Sensors

A Comprehensive Survey

Edited by
W. Göpel, J. Hesse, J. N. Zemel

Volume 8

Micro- and Nano-sensor Technology/ Trends in Sensor Markets

Volume Editors
H. Meixner and
R. Jones

Sensors

Volume 8

Micro- and Nanosensor Technology/ Trends in Sensor Markets

Sensors

A Comprehensive Survey

Edited by
W. Göpel (Universität Tübingen, FRG)
J. Hesse (Zeiss, Oberkochen, FRG)
J. N. Zemel (University of Pennsylvania, Philadelphia, PA, USA)

Vol. 1 Fundamentals and General Aspects
(Volume Editors: T. Grandke, W. H. Ko)
Vol. 2/3 Chemical and Biochemical Sensors, Part I/II
(Volume Editors: W. Göpel, T. A. Jones †, M. Kleitz, I. Lundström, T. Seiyama)
Vol. 4 Thermal Sensors
(Volume Editors: T. Ricolfi, J. Scholz)
Vol. 5 Magnetic Sensors
(Volume Editors: R. Boll, K. J. Overshott)
Vol. 6 Optical Sensors
(Volume Editors: E. Wagner, R. Dändliker, K. Spenner)
Vol. 7 Mechanical Sensors (Volume Editors: H. H. Bau, N. F. deRooij, B. Kloeck)
Vol. 8 Micro- and Nanosensor Technology/Trends in Sensor Markets
(Volume Editors: H. Meixner, R. Jones)

©VCH Verlagsgesellschaft mbH, D-69451 Weinheim (Federal Republic of Germany), 1995

Distribution:
VCH, P. O. Box 10 11 61, D-69451 Weinheim (Federal Republic of Germany)
Switzerland: VCH, P. O. Box, CH-4020 Basel (Switzerland)
United Kingdom and Ireland: VCH (UK) Ltd., 8 Wellington Court, Cambridge CB1 1HZ (England)
USA and Canada: VCH, 220 East 23rd Street, New York, NY 10010–4606 (USA)
Japan: VCH, Eikow Building, 10-9 Hongo 1-chome, Bunkyo-ku, Tokyo 113 (Japan)

ISBN 3-527-26774-3

Sensors

A Comprehensive Survey

Edited by
W. Göpel, J. Hesse, J. N. Zemel

Volume 8

Micro- and Nanosensor Technology/
Trends in Sensor Markets

Edited by
H. Meixner and R. Jones

Weinheim · New York
Basel · Cambridge · Tokyo

Series Editors:
Prof. Dr. W. Göpel
Institut für Physikalische und
Theoretische Chemie der Universität
Auf der Morgenstelle 8
D-72076 Tübingen, FRG

Prof. Dr. J. Hesse
Carl Zeiss,
ZB „Entwicklung“
Postfach 1380
D-73447 Oberkochen, FRG

Prof. Dr. J. N. Zemel
Center for Sensor Technology
University of Pennsylvania
Philadelphia, PA 19104-6390, USA

Volume Editors:
Prof. Dr. H. Meixner
Siemens AG
Zentralabteilung Forschung
und Entwicklung
Otto-Hahn-Ring 6
D-81739 München, FRG

Dr. R. Jones
Cambridge Consultants Ltd.
Science Park
Milton Road
Cambridge CB4 4DW
U.K.

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.

Published jointly by
VCH Verlagsgesellschaft mbH, Weinheim (Federal Republic of Germany)
VCH Publishers Inc., New York, NY (USA)

Managing Editors: Dr. P. Gregory, Dipl.-Phys. W. Greulich
Production Manager: Dipl.-Wirt.-Ing. (FH) H.-J. Schmitt
Indexing: Walter Greulich Publishing Services, D-69469 Weinheim

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

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Sensors : a comprehensive survey / ed. by W. Göpel ... -
Weinheim ; New York ; Basel ; Cambridge ; Tokyo : VCH.
NE: Göpel, Wolfgang [Hrsg.]

Vol. 8. Micro- and nanosensor technology; Trends in sensor
markets / [vol. ed.: H. Meixner ; R. Jones]. – 1995
ISBN 3-527-26774-3
NE: Meixner, Hans [Hrsg.]

©VCH Verlagsgesellschaft mbH, D-69451 Weinheim (Federal Republic of Germany), 1995

Printed on acid-free and chlorine-free paper

All rights reserved (including those of translation into 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 a 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.
Composition: Filmsatz Unger + Sommer GmbH, D-69469 Weinheim.
Printing: DiesbachMedien, D-69469 Weinheim.
Bookbinding: Großbuchbinderei J. Schäffer, D-67269 Grünstadt.
Printed in the Federal Republic of Germany.

Preface to the Series

The economic realities of productivity, quality, and reliability for the industrial societies of the 21st century are placing major demands on existing manufacturing technologies. To meet both present and anticipated requirements, new and improved methods are needed. It is now recognized that these methods must be based on the powerful techniques employing computer-assisted information systems and production methods. To be effective, the measurement, electronics and control components, and sub-systems, in particular sensors and sensor systems, have to be developed in parallel as part of computer-controlled manufacturing systems. Full computer compatibility of all components and systems must be aimed for. This strategy will, however, not be easy to implement, as seen from previous experience. One major aspect of meeting future requirements will be to systematize sensor research and development.

Intensive efforts to develop sensors with computer-compatible output signals began in the mid 1970's; relatively late compared to computer and electronic measurement peripherals. The rapidity of the development in recent years has been quite remarkable but its dynamism is affected by the many positive and negative aspects of any rapidly emerging technology. The positive aspect is that the field is advancing as a result of the infusion of inventive and financial capital. The downside is that these investments are distributed over the broad field of measurement technology consisting of many individual topics, a wide range of devices, and a short period of development. As a consequence, it is not surprising that sensor science and technology still lacks systematics. For these reasons, it is not only the user who has difficulties in classifying the flood of emerging technological development and solutions, but also the research and development scientists and engineers.

The aim of "Sensors" is to give a survey of the latest state of technology and to prepare the ground for a future systematics of sensor research and technology. For these reasons the publishers and the editors have decided that the division of the handbook into several volumes should be based on physical and technical principles.

Volume 1 (editors: T. Grandke/Siemens (FRG) and W. H. Ko/Case Western Reserve University (USA)) deals with general aspects and fundamentals: physical principles, basic technologies, and general applications.

Volume 2 and 3 (editors: W. Göpel/Tübingen University (FRG), T. A. Jones†/Health and Safety Executive (UK), M. Kleitz/LIESG-ENSEEG (France), I. Lundström/Linköping University (Sweden) and T. Seiyama/Tokuyama Soda Co. (Japan)) concentrate on chemical and biochemical sensors.

Volume 4 (editors: J. Scholz/Sensycon (FRG) and T. Ricolfi/Consiglio Nazionale Delle Ricerche (Italy)) refers to thermal sensors.

Volume 5 (editors: R. Boll/Vacuumschmelze (FRG) and K. J. Overshott/Gwent College (UK)) deals with magnetic sensors.

Volume 6 (editors: E. Wagner and K. Spenner/Fraunhofer-Gesellschaft (FRG) and R. Dändliker/Neuchâtel University (Switzerland)) treats optical sensors.

Volume 7 (editors: N. F. de Rooij/Neuchâtel University (Switzerland), B. Kloeck/CSEM (Switzerland) and H. H. Bau/University of Pennsylvania (USA)) presents mechanical sensors.

Each of these volumes is, in general, divided into the following three parts: specific physical and technological fundamentals and relevant measuring parameters; types of sensors and their technologies; most important applications and discussion of emerging trends.

Volume 8 (editors: H. Meixner, Siemens (FRG), and R. Jones, Cambridge Consultants (UK)) consists of two parts: one on micro- and nanotechnology, the other on sensor markets.

A comprehensive cumulative index published as a separate volume rounds off the series, facilitating effective use of the preceeding 8 volumes.

The series editors wish to thank their colleagues who have contributed to this important enterprise whether in editing or writing articles. Thanks are also due to the publishers, VCH Verlagsgesellschaft and their editorial/production staff P. Gregory and H.-J. Schmitt, and also especially to W. Greulich for their support in bringing this series into existence.

W. Göpel, Tübingen J. Hesse, Oberkochen J. N. Zemel, Philadelphia, PA

September 1995

Preface to Volume 8 of "Sensors"

Microtechnology, microfabrication, and integrated-circuits technology are key terms which continue to dominate discussions in all branches of sensor research and development. Microfabrication has reached a stage of serious application and is accepted as a good alternative to classical "macroscopic" technologies. It has provided us with the means of producing sensors characterized by, for example, high sensitivity, small size, integrated electronics and optics, and low cost.

New developments in solid-state physics, the technological application of quantum effects, materials research, electrooptical and electrochemical transducer technology, and information engineering have opened the door to the world of nanoscopic fabrication methods. Nanotechnology, for example the manipulation of structures on the atomic/molecular level, is still in its infancy, but will probably gain in importance in all fields of sensor application over the next ten to twenty years. Topics under current discussion include molecular electronics, molecular pattern recognition, optical and bio-computing, etc.

While the first seven volumes of "Sensors" have been devoted to various types of sensors from the viewpoint of techniques, tested materials, and well-understood principles, the present volume goes beyond this frontier. It deals with sensor-relevant micro- and nanotechnological aspects by giving an overview of the state of the art and also of future trends. Another intention of Volume 8 is to supply the reader with first-class information on the sensor markets. The development of commercially successful sensors is expensive and technically difficult and therefore requires sophisticated market knowledge. Central technical and commercial trends are presented from which a broad perspective of the emerging sensor industry may be acquired. Volume 8 rounds off the series, giving a comprehensive survey of the whole field.

The volume is divided into the technologically oriented part entitled "Micro- and Nanotechnology" and the part on "Sensor Markets". The first part, consisting of Chapters 1 to 11, begins with non-IC and IC technologies, proceeds with acoustic-wave devices and high-temperature microsensors, integrated optical sensors and optical microsensors and ends with molecular sensors and nanotechnology. The "market" part (Chapters 12 to 18) covers fields of application such as aerospace, process control, medicine/healthcare, environmental monitoring, automotive control and manufacturing technology.

We hope the book and the series will serve well as a reference source for many years to come. We hope that this book and the series as a whole will help all those concerned with sensor research, development and application to success in this fascinating field.

H. Meixner
R. Jones

W. Göpel
J. Hesse

Contents

List of Contributors XI

Part I: Micro- and Nanotechnology

1 Sensors in Micro- and Nanotechnology 1
H. Meixner

2 Approach to Microsystem Design 23
B. Kloeck, M. Derauwe

3 Sensors in Microsystems 51
H. Baltes, C. de Azeredo Leme

4 Three-Dimensional Microsensor Technology 79
H. J. Ache, W. Menz, J. Mohr, M. Strohrmann, W. Schomburg, B. Büstgens, J. Reichert, W. Hoffmann, W. Faubel

5 Acoustic Wave Devices (SAW and BAW) 135
G. Fischerauer, A. Mauder, R. Müller

6 High-Temperature Microsensors 181
J. Gerblinger, K. H. Härdtl, R. Aigner, H. Meixner

7 Integrated Optics Sensors 221
H. Teichmann

8 Optical Microsensors 259
H. Bartelt

9 Materials in Nanotechnology 275
H. Schmidt

10 Sensors and "Smart" Molecular Nanostructures: Components for Future Information Technologies 295
W. Göpel

11 Future Nanosensors 337
R. Wiesendanger

Part II: Sensor Markets

12 Trends in Sensor Technologies and Markets 357
R. Jones

13 Aerospace Sensors 365
R. Czichy

14 Process Sensing and Control 413
M. J. Scott

15 Medical and Healthcare Sensors 431
V. M. Owen

16 Environmental Sensors 451
K. Jones

17 Automotive Sensors 491
P. Cockshott

18 Sensors in Manufacturing and Quality Assurance 525
N. Pratt

Index 539

List of Symbols and Abbreviations 552

List of Contributors

Prof. H. J. Ache
Institut f. Radiochemie
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Dr. R. Aigner
Lehrstuhl für Technische Elektronik
TU München
Arcisstraße 21
D-80333 München
Germany

Prof. H. Baltes
ETH-Hoenggerberg
Physics Electronics Laboratory
HPT-H6
CH-8093 Zürich
Switzerland

Prof. Dr. H. Bartelt
Institut für Physikalische Hochtechnologie
Helmholtzweg 4
D-07743 Jena
Germany

Dr. B. Büstgens
Institut f. Mikrostrukturtechnik
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Dr. P. Cockshott
Peter Cockshott Associates
Broom Hawes
New Street
Broughton in Furness
Cumbria LA20 6JD
U.K.

Dr. R. Czichy
European Space Agency
ESA/ESTEC
Instrument Technology Division
Noordwijk
The Netherlands

Dr. C. deAzeredo Leme
ETH-Hoenggerberg
Physics Electronics Laboratory
HPT-H6
CH-8093 Zürich
Switzerland

Dr. M. Derauwe
CSEM Centre Suisse d'Electronique et de
Microtechnique SA
Rue Jaquet-Droz 7
CH-2007 Neuchatel
Switzerland

Dr. W. Faubel
Institut f. Radiochemie
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Dr. G. Fischerauer
Siemens AG
ZFE ST KM 11
Otto-Hahn-Ring 6
D-81739 München
Germany

Dr. J. Gerblinger
Siemens AG
ZFE ST KM 14
Otto-Hahn-Ring 6
D-81739 München
Germany

Prof. W. Göpel
Inst. f. Physikal. u. Theoret. Chemie
der Universität
Auf der Morgenstelle 8
D-72076 Tübingen
Germany

Prof. K. H. Härdtl
Institut f. Technologie d. Elektrotechnik
Universität Karlsruhe
Hertzstraße 16
D-76187 Karlsruhe
Germany

Dr. W. Hoffmann
Institut f. Radiochemie
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Dr. R. Jones
Cambridge Consultants Ltd.
Science Park
Milton Road
Cambridge, CB4 4DW
U.K.

Dr. K. Jones
Thorn EMI
CRL
Dawley Road
Hayes, Middlesex UB3 1HH
U.K.

Dr. B. Kloeck
CSEM Centre Suisse d'Electronique
et de Microtechnique SA
Rue Jaquet-Droz 7
CH-2007 Neuchatel
Switzerland

Dr. A. Mauder
Lehrstuhl für Technische Elektronik
TU München
Arcisstraße 21
D-80333 München
Germany

Prof. Dr. H. Meixner
Siemens AG
ZFE ST KM 14
Otto-Hahn-Ring 6
D-81739 München
Germany

Prof. W. Menz
Institut f. Mikrostrukturtechnik
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Dr. J. Mohr
Institut f. Mikrostrukturtechnik
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Prof. R. Müller
Lehrstuhl für Technische Elektronik
TU München
Arcisstraße 21
D-80333 München
Germany

Dr. V. M. Owen
Scitec Management Consultants Ltd.
1 Otter Close
Bletchley
Milton Keynes MK3 7QP
U.K.

Dr. N. Pratt
Consultant in Dimensional Metrology
19 Church Lane
Brofield on Green
Northants, NN7 1Ba
U.K.

Dr. M. J. Scott
Process Measurement Technology
27 West Green
Barrington
Cambridge, CB2 5RZ
U.K.

Dr. J. Reichert
Institut f. Radiochemie
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Dr. M. Strohrmann
Institut f. Mikrostrukturtechnik
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Prof. Dr. H. Schmidt
Institut für Neue Materialien gem. GmbH
Im Stadtwald
Gebäude 43
D-66123 Saarbrücken
Germany

Dipl.-Phys. H. Teichmann
Fachbereich 6 - Physik
Universität Gesamthochschule Paderborn
Warburger Straße 100
D-33095 Paderborn
Germany

Dr. W. Schomburg
Institut f. Mikrostrukturtechnik
Kernforschungszentrum
Postfach 3640
D-76021 Karlsruhe
Germany

Prof. Dr. R. Wiesendanger
Institut für Angewandte Physik
Universität Hamburg
Jungiusstr. 11
D-20355 Hamburg
Germany

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

1 Sensors in Micro- and Nanotechnology

HANS MEIXNER, Central Research and Development, Siemens AG, Munich, Germany

Contents

1.1 Overview-Scenario 2

1.2 Micro- and Nanotechnology for Sensors 3
1.2.1 Microtechnology – Microsystems Engineering 3
1.2.2 Systems-Development: Methods and Tools 5
1.2.3 Signal Processing 5
1.2.4 Testing and Diagnosis of Sensor Microsystems 6
1.2.5 Malfunction Susceptibility, Reliability 6
1.2.6 Constructive and Connecting Techniques 7
1.2.7 Housing Techniques 7
1.2.8 Microtechnologies 8
1.2.8.1 Micromechanics 8
1.2.8.2 Microoptics 9
1.2.8.3 Fiber Optics 10
1.2.8.4 Film-Coating Techniques 10
1.2.8.5 Surface Acoustic Waves 10
1.2.8.6 Techniques of Chemical Sensors 13
1.2.8.7 Biocomponents 13
1.2.9 Fields of Application 15

1.3 Nanotechnology 17
1.3.1 Product Prospects – Application Trends 17
1.3.2 Procedures and Techniques 20
1.3.2.1 The Making of Ultrathin Films 20
1.3.2.2 Creation of Lateral Nanostructures 20
1.3.2.3 Creation of Clusters and Nanocrystalline Materials 20
1.3.2.4 Principles of Self-Organization 21
1.3.2.5 Analysis of Ultrathin Films 21

1.4 Further References 22

1.1 Overview – Scenario

Concern about the future of industrial societies has stimulated intensive discussions around the world and has led to a number of recommendations. Leading the list is the necessity for a speedy and permanent turn towards resource-sparing and environmentally compatible technologies and innovations. Incentives for this development include the exponential growth of the world's population, the accelerating environmental burdens, and our diminishing primary resources. A solution to these problems presupposes ongoing improvements in our technical competence and its optimum utilization.

The following points are to be viewed as especially important fields of application for microtechnologies and nanotechnologies:

- communication,
- information/education/entertainment,
- transportation,
- energy,
- the environment,
- buildings/housing/industrial plants,
- production,
- safety/security.

For these fields of application the technological branches of

- new materials,
- sensor and actuator technology,
- microelectronics, optoelectronics, and information storage,
- electrooptical and electromechanical transducers,
- generation, distribution, and storage of electrical energy,
- information engineering, intelligence functions, and software

have key effects that are both wide-ranging and enduring. In coming years, research and development must concentrate on these technological branches in particular.

Industry is responsible first and foremost for innovations. It must above all:

- increase innovative power through more effective cooperation between research, development, production, and marketing,
- improve the climate of innovation by working in cooperation with science and government.

An improvement in the innovative climate requires a reorientation of the classical division of the research and development process and more effective forms of cooperation between government, national research establishment, and industry. Therefore, in order to convert our scientific competence better into economically measurable results, nationally supported research and research in the private sector must move appreciably closer together.

It is beyond human capabilities to anticipate technological developments over a relatively long period of time (eg, a century). Such capabilities are adequate enough at best to obtain

information for the next 10–20 years on the basis of currently existing knowledge and results, through the use of extrapolation and scenario techniques. The implementation probabilities decrease very rapidly with increasing period of consideration.

1.2 Micro- and Nanotechnologies for Sensors

Whereas in the 1970s microelectronics was one of the dominant strategic research and development goals, in the 1980s materials research and information engineering had priority. Then in the early 1990s, development work was started mainly in the field of miniaturization and integration of extremely small functional units within a system, in order to open up new technologies of the future, ranging from micro- and nanostructures down to molecular and atomic units, by utilizing also phenomena of quantum physics and quantum chemistry.

Sensors are of essential importance for most products and systems and for their manufacture. To some extent the development of sensors has not been able to keep pace with the tumultuous developments in microelectronic components. For this reason sensorics is in a restructuring phase in the direction of achieving increased miniaturization and integration of sensors and signal processing within a total system. This is giving substantially more importance to technologies that permit low-cost manufacture of both the sensors and the related electronics.

1.2.1 Microtechnologies – Microsystems Engineering

The basic philosophy of microsystems engineering can be described as using the smallest possible space to record data, process it, evaluate it, and translate it into actions. The special feature of this engineering is its combining of a number of miniaturization techniques or basic techniques.

Technical developments in the fields of sensorics, actuators, ASICs, and micromechanics are growing together into a "system". Innovations in the area of field-bus engineering and mathematical tools (computer logic) can improve these systems and optimize communication between them.

Thus a complex technology is available that autonomously processes information and directly translates it into actions in a decentralized fashion on peripheral equipment, without the need for large-scale central data processing.

Microsystems engineering is thus not only an enhancement of microelectronics, it represents also a qualitative innovation.

Microelectronics has entered in nearly all devices in which information is processed or processes are regulated or controlled, from the computer to the automobile and extending to self-sufficient robot systems. Why should not also other components and technologies be miniaturized and integrated on a chip and the intelligence of the system be expanded, with a simultaneously greatly reduced energy consumption? The sum total of these future changes brings a considerable advantage, which leads to a large number of new applications having great benefits for society.

The combination of a number of miniaturization techniques presupposes that the following problem is solved: in the design and realization of systems, a great deal of interdisciplinary knowledge concerning technical possibilities and technologies must come together. In the ideal case, this "knowledge" should come from one entity, because otherwise a high degree of cooperation is required. This can be successful in turn only when the exchange of information and the logistics are good, and it must be based on standardization and high quality.

A number of separate technologies have already been developed in recent years: high-frequency circuits, power semiconductors, and displays are still part of the field of electronic elements, and they expand the functional scope of the "classical" microelectronics. Micromechanics, integrated optics, electrooptics, chemosensors, biosensors, polymer sensors, sensors in thin-film and/or thick-film engineering, and radio-readable passive surface acoustic wave (SAW) sensors are opening up completely new dimensions.

Up to now, these technologies have only been pursued separately from one another, and in part they are also based on different materials to silicon, for example, on gallium arsenide, ceramics (eg, Al_2O_3), glass, or even monocrystals such as quartz or lithium niobate.

Nevertheless, today microsystems are frequently constructed in hybrid fashion from various different parts, from various technologies, with the goal of miniaturization going along with a simultaneously enhanced functionality. Therefore, new procedures are aimed at combining chips directly with one another, whether this is as a "chip on a chip" or as a "chip within a chip".

Microsystems engineering will provide manufacturing machinery which will enable much finer work than it is possible today. The door to the submicro world has in any case already been opened. The former magical limits of micrometer dimensions are being considerably

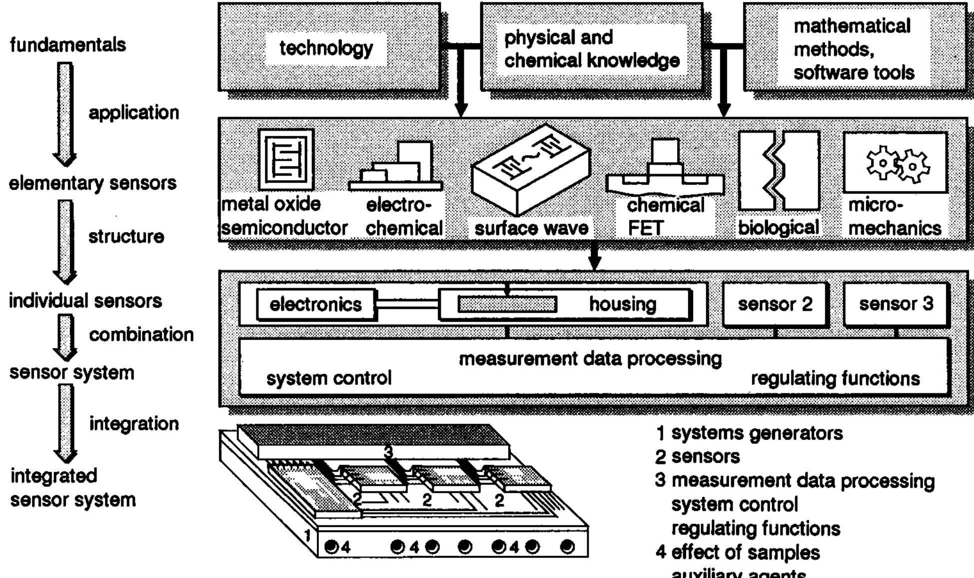


Figure 1-1. Sensors in microsystems engineering: from the elementary sensor to the bus-compatible sensor system.

lowered by the present-day memory chips and SAW components. Microsystems engineering leads us into nanotechnology, and micro-machines will be able to produce systems in molecular and atomic range.

Current research work is aimed first of all at bringing together sensors, actuators, and logic components into self-contained units (Figure 1-1). Here it is not a multitude of elements that is in great demand, but a multiplicity of functions. But to integrate this multiplicity means to bring together different production processes and technologies. Difficulties are necessarily associated with this task, since frequently many processes are "not compatible" with one another.

1.2.2 Systems-Development: Methods and Tools

Closely linked to these process-engineering difficulties are the problems of designing systems. Here, microelectronics has attained at a good position: nowadays logic chips can be designed right on the computer. All the important basic elements can be retrieved from libraries.

There are design tools also for other technologies, but still microsystems engineering is in lack of general design tools. This is true especially for the simulation of systems, which is much more difficult than in the electronics field. What in macroscopic range scarcely plays any role in systems or can be easily compensated has a tremendous effect in the microscopic range. Simulation programs must take into account all of these "cross-sensitivities". They must model electrical, thermal, and mechanical behavior in three dimensions, which presupposes the need for a complex mathematical description.

If there will ever be development tools that will have control over all aspects of microsystems, from design to the generation of masks and extending to simulation and testing, cannot be predicted from our present viewpoint. Technological leadership will have to depend more and more on powerful subsystem tools.

1.2.3 Signal Processing

In order to make use of the growing power of microelectronic systems, corresponding improvements in peripheral areas are required. Information storage devices such as cartridges, diskettes, or hard disks must make more information available along with ever smaller space requirements; sensors must be able to provide more information, and printers or display units must be able to pass on more detailed information. Just as the spatial extent of the storage device or the fineness of the printed or displayed picture are to increase, so also there must be shrinkage in the space needed by a single informational element. The geometrical dimensions of the components of the peripheral devices must follow.

