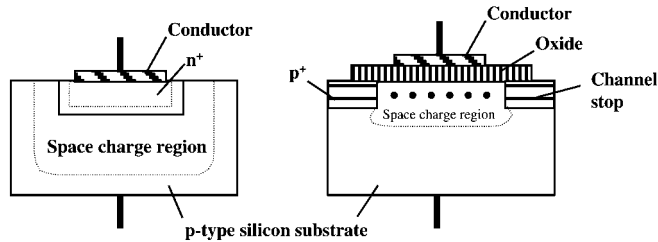


Fig. 7.8.3 Photosensitive structures: left, photodiode; right, metal-oxide semiconductor capacitance

intensity over many orders of magnitude (≥ 10). It is this strict linearity of photocurrent with incident light over a wide dynamic range that makes semiconductor photosensors so attractive for many applications in image sensors and optical measurement systems.

Fig. 7.8.3 illustrates the two most important photosensitive structures, the photodiode (PD, left) and the metal-oxide semiconductor (MOS, right) capacitor as used in the charge-coupled device (CCD) image sensors. Both devices are easily fabricated with standard semiconductor processes. A photodiode consists of a combination of two different conductivity types of semiconductor, as illustrated in Fig. 7.8.3, left. In the junction between the two types of semiconductor, an electric field exists in the space-charge region. At the same time, this space-charge region has a certain capacitance, varying with the inverse of the space-charge region width. Photodiodes are typically operated by biasing them to a certain potential and exposing them to light. Photocharge pairs entering the space-charge region are separated in the electric field of the PD, a photocurrent is produced, and the photocharge is accumulated in the PD capacitance. After the exposure time, the residual voltage is measured, and the voltage difference compared with the reset voltage level is a measure for the amount of light incident on the pixel during the exposure time.

The MOS capacitance illustrated in Fig. 7.8.3, right, consists of a thin layer of oxide on top of a semiconductor material. The oxide is covered with a conductive material (e.g., polysilicon). As in the case of the PD, the MOS structure is biased to a suitable voltage, leading to a space-charge region of a certain extent in the semiconductor. Again, photocharge is separated in the electric field and it is integrated in the MOS capacitance, collected at the interface between semiconductor and oxide.

A typical value for the PD and MOS area capacitance is $0.1 \text{ fF}/\mu\text{m}$. Once this storage capacity is exceeded, additional photocharge in the corresponding pixel starts to spill over to neighboring pixels. This effect is called blooming, and well-designed image sensors provide special collecting ('antiblooming') structures for a reduction of this effect [9].

The simplest type of photocurrent processing is integration of the photocurrent during a certain time: the exposure time. In this way an integrated charge is obtained that is proportional to the number of photons incident on the pixel during the exposure time. This functionality is very easy to implement by employing the capacitance of the device used for generating the electric field for photocharge separation.

7.8.4

Sensor Technologies

The classical approach to making video camera systems is based on using CCD imagers. This technique has matured during the last 25 years and CCD users have become accustomed to their limitations in resolution, dynamic range, temperature behavior, and the complex interfacing.

CMOS sensors feature several advantages with respect to the more generally used CCDs: they are fabricated in a fully standard VLSI (very large-scale integration) technology, which means low costs, taking advantage of submicron CMOS technology. Several functions can be integrated on the sensor substrate, including random access. They consume very little power as the circuitry in each pixel is typically active only during the readout and there is no clock signal driving large capacitance. Readout speed can be enhanced by parallel access to multiple taps of the pixel array. Because of these characteristics, CMOS sensors are the favored technology for demanding applications such as automotive uses.

7.8.4.1 CCD Sensors

In the case of CCDs, the photocharge is stored under a precharged MOS capacitance. The basic CCD idea is to combine a linear array of MOS capacitances, so that a stored photocharge can be moved laterally under the influence of appropriate MOS electrode voltage patterns. Photocharge pairs are generated in the semiconductor by incident light. Moving by diffusion and by drift, the photoelectrons can find their way to positively biased MOS electrodes (gates), where they are stored at the interface between semiconductor and thin oxide. The photogenerated holes are repelled by the positive gate voltage, and they move around by diffusion until they finally combine in the silicon substrate.

There are basically three types of CCD image sensors (imagers).

- Frame-transfer CCDs.
- Interline-transfer CCDs.
- Field-interline-transfer CCDs.

While frame-transfer CCDs are seldom used today, interline-transfer CCDs are the type used most in consumer applications. Fig. 7.8.4 shows the basic principle: the charges are sequentially and vertically transferred to a register. Field-interline-transfer CCDs are used for professional applications. CCD sensors suffer from a limited dynamic range and are therefore restricted to few automotive applications.

7.8.4.2 CMOS Sensors and Imagers

Recent improvement in CMOS process technology, pixel architecture, and sensor principles have launched a variety of imager types, both color and monochrome. Some development groups have concentrated on expansion of the pixel-array size, while others worked on improving the dynamic range or the speed of image acquisition. Common to all these developments is the pressure to minimize the diagonal of the pixel array size to benefit from the cost saving of smaller lens systems.

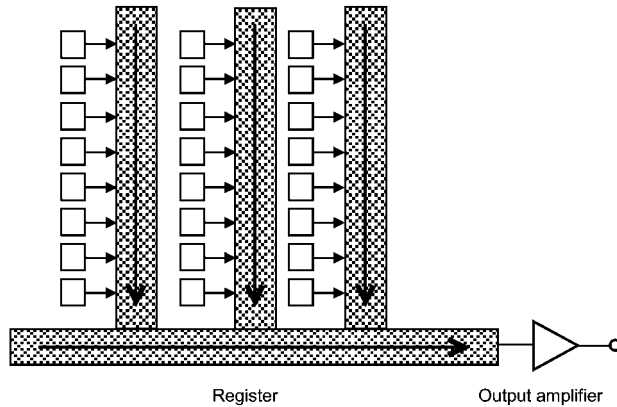


Fig. 7.8.4 Interline-transfer charge-coupled device with column light shields for vertical charge transfer

The human eye shows a nonlinear (close to logarithmic) sensitivity, and it would be helpful to mimic this behavior in a photosensor. Making CMOS pixels with logarithmic sensitivity is particularly easy: one can use the logarithmic relationship between gate voltage and drain current in a MOSFET operated in weak inversion. The resulting pixel architecture is easy to implement in a CMOS process because a pixel consists of just a photodiode and three MOS transistors. A typical photoresponse of about 40 mV per decade of optical input power is obtained with such logarithmic pixels, and their useful dynamic range exceeds 120 dB. The image of a light bulb with such a sensor, for example, shows the glowing filament as well as the background. In systems for night-vision improvement, the camera is often blinded by the headlights of an oncoming vehicle if using CCD chips. With nonlinear CMOS chips the headlights appear as light spots, but without blinding.

The dynamic behavior of such a logarithmic pixel depends on the photocurrent: the darker the scene, the lower the photocurrent, and the longer it takes until this MOSFET is in equilibrium. This is why logarithmic pixels react much slower at low, rather than high, brightness.

The sensing element in the pixel can be built as a photogate or photodiode. Even the use of polycrystalline silicon, deposited on top of the completed chip (TFA, thin film on ASIC), has shown a nearly 100% fill factor.

The opto-electric conversion characteristic can be designed as linear or nonlinear. A linear CMOS imager experiences numerous drawbacks that are typically known with CCDs, especially the limited dynamic range. Several developments of nonlinear imagers with an extended dynamic range have been presented.

7.8.4.3 CMOS Imagers

A CMOS imager consists of a two-dimensional arrangement of photodiodes, each provided with its own selection transistor, as shown in Fig. 7.8.5. For the description of the operation, assume that all photodiodes are precharged to a certain re-

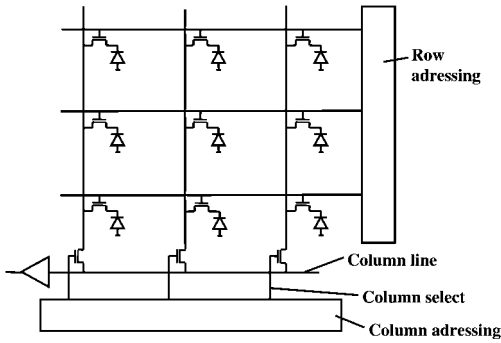


Fig. 7.8.5 Photodiode array image sensor with one photodiode and one selection transistor per pixel

verse bias voltage, typically 5 V. Under the influence of the incident light, each pixel is discharged to a certain level. A pixel is read out by addressing the corresponding row and column transistors, providing a conducting line from the pixel to the output amplifier. Using this line the pixel is charged up again to the same reverse bias voltage as before. The amplifier measures how much charge is required to do this, and this charge is identical to the photocharge (plus dark current charge) accumulated at the pixel site. In this way, each pixel can be read out individually, at random, and the exposure time is completely under the control of the external addressing electronics.

The random addressing freedom, however, comes at the price of a large capacitance of the conducting line between pixel and output amplifier of several pF. As is obvious from equation 1, this leads to noise levels one or two orders of magnitude larger than in corresponding CCD image sensors. For this reason, the use of such traditional photodiode imagers is restricted to applications in which the random pixel access is an absolute must [7].

With the so-called active pixel sensor (APS) imaging technology this noise behavior has been significantly improved by placing the first MOSFET into the pixel. The simplest APS image sensor pixel consists of one photodiode and three MOSFETs. APS technology is very attractive for several reasons.

- It can be produced in standard CMOS technology, opening the way to image sensors with integrated electronic functionality and even complete digital processors.
- The pixels offer random access similar to PD arrays.
- The pixel readout is non-destructive, and it can be carried out repeatedly for different exposure times.
- The exposure times can be programmed electronically.
- APS image sensors dissipate one or two magnitudes less electrical power than CCDs.
- APS imagers show less blooming (spilling of electronic charge to adjacent pixels).

CMOS sensors have achieved high spatial resolution. The pixel size is usually in the region of 5 μm . Resolution or imager size or format is a technical parameter that is only limited by die size and by the desired pixel optosensitivity. A small pixel or a high resolution implies a low sensitivity and vice versa.

Designing an imager for automotive applications necessitates balancing various factors. A high geometrical resolution (e.g., 4 k \times 4 k) can only be achieved by squeezing pixel pitch and in sequence the fill factor and sensitivity. The need for deep submicron CMOS processes shifts the maximum of the spectral response towards the blue end of the visual spectrum with small quantum efficiency.

Image quality strongly depends on the homogeneity of the material parameters over the pixel array that determines the opto-electric conversion function. Sensitivity to low light levels, characterized by quantum efficiency (ratio of the number of electrons collected by the pixel to the number of incident photons), depends on the wavelength of the light: there is a non-uniform spectral response. It also depends on the proportion of the pixel area that is light-sensitive, the 'fill factor'; so design and fabrication of the pixels is crucial. The ultimate limit on minimum resolvable signal is the noise introduced both within the pixel and in the external circuits. Noise comes from many sources.

7.8.4.4 Electronic Signal Detection

The basic task of electronic signal detection is the precise measurement of voltage signals offering low noise levels and a wide dynamic range. The output of the signal detection electronics is usually a voltage that should be proportional to the input voltage over a large dynamic range. An important property of the signal detection electronics is that its output should have very low impedance, that is, the output voltage should be stable and must not depend on the amount of current drawn. The electronic signal detection noise is the present limiting factor in increasing the sensitivity and dynamic range of an image sensor.

For a numerical description of the voltage or charge-noise performance of an electronic circuit, two values are often used: the signal-to-noise ratio (SNR) and the dynamic range (DR). SNR is defined by comparing an actual signal level V with its rms noise ΔV , according to

$$\text{SNR} = 20^{10} \log \frac{V}{\Delta V} \quad (1)$$

DR compares the maximum signal level ΔV_{max} with the minimum rms noise level (ΔV_{min}) in an image sensor typically obtained in the dark

$$\text{DR} = 20^{10} \log \frac{V_{\text{max}}}{\Delta V_{\text{min}}} \quad (2)$$

Three main noise sources can be identified in a source followed MOSFET.

Reset Noise

The channel of the reset transistor exhibits Johnson noise similar to an ordinary resistor. This causes statistical fluctuations in the observed reset voltage levels, which results in effective charge noise ΔQ_{reset} given by

$$\Delta Q_{\text{reset}} = \sqrt{kTC} \quad (3)$$

Flicker Noise

Statistical fluctuations in the mobility and charge carrier concentration of the source follower transistor's channel cause an effective charge noise $\Delta Q_{\text{flicker}}$ described by

$$\Delta Q_{\text{flicker}} \in C \sqrt{\frac{I^A B}{g_m^2 f C_{\text{ox}} WL}} \quad (4)$$

at frequency f , for current I , bandwidth B , transistor length L , and width W , oxide capacitance C_{ox} , process-dependent flicker noise constant A , which is typically between 0.5 and 2, and the transconductance of the transistor g_m .

Thermal Noise

Johnson noise in the channel of the source follower transistor can also be referred back to the input, resulting in thermally generated charge noise $\Delta Q_{\text{thermal}}$ given by

$$\Delta Q_{\text{thermal}} = C \sqrt{\frac{4kTBa}{g_m}} \quad (5)$$

using the same parameters as before.

In practice, both reset and flicker noise can be eliminated by appropriate signal processing, while the thermal noise depends on temperature, and can be reduced only by a reduction in temperature.

The main performance criteria for a CMOS image sensor to be used in an automotive environment are the following.

- Sensitivity.
- Optical detection range.
- Maximum pixel rate.
- Functionality.

The first of these denotes the accuracy of the sensor when converting light into signal. It comprises the photoconversion efficiency or signal generating capabilities of the sensor (pixel size, fill factor, quantum efficiency), and its signal degrading properties (readout noise, non-uniformity, dark current generation). This

whole performance aspect can be suitably compressed into the minimum and maximum light levels the sensor can detect simultaneously.

As many applications demand image rates beyond still imaging, even surpassing video rates, the pixel output speed of the sensor is an important parameter, when combined with the frame size.

7.8.4.5 HDRC Imagers

There are two ways to achieve a logarithmic transfer characteristic in a CMOS imager.

- One is to implement a log amplifier succeeding a linear-responding sensor element.
- The other is to attenuate a linear system response by an exponentially increasing attenuation.

Using log amplification with feedback, the signal bandwidth remains constant while the amplification factor is variable. A small photocurrent causes high amplification factors adding noise to the signal. A small photocurrent affects high gain settings resulting in considerably higher noise. Using the principle of exponential attenuation, the bandwidth is variable. The attenuation factor increases with the photosignal. The signal bandwidth of each pixel depends on the actual illumination. A small photocurrent results in a small bandwidth, and vice versa. Thus, a high noise potential is accompanied by a low signal bandwidth and a low noise potential is paired with a high bandwidth.

Practically this means that high signal frequencies are detected at bright illumination while poor illumination results in low-pass filtering in each pixel. Log amplification inserts an amplification noise proportional to the amplification factor. In contrast, exponential attenuation results in a constant SNR over most of the range because the noise is proportional to the square root of Δf . Fig. 7.8.6 shows the difference in dynamic behavior of a CCD and a nonlinear CMOS sensor.



Fig. 7.8.6 Comparison of the dynamic behavior of a charge-coupled device sensor (left) and a nonlinear complementary metal-oxide semiconductor sensor (right)

7.8.5

Video Camera

Fig. 7.8.7 shows the development steps of the Bosch camera module. A stereo video camera for automotive application can be produced as a compact unit to be mounted at the upper rim of the front window of the car. Fig. 7.8.8 shows the unit. It consists of two camera heads with a nonlinear CMOS sensor each, and the camera optics, both mounted into a housing, which can be hermetically sealed. The camera heads are connected to a powerful computer for image processing. The power required depends on the functions.

7.8.6

Video System

Video technology will first be introduced for to provide functions for driver warning or information. These functions will be transparent to the driver, who will be able to override them. As a second step, functions that intervene with vehicle dynamics through braking, steering, and acceleration, will open up new possibilities for collision mitigation and collision avoidance.

Fig. 7.8.9 illustrates the basic principle of a video sensor. The ECU extracts different predictors from the received image like the lane markings, speed limits, or the distance to objects ahead. These data enable the system to construct a virtual environment around the car. From the specified locations of objects in this environment, possible vehicle interactions can be derived.

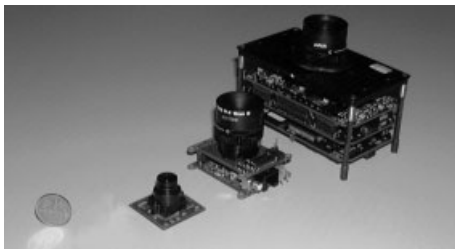


Fig. 7.8.7 Evolution of the Bosch video camera: right, first prototype; middle, second MCM-based generation; left, target size

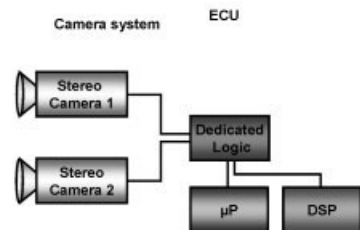


Fig. 7.8.8 Stereo video camera for automotive use and block diagram

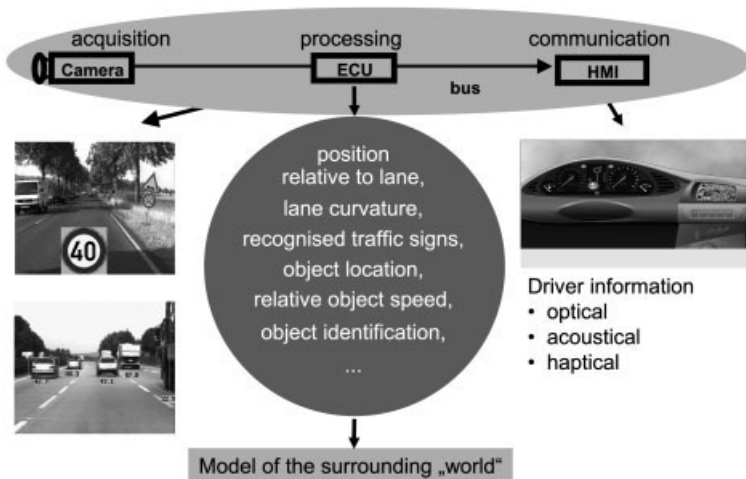


Fig. 7.8.9 Basic principle of a video sensor and functions being considered

The enormous potential of video sensing is intuitively obvious from our knowledge of human visual perception. Although computerized vision has nowhere near achieved similar performance yet, a respectable range of information and related functions can be achieved by video sensing.

- Lane recognition and lane-departure warning.
- Position of car within the lane.
- Traffic sign recognition (speed, no overtaking) with an appropriate warning to the driver.
- Obstacles in front of the car, collision warning.
- Vehicle inclination for headlight adjustments.

7.8.7

Image Processing

Fig. 7.8.10 shows six different steps in image processing. The first step shows the image, as it is seen by the camera. In the second step, relevant parts of the image are extracted from a model or features. Examples are lane detection or the geometry and the trajectory of a traffic sign. This traffic sign is recognized in the third step, from patterns that have been learnt by the system. Step four (which is explained in more detail below) describes the measurement of the position and the outlines of objects and markings, such as vehicles, obstacles, 3D lane curvature. Step five describes the object classification, that is, the discrimination between a vehicle or a pedestrian, for example, while step six describes the most challenging image processing, the interpretation of scenes, predicting the potential movement of other traffic in the vicinity.

The very essence of an environmental sensor for 3D geometry is the measurement of 3D positions in the real world. While this, from a theoretical point of

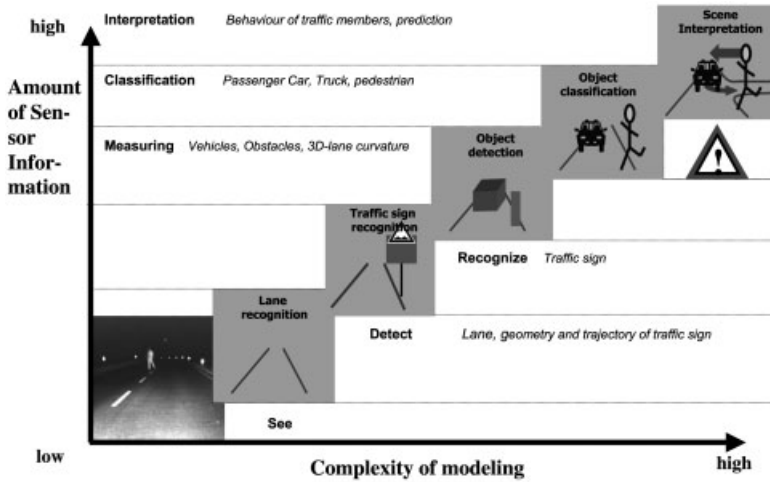


Fig. 7.8.10 Steps in image processing

view, can be achieved with a moving monoscopic camera system [10], automotive practice demands a stereoscopic vision system. Such a system exploits disparity information, as illustrated in Fig. 7.8.11. By identification of points in both images representing the same feature in the real world, that real world position can be reconstructed by triangulation [11]. The figure illustrates this technique for the example of a single pair of corresponding points. The distance Z to an object as depicted in Fig. 7.8.11 is given by

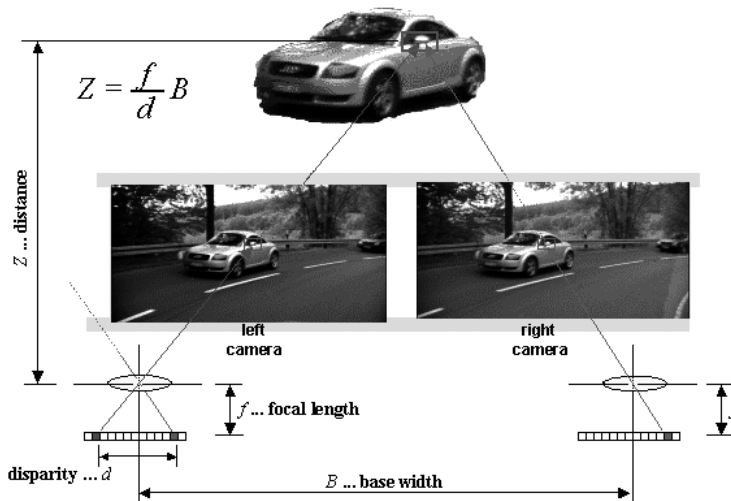


Fig. 7.8.11 Object detection by stereovision

$$Z = \frac{f}{d} B \quad (6)$$

where B denotes the base length of the stereo camera, f the focal length, and d the disparity of an observed point on the left and right camera chip, respectively.

These measurement points may be clustered for road surface detection and obstacle detection. Tracking these measurements over time, as implemented, for example by Kalman filtering, not only yields high reliability and accuracy, but also enables acquisition of information that is not directly observable in an image, such as speed or acceleration.

Beyond 3D geometry, vision sensors are also ideally suited to the detection and classification of 2D illumination patterns. Such patterns frequently possess important information in traffic situations. Traffic signs and lane markings are examples. It is worth noting that hardly any other sensor technology has the potential for acquisition of this type of information. So video technology is the complement to radar technology with its strength of accurate measurement of distance and relative velocity.

7.8.8

Examples

7.8.8.1 Night-Vision System

In 2000, a night-vision enhancement system was introduced to the US market. A specially designed camera, sensitive in the far infrared range of $7000 \mu\text{m}$, picks up an image. Warm objects appear light, while cold objects are hardly visible. FIR systems suffer from rather low resolution, and the difficulty that objects at ambient temperature are not visible to the camera.

Another approach, illustrated in Fig. 7.8.12, uses modified halogen headlamps radiating near infrared (NIR) radiation with a characteristic comparable to a visible high beam. CCD or CMOS cameras are sensitive in the NIR wavelength range, and can pick up the 'illuminated' scene in front of the car as it is seen by the driver. The camera image can be shown on a head-up display projected on the windscreen of the vehicle. In a second step, when object detection algorithms are available, objects can be detected and an appropriate warning can be given to the driver.

7.8.8.2 Lane Detection and Lane-Departure Warning

Lane-departure warning (LDW) systems can be operated with a mono or stereo camera. As shown in Fig. 7.8.13, the camera searches for lines in front of the vehicle. The upper image shows the camera image with search lines (see detail in the second image) and the crosses in the upper image show the lane course, which is calculated by the image-processing computer. To detect a line, the luminance signal within the search line is analyzed and, using a high-pass filtering algorithm, the edges of the line are detected. From this signal a warning can be derived if the driver is crossing the lane mark without having set his turning lights.

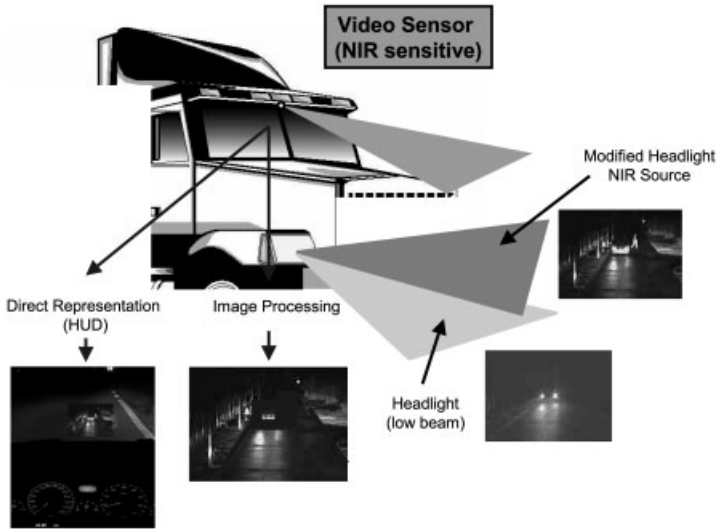


Fig. 7.8.12 Night-vision improvement system using near infrared radiation from modified headlamps [13]

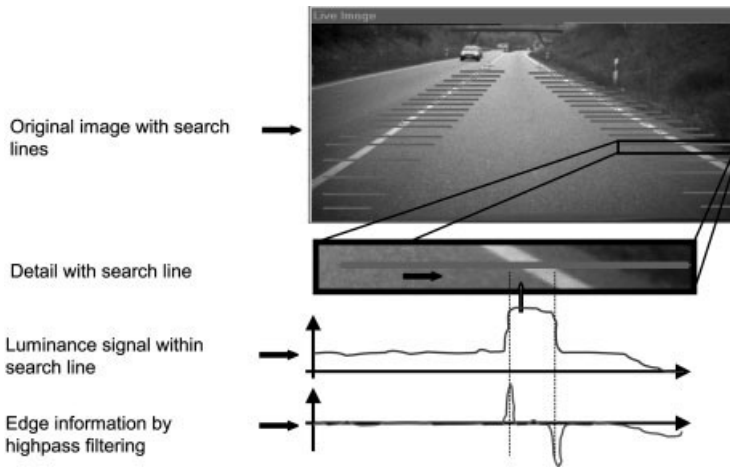


Fig. 7.8.13 Principle of lane detection

This warning may be acoustic or haptical giving a momentum on the steering wheel in the opposite direction.

7.8.8.3 Traffic-Sign Recognition

A traffic-sign recognition system consists of the above-described video system but with a mono camera. The image from the camera is shown in Fig. 7.8.14. While driving, the camera searches permanently for objects that could be a traffic sign.



Fig. 7.8.14 Traffic sign recognition: camera image with tracking lines and recognized signs

When such an object is detected, the object is tracked until the resolution of the camera chip is sufficient for reading the sign. Reading in this case means a comparison between the recognized sign and a sign library that has previously been learnt by the system. The recognized traffic sign can then be displayed on a small graphic display located in the instrument cluster or in the center console. It is easy to add an acoustic warning when the car speed is higher than the speed limit. Fig. 7.8.14 shows the camera image and an example of the information that could be displayed in the instrument cluster.

7.8.8.4 Object Detection for ACC Stop & Go

Object detection needs two cameras to achieve good performance under all conditions. Mono approaches suffer from poor reliability under extreme lighting conditions. In an advanced algorithm state, as it is shown in Fig. 7.8.15, the lanes as well as the objects are detected. Objects in front of the vehicle are recognized by the system, and their distance and a rough estimation of the size are calculated. The distance to objects in front and their relative speed is plotted in numbers, and the estimated outlines of the obstacles are indicated by squares depicted around the outlines of the appropriate vehicle. As can be seen in the image on the right, also the radius of the curve can be calculated ($R=1.389$ m). These data can be favorably be fused with the data from long-range radar allowing the so-called 'stop & go' feature, even in urban areas.

7.8.9

Summary and Outlook

Sensors for environmental sensing are under development and in rapid progress. Some applications are already on the market, others will follow soon. Meanwhile the complexity of the functions will grow steadily. The first significant step will be made with the introduction of a 24 GHz platform fulfilling a multi-usage concept for the sensors. The second step will follow a couple of years later when new CMOS imagers will be available. In the sequel, a significant extension of functions is expected. Ultrasonic reversing systems will stay on the market but will be

Extraction of information in the environment of the vehicle by evaluating image sequences

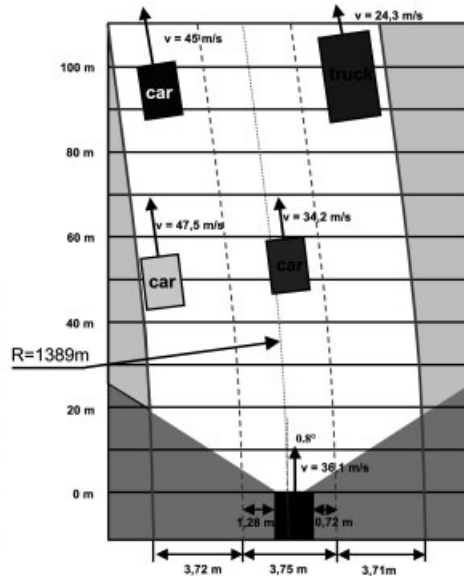
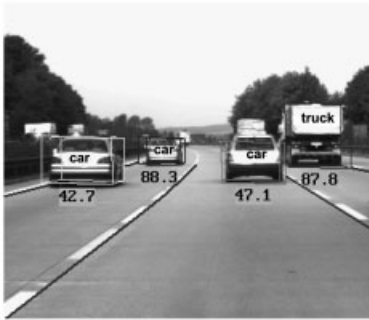


Fig. 7.8.15 Object detection: left, camera image with information about distance and size of obstacles; right, detailed information about relative speed and curve radius

used mostly in cheaper cars, while up-market cars will be equipped with microwave and video technology. In parallel to this scenario, sensors for night-vision systems will be developed further, and will exhibit a broadening distribution as lower price levels are achieved. Beyond the year 2005, the large complexity of functions and variety of sensors that is expected by then, will be accounted for by employing comprehensive sensor fusion [12].

7.8.10

References

- 1 BROGGI, A., BERTOZZI, M., FASCIOLI, A., CONTE, G., *Automatic Vehicle Guidance: The Experience of the ARGO Autonomous Vehicle*, World Scientific, Singapore, 1999.
- 2 DICKMANN, E. D., CHRISTIANS, T., in *Proc 2nd Conf. Intelligent Autonomous Systems*, IOS Press, Amsterdam, Netherlands, 1989.
- 3 SHLADOVER, S. E., *IATSS Res.* 1997, 21 (2), 40–48.
- 4 STILLER, C., HIPPEL, J., RÖSSIG, C., EWALD, A., *Image Vision Comput. J.* 2000, 18 (5), 389–396.
- 5 THORPE, C. E., *Vision and Navigation: The Carnegie Mellon Navlab*, Kluwer, Dordrecht, 1990.
- 6 *Strategy Analytics*, 2001, Dec.
- 7 JAEHNE, B. (ed.) *Handbook of Computer Vision and Applications, Vol. 1*, Academic Press, 1999.
- 8 SALEH, B., TEICH, M. *Fundamentals of Photonics*, Wiley, New York, NY, USA, 1991.
- 9 THEUWISSEN, A., *Solid State Imaging with Charge Coupled Devices*, Kluwer, Dordrecht, 1995.

- 10 ENKELMANN, W., GENGENBACH, V., KRÜGER, W., RÖSSLE, S., TÖLLE, W., in *Proc. Intelligent Vehicles '94 Symposium*, Oct 1994, pp. 99–102.
- 11 HARTLEY, R. I., STURM, P., *Computer Vision and Image Understanding*, 1998, 68 (2), 146–157.
- 12 SEGER, U., KNOLL, P. M., STILLER, C., in *Proc. Sensor Vision and Collision Warning Systems, Convergence Conference*, Detroit, 2000.
- 13 Photographs: Daimler Chrysler Research Institute, Ulm.

7.9

Wheel-Speed Sensors

MATTHIAS MÖRBE and GÜNTER ZWIENER

7.9.1

General Applications

Rotational speed sensors located at the vehicle wheel (either as a specific design or integrated into the bearing) are based on ABS (antilock braking system) development. To avoid a wheel locking, the control algorithm has to know the individual wheel speed in real-time conditions [1]. With additional functions of brake modulation systems like traction control (ASR=acceleration slip regulation) or vehicle stabilizing systems (ESP=electronic stability program) the importance of the wheel speed has increased.

This is because ASR and ESP systems need to know the wheel speed under all driving conditions. ABS uses the calculated vehicle speed and compares it with the wheel speed, only in emergency braking situations.

Wheel-speed sensors do not monitor speed directly; they sense the movement of the circumference of the tire. This is done with 48 or 32 pulses per turn, and some correction factor for tire sizes. This information comes from driven and non-driven wheels, and means four independent sources of information describing road surface, friction coefficient, cornering etc. In all cases the information is passed through the electronic control unit (ECU). ABS and all advanced systems are safety systems that do not allow the electrical signal to be split before it goes to the ECU. This is necessary to make a fail-safe system.

Other systems using wheel-speed information include the following.

- Transmission control.
- Engine management.
- Odometer.
- Distance measuring units (for taxis).
- Navigation systems.
- Windscreen wiper speed.
- Brake-force distribution.
- Roll-over protection

and all kinds of new features like 'stop and go', 'hill hold control', 'roll over mitigation', and collision avoidance systems.

7.9.2

Requirements

Each system has its own specific sensor requirements [2]. There are two main categories of wheel-speed sensor: passive and active.

7.9.2.1 Passive Sensors

All inductive principles and designs are covered by this term. Passive sensors do not need a power supply; the signals are generated with coils, magnets, and a cog. The change of magnetic flux density by the cog generates a current in the coil. The principle is shown in Fig. 7.9.1. Fig. 7.9.2 shows how airgap, modulation, and signal modulation are defined. As all mechanical modulations will generate electrical signals these have to be filtered at the input stage of the ECU.

7.9.2.2 Active Sensors

Active sensors require an external power supply for operation. This is the 5 or 12 V system of the ECU. The main technologies are the complementary metal-oxide semiconductor Hall-effect transducers (CMOS-Hall) or AMR (anisotropic magnetoresistive effect) plus silicon amplification [3].

7.9.2.3 Summary of Requirements

The requirements can be summarized as follows.

- Real-time pulses.
- Low jitter.

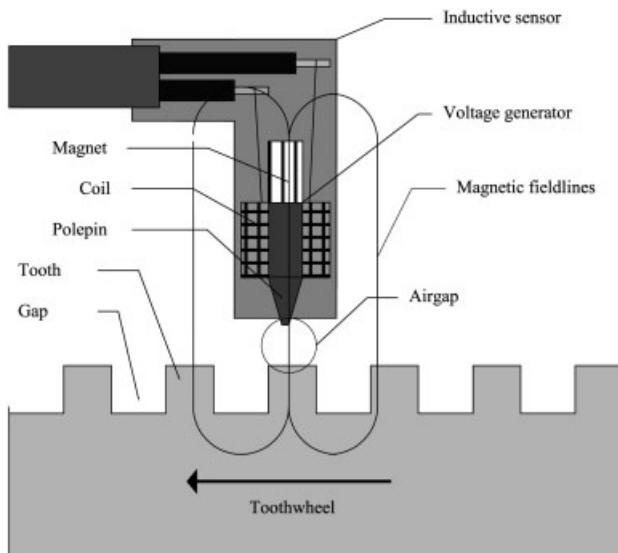


Fig. 7.9.1 Structure of passive sensors

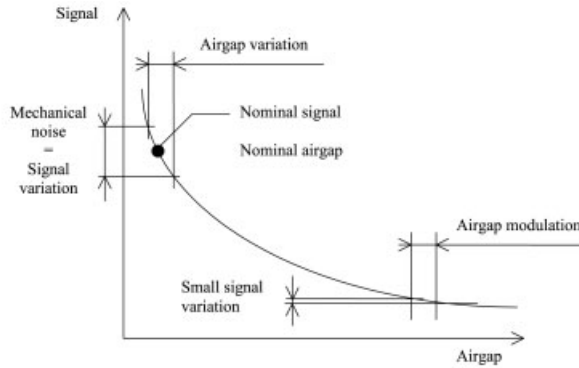


Fig. 7.9.2 Signal output and airgap modulation

- Large noise-to-signal ratio.
- Small size.
- Zero-speed signal.
- Robustness of mechanical design.
- Rotational speed direction.
- Phase correctness.
- True power on performance.
- High-temperature durability.
- Simple application conditions.
- Flexible design variations.

After more than 25 years of development, prices are now at a low point for this mature product group. It is foreseen that passive sensors will be replaced by active ones because of their zero speed and rotational direction capabilities.

7.9.3

Electrical Circuits

7.9.3.1 Circuits for Passive Sensors

Passive sensors operate as generators delivering an ac current, so a simple solution would be to use a differential amplifier. However, in automotive application there is no negative supply source available. This means that a special circuit is necessary. The increasing demand for immediate failure detection and diagnosis requires line monitoring. Any interruption in the cable harness, sensor, or ECU can be detected for each wheel separately. This is not possible just by looking at incoming pulses because at high speed the inductive sensor will generate a high voltage, which can trigger the sensitive input stage via a parasitic capacitance to ground.

The main compromise within all circuit designs is the sensitivity. The signal from the sensor at minimal speed has to exceed $U_{\text{trigger}} = \pm 35 \text{ V}$ with a tolerance of $\Delta U_{\text{trigger}} = 7 \text{ mV}$. To gain more input signal the passive part of the input circuit has to be designed carefully. Typical values are shown in Fig. 7.9.3 and Fig. 7.9.4.

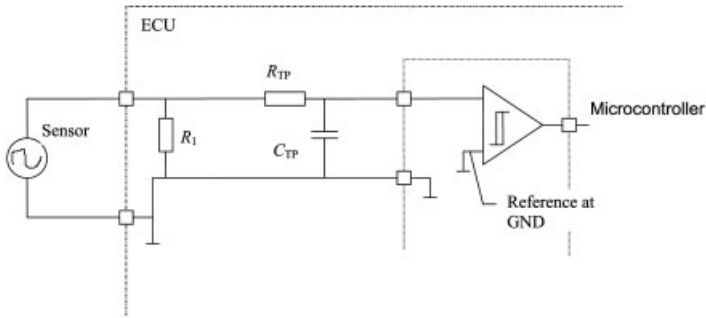


Fig. 7.9.3 Standard input circuit for passive sensors

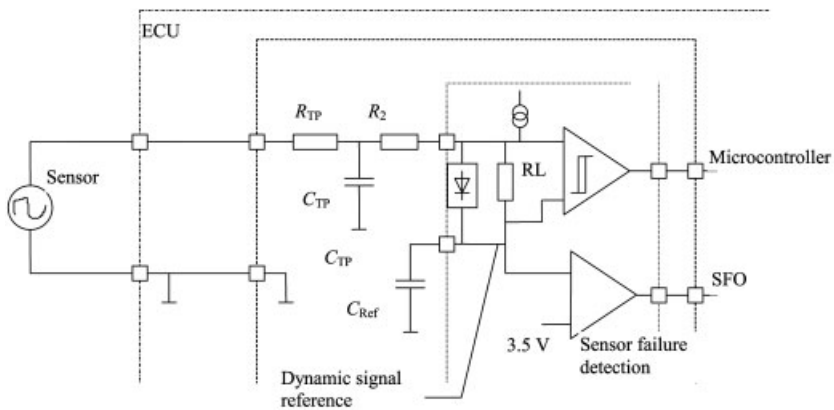


Fig. 7.9.4 Input circuit with dc current monitoring

For all calculations standard electrical laws can be used. To simulate a sensor including the magnetic circuit a sophisticated model and simulation tool is necessary. In all cases verification under real conditions is mandatory.

7.9.3.2 Circuits for Active Sensors

The input circuit of an ECU for active wheel-speed sensors has to provide

- Power supply.
- Signal amplification.
- Signal monitoring.
- Failure detection.
- Power-supply protection.

Power Supply

There are two variants of power supply for wheel-speed sensors on the market.

- 5 V regulated voltage from ECU.
- 12 V battery supply with protection in the ECU.

A 5 V supply has the benefit of less effort in sensor circuit design but all line disturbances in the cable harness are valid. This means it has the same specification for a technology as a 12 V battery supply. In terms of power consumption the sensor does not have the same range as with 12 V.

Battery voltage is $U_{\text{bat}}=5\text{--}24.5\text{ V}$, with a constant current output signal of $I_{\text{low}}=7\text{ mA}$ and $I_{\text{high}}=14\text{ mA}$ the sensor has to withstand a much wider power swing than with a 5 V regulated supply. For a realistic calculation and simulation, the following parameters are required: chip size, maximum junction temperature, thermal resistance of molding material, packaging in the hub or bearing, length and diameter of cable, and other conditions like post heating after long downhill braking. It is very important to verify all calculations with measurements during endurance testing.

Signal Amplification

After long discussions in the automotive and supplier industry the two-wire constant-current interface with switched levels as digital output is now the international standard. Fig. 7.9.5 shows the principle of an active sensor.

The sensor has an internal supply current for operation. This current is always smaller than 7 mA. Via current control circuits it is increased to this low output level and doubled to 14 mA for the high stage. Most circuits have an internal voltage regulator and in some cases also an external capacitor for protection purposes. There are two variants of input arrangements on the market. In one design all sensors are connected to ground and the other side to the input stage. This stage can be designed as a current mirror to avoid a sensing resistor. The other design is that all sensors have a common source at the battery line and the output current is passed to a sensing resistor in the ground line. This has the advantage that the ground pin of the ECU is protected to short circuits to 12 V lines in the cable harness. A typical input stage with sensor is shown in Fig. 7.9.6.

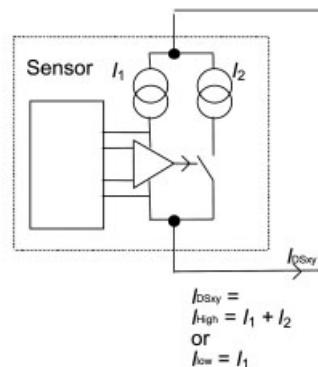


Fig. 7.9.5 Diagram for active sensors with current interface

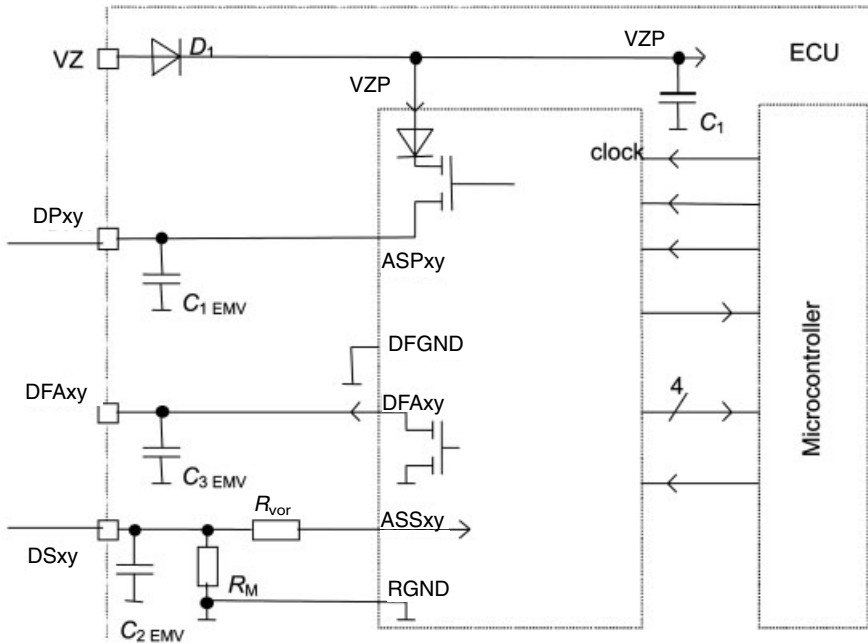


Fig. 7.9.6 Active wheel-speed sensor input stage

Signal Monitoring and Failure Detection

The output current tolerance and all other tolerances define an operational range as shown in Fig. 7.9.7.

By use of comparators, levels below or above those defined operational bands can be monitored for electrical failures. The switch in the supply line in Fig. 7.9.7 enables a short circuit to any other sensor line to be detected at 'power on'. A crossover between sensor lines would generate false output signals and as a consequence a false speed calculation. This is very critical for safety systems.

Passive sensors generate high output signals at high speed; at low speed the limit is defined by the maximal airgap. This information can be used to monitor the mounting condition in the car. In case of a line failure the signal itself could be high enough that in conjunction with a parasitic resistor or capacitor the input stage is triggered. To monitor those failures the system algorithm is used.

Active sensors allow for more failure monitoring. There are several operating conditions under control.

- Maximal airgap as warning level.
- Airgap too large as out-of-range signal.
- Airgap in defined range as original equipment check point.
- Rotational direction.
- Zero-speed detection after a defined standing still phase.
- Status of signals: valid or not valid.

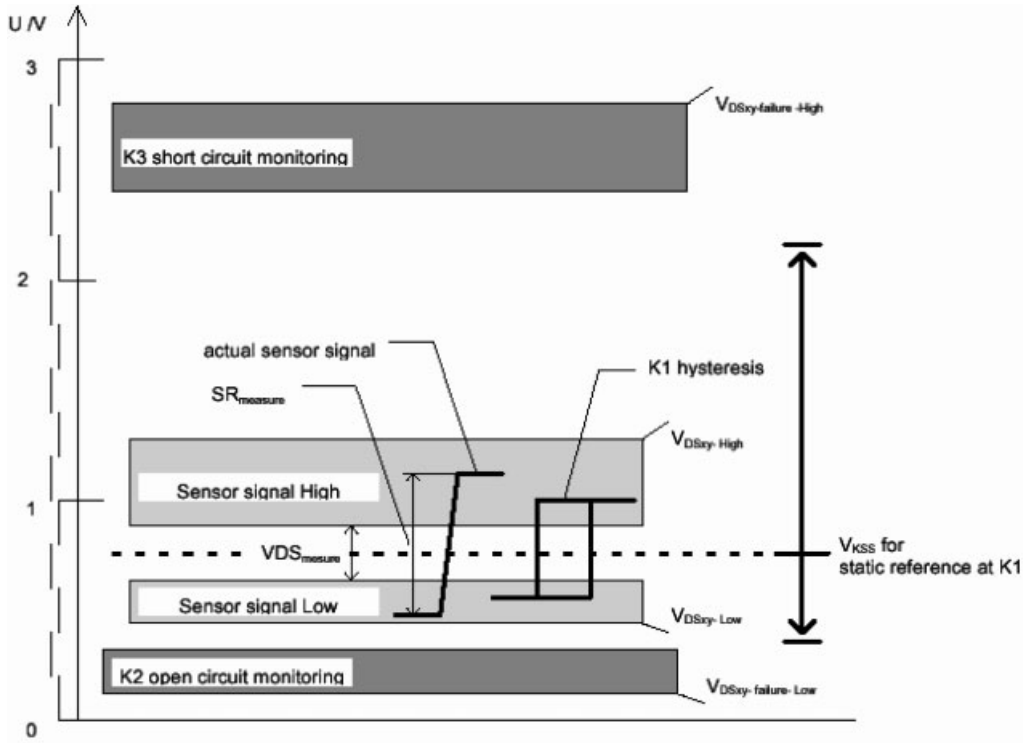


Fig. 7.9.7 Input operational range and signal monitoring

The major purpose of this information is to give an intelligent monitoring algorithm the chance to assess whether it is true signal or a failure. This requires intelligent filtering to allow for a high system availability.

7.9.4

Wheel-Speed Sensor Types

We distinguish passive and active wheel-speed sensors (WSS) according to their functional principle. The passive WSS generates a signal voltage using the variable reluctance. Magnetic field modulation is done by a cog wheel rotating with the speed of the vehicle. No external power supply is necessary.

Active WSS work on galvanomagnetic principles like the Hall or MR (magneto-resistive) effect. The sensor needs an external power supply. Compared with the passive WSS, the signal (amplitude) of active WSS depends not on the speed. The output level is constant.

A widespread output protocol is the two wire/two current interface, which transmits the signal by a two wire cable. A rectangle signal with I_{low} and I_{high} signal level is transmitted to the ABS ECU. For additional information transmission,

there are two current PWM (pulse width modulated) interfaces and three current serial data protocol (Manchester code) interfaces available.

The wheel-speed sensors have to cover the requirements of the car manufacturer both for function and for fitting. The sensor must fit in the specific axle and wheel bearing application. Use in demanding ambient conditions must be considered, so it is necessary to use suitable materials for the sensor housing withstanding a wide temperature range from -40 to $+150^{\circ}\text{C}$ and showing good resistance to chemicals used in a car and coming from the environment. The sensor cable and the connector must also fulfill these requirements.

The specific sensor construction takes into account the space requirements, the possible assembly hole dimension, and the sensor reading types: bottom read and side read. Therefore a big selection of sensor versions regarding the housing geometry are available. Fig. 7.9.8 gives an overview of possible versions for passive and active wheel-speed sensors.

The comparison of the housing geometry of active and passive sensors clearly shows the reduction of sensor dimension and weight reduction. The applied ASICs enable a large family of application-specific constructions because of their minimal packaging. Fig. 7.9.9 shows examples for further configurations.

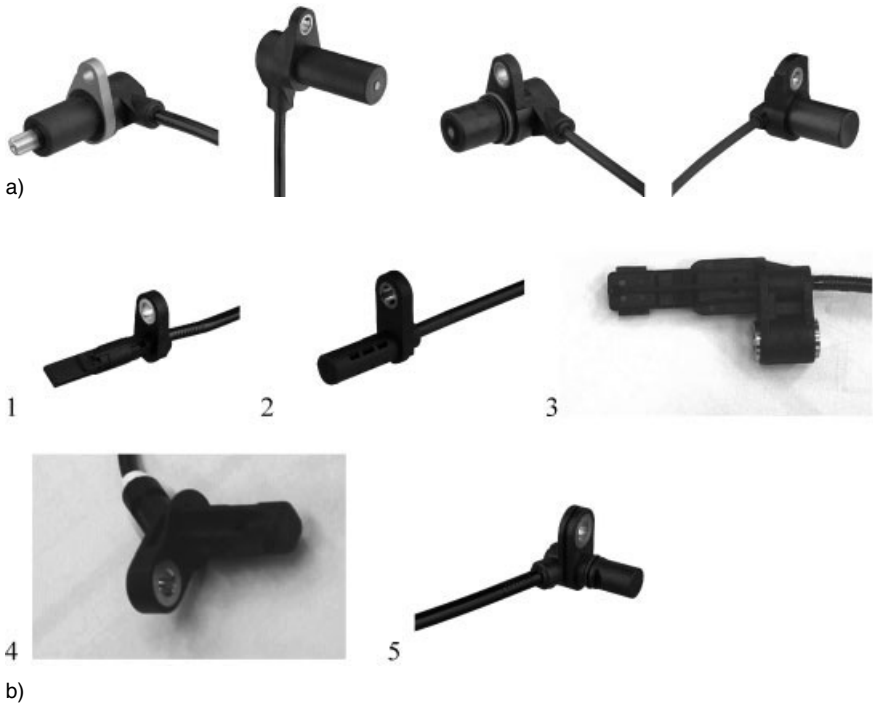


Fig. 7.9.8 a) Examples of passive wheel-speed sensors for steel tone wheel applications. b) Examples of active wheel-speed sen-

sors for steel tone wheel or multipole tone wheel application: 1, 3, side-read type; 2, 4, 5, bottom-read type

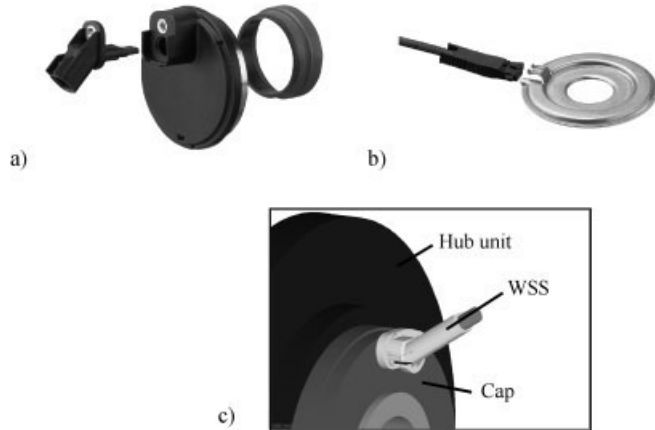


Fig. 7.9.9 a) Cap sensor consists of direct connected wheel-speed sensor, cap, and multipole tone wheel; b) clip sensor with depending fixture (without screw); c) cap-ring sensor

The cap sensor includes a direct connected sensor and a protection cap mounted to the hub. The clip sensor can be mounted by its special geometry without a flange and screw. Present wheel-speed sensors are manufactured by overmolding the sensor elements (ASICs) directly with polyamide thermoplastic material, which fulfills the current needs. The requirements regarding temperature will increase for future car projects. This means a challenge for the development of materials and joining methods.