Signal processing stands as a necessary connecting link between sensors and actuators and the superordinate levels of information systems. This processing treats and modifies an input signal of the sensor in such a way that finally the desired network information is made available. This is then converted by other processing measures into an output signal, which causes the desired action (by triggering an actuator) or is forwarded to a superordinate system for further processing.

The further development of signal processing in the next 10 years will involve primarily the bringing together of sensor and signal processing at the site of the measurement, and also the processing of measured values by intelligent devices. Problems that still need to be solved here lie in the field of manufacturing processes that will provide these integration steps, and in the mastering of increased reciprocal effects as a consequence of miniaturization. The availability of suitable signal-processing concepts and design tools will be of decisive importance to further developments here. In addition to signal processing on an electrical basis, implementations of a non-electrical type will be increasingly needed; for example, optical and acoustic signal processing and transmission, ultrasonic and surface applications, and biocybernetic systems. In this connection, high-temperature electronics will become increasingly more important. After the year 2000, complete, intelligent microsystems will be available. They will function as independent, teachable, or adaptable systems, will have geometries extending down into the nano range, will include non-electrical signal processing components, and will be based partly on materials from high-temperature electronics.

1.2.4 Testing and Diagnosis of Sensor Microsystems

Suitable self-testing and diagnostic components and the testing of supportive signal-processing algorithms can be anticipated even simultaneously with the design of sensor microsystems, and these will go beyond previous procedures used in the development of microelectronic circuits. For physical and chemical functions to be adequately tested in complex systems additional "test sensors" may have to be implemented. Via suitable diagnostic interfaces these test components are to be coupled with external diagnostic facilities in order thereby to test the functionality of the systems under real conditions.

Therefore, in the future, we must go over to unified and system-overarching strategies and signal processing concepts for the testing of microsystems.

1.2.5 Malfunction Susceptibility, Reliability

In the future, concepts going beyond this will be decisive in terms of staying competitive. One significant deficiency of microsystems is paradoxically their malfunction susceptibility. The significant advantage of an increased reliability by way of a homogeneous constructive and connective engineering should not be allowed to obscure the fact that in case of a breakdown, for example, of an interconnection or of a component, a microsystem can at best only detect its non-functionality. This basic malfunction susceptibility could possibly be removed by a massive parallelization of the processing chain. Materials with inherently intelligent properties are under discussion for this purpose (see Section 1.2.7). In the extreme case, these materials combine sensor, actuator, and signal processing. Two classes should be distinguished: materials that when exchanged for unintelligent materials increase operational reliability in familiar electronic systems, and materials that perform their function independently, without electronics. The latter have the advantage of being able to work totally without an electrical power supply.

1.2.6 Constructive and Connective Techniques

Constructive and connective engineering (AVT) encompasses the totality of process engineering and design tools that are needed for the implementation of microsystems in an extremely confined space, and therefore it actually forms the bridge between micro- and opto-electronic components as well as micromechanical components for the complete system. It is a significant factor in determining the functionality, quality, and economic efficiency of microsensors and microsystems.

In recent years, constructive and connective engineering has made decisive advances with respect to miniaturizability, fail-safe reliability, and ease of handling.

Major bottlenecks exist at present in the area of AVT for non-electronic components of microsystems engineering; for example, optical, mechanical, chemical, or biological micro-components. Therefore, in addition to progressive developments in the electronics field, in the next few years it will be above all important to provide corresponding technology in this sector as well.

The following basic developments are needed:

- provision of new materials with definite macroscopic and microscopic properties (alloys, polymers, material systems, and film systems),
- investigation of phenomena that arise from the combination of different materials, and the controllability of such phenomena (eg, interdiffusion processes, adhesive strength, thermo-mechanical correlation),
- non-material-dependent removable and permanent connection procedures (memory alloys),
- provision of simulation tools,
- provision of highly integrated fabrication technologies, availability of devices, reproducibility under fabrication conditions, and combinability of differing technology levels,
- system integration based on biological models,
- development of block-integration techniques (stacked chips, stacked modules),
- attendant studies on reliability and on the degradation behavior of the constructed sensor systems.

The power of modern AVT will largely determine the extent to which advances in the development of sensors can be converted into products.

1.2.7 Housing Techniques

Housings in microsensorics and microsytems engineering have the job of protecting components or systems and at the same time establishing the connection to the physical or chemical quantities to be measured. In sensorics and in microsystems engineering the "housing" cannot be regarded as independent. The housing is an indispensable component of the overall system.

There is a demand for multifunctional materials that on the basis of their properties also allow wide-ranging designer freedom in configuring the sensor system. Sensor housings with

sensor windows are needed in particular. These multifunctional materials include, among others:

- shape-memory alloys/polymers,
- piezoelectric ceramics, piezopolymers,
- diamond and diamond-like films,
- magnetoresistive materials,
- optically activatable gels,
- electro-rheological, magneto-rheological liquids,
- optomechanical (especially electrochrome) materials,
- materials with self-diagnostic and alarm functions,
- membranes with self-regulating permeability behavior.

Shape-memory alloys, piezoceramics, or magnetostrictive alloys, embedded in a filamentous or layer-like way in a matrix material, can affect in a well defined manner the state of stress and state of strain of a housing. Structures with a variable rigidity or with adjustable bending and deflection properties might be realizable.

Composite structures that are provided with piezoelectric foils/fibers or photoconductive fibers can perform sensor functions in addition to their structural-mechanics tasks, possibly even for the purpose of self-monitoring their material condition. This in situ damage recognition would be desirable precisely where there are reliability-relevant structural units.

There is not yet fully utilized developmental potential in connection with electro-rheological and magneto-rheological liquids, which react upon electrical or magnetic fields with a rapid and pronounced change in their viscosity, and thereby permit, for example, the coupling of forces. These materials are self-supplying, work without electric power, and are fail-safe, since even when substantial parts are destroyed the undamaged parts that remain can continue to function. Moreover, many such materials are self-adjusting, can survive overloading, and are self-repairing.

Special requirements are placed on housing engineering that involves sensitive windows that are exposed directly to caustic and corrosive chemical and biological media under measurement.

1.2.8 Microtechnologies

Microtechnologies and nanotechnologies create the technological prerequisites needed for miniaturization and for adjusting the individual functions or components of the system. Most microtechnologies have their origin in technologies of microelectronics or conventional sensors/actuators, and these are being further developed in a way that is geared to the interests of micro- and nanosystem engineering.

1.2.8.1 Micromechanics

The fabrication methods developed for microelectronics are not, of course, restricted to the production of electronic components. As a child of microelectronics, micromechanics was

able to rely from the beginning on a large reservoir of fabrication techniques that permitted the design of extremely fine, complex structures in the "batch" procedure.

The previously mentioned great application potential for micromechanics arises from the utilizability of five technical dimensions:

- materials,
- microstructures,
- complex arrays,
- combination (system-capability) with microelectronics,
- utilization of the entire scope of physical phenomena.

Among the available materials, silicon plays a predominant role. The treatment procedures familiar from microelectronics permit geometric three-dimensional forms to be structurized down to dimensions of a few microns.

Other materials besides silicon can likewise be used when appropriate shaping-treatment possibilities are available. For example, a large number of electrical non-conducting materials are of interest for shaping via heavy-ion bombardment.

Starting with lithographic procedures for transferring structures, three-dimensional shaping can be achieved by means of various micromechanical processes, also used in anisotropic etching of silicon, micro-metal plating, and laser processes, or by way of ion-tracking technology, which is still under basic development. With the LIGA procedure (lithography, electroforming, molding), microstructures can be implemented from a wide range of materials, such as metals, plastics, and ceramics. Silicon clearly dominates the materials used at present, but metals, plastics and glasses are gaining in importance.

Special advantages result from the possibility of combining this with microelectronics. Thereby local "intelligence" can be imparted to the micromechanical components. Sensors will be equipped with their own microprocessors. Of particular importance is the possibility of combining micromechanics with a multitude of physical phenomena. In connection with magnetism, micromechanical magnetic heads detecting the data of magnetic plates, can be used as non-contact potentiometers, or when applied as superconducting SQUIDs can accept information from brain waves.

Chemical sensors are being developed that can determine the compositions and concentrations of gases.

1.2.8.2 Microoptics

In addition to microelectronics and micromechanics, microoptics can also be seen as a basic technology. It permits the implementation of communication channels of very high transmission capacity in a very small space. Its goal is to make miniaturized optical components, integrated in wave guide structures, with the aid of planar technologies.

Microengineering methods play a crucial role in the production of microoptics components: from the production of semiconductor lasers to the adjusting of glass fibers at the output window of the laser or the low-loss splicing of glass fibers, and extending to the production of the optical elements, diffraction gratings, or holograms. Factors of great influence on the quality of the products are micron-precise adjustments and also the creation of new structures

and the inexpensive replication of the structures. In this context, certain material systems (eg, III-V semiconducters) and production processes (eg, molecular-beam epitaxy) are being used that go far beyond those employed in silicon technology.

Other basic elements of microoptics are wave guides, which for example can be made by ion exchange in special glasses, oxynitrides on silicon, by means of polyimides on glass substrates, or with special structures on the basis of III-V semiconductors. With wave guide structures such as branches and mirrors, complex beam-guiding elements can be achieved on a substrate.

1.2.8.3 Fiber Optics

This utilizes the guidance of optical and infrared radiation in fiber-optic cables. Accordingly, advantageous features familiar from optical transmission techniques can be utilized: immunity to electromagnetic effects, extremely low attenuation values, and considerable miniaturization possibilities.

The primary application areas of fiber optics in microsystems engineering are fiber-optic sensors and optical power supplies for microsystems. The fiber-optic solutions are at present characterized by a steady but somewhat gradual penetration into microsystems engineering, since they are relatively expensive they are used only where an above-average advantage can be expected or where other solutions are impossible.

1.2.8.4 Film-Coating Techniques

By using methods of epitaxy, it is possible to deposit various materials on each other with atomic precision. By this means, novel concepts for microelectronic components can be realized (see nanocomponents). From microelectronics, powerful technologies for miniaturization techniques are available, which in modified form are suitable for microsystems engineering.

Thick-film techniques are widespread in microsystems engineering. Thin-film technology is likewise used to a great extent, although the process engineering involved is more difficult. At present, developments are being pursued with the following objectives:

- double-sided lithography and assimilation of lithography engineering for three-dimensional micromechanical structures,
- coating and microstructurization of new materials (eg, biomaterials),
- combining various materials in film systems (thin-film stacking techniques, additive techniques),
- modifying applied film,
- ultra-thin impervious insulating films.

1.2.8.5 Surface Acoustic Waves

Acoustic waves that propagate on the surface of a solid similar to the waves on the surface of a pond, are called acoustic surface waves (SAW). Surface waves arise through transient or periodic distortions on the surface.

Electrical signals can excite surface waves on piezoelectric solids (quartz, lithium niobate ($LiNbO_3$), piezoceramics) or on materials provided with thin piezoelectric films. To that end use is made of thin (about 0.1 μm) strip-shaped metal electrodes, which are applied to the surface by vapor-deposition and photolithographic processes. When the signals are applied, electric fields are built up between the electrodes, and these fields are coupled to mechanical distortions by the piezoelectric effect.

Given a propagation velocity of about 3500 m/s, in the typically used frequency range of about 10 MHz to a few GHz for surface waves, microscopically small wavelengths and penetration depths are produced, for example, a magnitude of 35 μm at 100 MHz. The frequency range is limited upwards by the technologically feasible structural fineness of the electrodes.

The unrestricted accessibility of the surface in conjunction with the flexibility of the photolithographic process allows diverse electrode structures to be applied with high precision. The surface waves are affected by the selection of the crystal and the crystal cut and also by the shaping of the conducting structures on the surface. With this approach the transmission behavior for electrical signals that are transmitted via surface waves is determined in a highly reproducible fashion. A multitude of different surface-wave components of high quality with tailor-made electrical transmission properties that are stable over a long time is thereby made possible. In addition to the filters, resonators, correlators, etc. now being manufactured – in some cases in enormous unit numbers – sensor applications of surface-wave components are becoming more and more important.

By means of surface-wave resonators, oscillators can be built that are tuned with the introduction of other effects on the resonator, such as temperature or tensile stress (Figure 1-2). By this means the oscillator becomes a sensitive frequency-to-analog sensor. For example, the introduction of the appropriate force causes it to become a torque sensor that measures the torque on a rotating shaft, without the need for slip rings.

Surface-wave components can be used as passive identification flags, so-called SAW ID tags, for sensor systems designed for non-contacting identification. Radio pulses are in-

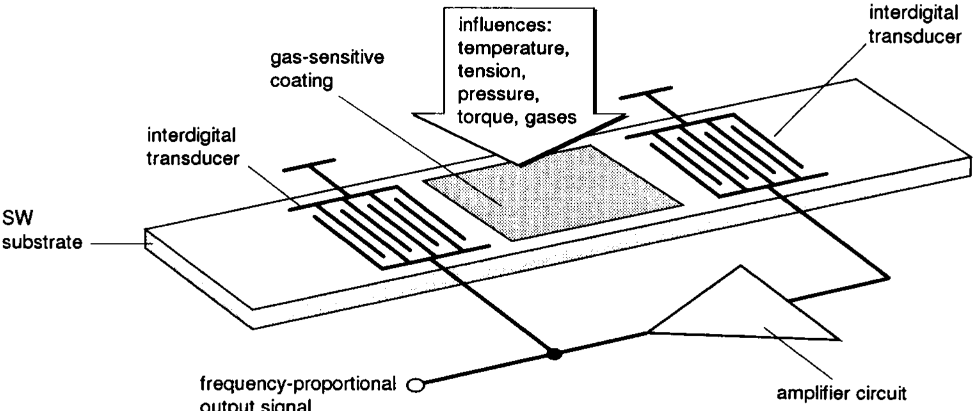


Figure 1-2. SW (surface-wave) sensors: changes in the parameters of SW propagation are converted into frequency changes of an oscillator.

tercepted by a simple antenna and converted into surface waves. Electrodes configured as reflectors are positioned in a characteristic sequence (similar to a bar code) on the surface of the crystal, where they reflect these waves. The reflected waves are re-converted into radio signals and received by the evaluation device. The arrangement of the reflectors is manifested in the response signal as a characteristic pattern in a chain of partial pulses, which for example embody 32 bits, giving 4 billion different distinct possibilities.

The advantages of this process, such as

- remote inquiry, even at inaccessible locations, such as in hazardous areas, through windows and walls, without the expense of an additional installation,
- passive functioning, no need for its own electric power at the ID tag (no battery needed, cost, maintenance, and safety aspects),
- small, lightweight, simple to install,
- well defined inquiry range, interference effects from echoes of other objects are limited because of a relatively short transit time,
- high reliability, "crystal-stable",

will open up a multitude of additional applications in the future for the SAW ID tag, both technical and economic. Interesting examples of applications in process control engineering would be

- identification of roadway and rail vehicles that are bringing in raw materials or are shipping out products,
- marking of containers, receptacles, semi-finished products, and workpieces, and tracking them within the process,
- checking the access authorization of persons, recording of the access of persons in hazardous areas,
- travel flags and orientation aids for autonomously navigating driverless vehicles, speed measurement, position determination, distance measurement,
- selective reflex gates, danger recognition, room security devices,
- malfunction indicators created by the setting of a switchable reflector ("read-me flag") by passive means, eg, upon damage from an impact or excess temperature, thereby preventing more extensive damage.

Another potential for innovative sensor systems arises from a combination of the sensor properties of surface-wave sensors and the radio of ID tags. Such radio-readable sensors can be positioned in an extremely flexible fashion and then can be interrogated via radio. They can be mounted at almost any site desired, even at locations that are not very accessible or are hazardous (eg, via a magnetic clamp), without a special installation effort being needed. In this way the measurement magnitudes of interest at these locations can be measured and passed on from a safe distance, even through windows and walls, by means of an interrogating device, which could, for example, be portable.

1.2.8.6 Techniques of Chemical Sensors

Chemical sensors are like measuring probes when they, in an unknown medium, identify a specific component and determine its concentration within the medium. Such a component can be, for example, nitric oxide or oxygen within the exhaust gas of a motor vehicle or carbon monoxide in room air. Such sensors must not have any cross-sensitivities, ie, they must not change their measured value when other substances are present. They must also respond rapidly, in order, for example, to regulate optimally the combustion process or to indicate malfunctions.

Important areas of application of chemical sensors are:

- medical diagnostics,
- workplace monitoring,
- chemical process engineering,
- environmental monitoring of water and air.

The chemical microsensors that have been used up to now do not satisfy the demands placed on them very well. From these unsatisfactory results it has been learned that only by means of system approaches successful results can be achieved. Thus, the chemical composition of the sensitive film of a sensor element is very crucial. The physical processes occurring on the surface or within the sensitive film likewise influence the result. In the case of semiconducting metal oxides, parameter changes can be obtained, for example, by doping the sensitive film or changing the working temperature. The structural architecture and the thickness of the sensitive film can significantly influence the result and especially the sensor's long-term stability.

Signal evaluation plays a decisive role in chemical sensorics. Often it is necessary to carry out the signal processing using a number of sensor elements of differing cross-sensitivity, in order to draw definitive conclusions about the concentration of the components from the different signals. So far there have been only insular solutions, and these are still not very perfect. Signal-processing concepts based on neural networks or fuzzy logic are still in their infancy.

Developmental approaches to complex chemical analysis systems are at the initial stage of research.

1.2.8.7 Biocomponents

Many people see a contradicition between what has naturally evolved and what has been created by human hand, a contradicition that supposedly would be difficult to bridge. Yet when capabilities are analyzed that in the course of evolution have been impressed by nature alone, it becomes clear that even the living world makes use of certain "technologies".

From combining certain biological building blocks that permit the organism to have perception together with components of mensuration and analysis engineering, a number of interesting applications are arising. Biosensors can be regarded as a bridge between biology and engineering (especially electrical engineering and electronics). They lead to measuring methods that can bring about important advances in medicine, in environmental protection, in food technology, and in connection with process control in other fields.

When the development of biosensors is considered, it can be seen that the trend of microsystems engineering – smaller and smaller, more and more precise – is being repeated here.

Over all, for biosensors the following lines of application are emerging:

- medical diagnostics, including patient monitoring in intensive medicine and bed-side analysis,
- environmental analysis, above all waste-water and water-pollution control,
- food quality control, for example, for purity and freshness,
- process control and substance testing, and other utilizations in the chemical-pharmaceutical industry,
- drug detection.

These applications become possible because biological probes are characterized by high sensitivity and selectivity, namely an accurate recognition of specific chemical compounds. Analyses can be provided considerably more rapidly than before, and also more reliably and almost continuously. One advantage is that these analyses can be made directly at the site of events instead of in the laboratory. Thereby not only a distinct reduction in costs per measurement, but also enormous savings in connection with the disposal of the most extremely poisonous analytes are possible even today.

In using biological probes, the capability of biological materials is exploited to find certain substances, so-called analytes, among a multitude of other materials, in accordance with the

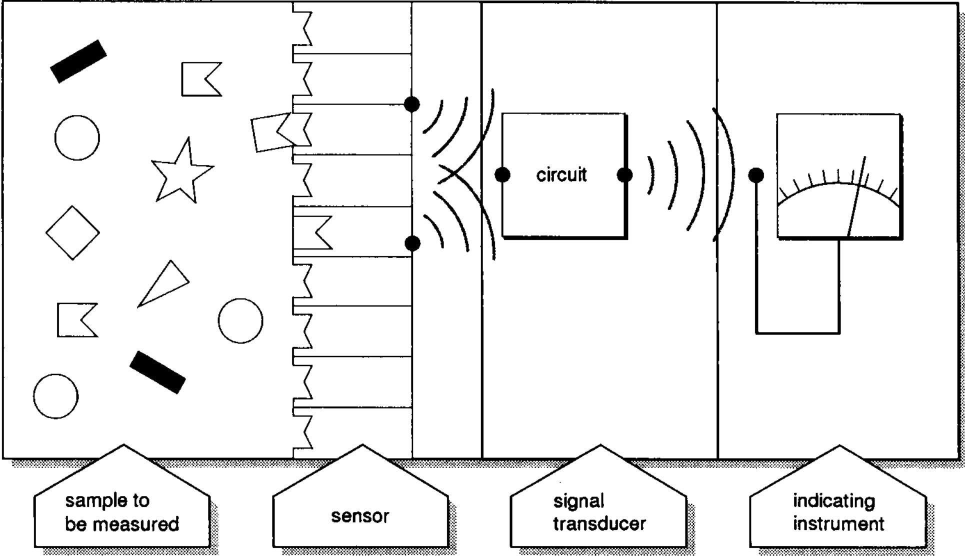


Figure 1-3. Functional principle of biosensors. According to the so-called lock-and-key principle, the biological measuring probe ("lock") reacts with a specific substance to be analyzed, the analyte ("key"). The physical magnitudes that change on account of the reaction are converted by the signal transducer into measurable signals, then amplified and read out. (Illustration: UNIVERS/Max Ley [5]).

lock-and-key principle of molecular biology, and to react with them. The number of possible analytes (keys) is virtually unlimited, and so are the biological recognition materials (locks). Among the analytes there can be both organic and also inorganic substances, as they are found, for example, in the living organism and in food. The chemical reaction of the biological material with the analyte affects physical parameters, such as the electrical potential connected with extremely small changes in mass, fluorescence, or temperature. These changes are converted into measurable signals and are amplified (Figure 1-3). In this way it is feasible to conduct analyses that were only possible with customary methods, if at all, by way of painstaking separation and measuring processes, and with the need to use experimental animals.

One disadvantage in contrast to this range of advantages is the limited stability of the biological measuring probe. Regardless of whether these are enzymes, antibodies, nucleic acids, cells, or even complete microorganisms, they exist in an aqueous milieu. As a rule, the biological components are confined (immobilized) between membranes. The analytes can diffuse out of the sample via one of the membranes and react with the biological component. A signal thereby generated is transformed by the signal transducer and is passed on.

At present biosensors have been described for about 120 different analytes. Among these are sensors for low-molecular-mass substances, sensors for enzyme components, and sensors for macromolecules, for viruses, and for microorganisms. However, most of these developments are still at the level of general solutions.

1.2.9 Fields of Application

For greater functionality, modern technical systems require more information. They become available by communication-capable (teachable, radio-readable) sensors. The availability of low-cost sensors plays an important role in ensuring that the implementation of these systems will be marketable, especially when large numbers of units are involved. This goal is usually achieved only via "batch-capable" microtechniques. Some of these technologies for miniaturization are already available.

At this point, the great potential of sensors now on the market must also be addressed, because here also the evolution of sensor techniques is going in the direction of miniaturization, multi-functionality, integration, and intelligence.

The fields of application that have a high demand for sensors have already been cited in the Introduction. Here the following fields of application are relevant:

- transportation,
- safety,
- modern conveniences,
- vehicle-drive management,
- vehicle running gear,
- traffic guidance,
- intelligent rail systems,
- navigation,
- roadway-route and obstacle recognition,
- health and usage monitoring.

Building engineering:

- illumination management,
- safety, property protection,
- energy management and air-conditioning management,
- emission and immission monitoring,
- intelligent household appliances,
- water and waste-water management.

Security:

- building access, ID tags,
- property protection,
- theft-proofing of motor vehicles,
- monitoring of national borders,
- fire detection and fire fighting.

Innovations in the field of measurement, control, and regulation engineering have advanced far enough so that sensors are available for these technologies in their various areas of application. But at the same time it must be noted that there is a shortage of efficient sensors at moderate prices that could promote rapid innovations of product and processes with the aid of microtechniques and microelectronics.

Whereas the automobile, the environment, and residential property are fields of application that call for inexpensive and rugged sensors, the fields of application of production and automation, security, health, and traffic-control engineering pay particular attention to the precision of sensors.

From the multitude of possible sensor applications, the following fields of application seem to be attractive in terms of demand, numbers of units, and complexity:

Traffic engineering:

- traffic carriers: automobiles, trains, aircraft, and ships.

Health:

- affordable health care,
- monitoring of TLVs (industrial threshold limit values),
- non-invasive analysis,
- minimally invasive surgery,
- exercise monitoring and maintaining physical fitness,
- drug dosing,
- intensive-care medicine.

Environment:

- monitoring of soil, water bodies, air, agricultural areas, dumps,
- leakage detection,
- material recognition and byproduct separation,
- emission and immission monitoring.

Production techniques:

- process control and regulation,
- process analysis,
- quality monitoring,
- testing,
- operational and environmental safety.

1.3 Nanotechnology

With respect to nanotechnological development work, what is innovative is the integrated application of various sciences and engineering disciplines, such as physics, chemistry, biology, process engineering, and communication and information science. However, this development presupposes successful basic research. At present, a corresponding market expectation cannot yet be seen. Moreover, all of the innovative production, characterization, and modification methods suitable for nanotechnology are oriented consistently towards the idea of "engineering on the atomic and molecular level". In this regard, not only the creation of complex molecular electronics components or biochips using neural networking are discussed at present, but also, for example, the implementation of nanomachines, which by means of physical, chemical, and biological methods could be built molecule by molecule or atom by atom into molecular systems having submicroscopic dimensions. A general view is shown in Figure 1-4.

Nanotechnology is gaining importance in all fields of application where smaller and smaller structures and greater and greater precision are becoming decisive to the market success of individual products, eg, in the fields of optics, information engineering and nano-electronics, solar and environmental engineering, biotechnology, sensorics, robotics, and medicine.