7.9.4.1 Installation Conditions: Arrangement to Tone Wheel

The wheel-speed sensor can be arranged to the tone wheel in different ways. Fig. 7.9.10a, shows a wheel bearing where the multipole tone wheel is a part of the bearing seal. The sensor is a side-read type. The reading principle is the axial reading type (sensor reading point parallel to the rotating tone wheel axis). Fig. 7.9.10b shows an inbearing solution, where the bearing includes the tone wheel. Different mounting positions are possible. Fig. 7.9.11, shows three examples considering the axle and the bearing. The application of steel tone wheels and magnetized multipole tone wheels is possible.

7.9.4.2 Wheel-Speed Sensor Installation, Cable Layout, and Connector

Wheel-speed sensors are available with or without (direct connected sensors) cable. Cable routing in the car is important and many types of sensor configuration (cable layout) are possible. Grommets and clips can fix the cable to the car body. Expansion bellows of plastic protect the cable from damage. Fig. 7.9.12 shows a complete WSS consisting of sensor head, connector, grommets, clips, and expansion bellows. The fixture to the car body with clips and grommets is shown in Fig. 7.9.13.

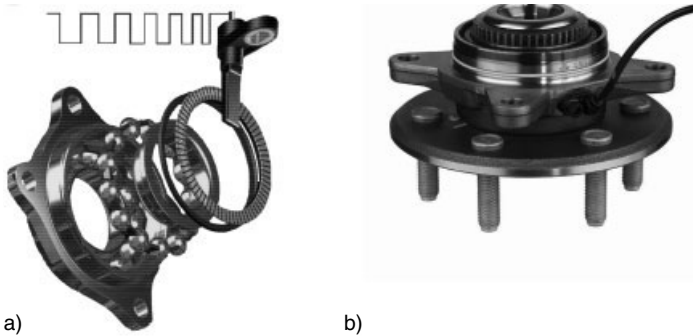


Fig. 7.9.10 a) Wheel bearing with integrated multiple tone wheel; b) wheel-speed sensor with in-bearing application (tone wheel not visible)

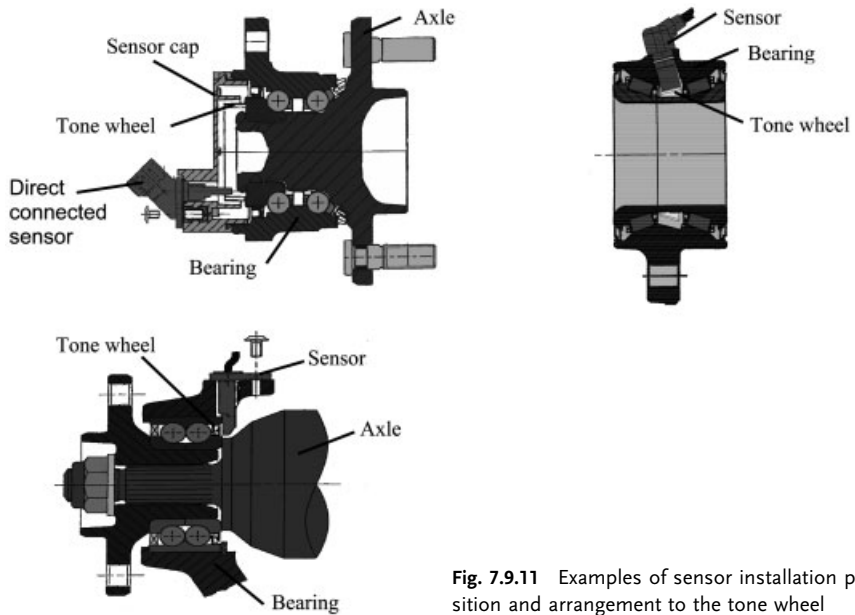


Fig. 7.9.11 Examples of sensor installation position and arrangement to the tone wheel

Joining the wheel-speed sensor to the cable tree is done with a connector. The demands on the connector are low contact-resistance, safe connection by connector pins under the influence of dynamic impact, and pin material resistant to corrosion. Depending on the installation area, the further requirement of tightness may be necessary. Fig. 7.9.14 shows three examples of connectors.

7.9.4.3 Test Equipment for Wheel-Speed Sensors

In addition to WSS testing and monitoring by the ABS control unit, the sensor function must be checked during manufacturing and service. The WSS manufac-

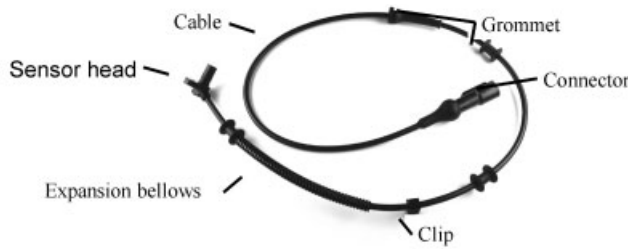


Fig. 7.9.12 Wheel-speed sensor with fixings on the cable

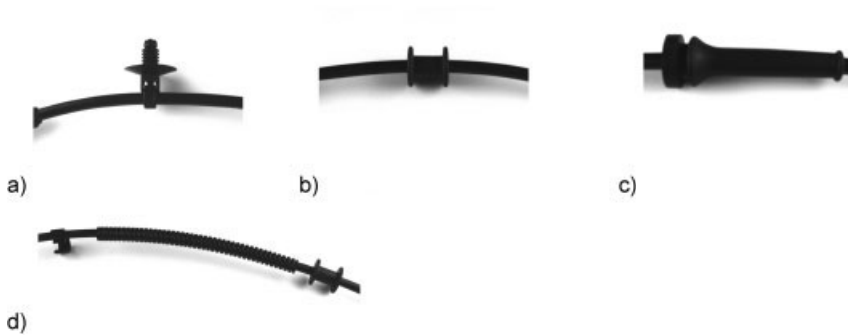


Fig. 7.9.13 Fixing parts to attach the sensor cable to the car body: a) cable clip; b) grommet; c) grommet; d) expansion bellows of plastic for cable protection

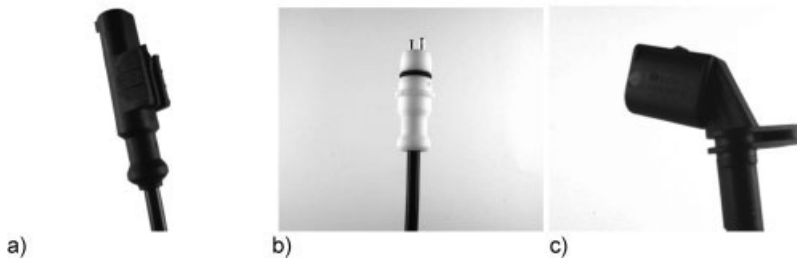


Fig. 7.9.14 Examples of wheel-speed sensor connectors: a) and b) molded connector; c) direct connection at the sensor head

turer offer special test equipment to test the sensors, so that output levels (voltage, current levels) and the additional function can be checked. Most of the test equipment uses the original input circuit of the control unit, in order to have same conditions.

7.9.5

Functional Limits**7.9.5.1 Angle Hysteresis**

The function is limited by hysteresis. The alternating magnetic field change caused by the rotating toothed or magnetized tone wheel (Fig. 7.9.15, sinusoidal signal), leads to an alternating electric signal to the sensor circuit. The signal is then formed by a comparator circuit into a rectangular signal (Fig. 7.9.15, rectangular signal) with defined switching hysteresis thresholds (Fig. 7.9.15, thresholds $+B_{\text{hysteresis}}$ and $-B_{\text{hysteresis}}$). Reaching the thresholds is necessary to switch the sensor output. Moreover the hysteresis denies unwanted switching. Unwanted switching can occur by vibration of the tone wheel or signal noise and lead to an evaluable signal for the control unit.

The switching hysteresis can be compared with a defined angle of rotation regarding the tone wheel. So there is a hysteresis regarding the angle: angle hysteresis. Fig. 7.9.16 shows the magnetized encoder. The movement of the tone wheel

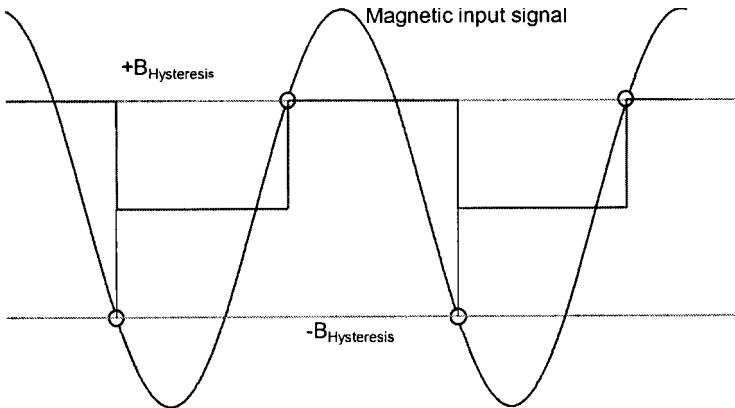


Fig. 7.9.15 Magnetic sensor input signal modulated by the tone wheel rotation (blue line). Switching thresholds for defined switching to prevent unwanted switching (green line). Generated sensor output signal by the comparator circuit (red line)

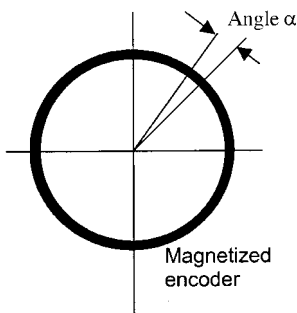


Fig. 7.9.16 The magnetized encoder. The movement of the tone wheel covers the angle α that describes the hysteresis. The circuit design has to specify the hysteresis thresholds to guarantee disturbance-free function

covers the angle α that describes the hysteresis. The circuit design has to specify the hysteresis thresholds to guarantee disturbance-free function.

7.9.5.2 Airgap and Dynamic Airgap Deflection

An important factor for the installation of WSS is the airgap between the tone wheel surface and the active area of the sensor housing. This is the mechanical airgap. The magnetic airgap is measured from the tone wheel surface to the sensing element (ASIC) in the sensor housing. Of course the axle and sensor dimensions and their tolerances have to be taken into account. So the arrangement of tone wheel and sensor can give a minimum and maximum airgap.

For correct sensor function, the magnetic conditions must be demonstrated. Knowledge of the magnetic distance law in the airgap range is necessary to calculate the safety margins regarding temperature influence, tone wheel differences, and sensitivity deviation of the sensor ASICs coming from the semiconductor process.

The specific mechanical application airgap is static, but during driving an airgap deflection occurs. The deflection depends on the specific car and of course on the dynamic forces on the axle. Fig. 7.9.17 shows the arrangement WSS/tone wheel. The deflection D in the X direction changes the airgap. Owing to the behavior of the magnetic field, the signal strength changes. Current ASICs can handle these by internal circuit measures.

7.9.5.3 Direction Detection

Active wheel-speed sensors have been available for years. New features in modern cars to enhance comfort require more information about the behavior of the car. One important piece of information is the rotational direction of the wheels, and hence the direction of the car. The Bosch DF11i was the first sensor to deliver this information, and it was used in the Bosch EHB system in 2001. The sensor transmits this additional information by modulation of the pulse width of the speed signal. The sensor contains three Hall elements: two for speed measurement by the difference principle, and a third arranged in such a way that an addi-

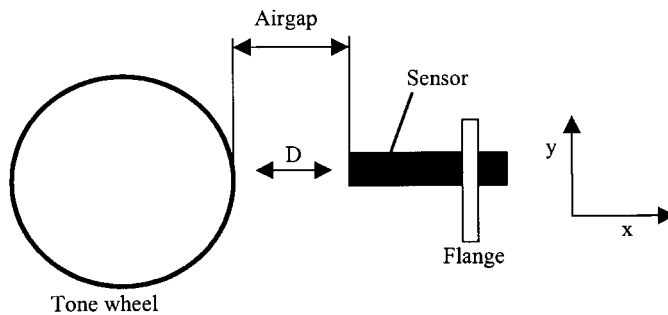


Fig. 7.9.17 Arrangement of tone wheel and wheel-speed sensor, airgap range, possible deflection D , and deflection direction

tional voltage is generated with a phase shift. Evaluating the sign of the phase shift and comparing it with the signal phase and the speed-signal phase, gives the direction of the rotation.

7.9.6

Future Development Trends and Objectives

Wheel-speed sensors are traded as a commodity so a good price is the main decision factor. For easier applications an airgap up to 3.5 mm could improve mounting conditions but it would be a compromise in terms of higher sensitivity against noise and false magnetic fields. If self-calibration is built-in, failure monitoring would be simplified over long periods, and there would be service advantages. The target has to be increased reliability and component quality, including the sensor head, connector cable, and all manufacturing processes. Operation at temperatures above 170°C, more information and signal conversion, and transmission from other wheel-based systems will increase the value of wheel-speed sensors in all vehicles in the future.

7.9.7

References

- | | |
|--|---|
| <p>1 Bosch, <i>Fahrsicherheitssysteme</i>, 2nd edn. Vieweg, Wiesbaden 1998.</p> <p>2 Bosch, <i>Kraftfahrtechnisches Taschenbuch</i>, 24th edn. Vieweg, Wiesbaden 2002.</p> | <p>3 V. GUSSMANN, D. DRAXELMAYER, J. REITER, T. SCHNEIDER, R. RETTIG, SAE Technical Paper 2000-01-C058.</p> |
|--|---|

7.10

Cam and Crank-Angle Sensors

IWAO YOKOMORI and YASUTOSHI SUZUKI

7.10.1

Introduction

The cam sensor and the crank sensor are installed opposite to the signal rotor provided on the cam axis and the crank axis. Fig. 7.10.1 shows the crank sensor installation position. The cam sensor detects information for cylinder identification from the cam axis, while the crank sensor mainly detects information on engine rpm and piston position from the crank axis. Nowadays, they are also used to detect misfiring caused by rpm variations.

The first crank sensor was put into service in parallel with the distributor-less ignition (DLI) system [1]. The use of crank sensors has increased along with the increasing use of the DLI system since the mid 1980s. Under OBD-II, a new on-board diagnosis standard, introduced in the US in the 1990s, manufacturers are

Fig. 7.10.1 Crank sensor installation position

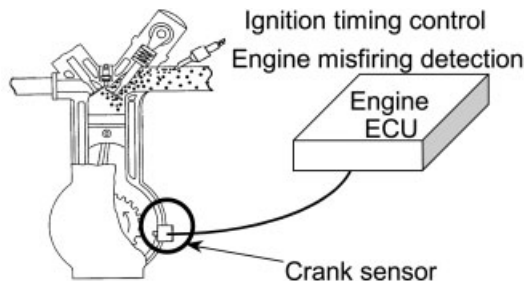
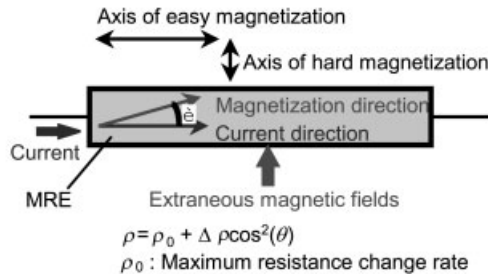


Fig. 7.10.2 Ferromagneto-resistive-effect sensor principle



obliged to adopt systems capable of diagnosing misfiring and alerting the driver; this accelerated the adoption of the DLI system, leading to a further increase in use of the crank sensor.

Cam and crank sensors are classified by the type of sensing element: MRE (ferromagneto-resistive effect) sensor, Hall (Hall effect) sensor, and MPU (magnetic pick-up coil) sensor. Tab. 7.10.1 provides a summary of the principles and characteristics of the cam and crank sensors according to which element is employed [2].

The MRE effect, peculiar to MRE films such as Ni-Fe, was discovered by William Thompson and Lord Kelvin in 1856. Fig. 7.10.2 shows the principle of the MRE sensor.

The Hall sensor utilizes the Hall effect, discovered by Edwin Hall in 1879. Fig. 7.10.3 shows the principle of the Hall sensor.

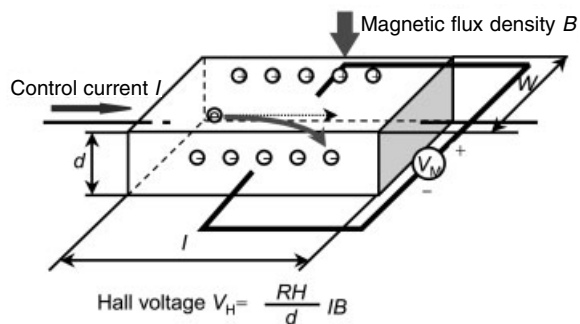
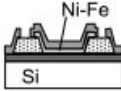

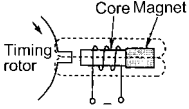
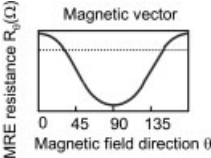
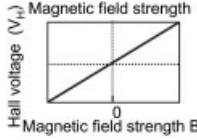
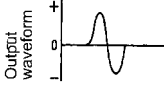


Fig. 7.10.3 Hall-effect sensor principle

Tab. 7.10.1 Principle and characteristic of cam and crank sensor

Principle	MRE (NiFe) Ferromagnetoresistance effect	Si Hall Hall effect	MPU Electromagnetic induction
Structure			
Sensor characteristics			
Sensitivity	High	Low	Low
Very low speed detection	Possible	Possible	Impossible
Integration	Easy	Easy	Difficult
LSI process	Easy	Easy	Difficult

The MPU sensor utilizes electromotive force generated in a coil by electromagnetic induction. Fig. 7.10.4 shows the principle of the MPU sensor.

Tab. 7.10.2 gives example specifications for the cam and crank sensor.

Since the cam and crank sensor is used in the engine compartment, it is required to operate over a wide temperature range. Power is supplied by a battery, which means a high maximum voltage rating. The sensor is also required to be accurate to within $\pm 1^\circ$ with a large air gap at rotor rpm more than 0.

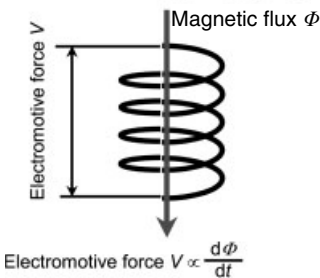


Fig. 7.10.4 Magnetic pick-up coil sensor principle

Tab. 7.10.2 Example specifications for cam and crank sensor

Parameter	Specification
Operating temperature	-40 to 135 °C
Detection gap	0.5 to 1.5 mm
Rotor RPM	More than 0 rpm
Absolute angular accuracy	±1 °
Maximum voltage rating	8 to 16 V
EMC	200 V/m
ESD	±25 kV

7.10.2

Technologies

Here the MRE crank sensor is given as an example; the technologies employed to achieve the sensor specifications given in Tab. 7.10.2 are described, in terms of the magnetic sensing element, the signal-processing circuit, and the package.

7.10.2.1 **Magnetic Sensing Element**

The magnetic characteristics of the MRE are determined by its material, film thickness, trace shape (width and length), and pattern angle. To make magnetostriction equal to zero, we select a permalloy (Ni:Fe, 81:19) as the material, and film thickness and trace shape so as to attain minimum saturation magnetic field strength and maximum sensitivity. Our MRE device structure features an MRE provided directly on the Al electrode to simplify the wafer process as much as possible. Fig. 7.10.5 shows a cross section of the MRE. The Al electrode is about 1 μm thick and is gently tapered so that NiFe film several tens of nanometers thick can be deposited on the Al with good step coverage.

The requirements for MRE device are as follows.

- The MRE device must be subjected to a magnetic field equal to or stronger than the saturation magnetic field strength of the device.
- The MRE device must accommodate various kinds of gears.

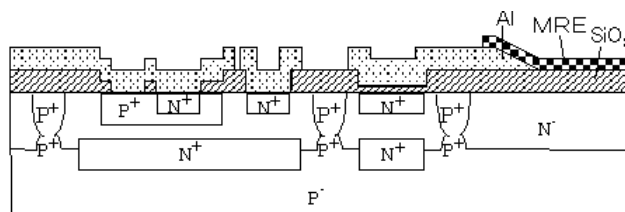


Fig. 7.10.5 Cross section of ferromagneto-resistive element sensor

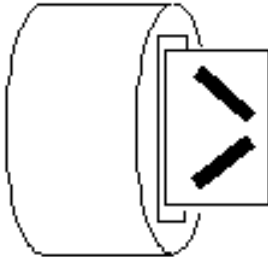


Fig. 7.10.6 Hollow-body magnet structure

We select a hollow-body magnet structure (a cylindrical magnet with a cavity in the middle) so that the MRE device may be subjected to a magnetic field equal to or stronger than the saturation magnetic field strength of the device. Fig. 7.10.6 shows a hollow-body magnetic structure. The hollow-body magnetic structure, with the close-set magnet and MRE device and strong magnetic strength, provides an advantage in the design of magnetic circuit.

The pattern angle varies depending on the magnetic circuit to be used. This is because the MRE sensor, placed in a magnetic circuit, detects the magnetic vector swings. To meet the second requirement, we place the MRE optimally so as to obtain the largest resistance change when the magnetic vector swings. We analyze the magnetic circuit by the software programs FEMAP (preprocessor) and ELF-MAGIC (solver). Fig. 7.10.7 shows a FEMAP model.

The FEMAP renders the gear and the magnet with a 3D solid model, and the area for MRE device placement with spatial entities (plates). Based on this FEMAP model, the magnetic flux distribution on the MRE device is analyzed using ELF-MAGIC, while the gear is rotated slowly. Magnetic vector swing and magnetic flux density are calculated from the analysis results obtained as magnetic flux components, B_x and B_y . Then the MRE device resistance variation is calculated from these values. Fig. 7.10.8 shows example analysis results.

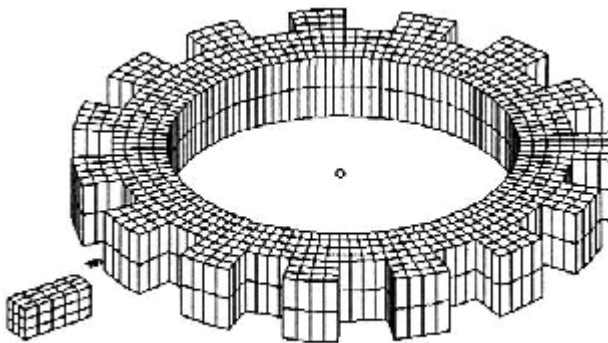
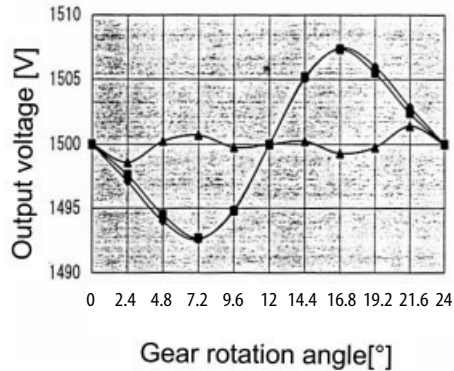


Fig. 7.10.7 FEMAP model

Fig. 7.10.8 Analysis results



7.10.2.2 Signal-Processing Circuit

The signal-processing circuit of the crank sensor consists of an amplification and electric noise-protection circuit and a wave-shaping circuit; the former convert the MRE output into an ac signal, which the latter converts into a binary signal appropriate for the gear pitches.

Amplification and Noise-Protection Circuit

The sensor operating on a 12 V battery must pass noise tests specific to the automobile such as load dump, field decay, noise injection, and reverse voltage test. Figure 7.10.9 and Fig. 7.10.10 show a block diagram and chip photograph of the amplification and noise protection circuit.

To prevent integrated circuit element breakdown, we incorporate protection resistors of NiFe film, the same material as the magnetic sensing element, in V_B and V_{out} terminals and use a chip capacitor. The output signals of the peak and

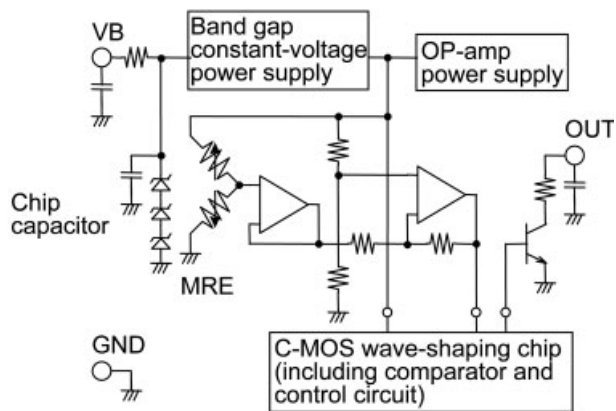


Fig. 7.10.9 Block diagram of amplification and noise-protection circuit

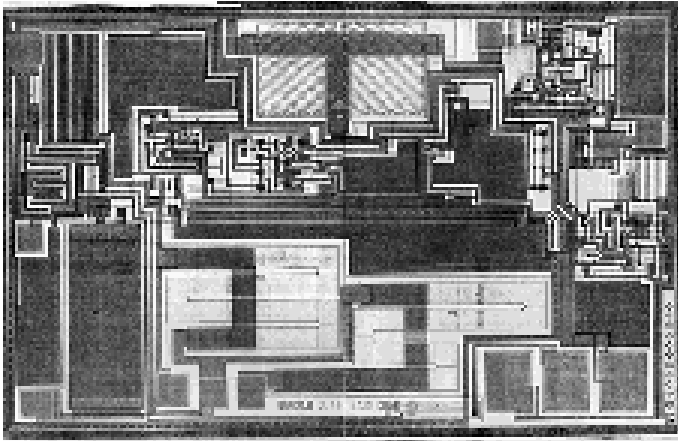


Fig. 7.10.10 Chip photograph

bottom hold circuit are not ratiometric to the supply voltage because it holds extreme values in its capacitors. Therefore the circuit delivers error pulses when a supply voltage variation occurs. As a measure to prevent error pulses, we provide a supply voltage filter in the noise protection circuit, which has another function of powering the wave-shaping circuit. The filter constant is designed based on extraneous noise frequency and amplitude. Since the amount of extraneous noise varies depending on the maker and the vehicle, it is necessary to ascertain the amount of noise by conducting a vehicle test and making sure that an adequate margin is provided.

Wave-Shaping Circuit

The wave-shaping circuit processes an ac signal from the amplification and electric noise-protection circuit and produces binary signals appropriate to the gear pitches. The offset level of the ac signal input to the wave-shaping circuit varies widely according to the characteristic variation between MREs and misalignment between the sensor and the vehicle, and drifts beyond the signal amplitude. Therefore the input-signal-tracking peak/bottom processor is important for angular accuracy. Fig. 7.10.11 shows a block diagram of the wave-shaping circuit.

Angular accuracy is affected by the following three circuit errors.

- Op-amp offset: to reduce the offset voltage of the op-amp, we take the following measures.
 - The input differential pair transistors are divided into two, and each pair is arranged to form a cross.
 - The transistors are arranged in a circle.
- High-temperature circuit leak. The basic peak/bottom hold circuit consists of an analog switch utilizing MOS transistors, a capacitor, and an op-amp. The transistors comprising the analog switch will leak at the pn junctions at high tem-

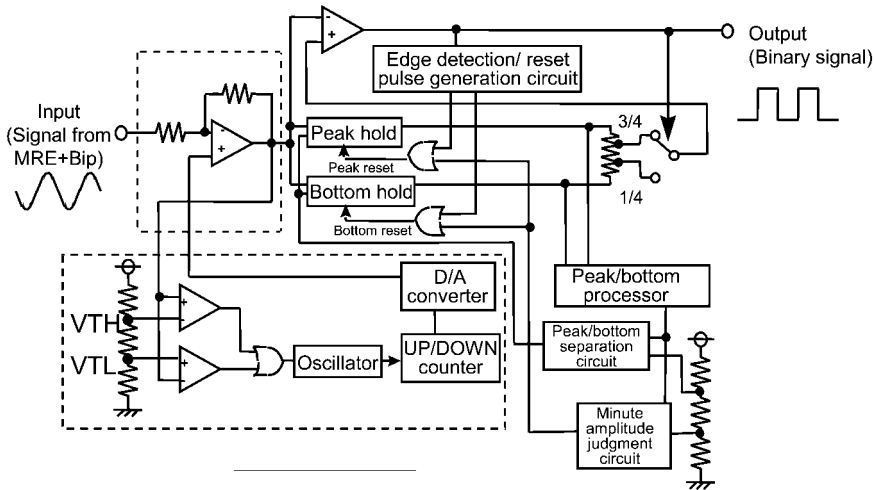


Fig. 7.10.11 Block diagram of wave-shaping circuit

peratures. The leak causes peak/bottom drift, resulting in an angular accuracy error. The capacitor size is therefore designed with a safety margin.

- Capacitor potential. With a built-in capacitor, the supply voltage detection circuit is designed to cause the peak/bottom processor to track the input signal until the power supply voltage reaches an adequate level for normal circuit operation.

Fig. 7.10.12 shows a simpler wave-shaping circuit than shown in Fig. 7.10.11. The minute ac signal from the MRE is amplified by the amplifiers, and then converted into a binary signal appropriate to the gear pitches based on the threshold

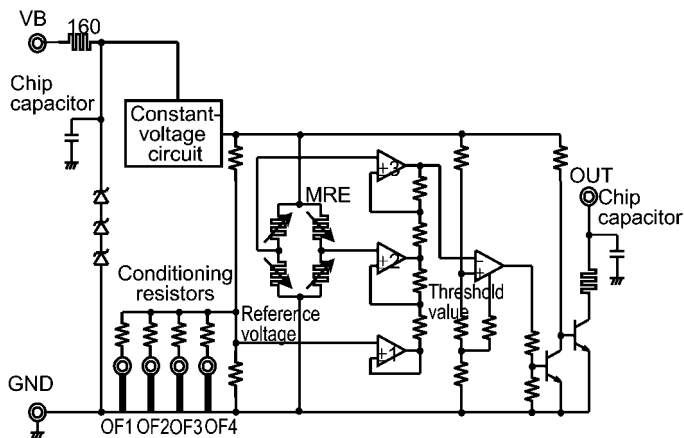


Fig. 7.10.12 Wave-shaping circuit with conditioning terminals

value determined by the relative proportion of resistances. The offset level of the ac signal input into the wave-shaping circuit varies widely by characteristic variation between MREs. To condition the MRE characteristic variation initially, this circuit is equipped with conditioning resistors. The four resistors are connected in parallel to the conditioning terminals, shown as OF1 to OF4 in Fig. 7.10.12, all of which are connected to a ground on the lead frame. The desired reference voltage is achieved by mechanically disconnecting terminals chosen based on the conditioning data.

7.10.2.3 Package

Fig. 7.10.13 shows the crank-sensor package. It consists of a magnet, a molded integrated circuit (MIC) with terminals, and a resin case. An MIC integrating an MRE device and a wave-shaping circuit is inserted into a hollow-body magnet. The MIC with the hollow-body magnet is inserted into a resin cap.

7.10.3

Another example

7.10.3.1 Hall sensor

This section describes examples of the gear detection sensor employing an Si Hall element. To achieve high sensitivity, it is important to make the smallest possible gap between the magnetic sensing element and the magnet when designing the package. Fig. 7.10.14 shows packages developed for the Hall sensor [3]. In the SG package (B), single-step molding eliminates the layer of plastic normally placed between the IC lead frame and the magnet and achieves the smallest possible gap. Improved heat dissipation realized through the reduced heat conduction path and elimination of the air entrapment that can occur during molding operation allow the realization of the package.

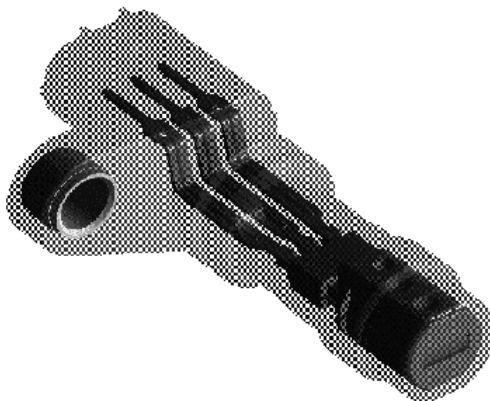


Fig. 7.10.13 Crank-sensor package

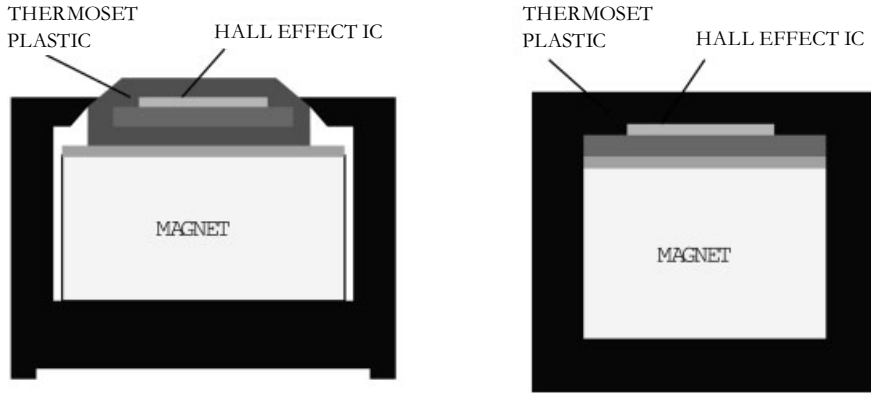


Fig. 7.10.14 Hall-sensor packages

7.10.3.2 Thin film magnetic field sensor utilizing MI effect

A novel thin film sensor sensitive to small magnetic field based on the Magneto Impedance effect has been proposed [4]. Figure 7.10.15 shows a schematic drawing of the thin film magnetic field detecting element. The sensor consists of half bridge of the individual detecting element with FeCoSiB/Cu/FeCoSiB multi-layer, which exhibits the large impedance change ratio more than 100% when an external magnetic field is applied. The detection resolution of 10^{-3} Oe order higher than those of any other conventional thin film sensor is obtained.

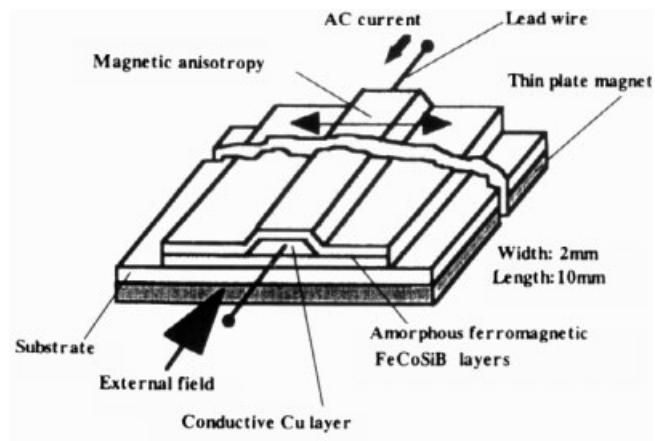

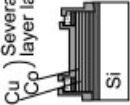
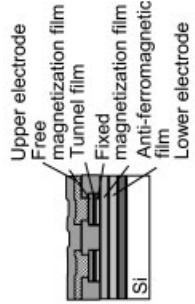
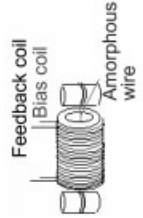
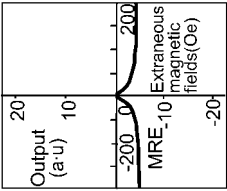
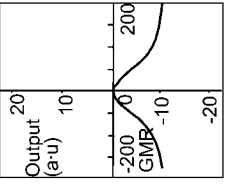
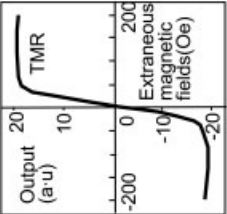
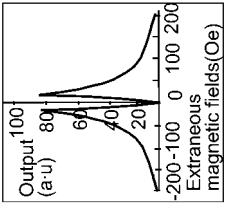


Fig. 7.10.15 Device photograph

Tab. 7.10.3 Structures and characteristics of magnetic elements

	MRE	GMR	TMR	MI
Structure				
Sensor characteristics				
Sensitivity (MR ratio)	3 to 5	5 to 10	10 to 20	50 to 80
Magnetic hysteresis	Low	High	High	Low
Integration with Si	Easy	Possible	Possible	Impossible
Power	DC	DC	DC	High frequency

7.10.4

Conclusions and Future Trends

Chassis manufacturers demand cam and crank sensors that are easy to install. It is therefore necessary that the air gap, the distance between the gear and the sensor, be widened with the same angular accuracy and at no additional cost. With this being the situation, an increasing number of new magnetic elements have been developed to improve the signal-to-noise ratio.

MRE has a rate of magnetic resistance change (MR ratio) of 2–4%. On the other hand, GMR (giant magnetoresistance), TMR (tunnel magnetoresistance), and MI (magnetoimpedance) films are known as magnetic elements with an MR ratio of 10% or more. Tab. 7.10.3 provides a summary of their structures and characteristics [4–6].

In an example, the drawback of the GMR element, the MR ratio decreasing at temperatures exceeding 250 °C, is overcome and a GMR element and a silicon signal processing circuit are integrated. For integration, the signal processing circuit is created first, and then the GMR. The problems are as follows.

- Satisfactory characteristics of the GMR element deposited on the interlayer insulation film (PSG film) of the signal processing circuit.
- Protection of the signal processing circuit against damage during GMR film creation and subsequent ion-beam etching.

The first problem is solved by ion-beam etching onto the PSG film surface, and the second one by providing protection diodes for the area susceptible to damage.

At present, nothing is superior to the MRE sensor in regards to the signal-to-noise ratio and ease of sensor device creation. In the case of the MRE sensor realized by applying the technology for the magnetic head to the sensor, however, the emergence of a superior sensor depends on whether materials and devices developed for the magnetic head can be applied to the sensor.

7.10.5

References

- 1 Engine Electrical Components Workshop, *Engine Electrical Components*, K. MITANI (ed.), Sankaido, 1996, Chapter 4.
- 2 Automotive Sensors Workshop, *Automotive Sensors*, M. OTA (ed.), Sankaido, 2000, Chapter 2.
- 3 D. DWYER, *J. Appl. Sens. Technol.*, September 2001, 14–24.
- 4 Y. NISHIBLE, H. YAMADERA, N. OHTA, K. TSUKADA, Y. NONOMURA, *Sensors and Actuators*, 2000, 82, 155–160.
- 5 S. I. KASATKIN, P. I. NIKITIN, A. M. MURAVJEV, et al., *Sensors and Actuators* 2000, 85, 221–226.
- 6 K. MOHRI, T. UCHIYAMA, L. V. PANINA, *Sensors and Actuators* 1997, A59, 1–8.

7.11

Steering-Angle Sensors

PETER KOFINK

7.11.1

Introduction

During the first half of the 20th century vehicles worked well without any electronics and sensors. A simple electric harness with a battery charged by an alternator, switches to operate indicators, light, fan, and ignition were enough. Safety and comfort considerations started to play a role in terms of passive safety using a safety belt or introducing a collapsible steering column, but everything worked fine without a need for electronics.

Electronic components, introduced in the 1970s, changed the way automotive components are designed. Heated-wire turn-indicator relays disappeared in favor of electronic timer relays, and engines performed much better with electronic control instead of electromechanical wear-sensitive ignition control. Starting with the development of environmentally friendly electronic controlled engines, sensors have played an increasing role in automotive technology. The main reasons for the introduction of electronic control were to reduce fuel consumption and emissions, and to increase comfort and power. Sensors are needed to measure engine temperature in order to control the proper amount of fuel to be injected for better starting behavior. Another example is the use of an oxygen (Lambda) sensor to inform an electronic control system about the oxygen content of exhaust fumes in order to reduce the emission of carbon monoxide. A catalytic converter further eliminates harmful gases to minimize pollution. ABS (antilock braking system) uses wheel-speed sensors on each wheel to prevent the wheels locking and making the car uncontrollable in case of full braking.

Steering-angle sensors (SAS) were developed in the mid 1990s as new brake-related functions such as ESP (electronic stability program) emerged. ESP needs steering-wheel angle, yaw rate, wheel speed, and lateral acceleration sensors to identify whether the car tends to oversteer or understeer. By braking individual wheels, ESP can regain stable driving conditions. After roll-over incidents with the Mercedes A-class in 1997, ESP became standard equipment in most cars.

Today SAS are being increasingly used in passenger cars and trucks in Europe. Other markets are following, and SAS will be used in cars for intelligent electronic control and assistance systems such as the following.

- ESP (electronic stability program).
- EPS (electric power steering).
- ABC (active body control).
- BCD (body control damping).
- ACC (adaptive cruise control).
- Navigation systems.
- Forward intelligent lighting.

- AFS (active front & rear steering).
- SbW (steer-by-wire).

7.11.2

Technology

Angle sensors, also known as rotary analogue/digital encoders in industrial applications, have come in a wide variety of designs since the 1990s, the most common are rotary potentiometers, capacitive, magnetic, and optical types. The optical angle sensor has become the most popular of these encoding methods because of its long life, simplicity of design, versatility, high accuracy, and high resolution.

Sensing a steering angle in automotive applications can be done in various ways using optical, magnetic, inductive, capacitive, or resistive sensor principles. Fig. 7.11.1 is a summary diagram that sorts different kinds of SAS known from patents or patent applications. The majority of SAS use optical principles, closely followed by magnetic principles based on the Hall-effect or magnetoresistive sensors including evaluation circuits, now available at reasonable cost. Inductive principles (especially screened inductance types) have been introduced recently for steering applications.

In SAS technology for automotive use there are various types of sensors. An incremental angle sensor, for example, is a sensor using a digital coding technique to digitize rotational movements into discrete steps. The resolution is directly proportional to the rotation from one digital step to another. Incremental coded angle sensors deliver a rising or falling edge on a digital output at every discrete angular step. A quadrature coded angle sensor also uses incremental coding, but has a second sensor positioned at 90° to the first sensor in order to deliver information about the direction of rotation (Fig. 7.11.7). Incremental and quadrature coded angle sensors are referred to as relative angle sensors. Absolute angle encoders benefit from absolute coding techniques such as Gray codes or chain codes [20]. Each angular step generates a complete word (several bits) corresponding to a new ab-

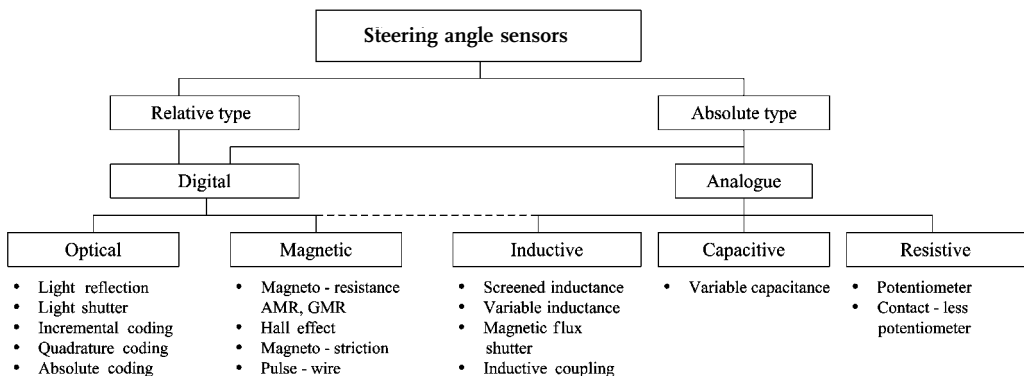


Fig. 7.11.1 Steering-angle sensor overview

solute angle. The digital coding technique makes these sensors more resistant to external distortions (e.g., electromagnetic fields).

There are also analogue angle sensors that produce a continuous signal (a voltage, phase shift, or frequency change) corresponding to the absolute angle of rotation. An A/D converter (ADC) is often used to generate a digital output, so it is not easy to tell from the external interface whether the sensor is an analogue or digital type. The resolution of the output signal is defined by the resolution of the ADC and the quality of the sensor signal. Some analogue sensors have a tendency to show signal drift, which is interpreted as a fault angle. The higher the resolution required, the more measures have to be taken in order to make the signal less sensitive and to reduce distortion or interference.

7.11.2.1 Measuring Steering Angle

Electronic control systems such as ESP, EPS, ABC, ECD, ACC, and navigation systems need steering-angle data as part of the information available on the vehicle network (CAN=control area network). Some ESP system suppliers and car manufacturers require an SAS to fulfill safety requirements. Fig. 7.11.2 shows a modular approach around an SAS core technology.

There are means other than SAS for counting several revolutions of a steering wheel, either mechanically or electrically. The electronic controls around an SAS must fulfill the automotive requirements: low current consumption, wake-up functions, and EMC requirements among others. When the SAS is integrated into the controls on top of the steering column (top column module, TCM) it has to share microcontroller resources and has to fit into the TCM architecture. The de-

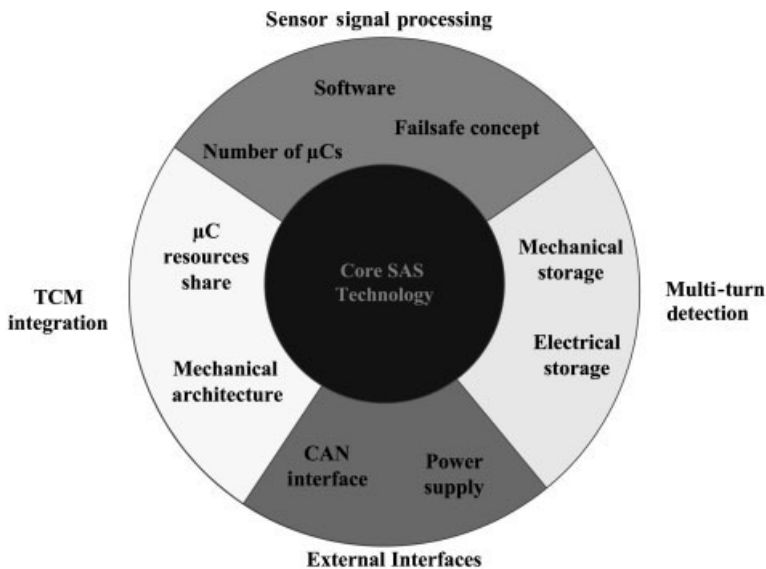


Fig. 7.11.2 Building a steering-angle measuring system (source: Valeo)

sign of the TCM architecture is essential to the performance of the SAS. Safety requirements and network capabilities have to be managed to give a fail-safe system, which means more than one microcontroller and appropriate software.

7.11.2.2 Function Modeling of Steering-Angle Sensor Principles

Function analysis system technique (FAST) is known as part of value analysis and engineering, originated by L. D. Miles [12]. Over the years, the function analysis method found its way into the TRIZ innovation methodology. TRIZ is a Russian acronym and can be translated in English as ‘theory of inventive problem solving’ (TIPS) [27]. There are TRIZ tools on the market that provide a function-modeling front-end, which is similar to the FAST approach but with additional functionality. This function-modeling tool is used to analyze products (or patents) at the beginning of a TRIZ session.

When studying the different SAS principles using the function-modeling tool, it can be found that there are basic structures common to almost all the concepts known today. One way of illustrating the commonality between the different SAS principles is to classify the contactless SAS into two sets: some with an active and fixed transmitter (Fig. 7.11.3) and some with a passive, rotating transmitter without modulator (Fig. 7.11.4).

Fig. 7.11.3 shows six main blocks. The rotating shaft is an external part such as the steering column, which rotates the modulator. The modulator gives a direct relationship between rotational angle and the modulated signal that is emitted by the transmitter. The receiver reads the signal generated by the modulator and the evaluation unit reads the receiver output representing the modulated signal. The evaluation unit drives the transmitter and can therefore compare driving signal with read signal finally to evaluate the steering angle and inform the system. Examples for this general structure can be found with almost all optical, inductive, magnetic, and capacitive sensor principles. Optical principles have a light transmitter to send light and a modulator to modulate the light according to a function

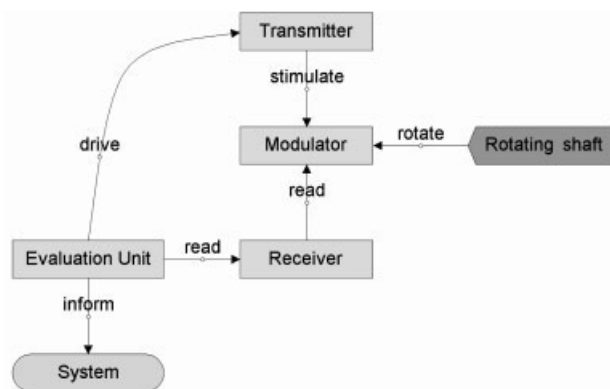


Fig. 7.11.3 Contactless steering-angle sensors with active and fixed transmitter

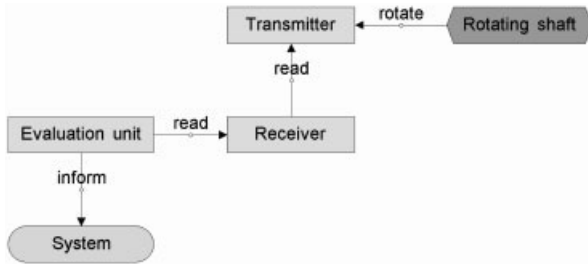


Fig. 7.11.4 Contactless steering-angle sensors with passive, rotating transmitter without modulator

of the rotational angle [3]. Screened inductance sensors have a transmitting and receiving coil shielded by a rotatable conductor [1, 2]. Capacitive principles are not very commonly used, but an example can be found [4].

Fig. 7.11.4 shows the structure without separate modulator. There is no need to drive the transmitter or actively to excite any fields. The transmitter performs as a passive, permanent transmitter that is pre-modulated. A constant magnetic field is provided by the use of permanent magnets, which are rotated by the rotating shaft so the magnetic field rotates correspondingly. Changes in orientation of the magnetic-field vectors can be measured by a magnetoresistive sensor, for example, as receiver [5].

For both models it is important to note that the interface between the SAS and the rotating shaft must be optimally designed to maximize overall accuracy.

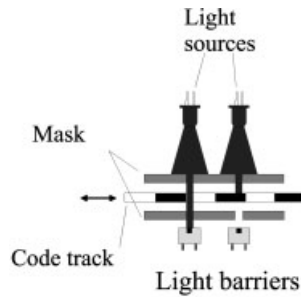
7.11.2.3 Optical Measuring Principles

Almost all optical rotation measuring applications are designed digitally because of the degradation effects in LEDs. Nevertheless, it would be possible to design an analogue optical SAS using a light damping material in the field of light intensity measurement [22]. Other optical effects known from physics could be used to modulate rotational movement (e.g., polarization of light). The optical measuring principles can be fully described by the laws of geometrical optics [14].

Examples of Optical Measuring Principles

Encoding of rotational angle can be achieved by using the effects of light reflection, absorption, and transmission combined with modulation of the rotational movement. Different kinds of photosensors, such as phototransistors, photodiodes, optical sensor arrays, or CMOS linear sensor arrays, are used to sense the modulated signals. SAS for automotive applications use 'single-ended' readout photosensors to measure light intensity in a light barrier configuration, for example. There are also 'push-pull' readout photosensors available that deliver a \sin/\cos output signal, in combination with an incremental code disc in order to compensate for the effects of degradation and the negative temperature gradient of LEDs [28].

Fig. 7.11.5 Optical barrier using the light shutter technique (source: Valeo)



All sensors using an optical measurement principle are contactless and, therefore, do not experience mechanical wear. Optical SAS consist of a light transmitter (e.g., an LED), a code disc to modulate the light path during rotation, and a receiver (e.g., a photosensor), to read the modulated light. There are two types of sensor: the reflective type and the light shutter type. In reflective sensors the code disc is patterned with reflecting and absorbing areas. Fig. 7.11.5 shows the light shutter technique.

This sensor principle is digital as the code track shuts the light in front of the receiver on and off. It is a quadrature encoded type as the two light barriers are at 90° . A mask is placed in front of the photosensors. Both code track size and mask slit size have to mate with each other to give a high-precision output signal. An example for this kind of sensor configuration (used in an SAS for ESP) can be seen in Fig. 7.11.6.

Here, the light barriers are formed of a plastic injection-molded part in which infrared LEDs and photosensors are fixed. The mask is part of the optical sensor head carrier to create a light path, that is shut on and off by the code and incremental tracks. There is a code disc having two tracks for code and incremental encoding formed as teeth and gaps. When rotating the code disc the four light paths are shut on and off by the teeth according to the predefined code. The code track is read by two light barriers and delivers a chain code with absolute-angle encoding. The incremental track is read by two light barriers and arranged as a quadrature sensor. With these tracks the sensor is able to recognize both the direction of

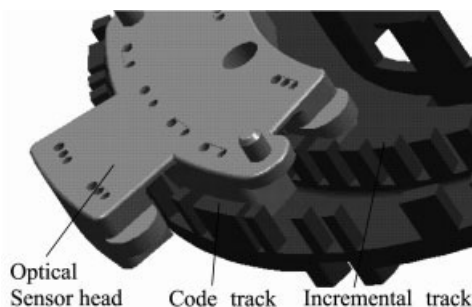


Fig. 7.11.6 Optical sensor head for incremental and code track (source: Valeo)

rotation and the absolute steering angle. Depending on the code disc position, the four light barriers deliver a special pattern seen in Fig. 7.11.7.