1.3.1 Product Prospects – Application Trends

The reduction of microscopic structural dimensions down to the nanometer scale, in order to utilize dimension-dependent properties and nanoscopic phenomena, is called a "*top-down strategy*". The most important advance here is from microelectronics to nanoelectronics. By these means, the aim is to utilize new physical interactive mechanisms whose causes are accounted for in terms of the transition from continuum physics to quantum physics. In order

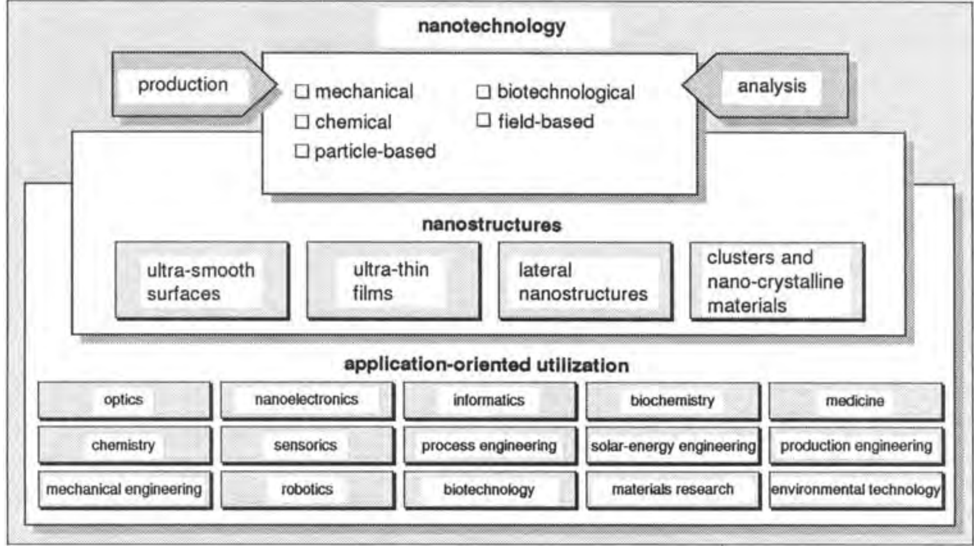


Figure 1-4. General view of the production, analysis, and application of nanotechnology structures.

to make such complicated components, it is also important to press on with the basic technologies needed for this development.

In *quantum electronics*, the making of layer arrangements for three-dimensional electron gases plays a major role. Chain arrangements or lattice arrangements of quantum dots – the building blocks of the individual electronics – permit, for example, under the influence of the Coulomb barrier, the correlated and controllable tunneling of individual electrons (single electron logic).

Since quantum dots can be viewed as artificial atoms with discrete energy levels, pronounced optical non-linearities can furthermore arise from external excitation by means of photons. Therefore, such structures are also suitable in *optoelectronics*. Furthermore, quantum dots can also be used as a *storage medium*.

For special applications (eg, high temperatures, high radiation exposure, high frequency), *vacuum electronics* can contribute to the opening up of new markets. One significant area of application is in display technology. In the area of *information storage*, nanotechnology will be the basis for future developments. The limit of storage density is technologically reached only when atomic dimensions are employed for structurization, with a storage density higher by a factor of 10^6 than of current magnetic recording methods.

Moreover, future storage technologies will take into consideration the actuating of three-dimensional storage mechanisms.

Another conceivable application is the interlinking of zeolite crystals with polymer wires and the subsequent local storing of chemical information. Also optical storage in bacteriorhodopsin, holographic storage technology, and electron-hole burning are other potential options for three-dimensional storage.

For *optics* it is possible to open wide fields of application from the development of structurization and super-polishing procedures for ultraprecision treatment, to nanometer-precise positioning and control methods necessary for this. In *chemistry* and *biology,* the applications of nanotechnology go beyond analysis (biosensorics) or the physical manipulation of molecules (tailor-made molecules). We could mention, for example, catalysis (by clusters or nanoporous materials) and filtration engineering (eg, biological ultrafilters with pores in the micrometer diameter range).

Visionary fields of application for nanotechnology can be seen in *medical engineering,* for example, in connection with the physical coupling of biological systems with microelectrodes, the use of nanometer-sized capillary tubes for micro-dosing devices, the use of micro-nanomachines in surgery, and the development of a silicon-retina chip.

In the field of *mechanical engineering* there is the possibility of minimizing wear and tear on workpieces by giving them a surface treatment of the highest quality.

In *solar-energy engineering* the development of granular films with tailor-made optical properties is conceivable.

Bottom-up strategy: another approach in nanotechnology consists in the selective building up of atomic or molecular aggregates into larger systems. Here the main distinctions made are those of principles of self-organization, organic/inorganic interface phenomena, and the selective chemical or physical joining of molecular systems to prepared surfaces.

The *Langmuir–Blodgett technique* represents a multi-disciplinary field in this regard. Material combinations of such monomolecular layers permit the selective preparation of materials with highly anisotropic electrical properties, good insulation properties, non-linear optical properties, or pyroelectrical activity.

In *molecular electronics,* through inspection and control, molecular processes and system concepts are to be developed that show a strong parallelism with information handling in their general tendency. Serving as a model here are properties of molecular and biological systems such as efficiency, packing density, self-organization, fault tolerance, reproducibility, and low energy consumption. With regard to the implementation of three-dimensional integrated circuits on a molecular level, technological upheavals are expected similar to those that came about in the course of supplanting the electron tube by the transistor.

Information being gained from *cluster research* also shows that there is a great range of applications here, for example, with respect to heterogeneous catalysis, microstructure physics, development of nanotechnological structure units, medicine, sensorics, and photovoltaics. Clusters of intermediate size show new properties between molecules and the solid state, whereby electrical and optical characteristics can be adjusted in a well defined way.

Analytics: both in the field of material and bio-sciences and also in the sector of micro- and nanoelectronics/sensorics, important scientific and technological advances will depend on whether, for example, molecular structures can be imaged in a non-destructive way and whether time changes on the atomic scale can be directly made visible. The currently maximum lateral resolution for the method now being used the most frequently, namely scanning electron microscopy, amounts to about 1–5 nm. With transmission electron microscopes, local resolutions of up to 0.1 nm can be achieved. An especially important role is played by scanning-probe methods. Their most important representative is the scanning tunnel microscope (STM). The STM is based on the distance-dependent electron tunnel current between a peak and a surface. With it, local resolutions in the range of sub-atomic diameters are possible.

With the development of scanning force microscopy (AFM), the particular advantages of STMs could be extended also to non-conducting samples. A disadvantage of this special scanning technique is that chemical elements cannot be unequivocally specified.

1.3.2 Procedures and Techniques

The listings offered below should give a general idea of the most important methods and techniques used in nanotechnology.

1.3.2.1 The Making of Ultrathin Films

By the construction and connecting of various ultra-thin films in the nm-range, complex combinations of mechanical, optical, electrical, or chemical properties can be created in a three-dimensional configuration with a function-oriented character and with an extremely high integration density. This is true not only for inorganic and organic substrate films but also for biological molecular structures, which can be used for making switches, storage devices, sensors, processors, actuators, membranes, catalysts, etc., and also for multivalent applications.

The requirements placed on the making of films in nanotechnology are characterized by the demand for sharp atomic interfaces and for controlling depositions in atomically thin layers. These are mostly vacuum methods, which are based either on molecular beam epitaxy (MBE) or precipitation from the gaseous phase (eg, atomic layer epitaxy (ALE), chemical beam epitaxy (CBE)).

For organic films the Langmuir-Blodgett method can be mentioned as an important preparation technique.

1.3.2.2 Creation of Lateral Nanostructures

In addition to the creation of extremely thin films, another central problem of nanotechnology is to make lateral structures and localized doping profiles.

The motivation for making smaller and smaller structures has come mainly from microelectronics. For the making of masks, in addition to optical lithography also direct-writing methods are being used, where focused laser, ion, or electron beams are employed. The resolutions are roughly 200 nm, 50 nm, and about 10 nm, respectively. Alternative structurization methods are holographic structurization and STM lithography. Nanolithography per se involves structural dimensions <100 nm.

One disadvantage of lithography that uses scanning-probe methods is its low writing speed and its small visual field. An array of structurization probes would provide a remedy for this.

1.3.2.3 Creation of Clusters and Nanocrystalline Materials

Clusters are agglomerates consisting of about 10–1000 atoms. With these, the dependence of their physical and chemical properties on cluster size is directly correlated with their major surface-to-volume ratio. There are many different techniques that can be used to make clusters

in the condensed phase, in the gaseous phase, or in vacuum. The scientific and technical interest in clusters is in essence based on their physical and chemical properties, which specifically for clusters consisting of up to about 100 atoms may depend greatly on their size. Thus, for example, structure, stability, electronic state density, optical absorption behavior, ionization potential, melting-point lowering, and also catalytic, magnetic, and chemical properties of cluster materials depend significantly on the number of particles in the cluster.

1.3.2.4 Principles of Self-Organization

With the use of atomic and molecular architectures, new material structures are being provided target-selectively for the execution of well defined chemical, physical, and biological functions. These systems have new capabilities. They are among those intelligent products that record and evaluate data and execute actions resulting from this. For example, molecules can become electric switches or can be used for information reception, storage, and transmission. Molecular architectures can also be used for pattern recognition, self-structurization, self-organization, self-reproduction, for the selective linking together of atoms, or for the creation or separation of molecular groups (clusters). By the fabrication of new structurally defined molecular species, metals can be transformed into semiconductors or into optical or "switchable" industrial materials. Filters or membranes become controllable, since they close their pores upon contact with certain molecules.

It can be expected that on this basis, the development of new

- sensor generations,
- biosensors,
- information storage devices and information conductors,
- biochips,
- molecular transistors, switches, and storage devices,
- structural and functional materials,
- materials for artificial intelligence systems,
- biocompatible materials,
- optical components,
- medical applications, pharmaceuticals, etc.,

will become feasible.

The general goal of molecular/atomic architecture is to see a certain property being realized in a molecule or material.

1.3.2.5 Analysis of Ultrathin Films

Highly developed procedures and devices that work with beams of electrons, ions, neutrons, neutral particles, and photons, with field-emission and tunnel effects, and also on the basis of acoustic, electrical, magnetic, and optical principles are used in nanoanalytics.

The detection of lateral element distribution is done with electron-beam or ion-beam techniques. Depth-dependent element analysis is done with ion sputtering.

A relatively new analytical procedure for nanotechnology can clearly be seen in the use of scanning tunnel microscopy/scanning force microscopy.

1.4 Further References

[1] Muller, R. S., Howe, R. T., Senturia, St. D., Smith, R. L., White, R. M., New York: IEEE Press, 1991. *Microsensors.*

[2] Muller, R. S., in: *Proceedings of Transducers,* San Francisco, 1991.

[3] Wise, K. D., Najafi, N., The Coming Opportunities in Microsensor Systems, in: *Proceedings of Transducers, San Francisco, 1991;* 1991; pp. 2-7.

[4] de Raolj, N. F., Current Status and Future Trends of Silicon Microsensors, in: *Proceedings of Transducers, San Francisco, 1991;* 1991, pp. 8-13.

[5] Benecke, W., Petzold, H. C., *Proceedings IEEE: Micro Electro Mechanical Systems, 1991;* Travemünde; 1992.

[6] *Mikrosystemtechnik [Microsystems Engineering];* Bonn: Bundesministerium für Forschung und Technologie, 1994.

[7] Stix, G., Trends in Micromechanics "Micron Machinations", *Sci. Am.* November (1992) 73-82.

[8] Howe, R. T., Muller, R. S., Gabriel, K. J., Trimmer, W. S. N., Silicon Micromechanics: Sensors and Actuators on Chip, *IEEE-Spectrum,* July (1990) 29-37.

[9] Kroy, W., Fuhr, G., Heuberger, A., Mikrosystemtechnik [Microsystems Engineering], *Spektrum Wiss.* May (1992) 98-115.

[10] Peterson, K., Dynamic Micromechanics of Silicon, *IEEE Trans. Electron Devices* **ED-25** (1978) 1241-1250.

[11] Heuberger, A., *Mikromechanik [Micromechanics];* Berlin: Springer, 1989.

[12] McEntee, J., Start Making Microsensors, *Phys. World* December (1993) 33-37.

[13] Kroy, W., Heuberger, A., Sehrfeld, W., Nerven und Sinne für die Chips [Nerves and Senses for Chips], *High Tech* **4** (1988) 22-36.

[14] Bley, P., Menz, W., Aufbruch in die Mikrowelt [Breakout into the Microworld], *Phys. Bl.* **49** (1993) 179-184.

[15] Menz, W., Bley, P., *Mikrosystemtechnik [Microsystems Engineering];* Weinheim: VCH, 1993.

[16] Bulst, W. E., Developments to Watch in Surface Acoustic Wave Technology, *Siemens Rev.* (1994) 2-6.

[17] *Biosensorik [Biosensors];* Bonn: Bundesministerium für Forschung und Entwicklung, 1992.

[18] *Produktionsintegrierter Umweltschutz [Protection-Integrated Environmental Protection];* Bonn: Bundesministerium für Forschung und Entwicklung, 1994.

[19] Scheller, F., Schubert, F., Pfeiffer, D., Biosensoren [Biosensors]; Enzym- und Zellsensoren: Anwendungen, Trends und Perspektiven [Enzyme and Cell Sensors: Applications, Trends, and Prospects], *Spektrum Wiss.* September (1992) 99-103.

[20] Scheller, F., Schubert, F., *Biosensoren [Biosensors];* Basel: Birkhäuser, 1989.

[21] Buck, P., *Biosensor Technology;* New York: Marcel Dekker, 1990.

[22] Renneberg, R., Molekulare Erkennung mittels Immuno- und Rezeptorsensoren [Molecular Recognition via Immuno-Sensors and Receptor-Sensors], *Spektrum Wiss.* September (1992) 103-112.

[23] Scheper, Th., Wärme- und Lichtsignale aus dem Bioreaktor [Thermal and Light Signals from the Bioreactor], *Spektrum Wiss.* September (1992) 103-144.

[24] Corcoran, E., Nanotechnik [Nano-Engineering], *Spektrum Wiss.* (1991) 76-86.

[25] Bachmann, D. G., *Nanotechnologie [Nanotechnology];* VDI, 1994.

[26] Reed, M. A., Quantenpunkte [Quantum Dots], *Spektrum Wiss.* (1993) 52-57.

[27] *Engineering, a Small World; From Atomic Manipulation to Microfabrication Science;* Vol. 254, 1991, pp. 1300-1342.

[28] Weisburch, C., Vinter, B., *Quantum Semiconductor Structures: Fundamentals and Applications;* New York: Academic Press, 1991.

[29] Kirk, W. P., Reed, M. A., *Nanostructures and Mesoscopic Systems;* New York: Academic Press, 1992.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

2 Approach to Microsystem Design

B. KLOECK, M. DEGRAUWE, CSEM Centre Suisse d'Electronique et de Microtechnique SA, Neuchâtel, Switzerland

Contents

2.1 Introduction 24
2.2 Microsystem Technologies 24
2.3 Market Demands 27
2.4 Efficient System Design to Reduce Time-to-Market 29
2.5 Examples of Advanced Beside-IC and In-IC Microsystems 30
2.5.1 Absolute Pressure Sensor 31
2.5.2 High-Resolution Accelerometer 36
2.5.3 High-Density, High-Speed Magnetotronic Actuator 42
2.6 Conclusions and Future Trends 47
2.7 Acknowledgements 49
2.8 References 49

2.1 Introduction

Microsystems are devices and complex functional units based on various microtechniques. They combine diverse functions relying on mechanical, electromagnetic, optical, biochemical, electronic and other principles. Usually one or more of the microsystem components are batch produced, using processes such as micromachining, photolithography, thin-film deposition, and etching. Many of these techniques originate from silicon technology. Owing to the multidisplinarity, microsystem design requires a well adapted design methodology.

This chapter starts with an overview of the three main classes of microsystem technologies and their characteristics, selection criteria, and application areas. In Section 2.3, the market demands for these technologies are discussed. Section 2.4 treats the design methodology, which must be adapted to the specific nature of microsystems, in order to reduce the time-to-market for a new product development. Section 2.5 gives three typical examples of microsystems, in order to illustrate the main points developed in this chapter. Section 2.6 takes a brief look into the future and summarizes the conclusions.

2.2 Microsystem Technologies

Over the last decade, a variety of technologies and devices have been developed for a wide range of microsystem applications. For a meaningful classification, two criteria can be used:

- according to the operation principle of the sensor or actuator;
- according to the fabrication technology of the microsystem with respect to sensor/actuator and electronics.

For the operation principle, four classes are distinguished corresponding to microsystems based on mechanical, magnetic, optical, or biochemical phenomena. The relationship between sensor or actuator on the one hand and electronics on the other is expressed as Beside-IC, In-IC or Above-IC. The definitions are as follows:

- *Beside-IC* The sensor or actuator is not on the same substrate as the associated signal processing electronics. Currently available devices of this type include bulk-micromachined accelerometers and pressure sensors and diffractive optical elements (DOE).
- *In-IC* The sensor or actuator and the associated electronics are built on the same substrate and simultaneously. Usually, an existing bipolar, CMOS or BiCMOS process is used as the starting point. Some intermediate processing steps or layers are modified or added.
- *Above-IC* The chip containing the associated electronics serves as a substrate for the micro-structures, which are built above the passivation layer of the integrated circuit. In other words, the electronic circuit is completely finished before the sensor or actuator is added in a number of post-processing steps.

The Beside-IC class is characterized by:

- relatively simple technology (less than five masks for the sensor or actuator process);
- high yield;
- no expensive production equipment is required;
- high assembly cost;
- potentially highly accurate devices (low noise);
- devices of all four operation principles are being designed and produced.

The In-IC class is characterized by:

- complex technology (more than 16 masks);
- low yield;
- expensive production equipment;
- low assembly cost;
- lower accuracy, since mechanical noise of the small masses becomes important;
- simultaneous integration of several devices is possible;
- demonstrators exist for all four operation principles.

The Above-IC class is characterized by:

- relatively simple technology (less than five masks for the sensor or actuator process);
- medium yield;
- no expensive production equipment is required;
- low assembly cost;
- medium-accuracy devices;
- simultaneous integration of several devices is possible;
- the signal conditioning circuit can be implemented in the best technology;
- only mecatronic devices have been realized up to now.

An important design parameter that is related to the choice of technology is the accuracy that can be obtained for mechanical microsystems. One of the fundamental limitations of the accuracy of a system is the noise level. For an electromechanical system, there is an electrical and a mechanical contribution to the signal-to-noise ratio:

$$\frac{1}{SNR_{\text{system}}^2} = \frac{1}{SNR_{\text{elec}}^2} + \frac{1}{SNR_{\text{mech}}^2} \tag{2-1}$$

For convenience, SNR is defined here as the direct ratio of the noise voltages, and not as the logarithm of the noise powers.

Using accelerometer design as a typical model, the mechanical signal-to-noise ratio can be written as [1]

$$SNR_{\text{mech}} = \frac{m \cdot a}{2} \sqrt{\frac{\pi \cdot a}{kT \cdot K}} \tag{2-2}$$

where m is the mass of the movable mechanical component, a is the acceleration range, α is the ratio of the bandwith of the mechanical component to the bandwith of the complete system and K is the spring constant of the suspensions.

The important conclusion that can be drawn from Equation (2-2) is that the *SNR*, and therefore also the accuracy that can be attained, decrease with decreasing mass of the mechanical component. Beside-IC and, to a lesser extent In-IC technologies make use of silicon bulk micromachining, where the dimensions of the movable parts are in the millimeter range. For these devices, the electronic noise is dominating. In-IC and Above-IC systems, on the other hand, exploit the advantages of silicon surface micromachining. The vertical dimensions are one to two orders of magnitude smaller than for bulk micromachining, and the lateral dimensions about one order of magnitude smaller.

Table 2-1 shows some typical values for dimensions and masses which are generally involved for microsystems of the three categories. Obviously, In-IC technology with polysilicon can work with very small mechanical parts, but the benefit of reduced dimensions is paid for by an increased noise level, and therefore lower accuracy. The Beside-IC category usually works with large masses, and a very high accuracy can be achieved. Above-IC covers both mass ranges, without reaching the extremes, and the accuracy will typically be between those of the other two. In conclusion, accuracy is an important criterion which has to be taken into account when selecting the appropriate technology for microsystem design.

Table 2-1. Mass range for the three microsystem technology categories (typical values).

Category	Technology	Surface area/mm^2	Thickness/μm	Density/kg/m^3	Relative mass
Beside-IC	Bulk micromachining (monosilicon)	1–10	10–380	2332	10^3–$4 \cdot 10^5$
In-IC	Surface micromachining (polysilicion)	0.01–1	1	2332	1–10^2
Above-IC	Galvanic deposition (eg, FeNi)	0.01–1	5–40	7800	$2 \cdot 10^1$–10^4

An important techno-economical parameter that is determined by the choice of technology is the fabrication yield of the devices. Obviously, the Beside-IC approach results in the highest yield, since the devices are fabricated separately. For Above-IC technologies, the yield of the two processes has to be multiplied, which reduces the overall yield. Still, since well established technologies can be used, at least for the electronic part, a reasonable yield can be achieved. For the In-IC category, however, the yield is much lower than mere multiplication of two individual yields, owing to the greatly increased complexity of the fabrication process.

The above considerations are important guidelines for the evaluation of the economic feasability of the different technologies and their market segments. This will be discussed in the next section.

2.3 Market Demands

Markets According to Technology Segmentation

In the previous section, a classification of microsystems was presented according to the applied technological approach. Here, the market implications for the three categories are briefly discussed.

In-IC microsystems are economically feasable only for mass markets, where the huge costs of developing a new technology and of pushing the production yield up to a reasonable level can be spread over a large number of devices. Some specific niche markets, which are able to bear extremely expensive devices, might also profit from In-IC technology.

Beside-IC technology is an appropriate choice for high-accuracy applications and for small markets, in particular when an extensive library of various inter-compatible microsystem modules is available, where the components can efficiently be combined into an effective microsystem using a top-down system design methodology. An intelligent design strategy results in a very short time-to-market.

The Above-IC approach has the advantage of combining most of the benefits of the other two categories. The time-to-market is situated somewhere between those for Beside- and In-IC technologies.

Silicon Demands

To estimate the volume ranges which are currently important for the microsystem community, the real industrial demands treated by a four-inch microsystem service centre during the past few years were evaluated. Figure 2-1 shows the relationship between the annual number of microsystems and the annual number of wafers for various industrial demands. The surface of the silicon chip is taken as a parameter. The number of microsystems ranges from ten to more than one million, and the number of wafers from 20 to 20000.

Some interesting conclusions concerning silicon needs of the microsystem business can be derived from Figure 2-1:

- A minimum of 200 wafers per year seems necessary to be able to guarantee a stable process.
- Only a few applications need much more than 200 wafers per year. The largest demand of 1 million microsystems per year had an annual need of 20000 wafers.
- The sum of all actual demands would not use more than 25% of the capacity of a normal semiconductor fabrication unit.

In spite of the relatively small wafer count, the number of microsystems which need to be assembled, packaged, and tested can be huge. These activities represent a large part of the microsystem cost, especially since they are not yet automated to any degree comparable to the microelectronic industry.

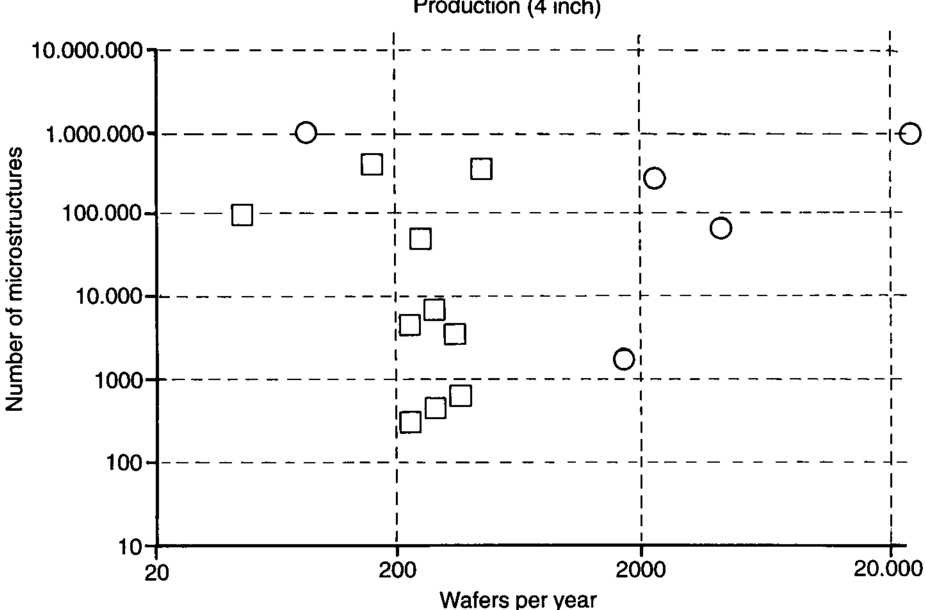


Figure 2-1. Annual number of microstructures and annual number of wafers for various recent industrial demands treated by a 4-inch microsystem service centre.

Companies and Markets

Table 2-2 shows the relationship between the size of the companies and the microsystem demands of Figure 2-1. The large companies are mostly looking for cost-effective replacement of existing products. They are able to bear the investments required for customized microsystems in Above-IC or In-IC technology. The medium-sized companies seem to be most innovative. They can effectively profit from Beside-IC technology to create new products which incorporate existing high-tech components. The small companies want to buy "standard" microsystems. They cannot afford new developments, and are mainly interested in adapting existing systems to their specific requirements.

Table 2-2. Microsystem demand and company size.

Annual number of microsystems / Company size	1000	10000	100000	1000000
Small	×	×		
Medium	×	×	×	
Large			×	×

The markets which were addressed with the demands of Figure 2-1 are as follows:

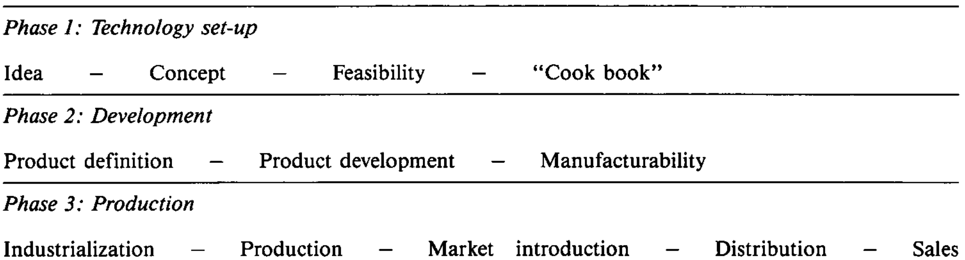
- consumer electronics;
- automotive;
- telecommunication;
- industrial control;
- instrumentation;
- biomedical.

2.4 Efficient System Design to Reduce Time-to-Market

Owing to their small size, microsystems are within reach of, and can be very profitable for, medium-sized enterprises. However, owing to its multifunctionality, microsystem design requires the same structural approach as for large complex systems, including project management methods such as concurrent engineering, usually in practice only in large companies. To deal with this contradiction, the microsystem community will need design houses to develop the right system design strategies, adapted to microsystems, and which are able to define the design flow with the shortest possible time-to-market for every individual product, taking into account issues such as technological risk and market price.

Since the discipline is relatively new, a standard methodology has not been generally adopted yet, but any systematic development cycle should go through the subsequent stages described in Table 2-3. Three main phases take place: technology set-up, development and production.

Table 2-3. Life cycle of a new product.



Phase 1: Technology set-up
Idea – Concept – Feasibility – "Cook book"
Phase 2: Development
Product definition – Product development – Manufacturability
Phase 3: Production
Industrialization – Production – Market introduction – Distribution – Sales

The technology set-up phase starts from the basic idea on which the new product will be built. This stage is concerned not only with the idea itself, but also with its originality and validity. Therefore, a literature and patent search will be done at this point. Further, all basic ideas of what will be needed for the new product are generated and listed, including issues such as packaging, assembly, manufacturability, testability, and reliability.

Second, a viable concept will be established and worked out in more detail. The idea is translated into a mathematical model, basically a set of equations describing all different

aspects of the behavior of the system, and their interactions. Simulations are carried out to verify the main objectives, and demonstrators are designed which will allow one to verify the most critical specifications and performances.

In the third stage of the first phase, the feasibility is proven by the construction of demonstrators. Extensive measurements on the demonstrators and comparison with the mathematical predictions will allow one to fine-tune the system model. Any new phenomenon observed at this stage must be explained and introduced in the model, if necessary. Statistics on a sufficient number of samples will give indications on the reproducibility of the critical parameters.

The last stage of the set-up phase, referred to as the "cook book", produces all documents required to start the actual product development. A list of design rules is established, which link the model equations to the customers' specifications. Sensitivities of the various system specifications to specific design parameters are described, and all logical, physical, and practical limitations are identified and listed.

The technology set-up phase has been described in some detail, since it has a significant influence on the time-to-market for new products. Indeed, a good design house will have a large library of "cook books" available, covering various generic microsystems and microsystem components, built up over the years. When a customer arrives with a specific demand, the set-up phase, and thus the total development time and risk, can be reduced dramatically.

The second phase of product innovation, as shown in Table 2-3, includes product definition, the actual product development based on the results and the experience gained in the first phase, and an in-depth study of all aspects of the manufacturability. The third phase, then, proceeds with industrialization and production. Also for these phases, a well structured design framework and methodology have to be put into practice, in order to guarantee efficient and effective execution of the project.

Among the different technology types described before, that with the shortest time-to-market potential is the Beside-IC technology. Indeed, microsystem product innovation is most efficient and cost-effective when all high-tech building blocks are available. The technology set-up phase is no longer needed, and efforts can be focused on an intelligent system concept, where all individual modules and components are tuned to the required system specifications. The importance of this approach will be illustrated in the next section, using a typical example.

2.5 Examples of Advanced Beside-IC and In-IC Microsystems

In support of the ideas put forward in the previous sections, three typical examples are given here which represent three different cases of every-day microsystem practice. All three concern recent microsystem innovations.

The first is a pressure sensor in Beside-IC technology. For this system, all critical components were available, and the innovation came from a well adapted system design methodology, which includes the definition of the most appropriate measurement principle, the selection of adequate components, and the calculation of all system parameters by a system-level simulation. Among the three examples discussed in this section, this one had the shortest development time.

The second example, a high-resolution accelerometer system, supports the claim that very high precision can be reached with Beside-IC technology. For this system, the measurement cell was available, but a new interface circuit had to be developed, adapted to the high precision requirement. This example also clearly shows that it is absolutely imperative to integrate both mechanical and electronic functions in one system simulation.

The third example is an In-IC electromagnetic actuator. The entire system had to be designed and developed from scratch, since an entirely new approach was devised to achieve unprecedented performance. The work included an extensive and innovative technological development.

2.5.1 Absolute Pressure Sensor

Silicon absolute pressure sensors are widely used in automotive applications for barometers and for altimeters. The established technology for these mass-produced sensors is based on piezoresistive sensing elements coupled to interface circuitry, including trimming capabilities for offset, span, and temperature compensation. Packaging is usually done using a silicone dye coating to protect the silicon pressure-sensing chip which is mounted on a plastic case [2]. Even though these technologies have reached a high degree of maturity, the manufacturing cost of these devices is still fairly high and is decreasing only slowly.

Alternative concepts, such as capacitive sensors, have attracted significant interest in recent years in order to obtain a sensible cost reduction for the same performance.

An important issue in piezoresistive pressure sensors is the need to adjust the temperature compensation on each sensor. It has been shown that capacitive pressure sensors have low and reproducible temperature coefficients [3, 4], which is a prerequisite for sensor concepts without temperature compensation.

New surface micromachining technologies allow for a significant reduction of the sensor chip size, essentially because the large chip surface necessary for the ⟨111⟩ sidewalls of the rear-side membrane etching can be avoided. These technologies are often based on a sacrificial layer technology using polysilicon or monocristalline silicon membranes [5].

The goal of this work was to design a low-cost pressure sensor microsystem based on today's known and proven technologies. The design work included the choice of the sensor technology, the interface circuit, a suitable packaging, and an efficient trimming concept. The objective was to obtain the same performance as the commercially available piezoresistive pressure transducers for automotive applications, eg, for measuring the absolute manifold pressure, at a significantly reduced manufacturing cost.

Sensor Design

An approach based on a capacitive measurement scheme was chosen for two reasons. First, capacitive measuring cells can be made with a low-cost process and with high stability. Capacitive sensors have very low temperature coefficients. This is an advantage over piezoresistive sensors, where the physical quantities, the resistances, and the piezoresistive coefficients involved have basically high temperature coefficients requiring highly balanced circuitry to achieve good results. This leads to a significant trimming effort especially for the temperature effects.

A second reason for choosing a capacitive approach is the ease of implementing high-stability interface circuits in CMOS technology for capacitor bridges [6].

The intrinsic large but predictable nonlinearity used to be a basic drawback of capacitive sensors. To deal with this problem, an efficient linearization scheme has been implemented in a self-balancing capacitor bridge [7]. Therefore, nonlinearity has been ruled out as a basic limitation of capacitive sensors.

The stability of silicon-based capacitive sensing cells can be severely degraded by packaging-induced stress. One way to solve this problem is to introduce mechanical decoupling schemes [4, 8, 9]. These solutions require a relatively large additional chip surface, which is not compatible with a low-cost sensor. As an alternative, a soft-mounting technique is used in this work to decouple mechanically the sensor chip from the substrate.

Different technological approaches with respect to chip size, maturity of the technology, and expected performance have been evaluated. Finally, the sensor concept as shown schematically in Figure 2-2 has been chosen. This sensor is composed of a structured silicon baseplate and a pressure-sensing plate, which together form the sensing capacitor. The total capacitor is composed of the central active part, varying with pressure, and the fixed part formed by the supporting rim. Considering the minimum width of the rim to assure a hermetic seal, a ratio of active to fixed capacitance of 1.4 could be obtained. The recess around the central active capacitor part etched into the baseplate allows a significant reduction of the fixed or little varying capacitance and an enhancement of the pressure reference cavity volume for good long-term stability.

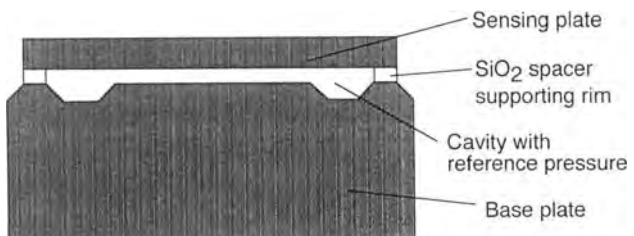


Figure 2-2. Schematic cross section of the capacitive silicon sensor chip (not to scale).

The manufacturing technology is based on a three-mask process using standard wet silicon anisotropic etching and silicon fusion bonding. A similar process was originally designed for silicon microphones [10]. The sensor chip has a surface area of 4.18 mm^2.

In accelerometer production, a cavity pressure below 100 Pa is achieved with high yield. The long-term stability of this pressure has been proven by systematically monitoring the device's dynamic behavior. Therefore, the same technology has been applied for pressure sensors. A pressure below 100 Pa is largely sufficient since the error due to the variation of the internal pressure with temperature is 30 Pa for a 100 °C temperature range. Thus, its contribution to the error budget is negligible.

Linearization Scheme

In a good approximation, the sensor can be described as two capacitors in parallel, $C_s(p)$ representing the central part of the sensor varying with pressure, and C_{sfix} describing the fixed capacitor of the rim.

With a self-balancing bridge as shown in Figure 2-3, a linear relationship can be obtained between the pressure and the output voltage. This bridge is an extension of the capacitor bridge cited in Ref. 6. It is based on a feedback loop adjusting the driving voltages to the different branches of the bridge in such a way that the net charge on the central electrode is zero. The output voltage is then given by

$$V_M = V_0 \frac{C_s(p) + C_{sfix} - C_{Ref} - C_{cl} + C_{c2}}{C_1(p) + C_{sfix} + C_{Ref} - C_{cl} - C_{c2}} \tag{2-3}$$

where V_0 is half the supply voltage. One can now choose the capacitors C_{Ref}, C_{cl} and C_{c2} in such a way that the output voltage is proportional to the pressure. This scheme corrects the basic nonlinearity of the capacitive effect and second-order effects such as the curvature of the membrane. From the above equation one notes that changing C_{Ref} and C_c will also change the offset and the span. This means that the sensor can be fully adjusted using only two parameters, one constant in the numerator, $C_{Ref} + C_{cl} - C_{c2}$, and one in the denominator, $C_{Ref} - C_{cl} - C_{c2}$.

Optimum linearity compensation can only be obtained for one set of fixed capacitors. However, for most low-cost applications, the error budget for nonlinearity is sufficiently large to trim of offset and span with the two constants. Evidently, this approach implies that the sensor design and the manufacturing tolerances have to be carefully optimized and monitored.

Since the circuit is implemented in CMOS technology, the capacitors C_{Ref}, C_{cl}, and C_{c2} are readily integrated using arrays of capacitors and suitable switches. Trimming is done at the final testing stage. The information of the particular combination of capacitors to be used is easily stored either in non-volatile or fuse-type memories.

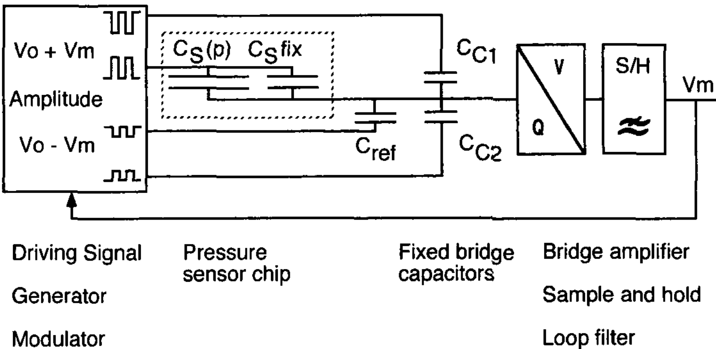


Figure 2-3. Schematic diagram of the self-balancing bridge configuration.

Packaging

A key feature of the packaging is the mounting technology of the silicon chip. The basic idea of this new "soft" packaging concept is to use materials with low Young's modulus for mounting and protection. A cross-section of the packaged sensor is shown in Figure 2-4. Suitable spacers are used to ensure proper distance from the substrate and to provide the

mechanical stability necessary for the wire bonding process. Owing to the low Young's modulus of the glue and the well defined geometry obtained by the spacer, the sensor chip is efficiently mechanically decoupled from the substrate.

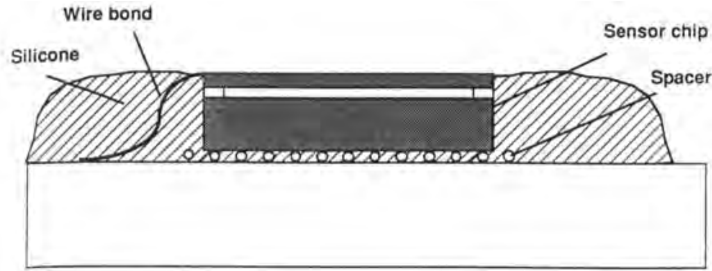


Figure 2-4. Schematic cross section of the mounted silicon chip.

The Microsystem Building Blocks

Capacitive pressure sensor ships have been mounted on TO18 headers with the soft mounting method. A mounted chip is shown in Figure 2-5. A typical curve of the temperature behavior is shown in Figure 2-6. A variation of less than 0.4% of the span is observed over a temperature range of 100 °C, with a parabolic-type behavior. The physical mechanisms giving rise to this low residual temperature effect have not been identified yet. However, the effects are sufficiently small that no temperature compensation is required.

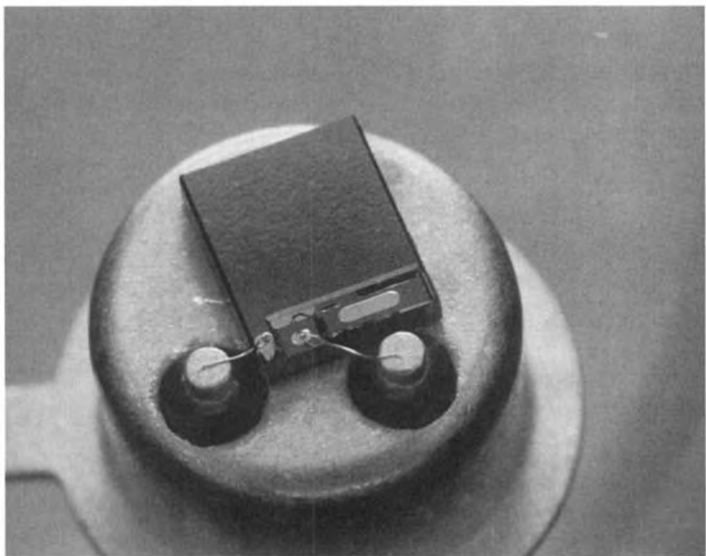


Figure 2-5. Prototype capacitive sensor chip mounted on a TO18 header. The chip protection is omitted for illustration purposes.

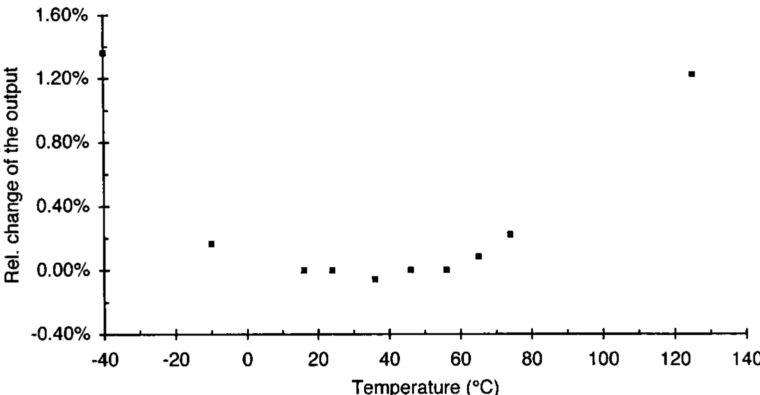


Figure 2-6. Variation of the capacitive silicon chip capacitor with temperature at a fixed pressure of 1 bar. The variation is expressed in terms of the corresponding pressure output of the sensor.

The electronics measurement scheme has been tested, using a previously developed self-balancing bridge ASIC and adding some external electronics components in order to implement the linearization scheme. Typical response curves and their deviation from a straight-line fit are shown in Figures 2-7 and 2-8.

Clearly, this measurement scheme allows to obtain a good linearity for the capacitive sensor output. One further notes from Figures 2-7 and 2-8 that a nonlinearity error budget of 1% allows one to adjust the sensitivity by as much as ±20% by digital trimming of the compensation capacitors.

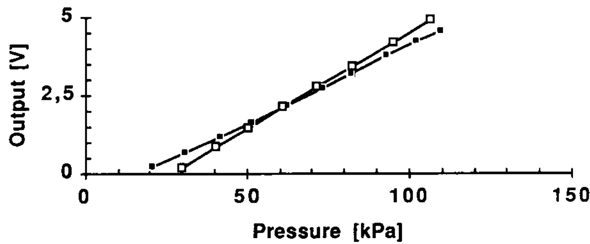


Figure 2-7.
Output signal of the complete prototype pressure sensor. The output voltage of the bridge (Figure 3-2) is amplified by a gain of 4. The two curves correspond to different values of the compensation capacitors.

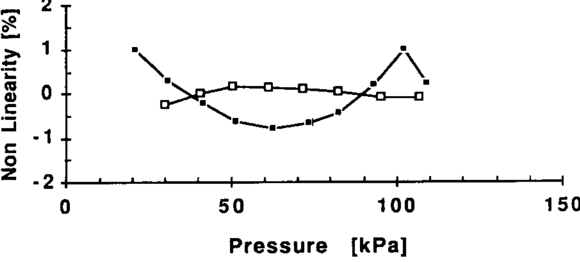


Figure 2-8.
Nonlinearity of the curves represented in Figure 3-6.
(□) Near optimum nonlinearity;
(■) compensation allowing 1% nonlinearity.

Conclusions

This work was based on a defined objective of a low-cost absolute pressure sensor. Using available microsystem components, it was possible to develop prototypes of a low-cost pressure sensor in a very short time. At the start of the project, no particular a priori choices were made for the technology approaches other than that they have to be available in the form of well characterized building blocks. A top-down microsystem design methodology allowed the identification of the key technology blocks required for this project:

- bulk silicon micromachining;
- low-pressure fusion bonding;
- soft-mounting technology;
- self-balancing capacitor bridge interface circuitry;
- digital trimming using arrays of capacitors.

All these technologies were well established and available. Low cost could be achieved by:

- a small silicon pressure sensor chip (4.12 mm^2);
- simple silicon technology (three-mask process);
- low-cost mounting;
- interface circuit of low complexity;
- digitally trimming two parameters at two pressures and at one temperature.

Measurements of the complete system showed the feasibility of the above-described concepts.

2.5.2 High-Resolution Accelerometer

Bulk silicon capacitive accelerometers have been known for several years. These accelerometers are often based on a movable plate suspended by a flexure bar between two fixed electrodes. The acceleration signal is derived either by measuring the plate deflection with a capacitor bridge [11] or by working in a force-balancing mode [12, 13]. These devices have already led to commercially available products for industrial and automotive applications. Results from measurements of microgravity vibrations on spacecraft have been reported more recently [14].

The work described here has improved the performance, mainly the stability, of the silicon accelerometers in order to make them suitable for high-precision applications such as inertial navigation and guidance or precision tilt measurements. The key performances are a working range of ± 1.5 g, an offset stability after thermal compensation of 10 μg, a temperature coefficient of the sensitivity of 100 ppm/K, a nonlinearity smaller than 1‰, a noise floor of 1 $\mu g/\sqrt{Hz}$ and a bandwidth of 300 Hz. Previously, this type of performance could only be obtained with the classical electromagnetic servoaccelerometers, which have drawbacks in fabrication cost, power consumption, and shock resistance.

Design of the Measurement System

The starting point of the design was an existing silicon capacitive accelerometer chip, called the measurement cell [11]. The acceleration is sensed by a spring-mass system in the form of a movable plate suspended by a flexure bar. Fixed electrodes on either side of the plate are used for capacitive detection of the plate position and for the application of electrostatic forces. A defined gas pressure surrounding the plate allows for a controlled mechanical damping.

For system design considerations, the measurement cell is described as a three-plate double capacitor with its middle plate moving according to a one degree of freedom spring-mass model.

A fundamental choice in the system design concerns the mode of operation of the accelerometer. In a trade-off study, the plate displacement measurement mode based on a balanced capacitance bridge [15] and different force balancing concepts have been compared. Two effects have been identified allowing force-balanced systems to obtain a better performance for offset stability and for linearity than balanced capacitor bridge systems. First, the mechanical offset drift with temperature and time, expressed as the variation of the sensor capacitance difference for zero acceleration, is almost independent of the spring constant. Drift is due to small deformations of the sensor structure, which is transmitted to the movable electrode through the spring, and causes small displacements of the movable electrode, which is then measured as a small capacitance drift. Typically, the spring constant can be chosen ten times smaller in a force-balanced system than in the plate displacement mode. Therefore, a given mechanical offset drift leads to an error in the output signal which is roughly ten times smaller for the force-balancing mode than the plate displacement mode. Second, the nonlinearity in the plate displacement mode is mainly due non-parallel plate displacement and parasitic capacitors in parallel with the working capacitors. As described previously for the pressure sensor, this working mode needs a linearization scheme to reduce this error by an order of magnitude. The force-balancing mode, however, is intrinsically linear when the loop working point is properly adjusted. For these two reasons, a force-balancing mode has been chosen to respond to the very high precision requirement.

From the different force balancing principles, such as based on variable voltages or pulsed systems, a measurement scheme called an "electromechanical delta-sigma converter" has been chosen. In this scheme, the balancing force is generated by electrostatic force pulses of constant amplitude and duration, applied on either side of the plate. A simplified schematic diagram is shown in Figure 2-9. The acceleration measurement system is composed of the measurement cell, represented by the capacitors C_1 and C_2, and the interface circuit, operating in the following way: in each measurement sequence, a signal proportional to the capacitance difference $C_1 - C_2$, thus to the plate position, is fed to an integrator. Depending on the sign of the integrator output signal, a force pulse is applied on one side of the movable plate. This system forms a closed loop maintaining the movable plate near a fixed position, called the working point. The output signal is then given by the pulse density that is applied to one side of the plate. Hence this system, which has a strong analogy to delta-sigma A/D converters, has a digital output in the form of a bit-stream.

The electronics implementation of this measurement prinicple is also shown in Figure 2-9. Acceleration forces up to 2 g can be compensated for with a single 5 V power supply. Comparing this electromechanical delta-sigma converter with a conventional second-order delta-sigma A/D converter [16], three major differences can be observed:

- The electromechanical second-order delta-sigma converter is composed of one mechanical and one electrical integrator, instead of two electrical integrators.
- The electromechanical converter combines continuous time (sensor) and discrete time (electronic) blocks, compared with an all-sampled data system.
- The electromechanical converter combines one force feedback path (electrostatic force) and one electronic feedback path, instead of two electronic feedback paths.

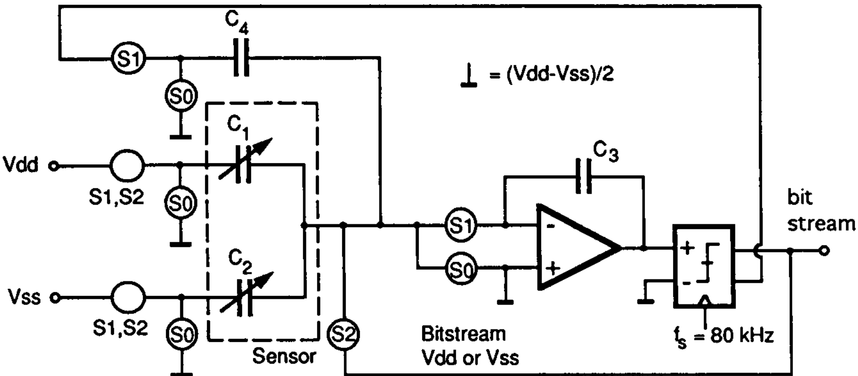


Figure 2-9. Simplified electrical diagram of the electromechanical delta-sigma converter.

In a simplified picture, the output signal is the ratio of the pulses applied on one plate to the total number of pulses. However, for correct signal conditioning, the high-frequency bit-stream has to be fed to a digital low-pass and decimation filter in order to obtain typically 15 or 16 bit words at low frequency.

With respect to previously reported similar schemes [13], the solution described here shows the following advantages:

- The sensor replaces the first integrator in a second-order delta-sigma converter; therefore, the stability of the system is guaranteed by choosing electrical parameters of the electronic feedback loop, and no additional filter in the forward path is needed.
- The mechanical cut-off frequency of the sensor is not critical and can be chosen one order of magnitude smaller than the closed loop system bandwidth (here 300 Hz).
- The electronic integrator in the electromechanical delta-sigma converter assures low residual excitations of the seismic mass, which is a major requirement for high linearity.
- The electronic integrator prevents a threshold effect at low level input signals, which would create dead zones in the acceleration transfer function [6].

Design of the Capacitive Measurement Cell and its Packaging

The high precision requirement of the system also imposes high demands on the offset stability of the sensor cell. Therefore, a chip was selected which includes a mechanical stress decoupling structure and, moreover, an improved mounting scheme was designed.

In a conventional capacitive silicon sensor, the mass-spring system and the electrodes form a hermetically sealed block which is glued with epoxy to a substrate. A key requirement for good stability is to avoid any packaging-induced stress. Finite-element analysis and experimental work have shown that any kind of mounting of this block by epoxy gluing or soldering leads to deformations of the measurement block, not compatible with the stability aimed at. Therefore, a previously designed silicon chip with a built-in stress relief structure was selected for this application. An exploded view of the silicon chip is shown in Figure 2-10. It is composed of three plates of silicon, separated by thin dielectric layers. The chip is composed of an inner part forming a hermetically sealed block containing the spring-mass system, and an outer mounting frame. The inner block is fixed to the mounting frame through a stress relief bar. This structure allows mounting of the chip on the outer frame using conventional epoxy technology. The unavoidable stress induced to the frame by this type of mounting does not affect the inner measurement cell. The stress relief bar is small enough to avoid packaging stress induced on one side being transmitted to the measurement block and large enough for the system to withstand overloads higher than 2000 *g*. This has been confirmed by detailed finite-element simulations.