These signals pass to a signal-conditioning unit and are then read by two microcontrollers. With this sensor-head configuration, microcontrollers are able to calculate the angle with a resolution of 1.5° . After a restart, the steering-angle sensor needs an initialization angle below 4.5° to determine the absolute angle [6, 7].

In contrast to the sensor principle described above, there is a solution that uses a Gray code encoding technique in combination with several light barriers, as shown in Fig. 7.11.8. The upper rotating disc carries teeth and gaps encoded as a 6-bit Gray code. All transmitter and receiver elements are assembled in the stator, which forms a set of optical barriers [15].

Another sensor principle designed as a compact optical sensor head is shown in Fig. 7.11.9. All the optical sensor elements are assembled, for example, on a printed circuit board.

The optical array has an unique 'W' shape optimized for a code disc with coding scheme having four code tracks. This method measures the absolute angle of rotation. In comparison with the first principle presented in Fig. 7.11.6, it has light barriers with 12 photosensors in order to perform a true power-on function, meaning the sensor does not need to be initialized. One sample taken by one flash from a transmitter LED (imaging the code disc onto the receiver) is enough to decode the absolute angle, as shown in Fig. 7.11.10 [9]. The characteristic of this code also allows complete single failure recognition of the encoding sensor scheme [11].

The SAS described are optimized for a resolution of $0.75\text{--}1.5^\circ$. For higher resolutions up to 0.1° it is necessary to use a different method because of the restriction in size for the plastic injection-molded code disc and mask. One example of an SAS with a higher resolution is shown in Fig. 7.11.11 [13].

Again there is a light transmitter (e.g., an infrared LED) to illuminate a code disc. The code disc is made of a translucent material and is illuminated by the transmitter. A spiral shaped prism in the surface of the translucent code disc reflects the light perpendicular to the surface. Directly under the code disc there is a CMOS line sensor detecting the reflected light. An additional reference prism in the code disc makes it possible to eliminate effects of a radial clearance.

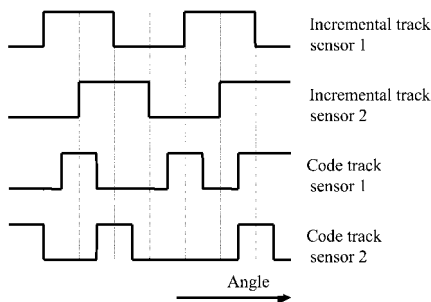


Fig. 7.11.7 Code pattern example (source: Valeo)

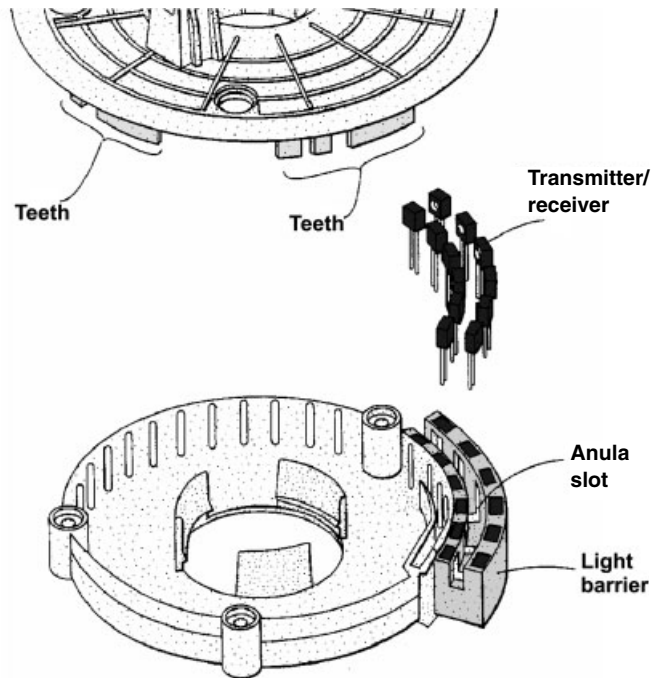


Fig. 7.11.8 Steering-angle sensor using Gray code technique [15]

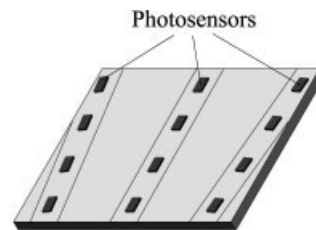


Fig. 7.11.9 Steering-angle sensor principle based on an optical array [9]

At first these kinds of sensors were based on CCD (charge-coupled device) line sensors as a subset of CCD matrix sensors known from image sensors used in digital cameras. SAS do not need as many sensitive pixels on a CCD line sensor as normally used for image processing applications. There are typically 128 pixels arranged in a line to pick up a code disc image [20]. CMOS line sensors are an interesting alternative for SAS applications because of their lower cost and lower power consumption than CCD line sensors. Both sensors capture the light intensity in each cell by accumulating it into a charge that can be captured by an ADC.

A different SAS solution based, for example, on a CMOS line sensor is shown in Fig. 7.11.12. It takes a picture of the teeth and gaps of a code disc illuminated by a light source. The image of the code pattern can be scaled by the use of a

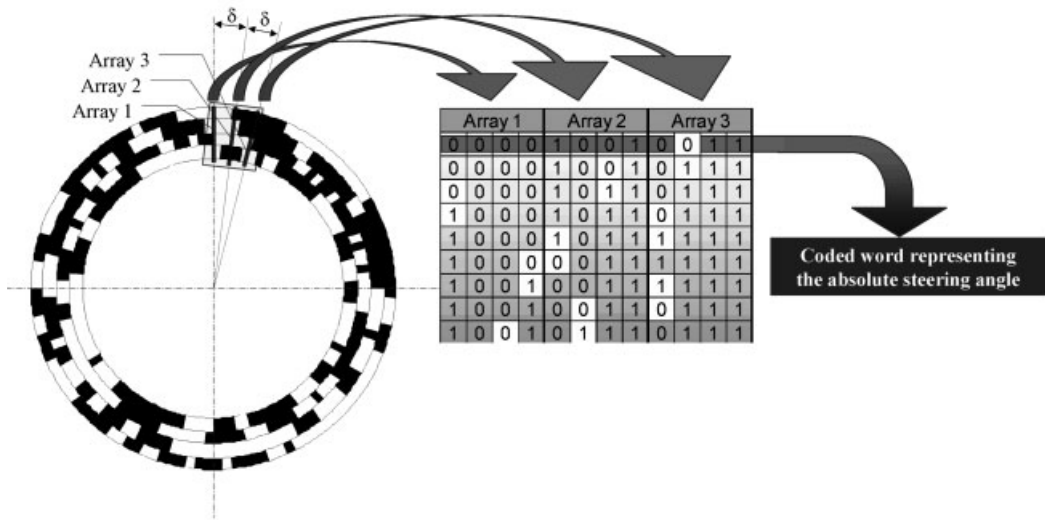


Fig. 7.11.10 Code pattern example used for the W-shaped sensor array (source: Valeo)

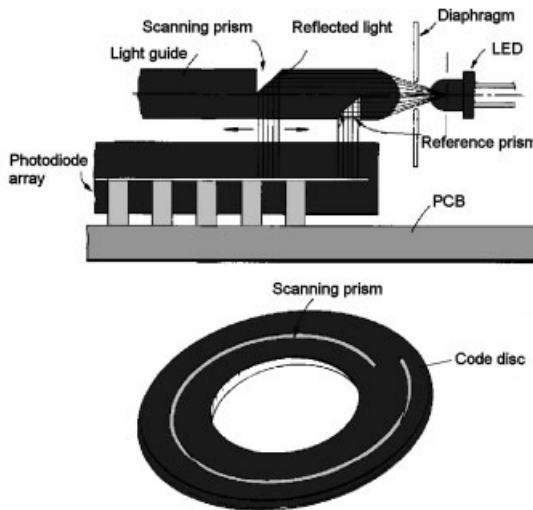


Fig. 7.11.11 Steering-angle sensor principle based on a CMOS line sensor [13]

lens [19]. A further example of an SAS based on a CCD line sensor is shown in Fig. 7.11.13 [10].

In this example a code disc carries an absolute code that looks like a bar-code. In this solution the transmitter (e.g., a LED) is located on the same side of a printed circuit board as the receiver (e.g., a CCD line sensor). By the use of a light guide, the light from the LED is guided evenly to the code disc, which is placed directly above the CCD line sensor, and shining through the code disc

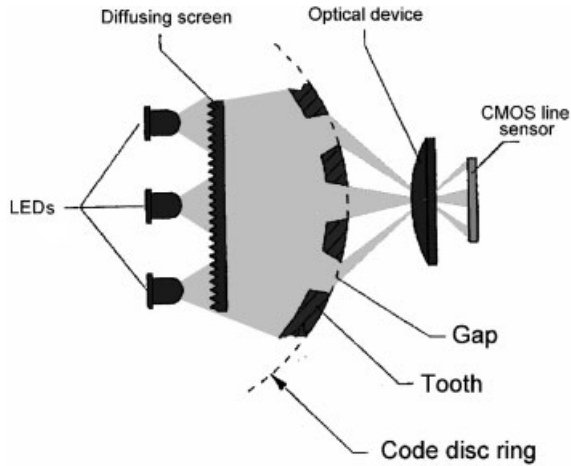


Fig. 7.11.12 Steering-angle sensor principle based on a CMOS line sensor [19]



Fig. 7.11.13 Steering-angle sensor principle as a separate optical sensor head (source: Valeo)

imaging a picture of the current code on the CCD line sensor. The advantage of this sensor configuration is in the possibility of gaining a higher resolution with the CCD line sensor than the code disc digital image on it. The CCD line sensor has, say, 128 sensitive pixels arranged in one line. The final resolution is achieved from the position of those individual light-sensitive pixels looking at rising or falling edges of the code picture imaged onto the line sensor [8].

7.11.2.4 Magnetic Measuring Principles

Magnetic sensors have similar advantages to optical sensors. They are contactless, wear free, and inexpensive. Magnetic sensors are less sensitive to contamination than optical sensors, but are more sensitive to external distortions such as external magnetic or electromagnetic fields. Magnetic fields from permanent magnets do not need an external power supply. When the magnet is rotated, the magnetic field changes relative to a stationary part such as a magnetic-field sensor or coil, and induces a voltage in the coil. This is an inductive sensor, because a current is induced. This example shows the close link between magnetic and inductive measuring principles. Magnetic sensors measure magnetic-flux density or magnetic-field orientation. Inductive sensors generally use the effects of electromotive force (EMF). The behavior of magnetic fields are governed by the fundamental set of Maxwell equations.

There are various magnetic-field sensors on the market. A Hall-effect sensor measures magnetic-flux density B perpendicular to its plane (Fig. 7.11.14). In one direction a constant current flows through the Hall-effect sensor. The external magnetic-flux density B creates a proportional voltage U_{Hall} . Various Hall-effect sensors are on the market, mostly with an evaluation circuit to deliver a digital output.

Another type of magnetic-field sensor is the anisotropic magnetoresistance (AMR) sensor element shown in Fig. 7.11.15. In contrast to the Hall-effect sensor, the AMR sensor measures the direction of the magnetization vector relative to the current-density vector, both lying in the area of the sensor plane. The magnetic-field strength is not important for AMR sensors as long as it is above a certain threshold and the magnetization vector follows the external magnet field. This is achieved when the external field strength is much higher than the internal magnetization of the AMR sensor [16].

Single AMR elements have some shortcomings, particularly temperature dependency. This can be easily overcome by using a Wheatstone bridge to compensate for the temperature effects and to improve measurement precision. One AMR element in an angular measurement application covers a 90° range. Therefore improvements have been made by adding a second AMR element at an angle of 45° to the first. As a result, this AMR sensor delivers a \sin/\cos signal, which can be further processed with an arc tangent algorithm (e.g., CORDIC). This sensor combination can measure an angular range of 180° (Fig. 7.11.16).

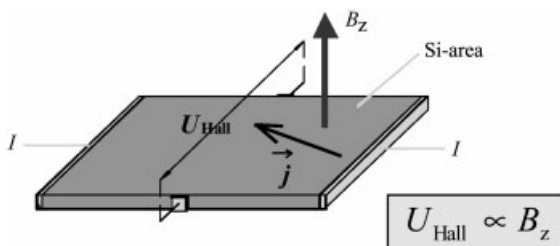


Fig. 7.11.14 Hall-effect sensor principle

Fig. 7.11.15 Anisotropic magnetoresistance sensor principle

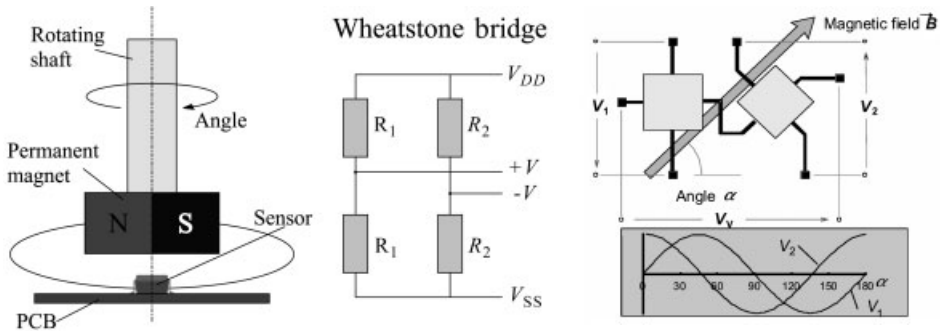
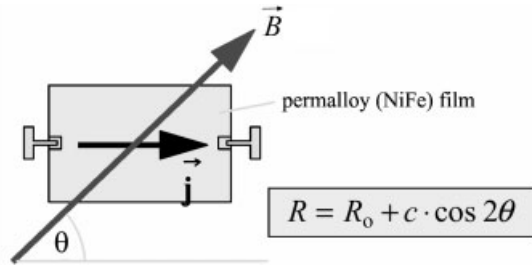
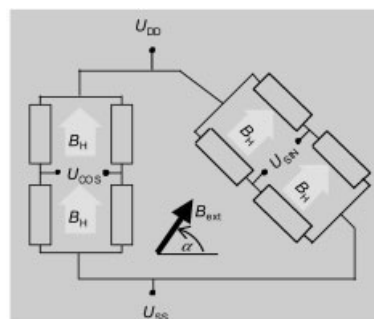


Fig. 7.11.16 Angle measurement using an anisotropic magnetoresistance sensor (source: Philips/Bosch)

It is now possible to measure a 360° angular range with one sensor having two AMR elements, by introducing an additional internal switchable magnetic field B_H in the sensor chip [17] (Fig. 7.11.17).

Magnetic methods can work similarly to the optical shutter method. A magnetic-field sensor is placed opposite a permanent magnet. With metal teeth on an encoded metal disc, it is possible to shield the magnetic field from the magnetic-field sensor. This kind of sensor is less precise than the optical shutter type because of the stray magnetic field, which has to be taken into account. A product

Fig. 7.11.17 Anisotropic magnetoresistance sensor principle for 360° rotation (source: Bosch)



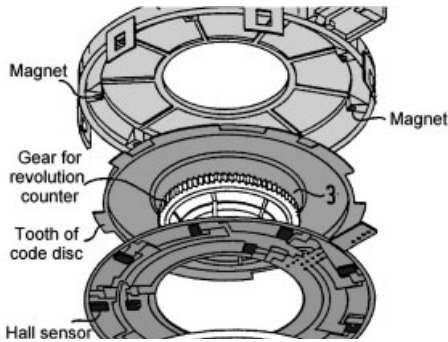


Fig. 7.11.18 Hall-effect sensor similar to the light shutter type [21]

based on this technique has been designed by Bosch, using the Gray code encoding technique with a resolution of 2° (Fig. 7.11.18) [21].

The code disc is made of ferromagnetic material. Its rotation modulates the magnetic field according to a Gray code encoding mechanism by shielding the magnetic field coming from stationary magnets. Each magnet is inserted into the sensor housing and placed opposite a Hall sensor. Several evenly distributed Hall sensors are arranged on an outer ring. Additional Hall sensors, positioned on an inner ring, are used to count rotations greater than 360° .

SAS for automotive applications, especially for ESP systems, have to measure an absolute angle of rotation over one rotation. The maximum angle of rotation is about 1040° , depending on the vehicle system specification. Therefore SAS must be capable of measuring several turns. Normally this is performed by a separate sensor arrangement called a revolution counter, or by storing information electronically. Another way of measuring a rotation up to 1040° is by gearing it down to a 180° rotation, but this would require a high-resolution SAS to compensate for the transformation. A different approach is to use a gear arrangement consisting of one driving gear with first and second driven gears, as shown in Fig. 7.11.19.

The second driven gear has one tooth more than the first. Consequently the second gear rotates slower than the first gear; so the phase relationship between first and second gear gives the turn identification information [18]. As a result this is a true power-on measuring method over several revolutions.

7.11.2.5 Inductive Measuring Principles

Inductive measurement methods are based on ac signals applied to an exciting- (transmitter) and receiving-coil arrangement. By means of an intermediate or juxtaposed inductive coupling element, the excited magnetic field is modulated dependent on the degree of rotation. The modulated result can be measured on the receiving coils because of a difference in the inductive coupling between exciting and receiving coils. An evaluation circuit is used to determine the angle of rotation via detection of amplitude values and/or phase relationship of the signals obtained from the receiving coils.

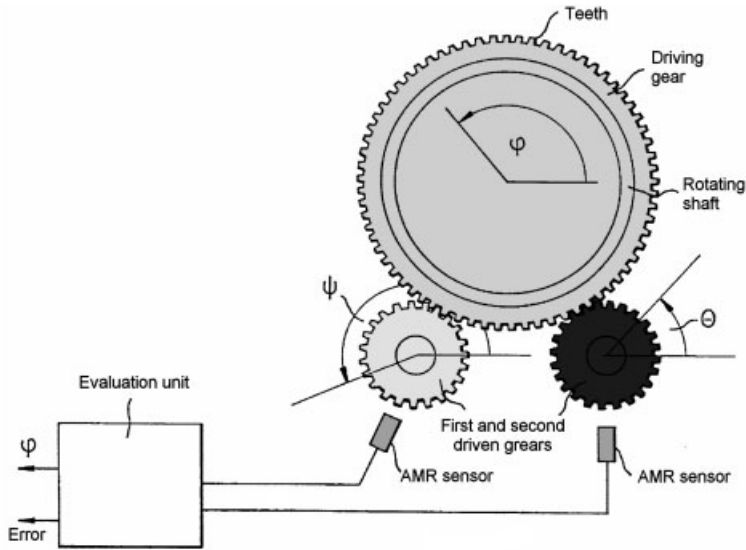


Fig. 7.11.19 True power on and absolute steering-angle sensor measuring principle (source: Bosch)

A particular problem with inductive measuring principles for automotive applications is sensitivity to electrical, magnetic, and electromagnetic fields, which can disturb signals picked up by the receiving coils. This results in a fault angle depending on the kind of disturbance. Appropriate measures have to be taken to compensate for the influence of external fields.

Inductive angle sensors are increasingly being used in motor vehicles instead of resistive sensors (e.g., throttle valve potentiometers). They are available for axles, throttle valves, accelerator pedal modules, and also as SAS. The angle measuring range is $10\text{--}360^\circ$. Greater angle ranges are achieved by multi-turn sensors. With measuring accuracy of $\pm 1\%$, which means 0.3° including temperature and aging drift, these sensors give a relatively exact and robust means of measurement at low cost.

There are presently two different measuring methods on the market, both using inductive coupling. The first consists of planar exciting and receiving coils building the stator element and a rotating element with a short-circuit loop. Fig. 7.11.20 shows the principle of this sensor [24].

All coils are placed on a printed circuit board, thus lying on a plane. The exciting (transmitter) coil generates an alternating magnetic field. An LC oscillator is used to create the necessary alternating current. There are three independent receiving coils ($S1/S2/S3$) twisted with 20° against each other. There is a rotor element with a short-circuit loop of the same shape as the receiver coils. Each receiver coil consists of several sub-coils, which are connected in reverse polarity leading to the unique meander shape. As a result the output voltage of each receiver coil is zero without the rotor being present [23, 24]. If the rotor element starts to

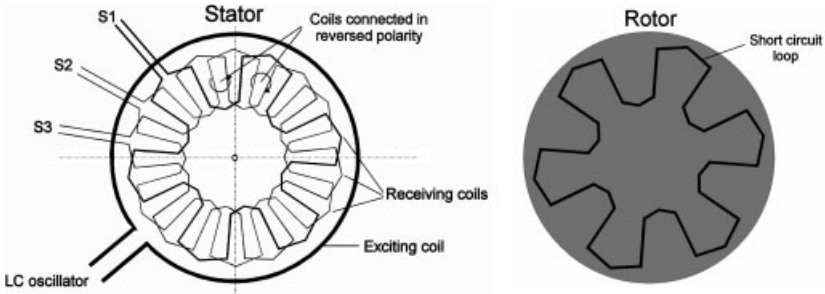


Fig. 7.11.20 Inductive coupling principle with rotor and stator [24]

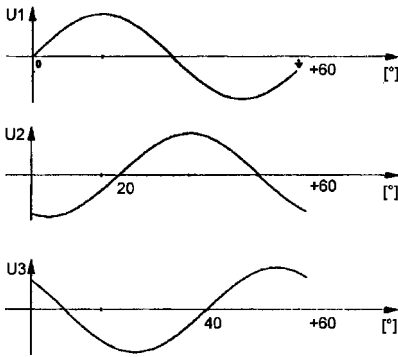


Fig. 7.11.21 Output voltage of inductive coupling principle [24]

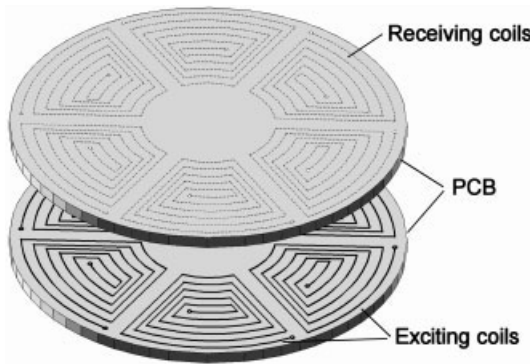


Fig. 7.11.22 Inductive coupling attenuated by a conductor not shown [25]

rotate from 0° to 60° , beginning from a position overlapping S1, the output voltage of the three coils ($U_1/U_2/U_3$) will be as shown in Fig. 7.11.21 because of a variation in the inductive coupling.

There is another way of measuring steering angle, by attenuating the inductive coupling between two coils facing each other with a conductor that is moving between both coils. Fig. 7.11.22 shows the coil arrangement with exciting and receiving coils facing each other.

This method is also called screened inductance as a specially shaped conductor (modulator) between the exciting and receiving coils attenuates the inductive coupling, that is, shielding the magnetic field from receiving coils [25, 26]. The conductor does not necessarily have to be positioned between exciting and receiving coils. Both exciting and receiving coils can also be located on one printed circuit board.

7.11.2.6 Capacitive Measuring Principles

Capacitive measuring methods are used, for example, to make a contactless potentiometer. Instead of a sliding contact there is a capacitive coupling element moving over a resistance track. These methods are pure analogue and need to be complemented by a proper signal-conditioning unit. Disturbance of the analogue signal will result in a corresponding fault angle.

7.11.3

Signal Processing

The necessary signal processing depends on the sensor method used. Fig. 7.11.23 shows how SAS signals are further processed in automotive applications. It shows an SAS version using a digital angle sensor for a 360° measurement and a digital revolution counter for a measurement over 360° , up to about 1040° . Most of the SAS for automotive applications consist of a microcontroller including a CAN interface. Besides the primary signal processing in SAS applications it is also very important to design a comprehensive fault diagnostic function for rapid failure analysis during production and service.

SAS for ESP systems must be designed as fault-silent units (FSU). This means that in the event of a failure no information, rather than an erroneous signal, must be sent to the ESP system. Special care has to be taken during development

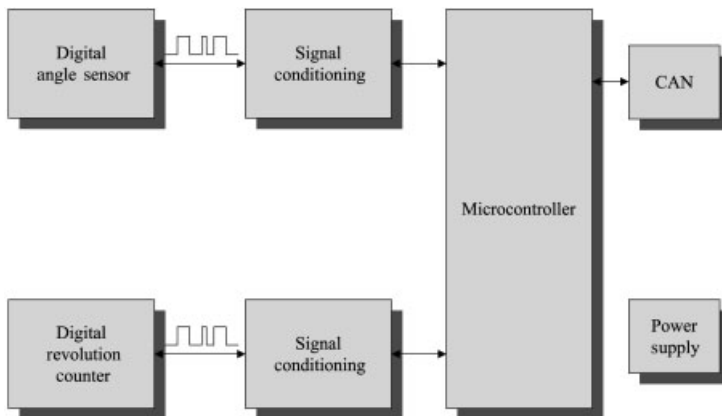


Fig. 7.11.23 Block diagram for a digital steering-angle sensor solution (source: Valeo)

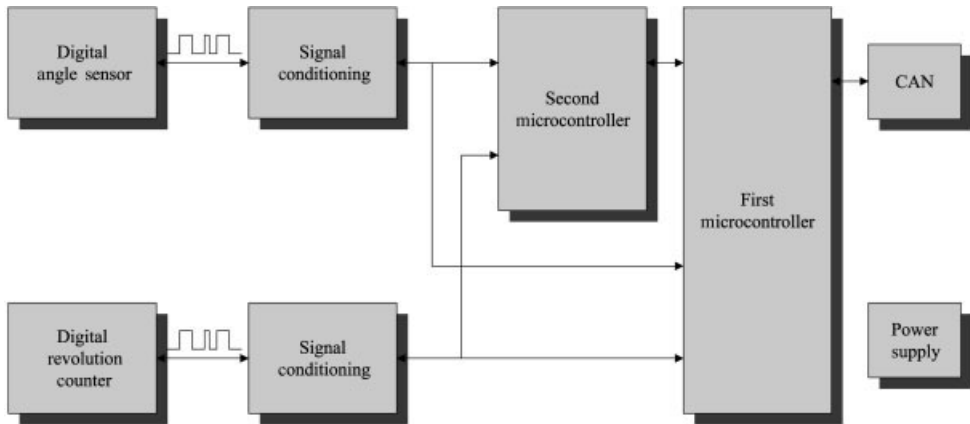


Fig. 7.11.24 Block diagram for a digital steering-angle sensor system with two microcontrollers (source: Valeo)

of such a product. Redundancy is used to protect against failure, and often two microcontrollers are used. An example can be seen in Fig. 7.11.24. Each microcontroller reads the sensor data and processes it to a corresponding steering angle. A mutual exchange of information decides whether or not the steering angle can be sent. In case of an error the ESP system is informed.

CAN is the serial communication network used in most intelligent SAS products. Future systems such as SbW will need fault-tolerant networks (FTU), which provide the necessary safety level for those systems, that is, being predictable and deterministic. There are two major protocols in development, TTP (time triggered protocol [32]) and Flexray, supported by two different consortia of car manufacturers and system suppliers. VW/Audi, PSA, Renault, Delphi, Visteon, United Technologies, Airbus, Honeywell, AMS C&C Electronics, and NEC are in the TTP consortium. DC, BMW, GM, Bosch, Philips, digital dna, Motorola, and partners are the Flexray group [30, 31].

7.11.4

Applications

7.11.4.1 Main Application Fields for Steering-Angle Sensors

SAS are designed for the use in different systems such as ESP, EPS, BCD, AFS, and others, but the majority of SAS are used for ESP. Today there are three different locations in the vehicle for an SAS system: near the steering wheel, in the steering column, or as part of the EPS system for column and rack-mounted type. EPS systems as shown in Fig. 7.11.25.

Most commonly the SAS is a steering-wheel mounted type, integrated in the TCM. There are different requirements for the SAS depending on the location and the systems for which they are used. EPS systems are more demanding, leading to a higher product cost. One major cost driver for SAS is the operating tem-

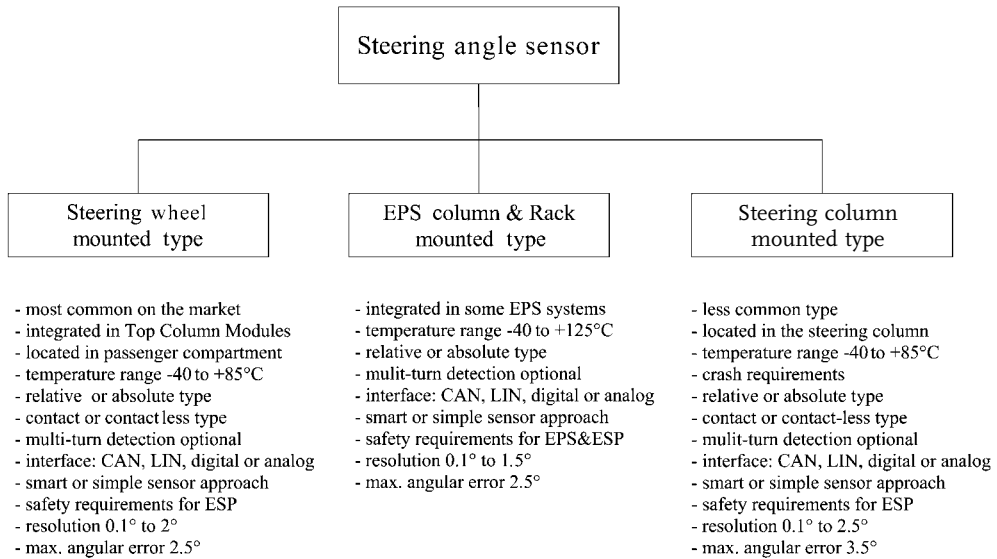


Fig. 7.11.25 Location of steering-angle sensors

perature range. An angular resolution of 1.5° for ESP systems is sufficient. Other systems such as AFS or EPS systems need a higher resolution up to 0.1°.

7.11.4.2 Steering-Angle Sensor: an Important Part of the Electronic Stability Program

The main application of SAS is ESP. This system extends longitudinal control of the vehicle (i.e., ABS, ASR) with a yaw-rate control. One major sensor input of the ESP system is the steering-wheel angle. This information is interpreted as the intention of the driver and serves as the important human-machine interface for the ESP system. The other inputs are signals for lateral acceleration, yaw rate, and wheel speed. With this additional information, the ESP can calculate the real behavior of the car. If the actual drive direction matches the intention, nothing happens. If there is a significant difference, the ESP system takes control of the car by braking individual wheels and intervening in engine management to force the car in the desired direction (Fig. 7.11.26). This illustration shows why an SAS should not send wrong angle information to the ESP system, whilst the ESP system secures its function by performing plausibility checks on all sensor data by means of sensor fusion.

7.11.4.3 Various Applications

SAS are basically relative and absolute rotary encoders especially designed for automotive use. Absolute SAS are produced in high volume for ESP systems. In Europe the most common versions are steering-wheel mounted types, being either a separate component such as in Fig. 7.11.28 or integrated into the TCM,



Fig. 7.11.26 Effect with and without electronic stability program (source: Continental)

as shown in Fig. 7.11.27. There are various suppliers offering these kinds of integrated TCM, such as Valeo, Kostal, Delphi, (Eaton), TRW, and Tokai-Rika.

The TCM shown in Fig. 7.11.27 consists of several parts: a shroud, a turn signal/wiper switch, a steering-column adjustment switch, an advanced cruise control and push-to-talk switch, an electronic module with integrated SAS, and a clockspring with steering-wheel heating option. A clockspring is a rotary connector connecting airbag and steering-wheel switches. The figure shows the TCM architecture with integrated SAS. Further comfort functions are included in the electronic module that leads to an overall lower system cost than a stand-alone SAS.

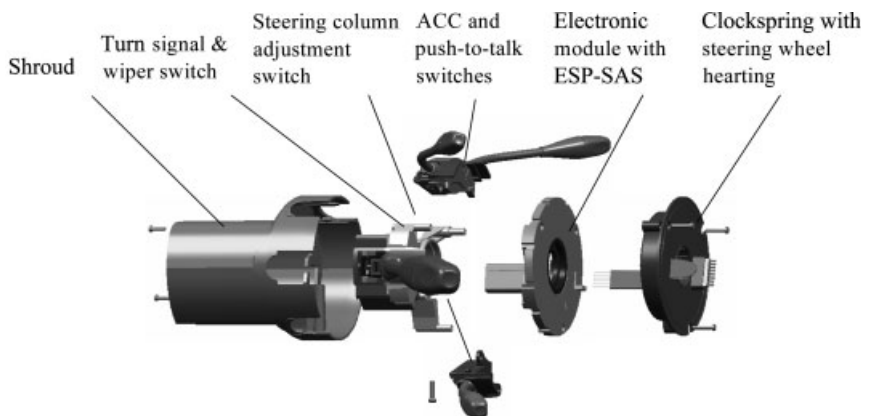


Fig. 7.11.27 Top column module with steering-angle sensor (source: Valeo)

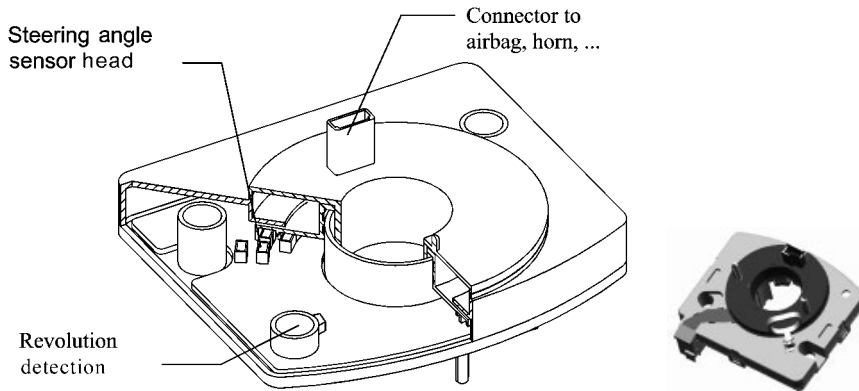


Fig. 7.11.28 Steering-angle sensor mounted beneath the steering wheel (source: Valeo)

The SAS is primarily designed for use in the ESP system and therefore has to fulfill safety requirements, taken into account by the use of redundant mechanisms (e.g., two microcontroller, redundant encoding, redundant messaging, hardware diagnostics).

Fig. 7.11.28 shows a typical SAS that fits atop a steering-column switch and connect to the steering-wheel electrical functions by a rotary connector. This SAS performs no further comfort functions: it is a stand-alone SAS.

An example of an SAS based on the 180° AMR sensor is illustrated in Fig. 7.11.29. The SAS has to measure several steering-wheel turns (up to 4.3) at a resolution of 0.1° while being true-power-on. After switching on the power supply the sensor has to deliver the absolute angle immediately. It was designed for the ESP from Bosch. Owing to this requirement there are two gears each carrying a magnet, which rotate over AMR sensors.



Fig. 7.11.29 Steering-angle sensor using the 180° anisotropic magnetoresistance sensor technology (source: Bosch)



Fig. 7.11.30 Example of a steering-angle sensor using anisotropic magnetoresistance (source: Bosch)

other, which leads to a permanent change in the phase relationship between both signals. This additional information is used to identify the rotational sections in which the individual AMR signals are repeated.

Other types of SAS are designed as stand-alone versions assembled in the lower part of the steering column, but still located in the passenger compartment resulting in less stringent environmental requirements. Newer versions are designed for integration in the EPS steering rack and therefore have to meet harsher requirements (e.g., -40 to 125°C temperature range). Fig. 7.11.30 shows a typical SAS integrated in the steering rack of an EPS system. These sensors benefit of the latest progress made in the field of AMR sensors.

The requirements arising from the various locations are different and have an influence on the final cost of SAS products. Stand-alone SAS will disappear in favor of integrated versions. Systems such as ESP, EPS, BCD, and AFS, are not handled separately anymore. System suppliers are starting to define interfaces in order to use available sensor information from other systems. Consequently SAS must be able to communicate via a vehicle network such as CAN.

7.11.4.4 The Mechanical Interface

The mechanical interface of an SAS is an important aspect to be taken into account during development of the SAS. If tolerances are not properly managed by the interface specification an additional superimposed nonlinearity will be the consequence. This additional fault angle is not created by the SAS itself, it is overlaid through its interface with other parts and their inherent tolerances (e.g., bearings, rotating shaft amongst others).

SAS that are integrated in TCM can use the existing mechanical structure of the TCM and its interface to the steering column, shown in Fig. 7.11.27. This leads to a lower overall system cost because of the use of existing components (e.g., electronics, housing).

7.11.5

Conclusions and Future Trends

SAS will be needed in future applications such as EPS, AFS, and SbW. These systems need a torque sensor to measure the steering torque applied by the driver in order to control an assisting torque. Hence a combination of SAS and torque sensor seems to be obvious. Products combining angle and torque measurement have been proposed (e.g., an opto-electronic angle and torque sensor for integration in EPS and EHPS systems [29]).

Later, steer-by-wire (SbW) systems will consist of a steering rack with electronic control including the necessary sensors, and a TCM for SbW, which also has electronics and sensors integrated. Further functions such as force feedback will then be necessary to form a proper interface for the driver. Both system parts are connected solely through wires. Tremendous efforts are still necessary to cover all safety related items regarding future SbW systems. It will no longer be sufficient to deliver angle sensor components with a conventional evaluation unit. The whole SAS has to be designed with respect to safety requirements in order to provide FSU and FTU capabilities.

7.11.6

References

- 1 US patent 4986124 **1991**.
- 2 US patent 6384597 **2002**.
- 3 US patent 6240652 **2001**.
- 4 US patent 2002/00114891 **2002**.
- 5 US patent 6064197 **2000**.
- 6 US patent 6158132 **2000**.
- 7 US patent 6311401 **2001**.
- 8 Patent Application: EP-0 862 728 **1998**.
- 9 Patent Application: EP-1 074 453 **2001**.
- 10 Patent Application: EP-0 922 627 **1999**.
- 11 Patent Application: EP-1 074 819 **2000**.
- 12 L. D. MILES, *Technique of Value Analysis and Engineering*, 3rd edn, Elenor Miles Walker **1989**.
- 13 US patent 6240652 **1999**.
- 14 KUCHLING, *Taschenbuch der Physik*, Deutsch, Thun **1981**.
- 15 US patent 5818038 **1998**.
- 16 K. C. J. DIETMAYER, SAE Technical Paper 1999-01-1040.
- 17 C. P. O. TREUTLER, *Magnetic Sensors for Automotive Applications*, Robert Bosch GmbH. EMSA 2000 conference
- 18 Patent Application: DE 195 06 938 **1996**.
- 19 Patent Application: WO 99/32345 **1999**.
- 20 TAOS, *Intelligent Opto Sensor Designer's Notebook*, Number 1; www.taosinc.com
- 21 US patent 5646523 **1997**.
- 22 T. I. BJENESCU, M. I. BZU, *Reliability of Electronic Components*, Springer, Heidelberg **1999**.
- 23 Patent Application: EP 0 159 191 **1985**.
- 24 Patent Application: DE 197 38 836 **1999**.
- 25 Patent Application: EP 1 083 408 **2001**.
- 26 US patent 4986124 **2001**.
- 27 R. HERB, *TRIZ Der Weg zum konkurrenzlosen Erfolgsprodukt*, Verlag Moderne Industrie, Landsberg/Lech **1998**.
- 28 *Application Note, Computer Optical Products Inc., Optical Encoder Applications*: www.opticalencoder.com
- 29 K. EISENHAEUER, J. ANDREW, J. BAXTER, D. BOYTON, S. ROBINSON, SAE Technical Paper 1999-01-1035.
- 30 The flexray group: www.flexray-group.com
- 31 The TTP group: www.ttgroup.org
- 32 H. KOPETZ, *Real-Time Systems, Design Principles for Distributed Embedded Applications*, Kluwer, London **1997**.

7.12

Force and Torque Sensors

MATTHIAS MÖRBE and CORD VON HÖRSTEN

7.12.1

Introduction

Force sensors are not used in general automotive applications. In agricultural machines a force sensor measures the load at the connection rod [1]. Some force sensors are specially made for measuring forces during vehicle testing and validation. Often metal strain gauges are welded to the devices under test. The data is sometimes needed for system development or to verify simulation results. The first and only mass production of torque sensors is in electrical steering systems [2].

7.12.1.1 General Physical Basics

Force applied to any object causes a reaction either as a movement or a change in magnetic characteristics, for example [3]. To measure this change requires a sensing element that sees this force or a part of it. There are four main parameters.

- Sensitivity.
- Operational range.
- Hysteresis.
- Limit values and overloads.

The distinction between a force sensor and a torque sensor is not clear, because in all cases a mechanical movement resulting from a force is converted into optical, electrical, or magnetic signals. This conversion defines the characteristics of the sensor. All sensors are made for limited ranges.

Sensitivity

Sensitivity is a general parameter for an application. The limitation depends on how the equipment converts a mechanical change into a visible signal. Therefore, the first thing that has to be decided is the maximum sensitivity of the system. This determines the technology of the transducer and the amplifier performance. The signal-to-noise ratio is important. For high sensitivity calibration is necessary. Zero point and absolute value calibration requires a master tool. In automotive applications, linearity has to be a design constant. In service a zero point calibration is accepted. Self-calibration algorithms are used and they are often commercially confidential.

Operational Range

While sensitivity defines the minimal response of a system, the operational range gives the spread between the minimal noise level and the harmful overload level. For a force or torque sensor these limits have to be defined for all circumstances in which the sensor will be used (Fig. 7.12.1).

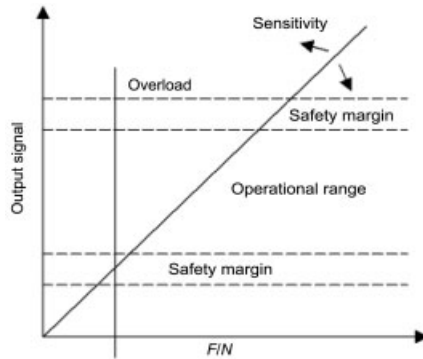


Fig. 7.12.1 Definition of sensitivity, operational range, safety margin, and overload

For safe working, a safety margin has to be defined at each end of the operational range. This gives the safe working range, which should encompass most of the likely signals. Signal filtering is discussed later.

Hysteresis

Reversible forces, F_{mess} , are forces that can operate in opposite directions; they can have all kinds of frequencies and amplitudes, which have to be converted into signals, F_{act} , without any additional or subtractive threshold, and offset or hysteresis F_h .

$$F_{\text{act}} = \pm F_h + F_{\text{mess}}$$

If there is measurable hysteresis, its time factor and thermal coefficient should be determined. The most difficult scenario is to have hysteresis that varies with amplitude, frequency, time, and temperature. Hysteresis requires detailed investigation when selecting transducers and amplifiers for a specific application. The limits should be verified for all likely circumstances, and possible failures should be identified. As well as hysteresis in the mechanical system, hysteresis in the amplifying circuit should be investigated. Fig. 7.12.2 shows that both kinds of hysteresis are possible.

Hysteresis effects are caused by the following factors.

- Saturation of spring rate.
- Technology.
- Friction.
- Mechanical play.
- Temperature.
- Aging of lubricants.
- Amplifier circuits.
- Parasitic electrical and signal line failures.
- Overload.
- Wear and tear.
- Dust and dirt.

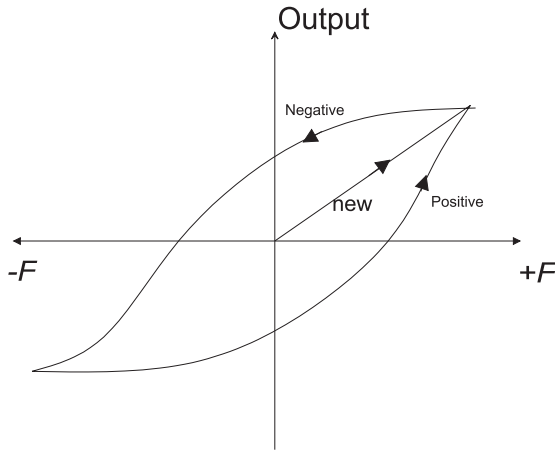


Fig. 7.12.2 Mechanical hysteresis

This summary is only a guide to searching in all areas influencing the measurement of torque or force.

Limit Values and Overloads

Fig. 7.12.3 is an illustrative diagram of a load-extension curve for tensile steel.

A force sensor combines a structure that responds to force changes with a sensor that converts these changes into a signal. A torque sensor similarly consists of a torsion or tension element detector, but it also has to transmit the force. Consider the torque for a electric power steering system any failure would immediately mean loss of control of the vehicle, so this must be avoided under all circumstances. Steering column torque requires careful finite element calculation, durability and performance testing to failure, and validation.

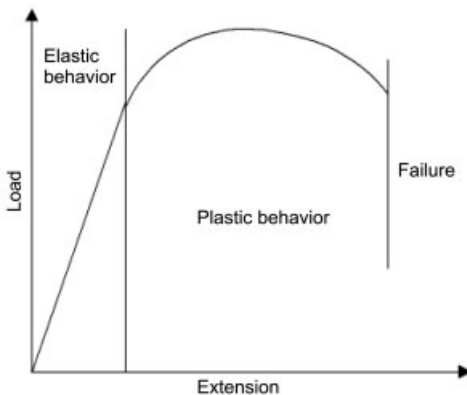


Fig. 7.12.3 Tensile behavior of steel

7.12.2

Automotive Applications of Torque Sensors7.12.2.1 **Powertrain**

The powertrain torque signal is used by all engine and transmission control units as well as by vehicle stabilizing systems. Engine control units presently in use estimate the torque at the end of the crank shaft from other sensed signals. To be commercially viable, a powertrain torque sensor must offer a higher accuracy than this estimated signal. Combining powertrain torque signals with continuously variable transmissions could lead to significant savings in fuel consumption, since the transmission ratio can be adjusted to operate the engine at the most economic point. So far, powertrain torque sensors are only available for development purposes: they are not yet in mass production.

7.12.2.2 **Steering**

Electric power-assisted steering systems are more and more replacing hydraulic systems. To improve fuel economy, the power assistance is provided by an electric drive. Such a system only consumes energy when power is supplied, unlike hydraulic systems. A torque sensor in the steering column provides the input signal to the drive control unit. An example of an electric power steering system is shown in Fig. 7.12.4.

Torque sensors measure angular displacement $\Delta\varphi$, which is proportional to torque. Fig. 7.12.5 shows the angular displacement caused by the torsion bar, which is fundamental to all automotive torque sensors. Many sensors that are established in the steering torque sensor market are based on potentiometer technology. These sensors experience a hysteresis of about 1% and a nonlinearity of about 3%. Major sensor manufacturers focus their research and development efforts on contactless torque sensors that have no mechanical link between the input and output shafts within the sensor. These sensors do not suffer from wear and their hysteresis and linearity performance are significantly better. Contactless steering torque sensors use one of three technologies: magnetic, optical, or inductive. Modern torque sensors are fully integrated into the steering gear design and support the installation requirements for a vehicle engine compartment.

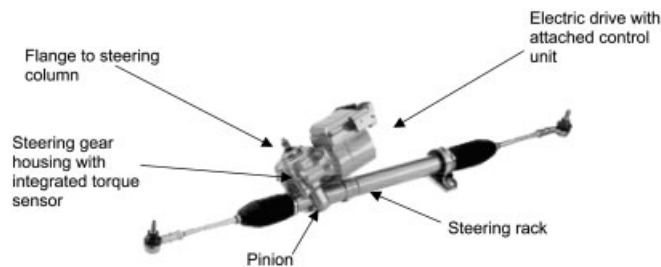


Fig. 7.12.4 Example of electric power steering

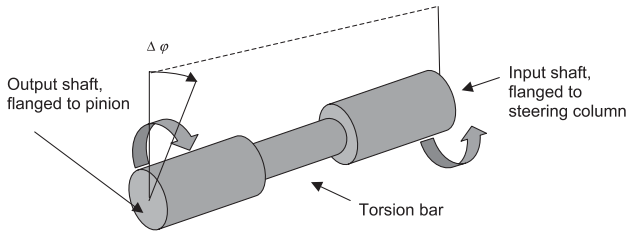


Fig. 7.12.5 Principle of torque sensors

One of the latest developments in the area of steering torque sensors is an optical concept created by Bishop Innovation with Bosch. As in conventional hydraulic steering systems, a torsion bar is a part of the steering column. A disc with a barcode on its surface is attached to each end of the torsion bar. A light source within the sensor module illuminates the surface of the coded discs, partially reflecting the light through a lens onto an optoelectronic ASIC.

The light intensity on the sensitive surface of the optical ASIC, an array of photodiodes, depends on the code of the discs. The intensity distribution over the photodiodes allows the optoelectronic ASIC to calculate the absolute angular position and the angular displacement of the steering shaft. Fig. 7.12.6 shows the sensor concept. This concept allows measurement of the absolute angular position of the steering wheel and the torque applied to the steering system.

ZF Lenksysteme GmbH developed another sensor for measuring steering torque. It has a torsion bar with a magnetic encoder ring on one end and an anisotropic magnetoresistive (AMR) element on the other. The AMR element detects the direction of the magnetic field, which depends on the angular displacement between the input and output shafts. Fig. 7.12.7 and Fig. 7.12.8 show the principle and the actual implementation, including a clock-spring. The number of turns of a steering system is limited, so an electrical connection between the electronic control unit and the sensing element is feasible using a clock-spring: no wireless transmission is required.

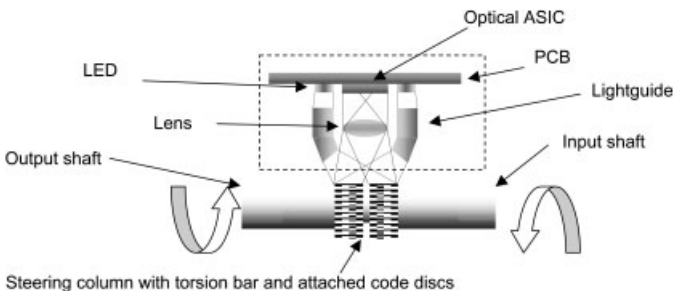


Fig. 7.12.6 Optical torque and angle sensor

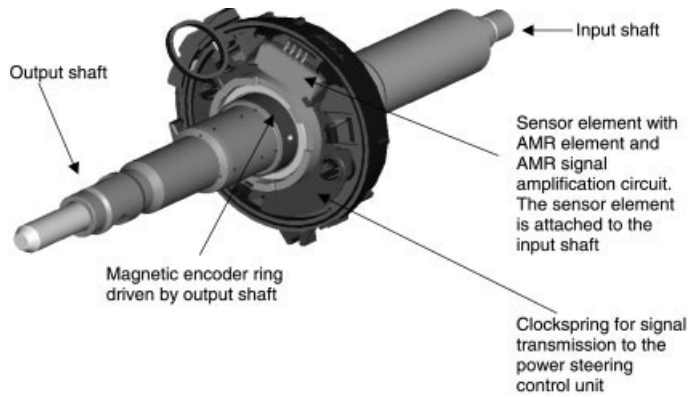


Fig. 7.12.7 Steering torque sensor using a magnetic encoder ring and an anisotropic magneto-resistive element

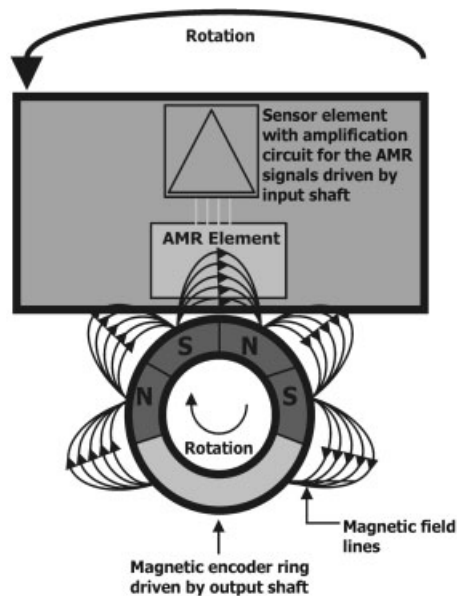


Fig. 7.12.8 Concept of torque sensor using anisotropic magneto-resistive technology

An inductive sensor was developed by Koyo Seiko Ltd. As shown in Fig. 7.12.9, the input and output shafts have detection rings attached. Detection ring 1 on the output shaft and detection ring 2 function together to change the inductance of the detection coil. Side effects such as temperature or electromagnetic radiation are compensated for via a compensation coil. The compensation coil inductance is torque independent due to the flush end of detection ring 2. Both coils are part of a Wheatstone bridge circuit. Applying a torque to the steering column varies the coil impedances, creating a voltage across the bridge circuit from which the torque can be determined. A more detailed explanation of the sensor is given elsewhere [4].