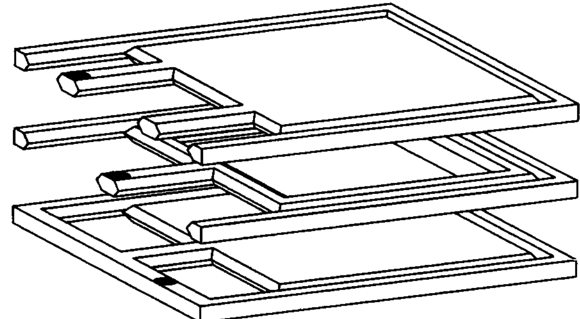


Figure 2-10. Schematic diagram of the three silicon plates.

Some relevant characteristics of the accelerometer cell are given in Table 2-4. Owing to the packaging stress relief structure, the offset drift related to the open-loop working range is one order of magnitude lower than sensors without this structure. The temperature coefficient of the offset of 12 μ*g*/K is sufficiently low in order to obtain an overall stability of 10 μ*g* after temperature compensation.

Table 2-4. Open-loop characteristics of the silicon capacitive measurement cells.

Open loop sensitivity	18.6(±0.7)	pF/*g*
Mechanical offset at 0 *g*	38(±22)	m*g*
Temperature coefficients:		
Offset	12(±4)	μ*g*/K
Sensitivity	−32(±22)	ppm/K
Capacitance (sum of both capacitors at 0 *g*)	<10 ppm/K	

Measurements on earlier chip versions have shown that the cell is sensitive to its surrounding pressure. It slightly deforms the fixed electrodes, leading to a change in the plate distance, and inducing also an effect on the offset. Furthermore, humidity affects the stability and the in-

sulation resistance of stray capacitors, eg, by introducing leakage currents. Therefore, a ceramic package sealed at reduced pressure and having a very low dew point was chosen. The interface circuit is mounted in the same package in order to favor miniaturization and to avoid stray capacitances between the sensor and the interface circuit. The complete package is shown in Figure 2-11.

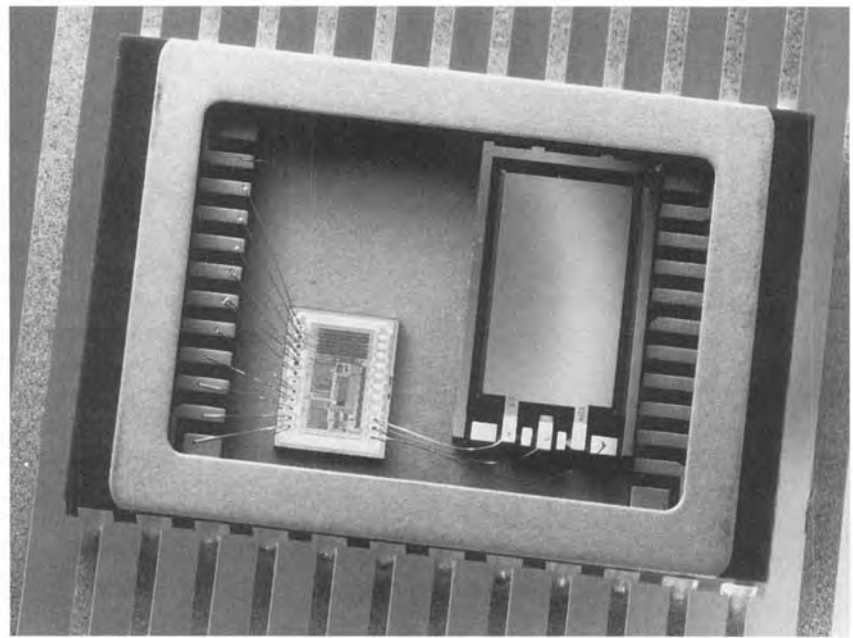


Figure 2-11. Complete packaged accelerometer system including the silicon capacitive measurement cell and the interface circuit.

Simulation of the Microsystem

The performance of the accelerometer system, described in the previous section, has been extensively simulated. Both the electrical and the mechanical aspects have been simulated simultaneously using a mixed-mode electrical simulator and a block-diagram oriented software for the simulation of dynamic systems. Particular attention has been given to simulate dynamic loop behavior, loop stability, and the analysis of the parameters that influence linearity and offset stability. An example of a simulated system response is shown in Figure 2-12. With such simulations, the mechanical deflection of the movable plate was studied in normal operation and after overload conditions.

The analysis of the different simulation results has allowed the design activity to be guided and the system parameters to be fixed. As an example, the overall offset stability is affected by the electrical offset of the integrator and by the mechanical stability of the sensing cell. Therefore, auto-zeroing techniques were chosen for the integrator and the improved mechanical design described above has been introduced. Another key issue studied in depth

is the nonlinearity. It is dominated by the choice of the working point, the position at which the movable plate is stabilized. Ideally, the plate should be perfectly centered. Any deviation from this position leads to a second-order nonlinearity. Consequently, a trimming capability for the working point has been introduced in the circuit. Simulations showed that a nonlinearity better than 0.1% could be obtained.

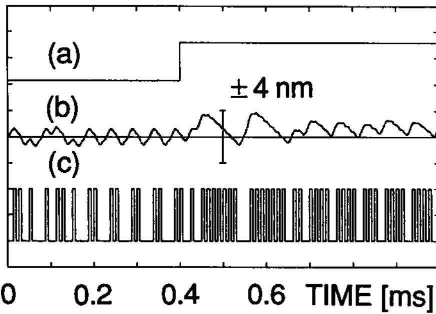


Figure 2-12.
(a) Simulated response of the system to an acceleration step from 0 to 1 *g*; (b) associated plate displacement; (c) force pulses applied to one side of the plate.

System Performance

The measured static and dynamic responses of the microsystem equipped with a the new CMOS ASIC are shown in Figures 2-13 and 2-14. The static response was obtained by measuring the bit stream with frequency counter. For the evaluation of the dynamic response, the bit stream was converted into an analog signal using an RC low-pass filter.

The static response agrees well with the circuit simulation. As expected, the dynamic response shows first-order behavior with a cut-off frequency of 98 Hz and a roll-off of 20 dB/decade. This bandwidth is much larger than the open-loop bandwidth of the measurement cell, which is near 3 Hz.

In agreement with the simulations, a very low nonlinearity of less than 0.1% was realized.

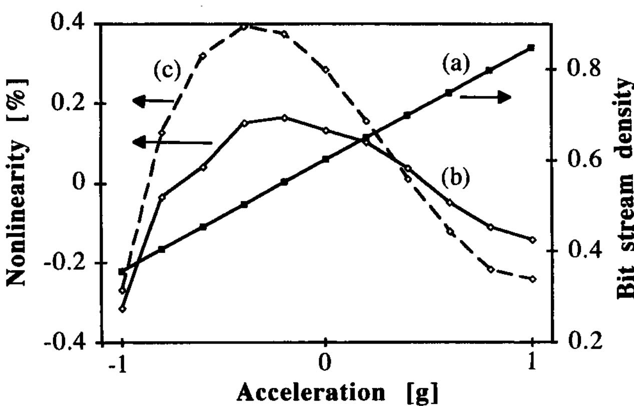


Figure 2-13. (a) Static response and (b) nonlinearity measured with the preliminary circuit and (c) calculated nonlinearity with the corresponding parameters.

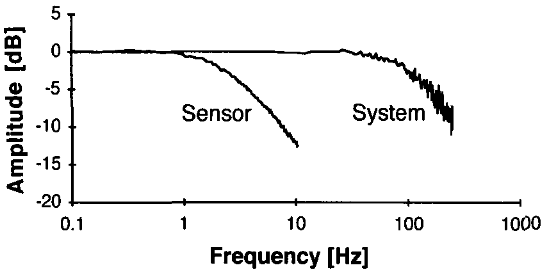


Figure 2-14.
Measured frequency response of the servoaccelerometer (sensor) and the open loop silicon capacitive cell (system).

Conclusions

A top-down approach has been applied to the design of a high-performance accelerometer microsystem with a digital output. The result is an extremely compact packaged accelerometer based on only one silicon electromechanical chip and an interface circuit of low complexity.

The accelerometer module, with a size of 15.3 mm × 10.2 mm × 1.2 mm, is based on an "electromechanical delta-sigma converter" including a silicon capacitive measurement cell and an interface circuit (ASIC), both packaged in a hermetic ceramic housing. The working range is ± 1.5 g, the offset stability 10 μg and the nonlinearity 0.1%. The silicon capacitive measurement cell has a built-in structure to reduce packaging stress. The chip size is 6.4 mm × 3.9 mm × 1.2 mm. The sensing cells have temperature coefficients of offset of 12 μg/K and of sensitivity of -32 ppm/K, proving the effectiveness of the stress relief structure.

The overall system design approach, based largely on simulation work, has allowed the simultaneous optimization of the mechanical and electrical aspects. Key issues, such as mechanical stability and the working point of the loop, have been identified, and optimized solutions were implemented. It has been shown that the objectives can be reached with this approach.

2.5.3 High-Density, High-Speed Magnetotronic Actuator

This section describes the different steps of the development of an In-IC magnetic microsystem for a magnetographic printing head. This development illustrates the life cycle of a new product as discussed in Section 2.4. First, the initial realization of the most critical part is described. The successful completion of this phase allowed the fabrication of a full prototype. Field tests with this device have allowed the definition of the final product. Finally, the development phase of the product presently in production is presented.

The principle of magnetographic printing is shown in Figure 2-15. A magnetic image is written on a rotating drum. The image pixels are written line by line by electromagnetic writing poles which are in contact with the drum, and which can be addressed individually.

Density in magnetographic printing depends on the size of the recording elements which form the magnetographic head. Conventional technologies used for densities up to 240 dpi currently rely on the assembly of single miniature devices [17]. In order to exceed this limit,

heads with integrated recording elements are required. In this work, a 480 dpi head was integrated on 1-inch wide silicon chips. The devices are composed of high-density arrays of recording electromagnets which are sequentially addressed by on-chip diode matrices. Fabrication involves the combination of standard electronic processes (for the diode matrices and the electrical circuits), microgalvanic technologies (for the magnetic and electrical circuits), and silicon micromachining (for precise location of the magnetic circuits).

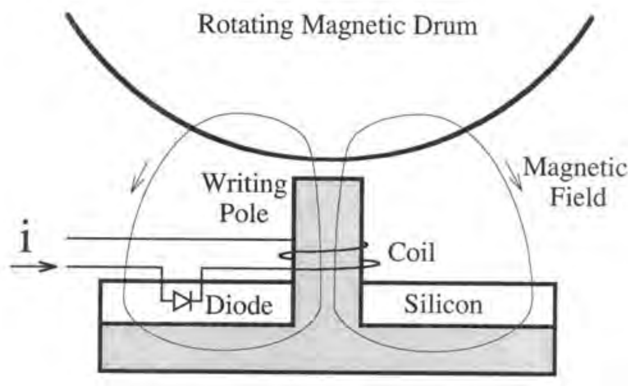


Figure 2-15. Principle of the magnetographic silicon head.

Fabrication of the Most Critical Part

The realization of magnetic elements on silicon chips was the most critical part of this development. In a first step, the technology was designed and developed with a simplified device consisting of 480 electromagnets integrated on a silicon chip without the addressing diode matrix.

The structure of an electromagnet is based on the perpendicular field magnetic printhead [18, 19]. It consists of a flat multi-turn coil surrounded by a magnetic circuit which is composed of a magnetic writing pole and a large back-plate providing a low reluctance for the flux to return. The writing pole concentrates and guides the magnetic flux generated by the coil across the recording surface of the drum. Individual electromagnets are arranged into a two-dimensional array of 6 rows of 80 elements. This arrangement provides enough space for a 10- or 12-turn flat coil to be built on the available chip width.

Fabrication of this device is based on the sequential integration of the electrical circuit including the coils, and the magnetic circuit including the writing poles. The electrical circuit was built using a standard electronic process whereas the magnetic circuit fabrication required the development of a dedicated process based on silicon micromachining and microgalvanic technologies [20]. A cross-section of this first design is presented in Figure 2-16.

A microelectronic double-metal process was used to built the electrical circuit on a 4-inch double-face polished ⟨100⟩ p-type silicon wafer. The metallic layers, both made of aluminium, were 0.7 μm and 1 μm thick. The coils were composed of four turns on the first metallic level and six turns on the second.

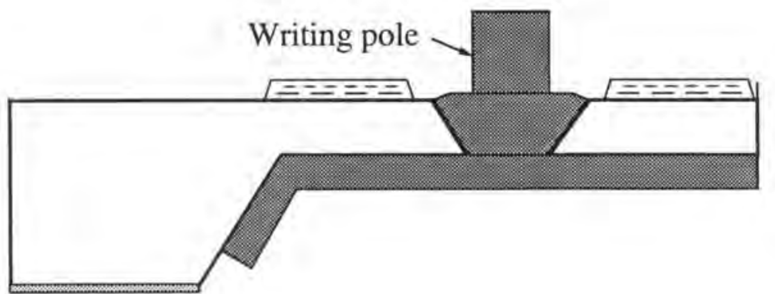


Figure 2-16. Cross section of the first magnetographic head design.

The thin membrane under the coils in Figure 2-16 was obtained by anisotropically etching the silicon wafer from the back. The magnetic back-plate under the membrane was realized by evaporation of an electroplating seed layer and by selective electrodeposition of the 50/50 FeNi alloy.

Holes were then anisotropically etched through the silicon membrane in the coil centers. Planarization of the front side followed by selectively filling the holes with galvanic FeNi alloy using the back plate as electroplating electrode. The FeNi writing poles were completed using a 40 µm thick polyimide resist mold patterned by standard TAB photolithographic techniques [21]. Finally, a tribological layer was deposited on the chip surface in order to minimize the wearing of the writing poles.

This first development step of a 480 dpi head has demonstrated the fabrication feasibility of recording electromagnets on silicon chips. Preliminary printing tests with the simplified head showed very good results.

Product Feasibility

The next phase, prototype development, was based on the structure of the simplified heads, with the following additional functions:

- diode matrices co-integrated with the electromagnet matrices allowing the 480 electromagnets to be individually and sequentially addressed;
- a low-resistance circuit to limit the power consumption and dissipation in the head.

The floorplan of the chip is depicted schematically in Figure 2-17. The diode connections organize the coils into 20 groups of 24 devices, thus allowing multiplexed addressing. Electrical connections to the chip are composed of only 24 data lines and 20 group-selection lines, which corresponds to 44 bonding pads.

The fabrication process of the prototypes included the addressing electronics, the low-resistance gold lines, and the magnetic circuit [22]. The fabrication of high-density, low-resistance gold conductors was performed using microelectrochemical deposition techniques. The 10-turn coils were made of 4 µm lines spaced by 4 µm, providing an 8 Ω coil resistance. The fabrication process of the magnetic circuit was similar to that for the simplified heads. The last step was again the deposition of the tribological layer. A partial view of the electromagnet matrix thus realized is presented in Figure 2-18.

The fabrication of prototypes was an important step in the development of the 480 dpi magnetographic head. Most of the prototype properties were close to the required product specifications. The following points have been successfully demonstrated:

- co-integration of magnetic elements together with electronics;
- individual addressing of the electromagnets leading to 480 dpi printing density.

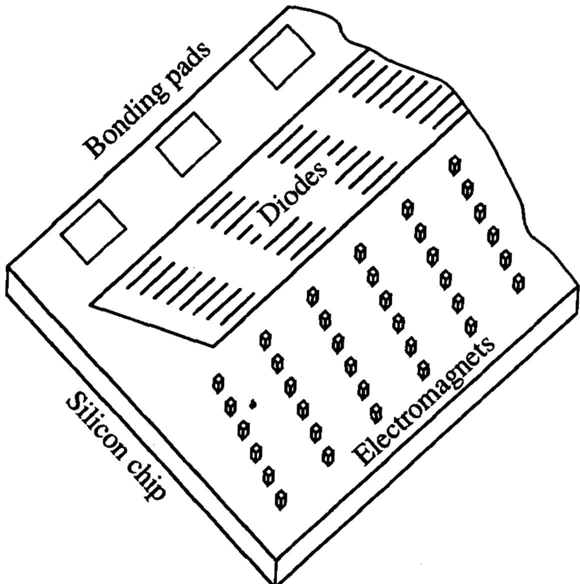


Figure 2-17.
Schematic floor plan of the electromagnetic chip.

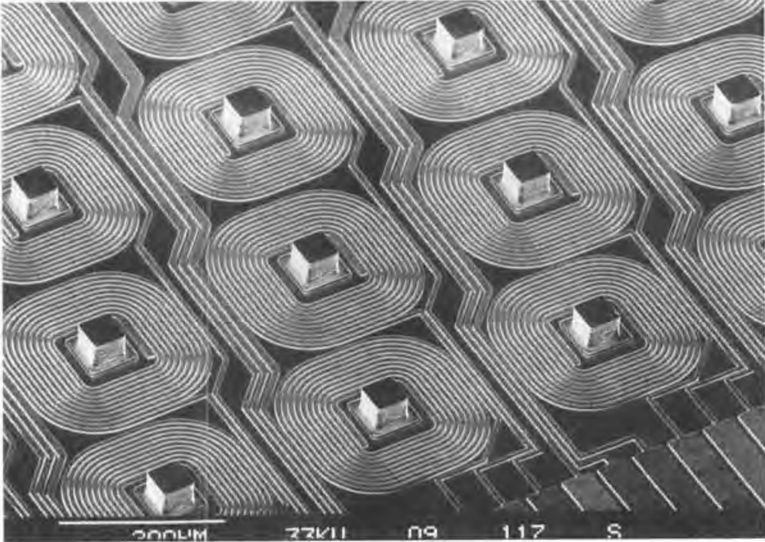


Figure 2-18. Part of the electromagnet matrix of the prototypes.

Product Development

In the third phase, the industrial product was developed. Compared with the prototypes, some improvements were introduced:

- reduction of power consumption, achieved by increasing the number of coil turns from 10 to 12 and by modifying the magnetic circuit in order to concentrate better the magnetic flux into the writing poles;
- reduction of fabrication cost, obtained by optimizing the chip lay-out and thus increasing the number of chips per wafer from 12 to 21.

Analysis of the protoype performance revealed a divergent magnetic flux in the writing poles, which reduces the magnetic density on the drum. The widening shape of the pole base (cf. Figure 2-16), due to anisotropic etching through the membrane, was identified as the reason for this divergent flux. Therefore, a new process, based on electrochemical etching of silicon in HF [23], was designed to improve the shape of these holes, leading to convergent paths. Figure 2-19 shows cross sections of this improved magnetic head.

For use in an industrial printer, the silicon chips were glued on aluminium plates and connected using wire bonding. Modules thus realized are assembled to form full-width printing heads.

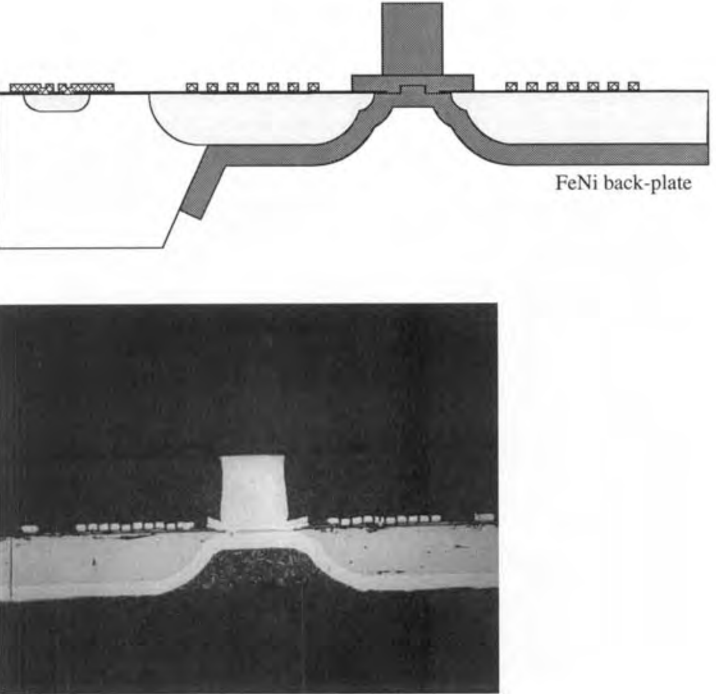


Figure 2-19. Schematic cross section (top) and photograph of the improved magnetic head structure (bottom).

Results

Complete head characterization and extensive printing tests have been performed. Good electrical and printing performances were demonstrated.
The main electrical performances are:

- electrical circuits support 1 A without damage;
- nominal driving current 0.3 A;
- recording pulse width 0.3 μs;
- time between pulses (group selection) 1.5 μs.

These results show an extremely fast response of heads when high-frequency signals are used. According to the electrical performances, the maximum printing speed achievable is 100 m/min. Shorter response time will be reached by both improving the performances of the printer drivers and reducing the stray capacitances of the modules.

The printing speed is at present limited by the power dissipation of the modules. A high printing speed of 60 m/min is reached under the following working conditions:

- page coverage rate 50%;
- maximum power dissipation per module 4.8 W.

The printed lines show good blackness (optical density 0.75).

Conclusion

This section has presented the different steps in the development of an innovative magnetographic silicon head in In-IC technology. The device is composed of high-density arrays of electromagnets which are sequentially addressed by on-chip diode matrices. This smart head is realized by a suitable combination of techniques belonging to standard silicon microelectronics, silicon microstructuring, and electrochemical deposition.

The main improvements over existing technology include batch production of the magnetic heads, significant size reduction of the devices, doubling of the pixel density to 480 dpi, and a high printing speed of 60 m/min.

This new product is at the leading edge of technology and meets the requirements of future generations of magnetic printers.

2.6 Conclusions and Future Trends

This chapter started with an overview of the three main classes of microsystem technologies and their characteristics: Beside-IC, In-IC, and Above-IC. Also the market demands for these technologies were discussed. Apart from specific system requirements, accuracy and fabrication yield were identified as two important criteria which have to be considered to make the appropriate choice. Beside-IC technology has the highest accuracy potential.

The design methodology must be adapted to the specific nature of microsystems. Three main phases take place: technology set-up, development, and production. It was shown that an efficient way to reduce the time-to-market for a new product development is to build up a library of generic microsystems and microsystem components. Obviously, this can only be done by a mature design house with a broad and up-to-date technology base.

Three typical examples of microsystems were analysed, in order to illustrate the main points developed in this chapter. All three concerned recent microsystem innovations. The first was a pressure sensor in Beside-IC technology. For this system, all critical components were available, and the innovation came from a well adapted system design methodology. This microsystem had a very short development time.

The second example, a high-resolution accelerometer system, was presented in support of the claim that very high precision can be reached with Beside-IC technology. For this system, the measurement cell was available, but a new interface circuit had to be developed, adapted to the high precision requirement. This example also showed that it is absolutely imperative to integrate both mechanical and electronic functions in one model for system design and simulation.

The third example was an In-IC electromagnetic actuator. The entire system had to be designed and developed from scratch, since an entirely new approach was devised to achieve unprecedented performance. The work included an extensive and innovative technological development and consequently, the development was relatively long and costly.

To look into the future, a selection of three typical trends for which the microsystem community is now preparing is presented here: galvanic deposition as a promising upcoming technology, the evolution of microsystem interconnection and communication, and the creation of microsystem design centers.

Microsystems based on galvanic deposition of metals are coming out of the research laboratories and finding their way to the market. It is expected that their place in microsystem design will be as important as bulk and surface micromachined silicon. The relatively new technology has been integrated as a typical representative of the In-IC and Above-IC microsystem technologies. Comparing galvanic deposition with polysilicon deposition, the following points should be noted:

- a wide choice of metallic materials allows to select a larger range of mechanical properties such as Young's module and mass density;
- fabrication requires only low temperature processes; therefore, it is IC-compatible, which makes it an ideal candidate for Above-IC microsystems.

Although the future of galvanic deposition is very promising, some more research is required, eg, concerning mechanical stress in movable parts.

Until recently, communication requirements for microsystems have been relatively simple. Since they have mainly been integrated in a local control environment, a DC voltage or a current loop were the most frequent demands. More and more, however, microsystems are designed to be integrated into digital networks or for remote control applications. Therefore, systems with digital output in the form of a frequently signal or a bit stream are rapidly gaining importance. In order to maintain high accuracy and low price, A/D converters should be avoided, and new measurement principles are being devised which directly integrate a digital function. Also, microsystems incorporating remote communication, based on RF or infrared

transmission, will soon become standard. These devices will often need on-board power supply, or wireless power transmission. In both cases, low power consumption will be a critical issue.

Finally, it has been mentioned before that design centers will take an important place in the microsystem world. Three competences will be decisive for a microsystem design center to be successful. The first is to identify and anticipate correctly future required microsystem components and technologies, in order to have a complete set available, adapted to the needs of the microsystem community. The second is a strong system design team, able to translate the given set of objectives and specifications into the most adequate system proposition, using an efficient top-down design methodology, and the available component box. Third, the design centers will need their own in-house production facilities, both for wafer fabrication and for assembly and testing of microsystems. This close proximity is imposed by the very intensive interaction between technological and design aspects of microsystem development and production.

2.7 Acknowledgements

Many scientists and engineers contributed to the technical work presented in this chapter. We acknowledge Y. de Coulon, T. Smith, J. Hermann, M. Chevroulet, J. Bergqvist, O. Nys, F. Porret, F. Cardot, J. Gobet, H. de Lambilly, Ph. Renaud, B. de Geeter, F. Rudolf, and J.-M. Moret.

2.8 References

[1] M. Pierre, *Power Consumption and Dynamic Range of Capacitive Accelerometer Interfaces,* Technical Report CSEM No. 595, 1993.
[2] Ristic, L. (ed.), *Sensor Technology and Devices:* Boston: Artech House, 1994.
[3] Pons, P., Blasquez, G., Behocaray, R., *Sensors Actuators A* **37-38** (1993) 112-115.
[4] Jornod, A., Rudolf, F., *Sensors Actuators* **17** (1989) 415-421.
[5] Shimaoka, K., Tabata, O., Kimura, M., Sugiyume, S., in: *Transducers'93, June 7-10, 1993, Yokohama, Digest of Technical Papers,* pp. 632-635.
[6] Leuthold, H., Rudolf, F., *Sensors Actuators A* **21-23** (1990) 278-281.
[7] de Coulon Y., de Geeter, B., Rudolf, F., Paper presented at Eurosensors VIII, September 25-28, 1994, Toulouse.
[8] Spiering, L. V., Bouwstra, S., Spiering, R. M. E. J., Elwengspoek, M., in: *Transducers'91, June 24-27, 1991, San Francisco, Digest of Technical Papers,* p. 982.
[9] Offereins, H. L., Sandmaier, H., Folkmer, B., Steger, U., Lang, W., in: *Transducers'91, June 24-27, 1991, San Francisco, Digest of Technical Papers,* pp. 986-989.
[10] Bergqvist, J., Rudolf, F., *Sensors and Actuators A* **45** (1994) 115-124.
[11] Rudolf, F., Jornod, A., Bergqvist, J., Leuthold, H., *Sensors Actuators A* **21-23** (1990) 297-302.
[12] Suzuki, S., Tuchitani, S., Sato, K., Ueno, S., Yokota, Y., Sato, M., Esashi, M., *Sensors Actuators A* **21-23** (1990) 316-319.
[13] Henrion, W., DiSanza, L., Ip, M., Terry, S., Jerman, H., in: *Proc. IEEE Solid-State Sensor and Actuator Workshop, Hilton Head Island, USA, June 4-7, 1990,* pp. 153-157.

[14] Kulzer, G., de Coulon, Y., Roussel, P., Trischberger, M., *Acta Astronaut.* **26** (1992) 543-549.
[15] Leuthold, H., Rudolf, F., *Sensors Actuators A* **21-23** (1990) 278-281.
[16] Candy, J. C., Temes, G. C., *Oversampling Delta-Sigma Converters;* New York: IEEE Press, 1992, pp. 1-25.
[17] Eltgen, J.-J. P., Magnenet, J. G., *IEEE Trans. Magn.* **MAG-16** (1980) 961-966.
[18] Keefe, G. E., Yarmchuck, E. J., Dove, D. B., *J. Imaging Sci.* **33** (1989) 69-72.
[19] Cherbuy, B., in: *Proc. IS&T 7th Int. Congress on Advances in Non-Impact Printing Technologies, Portland, OR, USA, October 1991.*
[20] Cardot, F., Cherbuy, B., in: *Proc. IS&T 10th Congress on Advances in Non-Impact Printing Technologies, New Orleans, Louisiana, USA, October, 1991,* pp. 534-538.
[21] Milosevic, I. I., Perret, A., Losert, E., Schlenkrich, P., *Semicond. Int.,* October (1988).
[22] Cardot, F., Gobet, J., Bogdanski, M., Rudolf, F., *Sensors Actuators A* **43** (1994) 11-16.
[23] Branebjerg, J., Eijkel, C. J. M., Gardeniers, J. G. E., van de Pol, F. C. M. in: *Proc. IEEE Micro-Electro-Mechanical Systems Workshop, Nara, Japan, 1991,* p. 221.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

3 Sensors in Microsystems

HENRY BALTES and CARLOS AZEREDO LEME, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

Contents

3.1 Introduction 52

3.2 IC Sensors 52
3.2.1 Just IC Technology 52
3.2.2 Additional Processing 53
3.2.2.1 Post-Processing Deposition 53
3.2.2.2 Post-Processing Etching 53

3.3 CMOS Sensor Elements 56
3.3.1 Humidity Sensors 56
3.3.2 Magnetic Sensors 57
3.3.3 Power Sensors 58

3.4 CMOS Sensors with On-Chip Circuits 59
3.4.1 Microsensor Interface Requirements 59
3.4.2 Microsensor Signal Processing Capabilities 60
3.4.3 Optimal Design of Microsensor Interfaces 61

3.5 Sensors with Oversampling Interfaces 62
3.5.1 Architecture Design 62
3.5.2 Capacitive Humidity Sensor 64
3.5.3 Magnetotransistor 67
3.5.4 Thermoelectric Power Sensor 69

3.6 Interface with Minimum Analog Content 71
3.6.1 System Architecture 71
3.6.2 Circuit Analysis 73

3.7 Conclusion and Outlook 76

3.8 Acknowledgments 76

3.9 References 76

3.1 Introduction

This chapter addresses the integrated circuit (IC) way to microsensors, ie, the design and fabrication of sensor elements with on-chip circuitry using industrial IC processes such as complementary metal oxide semiconductor (CMOS) technology with or without additional IC-compatible sensor-specific processing steps. This approach can lead to cost-effective, functional microsensor modules, which may serve as building blocks for sensor and actuator microsystems.

Rather than attempting a systematic review of the lively research area of microsensors with on-chip interface circuitry, we illustrate some essential concerns with specific examples. Microsensors made by IC technology are presented in Sections 3.2 and 3.3, notably CMOS humidity, magnetic, and power sensors. General sensor interface design issues are discussed in Section 3.4, while Section 3.5 concentrates on oversampled interfaces for humidity, magnetic, and power sensors. An approach to interfaces with minimum analog content is described in Section 3.6.

This chapter complements Chapter 2 in Volume 7 of this series [1].

3.2 IC Sensors

Sensors connect the world of physical and chemical measurands to the world of electronics and computers. Every sensor is based on a physical effect and a material showing that effect. It is desirable that practical sensors can be batch manufactured at low cost with high reliability and can be linked with integrated electronic circuitry. Thus the sensor designer may want to draw from the knowledge of not only solid-state physics but also IC fabrication technology and IC circuitry.

The task is to translate physical sensor effects into device structures to be manufactured by industrial IC technology. On-chip signal-conditioning circuitry is an overwhelming advantage of IC sensor technology. Such circuitry can enhance both the performance of the sensing element and the electromagnetic compatibility of the sensor system.

3.2.1 Just IC Technology

Silicon microsensors fabricated by industrial CMOS or bipolar IC technologies without further silicon processing have been demonstrated for magnetic field [2, 3], temperature [4], and radiation [5] measurands. Indeed, Hall chips (magnetic sensors based on the Hall effect in silicon with on-chip signal conditioning circuitry) have been commercially available for many years.

Recently developed bipolar interface circuitry for Hall sensors was reported in Volume 7 of this series [2]. A magnetotransistor with interface circuit in CMOS technology is described in Section 3.5.3.

3.2.2 Additional Processing

A number of microsensors can be realized by combining established IC technologies with additional processing steps specific to the sensor function and compatible with the IC process [1, 6, 7]. Preferably, the extra process steps are performed after completion of the regular IC process (post-processing), but pre- and between-processing (interrupting the IC process flow) has also been used [1, 8, 9, 10].

3.2.2.1 Post-Processing Deposition

The deposition temperature is crucial and must be below 400 °C in order to preserve the aluminium structures of the preceding IC process. This requirement may be relaxed for an IC process with tungsten metalization [11].

For example, deposition of moisture adsorbing or absorbing thin films such as porous ceramics or organic polymers is used for IC humidity sensors [12]. Photoimageable polyimide can be deposited and patterned on CMOS capacitive or bipolar resonant humidity sensors (see Section 3.3.1).

3.2.2.2 Post-Processing Etching

Inherent features of CMOS technology and etching properties of its (100) substrate allow the fabrication of open silicon dioxide microstructures (beams, bridges, or suspended membranes) by anisotropic silicon etching from the front of the wafer [1, 6, 7]. The preferred etchant is ethylene-diamine-pyrocatechol (EDP). By design, the passivation layer serves as an etching mask. This is illustrated by Figure 3-1, which shows a CMOS beam cross section before and after completion of the final etching step [7, 13, 14]. The dielectric beam supports

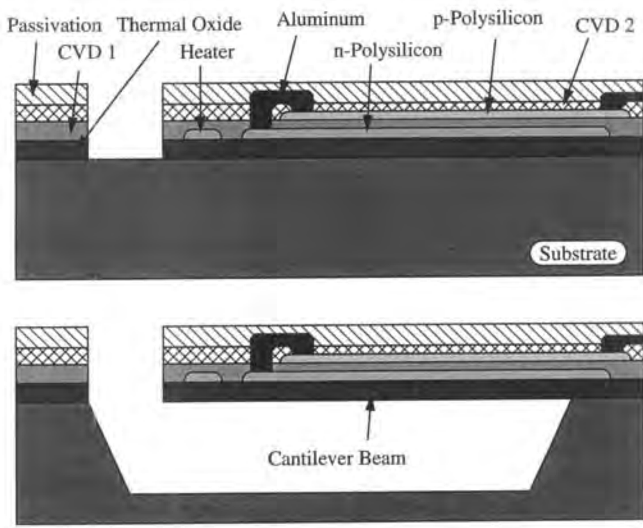


Figure 3-1. Cross section of CMOS beam with polysilicon thermopile before and after silicon etching [7].

an n-polysilicon-p-polysilicon thermopile. An SEM picture of the tip of the beam is shown in Figure 3-2. Removal of the silicon substrate is crucial for thermal insulation.

Closed dielectric membranes can be formed by etching from the back of the finished CMOS wafer [14, 15]. An extra mask is required in order to define the membrane size. The preferred etchant is KOH. The front of the wafer is protected from the etchant by a mechanical cover. A corresponding cross-section is shown in Figure 3-3.

A different kind of membrane based on epitaxial silicon layers can be obtained by post-processing etching using bipolar IC processes on (100) silicon [16]. This technology is provided

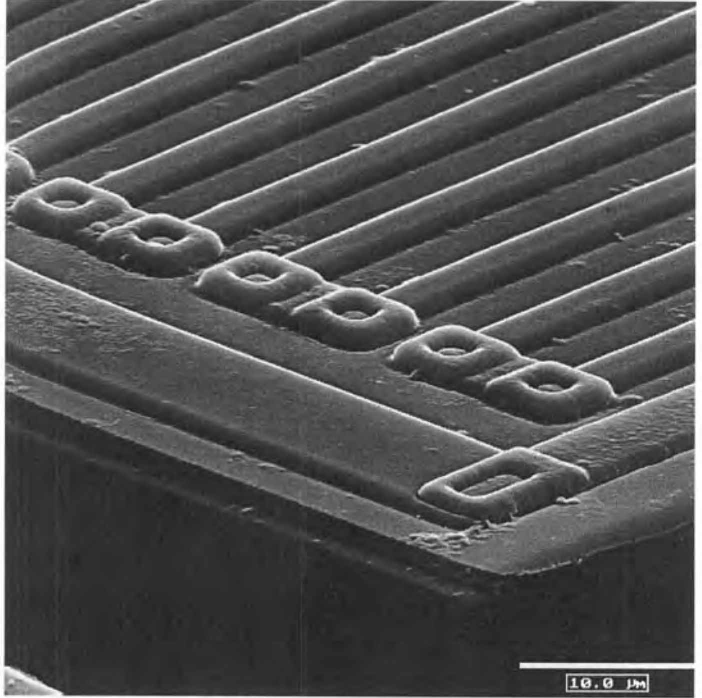


Figure 3-2. SEM picture of tip of CMOS beam polysilicon thermopile [13].

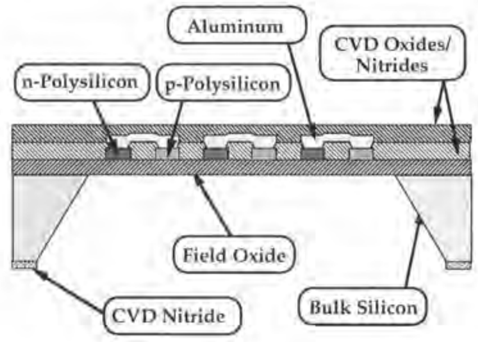


Figure 3-3. Cross section of CMOS membrane polysilicon thermopile [15].

by eg, Ascom Microelectronic (Bevaix, Switzerland) together with the final electrochemical back-etching process for releasing the membrane. Figure 3-4 shows a corresponding design for a thermally excited membrane resonator. The p-type diffusions are used for thermal excitation or piezoelectric detection of vibrations.

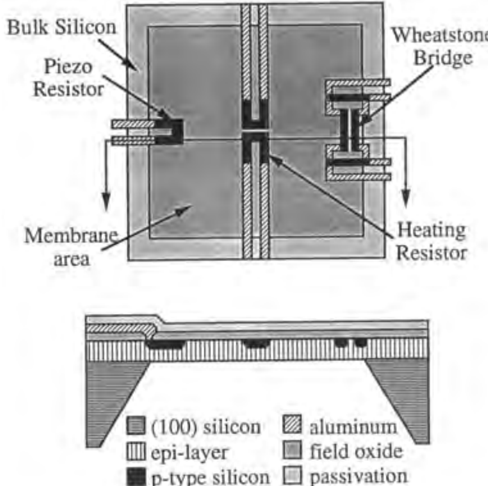


Figure 3-4.
Top view and cross section of thermally excited bipolar membrane resonator [19].

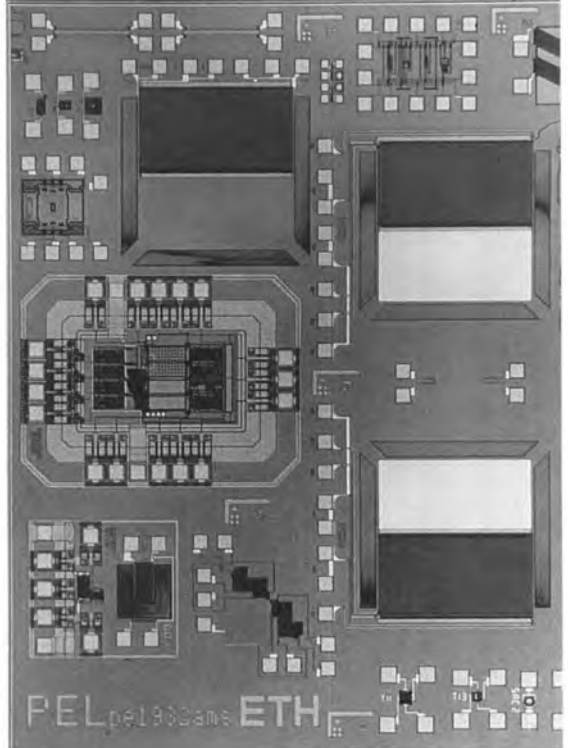


Figure 3-5.
SEM picture of CMOS membrane thermopile infrared sensors with amplifier circuit (center left) [21].

A variety of CMOS and bipolar microsensors have been demonstrated recently using the above post-processing etching procedures: thermal mass flow sensors [13-15], infrared sensors [17], ac power sensors (thermoconverters) [18], thermal and resonant vacuum sensors [19], and thermally excited mechanical resonators [16, 20]. As an example, Figure 3-5 shows the SEM picture of a CMOS membrane thermopile infrared sensor with amplifier circuit [21].

3.3 CMOS Sensor Elements

Four of the many possible types of CMOS IC sensor elements are discussed in this section: capacitive and piezoresistive humidity sensors for air, magnetotransistors, and ac power sensors.

3.3.1 Humidity Sensors

Capacitive humidity sensors manufactured in CMOS technology can be based on interdigitated or meshed capacitors with aluminium of polysilicon electrodes [12, 22]. In a post-processing step after completion of the CMOS IC process, such lateral capacitors are covered by a layer of hygroscopic polyimide (typically a few µm in thickness), which is also photosensitive and can be easily patterned and processed by photolithography. The polyimide curing cycle at 350 °C does not affect the CMOS structures. The dielectric permittivity of the polyimide varies from 3.2 (dry) to 4.0 (wet). A detail of such a humidity sensor is shown in Figure 3-6; its size is 380 by 380 µm.

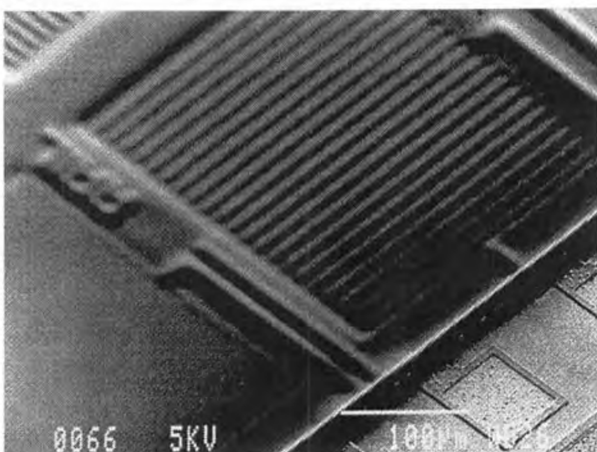


Figure 3-6.
SEM picture of CMOS capacitive humidity sensor with polyimide coating [22].

The typical capacitance of such devices is less than 1 pF, its response to ambient humidity is about 0.8% per 1% relative humidity, and the response time is less than 30 s. In view of the small capacitance modulation (a few femtofarads) to be detected here, on-chip signal con-

ditioning circuitry is indispensable (see Section 3.5.2). With such circuitry, a resolution of 1 % relative humidity has been achieved.

Piezoresitive humidity sensors compatible with CMOS or bipolar IC technology can be achieved by combining post-processing anisotropic etching and polyimide deposition [12, 23]. The device cross section is illustrated in Figure 3-7. Environmental humidity deforms the polyimide layer. The resulting stress on the membrane is detected by the change in resistance of the polysilicon piezoresistors integrated in the membrane.

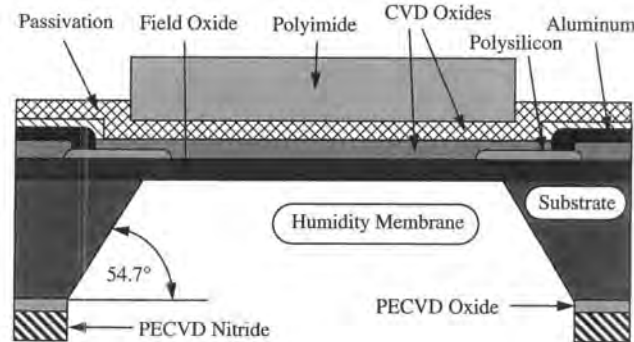


Figure 3-7.
Cross section of piezoresistive humidity sensor based on CMOS technology [12].

3.3.2 Magnetic Sensors

Many types of magnetic sensors can be made by CMOS or bipolar IC technolgoy [2, 3, 24]. Magnetotransistors (MTs), for example, are bipolar transistors whose collector current is modulated by a magnetic field. Most MTs have a dual-collector structure. At zero magnetic field, their operation is symmetrical with respect to the two collectors, which show identical currents, $I_{C1} = I_{C2}$. In the presence of a magnetic field, the Lorentz force creates asymmetries in the potential and current distribution, which translate into a collector current imbalance, $\Delta I_C = I_{C1} - I_{C2}$.

In lateral magnetotransistors (LMTs) the minority carriers flow laterally from the emitter to the collectors. While bipolar in function, LMTs can be made using CMOS IC technology. Figure 3-8 shows an example of an LMT cross section with one emitter (E), two collector (C)

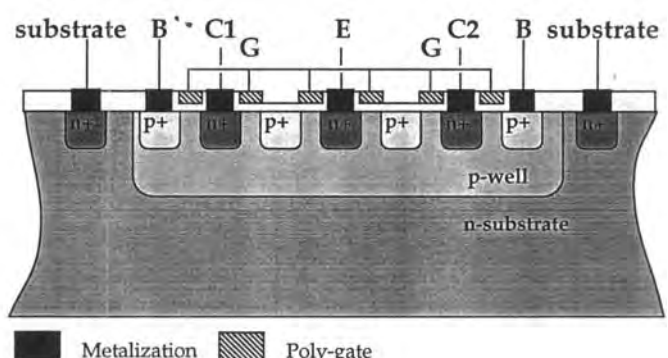


Figure 3-8.
CMOS lateral magnetotransistor [25].

and two base (B) contacts [25]. The additional p^+ diffusion between the emitter and collectors prevents the minority carriers from the shortest lateral path to the collectors and forces them to follow a curved trajectory far away from the surface. A typical sensitivity (ie, the relative collector current imbalance per magnetic induction) is 50% per tesla.

3.3.3 Power Sensors

In a thermoconverter or ac power sensor, electric power is converted into thermal power by a resistor. The resulting heat is measured with a temperature sensor. The CMOS power sensor shown in Figure 3-9 is based on two parallel polysilicon heating resistors and a thermopile on a 600 by 310 μm silicon dioxide microbridge. The heating resistors are covered by gold bumps provided routinely by the CMOS manufacturer EM Microelectronic Marin. In addition to providing improved protection against the post-processing etchant, the gold bumps increase the thermal time constant (20 ms).

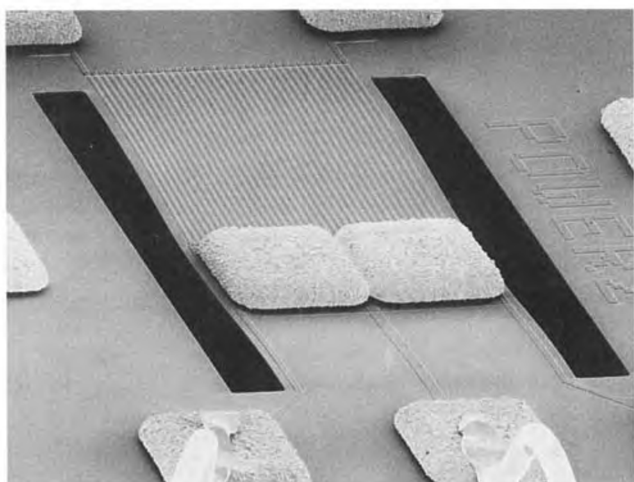


Figure 3-9. SEM picture of CMOS thermoconverter with gold bumps [18].

The ac power source to be calibrated is connected to the heating resistors. The dissipated heat is transformed into the thermopile output voltage. A sensitivity of about 90 mV/mW is achieved with 18 n-polysilicon-p-polysilicon thermocouples. The ac/dc transfer error between 10 Hz and 1 MHz is less than 700 ppm. The linearity error is less than 0.5% below 10 mW input power. The burn-out power is about 50 mW.

For further types of CMOS thermoconverters we refer to the literature [1, 18].

3.4 CMOS Sensors with On-Chip Circuits

The on-chip signal processing capability is a distinctive advantage of the IC sensor approach. Thanks to the underlying industrial IC process, well characterized high-performance IC components and libraries are readily available. This facilitates the co-integration of signal conditioning and interfacing functions on the sensor chip. Such interface circuits form an essential part of the microsensor system. However, the presence of the sensor, notably the post-processing and the environment to which the sensor system will be exposed, may affect the performance of the IC components, in particular the analog part. Moreover, some of the sensor interface requirements can be demanding.

3.4.1 Microsensor Interface Requirements

During to the vast areas of application and classes of sensors, interface specifications vary widely. We shall try to summarize the main requirements.

Most silicon sensor elements show only moderate performance. Process deviations and nonlinearity of the materials involved may affect the sensor response. Calibration and offset or scale-error compensation circuits can significantly improve the sensor quality. The resolution requirements on the interface are moderate, ie, in the range from 6 to 12 bits.

Also, speed requirements are not very demanding. Many micromechanical sensors have resonance frequencies between 10 kHz and several hundred kHz. Most sensors based on thermal effects have time constants between 100 μs and 10 ms. Some chemical sensors based on the reaction of a deposited sensitive film may even be slower and show response times above 10 ms. Thus speed specifications are usually in the range between 10 Hz and 200 kHz.

Figure 3-10 illustrates the speed/resolution requirements for most integrated silicon sensors. These do not seem excessively difficult to meet with conventional architectures for A/D conversion. Microsensor signals, however, can be very weak and this changes the picture completely. Output signals from the sensor are usually in the range from a few microvolts up to 10 mV. Thus the specifications in terms of noise and interference suppression are very severe. Low-noise design and filtering to the signal band are therefore highly desirable. Also, circuit

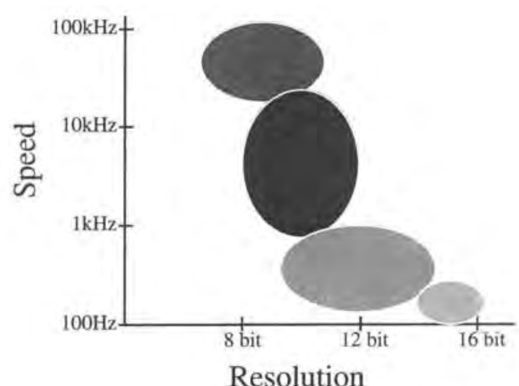


Figure 3-10.
Speed and resolution of silicon microsensors (mainly below 12 bit and 10 kHz, respectively).

techniques for the cancellation of $1/f$ noise, such as chopper stabilization [26] or correlated double sampling [27], should be applied when the signal band is very low and only MOS transistors are available for the input stage [28].

Another important noise source is electrical interference from external fields and on-chip digital signals. This problem can be alleviated by proper shielding and circuit partitioning in order to isolate the noise sources from the sensitive input stages. However, often such a procedure is not practical; this makes fully differential circuit topologies very attractive in view of their excellent interference immunity.

Because of the many different types of microsensors suitable for integration, the input signal can appear in a variety of formats to the interface. The input can be simply a voltage or current signal for an active sensor. It can also be a capacitance or resistance variation, such as in strain gauges. Other sensors produce a modulation in the transfer function of a two-port network, changing the coupling coefficient or the poles position. Moreover, the signal can be in either a single-ended or differential mode. Naturally, the connection of the sensor to the interface and its first-stage topology vary widely with the type of input signal. This complicates the design of the interface since its architecture must change dramatically depending on the sensor type.

3.4.2 Microsensor Signal Processing Capabilities

Figure 3-11 shows a microsensor interface with most functions that ideally would be co-integrated. Amplification is usually the first operation to be performed, since the signal levels delivered by an integrated sensor are relatively low. The amplifier design concentrates on low-noise and high immunity against interference.