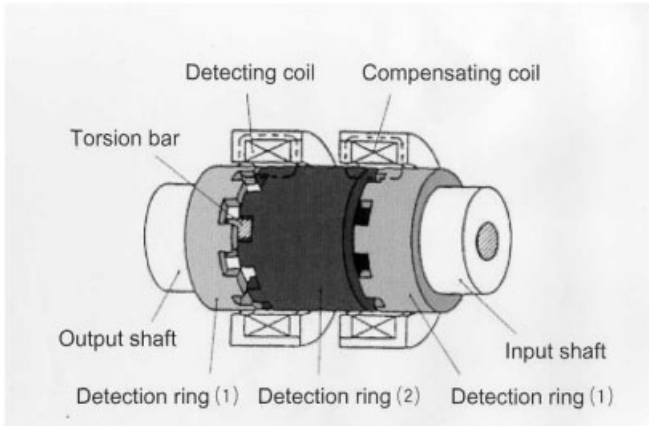


Fig. 7.12.9 Inductive torque sensor

Another optical sensor for measuring steering torque and absolute steering angle was developed by TRW Inc. Compared with the previously presented sensor by Bishop and Bosch, this is an alternative optical concept based on detecting the intensity of transmitted light. LEDs emit light through two different discs, each with a punched code. One of the discs is connected to the input shaft, the other to the output shaft. The intensity of the light that shines through the coded discs onto the optoelectronic integrated circuits varies with the angular displacement along the torsion bar. Fig. 7.12.10 shows the components of the sensor. The size of the punched gaps in the coded discs follows a sine and cosine distribution so that angle and torque are detectable.

A multipurpose torque sensor was developed by FAST Technology. The specifications of this sensor allows its use in electrically assisted steering systems. FAST Technology developed a method for permanently magnetizing ferromagnetic steel shafts. Fig. 7.12.11 shows the shape of the toroidal magnetic field. The magnetic field beside the shaft varies with the mechanical stress on the shaft. A sensor

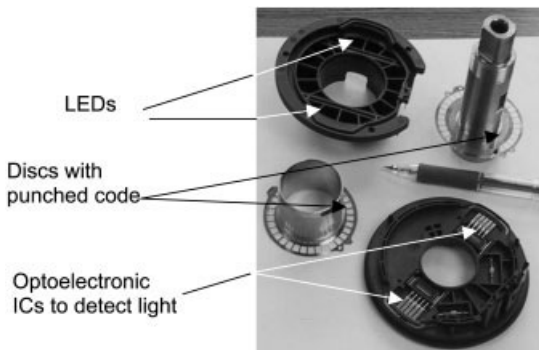


Fig. 7.12.10 Optical torque and steering-angle sensor

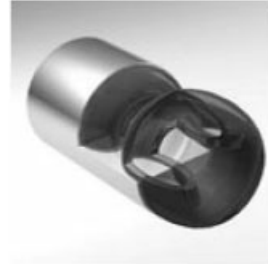


Fig. 7.12.11 Magnetized steel shaft for torque sensing

module that detects the magnetic field intensity beside the shaft permits torque measurement. The concept requires shielding against external magnetic fields.

7.12.3

Operating Conditions

There are various ways to design a torque sensor. They all have their specific areas of application. Beside those basic needs there are other values or influences of significant importance.

- Speed.
- Acceleration.
- Moved mass.
- All kinds of resonance effects.
- Dumping effects.

7.12.3.1 Speed

Speed limits the ability to transduce applied force or torque for all measurement methods. Apart from electrical limits of filtering and signal conversion effects like material internal friction, skin effects and non-homogeneous material behavior will define speed limits. In most cases the rotational speed defines the principle of signal transmission. This is one of the reasons why torque sensors are still not used with automatic transmission. The input torque is calculated by the engine management [5].

7.12.3.2 Acceleration

Normal accelerations do not affect sensor performance within the speed limits already mentioned. Endurance over long time periods, however, should be considered. Electrical stability is the main factor that might deteriorate in the long term, and corrosion should also be considered: a rusty torsion bar after 30 years in a car is thinner than a new one. At the other extreme, very short-term transient effects can be damaging. Acceleration spikes can cause damage to structures, which might not be recognized, if parameters are not adequately limited by low or high pass filters. Filtering is mainly electronic, but lubricants can also act as filter for applied forces.

7.12.3.3 Dynamic Mass and Acceleration

In a system in which applied force or resultant torque are being measured, the dynamic mass must be known. Precautions should be taken against damage resulting from a locking condition. A mechanical limiting unit has to guarantee safe operation. In automatic transmission gear boxes a rapid change of friction of the road surface can damage a sensor based on a torsion bar effect. A force sensor at a wheel hub unit can easily experience a force in excess of 100 g. If a vehicle wheel with an aluminum rim jumps with high brake force over a curb, such decelerations are possible.

7.12.3.4 Resonances

It is known that everything can be brought into resonance as long as there is the right frequency and damping forces are smaller than applied forces. Sensing elements can easily be destroyed if they go into resonance. Real-time testing and measuring is absolutely necessary to avoid damage over a lifetime. The actuation of a resonance condition can arise from either the sensor or its mounting. Using decoupling devices like rubber has a negative effect in terms of linearity or sensitivity.

7.12.3.5 Damping Effects

In most cases there will be an interface between the force transmitting object and the sensing element, except when the force transmitting object also produces the signal: for example, a magnetized bar is doing both.

The damping of an applied force has several consequences in terms of conversion into an output signal. The internal coupling factors of a rubber damping element have to be built into the dynamic modeling of the structure. Another area of damping effects arises from lubricant coupling, when the medium through which the force passes to the sensor influences the response.

The worst case is damping above the operational frequency. This means that the sensing element response includes a delay. For force and torque sensors there is no standard equipment on the market, as for electrical circuits. For signal transmission only radio-frequency based systems can avoid mechanical influences (Fig. 7.12.12).

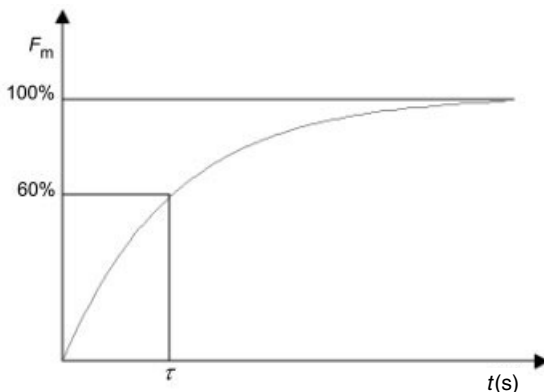


Fig. 7.12.12 Time constant of applied force/torque

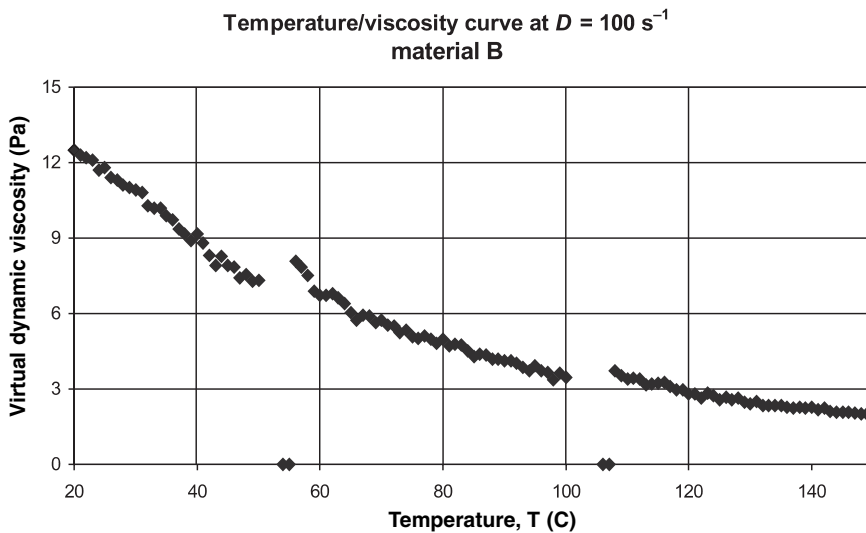
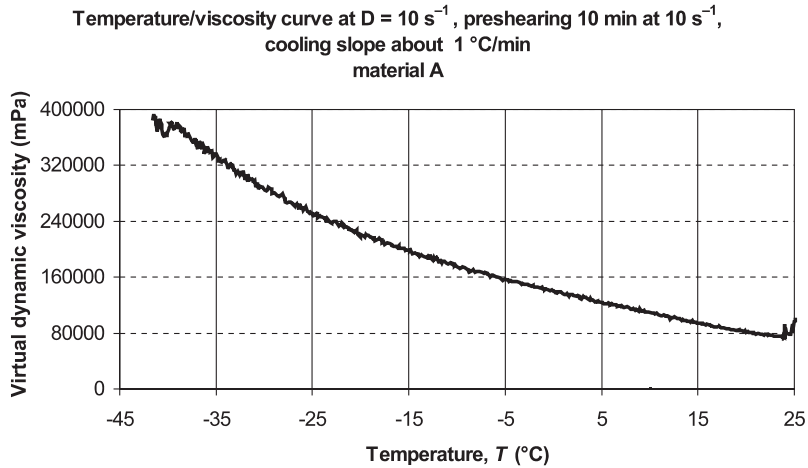


Fig. 7.12.13 Typical viscosity curve versus temperature for two different greases

Lubricants have a nonlinear viscosity coefficient against temperature. At very low or very high temperature this nonlinearity can create significant measurement failures. A careful analysis will show this effect within the transmission chain. Sticking is the worst case. It can happen that frozen water makes a short cut in torque sensors. In most cases damping factors are based on the electrical filter circuits (Fig. 7.12.13).

7.12.4

Control Systems using Torque Signals

One of the most important requirements for modern automotive control systems is the avoidance of applying vibration, harshness, or noise to the vehicle occupants. Sensors in the control loop have an impact on the performance of the control system because of their dynamic properties. In particular, steering torque control systems must avoid generating steering-wheel oscillations discernable by the driver when steering assistance is given. This determines the processing strategy for the electrical signals of the sensing elements and the data transmission to the controller unit. The time loop for calculating the control actuations determines the signal age requirements and the sensor refresh rate. Since the vehicle occupants' safety could be at risk through a malfunction of such a control system, a safe data transfer must be ensured.

7.12.4.1 Signal Age

Fig. 7.12.14 shows the four typical signal processing steps within automotive torque sensors. If the sensor is not equipped with a microprocessor, the duration of the sequence will not vary significantly. In that case some of the elements will be processed in the control unit. The signal age is defined as the time period from acquiring raw data to having a monitored torque value for the control algorithms inside the controller. In other words, each of these steps takes some time, the current torque value inside the controller reflects a torque that was captured by the sensor some time ago. In order to avoid any haptic impacts to the vehicle driver, the signal age should not exceed 1 ms for steering torque sensors.

7.12.4.2 Refresh Rate

The refresh rate is a measure of how often the sensor provides data to the control unit. If the signal processing within the sensor is designed to process older data while sampling current data, the refresh rate can be shorter than the signal age. Modern steering controllers require an update rate of 500 μs .

7.12.4.3 Common Interfaces of Torque Sensors

All automotive torque sensors form part of a system that could potentially lead to dangerous situations for the driver. Using incorrect signals that do not reflect the real torque must therefore be avoided under all circumstances. Various monitoring features within the sensors or the controllers are implemented. The data transmission from the sensor to the controller must fulfill these safety requirements, such that erroneous data transmissions are detectable. Another important

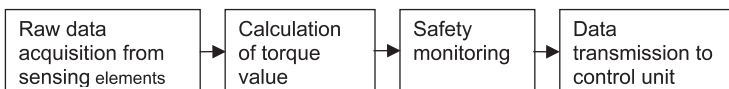


Fig. 7.12.14 Signal processing sequence as part of the control loop

requirement is the duration of the signal transmission. Torque-sensor interfaces can generally be split into analog and digital interfaces. In addition to the torque itself, modern torque sensors provide status information to the control unit. The status information indicates, for example, if the sensor is in normal operation mode or if the internal safety monitoring detected a failure.

Analog Signals

These interfaces are used on potentiometer-based torque sensors. The conversion of the current or voltage based signals into digital data is done by the microcontroller in the control unit. Disturbances to the signals, especially the voltage interface, can hardly be differentiated by the controller from undisturbed signals. Therefore the susceptibility to electromagnetic radiation should be carefully considered. Two-way communication for test and calibration purposes is only possible if additional equipment is connected with the sensor, which causes extra effort during manufacturing of the steering system.

Digital Serial Interfaces

The controller area network (CAN) became a standard for data communication of control units. This digital interface is also used to transfer torque sensor data to control units. The various monitoring options of the CAN protocol as well as the transmission of counters and checksums provides a reliable and safe data transmission. Other methods of serial digital data transfer are also an option for automotive torque sensors. The voltage difference between logic 1 and 0 should not exceed 2.5 V, otherwise the switching voltage of the driver circuits may limit the data transfer speed to 20 kBaud. The baudrate of serial data transfer for modern steering torque and angle sensors could be as high as 240 kBaud. Such a baudrate is required if 60 bits of data are required to represent the torque, sensor status, and possibly angle each 500 microseconds. This calculation also takes into account the 60 bit of checksums and other transmission control data required to ensure safe data transfer. Such requirements identify torque acquisition as a complex proposition for modern automotive control systems. Considerations must be made not just of the sensor itself, but also the entire signal processing and data transfer chain into the controller unit.

7.12.5

System Outlook

Sensors for force and torque measurement in automotive applications will always reflect the needs of the system in which they are used. For reasons of cost and reliability, most system designers try to reduce the number of sensors.

In the case of a torque sensor for traction control, output torque can be calculated by the engine management system from engine data. But when sensor designs become simple, very robust, and highly accurate, measurement with sensors will have advantages over calculation.

Beside the needs of the sensor module design the most important and sophisticated task is the selection of transducer or sensing element. All solutions include compromises. High sensitivity is not achieved without overload limitation or a blocking device has to be implemented, but such a possibility can only be selected if the system allows for it.

Generally the benefit from a force or torque sensor has to be significant in terms of system improvements in automotive applications. A replacement of hydraulic pressure sensors can be expected for the electromechanical brake system.

7.12.6

References

- 1 Bosch, *Kraftfahrtechnisches Taschenbuch*, 24th edn., Vieweg, Wiesbaden 2002.
- 2 H. STOLL, *Fahrwerktechnik: Lenkanlagen und Hilfskraftlenkungen* (ed.) J. REIMPELL, Vogel, Würzburg, 1992.
- 3 H. RUSER, U. TRÖLTZSCH, M. HORN, H.-R. TRÄNKLER, *11th ITG/GMA-Fachtagung*, March 2002, Ludwigsburg, VDE Verlag, 2002.
- 4 K. YOSHIDA, *Development of Custom IC for EPS Torque Sensor*, Koyo Engineering Journal English Edition E160, Nara, 2002.
- 5 G. LECHNER, H. NAUHEIMER, *Fahrzeugantriebe*, Springer, Berlin, 1994.

7.13

Light Sensors

IWAO YOKOMORI and YASUTOSHI SUZUKI

7.13.1

Introduction

An optical sensor that combines the functions of a twilight sensor and a solar sensor in one package is referred to as a light sensor. Light sensors have been on the market since 1997.

7.13.1.1 Twilight Sensor

Fig. 7.13.1 shows the essential functions of the twilight sensor. This sensor, intended to improve convenience and safety, automatically turns headlights and taillights on and off according to the ambient light level. The system also has the function of giving an alarm to remind the driver if headlights and taillights are on when the ignition is turned off.

7.13.1.2 Solar Sensor

The solar sensor is used in the air-conditioning system to compensate solar radiation. When an occupant is directly exposed to sunlight, he or she feels hotter than the actual temperature in the vehicle. The solar sensor therefore keeps the air temperature a little lower and supplies a slightly greater volume of air than the

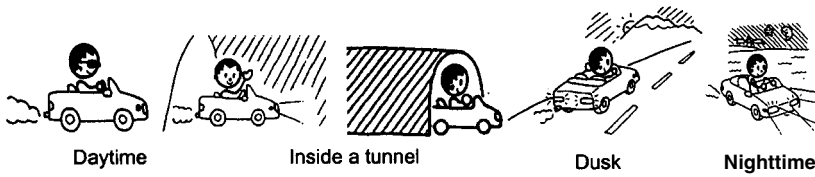


Fig. 7.13.1 Essential functions of twilight sensor

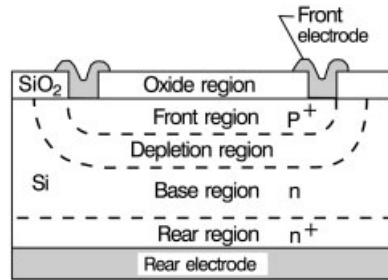


Fig. 7.13.2 Schematic diagram of photodiode

set values to increase occupant comfort. The light sensor uses a photodiode as the light-sensing element. Fig. 7.13.2 and Fig. 7.13.3 show a schematic cross section and the mode of operation of the photodiode [1, 2].

A reverse voltage is applied to a silicon p-n junction diode to make a thick depletion layer. Most of the light is absorbed in the depletion layer. The resulting electrons and holes are separated and caused to drift in the electric field within the depletion layer, producing a photocurrent.

Tab. 7.13.1 gives sample specifications for the light sensor. Since the light sensor is installed on the instrument panel, the upper limit of its operating temperature is 100°C. The light sensor is characterized by its wide light-intensity detection range and angular responses, in keeping with its use.

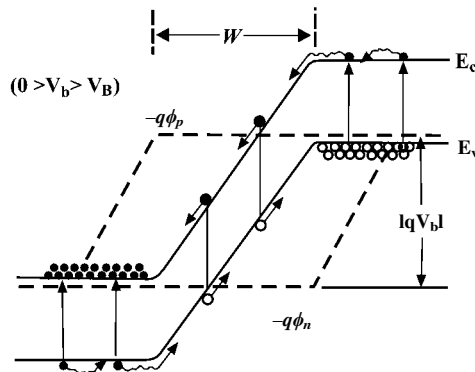


Fig. 7.13.3 Mode of operation of photodiode

Tab. 7.13.1 Example specifications for light sensor

<i>Parameter</i>	<i>Specification</i>
Operating temperature	-30 to 100 °C
Light intensity	0 to 100 000 lux
Angular response (solar sensor)	See Fig. 7.13.10
Angular response (twilight sensor)	See Fig. 7.13.10
Maximum voltage rating	8 to 16 V
EMC	200 V/m
ESD	± 25 kV

Since the driver perceives heat differently depending on the elevation of the sun, the solar sensor is required to possess an angular response appropriate to the vehicle's thermal load characteristics. When the sun is overhead, at an elevation of 90°, the vehicle's thermal load is rather small owing to the vehicle roof, while it is largest when the sun is at an elevation about 40°. Since the driver perceives sunlight even when the sun is on the horizon, at an elevation of 0°, the solar sensor needs to have a response capability even for that angle.

On the other hand, the twilight sensor needs to have near-zero horizontal response, in order to prevent the headlights erroneously turning-off when meeting an oncoming vehicle with bright headlights. The following section describes the technologies employed to achieve these responses.

7.13.2

Technologies

The three main factors relating to light-sensing technologies are the light-sensing element, the signal-processing circuit, and the package.

- The light-sensing element must have the function of detecting light intensity while having the optical characteristics required for automotive applications.
- The signal-processing circuit must be capable of covering a wide range of illumination, from twilight of only a few lux to full sunlight at 100 000 lux.
- A structure for installation on the instrument panel and filter design for desired optical characteristics are essential to the package.

The following sections, taking the one-direction light sensor as an example, give detailed explanations of the light-sensing element, the signal-processing circuit, and the package. The one-direction light sensor detects the sun's elevation. A dual-direction light sensor detects the sun's elevation and direction.

7.13.2.1 Light-Sensing Element

The light-sensing element is required to have the functions of detecting light intensity (light-intensity detection) and detecting the sun's elevation (elevation detection).

Light-Intensity Detection

A photodiode used as the light-sensing element provides an output current proportional to light intensity. To achieve the desired sensitivity, the sensor device is designed taking account of the structure and wavelength dependence of the photodiode.

The silicon photodiode output is analyzed using a device simulator intended for analyzing the semiconductor device electrical characteristics, with the added capability of analyzing the movement of electrons/holes when the silicon photodiode is exposed to light. Fig. 7.13.4, Fig. 7.13.5, and Fig. 7.13.6 show a photodiode analysis model, a mesh diagram, and an impurity density distribution map.

The silicon is finely meshed near its surface, where a diffusion layer is formed. For each mesh, the following values are calculated: potential ϕ , donor concentration in n-type semiconductor material N_n , acceptor concentration in n-type semiconductor material N_p , donor mobility μ_n , acceptor mobility μ_p , donor current density J_n , and acceptor current density J_p .

The photocurrent of the photodiode has a dependence on the wavelength of light. This is referred to as the spectral response. Fig. 7.13.7 and Fig. 7.13.8 show the analysis model and spectral response analysis result.

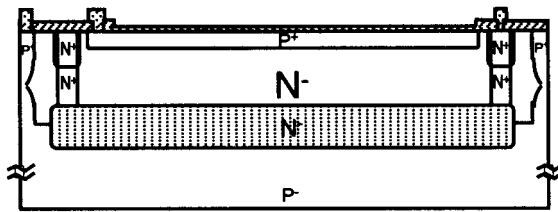


Fig. 7.13.4 Analysis model

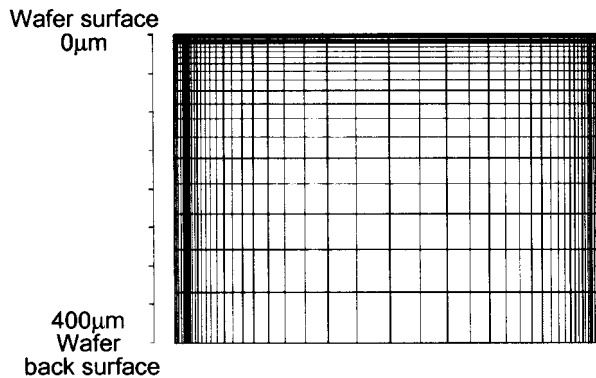


Fig. 7.13.5 Mesh diagram

Wafer surface

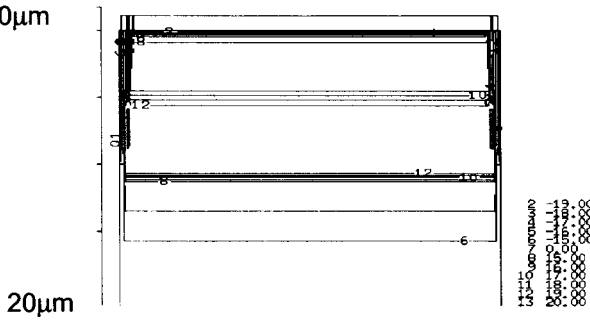
0 μm 

Fig. 7.13.6 Distribution map for impurity density

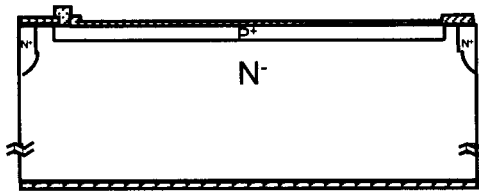


Fig. 7.13.7 Analysis model

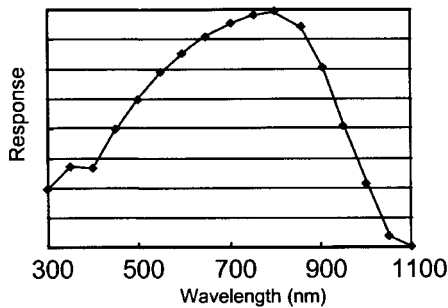


Fig. 7.13.8 Spectral response

A photodiode with a large N region, reaching near the underside of the silicon substrate, has a spectral response extending to the long wavelength range. For wavelengths of 1.1 μm or more, however, the photon energy is less than the band-gap energy of the silicon, so that few electron-hole pairs are generated.

Another factor affecting the spectral response is the surface reflectivity of the photodiode protection film, which is made of materials including SiO_2 . The refractive indexes for air, SiO_2 , and silicon are 1, about 1.45, and about 3.5, respectively. The differences in the refractive indexes result in the photodiode's spectral response being dependent on SiO_2 thickness. The photodiode including its protection film is therefore designed optimally to minimize the surface reflection effect.

Elevation Detection

The sensor's angular response is determined by a design in which a Fresnel lens is used as a filter. We use a simulator capable of analyzing sensor response for light from every direction to design a sensor structure with the desired angular response.

Fig. 7.13.9 and Fig. 7.13.10 show a light sensor assembly photographs and angular responses for both the solar sensor and the twilight sensor. The sensor is set to reach its maximum output when the sun is at an elevation about 40° , which is when the driver feels hottest.

7.13.2.2 Signal-Processing Circuit

Fig. 7.13.11 and Fig. 7.13.12 show the circuit block diagram and the system block diagram for the light sensor.

The control ECU software turns lights on and off in response to the output signal from the light sensor. The light sensor produces an output in the form of frequency. That means the sensor can be used not only for lighting control and solar radiation compensation, but also for DRL (daylight running lamp) dimmer control and HUD (head-up display) rheostat control. The sensor is designed to cover a wide range of light intensity from twilight to full sunlight thanks to its piecewise-linear characteristic.

The signal-processing circuit consists of the following four major circuit constituents.

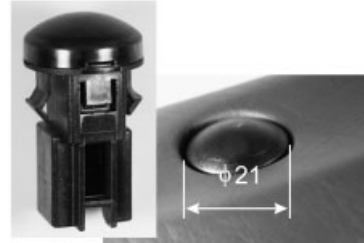


Fig. 7.13.9 Assembly

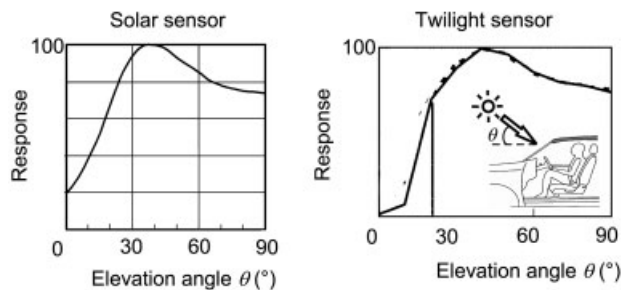


Fig. 7.13.10 Angular response

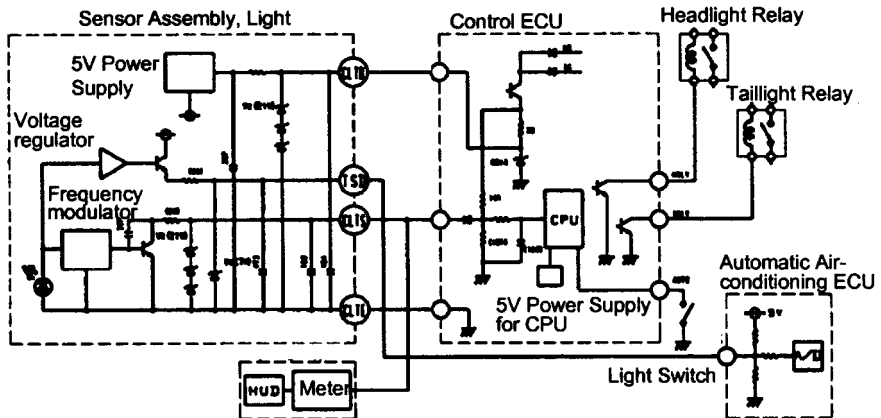


Fig. 7.13.11 Circuit block diagram

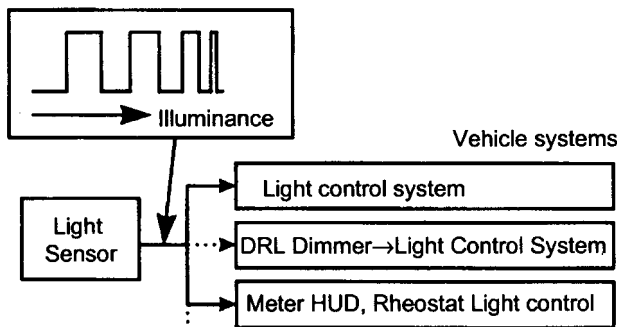


Fig. 7.13.12 System block diagram

- Current mirror: distributes the photodiode output current to the solar sensor output and the twilight sensor output.
- Circuitry for setting intercepts for piecewise-linear characteristics: sets the intercepts for piecewise-linear characteristic of the twilight sensor output.
- Circuitry for setting gradients for piecewise-linear characteristics: utilize a three-input OP amplifier to set the gradients for piecewise-linear characteristic of the twilight sensor output. Since the three inputs are at the lowest transistor base potential, the sensor's characteristics are as shown in Fig. 7.13.14 when the inputs are V1, V2, and V3.
- Circuitry for setting intercept for zero lux: sets an intercept so as to make the sensor output 50Hz at zero lux.

Fig. 7.13.15 shows a photograph of the light sensor device.

The signal-processing circuit is protected by an aluminum light shield to avoid accidental starting of elements such as a transistor when the sensor is exposed to sunlight.

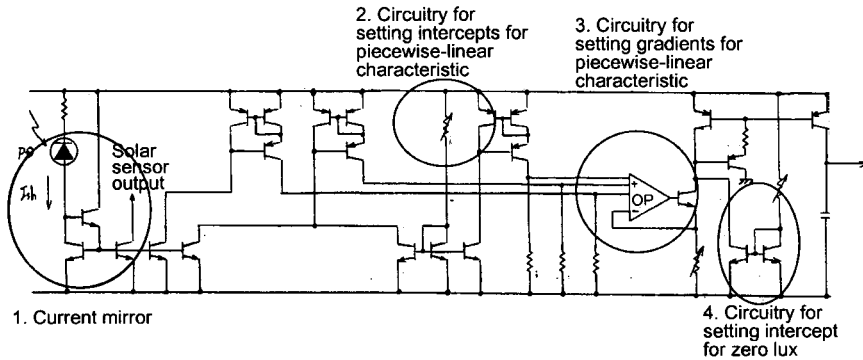


Fig. 7.13.13 Signal-processing circuit for light sensor

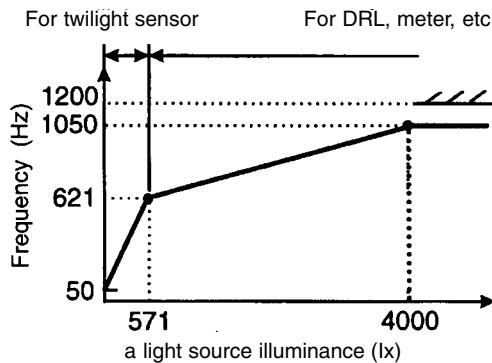


Fig. 7.13.14 Piecewise-linear characteristics of light sensor

7.13.2.3 Package

The light-sensing package features bare chip mounting for sensor installation on the instrument panel. Fig. 7.13.16 shows the assembly structure of the light sensor. The light sensor device is bonded to the middle terminal and then wire-bonded to the outer terminals. The top of the device is coated with gel to protect the Al wire. Chip capacitors for EMC are bonded between terminals with conductive adhesive.

The package also features a filter with two roles. One of the roles is to keep the light sensor spectral response from varying depending on the windshield spectral transmittance. The other is to achieve the desired angular response.

Nowadays windshields with low infrared transmittance have been increasingly used for heat absorption. We therefore select the filter material so as to make the light sensor unsusceptible to transmittance variations between windshields. We also regulate the intensity of the light entering the photodiode according to the light direction with a Fresnel lens, in order to achieve the desired angular response. Fig. 7.13.17 shows the principle of the Fresnel lens. Lights from overhead, from the horizontal direction, and from any other direction are regulated to 80, 50, and 100%, respectively.

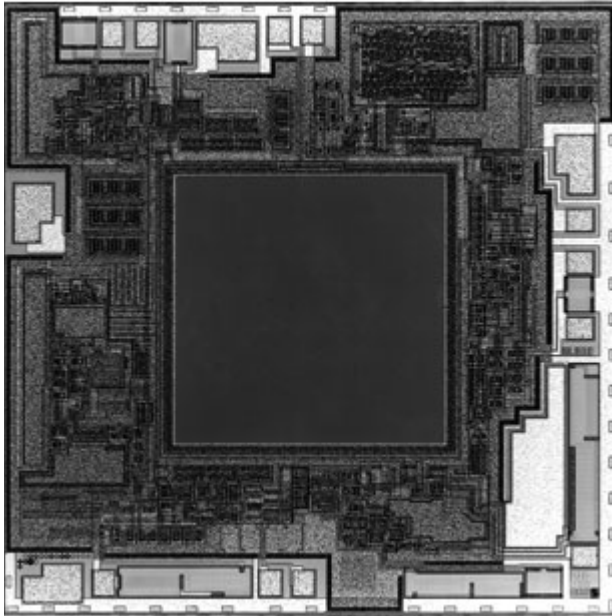


Fig. 7.13.15 Light sensor device photograph

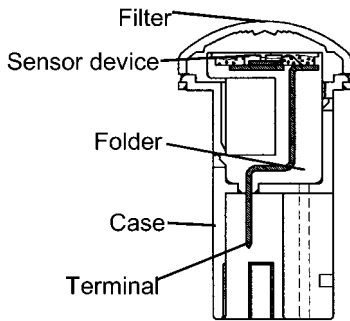


Fig. 7.13.16 Light sensor assembly structure

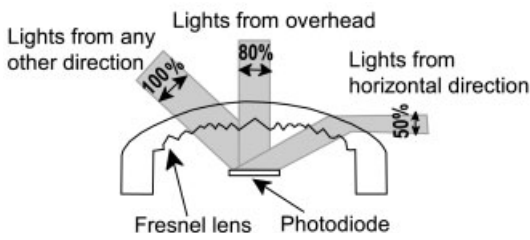


Fig. 7.13.17 Fresnel lens principle

7.13.3

Other Examples**7.13.3.1 Solar Twilight Sensor**

In this example, the desired angular response is achieved by improving the package. Fig. 7.13.18 and Fig. 7.13.19 show the package and angular response [3]. To produce a reduced response when the sun is overhead, the package has a diffuser called a 'retroreflector' molded into the outer lens cap.

7.13.3.2 Dual-Direction Light Sensor

The dual-zone climate-control system provides independent air conditioning temperature adjustment between the driver and passenger. The system needs a dual-

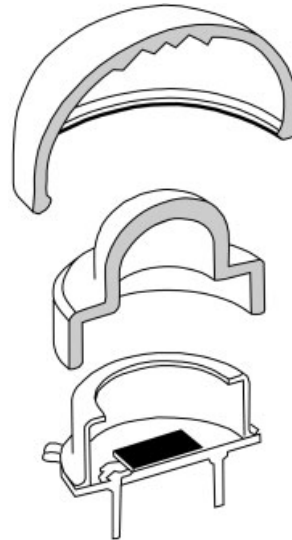


Fig. 7.13.18 Package

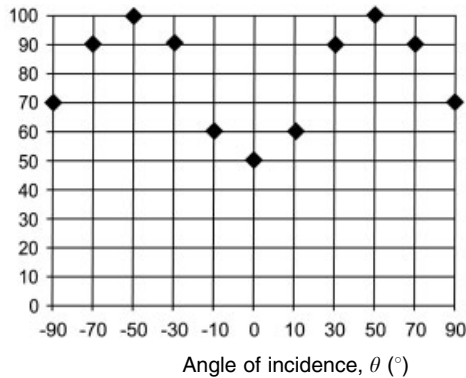


Fig. 7.13.19 Angular response

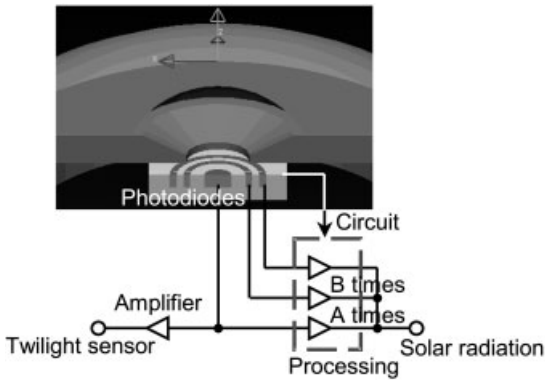


Fig. 7.13.20 Circuit configuration for desired angular response

direction solar sensor capable of measuring the light intensity on both the driver's side and the passenger's side simultaneously. The dual-direction light sensor integrates a dual-direction solar sensor and a twilight sensor into one chip. Fig. 7.13.20 shows the circuit for the desired solar sensor angular response.

Sunlight passes through the hole of the light-proof plate and enters the separated photodiodes, which are arranged in a ring. The horizontal direction of sunlight is determined based on the amounts of light received by the photodiodes. The circuit processes the output currents from the above photodiodes, proportional to the amounts of light, so as to achieve the desired angular response.

7.13.4

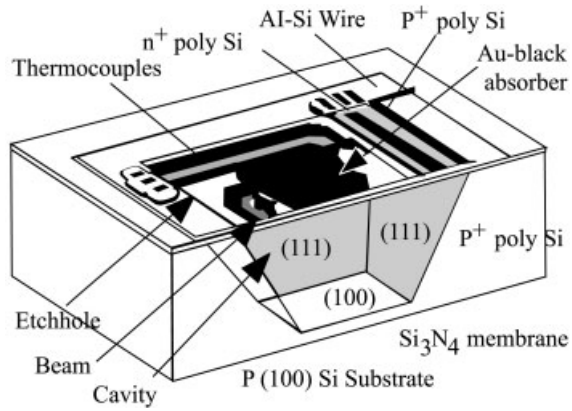
Conclusions and Future Trends

Conventional air-conditioning systems have effected solar radiation compensation by means of a solar sensor. However, systems that detect human body temperature with an infrared sensor are also coming along. The infrared sensor is capable of sensing energies radiated by hot bodies such as a human body, since those energies are in the infrared region.

Fig. 7.13.21 shows a cross section of a thermoelectric infrared sensor. This structure has potentially low cost because of its high compatibility with the widely used SiCMOS ICs [4]. To improve sensitivity, the sensor uses a swastika pattern with a curved beam and a gold-black absorber with good IR absorptivity so as to attain a UV absorption coefficient of 90% or larger.

Worthy of special note is the image sensor, one of the optical sensors employed as a safety feature. General-purpose CCD image sensors for daytime use are already available; however, an infrared image sensor for nighttime use will foreseeably come to be in demand. Night vision capability is already starting to be installed in vehicles. For it to gain wider use, technological innovation is required to reduce the cost of the sensor device and the signal-processing circuit.

Fig. 7.13.21 Cross section of thermoelectric infrared sensor



7.13.5

References

- 1 J. GEIST, *Sensor Technology and Devices* (ed.) L. RISTIC, Artech Norwood, 1994, p. 320.
- 2 H. YONEZU, *Optical Communications Device Engineering*, Kogakutosho, Tokyo, 1984, p. 314 (in Japanese).
- 3 W. BULTER, *Proc. Sens. Expo Baltim.* 1999, 183–190.
- 4 M. HIROTA, F. SATOU, S. MORITA, *Tech. Dig. Sens. Symp.* 2000, 17, 151–156.

7.14

Rain Sensors

ANDREAS SCHNEIDER

7.14.1

Introduction

Nowadays more and more sensors are placed in a vehicle to increase active and passive safety. One sensor with high market penetration is the rain sensor: for comfort and safety. If the car has a rain sensor, the driver does not need to adjust the windshield wiper setting to stop the wiper blades rattling or to have a better view through the windshield. When driving on wet roads the driver does not need to take a hand from the steering wheel to switch on the wiper if a truck splashes water onto the windshield. So the driver can concentrate completely on driving. If an additional 'assistant light switch' is installed, the headlamps will be switched on in dull wet conditions, without wasting fuel by having the lights on all the time. A common rain sensor is the Bosch RSM3ALS (rain sensor module including assistant light switch) shown in Fig. 7.14.1.

7.14.2

Operation of Rain Sensor

The optical system used by most rain sensors consists of a light emitting diode (LED), a light receiving diode (LRD), an ambient light sensor, an electronic control unit (ECU), and several lenses (Fig. 7.14.2).

Light from the LED is reflected at the outer surface of the windshield at the angle of total reflection. The angle has to be between 42° (glass-air) and 63° (glass-water). If water is on the windshield, some of the light is coupled out and this causes a reduced current from the LRD, which is evaluated by the electronics. The area on the windshield where light from the LED is reflected into the LRD is

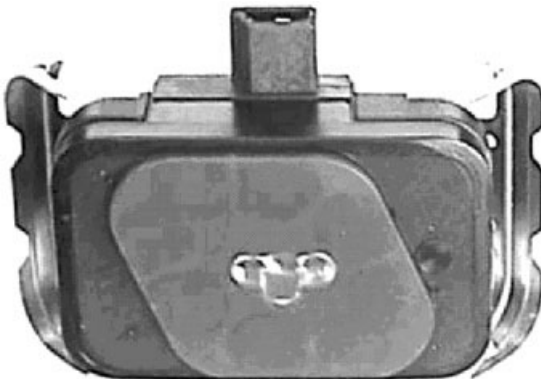


Fig. 7.14.1 Rain sensor module including assistant light switch

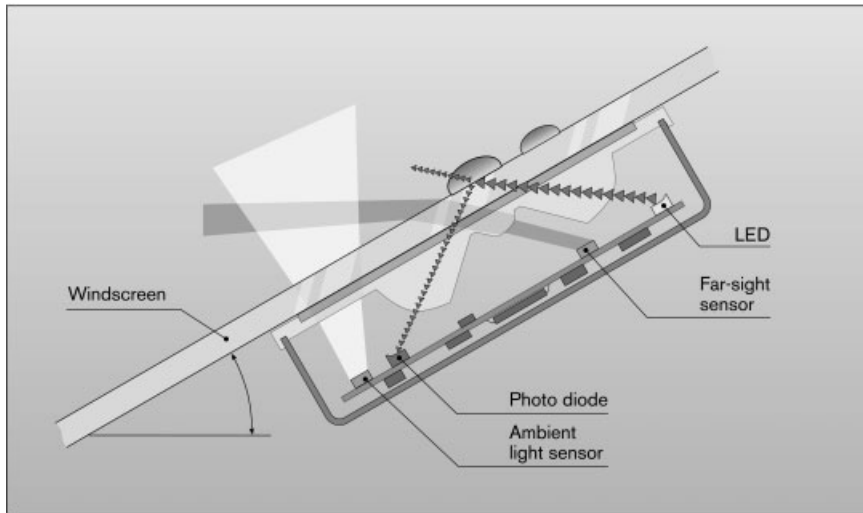


Fig. 7.14.2 Principle of rain sensor optics

called the 'sensitive area' of the sensor. Rain can be detected only when water drops hit this area. There must be a good ratio between the windshield area and the sensitive area in order to have a sensitive and reliable system.

7.14.3

Automatic Windshield Wiper System

An automatic windshield wiper system consists of a rain sensor, a wiper motor, a column switch, and relays or an ECU (Fig. 7.14.3). The rain sensor is mounted on the windshield in the wiped area to detect rain and wiper motion in such a way that the driver always has a good view. There are two different systems.

- Stand-alone rain sensors: in this system the rain sensor is directly connected to the wiper column switch, the wiper motor relays, and the wiper motor park signal.
- Rain sensors in a network: in this system the rain sensor is connected to a bus system, it receives all information and sends all commands via this network.

After the driver switches on the system, the rain sensor controls all wiping actions. It will command single wipes at low-speed wiping or high-speed wiping. Because every driver has a different expectation of how the wiping system should react, a sensitive setting is used to adapt the system to driver needs. With the new Bosch electronic wiper motors, it is possible to have continuous wiping with continuous variable speed.

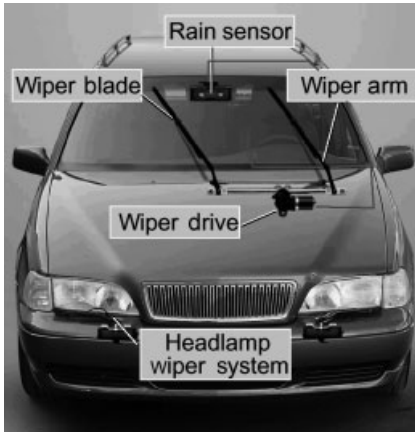


Fig. 7.14.3 Rain sensor integrated in a wiper system

7.14.4 Signal Evaluation

The diagrams in Fig. 7.14.4 illustrate how an analog signal can be evaluated by a microcontroller.

7.14.4.1 Transmitter

The transmitter consists of a digital-analog converter (DAC) and a current source, programmed by the analog voltage of the DAC, which determines the intensity of the light transmitted by the LEDs. Normally two to four LEDs are used. It is important to adjust the transmitter current, because the efficiency of the light conversion is very variable and has a temperature gradient. Normally light is pulsed

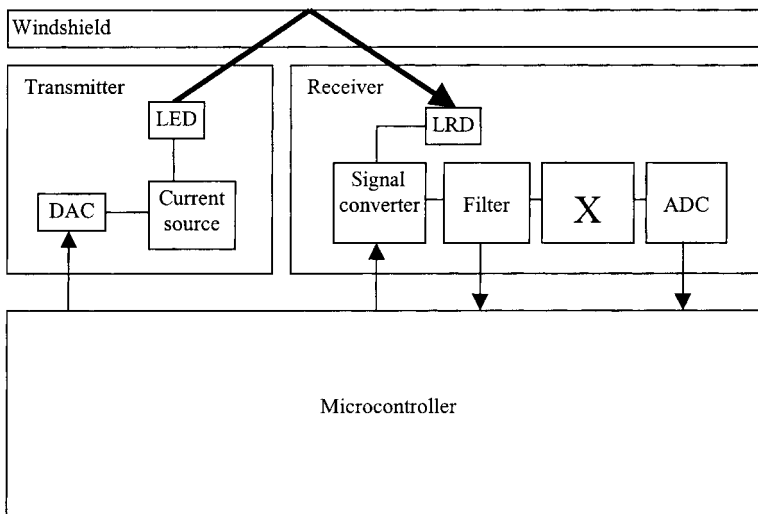


Fig. 7.14.4 Principle electronic schematic of the rain sensor module

to reduce power loss at the LEDs and to increase the current. Another reason for the wide range of transmitter currents used is that light transmission through the windshield is not well specified and has a wide tolerance. Common windshields are between 4 and 6 mm thick and have a certain transmission level for IR light perpendicular direction to the surface with a lower and upper limit, which can vary between suppliers.

7.14.4.2 Receiver

The receiver consists of several LRDs, a current-voltage converter, a filter to eliminate dc or low-frequency offsets, an amplifier, and an ADC, which is normally included in the microcontroller. The LRD is switched on or off by the microcontroller to control which optical path is active and to switch off an LRD if disturbance glare is present. The wide range of transmission is the reason for using a variable current-voltage converter to ensure good conditions for the rest of the receiver system. After conversion, the signal is filtered. All dc or low-frequency disturbance are eliminated to get the pure signal. Disturbances are normally caused by ambient light. The quantity of disturbance is measured by the microcontroller to decide whether the optical path is satisfactory or possibly saturated by the sunlight, and to give an indication of the accuracy of the signal. The filtered signal is amplified and converted by the ADC.

7.14.4.3 Microcontroller

The microcontroller controls the whole system and evaluates the signal. Before measurements are performed, the best operating point is evaluated. This is done by testing which ratio of the converter in the receiver is least needed to have a saturated signal at the ADC with maximum transmitter current. After the transmitter current is evaluated, it generates a signal at the receiver, which is between a defined upper and lower limit. The upper limit is determined by saturation effects and the lower limit by the accuracy of the ADC. The sensor starts working at this operation point. The disturbance and the signal is measured and evaluated frequently. The gradient of the receiving signal over time is used by the algorithm to evaluate the correct wiping strategy. In addition, ambient light is measured by an extra sensor to detect day or night conditions. At night, the sensitivity is increased because water drops have a bigger influence on the visibility, especially if other cars with bright headlamps are approaching.

7.14.5

Wiping Modes

Different wiping modes are used to optimize behavior for all rain conditions: Fig. 7.14.5.

7.14.5.1 Direct Mode

In the direct mode there is a single wipe at low wiper speed. This is the basic operating state when automatic mode is switched on and when the system detects a dry windshield. From this state, every (rain) event directly triggers a single wipe.

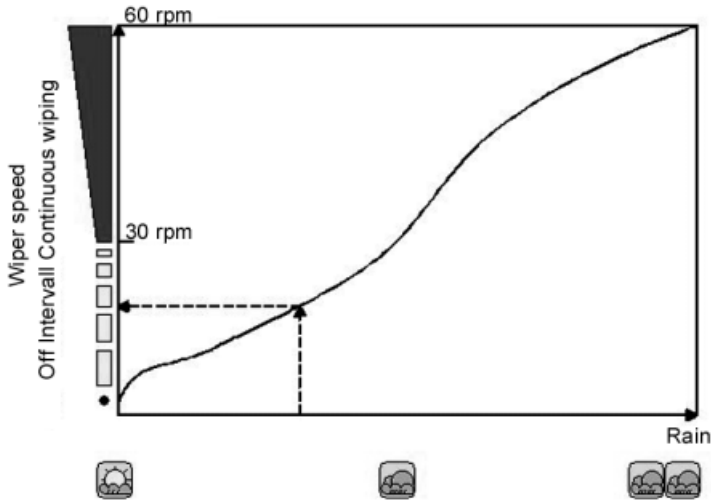


Fig. 7.14.5 Ratio wiping mode/speed over rain intensity

During this wiping process, the system decides to stop, to continue wiping at low speed, or to switch over to high speed.

7.14.5.2 Intermittent Mode

This mode is used in light rain, and delivers single wipes at low speed. The sensor detects droplets every few seconds. The time between two rain events determines the interval between two single wipes. At the end of every interval, the timing is recalculated. More rain generates a shorter interval and less rain a longer interval. In the calculation, former intervals are taken into account to get a harmonic appearance. Otherwise, 'nervous behavior' of the wipers will irritate the driver. If the calculated interval is longer than the maximum interval, the system switches to direct mode. During every wiping cycle the sensor checks whether to switch to continuous wiping.

7.14.5.3 Continuous Wiping

Rain events are counted during a wiping cycle and the sizes of the drops is evaluated. This is used to get information about rain intensity. Depending on this intensity, the rain sensor commands low speed, high speed, or a discrete wiping speed (e.g., 50 rpm). Dynamic hysteresis, which depends upon the events in the last cycles, prevents the system from switching rapidly from one speed to another. Using a time reference of the park signal of the wiper motor ensures correct operation under all conditions with all wiper motors. If a motor wipes at low speed, the detection time is reduced and the threshold between low speed and high speed is automatically increased. The rain sensor module always reduces wiping speed from high speed to low speed before stopping, to protect the wipers against mechanical stress.

7.14.5.4 Detection of Special Conditions

Splashes

A rapid change with a high degree of wetness (high-amplitude signal breakdown) can immediately trigger a high-speed wiping cycle in single wiping modes (direct mode or intermittent mode). This can be caused by a truck that splashes water onto the windshield and generates a critical situation for the driver. In this case, the rain sensor has to wipe the area as fast as possible. When splash-water is detected, the system switches over to high speed not only when the wipers reach the parked position but also during the current wiping process at low speed. The wiping process, which had already commenced at high speed is followed by another at high speed, and then a third one at low speed. The system then returns to the original state.

Smearing

A smear is a fine water film that more or less rapidly dries. A smear develops when a dirty or damaged rubber element strokes over a windshield with some rain drops. In direct and intermittent mode, smearing is detected when the sensor signal indicates an abrupt signal decrease caused by the distributed moisture. The drying film results in an increasing sensor signal (or decreasing wetness). In this case, only the detection of water drops remains active to avoid unnecessary wipe cycles.

Dirt

Contrary to a wet windshield, the system recognizes a dirty windshield when it can no longer detect any improvement from the strokes of the wiper blades. In this way, the system prevents a dirty windshield from being wiped.

7.14.6

Mechanical Construction

There are two ways to create a good coupling between the rain sensor and the windshield. The first is to fix the rain sensor with transparent tape to the windshield. The second involves the rain sensor having a soft coupling pad that is pressed to the windshield by springs mounted on a glued bracket. The gluing process, normally done by the glass manufacturer, has to be done in a vacuum area to avoid air bubbles, which lead to malfunction or at least poor appearance. If the windshield or the rain sensor is broken, both must be replaced with higher costs. The press-on-coupling needs some more space on the windshield for the bracket, but a standard PU gluing process is used. Other advantages are as follows.

- Different radiuses of the windshield are compensated by the soft coupling material.
- The complete sensor can be tested at the supplier.
- No extra bracket is needed for a beauty cover.

7.14.7

Additional Features

The position of the rain sensor on the front wind shield make it tempting to implement additional optical sensors. The most popular combination is rain and light sensors. Only an additional ‘far-sight’ sensor is needed. With information from the ambient light sensor and the far-sight sensor, the microcontroller can decide if it is night or day, whether the car is entering a tunnel or a bridge, and sends switching commands to an ECU that controls the lamps of the vehicle.

The ambient light information can used in the vehicle system to control the dashboard illumination and the air conditioning. An additional function of the rain sensor can be to detect sudden rain and close open windows or sun roofs: in conjunction with an ‘anti-pinch’ device to prevent occupant injury, of course.

7.15

Chemical Sensors for Oxygen Detection and Air/Fuel Ratio Control

JOHANN RIEGEL, HANS-MARTIN WIEDENMANN and HARALD NEUMANN

7.15.1

Automotive Exhaust-Gas System Requirements for Oxygen Sensors

Oxygen sensors in automotive applications are used to measure the air fuel (A/F) ratio of engine exhaust gases and to control the optimum A/F ratio for perfect exhaust gas after-treatment by catalytic converters [1]. Therefore, they are also known as lambda or A/F ratio sensors.

7.15.1.1 **Exhaust-Gas Composition and Normalized Air/Fuel Ratio Lambda**

The raw exhaust-gas composition of a spark-ignition engine strongly depends on the ratio between the intake air mass and the injected fuel mass (Fig. 17.15.1). Because of the importance of the stoichiometric mixture the normalized A/F ratio λ has been defined as

$$\lambda = \frac{(m_{\text{air}}/m_{\text{fuel}})_{\text{current}}}{(m_{\text{air}}/m_{\text{fuel}})_{\text{stoichiometric}}} \quad (1)$$

where

$\lambda > 1$, excess of air, lean mixture

$\lambda = 1$, stoichiometric air/fuel ratio

$\lambda < 1$, excess of fuel, rich mixture

For stoichiometric combustion approximately 14.7 kg air are necessary for 1 kg gasoline. However, even with stoichiometric mixtures the combustion will be never perfect. Beside the main reaction products CO_2 and H_2O , we find signifi-

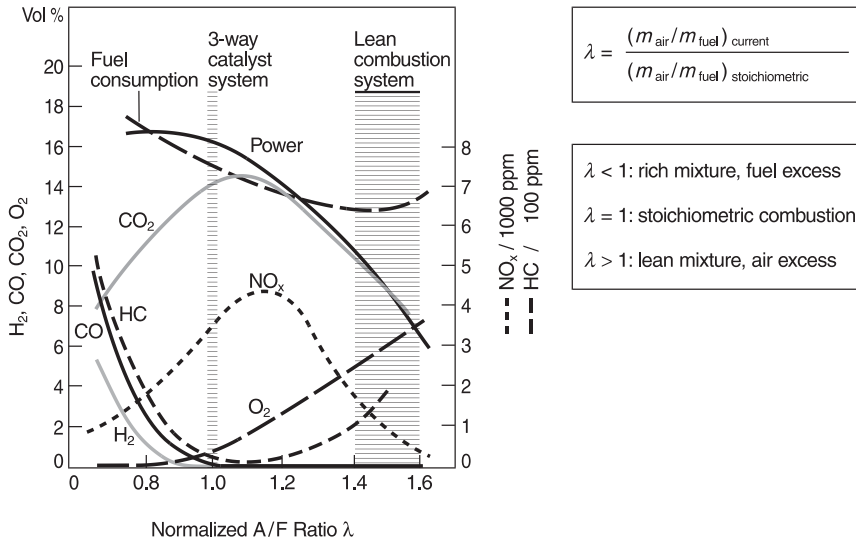


Fig. 7.15.1 Engine exhaust gases and normalized A/F ratio λ

cant concentrations of H_2 , CO, and HC as rich components, as well as NO_x and O_2 as lean components. To keep the harmful CO, HC, and NO_x below the legal limits at the end of the exhaust pipe the exhaust gas undergoes an after-treatment in a catalytic converter [2].

7.15.1.2 Air/Fuel Operating Conditions in Engines

Spark-ignition engines dominate the worldwide market for automotive applications today, and are commonly operated around the stoichiometric A/F ratio $\lambda=1$. A small, slightly rich range between $\lambda=0.995$ and 1.000 has been found to achieve the best emission conversion in combination with a three-way catalyst (TWC) intended to keep CO, NO_x , and HC simultaneously at a minimum (Fig. 17.15.2). Additionally it has been shown that a small continuous λ variation around $\lambda=1$ increases the conversion efficiency of an aged TWC. However, sometimes the overall optimum requires engine operation away from $\lambda=1$. The forthcoming emission legislation by the Californian air resource board, CARB (Tab. 7.15.1) and the extremely low emission limits of LEV (low-emission vehicles), ULEV (ultra LEV), and SULEV (super ULEV) are demanding faster light off of the catalysts to avoid the emission peak during the starting phase [3, 4]. To standardize the emission certification, CARB defined a typical driving cycle representing the driving pattern of an average Los Angeles resident. This driving cycle starts with a 20 s idle phase followed by a load/speed pattern typical for an urban environment. In the same way a typical European driving cycle beginning with an 11 s idle phase has been defined [5]. The length of the idle phase determines the required start-up time of both catalyst and oxygen sensor. This could only be achieved by dividing the catalytic converter into a low-volume pre-catalyst and a high-volume main catalyst

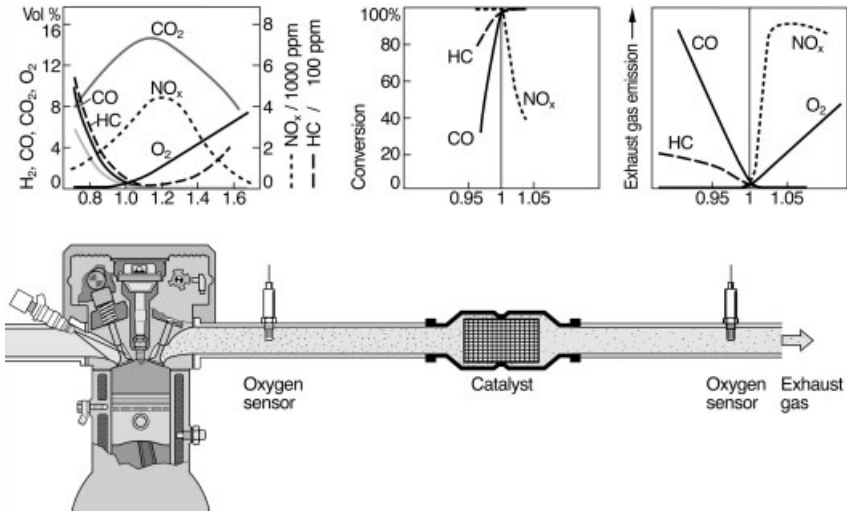


Fig. 7.15.2 Exhaust-gas after-treatment

Tab. 7.15.1 Exhaust emission limits

West European emission limits (ECE/EU driving cycle)

Standard	Introduction	CO (g/km)	HC (g/km)	NO _x (g/km)	HC + NO _x (g/km)
EU Stage 1	7.92	2.72	–	–	0.97
EU Stage 2	1.96	2.2	–	–	0.5
EU Stage 3	1.00	2.3	0.2	0.15	–
EU Stage 4	1.05	1.0	0.1	0.08	–

USA and California emission limits (FTP 75 driving cycle)

	Model year	Standard	CO (g/mile)	HC (g/mile)	NO _x (g/mile)
USA	1994	Tier 1	3.4	0.25	0.4
	2004	Tier 2	1.7	0.125	0.2
California	1997	LEV	3.4	0.075	0.2 (25% of fleet)
		ULEV	1.7	0.04	0.2 (2% of fleet)
	2004	SULEV	1.0	0.01	0.02

LEV=Low emission vehicle; ULEV=ultra low emission vehicle; SULEV=Super ultra low emission vehicle

(Fig. 17.15.3). Additional cold start strategies with operating conditions away from $\lambda=1$ are applied to accelerate the heating of the catalyst e.g. a rich gas mixture, ($\lambda \sim 0.7$) together with additional air injection by a secondary air pump upstream of the catalyst, or a lean mixture ($\lambda \sim 1.05$) together with retarded ignition to

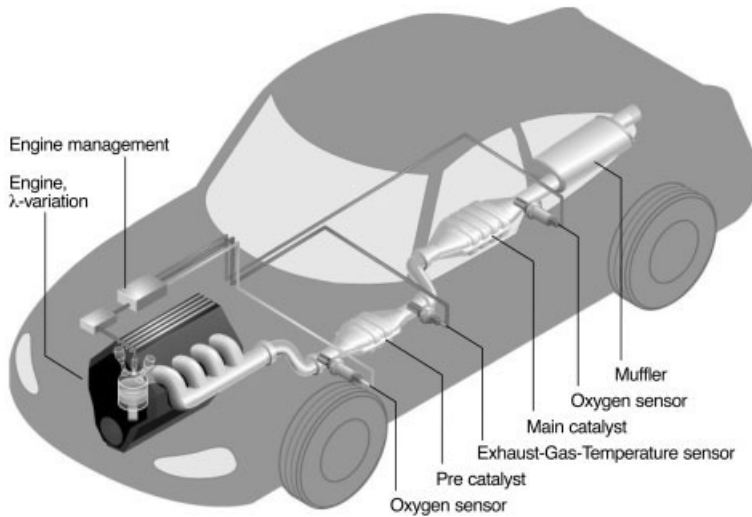


Fig. 7.15.3 Exhaust-gas system

partly combust the mixture in the manifold or the pre-catalyst. During a real driving cycle the air fuel ratio can diverge from $\lambda=1$, for example, at fuel cut-off the exhaust gas approaches pure air and for component overheating protection the cooling effect of operating conditions at $\lambda < 1$ can be necessary for a short time. Previously engines typically used fuel excess during acceleration because of the additional engine power at $\lambda \approx 0.8\text{--}0.9$, for example (Fig. 17.15.1). Recent emission legislation prohibits this A/F enrichment.

Lean-burn engines, such as gasoline direct-injection (GDI) engines [6, 7], offer a fuel consumption reduction of up to 15% and are operated at $\lambda=1.4\text{--}1.6$. Because of the weak NO_x conversion of TWCs under these conditions, special NO_x catalysts have been developed to meet the emission legislation requirements. If NO_x storage catalysts are used, cyclic rich phases down to $\lambda=0.7$ are necessary for catalyst regeneration. However, for better driveability and emission conversion of TWC converters, current lean burn, and GDI systems alter typically between lean and $\lambda=1$ operation. Therefore, an exhaust-gas after-treatment system needs both a TWC and a NO_x catalyst. Starting in 2002 oxygen sensors (especially wide-range A/F sensors) are also favorably used to control diesel engines with λ values between $\lambda=1.1$ and 8 [8, 9].

7.15.1.3 Measuring the Air/Fuel Ratio

Although the air/fuel ratio can be directly determined by measuring the fuel mass and intake air mass, the accuracy of this measurement is not sufficient to keep the exhaust gas λ in the very small window for optimum after-treatment with a TWC. Therefore, we have to use an indirect method by measuring the oxygen concentration in the exhaust manifold, which is strongly dependent on λ

(Section 5.6, Fig. 5.6.1) [10, 11]. However, the relationship is only applicable under equilibrium conditions. Unfortunately the rough combustion in a vehicle does not take place according to theory, and equilibrium has to be achieved within the sensor before the measurement. For $\lambda=1$ detection, the equilibrium partial pressure of oxygen can be measured with potentiometric ZrO_2 oxygen sensors or resistive TiO_2 oxygen sensors (Section 5.6.3 and Section 7.15.2), where the complete catalytic reaction occurs at the active surface of the sensor (catalytic electrode or semiconducting layer). In lean exhaust gases with $\lambda > 1$, the λ value is determined by the 'excess oxygen', which can be derived approximately from the H/C relation of the fuel [12]. For the measurement of 'excess oxygen' a different physical principle is advantageous as for limiting-current oxygen sensors, for example. For rich exhaust gases with $\lambda < 1$, the λ value is determined by the sum of the reducing gas components CO, HC, and H_2 and can favorably be measured by universal or wide-range oxygen sensors, which are most suitable for the whole lambda range.

7.15.1.4 Closed-Loop Lambda Control

Closed-loop lambda control was first introduced 1976 together with the ZrO_2 Nernst oxygen sensor and the TWC [13]. Integrated in the software of the engine-management system, the lambda control function regulates the optimum air/fuel ratio by measuring the exhaust-gas lambda value. The delay time between changing the injected fuel mass and the detection of the resulting lambda change determines the dynamic of the closed-loop lambda-control dependent on engine speed (rpm) and load, the installation position of the sensor, and the response time of the oxygen sensor itself.

7.15.1.5 Two Stage Lambda=1 Control

The majority of the present system concepts are based on the two-stage $\lambda=1$ control with a Nernst ZrO_2 oxygen sensor (Fig. 17.15.4). The step-like signal characteristic of the sensor leads to an oscillating lambda value around $\lambda=1$ with frequencies from 0.5 Hz at idle conditions up to 5 Hz at high rpm and a typical amplitude of $\Delta\lambda=0.01-0.05$. This lambda oscillation is attenuated by the TWC. The sensor voltage varying between 0 and 900 mV is compared with a set value of 450 mV, which corresponds to $\lambda=1$. Depending on the comparator output the injection time is step-wise increased or decreased using various algorithms. Due to the fact that all Nernst oxygen sensors show a small lean shift of the characteristic curve, a corrective factor is needed to control a mean value of $\lambda=1$. This is achieved by introducing a delay time, t_v , before reducing the fuel injection time after a signal change from lean to rich. Good pre-control of the air/fuel mixture avoids excessive lambda amplitudes and helps the controller to adjust varying operation conditions quickly.

For almost two decades this control algorithm satisfied the legislation requirements. When the emission standards became higher, an improved control algorithm was necessary, especially to guarantee the lifetime emission quality. The solution was the introduction of an additional trim control.

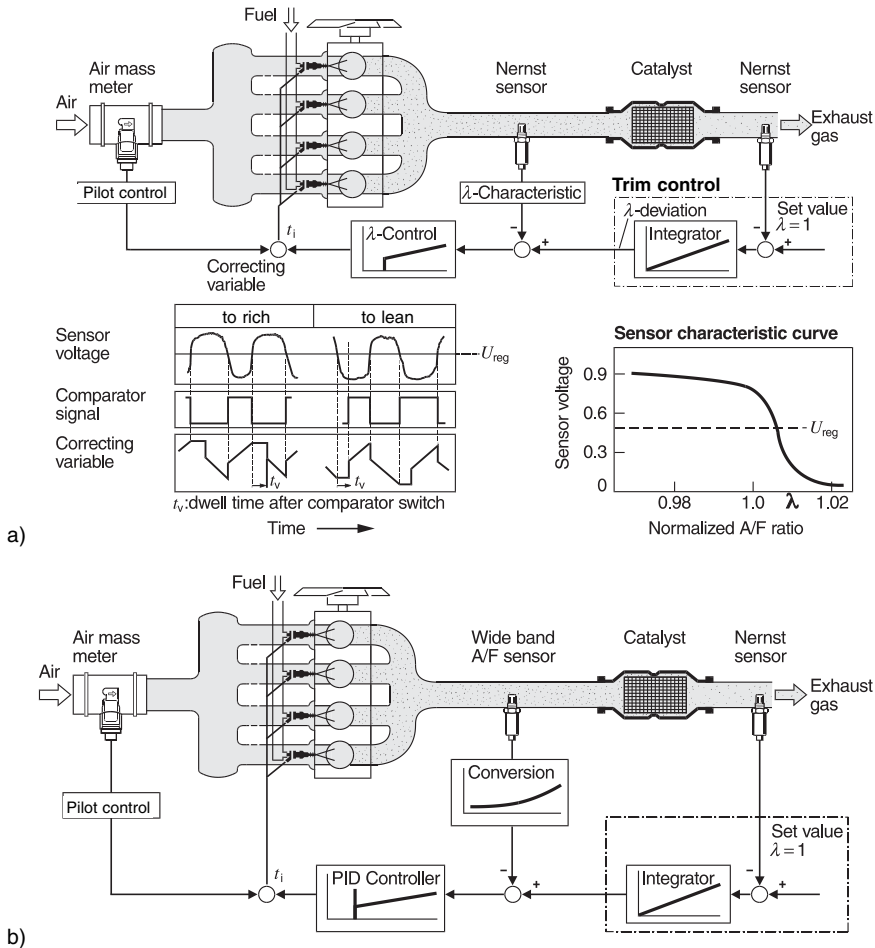


Fig. 7.15.4 a) Two-stage λ control with trim control. b) Linear λ control with PID control

A second sensor downstream of the catalyst, first introduced to meet the OBD diagnostic requirement such as catalyst monitoring (Section 7.15.1.7), is now also used to correct the aging behavior of the front sensor (Fig. 7.15.4). Protected by the catalyst against poisoning and high amounts of unreacted gas mixture, this rear sensor features a very stable signal characteristic over life. All new lambda control systems of recent years are based on this two-sensor concept.

7.15.1.6 Linear Lambda Control

The introduction of ULEV required additional measures to improve lambda control accuracy plus extended malfunction diagnostics [14, 15]. Furthermore, the development of gasoline direct-injection engines and advanced electronic diesel control needed linear lambda control using the whole lambda range of $\lambda = 0.7$ –air.

With the availability of a wide-range A/F sensor with a continuous lambda characteristic, a linear lambda control with an improved PID (Proportional-Integral-Differential) control behavior could be developed (Fig. 17.15.4). For $\lambda=1$ systems the small lambda deviations of the linear lambda control are advantageous, especially with aged catalysts and varying operating conditions, and in addition the wide-range lambda sensor offered several easy solutions for the necessary diagnostics.

7.15.1.7 On-Board Diagnosis

With the introduction of on-board diagnosis (OBD I+II), CARB required monitoring of the correct function of all emission-relevant components [15]. Special hardware and software measures in the ECU have been developed to detect failures of the electrical function of sensors and actuators immediately. Complex diagnostic algorithms use sensor signals to identify functional defects and inadmissible functional deviations of other components, especially when a direct evaluation or measurement is impossible, as for catalyst and oxygen-sensor deterioration. One way to detect malfunction of the catalytic converter and the upstream sensor is to use a second oxygen sensor downstream of the catalyst. The aging of the catalyst, and therefore the decrease of the simultaneous conversion efficiency of HC, CO, and NO_x , correlates with a loss of oxygen storage capability. With a new catalyst, the lambda amplitude of the raw exhaust gas forced by the lambda closed-loop control is suppressed by the catalyst storage leading to an almost constant A/F value downstream of the catalyst. The more the catalyst is aged, the higher is the lambda amplitude monitored by the rear sensor. As a second feature the lambda drift of the upstream sensor can be measured by comparison with the average lambda value of front and rear sensor. In both cases a failure code is set (MIL) if a threshold value is surpassed.

Since oxygen storage is only an indirect measure, the straightforward way to monitor catalyst efficiency is to use selective gas sensors to detect the breakthrough of NO_x and HC. After many years of research work, two physical principles based on ZrO_2 have been shown to be capable of selective gas detection under automotive exhaust-gas conditions: the mixed-potential potentiometric HC sensor and the dual-chamber limiting-current NO_x sensor [1], but only the NO_x sensor is currently on the market. For $\lambda=1$ systems furthermore, oxygen sensors are the most favorable and reliable sensors for monitoring TWC in GDI systems, to some extent NO_x sensors are used to diagnose the NO_x catalyst and control NO_x catalyst regeneration.

Beside linear lambda control, an essential advantage of wide-range A/F sensors is the possibility of managing extended diagnosis in the system. It improves catalyst monitoring due to the direct calculation of the O_2 storage capability. The measurement range up to air enables the diagnosis of the secondary air pump. For GDI systems inadmissible torque, and for diesel systems uncontrolled combustion during fuel cut off, can be detected.

7.15.2

Oxygen Sensors for Stoichiometric Air/Fuel Ratio Control

The equilibrium partial pressure of oxygen p'_{O_2} strongly depends on the A/F ratio and the temperature (Section 5.6, Fig. 5.6.1, p_{O_2} against λ). At the stoichiometric A/F ratio $\lambda=1$, p'_{O_2} against λ shows a steep step of several orders of magnitude that enables accurate $\lambda=1$ detection. In particular, two functional principles of oxygen sensors are obviously suitable for $\lambda=1$ detection: the potentiometric ZrO₂ Nernst sensors, and TiO₂ resistive oxygen sensors [16, 17], because they directly reflect this step function in their characteristic curve.

7.15.2.1 Potentiometric ZrO₂ Nernst Oxygen Sensors

The basic principle of the potentiometric ZrO₂ oxygen concentration cell is explained in Section 5.6.3.1, Fig. 5.6.2. The sensor consists of a ZrO₂ solid electrolyte body coated with two Pt electrodes, a measuring electrode exposed to the exhaust gas with an oxygen partial pressure of p'_{O_2} , and a reference electrode exposed to a reference atmosphere with a constant oxygen partial pressure of p''_{O_2} , typically ambient air.

According to the Nernst equation the internal cell voltage generated, V_S , depends on the ratio of the partial pressures of oxygen p''_{O_2}/p'_{O_2} at the Pt electrodes on both sides of the cell (Section 5.6.3.1, Fig. 5.6.1, and Fig. 5.6.2) and follows the steep slope of p'_{O_2} at $\lambda=1$. The resulting shape of the voltage characteristic being similar to the Greek letter λ gave the λ sensor its name (Fig. 17.15.5). A signal of $V_S=400\text{--}500\text{ mV}$ at the highest slope corresponds to the stoichiometric A/F ratio $\lambda=1$ with excellent accuracy. Even the temperature dependence can be neglected at $\lambda=1$, but affects the characteristic curve above and below $\lambda=1$ significantly. In lean exhaust gases ($\lambda > 1$) the temperature dependence of the Nernst voltage is de-

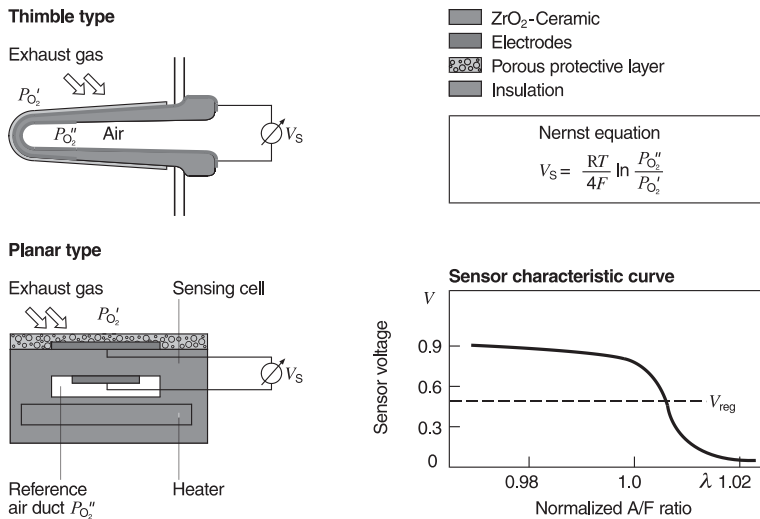


Fig. 7.15.5 Nernst oxygen sensors [19]

terminated by the prefactor of the Nernst equation, in rich exhaust gases ($\lambda < 1$) it is dominated by the gas reaction constants for the equilibrium reactions of the exhaust gas at the outer electrode (Section 5.6.3.1, Fig. 5.6.1).

For an accurate measurement of the equilibrium p'_{O_2} , an electrode with high catalytic activity at the exhaust-gas side is essential to guarantee a completed equilibrium reaction (Section 5.6.3.1, Fig. 5.6.2). Incomplete reactions increase p'_{O_2} leading to a lower voltage at $\lambda < 1$. Beside the chemically induced temperature dependence there is also another electronically induced one. The voltage measured at the sensor terminals also depends on the internal cell resistance (following the Arrhenius equation) because of the limited resistor load of the input circuit. Additionally, at low temperatures that means at high internal resistance in combination with a low-load resistor a polarization phenomenon at the electrode can further falsify the real Nernst voltage. Above 300–350 °C this effect disappears if electrodes with good catalytic activity are used. This is further supported by a porous protection layer on top of the sensor, limiting the gas diffusion to the measuring electrode, which also protects the sensitive electrode against abrasion by oil, ashes, or other particles such as iron oxides from the manifold and against catalyst poisons such as silicon, sulfur, or lead.

The disadvantage of this diffusion barrier is that it slightly changes the gas composition at the electrode because of the different diffusion coefficients of the constituents of the raw exhaust gas. On average the rich gas components, especially the small H_2 , are faster than the bigger O_2 or NO_x molecules causing a lean shift that is typical for all sensors. The closer the exhaust gas is to equilibrium, the less lean shift is observed. Downstream of the catalytic converter the lean shift completely disappears. Furthermore, precatalytic layers in front of the electrode, or catalytic materials inside the protection layer, reduce the effect.

A stable reference oxygen partial pressure p''_{O_2} is the precondition for a correct exhaust-gas measurement. Due to the logarithmic relation of the Nernst equation, the voltage is fortunately not very sensitive to minor changes in p''_{O_2} . Calculations show that even a change of one order of magnitude in p''_{O_2} gives only about 50 mV output difference maintaining a high $\lambda = 1$ accuracy, because of the steep slope of the characteristic curve. However, a reaction of the reference air with oxidizable components penetrating to the reference electrode can almost consume the total oxygen. Therefore, good gas-tightness of the seal between sensor element and housing is necessary to keep the reference air separate from the exhaust gas, and the tightness of the housing itself has to protect the reference air from environmental pollution (water, oil etc.), which would alter the reference oxygen partial pressure (reference air poisoning or characteristic shift down, CSD). The finite load resistor of the input circuit leads to a small continuous oxygen pumping current from the reference-air side to the exhaust-gas side, which also lowers p''_{O_2} after long operation. To ensure sufficient reference oxygen, a regular air exchange under all conditions currently needs to be maintained either through the wire harness or through oxygen-permeable membranes in the housing. The temperature profile during a typical driving cycle with phases where the sensor is heated up and cooled down ensures continuous 'breathing'.

Thimble ZrO₂ Oxygen Sensors

The most commonly used automotive exhaust-gas oxygen sensors in the world are thimble ZrO₂ sensors using ambient air ($p''_{O_2}=0.21$ bar) as reference atmosphere (Fig. 17.15.5).

A separate heating element inside the ceramic sensor element guarantees sufficient ceramic temperatures $>300^\circ\text{C}$ also in cold exhaust gas. The standard heating elements with heater power of 7–18 W allows start-up times of 25–60 s. New special ZrO₂ thimble elements with low thermal mass combined with heating elements providing concentrated high heat energy and good thermal contact to the sensing element have reduced the ‘light-off’ time to less than 10 s.

Various types of protective layers have been developed to improve poisoning resistance and thermal aging behavior against lifetime (Section 5.6.4, Fig. 5.6.7). A very good poisoning resistance could be achieved with a double protection layer. A fine porous cofired first layer protects against lead poisoning, while a more open porous second layer prevents against plugging and glass formation by silicon poisoning.

Planar ZrO₂ Oxygen Sensors

The given disadvantages, such as the ‘light-off’ time and the power consumption of the separate heater element, as well as the functional limitations of the conical thimble shape, initiated the development of a new generation of planar sensor elements based on ceramic multilayer technology (Section 5.5.4.3), which was first introduced for high-density circuit boards in computers [18–20]. This technology offers high variability of extended functionalities in complex three-dimensional designs, including an integrated heater in one compact monolithic structure. The small size, low thermal mass, and good thermal contact result in lower power consumption (5–7 W), faster heating up (10 s), and better temperature control, which are the advantageous properties of planar sensors.

Fig. 17.15.5 shows comparative cross sections of a standard thimble sensor (LSH) and the first mass-produced planar type sensor (LSF). Fig. 17.15.6 explains the three-dimensional layer structure. The functional layers of the planar element are stacked in three sheets (Section 5.6.4.3, Fig. 5.6.8).

- A heater sheet with a Pt heater layer encapsulated within Al₂O₃ insulation layers.
- A reference air duct sheet with a reference air channel to the communicating rear end of the element.
- A sensor sheet with a reference electrode exposed to the reference air duct and the outer measuring electrode covered with a porous protective layer.

The electrical contact pads at the rear end are connected to the inner electrode and heater leads by vias (holes with conducting material inside, used in microhybrids). The signal characteristic is similar to the thimble sensor. Since 1994 the annual mass production volume has reached more than 18 million sensors.

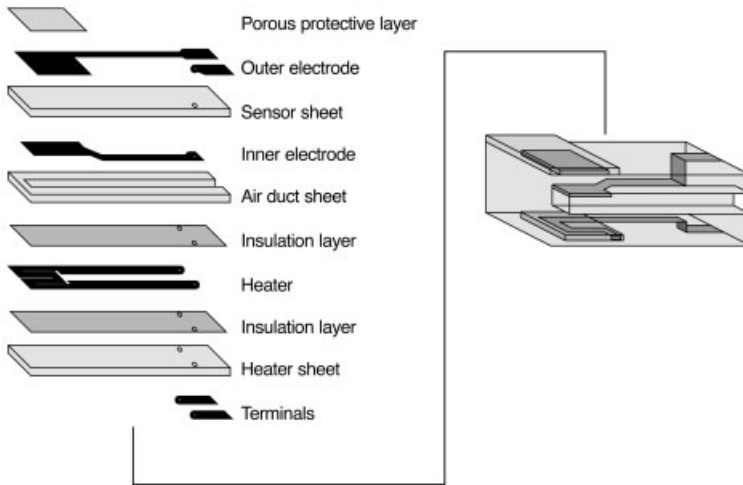


Fig. 7.15.6 Thickfilm technology of planar zirconia Nernst sensor [19]

Potentiometric ZrO_2 Oxygen Sensors with Pumped Reference

To improve the quality of the reference atmosphere and to increase the robustness against reference air poisoning, a superimposed small oxygen ion current of a few μA is used to pump oxygen from the measuring electrode to the reference electrode using a pull-up resistor to a voltage source. In this case the reference air duct can be replaced by a printed porous layer with defined gas diffusion resistance to guarantee the correct reference partial pressure p''_{O_2} . The sensor is typically operated at ceramic temperatures $>600^\circ C$, maintaining a low internal resistance and avoiding any polarization phenomena, which keeps the voltage offset due to the pumping current almost neglectible. As a side effect the elimination of the large air duct improves further the thermal contact between heater and sensor cell and allows a thinner and more symmetrical design with less thermal stress inside.

7.15.2.2 Semiconducting TiO_2 Resistive Oxygen Sensors

Beside the ZrO_2 solid electrolyte potentiometric and amperometric oxygen sensors, only the TiO_2 resistive type has been manufactured in large quantities [21]. The physical basics of the material are explained in Section 5.6.3.2. The sensor utilizes also the planar multilayer technology based on Al_2O_3 substrates (Fig. 17.15.7). A porous semiconducting TiO_2 thick film deposited on two parallel electrodes acts as the sensitive element changing its resistivity according to the oxygen partial pressure in the exhaust gas (Section 5.6.3.2, Fig. 5.6.4 and Fig. 5.6.5). The equilibrium p_{O_2} at the surface is achieved by the catalytic activity of TiO_2 itself supported by Pt or Pt/Rh catalytic impregnation. The voltage (about 1 V) for the resistance measurement is derived from the heater circuit leading to a three-wire design for the complete sensor. In contrast to potentiometric ZrO_2 sensors, resistive sensors do not need reference air. The TiO_2 sensor shows a very

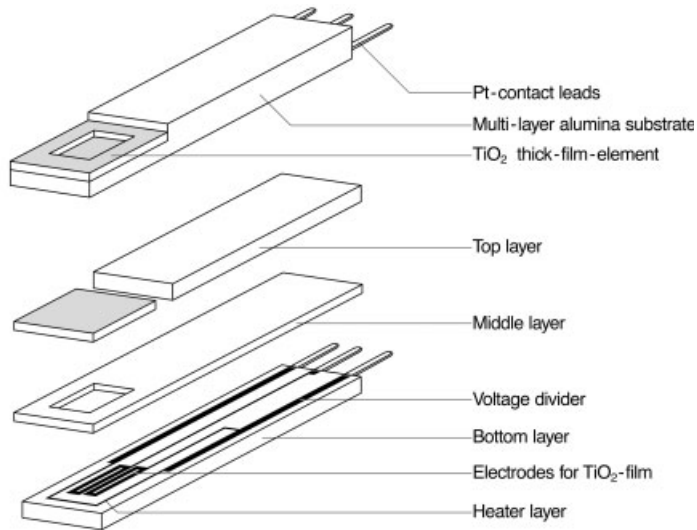


Fig. 7.15.7 TiO₂ resistive oxygen sensor [21]

good sensitivity and good lead-poisoning resistance, but has not reached the high market volume of the various ZrO₂ designs.

7.15.3

Oxygen Sensors for Non-Stoichiometric Lambda Control

The low slope of the characteristic step function at rich ($\lambda > 1$) and lean ($\lambda < 1$) voltages limits the application of switching potentiometric ZrO₂ oxygen sensors to a maximum range of $\lambda = 0.9\text{--}1.3$. Even in this small range, special treatments and measures are needed to compensate for temperature and voltage drop effects to give an accurate measurement [22]. Therefore, amperometric oxygen sensors are preferred for wide-range applications because of the proportionality of the signal current against the oxygen concentration.

7.15.3.1 Amperometric Limiting-Current Oxygen Sensors

The principle of a limiting-current oxygen sensor (Fig. 17.15.8) is based on the correspondence of the ionic oxygen (O^{2-}) current through an electrochemical cell where both electrodes are exposed to the exhaust gas and the molecular oxygen current through a diffusion barrier [23, 24]. If, at elevated temperatures above 600°C, an external pumping voltage, V_p , is applied to the two electrodes of a solid electrolyte ZrO₂ cell, a remarkable oxygen-ion flow is initiated from the cathode to the anode leading to a pumping current, I_p (Fig. 17.15.8). The oxygen-ion current corresponds to the number of oxygen molecules diffusing to the cathode in the exhaust gas. With increasing pumping voltage V_p , the current increases according to the internal cell resistance. If a diffusion barrier, such as a small channel or a

$I_L = \text{const} \cdot D_{O_2} \cdot c_{O_2}$
 D_{O_2} : diffusion coefficient of O_2
 c_{O_2} : concentration of O_2

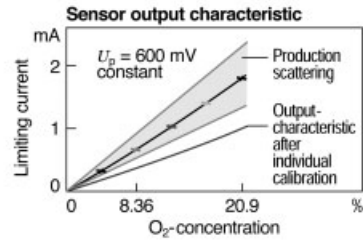
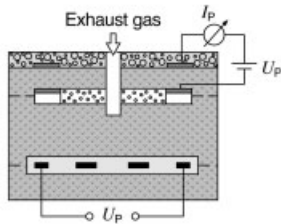
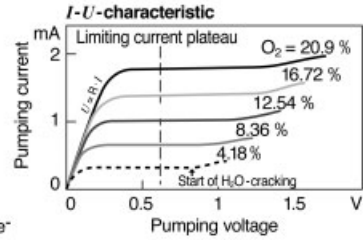
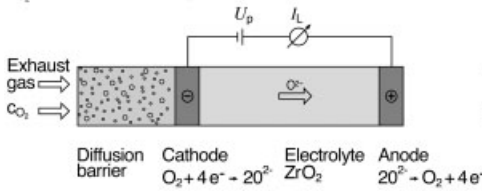


Fig. 7.15.8 Limiting current oxygen sensors

porous ceramic layer, in front of the cathode limits the flow of oxygen molecules to the electrode, the oxygen concentration at the cathode approaches zero, resulting in a current saturation according to the concentration gradient dc_{O_2}/dx above a certain pumping voltage threshold. The resulting limiting current, I_L , is roughly proportional to the exhaust-gas oxygen concentration, c_{O_2} .

$$I_L = 4FD_{O_2} \frac{Q}{L} c_{O_2} \tag{2}$$

where

- F = Faraday's constant
- D_{O_2} = diffusion coefficient
- Q = effective diffusion cross section
- L = effective diffusion length.

If the pumping voltage is further increased, the pumping current starts to increase too by decomposing H_2O and CO_2 in the exhaust gas and later by taking oxygen from the ZrO_2 lattice, which could lead to blackening, a partial damage of the ceramic.

After explaining the effects of the ionic pumping current we need also to consider the diffusion part of the amperometric sensors to fully understand their properties. The mix of the two diffusion mechanisms, gas phase diffusion and Knudsen diffusion, determines the sensitivity against temperature and pressure. Reducing the pore diameter of the diffusion barrier promotes Knudsen diffusion, leading to less temperature dependence but higher pressure dependence, while large pores show the opposite effect [24, 25]. Another functional disadvantage of

the simple pumping cell is that at stoichiometric or rich gas conditions, the pumping starts to increase again due to decomposition effects at the cathode, which limits the application to lean exhaust-gas applications only.

7.15.3.2 Single Cell Wide-Range A/F Ratio Sensors

The effect of an ambiguous pumping current characteristic at $\lambda \leq 1$ can be avoided by exposing the anode to a reference atmosphere, favorably to ambient air (Fig. 17.15.9) like a potentiometric cell generating a superimposed Nernst voltage, V_N , which is compared with an applied bias voltage of typically $V_B = 0.5$ V. The pumping voltage, V_P , is determined by the difference between the bias and the Nernst voltage

$$V_P = V_B - V_N \tag{3}$$

Because the maximum Nernst voltage varies in the range 0–1 V and is 0.5 V at $\lambda = 1$, we can distinguish three cases

- at $\lambda = 1$, $V_P = 0$ and subsequently $I_P = 0$
- at $\lambda > 1$, $0 < V_P < + 0.5$ V
- at $\lambda < 1$, $0 > V_P > -0.5$ V

To guarantee reliable operation the limiting current condition should be maintained, so the internal resistance needs to be kept low to give sufficient effective pumping voltage, V_{Peff}

$$|V_{Peff}| = |V_P| - |V_R| \tag{4}$$

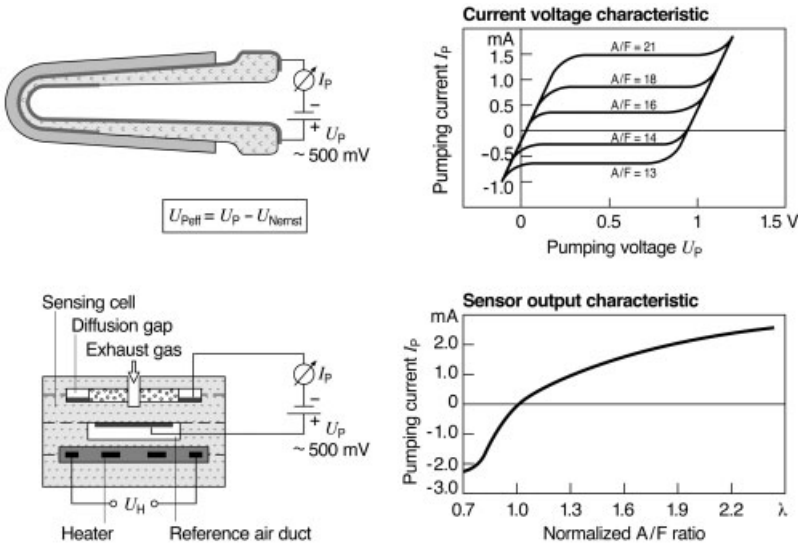


Fig. 7.15.9 Single-cell wide-range A/F sensors [34]

This requires high sensor temperatures, low electrode resistances and aging effects, and good electrode-poisoning protection. The pumping current in rich exhaust gas is limited by the cathodic limiting current and the oxygen supply to the reference air electrode by the reference air duct [26]. These effects typically cause restrictions to the sensitivity and λ range of this sensor compared with dual-cell types. This sensor is mass produced in a thimble shape and a planar design [26, 27].

7.15.3.3 Dual-Cell Wide-Band A/F Ratio Sensors

For applications with A/F ratio ranges from $\lambda=0.65$ –air, dual-cell wide-range A/F sensors are used [19, 20, 28–30]. This sensor is made using a combination of an amperometric and a potentiometric ZrO_2 sensor (Fig. 17.15.10). The inner pumping electrode and the measuring electrode of the Nernst cell are placed opposite each other in a diffusion gap, which is separated from the exhaust gas by a diffusion barrier. An electronic controller circuit regulates the pumping current by comparing the Nernst voltage with a set value of 450 mV. In lean exhaust gas with $V_N < 450$ mV, the pumping voltage is positive; in rich exhaust gas with $V_N \geq 450$ mV, the pumping voltage is negative. The resulting pumping current I_p maintains the limiting current conditions as described above resulting in a linear characteristic with O_2 across the whole A/F range with positive currents at $\lambda > 1$ and negative ones at $\lambda < 1$. At $\lambda = 1$ the pumping current is zero.

This sensor type is made in a planar shape only, which enables light-off times of < 20 s necessary for ULEV systems. The fast response time of this sensor, < 30 ms, allows an individual cylinder measurement for cylinder balancing, which

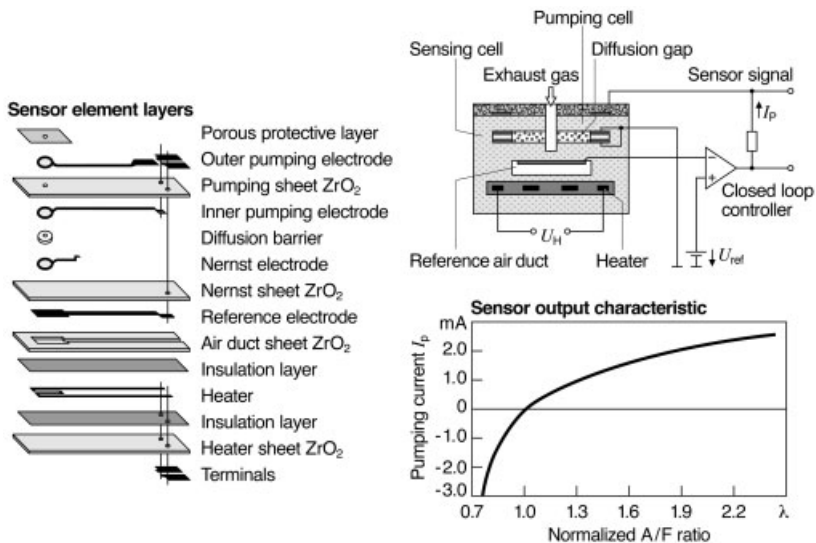


Fig. 7.15.10 Dual-cell wide-band A/F sensor: LSU [34]

helps to meet future emission legislation, such as SULEV, with a higher degree of freedom in the design of the after-treatment systems.

Because the O_2 signal is derived from the interaction of pumping and Nernst cells, this sensor requires a special electronic controller circuit. This ASIC (application-specific integrated circuit) includes the basic pumping current control, internal resistance measurement, and extended hardware diagnostics for the sensor. The calibration of the characteristic curve is done by a current divider with a trimming resistor in the sensor connector, for example. This sensor type has been in mass production since the mid 1990s, with production volumes of millions.

An advanced version of this sensor reaches fast light-off times of <10 s, even at lean and rich A/F ratios. This would allow diagnostic functions during the first 20 s idle phase of the Californian FTP emission test as well as catalyst light-off strategies with closed-loop control. Since 2002 this wide-band lambda sensor has been used in diesel systems offering EU 4 emission performance including advantages in drivability, fuel consumption, and NO_x storage catalyst regeneration by fuel mean-value adaptation, full-load smoke limitation, and monitoring of irregular combustion [8, 9]. Similar to the planar $\lambda=1$ sensor, the dual-cell wide-range A/F sensor can be designed with natural or pumped reference [30].

7.15.4

Assembly and Mechanical Design of Automotive Exhaust-Gas Sensors

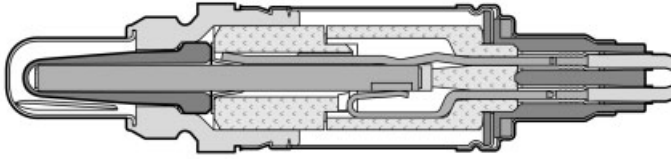
The essential tasks of the sensor housing are to hold the sensing element in place and to protect it against mechanical impacts or aggressive chemicals. Furthermore, it provides the electrical connection and keeps the reference side clean and separated from the exhaust side. Thimble and planar designs naturally differ from each other because of the tubular or rectangular geometry of the ceramic elements and the fact that the planar element typically includes a heater while the thimble heater is a separate part (Fig. 17.15.11).

The thimble element is clamped between housing and ceramic bushings. To provide a potential-free signal the measuring electrode needs to be electrically insulated from the housing by an insulation layer. The planar element is held by a pressed ceramic powder, which also does the insulation.

Another sensitive portion is the electrical connection between sensor element and wires, and the seal of the wire harness because of the high temperatures in the direct vicinity of the manifold. Various technologies and materials can be used to make a good high-temperature contact. The most common method is to pair the platinum-based electrode with stainless steel wires by clamping. Wire insulation and cable outlet seal are made of PTFE withstanding temperatures up to 280°C without deterioration or leakage (to withstand water submersibility if installed under the floor).

At the exhaust side, a tube protects the sensor element against abrasion by rust, oil, ashes, or condensed water droplets carried by the exhaust-gas stream. Newer designs use double protection tubes with shaded holes. Such a labyrinth improves the efficiency of the filter.

Thimble type



Planar type

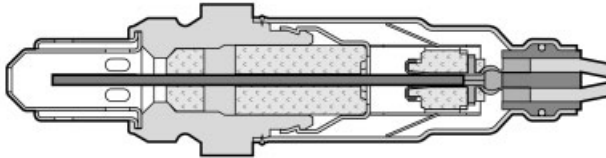


Fig. 7.15.11 Assembly of exhaust-gas oxygen sensors

7.15.5

Durability, Poisoning, and Aging Behavior

7.15.5.1 **Environmental Conditions and Durability Requirements**

Several stress factors affect an exhaust-gas sensor during its lifetime (Fig. 17.15.12, Tab. 7.15.2). Inside the exhaust pipe at gas velocities up to 80 m/s, high temperature, catalyst poisons, particles, and water are assaulting the sensor. From outside thermal, mechanical, and chemical impacts attack the housing, cable outlet, wire harness, and connector depending on the installation position

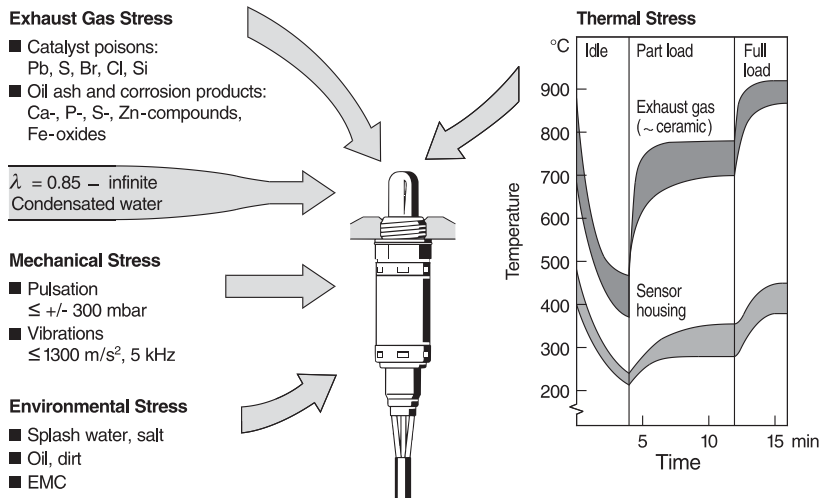


Fig. 7.15.12 Stress factors for exhaust-gas sensors [19]

Tab. 7.15.2 Overview of environmental stress factors for exhaust-gas sensors

<i>Thermal stress</i>	<i>Mechanical stress</i>	<i>Stress caused by environmental conditions</i>	<i>Stress caused by exposure to exhaust gas</i>
Exhaust gas temperature (150–1000 °C)	Vibrational stress (<1300 m/s ²) by engine vibration	Splash water, gush of water (in case containing NaCl, CaCl ₂ , NgCl ₂)	Exhaust gas composition between lambda=0.65 and lambda infinite (fuel cut off)
Exhaust gas temperature gradients at cold start (<500 K/s) at overrun fuel cutoff (<200 K/s)	(<5 kHz) by pulsation of the exhaust gas (<+300 mbar... >+4.5 bar at turbo) by ambient wind (cable, <10 Hz)	Dust (organic, inorganic origin)	Catalyst poisoning from fuel additives: permissible lead content: <0.005 g/l sulfur content: <0.1% Br-, Cl-compounds
Temperature increase after engine stalling (<300 °C at cable outlet)	Stone impact (<1.5 Nm)	Oil, dirt undercarriage protection	Oil ash (Ca-, P-, S-, Zn-compounds up to 1 kg within ≈ 60 000 miles)
Radiant heat (may require corresponding protective measures)	Installation/removal (torque about 50 Nm when assembled)	Fluctuation of the electrical supply system (10.5–16 V)	Miscellaneous corrosion products (e.g., Fe-oxides)
	Wire pull force (>70 N)	EMC (<200 V/m)	Condensation water
	Handling (shock load up to 1000 g)	Variation of climatic conditions (–40 to 50 °C, 10–100% rel. humidity)	

and exposition of the sensor. The current legislation for SULEV and PZEV (partial zero emission vehicles, including exhaust emissions and fuel evaporation) demands a lifetime over 150 000 miles and 15 years. This requires continuous engine testing and end-of-life durability tests of more than one year.

7.15.5.2 Factors Affecting Sensor Function

The interaction of the exhaust-gas composition with the catalytic activity of the electrode and the diffusion of the gas through the porous layers is the most important factor affecting the static and dynamic signal characteristics of the sensors.

Potentiometric ZrO₂ Oxygen Sensors

Poisoning and thermal aging are the main reason why the lambda characteristic and dynamics of the sensor changes with lifetime [1, 31–33]. Plugging of the porous electrode protective layer by oil, ash, or silicon oxide favors the diffusion of hydrogen to the electrode, which leads to a lean shift in the static characteristic curve [31].

Cracks and flake-off of the protective layer lead to a rich shift in the static characteristic curve. Deactivation of the electrode, especially due to lead poisoning, results in a flatter characteristic curve, also shifting the threshold value of 450 mV

towards rich. Reference air poisoning, which shifts the whole characteristic curve to lower voltages, has a similar effect.

Deterioration of the sensor dynamics changes the control state, especially if the response times lean/rich and rich/lean are asymmetrically increased. Adsorption of CO and HC and plugging of the protective layer will slow down the rich/lean response, while electrode poisoning (lead poisoning), which leads to reduced catalytic activity, does the same with the lean/rich response. Symmetrically delayed response times have no effect on the control state, but on the emissions by increasing the control amplitude, which diminishes the conversion rate of the catalyst [1].

Wide-Range ZrO_2 A/F Sensors

The $\lambda=1$ accuracy of single-cell wide-range A/F sensors and the Nernst cell in dual-cell wide-range A/F sensors are affected by the same influences, with the difference that the measuring electrode is protected by the diffusion barrier, which normally offers better protection.

For the rich or lean λ characteristic, plugging of the diffusion barrier leads to a decreasing pumping current, which rotates the characteristic curve around $\lambda=1$. In contrast to this, cracks and delamination within the diffusion barrier layers lead to an increased pumping current causing an opposite rotation of the characteristic curve.

Poisoning, plugging, or flake-off of the outer pumping electrode cause a higher electrode resistance resulting in a polarization of the pumping cell. This destroys the linearity of the oxygen dependence especially at higher pumping currents. Changes of the electrode activity and internal coupling effects of the electrochemical cells influence the dynamic behavior of the sensor.

7.15.5.3 Environmental Influences

Excessive vibrations and stone impact can cause damage to mechanical parts. Thermomechanical overloading mostly lowers the spring load and the sealing at the various joints, causing after-effects such as corrosion of contacts or functional deviations. Chemical influences lead to material variations and corrosion effects.

7.15.5.4 Durability Testing

To ensure the required specification over the whole lifetime, various durability tests are necessary. Engine endurance runs over more than 3000 h, engine poisoning, vibration testing, temperature cycling, and storage and environmental tests are typical proof tests. The extensive durability testing required, is now the limiting factor in the development time of exhaust-gas sensors.

7.15.6

Application of Oxygen Sensors for Exhaust Emission Control

The installation position of the sensor is one of the most important parameters to guarantee that the sensor measures the correct exhaust-gas composition. Calibration engineers talk about a 'stringy' exhaust gas, which means that the gas

streams of the different cylinders are not randomly mixed through the cross section of the manifold. This can falsify the lambda value significantly. Extreme pressure pulsations and turbulences can also influence the measurement result. Beside the measurement function, the protection of the sensor against prohibited operating conditions (e.g., heating up strategy and exhaust water protection) are essential for reliability.

With the growing OBD requirements the diagnosis strategies are getting more important. The complex interaction of the diagnosis functions and the sometimes strong dependence on the operating conditions require expensive and time-consuming application procedures. Because of its high functionality the wide-range A/F sensor is increasingly used for diverse diagnosis, which makes robust applications more difficult.

7.15.7

Future Developments and Trends

Future trends will be determined by improved performance and durability, extended functionalities, new design concepts, and sensor principles [34]. Sensors with light-off times of < 10 s offer emission advantages for future legislation. Individual-cylinder balancing increases the degree of freedom for new exhaust after-treatment concepts and will become standard. The introduction of the lambda control in diesel systems secure their future by competing with gasoline concepts regarding emissions. Installation conditions in closed-coupled catalyst systems, especially between pre- and main catalyst, favor the development of miniaturized sensor designs with higher admissible temperatures. The introduction of 42 V power nets at the end of this decade will challenge the future development of materials and technologies. The increasing complexity and interaction of sensor, electronic control unit, and system functions require the development of complete component packages comprising sensor, ASIC, and software. System know-how becomes important to offer optimized performance. New system concepts include up to eight exhaust-gas sensors in a vehicle to meet SULEV requirements at higher diagnostic standards. Therefore, the automotive exhaust-gas sensor market remains a growing market, by about 30% in this decade, to over 100 million sensors per year.

7.15.8

References

- 1 H.-M. WIEDENMANN, G. HÖTZEL, H. NEUMANN, J. RIEGEL, F. STANGLMEIER, H. WEYL, *Exhaust Gas Sensors in Automotive Electronics Handbook*, 2nd edn. R. K. JURGEN (ed.) McGraw Hill 1999.
- 2 R. ZECHNALL, G. BAUMANN, H. EISELE, SAE Technical Paper 730566.
- 3 K. KOLLMANN, J. ABTHOFF, W. ZAHN, SAE Technical Paper 940469.
- 4 K. ANTONIUS, A. GARNER, S. GARRETT, Honda Press Release, Jan 1995.
- 5 Bosch, *Gasoline-Engine Management*, 1st edn., SAE Society of Automotive Engineers and Robert Bosch GmbH, Stuttgart, 1999.