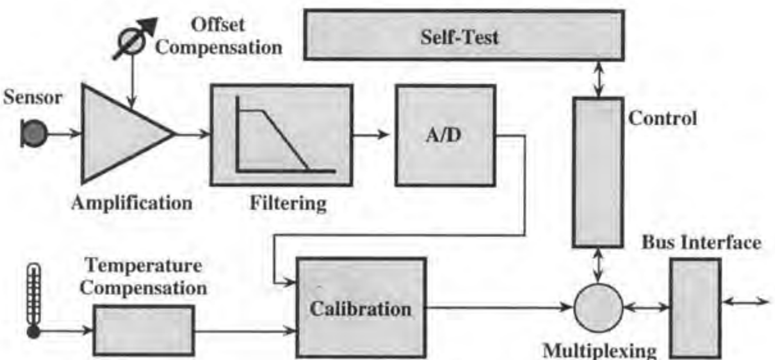


Figure 3-11. Microsensor interface with most important functions.

Microsensors often show a large offset. This should be compensated as soon as possible in the signal flow in order to prevent consumption of valuable dynamic range of subsequent circuits, namely the A/D converter.

The A/D converter is a highly desirable circuit block to include. By working in the digital domain, high interference immunity and robustness against component degradation are ob-

tained. Therefore, it should be placed as close as possible to the sensor. However, the performance of the whole system will depend very strongly on the quality of the A/D converter and, accordingly, it should be designed with great care.

Filters are also an important circuit block. Both an anti-aliasing filter and an out-of-band noise rejection filter may need to be integrated depending, respectively, on the type of A/D converter used and the spectrum of unwanted interfering signals.

A number of additional functions should be included, whenever possible, that can significantly improve the performance and reliability of the microsensor system. Control of driving power is important in thermal sensors [29–31]. Calibration can correct sensor nonlinearities, thereby increasing the measurement resolution [32, 33]. User friendliness can be obtained by including sophisticated digital control circuits. A self-test capability is an important issue in most modern applications as it provides high reliability and operating security. The presence of a bus interface can considerably simplify the connection to a sensor network. Morerover, by multiplexing both sensor output signals and digital control signals, the pin-out can be reduced, which is important in remote or implanted sensors.

3.4.3 Optimal Design of Microsensor Interfaces

The design of a microsensor interface is a challenging task. On the one hand, the requirements are stringent and the number of functions to implement numerous. On the other hand, the side-effects of post-processing and the harsh environment to which the sensor chip may be exposed can reduce the performance of the IC components to some extent. Therefore, full integration of the complete microsensor system on one chip has not often been achieved hitherto. A compromise must be found where either the sensor sophistication is reduced or the system is partitioned with only a minimum of on-chip functions. In either case a loss of performance may generally result.

In order to approach the ideal of complete system implementation without compromising overall performance, new design techniques have to be selected which allow high robustness against deviations in component characteristics and do not rely on high accuracy component matching. In the following we present several design guidelines for optimizing sensor interface design.

Let us first emphasize the necessity to design the whole system by considering the sensor element, interface, and operating conditions together. The particular environmental and post-processing conditions determine the complexity and stability of the functions that can be implemented. Also, by designing the sensor and interface as a whole, a closed loop system can often be devised where signal processing and negative feedback can substantially improve sensor linearity and stability.

The sensor element should be optimally coupled to the interface circuit; ideally it should be included as an appropriate internal circuit element. This allows minimization of the analog content in the interface, which is the second important point to highlight. Analog circuits are very sensitive to perturbations and, therefore, most functions should be realized in the digital domain and the A/D converter should be placed as early as possible in the signal flow.

The third point is the use of circuit techniques that possess high robustness against component mismatch and performance loss. This rules out most architectures for A/D conversion at the Nyquist rate. Most promising candidates are the oversampling A/D converters [34].

They achieve accuracy and resolution at the expense of speed in the analog modulator and logic in the decimating filter [35]. They usually come with noise-shaping techniques and negative feedback that provide low sensitivities to component deviations and require only simple analog circuits.

Another point to stress is the advantage of having an easily reconfigurable first stage. This eases redesign of the interface for new types of sensors requiring different input formats. Here, oversampled modulators are in a good position as their input stage consists of an integrator which can easily accept practically any input signal format, such as voltage or current and capacitance or resistance variation.

3.5 Sensors with Oversampling Interfaces

Our goal is to achieve a versatile and robust interface which can be used with a variety of sensors and withstand environmental and processing hazards.

To this end we present basic sigma-delta modulators with switched-capacitor techniques. The most appropriate topologies are the first- and second-order modulators. The first-order noise-shaping modulator has the advantage of simplicity since it consists of a loop containing only an integrator and a comparator, which can be easily integrated in most sensor processes. However, this architecture is limited to rather low band widths; for a 15 bit resolution an oversampling ratio of 2048 would be required. With a second-order noise shaping modulator, the oversampling ratio could be reduced to 128, allowing more than an order of magnitude increase in the speed of operation. Also, in-band noise tones, which are a problem with first-order modulators, are more easily suppressed. Finally, when a high gain is needed in the loop for high sensitivity, the capacitance spread can be greatly reduced by distributing the gain between the two integrators. Therefore, the second-order modulator is preferred [36].

3.5.1 Architecture Design

The circuit (Figure 3-12) has a fully-differential topology for achieving a good power supply rejection ratio (PSRR), interference suppression, and clock feedthrough compensation. Additionally, for effective reduction of clock feedthrough, dummy switches are used extensively for charge injection compensation [37]. These measures are essential since, even when low resolution is required, the signal level may be low.

The clocks provided to the integrators turn off the switches coupled to the summing node before the switches on the opposite side of the capacitors. Thus the charge from the switches that are driven by the input, V_{REF}, or amplifier output are not integrated on to the integration capacitor, and the effects of voltage-dependent switch charge injection are eliminated [38]. The summing node switches still inject charge, but their sources and drains are all at the same potential. Since the charge injected is the same for both sides of the fully differential circuit, it appears as a common-mode voltage and, therefore, does not affect the circuit performance.

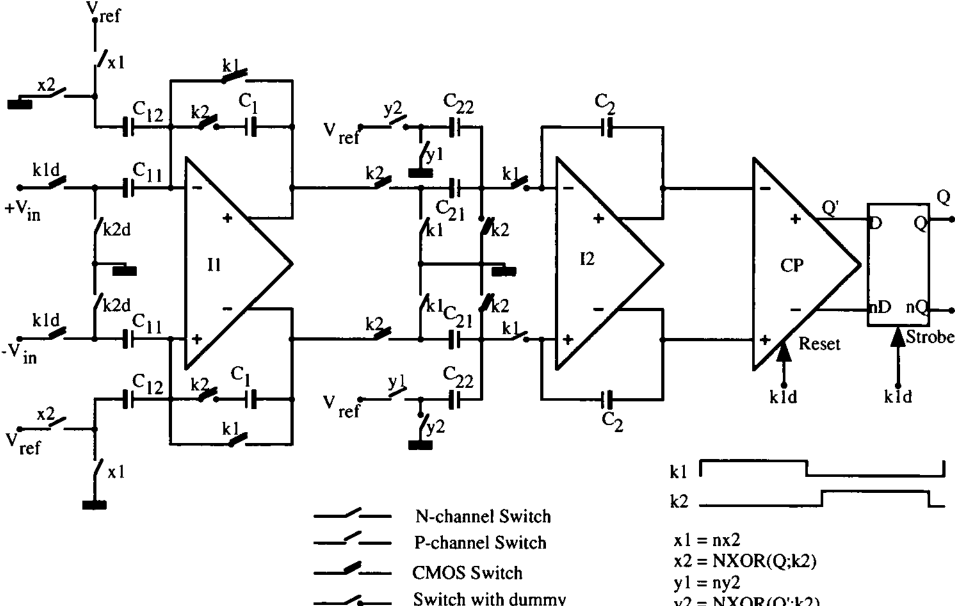


Figure 3-12. Basic modulator structure.

An offset compensation topology is chosen for the first integrator which has the additional advantage of eliminating the low-frequency flicker noise generated by its amplifier [27]. The second integrator, on the other hand, has a conventional topology which permits the operation of the circuit with only a positive reference voltage and a two-phase clock.

Low input level also means that an amplification of the signal will have to be carried out inside the loop. This is achieved by scaling the value of the reference voltage fed to each of the integrators together with their gains. The obtained gain, g, from the input to the output of the second integrator is simply given by

$$g = \frac{C_{11}}{C_1}\frac{C_{21}}{C_2} \tag{3-1}$$

with C_1, C_2, C_{11} and C_{21} as in Figure 3-12. Nevertheless, the transfer function of the loop, Equation (3-1), must be preserved [39] which leads to the design condition

$$\frac{C_{22}\,C_1}{C_{21}\,C_{12}}\frac{V_{\mathrm{REF2}}}{V_{\mathrm{REF1}}} = 2 \tag{3-2}$$

with V_{REF1} and V_{REF2} as in Figure 3-12. Using this scheme, arbitrary input sensitivities can be implemented without the need for previous amplification. This allows the inclusion of the input source, or sensor, inside the modulator loop.

The thermal noise due to switch resistance has to be considered when choosing the capacitor values in the modulator input stage. Each switched capacitor has a total noise power of $2kT/C$

that is aliased back into a band width equal to half the sampling frequency [40]. Therefore, in addition to the quantization noise, given by [35]

$$P_{B_{QUANTIZATION}} = \frac{\Delta^2 \pi^4}{60 M^5}, \tag{3-3}$$

where Δ denotes the step size, one has to consider the thermal noise in the baseband given by

$$P_{B_{THERMAL}} = \zeta \frac{kT}{C_{IN}} \frac{1}{M} \tag{3-4}$$

where C_{IN} is the input switched capacitance and M is the oversampling ratio. The degradation factor ζ depends on the amplifier and, for a proper design, its value lies around 2. This noise power is referred to the point to which the capacitor is switched. This should be taken into account so that Equation (3-4) is coherent with Equation (3-3) for the quantization noise power in the baseband. Thus, C_{IN} should be the switched capacitor in the outer feedback loop, C_{12} in Figure 3-12. Generally, owing to the very low input signal level, the system resolution will be determined by this noise component rather than the quantization noise.

One such conventional sigma-delta modulator can be directly applied to classical voltage output sensors. A wide range of sensitivities can be accommodated by appropriately scaling the voltage references and the capacitance ratios. More interesting is the application of this same structure to different types of sensors. This can be achieved by simple reconfiguration of the first integrator in the loop [32].

For input sources producing a capacitance variation, the capacitor element is inserted as an input switched branch of the first integrator. The remaining structure of the modulator is kept unchanged.

For the case of a current input signal, the input stage of the modulator is reconfigured to accept a current input. This is done by modifying the operational amplifier in the first stage. We use the fact that a transconductance amplifier is used as the active device in the integrator.

In the case of the power sensor based on a micromachined thermoconverter, one can take advantage of the inherent integrating function of the device. By judiciously combining two thermoconverter devices, the whole first stage of the modulator has been substituted.

These cases are now illustrated with sensor implementations.

3.5.2 Capacitive Humidity Sensor

Here, the output signal is a change of capacitance, C_s, related to the input measurand X_{IN} by

$$C_S = C_0 + \xi X_{IN} \tag{3-5}$$

where C_0 denotes the nominal element capacitance and ξ the sensitivity coefficient.

Owing to the switched-capacitor nature of the designed modulator, its reconfiguration for capacitive sensors is straightforward. The sensitive capacitive element of the sensor, C_{sensor}, can replace the input capacitor, C_{11}, in the non-inverting input of the first integrator stage.

The inverting switched-capacitor input can be used with an additional sensor element without passivation opening and, therefore, without humidity sensitivity. This takes advantage of the differential topology of the modulator which responds only to the difference between the two inputs. Therefore, the offset component of sensor output, C_0, can be readily subtracted by this reference element. In addition, by properly matching the sensitive and the reference sensor elements, a first-order compensation of temperature dependence and drift over time on C_0 can be obtained.

This resulting reconfigured first stage of the modulator is shown in Figure 3-13. The output code, represented as a variable D between 0 and 1, is given by

$$D = \frac{V_R\, C_{ref}}{V_{REF}\, C_{12}} - \frac{V_S\, C_{sensor}}{V_{REF}\, C_{12}} . \tag{3-6}$$

Using Equation (3-5), we have

$$D = \frac{-V_S}{V_{REF}\, C_{12}}\, \xi X_{IN} + \left(\frac{V_R}{V_{REF}\, C_{12}}\, C_{ref} - \frac{V_S}{V_{Ref}\, C_{12}}\, C_0 \right) . \tag{3-7}$$

Therefore, V_S determines the overall sensitivity and V_R allows offset compensation.

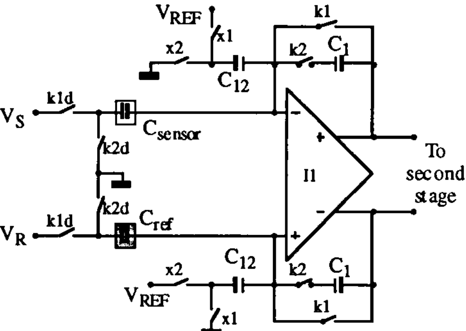


Figure 3-13.
Reconfigured first stage of the modulator for capacitance input signals.

The basic modulator presented in the last section has been redesigned in order to incorporate the reconfigured first stage of the modulator and integrated together with a capacitive humidity sensor. This consists of an "upper" aluminium electrode and a "lower" polysilicon electrode, each containing 32 fingers in an interdigitated geometry. Between the two electrodes the CVD oxide and silicon nitride layers, provided by the CMOS process, guarantee good electrical isolation. Both the inter-electrode distance and the electrode widths are designed with the minimum feature size of the process, 3 µm. The whole device occupies an area of 380 by 380 µm which is placed over a ground-connected P-well that serves as a shield against substrate noise. By appropriate definition of the layout layers, the passivation oxide is opened leaving the device exposed to the ambient environment. Two such structures have been included, one as the sensitive element and another, without the passivation opening, as the reference element, as shown in Figure 3-14.

The reconfigured modulator with the two capacitive sensors, one being the reference, has been fabricated. Since the digital decimator was not included on-chip, an alternative reading

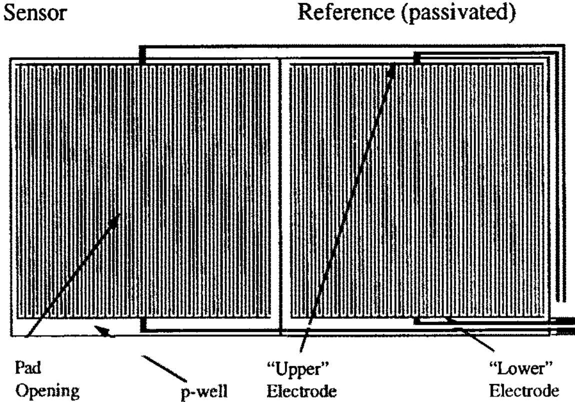


Figure 3-14.
Layout of capacitive humidity sensor and reference designed for 3 µm CMOS process [12].

approach was used. The response of the sensor is very slow, with an absorption time constant of 6 s. Therefore the oversampling ratio can be very large, allowing a simple first-order digital filter to eliminate the high-frequency noise. The set-up is shown in Figure 3-15. A frequency counter with an integration time of 1 s works as a counter and reset circuit. The bit stream

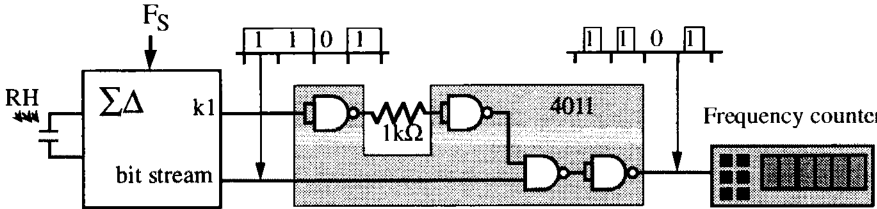


Figure 3-15. Reading set-up giving a frequency output.

is first passed through a shaping circuit that produces a return to zero format using the clock phase *k*1 properly delayed (see Figure 3-13). The only additional components needed are a standard CMOS 4011 IC and a resistor. Thus the measured value is a frequency F_{OUT}, given by

$$F_{OUT} = DF_S \tag{3-8}$$

where D is defined by Equation (3-7).

Using a clock frequency of 500 kHz and the control voltages $V_{REF} = 200$ mV, $V_R = 2.5$ V, and $V_S = 2.11$ V, characterization of the sensor in a calibrated chamber [12] leads to the results shown in Figure 3-16. The ambient temperature corresponding to this measurement is 60 °C. The measurement was performed both with rising (curve RH+) and falling (curve RH−) humidity values. The hysteresis and the non-linearity error are below 1 %, which is the limit of the humidity chamber.

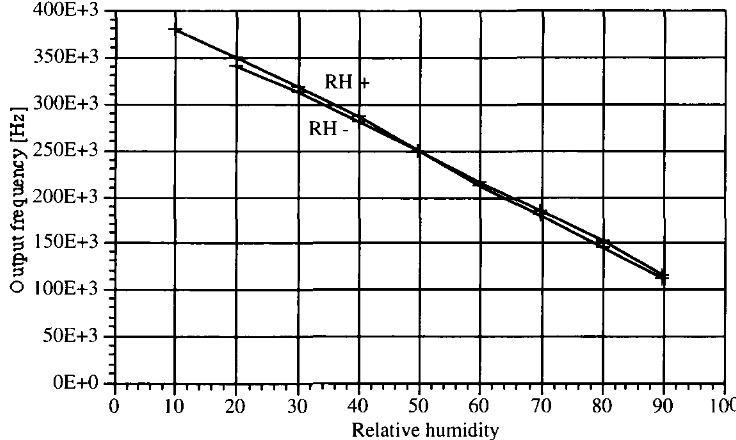


Figure 3-16. Measured response of capacitive humidity microsensor.

3.5.3 Magnetotransistor

The lateral magnetotransistor introduced in Section 3.3.2 operates at the collector current level of 100 μA. The current gain is smaller than one and poorly controlled. Therefore, an auxiliary conditioning circuit is required for both biasing the device on the proper operating point and reducing the strong common mode (100 μA) current that masks the small current difference signal [25]. This circuit is shown in Figure 3-17. The operating point is defined by a reference current, I_{REF}, externally supplied to the circuit. This current is compared with the average value of the two lateral collector currents by means of current mirrors $M_{1,2}/M_{3,4}$, and M_5/M_6. Transistor M_7 closes the loop and produces the necessary base current. Therefore, the circuit is biased at the desired operating point:

$$\frac{I_{C1} + I_{C2}}{2} = I_{REF} \ . \tag{3-9}$$

The differential current signal is extracted by the current subtraction part of the circuit. The device output currents I_{C1} and I_{C2} are mirrored into transistors M_8 and M_9. In the clock

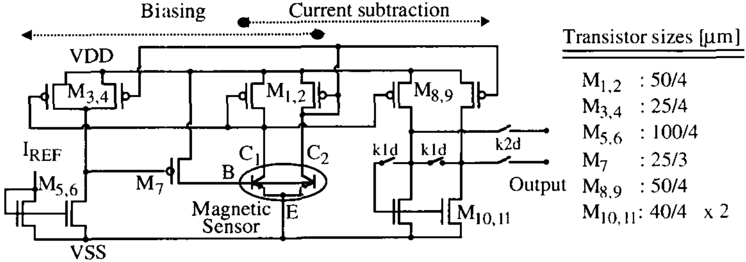


Figure 3-17. Front-end circuit for the magnetic sensor.

phase *k1d* the sum of these currents is stored on M_{10} and M_{11}. Then, in phase *k2d*, these transistors subtract from the output, ideally, the common mode component of the current. There is a residual common mode component, however, produced by charge injection into the gates of M_{10} and M_{11} and by channel length modulation due to the actual voltage on the nodes to which the output current is supplied. To reduce these effects, charge injection compensation schemes were used and the charge memory cell (M_{10} and M_{11}) is actually composed of two pairs of transistors in cascade topology.

The first integrator on the sigma-delta modulator is actually built around an operational transconductance amplifier (OTA). Therefore, the current signal, I_{IN}, produced by the sensor can be applied to this stage directly at the output of the amplifier. This is equivalent to generating a signal-dependent offset voltage at the input of the OTA. Since this stage is offset-compensated, the input current must be conveniently chopped. To this end, output switches are included in the auxiliary conditioning circuit in Figure 3-17. By simple analysis of the circuit, the effect is equivalent to an input voltage signal at the reference voltage input of the modulator given by

$$V_{\mathrm{INequiv}} = \frac{I_{\mathrm{IN}}}{G_{\mathrm{MOTA}}}\left[\frac{C_1}{C_{11}}\left(\frac{C_{11\mathrm{p}}}{C_1}+1\right)\right] \tag{3-10}$$

where G_{MOTA} is the transconductance of the OTA and C_{11p} is the parallel combination of C_{11} and the parasitic capacitance to ground at the OTA input node. According to the capacitance values designed, the term in square brackets is around 7. The resulting output code, represented by the variable D between -1 and $+1$, is expressed as

$$D = \frac{S I_{\mathrm{C1,2}}}{V_{\mathrm{REF}}\, G_{\mathrm{MOTA}}}\left[\frac{C_1}{C_{11}}\left(\frac{C_{11\mathrm{p}}}{C_1}+1\right)\right]B \tag{3-11}$$

where S is the magnetic sensitivity, B is the magnetic induction density and $I_{C1,2}$ stands for the average of I_{C1} and I_{C2}.

A more efficient way of applying the input current signal to the modulator is to use the drain nodes of the OTA input differential pair, when they are available. The reason is that these nodes present both a lower impedance and a lower signal swing than the output nodes. This is the solution we have implemented. The amplifier in the modulator was redesigned with the additional current input, and a transconductance $G_{MOTA} = 140$ µS.

One drawback of this approach is that the transfer characteristic reflects the non-linearity of the amplifier differential pair. This is due to the fact that the input current signal is applied at a different circuit point than the compensating feedback. Therefore, the amplifier input is not a virtual ground. In order to minimize the distortion, the input transistors were biased with a large overdrive voltage of 0.8 V in order to maximize the linear region.

For testing the prototypes, we set the clock frequency to 500 kHz. The circuit sensitivity depends on the value of the reference voltage used as predicted by Equation (3-11). For $V_{REF} = 0.1$ V, we obtain a full-scale range from about -400 to $+400$ G as illustrated in Figure 3-18. The offset of about 100 G is compensated by externally injecting a correcting current in the output of the conditioning circuit of Figure 3-17, before the output switches. With the sensor switched off, the circuit shows no magnetic field response even with a field as high as 1 T.

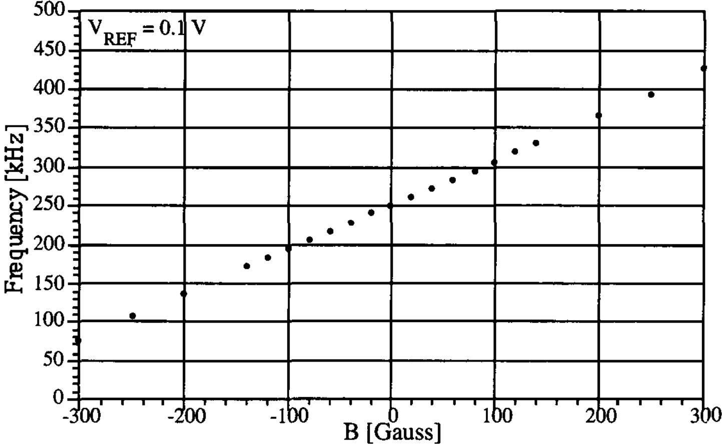


Figure 3-18. Measured response of magnetotransistor with 0.1 V reference voltage.

3.5.4 Thermoelectric Power Sensor

Mechanical and thermal sensors can implement an integrating function. By judicious design of such sensors, the whole first integrator stage of the sigma–delta modulator can be replaced by the sensor element.

A specific example is the micromachined ac power sensor or thermoconverter introduced in Section 3.3.3. Through the Seebeck effect, a voltage proportional to the mean square value of the input signal (heating voltage V_{IN}) is produced at the output. Since the thermal time constant of the device is much larger than the period of the sampling clock, the sensor can act as an integrator with a small loss when inserted in the modulator loop.

Two thermoconverters can be combined, one driven by the input signal and the other by the feedback signal of a sigma–delta modulator, as illustrated in Figure 3-19. This combination completely substitutes the first integrating stage of the modulator. The inner feed back loop present in the conventional architecture for stability reasons is no longer required as the integrating devices themselves have loss. The resulting modulator, however, is fundamentally

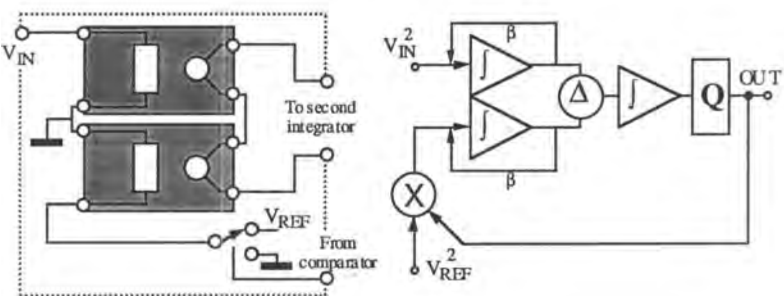


Figure 3-19. Left: Interconnection of thermoconverters for implementing the first integrator stage. Right: Equivalent block diagram of resulting modulator.

different from the conventional second-order sigma-delta modulator. However, for small integrator losses, the transfer function for the quantization noise retains the desired second-order characteristic. A circuit analysis is not practical since it contains both continuous-time and discrete-time circuit blocks. However, through extensive computer simulations on MATLAB, the proper operation with respect to linearity and stability has been verified.

Owing to the low output data rate required from a power sensor, the oversampling ratio can be made very large. This allows the digital decimator to be implemented with a simple counter and reset circuit. In this case, the actual noise-shaping function is irrelevant.