- 6 K. WINKLER, M. KÜSELL, E. SCHNAIBEL, W. STREHLAU, U. GÖBEL, J. HOHNE, W. MÜLLER, in *Proc. FISITA World Automotive Congress*, 27 Sept – 1 Oct, Paris, 1998, F98T218.
- 7 M. KÜSELL, W. MOSER, M. PHILIPP, SAE Technical Paper 99011284.
- 8 M. BIRK, G. ENGEL, R. LEONHARD, in *Proc. Symp. Steuerungssysteme für den Antriebsstrang von Kraftfahrzeugen*, Berlin 1997.
- 9 W. GERWING, W. BOEHNER, in *Proc. 4 Int. Stuttgarter Motorensymposium*, 20–22, Stuttgart, Feb 2001.
- 10 R. S. SPINDT, SAE Technical Paper 650507.
- 11 J. BRETTSCHEIDER, *Bosch Technische Berichte* 1979, 6 (4), 177–186.
- 12 R. PISCHINGER, G. KRASSNIG, G. TAUCAR, TH. SAMS, *Thermodynamik der Verbrennungskraftmaschine*, Springer, New York, 1989.
- 13 G. T. ENGH, S. WALLMAN, SAE Technical Paper 770295.
- 14 G. PLAPP, O. GLÖCKLER, E. SCHNAIBEL, in *Proc. 3 Aachener Kolloquium Fahrzeug und Motorentechnik*, 15–17 Okt, Aachen, 1991.
- 15 O. GLÖCKLER, M. MEZGER, *VDI-Fortschritt-Berichte*, VDI, Düsseldorf, 1994, 205, 1–17.
- 16 H. U. GRUBER, H. M. WIEDENMANN, SAE Technical Paper 800017.
- 17 W. J. KAISER, E. M. LOGOTHETIS, SAE Technical Paper 830167.
- 18 T. OGASAWARA, H. KURACHI, SAE Technical Paper 880557.
- 19 J. RIEGEL, G. HÖTZEL, H. NEUMANN, in *Proc. 44th Mtg International Society of Electrochemistry*, Berlin, 5–10 Sept 1993.
- 20 H. NEUMANN, G. HÖTZEL, G. LINDEMANN, SAE Technical Paper 970459.
- 21 A. TAKAMI, *Ceram. Bull.* 1988, 67 (12), 1956–1960.
- 22 H.-M. WIEDENMANN, L. RAFF, R. NOACK, SAE Technical Paper 840141.
- 23 H.-M. WIEDENMANN, *VDI-Berichte*, VDI, Düsseldorf, 1985, 578, 129–151.
- 24 H. DIETZ, *Solid State Ionics* 1982, 6, 175–183.
- 25 K. SAJI, *J. Electrochem. Soc.: Electrochem. Sci. Technol.* 1987, 134 (10), 2430–2435.
- 26 K. MIZUSAWA, K. KATOH, S. HAMAGUCHI, H. HAYASHI, S. HOCHO, SAE Technical Paper 970843.
- 27 M. NAKAE, T. TSURUTA, R. MORI, S. INAGAKI, SAE Technical Paper 2002-01-0474.
- 28 S. SOJIMA, S. MASE, SAE Technical Paper 850378.
- 29 T. TAKEUCHI, *Sensors and Actuators* 1988, 14, 109–124.
- 30 T. YAMADA, N. HAYAKAWA, Y. KAMI, T. KAWAI, SAE Technical Paper 920234.
- 31 B. W. HOLLEBOOM, S. W. HAWES, E. L. KER, SAE Technical Paper 860478.
- 32 P. S. BRETT, A. L. NEVILLE, W. H. PRESTON, J. WILLIAMSON, SAE Technical Paper 890490.
- 33 F. UEDA, S. SUGIYAMA, K. ARIMURA, S. HAMAGUCHE, K. AKIYAMA, SAE Technical Paper 940746.
- 34 J. RIEGEL, H. NEUMANN, H.-M. WIEDENMANN, *Solid State Ionics*, 152–153 (2002) 783–800.

7.16

Chemical Sensors for Emission Control

DIETMAR SCHMITT

7.16.1

Introduction

A three-way catalyst (TWC) is used for the conversion of exhaust emissions from gasoline-fuelled vehicles. The TWC converts toxic emissions (like CO, NO_x, hydrocarbons) into their non-toxic components: N₂, O₂, H₂O, and CO₂. The TWC gives the best conversion at the stoichiometric air/fuel (A/F) ratio (=14.6/1), which is

called the lambda, or $\lambda=1$ value. To control the combustion at the $\lambda=1$ point, two main types of oxygen sensor are used. Upstream of the TWC a wide-range O_2 sensor is used and downstream of the TWC there is a binary λ sensor. The wide-range O_2 sensor controls the engine combustion to the optimal A/F ratio. The binary λ sensor monitors the function of the catalyst (in accordance with OBD II standards=on-board diagnosis).

For diesel engines, the use of an oxidation catalyst does not require a sensor. Such an oxidation catalyst mainly converts CO and hydrocarbons (HC). The limits of NO_x are still rather high and the NO_x emission requirements can be fulfilled without any active sensor control. However, when the new EURO IV (2005) or EURO V (2008 for heavy duty trucks) legislation is implemented, the limits of NO_x will get more and more stringent and require new-sensor controlled after-treatment systems.

The most feasible catalyst system for the future is the NO_x storage catalyst. NO_x adsorbs on this catalyst during lean operation of the engine. To regenerate the catalyst, the engine is switched for a short period to rich operation. Then the adsorbed NO_x is decomposed to N_2 and H_2O . In this after-treatment system a NO_x sensor is installed after the NO_x storage catalyst. By detecting the NO_x concentration and the λ value, the NO_x sensor can optimize and monitor the exhaust after-treatment operation. Deterioration of the NO_x storage catalyst by poisoning or thermal aging can also be detected. Fig. 7.16.1 shows the typical arrangement of such an exhaust system.

There are two main driving forces for the development of NO_x sensors.

- The more and more stringent legislation on exhaust gas emissions from combustion engines (especially NO_x emissions).
- The increasing market share of direct injection engines and diesel engines, for the reduction of fuel consumption, which need new technologies in the exhaust after-treatment system.

Therefore three main applications can be emphasized.

- More accurate control of the exhaust system by on-board measurement [1].
- Permanent self-monitoring and control by a modeling system [2].
- Diagnosis of the exhaust system by logical self-checks.

These applications require a reliable sensor signal during the whole lifetime of the vehicle. So the sensing element, the assembly of the sensing element, the mechanical fixing of the sensor at the exhaust pipe, the cable, and the control

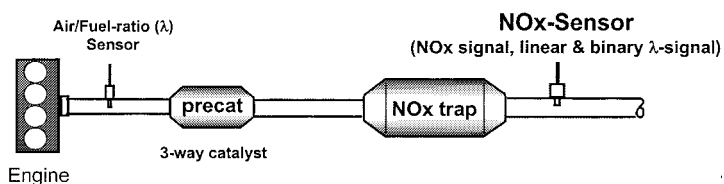


Fig. 7.16.1 Arrangement of exhaust after-treatment system using NO_x sensor

electronics must all be stable against ambient environmental stress and exhaust gas conditions. For example, the sensing element is exposed to high temperatures (up to 950°C), rapid temperature change of the exhaust gas, hard vibration conditions, an aggressive chemical environment, and hard mechanical impacts.

The NO_x sensor has been developed to detect directly the NO_x emissions in the exhaust gas of automotive engines. The basic design is derived from the thick-film based universal exhaust gas oxygen (UEGO) sensor.

7.16.2

Structure and Theory

7.16.2.1 Theory of λ Measurement

An oxygen pumping cell operates by active pumping of O²⁻ ions from the negative to the positive electrode when an electrical charge is applied. To adjust a fixed O²⁻ concentration in a cavity, a certain electrical charge has to be applied (EMF). This electromotive force (E) is described by the Nernst equation.

$$E = \frac{RT}{4F} \ln \frac{p_{O_2, \text{air}}}{p_{O_2, \text{exh}}} \quad (1)$$

where

E = electromotive force (V)

T = absolute temperature (K)

R = gas constant (JK⁻¹mol⁻¹)

$p_{O_2, \text{air}}$ = O₂ partial pressure in air (bar)

4 = moving electrons by one O₂ molecule

$p_{O_2, \text{exh}}$ = O₂ partial pressure in exhaust gas (bar)

F = Faraday constant (Cmol⁻¹)

For example, if a certain oxygen concentration in an internal cavity (1st or 2nd cavity in Fig. 7.16.2) is adjusted, O₂ molecules have to be pumped.

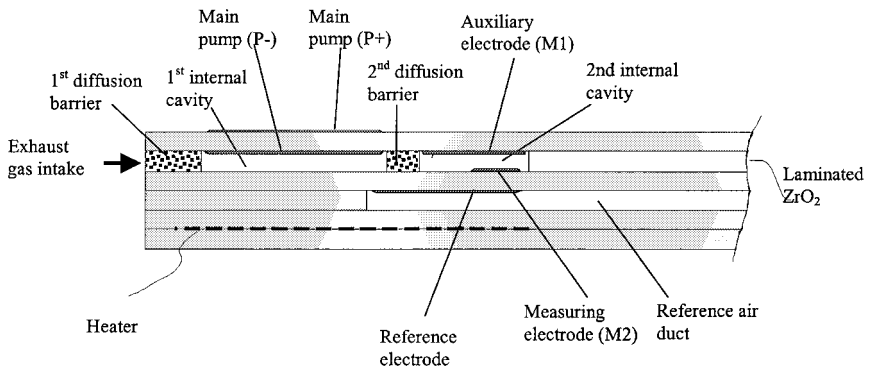


Fig. 7.16.2 Cross section of the NO_x sensor element

- If rich gas diffuses into the cavity, O_2 must be pumped in.
- If lean gas diffuses into the cavity, O_2 must be pumped out.

The resulting pumping current I_{p0} is directly proportional the oxygen concentration of the exhaust gas.

$$I_{p0} = \frac{4FDA}{RTL} (p_{O_2,ex} - p_{O_2,in}) \quad (2)$$

- I_{p0} = pumping current (1st cavity)
 A = surface area of diffusion path (m^2)
 L = length of diffusion path
 $p_{O_2,ex}$ = O_2 partial pressure in the exhaust gas
 D = diffusion coefficient (O_2)
 $p_{O_2,in}$ = oxygen partial pressure in the 1st cavity

Such active O^{2-} pumping through yttrium-doped zirconium oxide is used to measure both O_2 concentration (I_{p0} , Fig. 7.16.3) and NO_x concentration (I_{p2} , Fig. 7.16.3) in the exhaust gas.

7.16.2.2 Arrangement of NO_x Sensor Element

Fig. 7.16.2 shows a cross section of the sensing element. The sensing element mainly consists of six layers of yttrium-doped zirconium oxide, which has a high capability against thermal stress and promotes O^{2-} conductivity. The basis for such O^{2-} conductivity is the temperature of more than $600^\circ C$. The pumping electrodes mainly consist of platinum.

The sensor has a resistance-controlled heater to establish isothermal operation of the sensor, which is important for well-controlled O^{2-} pumping and reliable operation over the lifetime of the sensor. The sensing element is typically controlled at constant temperature of $800^\circ C$ by monitoring the heater resistance.

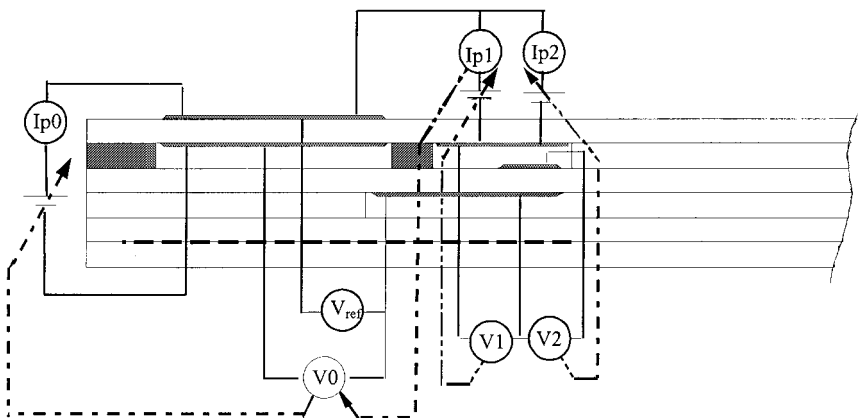


Fig. 7.16.3 Cross section of the NO_x sensor element and control diagram

7.16.2.3 Operation of the NO_x Sensor

The NO_x sensor provides three different signals in parallel.

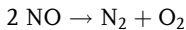
- Binary λ signal (V_{ref}).
- Wide-range λ signal (I_{p0})
- NO_x signal (I_{p2}).

Binary λ Signal

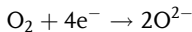
The binary λ signal is obtained from measurement of the EMF between the reference electrode and the main pump electrode (P+). The signal is shown in Fig. 7.16.3 as V_{ref} . By recording the binary signal, rich or lean exhaust gas can be distinguished. The potential application would be detecting the regeneration stop of the NO_x trap by detecting breakthrough of rich exhaust gas. Another potential application is the trimming control of the engine under conventional $\lambda=1$ running conditions.

Wide-Range λ Signal and NO_x Signal

To understand the operation principle of the NO_x sensor it is necessary to understand the measurement of NO_x components. The NO_x gas is detected and measured by the measuring electrode M₂ (Fig. 7.16.2). The measuring principle is an indirect method: NO_x adsorbs at the M₂ electrode and decomposes to N₂ and O₂.



In the next step O₂ is reduced at the surface of the M₂ electrode and the O²⁻ ions are pumped to the P+ electrode.



This pumping current is I_{p2} and is directly proportional to the NO_x concentration in the exhaust gas. This is achieved by controlling V_2 at a constant level and adjusting I_{p2} (Fig. 7.16.3).

Based on this principle there are two important requirements.

- To guarantee a high accuracy of NO_x measurement (less than ± 10 ppm) under high dynamic driving conditions, very precise control of the oxygen concentration in the 2nd cavity must be achieved.
- The second requirement is the adjustment of a low O₂ concentration in the second cavity to keep the offset of the NO_x signal small.

These conditions are fulfilled with the I_{p1} constant control method, achieved by adjusting V_0 [3]. To understand this control mechanism, consider an example.

A car drives on the road at full speed, so the exhaust gas has a rich composition. Suddenly the driver removes his foot from the accelerator; the amount of oxygen in the combustion chamber immediately jumps up, and the amount of fuel injected decreases. The exhaust gas changes from rich to lean conditions, so

a lot of oxygen diffuses into the second cavity of the NO_x sensor. The incoming O_2 causes V_1 to change from the original setpoint to a more lean value (e.g., 200 mV). To adjust the O_2 concentration back to the original setpoint, the oxygen pump cell in the first cavity starts to pump oxygen out. This feedback control loop establishes precise control of the O_2 concentration in the second cavity.

If there is a sudden A/F ratio change in the exhaust gas, it might be supposed that it takes a long time for the A/F ratio signal to stabilize. But this is not the case. From measurements with a conventional engine, on changing from rich to stoichiometric conditions, it can be shown that the I_{p0} signal has a response time ($t_{33-66\%}$) of about 300 ms. This is equivalent performance to a commercial UEGO sensor; $t_{33-66\%}$ is the time taken for the signal to change from the original value to the final value in the 33% to 66% range.

The I_{p2} (NO_x concentration) has a response time ($t_{33-66\%}$) of about 400 ms. The difference between the I_{p0} and I_{p2} response time is due to the time taken for diffusion of the exhaust gas from the first cavity into the second cavity (Fig. 7.16.2).

7.16.3

Accuracy and Cross Sensitivity of NO_x Sensor

In order to have a reliable signal during different driving conditions, the NO_x sensor must achieve good linearity; Fig. 7.16.4 shows measurements with a conventional diesel engine dynamometer. The engine speed was varied from 1000 to 4000 ppm, the engine torque was varied from 50 to 200 Nm. Compared with a

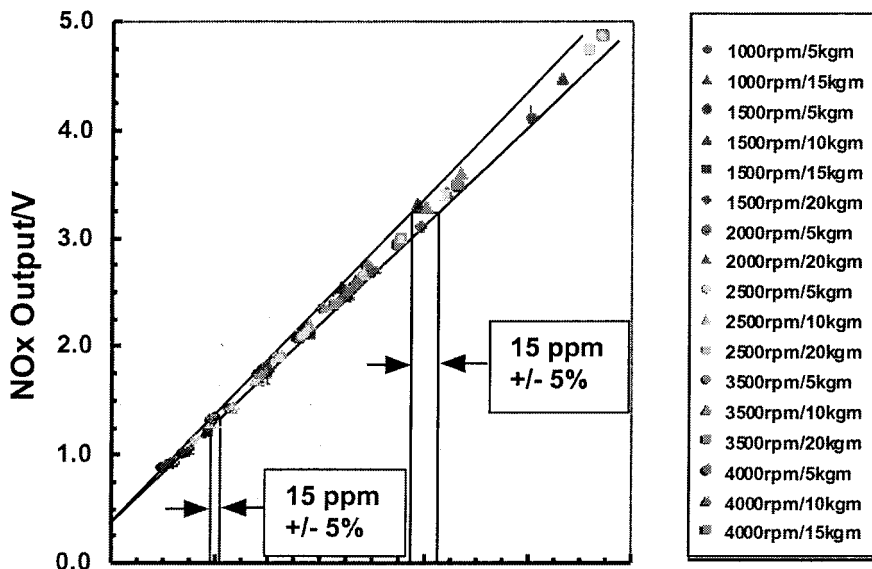


Fig. 7.16.4 Linearity of NO_x output signal at 17 different operating conditions of a diesel engine

CLD (chemiluminescence detector) it is shown that an accuracy of 5% can be ensured under 17 different engine driving conditions. The accuracy depends on the exhaust-pipe system and the engine, but is typical for a commercial car.

This test demonstrates the low cross sensitivity to gaseous components such as CO, O₂, HC, SO₂, and H₂O, and the small influence from exhaust gas overpressure (up to 300 mbar). The experiment varied gas temperature in the range 190–790 °C as well as gas composition.

To understand the influence of each gas component, further tests were carried out with the results shown in Fig. 7.16.5. The gas flow rate was 5 L/min and the gas temperature was kept constant at 300 °C. As the base gas N₂, 3% H₂O, and 500 ppm NO was used, with additions of the gas being tested. Fig. 7.16.5a shows the dependency on O₂ in the range 0–18%; this represents lean conditions typical of a diesel exhaust. The NO_x output signal increases slightly (+0.3%) with increasing oxygen concentration. The dependency on rich exhaust gas components such as CO and hydrocarbons (propane, C₃H₈) is shown in Fig. 7.16.5b and Fig. 7.16.5c; the biggest cross sensitivity is only –0.9%.

Finally, the cross sensitivity to SO₂ was measured. High SO₂ concentrations do not occur under normal driving conditions with fuels available in Western Europe, North America, or Japan, but if we consider the fuels available in emerging markets or the exhaust composition during desulfuration of a NO_x storage catalyst, SO₂ concentrations of more than 500 ppm can occur. The results in Fig. 7.16.5d show a small cross sensitivity to SO₂.

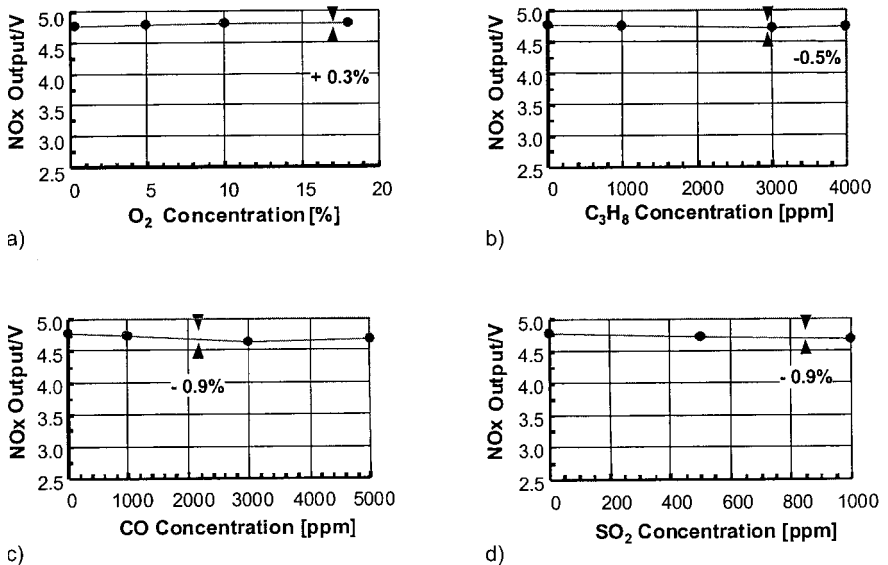


Fig. 7.16.5 Cross sensitivity of NO_x signal to: a) O₂, b) C₃H₈, c) CO, d) SO₂

7.16.4

Application of NO_x Sensors

A typical application of the NO_x sensor is shown in Fig. 7.16.1. This exhaust after-treatment system is an application in direct injection engines. During the lean running operation of the engine the NO_x emissions are adsorbed on the NO_x trap catalyst. The loading level is monitored by the NO_x sensor. If the NO_x loading level increases, less NO_x adsorbs at the catalyst, which means the passing NO_x increases. If a certain threshold of detected NO_x emissions are exceeded, the regeneration phase starts. The engine combustion conditions change from lean to rich and the NO_x trap catalyst is regenerated.

In Fig. 7.16.6 typical loading (lean) and regeneration (rich) phases of a NO_x storage catalyst are monitored. During the loading phase the engine runs at $\lambda=1.9$ (first lean phase). The generated NO_x is stored by adsorption in the NO_x trap catalyst. During this phase the NO_x sensor measures only a negligible amount of NO_x gas. After a certain time the NO_x trap gets more and more saturated and more and more NO_x passes the catalyst. If the NO_x concentration exceeds the threshold of 80 ppm, the regeneration starts. (The value of 80 ppm is a typical value, the threshold is not fixed and depends on the after-treatment system, the engine, and the legislation limit.) The engine combustion changes from $\lambda=1.9$ to $\lambda=0.8$ (rich phase). The adsorbed NO_x is converted by rich gas components (for example H₂, HC, or CO) to H₂O, N₂, and CO₂. The jump in the NO_x signal is caused by desorption of NO_x and ammonia. Ammonia is generated by the catalyst during the rich phase. At the end of the rich phase there is a small spike in the λ signal.

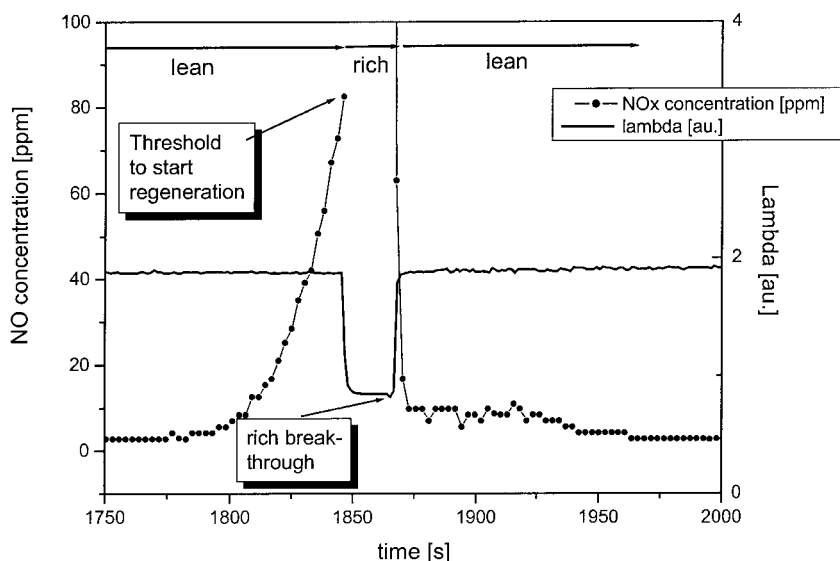


Fig. 7.16.6 Control of NO_x storage catalyst with NO_x sensor



Fig. 7.16.7 The 'smart NO_x sensor' made by NGK Insulators (left hand side); electronic control unit made by SiemensVDO Automotive (right hand side)

This spike is caused by the breakthrough of the rich gas components. Then the fuel combustion changes from a rich A/F ratio back to a lean A/F ratio, and loading of the NO_x storage catalyst starts again.

Fig. 7.16.7 shows the 'smart NO_x sensor' jointly developed by NGK Insulators (sensor) and SiemensVDO Automotive (control unit). The sensor element is packed in a metal housing. A 50 cm cable connects the sensor with the electronic control unit, which controls the sensing element and reads out the output data: NO_x signal, wide-range λ signal, and the binary λ signal. These signals are transmitted to the engine ECU through the CAN bus. The advantage of the smart NO_x sensor is that each sensor is calibrated separately and can be operated independently of the application system itself. It means there is no special optimization to the engine management system required.

Beside the three NO_x sensor signals, the OBD signal is transmitted to the ECU (Fig. 7.16.8). The OBD function can be understood as a self-check of the sensor and can detect short circuit and open wire. The communication of the smart NO_x sensor with the engine ECU through the CAN bus is bi-directional. This means the sensor submits the measured signals to the engine ECU, and the engine ECU submits to the sensor ECU the heat-up flag for starting the heater, this is the so called dew point signal. It is important that the exhaust gas system is free of liquid

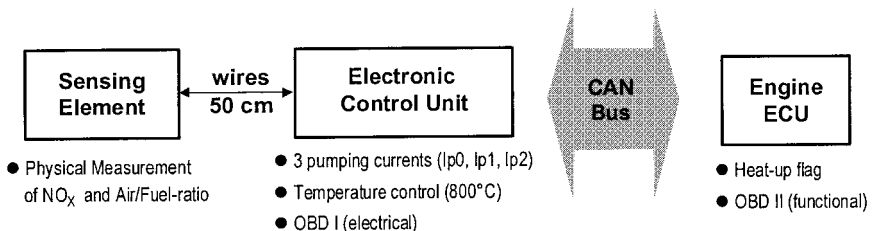


Fig. 7.16.8 Principle of the 'smart NO_x sensor' and communication with ECU

water before heating the sensor, because if there is any water inside the exhaust after-treatment system, there is a high risk of a permanent damage to the sensing element. As a consequence the safety operation of a heated gas sensor (binary λ , linear O_2 , or NO_x sensor) is only possible with an adequate dew point strategy. The engine ECU also communicates an OBD function with the control electronics of the sensor, which monitors the correct operation of the exhaust gas after-treatment system. This function is likely to be mandatory in future legislation.

7.16.5

References

- 1 B. HEIL, C. ENDERLE, G. KARL, P. LAUTENSCHÜTZ, M. MÜRWARD, in *Proc. 23rd Int. Wiener Motorensymposium*, Presentation 2, pp. 25–53, 2002.
- 2 R. WURMS, M. GRIGO, J. BÖHME, H.-D. ERDMANN, J. GESSLER, J. GÖSCHEL, W. HATZ, *Verkehrstechnik/Fahrzeugtechnik*, Fortschritt-Berichte VDI, series 12, no. 490, vol. 1, pp. 23–40, 2002.
- 3 N. KATO, H. KURACHI, Y. HAMADA, SAE Technical Paper 980170, pp. 69–77.

7.17

Chemical Sensors for Air Quality

KLAUS DIETER FRERS

7.17.1

Introduction

A new technology that has found its way into most European up-market cars uses an air-quality sensor (AQS). When detecting high levels of pollution, this sensor instructs the climate control system to close the fresh air inlet door to keep poisonous gases out. The multitude of different climate control systems on the market and the unpredictability of interface changes exposes an AQS subsystem to the problem of providing a flexible yet easy-to-integrate interface. The stand-alone design of such a system with a microcontrolled interface provides an advanced solution.

Today, environmental air-quality is generally improving. Filters are installed in industrial facilities, cars are equipped with catalytic converters, and so forth. However, examination of the air we breathe in different places throughout the day shows that the air we breathe while driving in traffic is one of the worst microenvironments we can possibly encounter in everyday life.

Clouds of pollution are encountered when crossing major intersections, passing through tunnels, and when following vehicles without catalytic converters, such as old buses or cars. Such pollution ‘hot spots’ have high concentrations of CO and/or NO_x , which can be 100–1000 times the ambient levels.

The ventilation system of a car then carries all this pollution into the vehicle cabin and exposes the driver and the passengers to an unhealthy, or even danger-

ous, microenvironment. To maintain good health and well-being it is important that we breathe air of good quality; air that is polluted with harmful gases can do serious damage to human health over time.

Surprisingly, passengers in a closed vehicle are exposed to greater quantities of polluted air than people outside. The reason is that gases are blown into the passenger cabin with the help of the fan, and then diffuse throughout the cabin. The induction of fresh air in this situation is similar to 'cleansing by dilution' (i.e., the same as trying to water down ink until it has vanished: it takes a very long time!).

7.17.2

History

The first air-quality sensors for cars used a Japanese sensor element in which a sensitive tin-oxide material was layered onto a ceramic carrier. This sensor matches very well the requirements for a sensor for controlling air quality in petrol-driven vehicles. In 1986 the first generation was completed; it used an SnO_2 sensor and detected more than 85% of exhaust gases. However, diesel engines also produce nitrogen oxides, which can not be detected by tin-oxide sensors.

This problem was solved by means of phthalocyanin sensors. Phthalocyanin, an organic semiconductor, was deposited as a thin film on a microstructured silicon substrate. This could detect nitrogen oxides, ozone, and sulfur dioxide; it took less than 1 s to detect concentrations below 1 ppm.

The combination of these two sensor types gives two-sensor systems, which detect petrol exhaust fumes (oxidizable gases) and those from diesel engines (reducible gases). In 1989 the second generation as a two-sensor system was launched.

Then the mixed-oxide sensor was developed. This type shows a high sensitivity and selectivity, and is able to work as an in-situ sensor. The paragon company began experimenting with these mixed-oxide sensors in early 1994. The initial design was the same as the tin-oxide sensors already developed, but by finding the right composition of mixed-metal oxide, we created a sensor element that is able to react to both reducible as well as oxidizable gases, thus eliminating the need for two combined sensor elements and handling two signals. The third generation of the sensor system, a metal dioxide sensor, was put on the market in 1994.

The following step was to optimize the signal processing in order to detect, with high reliability, tiny amounts of toxic substances with concentrations lower than 10 ppm. This was done with assistance from the European Union and the German government.

7.17.3

Background

The gases that the passenger of a car most needs to be protected from are CO and NO_x : detection of CO is considered to be the most important.

NO_x is toxic and causes cancer, it reacts with humidity to form nitric acid, which damages mucous membranes and impairs immunity. The thickness of mu-

cous membranes is reduced by frequently exposure to NO_x and their protective function decreases. The threshold for smelling NO_x is 50 ppb, so a human nose is quite a good sensor for this substance, but to sense it the driver has first to take a nose full of it, with a corresponding risk to their health.

But CO does not smell; CO reacts with hemoglobin and prevents it functioning as an oxygen carrier. This process is irreversible, but as the human body continuously produces new blood cells, this process does not lead to permanent damage. Without a sensing system the driver does not realize that their health is in danger. Very extreme concentrations (>100 ppm) can lead to damage or even death, but these concentrations do not occur in a vehicle unless in a closed garage with the engine on.

Additional pollution is caused by interior of the car. Everyone knows the smell of a new car: it comes from all the plastics used for the fittings. More than 1000 substances evaporate from a new car; this is called 'fogging'. It often causes headaches and nausea, and sometimes allergic reactions. Recently, the number of volatile substances has been reduced, so the 'fogging' effect is less severe, but it lasts longer. Another source of bad smells is the fuel tank; carcinogenic benzoles can get inside the car.

If NO_x is detected at 50 ppb, the level of CO is found to be 500–1000 ppb. This means that in a normal traffic situation the fresh air inlet should be closed all the time, which does not make sense at all. To achieve a true health benefit the whole spectrum of toxic gases needs to be considered.

7.17.4

Technology

The microcontroller reads and interprets the signals from the sensor element and reports to the climate control system using a PWM pulse-width modulation signal. This signal communicates the level of pollution to the climate control, which in response initiates the actuation of the fresh-air inlet. The use of a microcontroller enables us to enhance the performance of the climate control system by programming it to behave in accordance with the wishes of the customer. The software determines, for example, the total fresh air inlet closed versus open time, the structure of the PWM signal output, sensor behavior at start up, and many more variables.

The ability to detect gases harmful to health is not the only necessary attribute of the AQS microsystem, it must also be capable of being integrated into the complete automotive electric system.

The AQS is designed as a stand-alone modular system. It incorporates all the functions necessary to make it an independent, yet adjustable, device within the heating, ventilation, and air-conditioning system. It features a microcontroller that can evaluate the sensor signals and deliver a PWM signal to the climate control unit. The climate control unit in turn decides the relevance of the signal and what to actions to implement.

This is a good example of shared responsibilities and priorities within the system. The AQS is responsible for reporting the level of air quality, no matter what

situation prevails. However, the climate control finally determines how to respond to the AQS signal, which makes the AQS a true and responsible subsystem.

One may assume that the AQS is a sort of 'electronic nose'. The human nose reacts immediately to bad smells to protect us from rotten food or harmful gases. However, a nose functions rather differently from the AQS. The nose detects primarily ketenes, aldehydes, mercaptans, mecaptans, not mercaptanes and similar substances, often together with carriers such as methane. The nose can detect some gases in incredibly small concentrations of only a few ppb. The evaluation of these gases then is subject to a very complicated process that is determined by a lot of different factors that make it a very subjective, even culturally dependent, process. So a standard or normative evaluation is not possible.

Most sensors are not able to detect gas concentrations in the ppb area as pollutants are usually oxidizable substances that react to slight changes in parameters such as humidity.

But why should one use an AQS instead of filters? Filters have quite narrow limitations. Harmful particles in exhaust fumes are very tiny, about 1 μm , which are difficult to filter out, and filters have to be replaced about every 6 months.

The AQS automatically closes the air inlet door, in order to reduce the amount of pollution carried into the vehicle cabin when the vehicle enters high pollution areas. This provides significant health, safety, and comfort benefits to drivers and passengers, and significantly extends the life of cabin air filters. The health benefit comes from the reduction in exposure to harmful substances, and the comfort and safety benefit is from the automatic operation of the system, allowing the driver to concentrate on the driving tasks. The automatically reopening of the air inlet eliminates the possibility that the driver might forget to do this. Forgetting to reopen the inlet could cause fogging of the windows, leading to lower visibility, or cause low oxygen levels in the cabin, leading to drowsiness, which can be the cause of accidents. The AQS senses the unhealthy gases in the environment before they enter the cabin. The driver cannot detect many harmful gases directly because they are odor free; they can only detect the odor of some gases that are coincident with the harmful gases. Leaving detection to the driver would also mean that a significant amount of harmful substances would be in the cabin before the driver could operate the recirculation switch.

The paragon AQS is a fully integrated system consisting of a sensor element, microcontroller-based signal processing, and an interface. The AQS detects both reducible gases, such as NO_x , and oxidizable gases, such as CO, with only one sensor element. The sensor element is fabricated utilizing thick-film technology, which is well recognized as very robust. The special mixed oxide sensor element is not limited to the detection of only a few gases. This sensor is unique because it is a multi-gas sensor that senses a broad range of the gases commonly found in everyday traffic environments. It senses oxidizable and reducible gases such as CO, benzole fumes, gasoline, *n*-hexane, *n*-heptane, *n*-octane, sulfur dioxide, hydrogen sulfide, and carbon disulfide.

A highly sophisticated electronic circuit enables the dual element to detect the broad range of gases. This unique capability is the reason for the unparalleled

cost effectiveness. The AQS signals the electronic climate control to close the air inlet when detecting a pollution peak. If necessary, the climate control system can override the AQS signal, if other functions take priority, such as fast cool down. The AQS is designed to detect very low concentrations of gasoline and diesel exhaust gases, which is beneficial to the health, safety, and comfort of the vehicle occupants. The AQS should be placed near the air inlet in the cowl or plenum under the base of the windshield. Here it would be exposed to the flow of air that would be pulled into the vehicle when the air inlet is open.

The system offers dynamic adaptation to various driving environments such as a city, rural, traffic jams, and tunnels. The sensor software provides for adaptable sensitivity control to adapt continuously to the ambient pollution levels. The sensor incorporates a self-learning feature to provide continuous adjustments that yield consistent sensor performance over the life of the vehicle.

The system's sensibility control acts as an automatic compensator for sensor tolerances during the lifetime of the car. The AQS controls the temperature of the sensor element with complex software algorithms to provide for a 'soft start' for long sensor life and algorithms to determine the dynamic adaptation to the driving situation. This software algorithm allows the sensor temperature to be regulated to give consistent function over a wide ambient temperature range without being influenced by the speed or temperature of the air stream.

The AQS incorporates self-diagnostics. The microcontroller monitors the function of the sensor. If there is a malfunction, an automatic new initialization occurs. This function is a safety algorithm eclipsing the normal program. Diagnosis data are sent to the ECC. The AQS microcontroller can meet a variety of interface specifications for PWM or bus compatible applications, with appropriate output overload protection. It generates the interface signals according to the interface definition, and it controls current and voltage of the interface and deactivates the interface in case of overload.

The AQS architecture provides clear system responsibility as a complete stand-alone device, so the manufacturer is responsible for the complete system and test runs. AQS comes with excellent future prospects, since potential further developments can be integrated without any problems due to the unchanged interface once chosen, and compatibility with older releases is secured. The system configuration allows great savings in the ECC.

7.17.5

System Flexibility

AQS is a flexible system that can easily adapt to each type of car. The type of interface can be chosen: digital, quasi-analogue (air-quality levels), or analogue. We use a data telegram PWM protocol or any bus definition. The motor of the ventilation flap can even be controlled directly. The customer can choose 12 or 5 V supply, and different water protection levels.

The AQS MK III Universal is dedicated for use in trucks and for after-sales use, it has a standard, easy-to-fix casing, and the cable-harness connectors can be selected.



Fig. 7.17.1 All different generations of air-quality sensors

The AQS MK III+ is a variant intended for use in up-market cars; it gives the highest levels of comfort and security, because of its 100% adaptation to the driving circumstances. Various AQS are shown in Fig. 7.17.1.

The AQS Mk IV is the smallest fully integrated AQS, designed for compact and medium-sized vehicles as a universal device (Fig. 7.17.2). It is waterproof (IP 69 standard) and comes with an integral bayonet mount and an optional mounting clip, both allowing assembly without tools. The small size and cost efficiency make it suitable for high volume production. The Mk IV software allows for application-specific variations by offering selectable parameters. This allows for sensors that meet the specific needs of various vehicle models while maintaining identical physical attributes. The Mk IV provides a high performance at a power rating of only 0.8 W. It warms up automatically for 35 s, then it starts analyzing the incoming air; pollution spikes are detected in 1 s or less. It weighs only 14.5 g, which is lighter than other systems. Because of its robust technique, controlling, and self-diagnostic features it offers a trouble-free working lifetime of more than 10 years. The standard integral bayonet mount and optional mounting clip mean that it



Fig. 7.17.2 The fourth generation of AQS – the smallest fully-integrated air quality sensor worldwide

can be fitted in about 2 min: the worker has to stick it to the base and make the electrical connection, no tools are required. So the assembly can be easily integrated into existing production lines.

The AQS MK IV MF (multi-function) as the latest product of the AQS portfolio, which has a LIN interface and software that can adapt, or self-learn, over a number of duty cycles. It includes a subroutine called 'city-detect', which adopts the software routine best optimized for the special requirements of inner cities. It is also particularly sensitive to tunnel entry/exit. The software takes care that the air inlet door only opens after the tunnel exit. AQS MK IV MF even offers a skunk detection function, which is embedded in the software and uses a second sensor element especially to detect that special smell.

7.17.6

Conclusions

Standardization of the AQS system enables paragon to provide sensors of superior value and quality, which enables car manufacturers to supply comfortable cars with a high performance air-quality management system. As for other components, the trend is that more and more compact cars are equipped with air-quality sensors to improve safety, health, and comfort. The customer is offered an effective air-quality management, with plenty more potential for the future.



Fig. 7.17.3 The latest generation of air quality sensors with additional, multi-functional features

7.18**Chemical Sensors for Liquid Media**

BERNHARD JAKOBY and FALK HERRMANN

7.18.1**Introduction****7.18.1.1 Motivation**

The monitoring of liquids such as engine oil, gear box oil, fuel, and battery liquid has received increasing attention in automotive applications because of the increasing availability of robust liquid sensor technologies. Liquid sensors can be used to detect the quality of refilled liquids because of the potential impact on exhaust-gas composition (in the case of fuel and, to a certain extent, engine oil). Many automotive liquids become exhausted, and it is important to know when the liquids, such as engine oil, should be changed or refreshed. Other liquids are usually 'life-time fillings' such as gear oil or battery liquid. In this case aim is the detection of abnormal states (increased particle contamination in gear oil or decreasing charge in battery liquid).

7.18.1.2 Physical Chemosensors

At first sight, sensing tasks in a liquid environment seem to require the use of dedicated chemical sensors. A typical chemical sensor uses an interface material that transforms the chemical information (e.g., the concentration of a certain substance) into a physical quantity that can be sensed by a corresponding physical sensor (Fig. 7.18.1 a). A typical example would be a (selectively) adsorbing material that changes its density, conductivity, or another physically measurable parameter as a consequence of chemical adsorption. However, chemical interfaces often feature adverse properties such as lack of reversibility of the chemical reaction, potential 'poisoning' of the material, and poor long-term stability. In automotive and other applications, so-called 'physical chemosensors' [1] are preferred whenever feasible. A physical chemosensor is defined as a purely physical sensor (or sensor system) for determining a 'chemical' material property. Examples would be the determination of liquid mixing ratios or the observation of fermentation processes by means of density or viscosity measurements. Of course, as removing a chemical interface potentially means a loss in selectivity, the accuracy of the method is determined by knowledge about the underlying process, such as, e.g., the components used in mixing or the kind of fermentation process considered. Alternatively, an array of physical sensors with a subsequent algorithm processing the sensor signals represent a physical chemosensor (Fig. 7.18.1 b). For automotive applications a typical example of this latter type is the engine oil condition sensor discussed below.

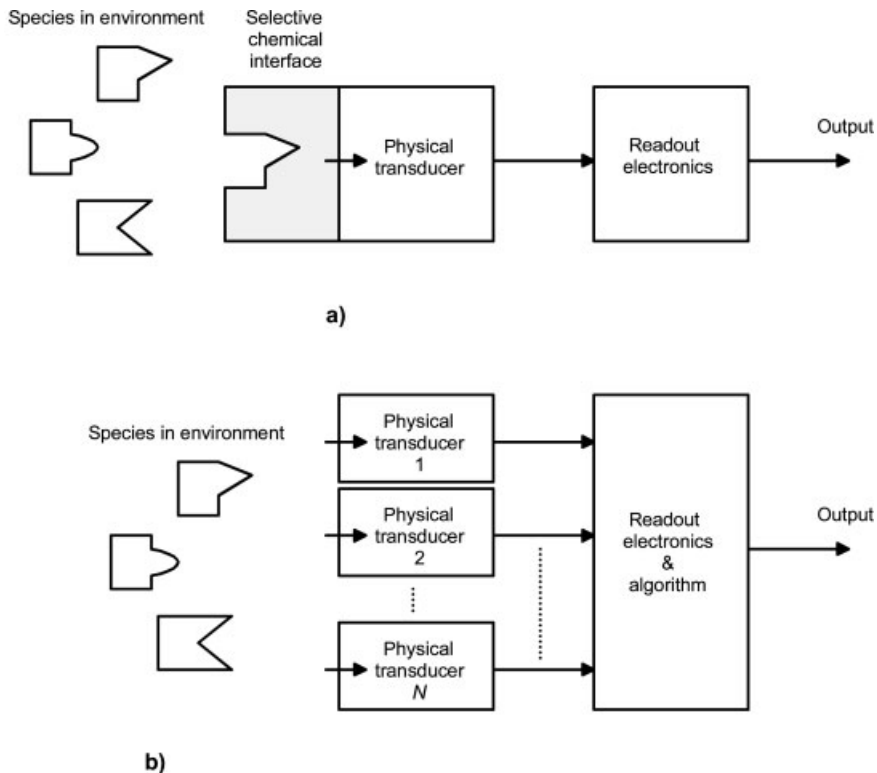


Fig. 7.18.1 The concept of physical chemosensors replacing traditional chemosensors: a) conventional chemical sensor; b) array implementation of a physical chemosensor, in

which a number of different physical sensors are used. The outputs of these sensors are processed by a designated algorithm in order to obtain the desired chemical information

7.18.2

Sensor Principles for Automotive Liquids

Concentrating on physical chemosensors for automotive liquids, we can set up a matrix indicating sensor principles applicable for certain applications. There are a number of physical parameters that are potentially interesting for physical chemosensor applications. They can be categorized according to the energy domain.

- Electrical liquid properties (e.g., conductivity, permittivity).
- Thermal liquid properties (e.g., thermal conductivity, heat capacity).
- Mechanical liquid properties (e.g., viscosity, density, sound velocity).

In Tab. 7.18.1 we indicate the suitability of four important physical parameters currently considered in sensor research and development (conductivity, permittivity, viscosity, and density) for some important automotive liquid sensing tasks. Although conductivity and permittivity represent parameters that in general show different

Tab. 7.18.1 Suitability of physical liquid parameters for different sensing tasks

<i>Liquid</i>	<i>Liquid property to be determined</i>	<i>Conductivity/permittivity</i>	<i>Viscosity</i>	<i>Density</i>
Fuel (gasoline, diesel)	Type	+	(+)	(+)
	Level	+		
Engine oil	Type		+	
	Degree of deterioration	(+)	+	
	Level	+		
Transmission oil	Degree of contamination	(+)	(+)	(+)
Battery liquid	Charge	+	(+)	+
Brake fluid	H ₂ O contamination	+	+	

+ = parameter is suitable; (+) = parameter is suitable to a certain degree

correlations to the properties to be sensed, they can be measured using similar sensor (electrode) configurations and thus the table lists them in one column.

The above discussion concerns selected applications and physical parameters but it illustrates the general implementation strategy for physical chemosensors.

- Consideration of a set of reliably implementable sensor principles and evaluation of their correlation with the considered ‘target property’ of the liquid.
- Implementation of these sensors together with a suitable readout concept (e.g., including an algorithm) allowing for the evaluation of the target property based on the measured physical parameter(s).

Most of the listed applications represent examples that are still in research and development.

7.18.3

Fuel Sensing

7.18.3.1 Gasoline, Flex Fuel

Background

Since the oil crisis in the 1970s, alternative fuels have received attention and have been used to a certain extent. One interesting alternative to gasoline is a mixture of gasoline and ethanol or methanol. In Brazil up to 30% ethanol (gained from sugar cane) is blended into gasoline. Because of the different calorific values of those blended fuels, the engine control algorithms would benefit from having information about the mixture in order to adapt the engine control accordingly.

Sensor Technology

For the determination of alcohol content in gasoline, the associated changes in the permittivity or the index of refraction can be utilized. Optical sensors suffer from sensitivity to contamination and thus the use of permittivity sensors is near at hand [2, 3]. Because of the temperature dependence of the permittivity, an addi-

tional temperature sensor is required for compensation purposes. The permittivity can be measured by conventional capacitive probes or alternatively by microstructured electrodes, in which the latter are more vulnerable to contamination. Currently flex fuel systems including corresponding sensors are mainly offered by North American car manufacturers [4].

For the sake of completeness we mention that there is also some interest in the determination of sulfur content of gasoline. However, the corresponding concentrations are very low, which, in our opinion, cannot be sensed reliably by means of the physical parameters discussed here.

7.18.3.2 Diesel, Biodiesel

Background

For ecological and political reasons, so-called biodiesel, most often in the form of RME (rapeseed oil methyl ester), has received much attention, especially in Europe. Biodiesel can be used as a replacement for conventional diesel fuel, however, in certain cases it can be beneficial to adapt the engine control if biodiesel is used [5]. Thus there is interest in a sensor detecting the presence of biodiesel and, for a mixture with conventional diesel, the mixing ratio in the fuel tank.

Sensor Technology

When evaluating physical sensors regarding their suitability for the discrimination of diesel types, viscosity, density, and permittivity can be considered (Tab. 7.18.1). Whereas differences in density are not significant, RME biodiesel shows a higher viscosity than conventional diesel: at 20°C it can have up to twice the typical viscosity of ordinary diesel. However, the viscosity may vary in a relatively broad range, for example the German standard DIN V 51606 prescribes the kinematic viscosity, ν , of biodiesel at 40°C to be 3.5–5 mm²/s. Thus the physical parameter viscosity appears less useful for the task of determining mixing ratios. On the other hand, the permittivity, ϵ_r , of (bio)diesel fuel seems to be better confined to certain values, although the permittivity is not a value prescribed in standards. Thus it appears near at hand to use a permittivity sensor for the discrimination between diesel and RME fuels [5].

Fig. 7.18.2 shows measured permittivities for three different kinds of ordinary diesel fuels, kerosene, and RME. The sensor has been calibrated with air (relative permittivity 1) and a standard liquid of known permittivity (cyclohexane, $\epsilon_r=2.023$ at 20°C). The measurement frequency was 100 kHz. It can be seen that all diesel fuels roughly show relative permittivities confined in a fairly narrow interval around 2.1 whereas RME yields values well above 3.1. Accordingly a mixture should yield permittivities between these values, which allows for the determination of the mixing ratio. Strictly this holds only as long as no emulsion is formed [9].

Fig. 7.18.3 shows the output of a prototype capacitive sensor fabricated by Bosch, showing a linear dc output voltage depending on the RME content in the fuel mixture. The sensor utilizes a coaxial capacitive probe and a dedicated read-out circuit operating on a 5 V supply.

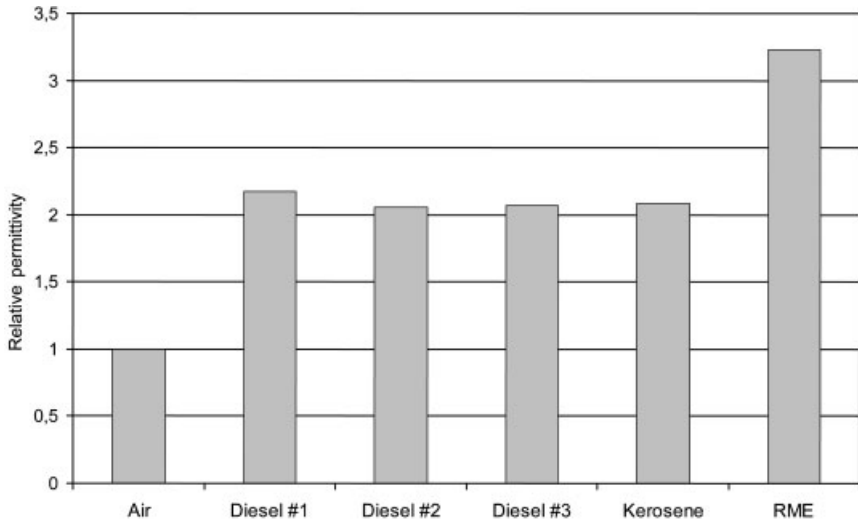


Fig. 7.18.2 Relative permittivities of three different types of diesel fuel, kerosene, and RME measured at 100 kHz compared to air (reference value). The variation between different

types of ordinary diesel fuels (and kerosene) is much smaller than the difference between RME and the diesel fuels

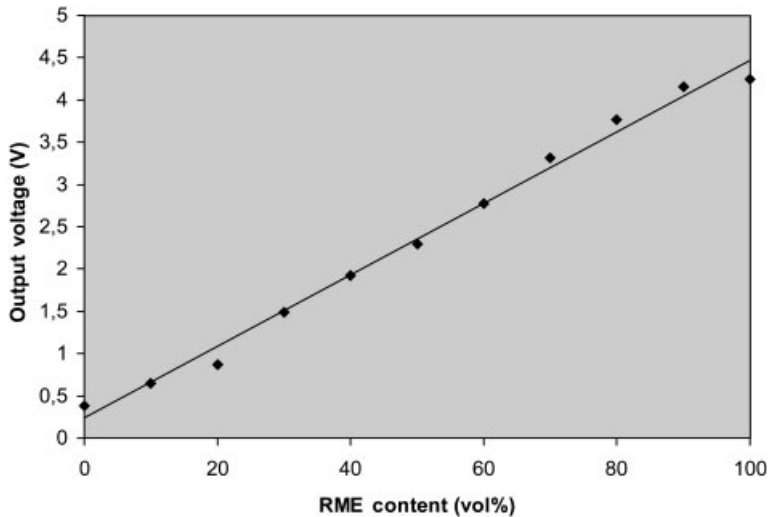


Fig. 7.18.3 Output of a prototype sensor for RME (biodiesel) detection in diesel fuel

Depending on future political decisions about the release of alternative diesel fuels (especially within the European Union), such diesel-type sensors may be implemented in a broad range of cars.

Above we briefly mentioned the application of viscosity sensors for diesel fuel, which appeared to be less suited for biodiesel detection. However, as the injection

time and amount in common-rail systems depends on the viscosity of the used diesel, it can be beneficial to measure this parameter in order to adapt the parameters in the injection control [6]. The microacoustic sensors described in the following section are suitable candidates for such an application.

7.18.4

Oil Condition Sensing

7.18.4.1 Background

Monitoring the engine oil condition allows increased oil drain intervals. Moreover it provides an insight into the state of the engine, which enables the detection of possibly approaching engine failures, and gives information about the performance of 'fresh' engine oils of varying quality.

The current state of the art in oil condition monitoring relies on algorithms for indirect evaluation of the actual oil condition [7]. This is achieved by means of processing driving parameters such as elapsed mileage since the last oil change and the number of cold starts. Currently some of these algorithms are supported by sensors measuring the permittivity [8] or conductivity of the oil. The latter parameters correlate with a number of important oil properties such as water content [9] or acidity. However, in order to increase oil drain intervals even more and protect the engine against overly deteriorated oil at the same time, further physical oil parameters need to be considered.

Basically an algorithm evaluating the actual oil condition (leading to a 'change oil' signal at a certain instant) can be structured as depicted in Fig. 7.18.4. The next generation of algorithms will utilize the indirect driving parameters provided by the engine management together with signals from dedicated oil condition sensors. The associated algorithm can conceptually be split into an engine-plat-

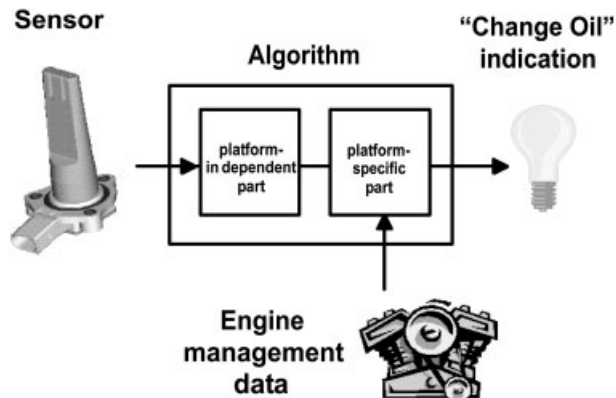


Fig. 7.18.4 Basic structure of an oil-condition algorithm including a sensor and data from the engine management. For a multi-sensor, this arrangement resembles the configuration shown in Fig. 7.18.1 b supplemented by data from the engine management

form-independent part processing the sensor signals whose output will be fed into a platform-specific part, which also considers driving parameters obtained from the engine management. While the traditional approach of oil condition evaluation does not feature a sensor (and the associated part of the algorithm), future efforts will aim at completely eliminating the platform-specific part, which would save resources in platform-specific software implementation. This would furthermore allow the integration of the entire signal processing into the sensor unit rather than the engine control unit (ECU). This approach is considered already nowadays in an oil-condition sensor based on a conductivity measurement [10]. But since a mere conductivity measurement only yields a very vague picture of the oil condition, the safety margins in the algorithm have to be so high, that no increased drain intervals can be achieved, which makes this solution less attractive, at least for the European market.

7.18.4.2 Multi-Functional Oil-Condition Sensor

Here we describe a novel sensor under development by Bosch, which measures the viscosity, permittivity, temperature, and level of the oil [11]. The sensor housing is designed for insertion at the bottom of the oil pan where the opening in the oil pan is sealed by means of an O-ring. The measured viscosity and permittivity are the primary quantities supporting the oil condition evaluation, the temperature measurement is necessary to compensate the temperature-dependence of the sensor elements and because the measured parameters are temperature-dependent itself (especially the viscosity). Monitoring the oil level represents an extra feature, which makes the sensor a suitable replacement for already existing oil level sensors. This kind of sensor will be introduced to the market in 2004.

The viscosity sensor uses a microacoustic sensor device [11], in which the piezoelectric effect in a crystal is used to excite (and detect) acoustic vibrations. For microacoustic liquid sensing, special wavetypes are used in order to avoid unwanted radiation losses due to the excitation of compressional waves in the liquid. This



Fig. 7.18.5 Housing of the multi-functional oil condition sensor (to be inserted at the bottom of the oil pan)

can be achieved by using shear-polarized waves, which do not couple to compressional wave modes.

Liquid loading of a device performing a shear oscillation leads to an entrainment of a thin liquid film with exponential decay of the entrained shear movement (Fig. 7.18.6), in which the decay length δ is given by [12]

$$\delta = \sqrt{\frac{2\eta}{\rho\omega}}$$

Here η and ρ denote the dynamic viscosity and the mass density of the liquid, and ω is the angular frequency of the oscillation. The entrainment leads to a mass-loading of the device causing a change in resonance frequency (in case of a resonator) or a change in phase velocity (in case of propagating wave) and a damping of the oscillation of the wave. Both effects are to first order proportional to $(\omega\rho\eta)^{1/2}$ and thus a sensor for the viscosity-density product can be made. For non-smooth surfaces, parts of the liquid can be 'trapped' and lead to an additional pure mass-loading effect according to the density of the liquid. Thus using two sensors, one with a smooth and one with a corrugated surface, makes it possible to distinguish between the effects of viscosity and density [12, 13]. However, in many applications such as oil sensing, changes in viscosity outweigh changes in density such that sensors with smooth surfaces represent efficient monitoring devices for the viscosity.

A more detailed description of the application of microacoustic viscosity measurements to engine oil is given elsewhere [14], and a general overview on microacoustic liquid sensors has been published in [15]. Fig. 7.18.7 shows the viscosity signal of a sensor prototype utilizing an designated microacoustic device, which delivers an output signal being approximately proportional to $1/(\omega\rho\eta)^{1/2}$. The prototype also comprises a temperature sensor close to the viscosity sensor element, allowing to record viscosity-temperature profiles. Fig. 7.18.7 shows the output characteristics for three new oils (Esso 0W30, 10W40, and 15W40) and it can be

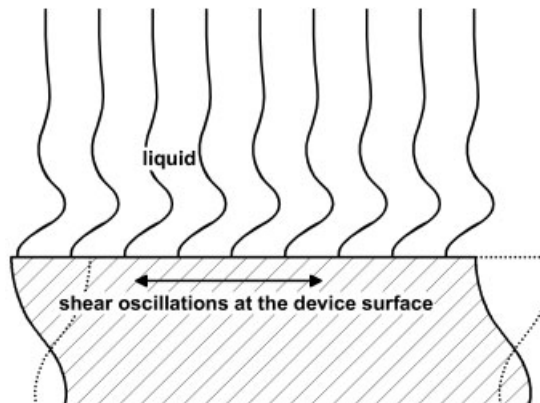


Fig. 7.18.6 Entrainment of liquid with a shear oscillation

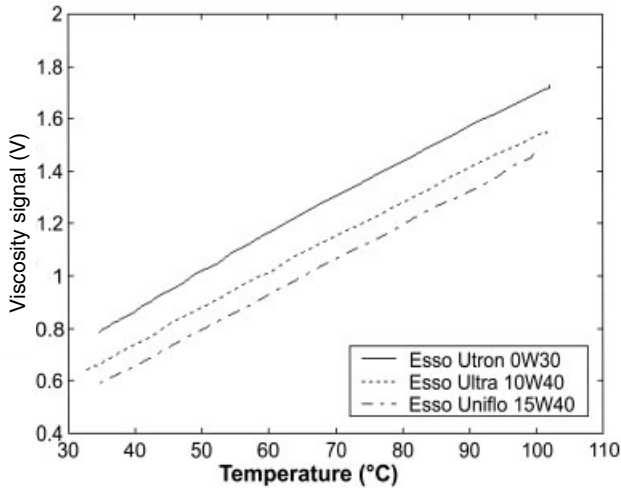


Fig. 7.18.7 Viscosity signal of a prototype sensor versus temperature for three oils of different SAE classes

seen that each oil approximately corresponds to a straight line in the plot, allowing interpolation if only a few points of the characteristics have been determined. As expected, the oil with highest nominal viscosity class (15W40) yields the lowest sensor signal (indicating high viscosity) and vice versa.

Fig. 7.18.8 shows sensor-based measurement results for oil samples obtained from diesel engines, which have undergone different engine tests. All measurements

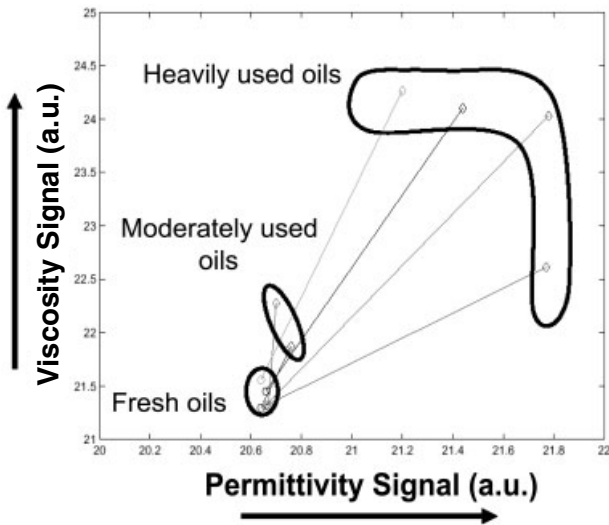


Fig. 7.18.8 Simultaneous measurement of permittivity and viscosity for oil samples obtained from diesel engines

have been made at a single temperature (25 °C). The ‘fresh oil’ reference measurements are connected by straight lines with the associated measurement points for the deteriorated sample. It can be seen that in general increasing deterioration yields increasing permittivity and viscosity, but the effects are different for different tests showing the necessity of measuring at least two parameters. Note that deterioration does not always yield increasing viscosity, for instance oil in gasoline engines can become considerable diluted by fuel, which lowers the viscosity. Such opposite trends need to be considered by the processing algorithm (Fig. 7.18.4).

7.18.5

Other Automotive Liquid Sensors

Fuel- as well as oil-level sensors using capacitive and thermal principles are state of the art and are thus not further discussed in this contribution.

In contrast to engine oil, transmission (or gearbox) oil usually represents a lifetime filling unless unexpected contamination due to failures occurs, which, depending on the kind of contamination, may be detected potentially by all the physical parameters considered. For the detection of metallic wear particles in transmission oil, both ferromagnetic as well as non-ferromagnetic, the utilization of magnetic sensor principles would be near at hand [16, 17].

The state of charge of automotive batteries can be determined by monitoring the acidic concentration within the battery cell. As more and more safety-relevant features crucially rely on the availability of battery power, it can be expected that on-board battery monitoring will become a necessity in the near future. One way to determine the acidic concentration would be the determination of electrolyte density, which varies from typically 1100 kg/m³ for a discharged battery (15% H₂SO₄) to about 1280 kg/m³ for a fully charged battery (42% H₂SO₄). In the same range the dynamic viscosity changes from 1.2 mPas to about 3 mPas (at 25 °C) [6]. Thus the utilization of a microacoustic viscosity sensor as described above is near at hand. Because of the considerable and also varying conductivity of the electrolyte, care has to be taken to avoid acousto-electric cross-coupling or unwanted electric influence on acoustic transducers [18]. Cernosek et al. have published corresponding experimental results using transverse shear resonators (or QCM devices) [19].

Finally we mention the interesting application of monitoring brake-fluids. As the formation of vapor bubbles dramatically influences the hydraulic performance of the brake fluid, its boiling point must lie above 140 °C. Since brake fluids tend to be hygroscopic, water diffusion through the hydraulic tubing lowers the boiling point and thus brake fluids have to be renewed in regular intervals. To avoid unnecessary liquid changes or provide warning signals in case of a rapid increase of the water content (in case of failure), monitoring of the water content would be beneficial. In principle, permittivity sensors could be used for this purpose if they meet the mechanical and physical requirements for the integration into the hydraulic system. Apart from these boundary conditions, sensitivity, and calibration issues (because of the varying permittivities of different brands of brake-fluids) would have to be solved.

7.18.6

Conclusions

Regardless of the wealth of research that has been performed on liquid sensors recently, so far only a few liquid sensors have made progress towards actual utilization in automotive applications. The oil condition sensor described above represents the cutting edge in automotive liquid sensing and its introduction into the market is expected for 2004.

The reason for this slow-going progress lies in two, often mutually contradictory, requirements for automotive equipment: high-tech devices should be provided at consumer prices. Furthermore reliability requirements including specifications regarding aging, drift, and robustness in a harsh environment represent tough demands on devices and systems. Hence, from our point of view, physical chemosensors, which by definition avoid the introduction of fragile chemical interfaces into the automotive environment, represent a robust concept suitable for a number of applications described above.

7.18.7

Acknowledgments

The authors would like to thank Matthias Buskies, Heinz Eisenschmid, Gerald Hamm, Markus Niemann, Jeroen Nieuwenhuis, Oliver Schatz, Monika Scherer, Michiel Vellekoop, and Matthias Ziebell for many stimulating discussions.

7.18.8

References

- 1 M. J. VELLEKOOP, in *Proc. Transducers'01*, Munich, Germany, June 2001, pp. 770–775.
- 2 A. H. MEITZLER, G. S. SALOKA, SAE Technical Paper 920699.
- 3 R. A. PINNOCK, in *Proc. IEE Colloquium on Automotive Sensors*, pp. 4/1–4/3, 1992.
- 4 www.fueleconomy.gov/feg/flextech.shtml
- 5 German Patent Application DE 10042490 A1, 2000.
- 6 F. HERRMANN, Doctoral Dissertation, Braunschweig University of Technology, published by Shaker, Aachen, Germany, 2000.
- 7 R. THOM, K. KOLLMANN, W. WARNECKE, M. FREND, SAE Technical Paper 951035.
- 8 E. IRION, K. LAND, T. GÜRTLER, M. KLEIN, SAE Technical Paper 970847.
- 9 B. JAKOBY, M. J. VELLEKOOP, in *Proc. Euro-sensors XVI* (to appear), Prague, Czech. Rep., Sept. 2002, pp. 53–56.
- 10 S. S. WANG, *Sens. Act. B* 2001, 73, 106–111.
- 11 B. JAKOBY, M. BUSKIES, M. SCHERER, S. HENZLER, H. EISENSCHMID, O. SCHATZ, in *Advanced Microsystems for Automotive Applications 2001* (S. KRUEGER and W. GESSNER, eds.), Springer, Berlin 2001, pp. 157–165.
- 12 S. J. MARTIN, G. C. FRYE, K. O. WESSENDORF, *Sens. Act. A* 1994, 44, 209–218.
- 13 F. HERRMANN, D. HAHN, S. BÜTTGENBACH, *Sens. Act. A* 1999, 78, 99–107.
- 14 B. JAKOBY, M. SCHERER, M. BUSKIES, H. EISENSCHMID, in *Proc. IEEE Sensors 2002*, Orlando, FL, June, 2002, pp. 1587–1589.
- 15 F. HERRMANN, B. JAKOBY, J. RABE, S. BÜTTGENBACH, in *Sensors Update*, Vol. 9, (BALTES, HESSE, KORVINK, eds.), VCH, Weinheim 2001, pp. 105–160.

- 16 J. L. MILLER, D. KITALJEVICH, in *IEEE Aerospace Conference/Proceedings* (published by IEEE), 2000, vol. 6, pp. 49–56.
- 17 German Patent Application DE 39 31 497 A1.
- 18 B. JAKOBY, M. J. VELLEKOOP, *IEEE Trans. UFFC* 1998, 45 (5), 1293–1302.
- 19 R. W. CERNOSEK, S. J. MARTIN, K. O. WESSENDORF, in *Proc. Acoustic Wave-Based Sensors Symp., 186th ECS Mtg.*, Miami, FL, 1994.

7.19

Electric-Current Sensors

HENRIK SIEGLE

7.19.1

Introduction

In 1998 the world market for electric-current sensors was 168 million Euro and it is expected to grow to about 260 million Euro by 2008. The automotive sector will contribute about 20% [1].

Contactless magnetic current sensors, in particular, are attracting attention from system suppliers and vehicle manufacturers for a large variety of applications. The spectrum ranges from simple monitoring functions (such as electronic resettable fuses, over-current protection, or current control in the after-burners of modern diesel catalyst systems) to precise and very fast measurement in current-controlled motors and electric drives. The latter is becoming a major field of application as more and more electrically-powered actuators are being used in cars. In these systems the current sensor acts both as a feedback element for torque control and as a fault-condition monitor. Examples are active chassis systems, power steering (electric or hydraulic), and combined starter-alternators. Because of the growing number of power-consuming devices, energy management is becoming increasingly important, and current sensors are necessary for functions such as battery diagnostics, charging control, or automatic system-deactivation when the engine is turned off.

The wide variety of applications gives rise to a wide variety of sensor requirements. Current is measured over a range from milliamps to hundreds of amperes with application-specific accuracy and frequency bandwidths. In some systems, such as energy management, the challenge is both to resolve fractions of milliamps and to withstand over-currents of thousand amperes and more. Current frequencies are typically between dc and some 10 kHz.

The requirements regarding ambient temperature, shock resistance, media resistance, and electromagnetic compatibility (EMC) depend strongly on where the sensor will be installed in the car. Typically, sensors in the car interior are exposed to temperatures between -40 and $+85^{\circ}\text{C}$ while temperatures under the hood and near the engine can reach $+125^{\circ}\text{C}$ or more. Sufficient EMC essential when sensing low currents because even the smallest interfering fields can result in large

deviations. Other requirements are low power consumption, compact size, and low system costs including assembly and packaging, installation and calibration, as well as signal processing.

7.19.2

Technologies

There are two principal methods of detecting electric current (Fig. 7.19.1). In the first, the voltage drop across a shunt resistance placed in the current path is measured. This is the basis for *resistive current sensing* or *resistive shunt* (Chapter 6.2). In the second, electric current is derived from the magnetic field it generates (*magnetic current sensing*), which is oriented perpendicular to the current direction. The magnetic sensors used for this method are described in Sections 7.19.2.2–7.19.2.4. They are in parts based on the same technologies as those magnetic sensors for detecting speed, angle, or position (Chapters 5.7, 7.9–7.12), but differ in sensor arrangement, measuring range, and sensitivity, as well as signal formation. The optic method utilizing the Faraday effect in optical fibers is only suitable for very high currents, such as in high-voltage engineering [2].

Each method and technology has its own advantages and disadvantages. The major advantage of magnetic current sensing is the galvanic separation from the circuit. An overview of current-sensing technologies and a selection of suppliers is given in Tab. 7.19.1.

7.19.2.1 Resistive Current Sensing and Resistive Shunt

Resistive current sensing is the oldest method of measuring electric current. It has been known for decades, is well understood, and is widely used in all kinds of measuring equipment. In recent years, this method has been used more and more in automobiles because of its simple setup, rigidity, and standard technology.

The resistive shunt is a calibrated resistor placed in the current path (Fig. 7.19.2). According to Ohm's law the voltage drop across the resistor is directly proportional to the current. Highly linear and stable resistors are used as shunts to minimize deviations from Ohm law. A differential amplifier supplies the signal for further processing. This method can be used to measure both direct current (dc) and alternating current (ac).

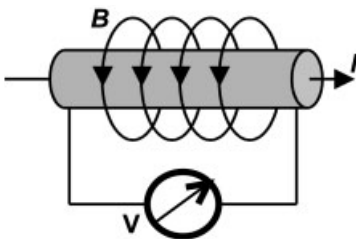


Fig. 7.19.1 The voltage drop and the magnetic field generated by an electric current can both be used to detect the magnitude and direction of the current

Tab. 7.19.1 Overview and comparison of current sensing technologies and suppliers

	<i>Shunt</i>	<i>Inductive</i>	<i>Open-loop Hall</i>	<i>Closed-loop Hall</i>	<i>AMR/GMR</i>
Galvanic isolation	-	+	+	+	+
dc/ac	+/+	-/+	+/+	+/+	+/+
Magnetic circuit/core needed	+	-	-	-	+
Power consumption	-	+	+	-	+
Size, weight	+	-	-	-	+
Costs	+(-) ¹⁾	+	+	-	+
Accuracy	+	-	-	+	+
Measuring range	-/(+) ¹⁾	-	-	+	+
Temperature range	+	+	-	+	+
Suppliers	Isabellenhütte		Allegro, EMCORE, Honeywell, Infineon, LEM, Melexis, Micronas, Yazaki		Honeywell, HL Planar, NVE, Sensitac

1) Depends on evaluating circuit.

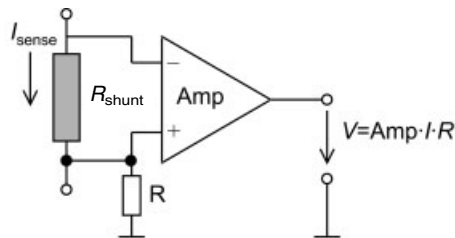


Fig. 7.19.2 Resistive current sensing

In many automotive applications the same conductor has to carry both low and high currents (100 A or more). Power consumption is given by $P = I^2/R$, so to reduce the power consumption, the resistance has to be in the region of a few milliohms, producing only a weak signal at low currents and thus necessitating a very high amplification. Even if the resistance is only 1 m Ω , the power dissipation can easily exceed 10 W, resulting in self-heating of the sensing element. Together with high ambient temperatures this may lead to a deviation from Ohm's law and the voltage across the resistor will no longer be a linear function of current. Contact resistances, with their temperature drifts, also contribute to these nonlinearities. Since changes in temperature are not instantaneous, nonlinearities will be a function of time and past current levels and therefore be very difficult to compensate. Consequently, care has to be taken in proper thermal design, heat sinking, and in using resistor materials with extremely small temperature coefficients. For permanent high currents there are precise shunt resistors on the market in the form of current bars made of alloys from Cu, Mn, Ni, and Sn. They combine low resis-

tance ($< 1 \text{ m}\Omega$), small temperature coefficients ($< 10 \text{ ppm/K}$), and low inductance values with the necessary long-term stability [3].

In addition to the characteristics of the shunt, the subsequent signal conditioner with its offset, amplification, and frequency bandwidth determines sensor performance. One problem arises because the amplifier and the current circuit use same potential. This may falsify the desired signal because of transient spikes, ground drifts, or electromagnetic coupling. In many applications the signal will be fed directly to a microcontroller or digital signal processor. For this purpose a potential-free measurement is required. This is generally not possible with shunt, necessitating corrective measures that increase sensor costs.

In summary, resistive current sensing is the simplest and the least expensive method of measuring mid-range currents. However, this technique has several drawbacks, especially when measuring both low and high currents as is the case in many automotive applications. Major disadvantages are the insertion loss (wasted power) and the lack of isolation between source and sensor. High accuracy over a broad temperature range requires the use of power resistors with flat temperature qualities and an adjusted high-end signal conditioner essential, changing the initially low-cost device into a high-cost device. Another disadvantage is that installing the shunt requires cutting and reconnecting the conductor, which is not always possible, for example in printed circuit boards (PCB) or high-voltage circuits.

7.19.2.2 Open-Loop Hall-Effect Current Sensing

Hall sensors utilize the Hall effect, discovered in 1879. When a current-carrying conductor is placed in a magnetic field, the Lorentz force deflects the current carriers generating a voltage perpendicular to both the current direction and the applied magnetic field. Hall elements are made of semiconductors, typically GaAs, InAs, InSb, or silicon, formed either as thin plates or, for silicon, monolithically integrated with an evaluating circuit. Despite the relatively poor material properties of silicon compared to the III–V semiconductors (regarding the Hall effect)

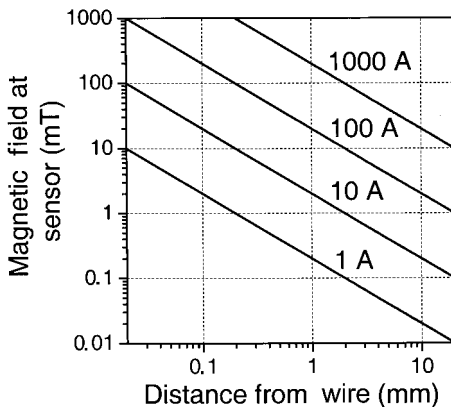


Fig. 7.19.3 Magnetic field of current-carrying wire as a function of radial distance

the possibility of integrating silicon Hall elements in CMOS technology is the deciding cost advantage.

The most direct approach to measuring electric current with Hall sensors is to place a Hall element a known distance from the to be monitored. One can deduce the current flowing through the conductor from the linear relationship (for small fields) between current, induced magnetic field, and the Hall signal. In practice, however, the magnetic flux going through the Hall element declines rapidly with increasing sensor-conductor distance. A current of 1 A generates a magnetic field similar to the Earth's field ($30 \mu\text{T}$) at a distance of 10 mm: too small for the not-very-sensitive silicon Hall elements. Manufacturing tolerances give a variation in the distance between sensor and conductor that diminishes the accuracy of the measurement. In addition, stray fields generated by currents through nearby devices (or even the Earth's magnetic field) will introduce more inaccuracies.

These are the reasons why almost all Hall-effect current sensors require the use of a magnetic circuit to direct the magnetic flux to the sensing element (Fig. 7.19.4). These circuits can be toroidal rings of highly permeable materials that trap magnetic flux. Iron, for example, has a relative permeability of about 2000 with respect to air. It is therefore 2000 times more able to carry magnetic fields and fields will be concentrated in this kind of material. The permeability of the usual core materials can even exceed 50 000 [4]. Magnetic circuits can also be formed as a laminated stack of magnetic materials. Since these layers are much thinner than solid cores, they can carry magnetic flux of higher frequency.

The magnetic sensor is placed into a slot cut into the core and the current carrying conductor or wire is placed through the hole. The magnetic field from the current is then trapped in the core and crosses the sensor at the slot, yielding an analog Hall output signal proportional to the current. This type of direct current sensor is called an open-loop device.

As is the case for all magnetic current sensors, these devices have the advantage of providing galvanic separation between the conducting and sensing circuit. They allow measurement of both dc and ac currents. They have low power consumption and no insertion loss. Because of the magnetic circuit the sensor is insensitive to the position of the current-carrying conductor and the magnetic field available to the magnetic sensor is drastically increased. The sensitivity is inversely proportional to the slot width. A further increase can be obtained by increasing the number of times the conductor passes through the magnetic circuit: two turns provides twice the signal of one turn. But this also increases the inductance

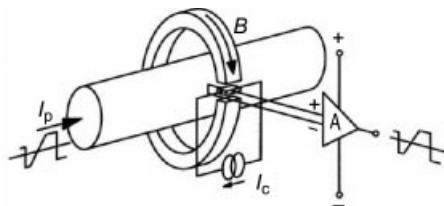


Fig. 7.19.4 Open-loop Hall sensor arrangement

and thus restricts the dynamic range. Similarly, it is possible to measure dc between two lines, since the flux from each current will be trapped in the magnetic circuit and cancel to leave just the difference. Another very important advantage of using a magnetic circuit is the shielding against interfering fields: stray fields from other currents or the Earth's field, because external field lines would rather remain in the magnetic material than jump the slot through the sensor. The slot has a much higher magnetic reluctance.

There are, however, also some drawbacks to these devices. Because of their poor sensitivity, silicon Hall elements provide only a very small output voltage, which must be highly amplified. Their inevitable offset voltage is subject to significant fluctuations because of amplitude and temperature coefficients, and the strong temperature-dependent sensitivity requires adequate compensation. Thus, similarly to the resistive current principle, the amplifying circuit often defines the quality of the sensor.

Also the use of a flux concentrating magnetic circuit is not without its problems. If the current (and thus the magnetic field) is too large, the material can become saturated, limiting the linear range of the sensor. When the flux concentrator is magnetically driven in one direction and the magnetic drive is removed, the material retains a small amount of field, the remanent flux, in the direction of the original drive. This results in a shift of the zero-current output, the magnitude of which depends on the history of the sensor. In order to minimize the hysteresis effect, soft magnetic materials are used as the flux concentrating material with remanent fields below 0.1 mT. Two further aspects to be considered are the additional weight and the bulky size of magnetic circuits.

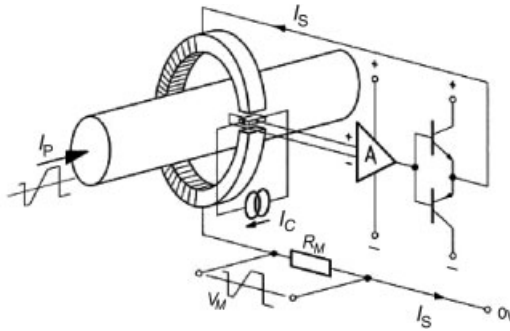
7.19.2.3 Closed-Loop Hall-Effect Sensors

Balancing the current being measured with a known current is the idea behind the so-called closed-loop sensors. This compensation principle avoids some of the nonlinear behaviors of the magnetic circuit and of the sensing element in the direct measuring mode described above, thus significantly improving the performance of the sensor.

In a closed loop, the sensor is placed in a compensating field that drives the flux through the sensor to zero. A coil with a set turns ratio generates the compensating field. Therefore, the current in the feedback coil is proportional to the field applied from the primary current scaled by the turns ratio. Turn ratios of 1000:1 and more can be achieved. Therefore, the sensor can be designed to measure currents of hundreds of amperes while only needing a few tens of milliamps of balance current. Since the net flux through the magnetic circuit is close to zero, saturation effects do not occur. Typically, a load resistor is placed in series with the coil to measure a voltage proportional to the input current. An example of how this could be done is shown in Fig. 7.19.5.

Apart from galvanic separation and the possibility of measuring both dc and ac currents, the closed-loop method has several other advantages. The range of currents being measured is much larger than in open-loop devices and only limited by how much current the feedback circuit can deliver. Since the sensor always

Fig. 7.19.5 Closed-loop Hall sensor arrangement



works close to zero field, the linearity is very high; the sensor exhibits an excellent accuracy and a low temperature drift. The bandwidth of closed-loop systems is generally higher and they allow faster response times than open-loop systems. In addition, the high loop gain in the feedback loop makes the performance insensitive to component variations.

Constraints of these kind of current-sensing devices are mainly their high power consumption due to the compensating field, and their high costs. Similarly to the open-loop systems, the weight and the large outline of the magnetic circuit are not advantageous. Also, the offset problem remains. But in summary, closed-loop current sensors are well suited for applications in which high accuracy and broad frequency range are needed and cost pressure is not the deciding factor.

7.19.2.4 Current Sensing with Magnetoresistive Elements

Magnetoresistive sensors are well suited to measuring both field strength and field direction. These devices are typically made of a nickel-iron (permalloy) thin-film deposited on a silicon wafer and patterned as a resistive strip or meander. The anisotropic magnetoresistive effect (AMR) causes the film to change its resistance by 2–3% in the presence of an in-plane magnetic field. To reduce temperature effects, typically four of these AMR resistors are connected in a Wheatstone bridge arrangement. In order to linearize the output signal, so-called barber-pole biasing is applied in which low-resistance shorting bars are placed across the film's width at 45° (for details see Chapter 5.7). AMR sensors are manufactured in batch processes on silicon and mounted in standard IC packages. They are small compared to Hall devices. Regarding electric-current sensing their main advantages are their high sensitivity, thus eliminating the need of field concentration, and their large bandwidth, which is typically up to 15 MHz. Another advantage is their good temperature behavior. Ambient temperatures of up to $+150^\circ\text{C}$ and higher are possible with much smaller offset voltage than silicon-Hall devices. AMR sensors can be designed for very high sensitivity and noise immunity, allowing the detection of extremely small fields even in the nanotesla range.

Similarly to the Hall-based current sensing techniques, the simplest way to measure an electric current with MR devices is to sense the magnetic field around the current conducting wire directly. Since they are sensitive in the film-

plane these sensors can be placed directly above the current track, which is especially useful in the case of current detection on PCBs. This geometry allows currents as low as fractions of milliamps to be measured because of the high sensitivity. In this direct approach, interference can arise from external magnetic fields, for example, the Earth's magnetic field or fields from nearby current-carrying parts. As with the Hall-sensor principles, this interference can be minimized by using magnetic flux concentrators but of course with the same disadvantages: hysteresis effects and bulky geometries.

Another way to overcome this problem is to measure in a differential mode either with a dual-sensor geometry or with a gradiometer layout AMR element. Fig. 7.19.6 demonstrates such a dual-sensor geometry in which the interfering field becomes a common mode signal that may easily be removed from the differential signal [5]. In one sensor the interfering field and the sensed field are summed, in the second sensor the fields are subtracted. This approach works for those cases in which the interfering fields are uniform over both sensing regions and low enough not to drive the sensors out of their linear range. Similar measuring geometries can be obtained with two sensors placed in a slot in a straight conductor [6] or on a U-shaped current path [6, 7].

Instead of two sensors it is also possible to use an AMR gradiometer, in which the two arms of the bridge are spatially separated. The barber poles have the same orientation in the two arms, so that only a field difference between the two positions is sensed. This configuration is not disturbed by interfering fields and has the advantage that it contains only one device. Therefore, it is not affected by element-to-element variations in sensitivity, offset, or temperature behavior.

Since for some applications the linearity even with barber-pole biasing is not sufficient, the use of compensating coils as described for the Hall sensors are necessary. In AMR devices these coils can be directly integrated above the sensing structures, so that the advantageous compact size of these elements remain [6, 7].

In 1988 the so-called giant magnetoresistance effect (GMR) was discovered by a French and a German team of scientists [8]. This effect occurs in stacks of alternating ultra-thin magnetic/non-magnetic layers and has between twice and ten times the size of the AMR effect, depending on the material system used. It is already applied in read-heads of hard-disk drives where it replaced AMR technology because of its much higher sensitivity. The sensing configurations of GMR sen-

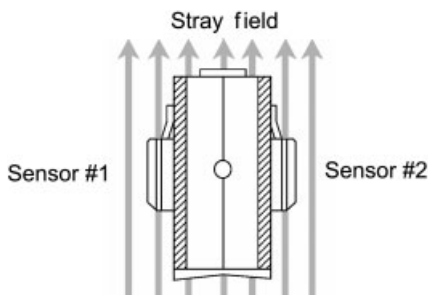


Fig. 7.19.6 Dual arrangement of magnetoresistive sensors

sors are similar to those of AMR, since the sensing direction is also in the thin-film plane. The first electric-current sensors based on this effect are now on the market [9] and it is expected that in many applications GMR will take the place of AMR technology, in the field of hard-disk drives.

In summary, AMR and GMR current sensors provide galvanic separation, high sensitivity, large bandwidths, and compact size. When used in differential mode no flux concentration is necessary to suppress interference from stray magnetic fields. Furthermore, they give good linearity, especially when combined with the compensation method. As is the case for all compensating principles, this implies higher power consumption.

7.19.3

Summary and Outlook

The sensing and management of electric currents is becoming more and more important because of the increasing number of electronic systems, electric drives, and other main electric power consumers in cars. We have given an overview of the main technologies for electric-current sensing that are on the market today. For automotive use, GMR technology is still in its infancy but is expected to penetrate this field in the near future. Until then, shunt, Hall, and AMR based-sensors will continue to be the main types of electric-current sensors. Each of these technologies has its own advantages and disadvantages. The application decides which sensing method is the most suitable.

7.19.4

References

- 1 Intechno Consulting, *Intechno Weltreport: Sensormärkte 2008*, Intechno Consulting, Basel 1999.
- 2 T. SATO, I. SONE, *Optical Review* 1997, 4, 35.
- 3 Precision resistances made of Manganin alloy from Isabellenhütte, Germany, exhibit a temperature coefficient below $10^{-5}/\text{K}$; see www.isabellenhuette.de.
- 4 The permeability of cores made of NiFe alloy exceeds 90 000. See also G. HERZER, *IEEE Trans. Magn.* 1990, 26, 1397.
- 5 Philips Data Sheet, *General Magnetic Field Sensors*, 1998; Honeywell, *Application Notes, AN-209*, 1999.
- 6 H. LEMME, A.P. FRIEDRICH, *Sensors*, 2000, 17 (5), pp. 82–91; Sensitec, *Data Sheet CMS2000*, 2001.
- 7 B. DRAFTS, *Sensors*, 1999, 16 (9).
- 8 M.N. BAIBICH, J.M. BROTO, A. FERT, F. NGUYEN VAN DAU, F. PETROFF, P. ETIENNE, G. CREUZET, A. FRIEDERICH, J. CHAZELAS, *Phys. Rev. Lett.* 1988, 61, 2472; G. BINASCH, P. GRÜNBERG, F. SAURENBACH, W. ZINN, *Phys. Rev. B* 1989, 39, 4828.
- 9 Nonvolatile Electronics, *Engineering and Application Notes, GMR Sensor Application Notes*, 9/10/1998.

7.20

Tire-Pressure Sensors

NORBERT NORMANN, GUNTER SCHULZE and WOLFGANG KELLER

7.20.1

Technical Status

Present tire-pressure monitoring systems provide the following functions.

- Identification and location of wheel positions.
- Monitoring of tire pressure.
- Visualization of tire pressure on starting vehicle.
- Early-warning of pressure loss, tires running at low pressure, and tire failure.
- Prevention of errors when setting the nominal pressure and when inflating tires.
- System diagnosis.

A typical tire-pressure measurement system is shown in Fig. 7.20.1.

Mass-produced tire-pressure monitoring systems perform these functions by using battery-powered wheel electronics that are able to measure the air pressure and temperature inside the tire. Together with coded information for individual wheel identification and data about battery lifetime, these values are transmitted as radio frequency data messages to an RF antenna mounted in the wheel arch. The data messages are sent independently by all the wheel electronics in a certain time-slot pattern or extracted on demand by using a transmission trigger. A central control unit evaluates the data, identifies the tire, and decides whether the driver should be informed. Each tire is monitored separately. The related air pressure is converted to the standard pressure by means of a temperature characteristic.

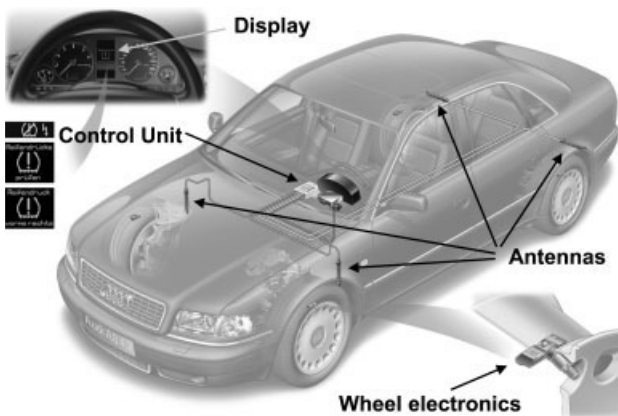


Fig. 7.20.1 Tire-pressure measurement system

The nominal tire pressure is provided either by the driver using a push button or by the manufacturer of the car. If the driver makes this input, the system checks it for plausibility to avoid set-up errors, and then the value is used for tire pressure monitoring.

Pressure monitoring is still possible after installing new or different wheels (for example, winter tires equipped with wheel electronics). The RF signals coming from the tire pressure-monitoring system of a near-by car do not influence the function because the system is capable of identifying its own wheels. The spare wheel can be integrated into the pressure-monitoring system too.

7.20.2

Components

Fig. 7.20.2 shows the components of the tire pressure measurement system.

7.20.2.1 Wheel Electronics

The wheel electronics consists of a sensor, the signal processing circuit, the RF-transmitter stage, RF-antenna, the battery, and the housing with filter: only a small antenna is needed to pick up the trigger signal (Fig. 7.20.3).

The design of the wheel electronics and the mounting technique on the rim makes it able to withstand the extremely severe operational conditions of the tire for a period of 7–10 years: temperature from -40 to $+150^{\circ}\text{C}$ or even 170°C for short periods, dynamic peak acceleration values up to $2000g$, extreme moisture and condensation, and also other solid or liquid substances in the tire such as those used to assist tire fitting (Fig. 7.20.4).

The key element in the wheel electronics is the sensor. Specially developed components for this application nowadays combine the pressure sensor, a temperature sensor, and the complete signal processing and data interface in the same hous-

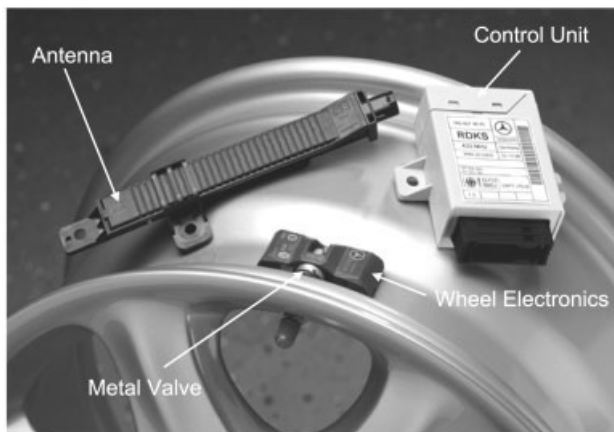


Fig. 7.20.2 System components

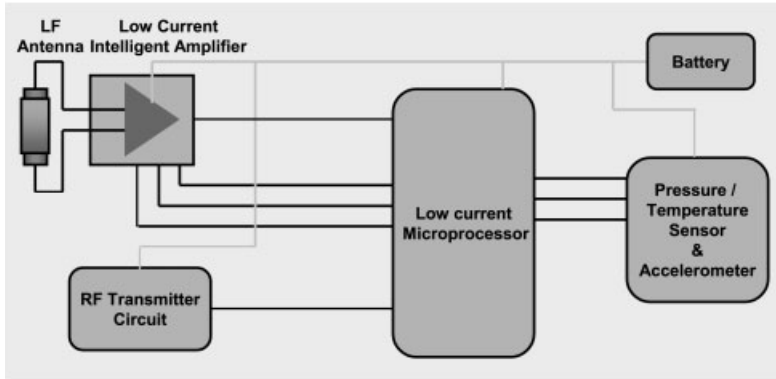
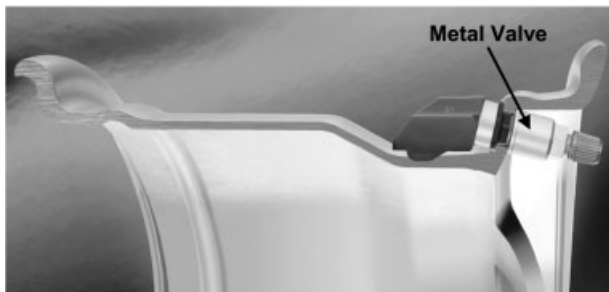


Fig. 7.20.3 Wheel electronics architecture



Universal applicability for various wheel rims and valve types using a metal-valve as a standard universal mounting method

Use of only one wheel electronic design across all vehicle platforms and wheel rim type combinations

Fig. 7.20.4 Wheel electronics mounting position

ing. A low-power processor controls the operating conditions of the wheel electronics and directly activates a digital module with integrated RF transmitter stage.

The power supply for the sensor and the transmitter is normally a lithium battery specially designed to fulfill the very high requirements for resistance against temperature, temperature changes, acceleration forces, and to guarantee an extra-long lifetime in these operating conditions. Technical data are shown in Fig. 7.20.5.

Pressure range	50 – 637,5 kPa, abs
Resolution	± 2.5 kPa
Accuracy	± 7.5 kPa @ -20 – 70 °C ± 15 kPa @ -40 – 100 °C ± 30 kPa @ 100 – 120 °C
Operating temperature	-40 – +120 °C
Peak temperature resistance	-40 – +170 °C
Measuring interval, normal mode	3.14 s
Transmission interval, normal mode	54 s
Measuring interval, rapid mode	0.85 s
Transmission interval, rapid mode	0.85 s
Pressure difference for transition normal to rapid mode	± 20 kPa / 54 s
Measurement independent of rotational speed	Sensor is always measuring
Battery-lifetime	7 – 10 years

Fig. 7.20.5 Wheel sensor technical data

7.20.2.2 Control Unit

The electronic control unit of a pressure-monitoring system consists of a micro-controller, data bus interface, and diagnostics. It evaluates the data it has received and transmits messages to the driver information system if necessary.

Control unit functions can be divided into two groups: tire pressure monitoring (the system section) and the specific data bus connection of the vehicle with the associated control and warning routines and integrated diagnostic possibilities (the communication section). These two functional sections are realized with two separate, autonomous program modules. This makes it relatively simple to customize the system according to the needs of different car manufacturers and types of vehicles.

7.20.2.3 Antenna

The RF antenna is designed for reception very close to the vehicle. The RF signals received by the antenna are converted directly into digital information. This avoids the need for an RF wiring harness. For calling-up data from the wheel sensor, the wheel casings are equipped with simple trigger transmitters which, in response to the demands of the control unit, causes the wheel electronics to transmit. The trigger function is performed by a simplified transponder technique. The basis wiring diagram is shown in Fig. 7.20.6.

7.20.3

Functions

7.20.3.1 Data Recording

A tire-pressure monitoring system has to perform two main functions: measurement of the absolute tire pressure and identification of pressure loss. For the measurement of the absolute value, data can be transmitted at relatively infrequent intervals, whereas a loss of pressure has to be transmitted immediately. In systems

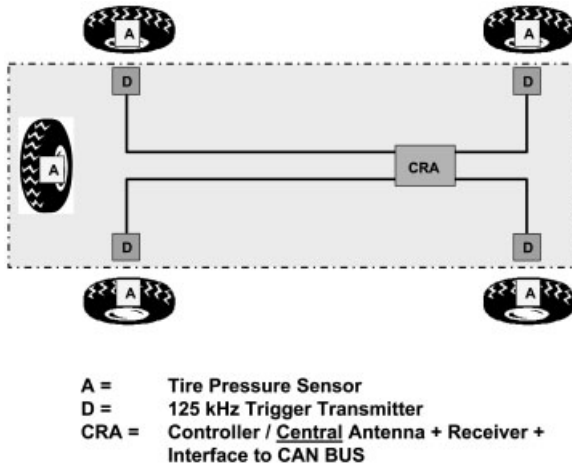


Fig. 7.20.6 Basic wiring diagram

without a trigger transmitter the wheel electronics themselves decide when data should be transmitted (for example, every 54 s). In systems with automatic wheel identification the transmission of data is triggered by the trigger transmitter. Each wheel electronics measures the tire pressure and the air temperature within the tire every 3 s. At a pressure loss of more than 0.2 bar, compared with the previous measurement, the wheel electronics sending their signals for a certain time at intervals of 0.8 s (fast transmitting mode) in order to provide sufficient data for a fast evaluation of the pressure status.

7.20.3.2 Data Transmission

An unidirectional RF data transmission principle is used from the wheel to the vehicle. The RF system is designed to ensure reliable transmission of the data in a wide variety of conditions such as statistical fluctuations of reception amplitude caused by varying reception conditions at the antenna. The transmissions take place in the 433 MHz range, the so-called ISM waveband; in some territories the 315 MHz waveband is used. Premium systems (Fig. 7.20.7) are able to process the data transmitted by the wheel electronics directly at the reception points by microcontrollers in the antennas. The signals are converted into a suitable data bus format and sent to the control unit by a LIN bus (local interconnect network), for example.

Data on Demand

By having the possibility for data to be called up at any time, the transmitting intervals of the wheel electronics (54 s is standard) can be considerably lengthened. The system, normally operating at 1 min transmissions, is able to detect possible wheel electronic defects after a certain period that must be long enough to bridge unavoidable gaps caused by 'lost' transmissions. The right timing is determined by the car manufacturer. Typical values are between 8 and 10 min. The integra-

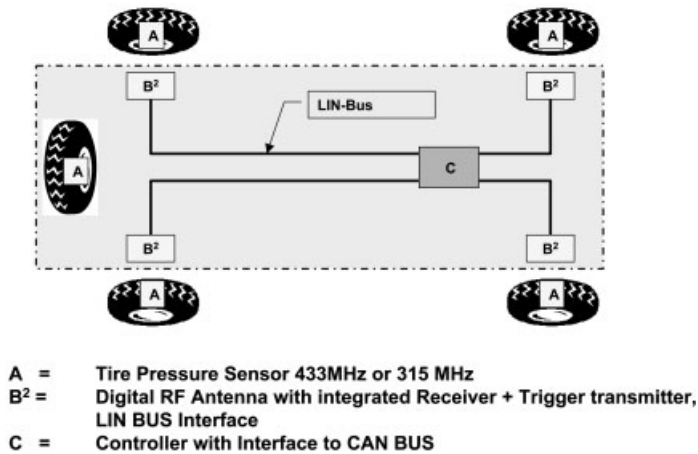


Fig. 7.20.7 Premium system wiring diagram

tion of a trigger function makes it possible to determine the system status at any time if messages are not received, to ensure that the wheel electronics are still working correctly. This is done by calling for a series of transmission operations. The normal transmission interval can thus be reduced to 2 min, for example. This halves energy consumption without reducing the safety of the system.

7.20.3.3 Nominal Pressure Input

The actual tire pressure after adjustment is either initialized by the driver pressing a button or the system adopts nominal pressure for the vehicle tires stored in system memory. The system checks the plausibility of pressure inputs, for example front/rear or left/right or the minimum pressure. If tires have been inflated correctly, the pressure values are monitored. If not, the driver is asked to adjust the pressure. Tire pressure is essential for driving comfort in the sense of low rolling noise, freedom from sharp, vertical movements and sensitivity to transverse road-surface joints.

After the wheels of the vehicle have been changed (for example, winter tires fitted), the system is able to identify the new wheels and their positions.

7.20.3.4 Pressure Low Warning

In regular vehicle operation, tire pressure changes only gradually. The typical slight loss of pressure due to air diffusion is reliably identified by the system as soon as the corresponding warning thresholds are exceeded. Tire failures as a result of excessive carcass flexing, caused by driving fast with insufficient tire pressure, can be avoided in this way: 85% of all tire failures are a result of this kind of gradual pressure loss (Fig. 7.20.8).

In the case of tire failure caused by a puncture, the driver is warned immediately (Fig. 7.20.9). If a significant loss of pressure occurs while the vehicle is not used, the driver is warned as soon as the vehicle is next driven.

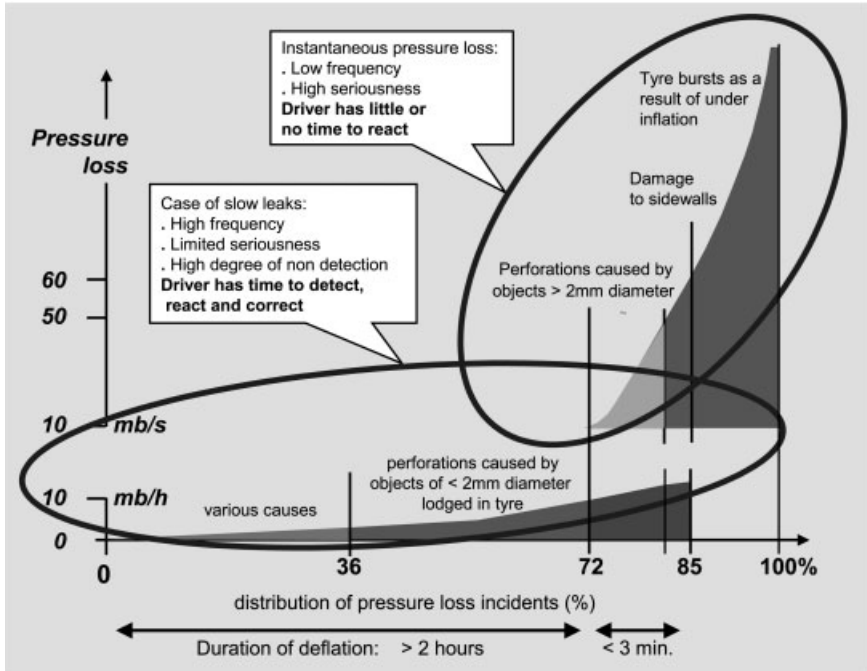


Fig. 7.20.8 Accidental pressure losses

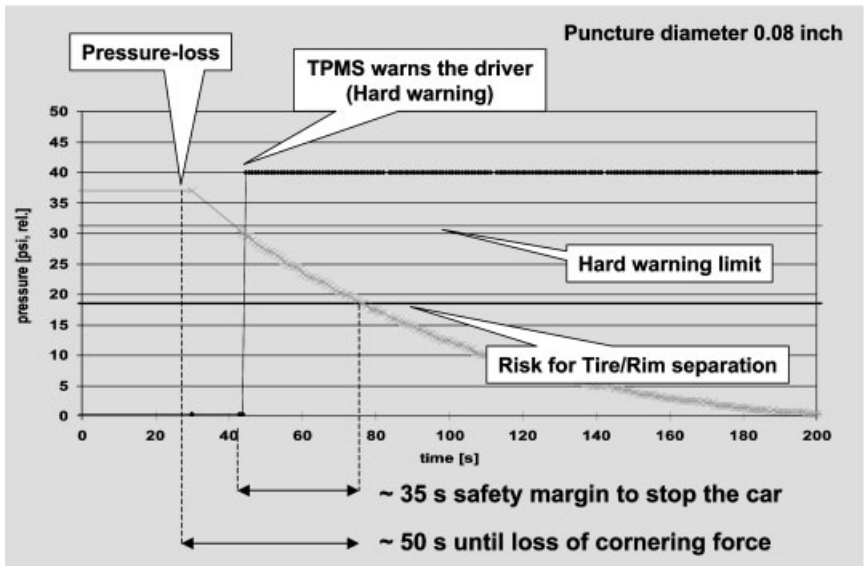


Fig. 7.20.9 Dynamic response to extremely rapid pressure loss

7.20.3.5 Warning Limits

The isochoric relationship between pressure and temperature is used when computing the warning limits (Fig. 7.20.10). The same warning effect can thus be achieved over a broad temperature range. With the aid of the nominal pressure and the temperature when calibrated, the system calculates the limits valid for the actual tire temperature. If they are exceeded, the control unit transmits signals to the driver information system according to the degree of pressure change.

A distinction is made between two types of low-pressure warning: ‘reminder’ (a ‘soft’ warning) and ‘failure’ (a ‘hard’ warning) as shown in Fig. 7.20.11. In the case of ‘reminder’ warnings, the vehicle remains safe to drive. While still driving or when the ignition is next switched on, the driver is reminded to correct the tire pressure at the next opportunity. In the case of ‘failure’ warnings, the vehicle may not be safe to drive. The driver is instructed to stop immediately and to check the condition of the tire. It is then up to the driver to decide if it is possible to continue driving at low speed or to change the wheel or to repair the tire with a sealant.

7.20.4

Future of Tire-Pressure Monitoring Systems

Tire pressure is an important factor for road safety, tire life, comfort, and the fuel consumption of the vehicle. Despite the evident importance and manifold approaches, systems for monitoring of tire pressure are mostly only available in luxury cars. Future development will be characterized by further innovation of the sensor technique. Higher integrated and more intelligent sensor components will

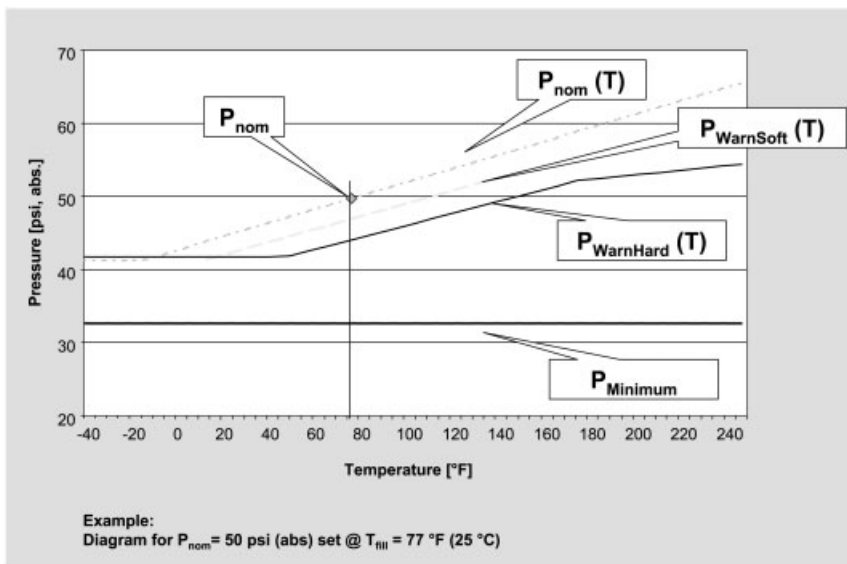


Fig. 7.20.10 Influence of pressure against temperature

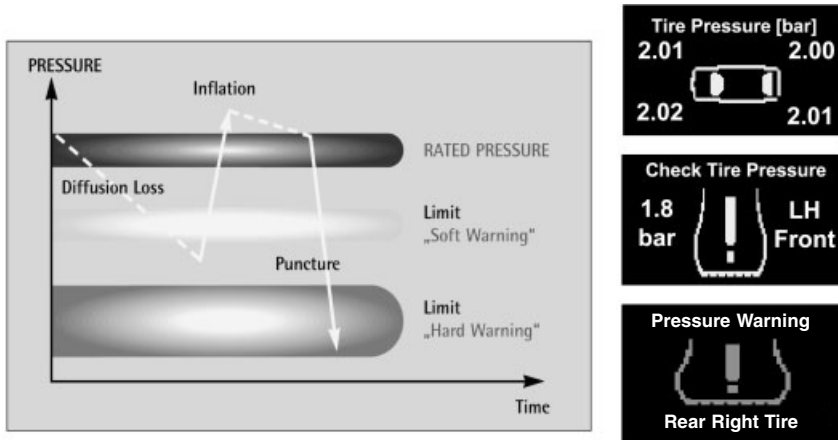


Fig. 7.20.11 System function: warnings

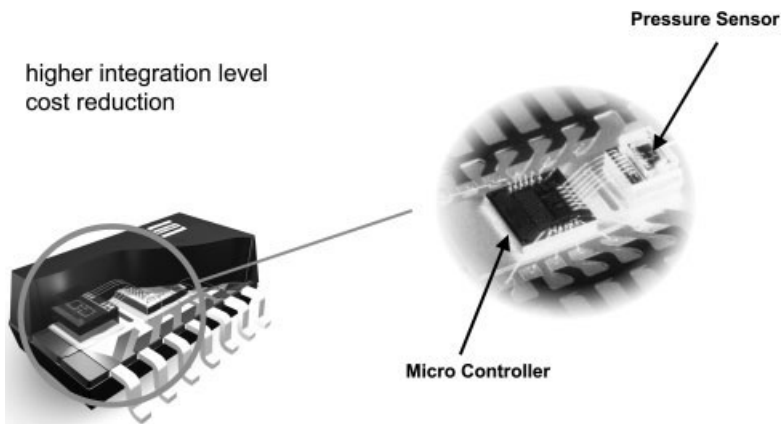


Fig. 7.20.12 Next generation technology

enable low-cost wheel electronics in combination with further improved functionality (Fig. 7.20.12).

The integration of the control unit into other car systems, ABS/ESP for example, will make it possible to reduce the system costs essentially and new safety features can be realized by the tight linkage of the system functions at the same time. Basic functionality of tire pressure monitoring with a significant reduction of the number of components needed can be reached by the use of the control unit and RF technique of the electronic car locking system to receive and process the data from the wheel electronics. Therefore the benefits of the tire-pressure monitoring systems in terms of safety, economy, and comfort will be available for drivers of the volume models as well in the near future. The introduction of tire pressure-monitoring systems for trucks will be an important contribution to improve safety on our roads.

Index

- ABAQUS 61
- acceleration sensors 12, 119 f, 301
 - torque 457
- accelerometers 93, 253, 269 ff
- acceptors 465
- accuracy 21, 82
 - accelerometers 283
 - chemical sensors 504
 - crankle angle sensors 419
 - electric current sensors 529
 - evaluation circuits 263
 - high pressure sensors 338
 - platinum sensors 350
 - pressure sensors 317
 - surface micromachining 123
 - tire pressure sensors 539
- activation 170
- active body control (ABC) 428,
- active front & rear steering (AFS) 429, 444
- active pixel sensor (APS) 392
- active wheel speed sensors 404
- AD converter (ADC) 348, 430
- adaptive cruise control (ACC) 2, 16
 - radar sensors 372, 379, 382
 - steering-angle sensors 428
 - stop&go 382, 401
 - video sensors 386
- adhesion 145
- adhesive bonding 84
- adhesives, fluorine-based 326
- adjustable sources, TCS/TCO 261
- ADXL 50 269
- aftertreatment 482, 501
- Agilent's advanced design system (ADS) 59, 69
- aging 496
- air bridges 241
- air conditioner 314
- air flow metering 360 ff
- air quality sensors (AQS) 509–515
- air/fuel (A/F) ratio control 480–500
- airbag passenger protection 2–9, 269, 383
- airgap deflection, wheel speed 415
- Al₂Au phase 137
- ALGOR 61
- alignment tolerances 95
- alkaline solution 285
- allowed tolerance band (ATB) 50
- alumina 165
- aluminum 154, 347, 354
- ammonia 507
- Ampere law 29
- amperometric limiting current oxygen sensor 491
- amplifiers
 - crankle angle sensors 421
 - evaluation circuits 247, 260 ff
 - force sensors 451
 - mass-flow sensors 366
 - wheel speed sensors 406 f
 - yaw rate sensors 302
- analog evaluation circuit 259
- analog hardware description language (AHDL) 59
- analog output air flow meter 366 f
- analog signal processing 259, 461
- analog steering-angle sensors 429
- analytical modeling 46
- anchoring 94 f, 97 f, 105
- angle hysteresis, wheel speed sensors 414
- angle sensors
 - AMR 177 f
 - GMR 184
 - market 15
- angular displacement, torque sensors 453
- angular momentum conservation 46
- angular rate sensors 45, 51, 297–314
 - market 13
 - packaging 203
 - rotationally vibrating 43
 - surface micromachining 119
- angular response, light sensors 467, 471
- anisotropic etching 74, 103, 110
 - pressure sensors 320
- anisotropic magnetoresistive effect (AMR) 173 ff, 438, 454, 533
- annealing 98, 109, 126, 150
- anodic bonding 82, 324
- anomalies 233
- Ansoff's HFFS 61, 69
- ANSYS 61
- antenna 539
- antiblooming 389
- antilock braking system (ABS)
 - magnetic field sensors 172
 - market 2, 12
 - steering-angle sensors 428, 444
 - wheel speed sensors 403
- antistiction coating 99

- antitheft system, market 12
- applications 1, 269–296
- applied wave research's (AWR's) microwave office 59
- AQS Mk IV 514
- aqueous alkaline 74
- Arrhenius model 218
- ASIC
 - evaluation circuits 264
 - exhaust gas sensors 187
 - high-pressure sensors 335, 340
 - radar sensors 377
 - reliability 224
 - torque sensors 454
 - wheel speed sensors 410
- assembly 191–203
 - electric current sensors 528
 - oxygen sensors 495
 - phases 206 ff
 - process 288
- ATHENA 79
- atmosphere, temperature sensors 352
- atmospheric pressure 13
- automatic distance control 297
- automatic gain control (AGC) 45
- average annual growth rate (CAGR) 5

- back annotation 43
- backend processing 288
- backflow detection 361
- backlap/polish 85
- ball in tube type sensors 269
- bandgap, platinum 155
- bandpass filter 279
- bandwidth 33, 247, 284
- basic functions 21–38
- batch processing 124, 194, 305
- batteries 516, 526
- bending mode, gyros 47
- benzoles 511
- BESOI wafers 84
- Bessel function 278
- biasing 248
- binary lambda signal 504
- biodiesel 519
- biometric sensors 18
- bipolar and CMOS 94 ff
- bit synchronous interface (BSI) 267
- blended fuels 518
- blind spot detection 382
- block diagrams
 - amplification circuit 421
 - analog signal processing 259
 - camera system 387
 - capacitive pressure sensors 314
 - exhaust gas sensor 352
 - ferroMRE sensor 417
 - FMCW radar sensor 375
 - Hall effect sensor 417
 - magnetorestrictive sensor 439
 - mass-flow sensors 363
 - MPU sensor 418
 - NO_x sensor 503
 - oil conditioning 521
 - pulse radar 378
 - rain sensors 475
 - steering-angle sensors 430, 441
 - temperature sensors 346
 - tire pressure sensors 540
 - torque sensors 453
 - wave shaping circuit 423
 - wheel speed sensors 404, 407
 - yaw rate sensors 301
- blood cells 511
- blooming 389, 393
- board mounting 282
- body control damping (BCD) 428, 444
- boiling point, brake fluid 525
- Boltzmann constant 163, 251, 272
- bonding 73 ff, 82 f
 - adhesive 84
 - aluminum layers 154
 - anodic 82
 - eutetic 84
 - fusion 83
 - pressure sensors 324
 - thermocompression 84
- boron doping 76, 130
- borosilicate glass 82
- Bosch 77 Ghz sensor 376
- Bosch deep reactive ion etching 111
- Bosch DF11i 415
- Bosch microsilicon sensor 361
- Bosch MM2 rate gyro 44
- Bosch RSM3ALS 474
- Bosch trench process 142
- Bosch video system 396
- Bosch XY lateral accelerometer 270
- boundary element method (BEM) 41, 61
- brake fluid 525
- brake pressure 13
- braking request 333
- brass 347, 354
- breakdown 206, 210
- breakdown voltage 127, 292

- bridge offset, temperature coefficient (TCO) 256 ff
- bridge supply voltage 260
- Brownian noise 253, 272
- Bryan factor 311
- buckling 109, 115, 143
- buffered oxide etch (BOE) 113
- build-in self test function (BITE) 225, 310
- bulk conductors 163
- bulk micromachining 25, 73–92
 - accelerometers 285
 - vibrating system 305
- buried feed throughs 85
- buried polysilicon 116 ff
- burst pressure 335
- business models 7
- bypass shape, mass-flow sensors 366

- cable layout, wheel speed sensors 411
- Cadence’s analog environment 59
- calcium 160
- calibration 262
 - electric current sensors 528
 - force sensors 450
 - mass-flow sensors 368
 - oxygen sensors 498
- cam sensors 416–427
- camera system 387
- cancellation algorithm 256
- cantilevers 104 ff, 143
- cap 200
- capacitance changes 30
- capacitance element 276
- capacitance-voltage converter 276, 281, 301
- capacitive accelerometer 88
- capacitive bulk machining 285
- capacitive coupling 112
- capacitive evaluation circuits 237 ff
- capacitive pressure sensors 314
- capacitive steering-angle sensors 429, 443 f
- capacitor 389, 423
- carbon disulfides 512
- carbon monoxide 500, 512
- carcinogenic benzoles 511
- carrier mobility 189
- catalytic converters 509
- cavities 85 f
- cell phone radiation 191
- ceramic capacitance 315
- ceramic catalysts 163
- ceramic materials 159–172, 361
- ceramic modules 196

- CFD ACE+ coventor ware analysis 61
- change oil indication 521
- charge carriers 129, 388
- charge coupled device (CCD) 389 ff, 435
- chassis systems 18
- chemical failure 210, 218
- chemical input signals 28
- chemical sensors
 - air quality 509–515
 - emission control 500–509
 - exhaust gas 480–500
 - liquid media 516–526
- chemical vapor deposited oxides/nitrides 146 ff
- chemiluminescence 29, 506
- chemimechanical polishing 108
- chip-on-board 194
- chopper stabilization 248
- circuit boards 199, 340
- circuits
 - active sensors 406
 - evaluation 237–267
 - simulators 59
- clamped structures 82, 143
- cleaning 7, 83, 126
- climate control 2, 13, 343
- closed-loop Hall sensors 529, 532 f
- closed-loop lambda control 484 f
- coatings
 - antistiction 99
 - protective 80
 - resist 139
 - temperature sensors 347
 - thimble type sensors 167
- cobalt 343
- CoCrPt layer 184
- code track, steering-angle sensors 433 f
- coercitivity 176
- Coffin–Manson model 218
- cofired zirconia-based sensors 169
- cold start strategies, exhaust gas 482
- collision avoidance system 372, 403
- collosal magnetoresistance (CMR) 186
- color imagers 390
- columnar growth 108
- comb interfaces 244
- combustion control 314
- comfort systems 1, 9
- common mode rejection ratio (CMRR) 247
- common rail system pressure sensors 328
- communication interfaces 266
- complementary metal oxide semiconductor (CMOS)
 - pressure sensors 322

- steering angle sensors 435
- surface machining 93
- video sensors 390 ff
- complementary parallel plate interfaces 242
- computer aided design (CAD) 58 ff
- concept phase 206
- conditioning 170
- conducting layers 136
- conductivity 517
- conductors 85, 347
- connectors, wheel speed sensors 411
- conservation laws 23
- construction guidelines 22
- contactless magnetic current sensors 527
- contacts 136, 176
- continuous wiping, rain sensors 478
- control area network (CAN)
 - chemical sensors 508
 - interfaces 267
 - steering-angle sensors 430
 - torque sensors 461
 - yaw rate sensors 307
- control systems
 - market 12
 - radar sensors 381
 - tire pressure sensors 539
 - torque sensors 460
- conversion
 - in/output signals 26 f, 32 f
 - optoelectronic 393
- copper 181, 247, 343
- Coriolis effect 45, 226, 298 f, 309 ff
- correlated double sampling (CDS) 248 ff
- corrosion 123, 210, 218
- COSMOS 61
- cost reduction 10, 191 ff, 338
- coupling, magnetic 180
- cracking 138
 - oxygen sensors 497
 - polysilicon layers 154
- crankle angle sensors 416–427
- crash detection accelerometers 269, 280
- cross sensitivity 33, 505
- cryogenic temperatures 111
- crystal lattice 160
- crystal orientations, polysilicon 130
- CuBe spring 269
- CuNi, thin films 133
- Curie–Weiss law 29
- current mirror, light sensors 468
- current trends 9
- customer requirements 192
- cycling 290
- cyclohexane 114, 519
- cyclotron resonance mode 112
- dam processing, pressure sensors 330
- damages, accelerometers 284
- damping
 - coefficients 33
 - g-cell 278
 - torque sensors 457
- data recording, tire pressure sensors 539 f
- daylight running lamp (DRL) 467
- deep-reactive ion-etching (DRIE) 110, 270
- defects 127, 210, 233
- deflection grids, mass-flow sensors 370
- deformations 126
- degradation, LEDs 432
- DENSO 317
- depletion 79
- deposition techniques 137
- deposition temperatures 107
- design
 - flow analysis 61 f
 - magnetic field sensors 176
 - methodology 1, 39–72
- detection gap, crankle angle sensors 419
- detection mode, gyros 47
- detection unit 298
- diamonds 156
- diaphragms
 - high-pressure sensors 335
 - mass-flow sensors 361
 - pressure sensors 314, 319
- dielectric properties 127
- diesel 519
- diesel direct injection 333
- diffused conductors 85
- diffused resistors 314
- diffusion coefficient, exhaust gas 492
- digging 73
- digital evaluation circuit 264
- digital output air-flow meter 368 f
- digital serial interfaces 461
- digital signal processors (DSPs) 59, 322
- digital steering-angle sensors 429, 443
- diode zapping 262
- direct crossing 85
- direct wiping, rain sensors 477
- direction detection, wheel speed sensors 415
- discrete surface micromachining 103–123
- displacement sensors 237
- distortion 278, 324
- distributor-less ignition (DLI) 416

- disturbances, electric 366
- domains, magnetic field sensors 175
- donors, light sensors 465
- doping 108, 117, 130
- doping, polysilicon layers 154
- Doppler effect 374, 378
- drifting 210
- drilling 81
- drive mode, gyros 47
- drive unit 298
- driver assistance system 2, 373
- drivetrain control 9, 17
- drop shock, accelerators 290
- dry etching 79 f, 96
- dry processes 274
- dual-cell wide band A/F ratio sensors 494
- dual-direction light sensor 471
- dual-sensor geometry 534
- durability
 - high-pressure sensors 341
 - oxygen sensors 496
 - wheel speed sensors 405
- dust particles 7, 233
- dynamic airgap deflection 415
- dynamic mass, torque sensors 458
- dynamic range, video sensors 393

- EDP* etchant 74, 110
- EFL-MAGIC 420
- elastic behavior, steel 452
- ELDO 59
- electric current sensors 18, 527–535
- electric function, Bosch air flow meter 367
- electrical circuits, wheel speed sensors 405
- electrical connections 142
- electrical evaluation 99
- electrical failure 210, 218
- electrical input signals 28
- electrical liquid properties 517
- electrical noise 44, 322
- electrochemical etch stop 76
- electrode reactions, ion conductors 160
- electrodes, layer coatings 165 ff
- electrohydraulic brakes (EHB) 333
- electrolyte density 525
- electromagnetic compatibility (EMC)
 - crankle angle sensors 419
 - electric current sensors 527
 - light sensors 463
 - pressure sensors 322
- electromagnetic interferences (EMI)
 - accelerators 276, 290
 - packaging 191, 195
- electromagnetic simulation 69
- electromotive force (EMF) 502
- electron hole pairs 388, 466
- electron hopping 186
- electronic control unit (ECU) 3, 21, 191 ff, 198
 - chemical sensors 508
 - oxygen sensor 486
 - packaging 191 ff, 198
 - pressure sensors 318
 - radar sensors 379
 - rain sensors 474
 - temperature sensors 348
 - wheel speed sensors 403
- electronic noise 250
- electronic signal detection 393 f
- electronic stability program (ESP) 2, 8, 11
 - accelerometers 297
 - high-pressure sensors 333
 - radar sensors 379
 - steering-angle sensors 428, 444 ff
- emission control, chemical sensors 500–509
- emission limits 482
- emission requirements 343
- encapsulation processes 330
- energy conservation 3
- engine combustion 17
- engine managements 1
- engine oil 516
- engine speed 14
- entropy 23
- environmental conditions 7 f
 - high-pressure sensors 334
 - oxygen sensors 496
- epipolysilicon 150 ff, 270 f, 285 f
- epitaxial polysilicon 107, 117
- epithickness 52
- epoxies temperature sensors 347
- EPROM 262, 322
- errors 33 ff
 - magnetic field sensors 177
 - pressure sensors 321
 - tire pressure sensors 536
 - yaw rate sensors 312
- etch stop techniques 75 ff, 112
- etching 73 ff, 105, 110 f, 139
 - anisotropic 320
 - mass-flow sensors 361
 - pressure sensors 320
 - wet 103, 110
- ethanol, liquid media sensors 518
- Ettinghausen effect 29
- Euler buckling 109, 115

- EURO4 343
- eutetic bonding 84
- evaluation circuits 237–267
- evaluation detection, light sensors 467
- evaporation 135
- exhaust gas sensors 3
 - ceramic materials 159–172
 - market 15
 - oxygen 480–500
- exhaust gas temperature 343, 352
- extensive signals 24 f
- extracting parameters 229
- Eyring model 218

- fabrication, accelerometers 285
- failure 1
 - force sensors 451
 - mechanical 82
 - oxygen sensor 486
 - packaging 191 ff, 194
 - reliability 208 ff
 - surface micromachining 99
 - tire pressure sensors 536
 - wheel speed sensors 406 ff
- fan speed control, temperature sensors 343
- far infrared (FIR) sensors 386
- Faraday constant 161, 492
- Faraday effect 29
- fault tolerant network (FTU) 444
- feed throughs 85
- FEMAPP 420
- ferromagnetoresistive effect 416 ff
- fiber optic gyroscope 298
- FIDAP 61
- field interline transfer CCDs 390
- fill factor, video sensors 393
- films
 - crankle angle sensors 417
 - deposition 95
 - etching 99
 - thin *see*: thin films
- filters 15
 - accelerometers 276 f
 - air quality sensors 512
- finite element method (FEM) 41, 61, 68, 242, 318, 335
- FIRE 61
- firing techniques 170 f
- flake off, protective layers 497
- flex fuel, liquid media sensors 518
- flexibility, air quality sensors 513
- Flexray 444

- flicker noise 251 ff, 394
- fluorine-based adhesives 326
- fluorine-based plasma processes 110
- fluorine compounds 80
- fogging effect 511
- force sensors 450–461
- forward intelligent lightening 429
- Foucault pendulum 298, 311
- fracture strength 143 f, 153
- fracture toughness 165, 206
- frame transfer CCDs 390
- frequency-modulated continuous wave radar (FMCW) 375
- frequency ranges, accelerometers 283
- frequency shifts 30
 - interfaces 238
 - radar sensors 374
 - yaw rate sensors 302
- frequency transfer function 32
- Fresnel lens principle 470
- friction 29, 451
- fuel 7, 516
- fuel level sensors 525
- fuel pressure sensor 13, 324, 327
- full scale deflection (FSD) 317
- full scale pressure 333
- full span 258
- fully-stabilized zirconia (FSZ) 162
- function affecting factors, oxygen sensors 497
- function analysis systems technique (FAST) 431
- functional layers 104, 142
- functional limits, wheel speed sensors 414 f
- functional parameters 49, 227 ff
- functionality 10, 205, 395
- fusion bonding 19, 83

- g cell 271, 276 ff, 282 f
- gages 128 ff, 264
 - pressure sensors 314, 318, 329
- gallium arsenide 187
- Galvanic isolation 529
- gas plasmas 79
- gas reactions 161
- gas sensors 18
 - see also*: exhaust gas sensors
- gasket, temperature sensors 347, 354
- gasoline 512, 518
- gasoline direct injection (GDI) 333, 483
- gasoline leak 314
- Gaussian behavior, process parameters 50
- gear box oil 516

- gear box temperature 343
 - gear lever position 5
 - gear shaft speed 14
 - geographical position system (GPS) 297
 - geometry level 41
 - geometry parameters testing 229
 - germanium layers 156
 - giant magnetoresistance (GMR) 179 ff, 534
 - glass
 - etchants 81
 - mass-flow sensors 361
 - micromachining 81 f
 - pressure sensors 324
 - undoped 97
 - Golay cell 28
 - gold layers 137, 156, 165
 - gradiometer 184
 - grain properties 100, 130 f
 - granular systems 186
 - gravitational acceleration 7
 - Gray code technique 435, 440
 - grids, mass-flow sensors 370
 - ground plate polysilicon 97
 - growth
 - columnar 108
 - epitaxial 117
 - silicon oxides 146
 - Gunn oscillator 376
 - gyros 44 f, 227
 - gyroscope 109, 250, 297–314
-
- H bridge circuit 264
 - Hall effect sensors
 - electric current 529
 - magnetic field 186
 - steering-angle 439
 - wheel speed 417, 424
 - Hall plate 28
 - Hall transducers 404
 - Hall/AMR technology 3
 - handling 195
 - hardening 126
 - HDRC imagers 395 f
 - head-up display (HUD) 467
 - headlight leveling, market 12
 - heat conduction filler 347
 - heaters 169, 366
 - hemoglobin 511
 - heptanes 512
 - hermetic glass 327
 - hermetic packages 194
 - hexanes 512
 - high-boron dose etch stop 76
 - high-frequency electromagnetic simulation 69
 - high-pressure sensors 53 ff, 333–342
 - high-temperature leakage, crankle angle sensors 422
 - high-volume applications 142
 - hill-hold control 403
 - Hitachi pressure sensor 87
 - HNA etchants 74
 - housing 191, 203
 - accelerometers 282
 - liquid media sensors 522
 - mass-flow sensors 365
 - pressure sensors 327
 - temperature sensors 347, 352–359
 - yaw rate sensors 307
 - humidity 218, 290, 335
 - HVAG 9, 11
 - hybrid electronic packaging 81
 - hybrids, high-pressure sensors 340
 - hydraulic mounting interface 334
 - hydrocarbons 501 ff, 510 ff
 - hydrofluoric acid 98, 114
 - hydrofluorocarbons 111
 - hydrogen 352
 - hydrogen impurities 147
 - hydrogen sulfides 512
 - hydrolysis, silane 107
 - hygroscopic brake fluids 525
 - hysteresis 34, 450
-
- imagers 390 ff, 397 f
 - implantation mask 78
 - impurity, light sensors 465
 - in-band overload 279
 - In Sb/As, magnetic field sensors 187
 - incremental track 433
 - inductance changes 30
 - induction, electromagnetic 417
 - inductive steering-angle sensors 429, 440 ff
 - inductive torque sensors 456
 - inertial sensor inputs 26
 - influence strengths 49 f
 - infrared range 374, 387
 - inprocess measurements 54
 - inprocess striction 114, 273 ff
 - input circuit, wheel speed sensors 406
 - input signals 24 ff
 - insertion loss 66
 - installation
 - crankle angle sensors 416
 - electric current sensors 528

- mass-flow sensors 365
- oxygen sensors 498
- wheel speed sensors 411
- integrated circuit (IC) processes 92, 103 f, 207
 - electric current sensors 533
 - testing 225
- integrated surface micromachining 92–106
- IntelliSuite 61
- intensive signals 24 f
- interaction points 23
- interconnection phase 206
- interfaces
 - air quality sensors 509
 - evaluated circuits 238 ff
 - silicon/glass 85
 - steering-angle sensors 448
 - torque sensors 460
- interfering factors 29, 33, 82
- interline transfer CCDs 390
- intermittent wiping, rain sensors 478
- intrinsic stress, layers 147 ff
- ion-beam sputtering 182
- ion-conductors 160
- ion-implantation 108
- iron 343
- islands 76
- isolation layer 126
- isotropic etching 74

- jig processing 124 f, 139
- jitter, wheel speed sensors 404
- junctions 76 f

- Kirkendall effect 137
- KOH 320

- lambda sensor 9
 - ceramic materials 159 ff
 - exhaust gas systems 480 ff, 502
 - steering angle sensors 428
 - see also*: oxygen sensor
- lane-departure warning (LDW) 399
- lane detection 397 ff
- Laplace transform 33
- laser cutting 262
- laser drilling 81
- laser welding 201
- lateral combs 244
- layers
 - angular rate sensors 52 ff
 - bulk machining 73 ff
 - conducting 136
 - multiple 103
 - surface machining 103 ff
 - surface machining 97 f
 - thimble type sensors 167
- lead 347
- leakage
 - crankle angle sensors 422
 - diodes 78
 - evaluation circuits 248
 - high-pressure sensors 335
- lean-burn engines 483
- LIDAR 2, 373
- lifetime
 - air quality sensors 513
 - high-pressure sensors 338
 - oxygen sensors 497
 - reliability 207 f, 212 f
 - tire pressure sensors 538
- light emitting diodes 432, 474
- light receiving diode (LRD) 474
- light sensors 462–473
- limiting current oxygen sensor 491
- linear lambda control 485
- linear vibrating system, quartz 303
- liquid media conditions 17
- liquid media sensors 516–526
- lithium batteries 528
- lithography 105, 117
- local interconnect network (LIN) 267
- long range radar system 373, 375 f
- long term stability 123
- Lorentz force 186 f
- losses 66
- low-emission vehicles (LEV) 481
- low jitter, wheel speed sensors 404
- low-pass filters 302
- low-pressure chemical vapor deposition (LPCVD) 94, 97
 - layers 148 ff
 - surface machining 107, 115 ff
- lubricant aging, force sensors 451
- luminance 400
- luminescence 29

- MAFIA 61
- magnetic-current sensors 528, 533
- magnetic-field modulation wheel speed sensors 409
- magnetic-field sensor technologies 172–190
- magnetic hysteresis 426

- magnetic input signals 28
- magnetic pick-up coil sensors (MPU) 417
- magnetic RAMs 186
- magnetic steering-angle sensors 429, 438
- magnetoresistive effect (MRE) 416
- magnetoresistive sensors, AMR 17
- magnetoresistors 188
- magnetostriction 176
- magnetron sputtering 135
- manganese 343
- manufacturability 192
- manufacturing 1, 7
 - ceramic materials 166
 - defects 210
 - GMR sensors 182
 - magnetic field sensors 176
 - thin films 124
- markets 1, 5–20, 192
- masking, protective 95
- mass conservation 23
- mass flow 26
- mass flow sensors 360–371
 - market 13
 - testing 231
- mass loading, liquid media sensors 523
- mass-spring dashpot model 271, 282
- MAST 59
- MATLAB 59
- maturity 192
- mean time to failure (MTTF) 204, 213
- measurement methods 21–38, 143
- mechanical failure 82, 210, 217
- mechanical input signals 28
- mechanical interfaces 448
- mechanical liquid properties 517
- mechanical noise 44
- mechanical overload 191, 290
- mechanical overtravel stops 284
- mechanical strength 145
- mechanical stress 496
- mechatronic systems 40
- media-compatible pressure sensors 329
- membranes 54, 142, 206
- metal catalysts 165
- metal contacts 136
- metal housing
 - pressure sensors 327
 - temperature sensors 357
- metal oxide semiconductor (MOS) 389
- metal packages 197
- metallization 176
- methanol 518
- microcontroller, rain sensors 477
- microelectromechanical systems (MEMS)
 - 10, 39
 - failure 204 ff, 269
 - surface micromachining 92 f, 103
- microfused strain gages 339
- micromachined rate sensing gyros 45
- micromachining 3
 - bulk 73–92
 - glass 81 f
- micropackaging 81
- microsilicon elements 361
- microspring 86
- microstructure stabilization 98
- microsystems (MST) 10
- MIL failure, oxygen sensor 486
- miniaturization 10, 25 ff, 338
- mobility
 - charge carriers 394
 - light sensors 465
 - thin films 129
- modeling 41 ff
 - analytical 46
 - MEMS 58 f
- modulating transducers 29
- modulation, circuits 248
- modules 27, 195
- molded integrated circuit (MIC) 424
- momentum transfer 29
- monochrome imagers 390
- monolithic microwave ICs (MMICs) 17
- monolithic three axis accelerometer 253
- monostatic lens antenna 376
- Monte Carlo methods 50, 55
- MOSFET resistors 249
- motor control 17
- Motorola Z axis accelerometer 270
- mounting
 - light sensors 469
 - mass-flow sensors 363
 - rain sensors 479
 - tire pressure sensors 538
- mucous membrane damage 510
- multi-functional oil conditioning 522
- multilayers 103, 168 ff, 180
- multiple sensors 19
- Nasicon 162
- navigation systems 2, 172
 - accelerometers 297
 - steering-angle sensors 428
- near infrared (NIR) radiation 399
- Néel temperature 185

- negative thermal coefficient resistors (NTRs) 343
- Nernst equation 161, 502
- Nernst oxygen sensor 487 f, 493 f
- network level 42
- network rain sensors 475
- network type analysis 50
- networked satellite sensors 290
- Newton law 26, 272
- nickel 156, 343
- nickel-iron thin films
 - crankle angle sensors 417 ff
 - electric current sensors 533
- NiCr thin films 124–136, 337
- night vision systems 17, 399
- nitride removal 98
- nitrogen 147, 352
- noise 44
 - accelerometers 283
 - crankle angle sensors 421
 - evaluation circuits 250 ff
 - pressure sensors 322
 - video sensors 393 f
- noise aliasing 249
- nominal pressure input 541
- nonlinearity 34
 - sensitivity 321
 - temperature coefficients 49
- nonsilicon-based high-pressure sensors 53
- nonstoichiometric lambda control 491
- notching effects 112
- npn* bipolar transistor 78
- nucleation rate, polysilicon 101
- Nyquist band 264

- object detection 401
- octanes 512
- offset 34, 49
 - accelerometers 283
 - crankle angle sensors 422
 - high-pressure sensors 338
 - mass-flow sensors 366
 - pressure sensors 322
- oil 7
 - conditioning 17, 203, 521
 - level sensors 525
 - liquid media sensors 516
 - pressure 13
- oil-filled high-pressure sensors 341
- on-board diagnosis (OBD) 1
 - chemical sensors 486, 508
 - temperature sensors 343
- yaw rate sensors 314
- open-loop Hall sensors 529 f
- operating conditions
 - force sensors 450, 457 f
 - oxygen sensor 481 ff
- operating point (OP) 51
- operating principle, bulk conductors 163
- operating temperature
 - accelerators 290
 - crankle angle sensors 419
 - light sensors 463
 - tire pressure sensors 539
- operation mode, rain sensors 474
- optical detection range, video sensors 395
- optical filters 28
- optical steering-angle sensors 429, 432
- optical torque sensors 454
- optimization, gyros 47, 51
- optoelectric conversion function 393
- orange peel coupling 181
- ordinary differential equations (ODEs) 60
- original equipment manufacturers (OEMs) 7
- outband overload 279
- output signals 32 ff
- overloads 279, 450 f
- overtravel stops 273
- oxide etch, buffered 113
- oxide removal 97 f
- oxide thickness 52 f
- oxide zapping 262, 265
- oxygen 352
- oxygen sensor 428, 480–500
 - zirconia-based 159 ff
- *see also*: lambda sensors, chemical sensors
- ozone 7

- packaging 81, 191–203, 208
 - accelerators 273 ff, 282 f
 - ceramics 188
 - crankle angle sensors 424
 - electric current sensors 528
 - high-pressure sensors 340
 - light sensors 462, 469 f
 - mass-flow sensors 363 f
 - pressure sensors 318, 323 f, 325 f
 - processes 26
 - yaw rate sensors 310
- palladium 137, 165
- parallel plate interfaces 238, 242
- parameter testing 227 ff
- parasitics 82, 238 ff
 - force sensors 451

- parking aid 382
- partially stabilized zirconia (PSZ) 161
- parylene 274
- passivation 76, 96, 111
- passivation layer 137, 142
- passive wheel speed sensors 404
- patterned polymeric dam processing 330
- patterning 73, 105, 138
- PDE solvers 61
- Peltier effect 29
- performance
 - gyros 52
 - magnetic field sensors 176
- peripheral acceleration sensors (PAS) 267
- permalloy 174, 533
- permittivity, liquid media sensors 517 ff
- perovskite structure 164
- phase-locked loop (PPL) 45
- phase sensors, market 14
- phases, polysilicon 100
- phosphorous silicon glass (PSG) 97
- phosphorus doping 108, 154
- photocharge generation 388
- photocurrent 388
- photodiodes 389 ff, 463
- photolithography 73, 115, 138
- photoluminescence 29
- photoresist 96, 110, 116 f, 274
- photosensing 388 f, 435
- phthalocyanin sensor 510
- physical chemosensors 516
- picture frame design 319
- piezoelectric pressure sensors 314
- piezoelectricity 29, 128
- piezoresistive bulk machining 285
- piezoresistive pressure sensors 314
- piezoresistive Wheatstone bridge 256
- piezoresistive/electric interfaces 237, 240
- pinning 181, 184
- pixel rate 395
- placement 200
- planar sensors, layers 168
- planar zirconia oxygen sensor 489 f
- plasma deposition 135
- plasma enhanced chemical vapor deposition (PECVD)
 - layers 146 ff
 - surface micromachining 97
 - thin films 127, 133, 138
- plasma etching 79 f, 105, 110
- plastic behavior, steel 452
- plastic housing 347, 357
- plastic mold packages 195
- plated steel 347
- platforms 19
- plating 95
- platinum heaters 169
- platinum layers 155 ff
- platinum resistors 361
- platinum sensors 350 ff
- PLCC housing 310
- plug in module 364
- pn* junctions 76
- poisoning
 - liquid media sensors 516
 - oxygen sensors 496
- Poisson ratio 129
- polishing
 - chemimechanical 108
 - oil/water 126
- pollution 509
- polybutylene terephthalate (PET) 199
- polymeric dam processing patterned 330
- polysilicon
 - anodic bonding 85
 - buried 116
 - deposition 100 f
 - disk 43
 - epitaxial 107
 - high-pressure sensors 337
 - processing 100, 104 ff, 149 ff
 - surface machining 93 ff, 96 f
 - thin films 128, 130 ff
- position sensors 15
- potentiometric zirconia Nernst oxygen sensor 487 f
- powder mixtures, temperature sensors 343
- power consumption 22, 529
- power supply, wheel speed sensors 406
- power supply rejection ratio (PSRR) 247
- powertrain, torque sensors 453
- preaging, high-pressure sensors 340
- precipitation hardening 126
- precrash roadmap, radar sensors 383
- pre mold packages 196
- press contacts, anodic bonding 85 f
- press fit 200
- press-on coupling, rain sensors 479
- pressure loss, tire pressure sensors 536
- pressure low warning 541
- pressure nonlinearity (Nlp) 321
- pressure sensors 314–332
 - Hitachi 87
 - market 13
- pressure-volume product 24
- printed circuit boards 199, 249

- printed heaters 169
- printed wiring board (PWB) 282
- probe mass, gyros 46
- process flow, surface machining 107 ff, 116
- process parameters 46, 49, 52
- process properties, MEMS sensors 42
- process temperatures 146
- processing, ceramic materials 166
- production tolerances 264, 300
- production volumes 22
- PROM array 265
- propane 506
- protective coatings 80
- protective layer, flake-off 497
- protective masking 95
- PSG films 427
- PSG sources 108
- PtIr thin films 133
- pull-in effects 241
- pulse radar 378
- pulse-width modulation (PWM) 410, 511
- pumped reference, zirconia oxygen sensor 490
- purple phase 137
- Pyrex 7740 82

- Q factor 59, 302
- quadrature signal 301
- quality 10, 192, 272
- quantization errors 34
- quantum efficiency 388, 393
- quantum mechanical effect 173
- quartz technology 98, 303

- RADAR** 2
- radar antenna 376
- radar sensors 372–385
 - market 16
- radiant input signals 28
- radiant noise 323
- radio frequency (RF) 59, 65 f
- radio frequency (RF) antenna 539
- rain sensor modulus including assistant light switch (RSM ALS) 474
- rain sensors 474–480
- random addressing freedom 392
- rapeseed oil methyl ester (RME) 519
- rate gyro, Bosch MM2 44
- reactive ion-etching (RIE) 79, 110
- readout circuits 253
- real-life conditions 219
- real-time pulses, wheel speed sensors 404
- receiver, rain sensors 477
- recrystallization 98, 132
- redundance 216
- reflectivity, light sensors 466
- refractive index, light sensors 466
- refresh rate, torque sensors 460
- refrigerant pressure 314
- regeneration phase, chemical sensors 507
- regional trends 11
- release cycles, switches 67
- release stiction 273
- reliability 22, 82, 191, 204–223, 290
- reset noise, video sensors 394
- resins, temperature sensors 347
- resist coating 39
- resistance 335
- resistance changes 30
- resistive evaluation circuits 256 ff
- resistive shunt, electric current sensors 528
- resistive steering-angle sensors 429
- resistive zirconia oxygen sensor 490
- resistivity 127, 135
 - platinum 155
 - polysilicon 131
- resistor values, NTCs 344
- resistors, metal 53
- resolution 237, 539
- resonance frequency 46
- resonances
 - accelerometers 272
 - torque sensors 458
- response times, mass-flow sensors 361
- return loss 66
- Reynolds number 364
- RFI accelerators 290
- rhodium 165
- rinse processes 274
- robustness 48
 - packaging 195
 - thin films 142
 - wheel speed sensors 405
- rod-shaped heaters 169
- roll-over mitigating 403
- roll-over protection 297
- rotational speed sensors 14, 403
- rotational vibrating system 309
- rotationally vibrating angular-rate sensors 43
- roughness 126, 154

- SABRE** 59
- sacrificial etch processes 274
- sacrificial layers 104 f, 116 ff, 142

- sacrificial oxide etch 113
- safety systems 1, 10, 224
- salt spray 7, 335
- scanning electronmicrograph (SEM) 43
- scattering 169 f
 - spin-dependent 180
- scratch-resistant covering 388
- sealing 200, 328
- search lines, lane detection 400
- seat position 15
- Seebeck effect 29
- seed layer 117
- selfcalibration algorithm, force sensors 450
- selftest, accelerometers 273
- semiconducting ceramics 163
- semiconducting zirconia oxygen sensor 490
- semiconductor manufacturers 8
- semiconductor processing 104
- semicustom trim/test systems 290
- sensing layer 128
- sensitivity 32
 - accelerometers 272, 283
 - chemical sensors 505
 - crankle angle sensors 418, 426
 - evaluation circuits 239
 - force sensors 450
 - light sensors 465
 - magnetic field sensors 176
 - pressure sensors 314
 - silicon 243
 - temperature coefficient 256 ff
 - video sensors 393 ff
 - wheel speed sensors 405
- SensoNor 88
- serial peripheral interface (SPI) 267
- series switch plates 68
- set speed control, radar sensors 379
- shadow mask technique 139
- shallow junctions 77
- shaping, thimble type sensors 166
- shifts 238
- shock type loads 7, 210, 217
- shocks 278, 284
- short-range radar sensing (SRR) 16, 377, 382
- shunt, resistive 528
- side impact sensors 290
- signal conditioner 275, 301, 321
- signal conversion 21, 32
- signal evaluation circuit 47
- signal extraction 21
- signal flow simulator 59
- signal formation 23
 - mass-flow sensors 366
 - temperature sensors 348, 353
 - wheel speed sensors 406 ff
- signal path 28
 - air flow meter 367
- signal processing 21, 24 ff
 - analog 259
 - crankle angle sensors 421
 - light sensors 467 ff
 - pressure sensors 318 ff
 - rain sensors 476
 - steering-angle sensors 430
 - torque sensors 460
- signal to noise ratio (SNR)
 - accelerometers 275, 280
 - evaluation circuits 254
 - pressure 319
 - video sensors 393
 - wheel speed sensors 405
 - yaw rate sensors 300
- silane 107, 127
 - thin films 146, 149
- silica, light sensors 466
- silicon 104 ff, 243
 - crankle angle sensors 417
 - magnetic field sensors 187
 - mass-flow sensors 361
 - micromachining 191 ff
 - pressure sensors 314, 324
 - thin films 124, 141–158
- silicon dam processes 331
- silicon dioxide 98, 388
- silicon etchants 74
- silicon fusion bonding (SFB) 83, 87
- silicon-germanium layers 156
- silicon Hall effect sensor 417, 424
- silicon micromachining 304
- silicon nitrides 148 ff, 388
- silicon on insulator (SOI)
 - elements 95
 - layers 154 ff
 - MEMS 270
 - wafers 83 ff
- silicon oxide 146 ff
- silicon oxynitrides 156
- silicon *pn* junction diode 463
- silicon RIE 142
- silicon wafers 25, 74
 - mass-flow sensors 361
 - yaw rate sensors 305
- SIMOX wafers 84
- simulation methods 41 ff
- Simulink 59
- single-cell wide range A/F ratio sensors 493

- single-chip, steel diaphragms 320
- single-chip systems 92
- single-crystal silicon
 - layers 155
 - pressure sensors 314, 328 f
- single-ended parallel plate interface 238
- smart cut 84
- SMD resistors 344
- smearing, rain sensors 479
- snap-in design 358
- sodium ion conductor 162
- solar sensor 462, 471
- solarimeter 28
- soldering 200, 347
- solution annealing 126
- spacers 181
- spark ignition engines 480 ff
- specifications
 - accelerometers 283
 - crankle angle sensors 419
 - light sensors 463
 - MEMS sensors 42
 - packaging 23, 195
 - yaw rate sensors 298
- Spectre 59
- speed sensor, GMR 183
- spin-dependent scattering 180
- spin orientation 173
- spin valves 181
- splashes, rain sensors 479
- spring rate, force sensors 451
- springs 206, 271
- sputtering 117, 176
- squeezing pixel pitch 393
- SQUID 188
- SSUPREME 79
- stability
 - high-pressure sensors 338
 - polysilicon layers 151
 - thermal 127
 - thin films 123, 135
- stabilization, microstructures 98
- stainless steel
 - 17-4 PH 125
 - temperature sensors 347, 352
 - thin films 123
 - wafers 124
- stand-alone rain sensors 475
- standard deviation 253
- standard input circuit, wheel speed sensors 406
- standard manufacture route, thimble type sensors 166
- standard processes, bonding 88
- standardization
 - input signals 27
 - packaging 192 f
- steel
 - diaphragms 335
 - temperature sensors 347
 - thin film technology 123–140
- steering, torque sensors 453
- steering angle 15
- steering angle sensors (SAS) 428–449
- steer-by-wire (SbW) 429
- stem, pressure sensors 326
- sticking 233
- stiction 114, 210, 273
- stoichiometric air/fuel control 487
- stop & go (roll), radar sensors 382
- strain gages
 - high-pressure sensors 336 ff
 - thin films 123 ff, 126 ff
- stray capacitor 250
- stray electric signals 238
- strength
 - mechanical 145
 - process-dependent 49 f
 - substrates 165
- stress, residual 206
- stress factors, oxygen sensors 496
- stress gradient 52, 109, 143
- stress-strain product 24
- stringy exhaust gas 498
- stripping 139
- structural materials, exhaust gas sensors 165
- structures, surface machining 94
- sulfur 352
- sulfur dioxides 512
- sumpoints 260
- super-ultra-low emission vehicles (SULEV) 481
- superposition, sensitivity/offset 34
- suppliers competition 192
- surface conductors 164
- surface micromachining 73, 92–123, 285
 - vibrating systems 308
- surface roughness, polysilicon layers 154
- surround sensing 374, 387
- suspension control, market 12
- suspension pressure sensors 314, 318
- switches 66
- switching hysteresis 414
- system diagnosis, tire pressure sensors 536
- system level, modeling 42
- system reliability 215
- system suppliers 8

- systematic errors 33
 - see also*: errors

- tape casting, ceramic 168
- target selection, radar sensors 380
- Taylor series 321
- Teflon type layers 111
- telemetric sensors 18
- temperature changes 34
- temperature coefficients 35, 49
 - resistivity 131, 155
 - sensitivity (TCS) 256
- temperature compensation circuit 320
- temperature-moles product 25
- temperature nonlinearities 321
- temperature ranges 3, 7
 - accelerometers 284, 290
 - cryogenic 111
 - electric current sensors 529
 - exhaust gas 159
 - high-pressure sensors 338
 - packaging 191
 - pressure sensors 317
 - reliability 204
 - silicon processes 146
 - surface micromachining 107
- temperature sensors 343–360
- temperature shock 191
- tensile strength, steel 125
- tensile stress, residual 98
- testing 224–236, 290
 - high-pressure sensors 341
 - oxygen sensors 498
 - wheel speed sensors 412 ff
- tetragonal stabilized zirconia (TSZ) 162
- tetrathylorthosilicate (TEOS) 117, 127, 146
- theory of inventive problem solving (TRIZ/TIPS) 431
- thermal cycling 217
- thermal expansion coefficient 127, 165, 324
- thermal failure 210, 217
- thermal input signals 28
- thermal liquid properties 517
- thermal material properties 231
- thermal noise 44, 251
 - video sensors 394
- thermal oxidation 117
- thermal oxides 146
- thermal processing, oxide layers 105
- thermal shock, accelerators 290
- thermal stability 127
- thermal stress, oxygen sensors 496
- thermocompression bonding 84
- thermocouples 28, 350
- thermodynamic equilibrium 23 ff
- thick films 167, 231
- thimble type sensors 166, 487 ff
- thin films 3
 - electric current sensors 533
 - magnetic field sensors 176
 - mass-flow sensors 361
 - metal resistors 53
 - on silicon 141–158
 - permalloys 174
 - pressure sensors 314
 - steel 123–142, 336
 - strain gages 336
 - thimble type sensors 167
- three axis accelerometer 253
- three-layer polysilicon surface 285
- three-omega technique 231 f
- three-phase boundary 161
- three-way catalysts (TWC) 484, 500
- throttle valve angle 15
- thyristor zapping 261
- tilt oscillation 309
- time-triggered protocol (TTP) 444
- tin film resistor 322
- TIPS 431
- tire pressure sensors 536–544
- titania 163
- tone wheels 411
- topology limitations, radar sensors 381
- torque sensors 450–461
- toxic emissions, chemical sensors 500, 510
- traction control, wheel speed sensors 403
- traffic sign recognition 397, 400 f
- transducers 3, 21–28, 45, 53
 - accelerometers 271 f
 - chemical 3
 - convective heat transfer 286
 - interfaces 238 ff
 - modulating 29
 - substrates 125
- transistor, *npn* bipolar 78
- transition metals, ferromagnetic 173
- transmission laser welding 201
- transmitter, rain sensors 476
- transmitter antenna 377
- transparent covering 388
- trapping 86
- trench edge loss 52
- trench etching 110 f, 118
- trenches 75
- trigger point 277

- trimming
 - accelerometers 290
 - evaluation circuits 262, 265
 - high-pressure sensors 337
 - packaging 192
 - pressure sensors 322
 - thin films 139 f
- TRIZ 431
- tuning fork, yaw rate sensors 304
- tunneling 237
- tunneling magnetoresistance (TMR) 186
- turbocharger pressure 13
- twilight sensors 462
- two step lambda control 484

- ultra low emission vehicles (ULEV) 2, 343, 481, 494
- ultrasonic drilling 81
- ultrasonic radar sensing 374
- undercutting 76
- unexcited states, gyros 47
- unibond wafers 84
- universal exhaust gas oxygen (UEGO) 502
- upfront sensors (UFS) 198
- utilization 192

- V grooves 75
- vacancies 160
- valves 181
- vapor etching 105
- vector flipping 175
- vehicle dynamics control (VDC) 172, 297, 312
- vehicle dynamics control systems 12
- vehicle interference, radar sensors 384
- velocity-time profile 277
- ventilation systems 509
- vibrating shell, yaw rate sensors 311
- vibrating system, quartz 303
- vibrations 7, 217, 335
- video camera 396
- video radar sensing 374
- video sensors 16, 386–402
- video system 396
- viscosity
 - air 364
 - liquid media sensors 518 ff
 - torque sensors 459
- vision systems, components 387
- voltages, accelerometers 284

- voltaic effect 29
- volume ramp-up 192
- VTI Hamlin 88

- wafer bonding 73 ff, 81 ff, 286
- wafer level package 274
- wafer level testing 226
- wafers, SOI/SIMOX 84
- warning limits, tire pressure sensors 543
- warning systems, radar sensors 384
- water 7
- water protection, mass-flow sensors 370
- wave shaping circuit, crankle angle sensors 422
- weather limitations, radar sensors 381
- Weibull plot 214, 221
- weight 22
- welding 201
- wet etching 73 f, 103, 110
- Wheatstone bridge 53
 - high-pressure sensors 337, 340
 - mass-flow sensors 362
 - piezoresistive 256
 - thin films 128, 135 f, 139
- wheel positions, tire pressure sensors 536
- wheel speed 14, 172
- wheel speed sensors (WSS) 403–416
- wide-range A/F ratio sensors 493 ff, 498
- wide-range lambda signals 504
- windshield 474 ff
- wiper systems 475 ff
- wire-wound resistors 351
- work flow, MEMS sensors 42 f
- working design, accelerometer 119

- X-by-wire applications, pressure sensors 331

- yaw rate sensors 297–314
- yield losses 194
- yields 55
- Young modulus 165, 206, 324
- yttrium 160

- zapping 261 ff
- zeolites 165
- zinc 343
- zirconia-based oxygen sensor 159 ff
- zirconia Nernst oxygen sensor 487 f