This configuration of the two power converters has some important properties. The circuit operation forces both devices to the same input power level. Owing to their close proximity on the chip, any deviation in the transfer characteristic is mutually cancelled. In particular, the non-linearity appearing at high power levels due to the thermal coefficient of the heating resistor is compensated. The output bit stream is a digital representation of the mean square value, or power, of the input voltage signal.

Care must be taken in the design of the second modulator stage. First, the signal level is very low, and second, any offset on the integrator will impose a limit at the low-power range. To this end, the complete modulator must be redesigned [18]. The modified circuit is shown in Figure 3-20. The transconductance amplifier uses a dynamic offset compensation. Capacitors C_1 and C_2, at the auxiliary input of the amplifier, store the offset voltage evaluated during clock phase ϕ_1. The amplifier transconductance, G_M, together with the load capacitors C_3 and C_4, constitute the integrator. The AND gate serves the purpose of shaping the digital signal driving the feedback device as a return-to-zero waveform. This compensates for the rise/fall edge asymmetry on the drive waveform. High digital interference suppression is obtained thanks to the fully differential topology and the clocking scheme, where great care was taken to minimize charge injection effects.

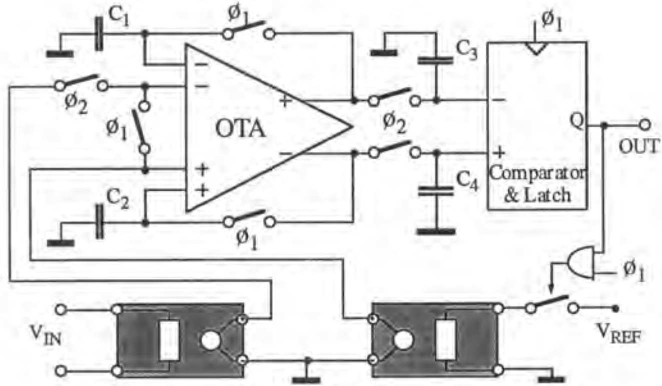


Figure 3-20. Schematic diagram of modulator containing two thermoconverters.

The system has been integrated in two separate chips, one with the thermoconverter device and the other with the circuit. The resulting characteristic is shown in Figure 3-21 for a clock frequency of 100 kHz and a reference voltage of 1.25 V. The reading set-up is the same as in Figure 3-15.

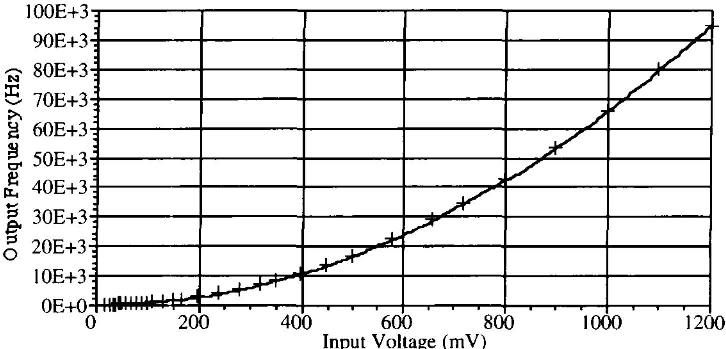


Figure 3-21. Experimental characteristic of power sensor circuit compared with square fit.

3.6 Interface with Minimum Analog Content

In this section we present a new modulator architecture for A/D conversion. This modulator allows a great simplification of the analog front end at the expense of a more complex digital signal processing for the bit stream decoding. Moreover, the system can be partitioned into two blocks communicating through noise-immune signal lines. These two features are desirable for sensor interfaces in remote locations and when only a minimal amount of circuitry can be co-integrated with the sensor.

In certain sensor applications, it is not possible to include more than only a minimal amount of circuitry on-chip. Examples are sensors using sophisticated fabrication processes. In these cases, the process is not optimized for IC devices and their parameters are not well controlled. Another example is a sensor located in harsh environment causing appreciable degradation of the on-chip IC devices.

Therefore, there is a need for an interface architecture that keeps the performance of the second-order sigma-delta modulator, but uses a less complex circuit, like that of the first-order modulator.

3.6.1 System Architecture

The circuit principle of such a system is illustrated with the block diagram shown in Figure 3-22. It is similar to the conventional second-order sigma-delta modulator. The difference is that the comparator has been moved to the input of the second integrator. This allows a digital implementation of this integrator by a simple up/down counter. The analog content of the system is reduced to a single integrator as in the first-order modulator. The drawback is that the feedback D/A converter is no longer an inherently linear 1-bit circuit. In order to make this architecture of practical value, an alternative implementation of the feedback loop must be found.

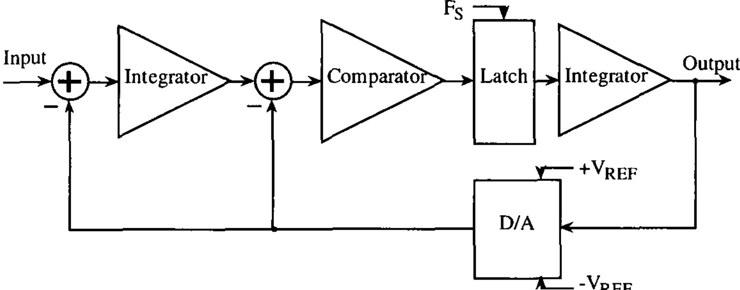


Figure 3-22. Block diagram of minimal analog interface.

Our solution is shown in Figure 3-23. Two different D/A converter circuits are used, one in each feedback loop [41, 42]. For the inner feedback loop, we employ a conventional multi-bit D/A converter. Any non-linearity in this component is divided by the first integrator gain, when referred to the input. This factor is at least equal to the oversampling ratio. Therefore, no stringent linearity requirements are demanded on the D/A converter.

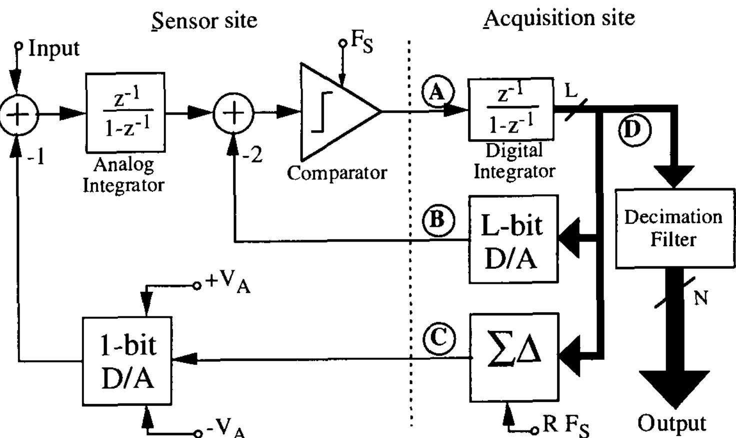


Figure 3-23. Implementation of minimal analog interface.

On the other hand, the D/A converter in the outer feedback loop determines the overall system linearity. It is implemented with a 1-bit D/A converter and a truncator circuit that translates the L-bit word into a 1-bit word at a higher sampling frequency, RF_S. This truncator circuit consists of a digital first-order sigma-delta modulator. This arrangement makes possible the implementation of an inherently linear D/A converter at the expense of a higher operating speed and an additional high-frequency quantization noise source. The baseband component of this noise can be made sufficiently low by increasing the secondary oversampling ratio R.

This architecture has the additional property that it can be segmented into two separate blocks, one at the sensor site and the other at the data acquisition site. The block at the sensor

site consists of a simple analog circuit, which can be easily co-integrated with the sensor device. The second block is the high-speed digital signal processor and is implemented at the data acquisition site. In this way, the communication between the two blocks is performed solely through two digital lines and a non-critical analog line, nodes A, B, and C. A high degree of immunity to interference is thus obtained which makes possible the placement of the interface front-end on a remote and noisy location.

This approach offers the following advantages desirable for sensor systems in remote and aggressive environments:

- the circuit at the sensor site can be easily co-integrated on the same chip since it requires only very simple analog components and its operation is very robust;
- the sensor site can be located at a physically remote point as the communciation with the data acquisition site is carried out already in a noise-immune digital format.

The linearized signal flow graph of this structure is shown in Figure 3-24, where the comparator is modeled as an additive noise source and the D/A converters are considered ideal. It can be immediately recognized that this structure is similar to a conventional second-order noise-shaping modulator where the quantizer has been moved to the input of the second integrator stage. Therefore, the signal transfer function is unchanged and remains unity. The main difference is that now the quantized variable is the first derivative of the input signal. As will be shown next, the signal-to-noise ratio in the baseband for this architecture is very close to that obtainable with the conventional one.

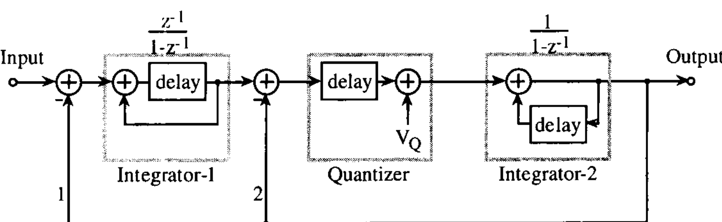


Figure 3-24. Linearized signal flow graph of interface in Figure 3-23.

3.6.2 Circuit Analysis

The circuit analysis will be divided into three parts. First the quantization noise is calculated, then the input signal amplitude limits are defined and finally the effect of the secondary quantization noise introduced by the oversampled D/A converter is studied.

Quantization noise. For a comparator step size Δ (meaning that the comparator output states are $+\Delta/2$ and $-\Delta/2$), the quantization noise power is

$$E_{Q1} = \frac{\Delta^2}{12} \tag{3-12}$$

which we assume to be white. The transfer function to the output is

$$T_{Q1}(z) = (1 - z^{-1})\, z^{-1} \,. \tag{3-13}$$

We use a positive representation of frequency and the power of a sampled signal is assumed to be contained in the frequency band 0 to $F_S/2$. The power spectrum of the quantization noise at the output is then

$$n_1(f) = \frac{E_{Q1}}{F_S/2}\left[2\sin\left(\frac{\pi f}{F_S}\right)\right]^2 \tag{3-14}$$

and the in-band power P_1 assuming an oversampling ratio $M = F_S/(2B)$ is

$$P_1 = \int_0^B n_1(f)\,\mathrm{d}f \approx \frac{\Delta^2\pi^2}{36\,M^3} \tag{3-15}$$

where the first three order terms in the series expansion of Equation (3-14) were kept.

Input limits. Since the first derivative of the input signal is quantized, it is the signal slope that is limited and not its amplitude. The maximum slope occurs for a sinusoid of maximum amplitude V_A at the baseband limit and it must satisfy the condition

$$V_A\,2\pi B \le \frac{\Delta}{2}F_s \tag{3-16}$$

or

$$V_A \le \frac{M}{2\pi}\Delta\,. \tag{3-17}$$

The full-scale signal power is then given by

$$E_0 = \frac{1}{2}\left(\frac{M}{2\pi}\Delta\right)^2. \tag{3-18}$$

Secondary quantization noise. The shaped quantization noise generated by the secondary modulator can be considered as an additive noise source at the system input. We consider a first-order loop with step size equal to $2\,V_A$ and running at an oversampling rate RF_S. Its power spectrum, according to Equation (3-12), is given as

$$n_2(f) = \frac{E_{Q2}}{F_S/2}\left[2\sin\left(\frac{\pi f}{F_S}\right)\right]^2 \tag{3-19}$$

with

$$E_{Q2} = \frac{(2\,V_A)^2}{12}\,. \tag{3-20}$$

At the input of the first integrating stage, the sample rate is reduced down to F_S by averaging every R samples. This operation is equivalent to $R:1$ decimation, preceded by the filter

$$T_{Q2}(z) = \frac{1}{R}\,\frac{1-z^{-R}}{1-z^{-1}}\,. \tag{3-21}$$

This corresponds to the transfer function in frequency

$$T_{Q2}(f) = \frac{1}{R} \frac{\sin\left(\frac{\pi f}{F_S}\right)}{\sin\left(\frac{\pi f}{RF_S}\right)} . \tag{3-22}$$

The effect of decimation is to fold back the frequency range [0; RF_S] into the frequency range [0; F_S]. The transfer function from the input of the first integrator to the output of the system is unity. Hence the power spectrum of the secondary quantization noise at the output is given by

$$n_{Q2}^{OUT} = \sum_{i=0}^{R-1} n_2(f - iF_S) \left[\frac{1}{R} \frac{\sin\left(\pi \frac{f - iF_S}{F_S}\right)}{\sin\left(\pi \frac{f - iF_S}{RF_S}\right)} \right]^2 = \frac{E_{Q2}}{RF_S/2} \frac{1}{R} \left[2 \sin\left(\frac{\pi f}{F_S}\right)\right]^2 \tag{3-23}$$

and the power in the band is

$$P_{Q2}^{OUT} = \int_0^B n_2^{OUT}(f)\, df \approx \frac{\Delta^2}{36MR^2} . \tag{3-24}$$

Design guidelines. The design parameters for this system are the oversampling ratios, M and R, and the comparator step size, Δ. The system specifications are defined by the maximum input amplitude, A_{MAX}, the desired signal-to-noise ratio, *SNR*, and the bandwidth, B. The design equations then read

$$SNR = \frac{E_0}{P_{QI} + P_{Q2}^{OUT}} = \frac{\frac{M^2}{8\pi^2}}{\frac{\pi^2}{36M^3} + \frac{\Delta^2}{36M\,R^2}} \tag{3-25}$$

and

$$V_A = \frac{M}{2\pi} \Delta . \tag{3-26}$$

A conventional second-order modulator has a signal-to-noise ratio [35]

$$SNR = \frac{15M^5}{2\pi^4} . \tag{3-27}$$

As a comparison, in the case when the primary quantization noise, P_{QI}, is dominant, Equation (3-25) simplifies to

$$SNR = \frac{9M^5}{2\pi^4} \tag{3-28}$$

showing a degradation of a mere 2.2 dB relative to the conventional sigma-delta modulator.

We finally note that the optimum *sinc* filter in the digital decimator is of order two since the quantization noise is first-order shaped. Therefore, the complexity of the decimator filter is relaxed when compared with the conventional second order sigma-delta modulator that requires a third-order *sinc* filter.

3.7 Conclusion and Outlook

We have illustrated the IC way to sensor modules comprising sensing elements and on-chip signal conditioning circuitry by exemplary microsensors for humidity, magnetic field, and ac power made by CMOS technology combined with post-processing deposition and etching. The crucial role of on-chip circuitry was illustrated by an indispensable capacitance meter for humidity sensors, biasing and readout circuit for dual-collector magnetotransistors, and a comparator-integrator arrangement for the ac power sensor. We demonstrated the promise of sigma-delta modulator and oversampling schemes. The trend to less analog content was illustrated in terms of an alternative modulator architecture for A/D conversion.

In view of the growing sophistication of IC technology (notably more robust metal connections and pads) and interface circuit design, we expect a rapid expansion of IC microsensors. Future work should address the measurement of process-dependent properties of the IC materials, on-wafer test procedures for sensor functions, low-stress dicing for chips involving micromechanical structures, and IC sensor chip packaging.

3.8 Acknowledgments

The authors thank Dr. T. Boltshauser, R. Castagnetti, D. Jaeggi and P. Malcovati of ETH Zurich for contributing to this chapter. The Swiss Foundation for Research in Microtechnology (FSRM), the Committee for the Support of Scientific Research (KWF), and the Swiss Federal Priority Program LESIT are acknowledged for financial support.

3.9 References

[1] Baltes, H., Moser, D., Völklein, F., in: *Sensors*, Vol. 7, Bau, H. H., de Rooij, N. F., Kloek, B. (eds.); Weinheim: VCH, 1993, pp. 13-55.
[2] Maenaka, K., in: *Sensors*, Vol. 7, Bau, H. H., de Rooij, N. F., Kloek, B. (eds.); Weinheim: VCH, 1993, pp. 173-204.
[3] Baltes, H., Castagnetti, R., in: *Semiconductor Sensors*, Sze, S. M. (ed.); New York: Wiley, 1994, pp. 205-269.
[4] Krummenacher, P., Oguey, H., *Sensors Actuators A* **21-23** (1990) 636-638.
[5] Kramer, J., Seitz, P., Baltes, H., in: *Transducers 91 Digest of Technical Papers;* New York: IEEE, 1991, pp. 727-729.
[6] Baltes, H., *Sensors Actuators A* **37-38** (1993) 51-56.
[7] Baltes, H., *Phys. Scr.* **T 49** (1993) 449-453.

[8] Yoon E., Wise, K. D., *IEEE Trans. Electron Devices* **ED-39** (1992) 1376-1386.
[9] Riethmüller, W., Benecke, W., Schnakenberg, U., Wagner, B. *Sensors Actuators A* **31-32** (1991) 121-124.
[10] Sauer, B., Gottfried-Gottfried, R., Hase, T., Kück, H., *Sensors Actuators A* **41-42** (1994) 582-584.
[11] Nguyen, C. T.-C., Howe, R. T., *IEDM Technical Digest,* Piscateway: IEEE, 1993, pp. 199-202.
[12] Boltshauser, T., *CMOS Humidity Sensors,* PhD Thesis No. 10329; Zurich: Physical Electronics Laboratory ETH, 1993.
[13] Moser. D., Baltes, H., in: *Micromechanical Systems,* **DSC-Vol. 40,** Cho, D., Peterson, J. P., Pisano, A. P., Friederich, C. (eds.); New York: ASME, 1992, pp. 91-101.
[14] Moser, D., *CMOS Flow Sensors,* PhD Thesis No. 10059; Zurich: Physical Electronics Laboratory ETH, 1993.
[15] Moser, D., Baltes, H., *Sensors Actuators A* **37-38** (1993) 33-37.
[16] Brand, O., Baltes, H., Baldenweg, U., in: *Transducers 93 Digest of Technical Papers;* Tokyo: IEE Japan, 1993, pp. 646-649.
[17] Lenggenhager, R., Baltes, H., in: *Transducers 93 Digest of Technical Papers;* Tokyo: IEEE Japan, 1993, pp. 1008-1011.
[18] Jaeggi, D., Azeredo Leme, C., O'Leary, P., Baltes, H., in: *Transducers 93 Digest of Technical Papers;* Tokyo: IEE Japan, 1993, pp. 462-465.
[19] Brand, O., Lenggenhager, R., Baltes, H., in: *IEDM Technical Digest;* Piscateway: IEEE, 1993, pp. 195-198.
[20] Brand, O., Baltes, H., Baldenweg, U., in: *Proc. IEEE MEMS 1994,* pp. 33-38.
[21] Lenggenhager, R., *CMOS Infrared Sensors,* PhD Thesis; Zurich: Physical Electronics Laboratory ETH, 1994.
[22] Boltshauser, T., Azeredo Leme, C., O'Leary, P., Baltes, H., *Sensors Actuators B* **15-16** (1993) 75-80.
[23] Boltshauser, T., Häberli, A., Baltes, H., *Sensors Mater. 5* (1993) 125-134.
[24] Maenaka, K., Maeda, M., *Sensors Mater.* **5** (1994) 265-284.
[25] Castagnetti, R., Azeredo Leme, C., Baltes, H., *Sensors Actuators A* **37-38** (1993) 698-702.
[26] Hsieh, K., Gray P., Senderowicz, D., Messerschmitt, D., IEEE J. *Solid-State Circuits* **16** (1981) 708-714.
[27] Pimbley, J., Michon, G., *IEEE Trans. Circuits Syst.* **38** (1991) 1086-1090.
[28] Malcovati, P., Azeredo Leme, C., Lenggenhager, R., Maloberti, F., Baltes, H., (1994), in: *Proc. IMTC 94,* Piscateway: IEEE, 1994, pp. 652-655.
[29] Azeredo Leme, C., Filanovsky, I., Baltes, H., *Electron. Lett.* **28** (1992) 1153-1155.
[30] Azeredo Leme, C., Lenggenhager, R., Baltes, H., in: *Proc. IEEE International Symposium on Circuits and Systems;* Piscateway: IEEE, 1992, pp. 1844-1846.
[31] Filanovsky, I., Azeredo Leme, C., Piskarev, V., Baltes, H., in: *Proc. IEEE International Symposium on Circuits and Systems;* Piscateway: IEEE, 1992, pp. 1836-1839.
[32] Azeredo Leme, C., Baltes, H., in: *Proc. IEEE International Symposium on Circuits and Systems;* Piscataway: IEEE, 1993, pp. 1397-1400.
[33] Malcovati, P., Azeredo Leme, C., O'Leary, P., Maloberti, F., Baltes, H., IEEE *J. Sol. St. Circuits* **29** (1994) 963-966.
[34] Dijkmans, E., Naus, P., in: *Proc. 15th European Solid-State Circuit Conf.;* 1989, pp. 35-63.
[35] Candy, J., Temes, G., in: *Oversampling Delta-Sigma Data Converters: Theory, Design and Simulation;* New York: IEEE 1991, pp. 1-29.
[36] Azeredo Leme, C., Chevroulet, M., Baltes, H., in: *Proc. IEEE Interrnational Symposium on Circuits and Systems;* Piscateway: IEEE, 1992, pp. 1828-1831.
[37] Eichenberger, C., Guggenbühl, W., *IEEE Trans. Circuits Syst.* **37** (1990) 256-264.
[38] Shieh, J., Patil, M., Sheu, B., *IEEE J. Solid-State Circuits* **22** (1987) 277-281.
[39] Huang, Q., Maguire, P., in: *Proc. 11th European Conference on Circuit Theory and Design;* Amsterdam: Elsevier, 1993, pp. 1355-1360.
[40] Dias, V., Palmisano, G., O'Leary, P., Maloberti, F., *IEEE Proc. G* **139** (1992) 27-32.
[41] Azeredo Leme, C., Malcovati, P., Baltes, H., in: *Proc. IMTC 94,* Piscateway: IEEE, 1994, pp. 652-655.
[42] Azeredo Leme, C., Malcovati, P., Baltes, H., in: *Proc. IEEE International Symposium on Circuits and Systems;* Piscateway: IEEE, 1994, Vol. 5, pp. 381-384.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

4 Three-Dimensional Microsensor Technology

H.-J. ACHE, J. BÜRCK, W. FAUBEL, W. HOFFMANN, and J. REICHERT,
Institut für Radiochemie (IRCH),
Kernforschungszentrum Karlsruhe, Karlsruhe, Germany,
W. MENZ, B. BÜSTGENS, J. MOHR, C. MÜLLER, W. SCHOMBERG, and
M. STROHRMANN, Institut für Mikrostrukturtechnik (IMT),
Kernforschungszentrum Karlsruhe, Karlsruhe, Germany

Contents

4.1 The Basic Concepts of the LIGA Technique . . . 81
4.1.1 Mask Making . . . 82
4.1.2 X-Ray Lithography . . . 82
4.1.3 Electroplating . . . 83
4.1.4 Plastic Molding . . . 83

4.2 Modifications and Extensions of the LIGA Technique . . . 85
4.2.1 The Sacrificial Layer Technique . . . 85
4.2.2 The Microstructure with Different Level Heights . . . 85
4.2.3 Microstructures Combined with Membranes . . . 85

4.3 Microstructures in the LIGA Technique . . . 86
4.3.1 Examples of Optical Structures in PMMA . . . 86
4.3.2 Examples of Rigid Metallic Structures . . . 87
4.3.3 Examples of Flexible and Rotatable Structures . . . 88

4.4 The Microspectrometer in the LIGA Technique . . . 89
4.4.1 Introduction . . . 89
4.4.2 Theoretical Aspects for the Layout of a Grating Spectrograph . . . 89
4.4.3 Fabrication and Performance of a Microspectrometer Prototype . . . 91
4.4.4 Experimental Results . . . 93
4.4.5 Application Fields: The Microspectrometer as a Component of an Analysis System . . . 97

4.5 The Acceleration Sensor in the LIGA Technique . . . 98
4.5.1 Introduction . . . 98
4.5.2 Thermal Model . . . 99
4.5.3 The Concept of an "Intelligent" Sensor System . . . 102

4.6 Micropump Manufactured by Thermoplastic Molding . . . 109
4.6.1 Micropump Design . . . 109
4.6.2 Manufacturing Process . . . 109
4.6.3 Measurements . . . 111

4.7 Optochemical Sensor Systems for Detection of Toxic Substances in Gases and Liquids . . . 113
4.7.1 Introduction . . . 113

4.7.2 System Components 113
4.7.3 Analytical Microsystem 116

4.8 Electrochemical Microanalytical System (ELMAS) for the Ionometry of Liquid Media 117
4.8.1 Introduction 117
4.8.2 The ELMAS Concept 118
4.8.3 Potentiometric Microsensors 120
4.8.4 ELMAS Functional Model 122

4.9 Concepts for Miniaturized Analytical Microprobes: Integrated Optical NIR Evanescent Wave Sensor System for Chemical Analysis 123
4.9.1 Integrated Optical Evanescent Wave Sensor 123
4.9.2 Evaluation of the Sensor Performance 125
4.9.3 Spectroscopic Microsystem for Chemical Analysis 126

4.10 Near Field Thermal Lens System 127
4.10.1 The Thermal Lens Effect 128
4.10.2 Apparatus 129
4.10.3 Concept of Miniaturized NFTL System 130

4.11 Chemical Microanalyzer Systems: Their Role and Importance in Modern Chemical Analysis in Environmental Workplace and Process Control and Medical Diagnosis: a Résumé 131

4.12 References 132

The motivation for the development of microsystems engineering lies in the application of the concepts of microelectronics to mechanics, optics, fluidics and, ultimately, chemistry, biochemistry and medicine while trusting that the same enhancement in performance accompanied by a reduction in costs can be achieved as in microelectronics. However, attention should be paid to an essential difference between a microelectronic circuit and a mechanical microstructure: whereas the first generally extends in two dimensions only (the lateral extension is greater by powers of ten than the vertical one), techniques had to be developed for microstructure generation which also allow the third dimension of a structure to be shaped.

While the means of lateral pattern generation could be taken over from microelectronics in the form of photolithography, mostly with minor modifications, the individual techniques differ mainly in the methods applied to shape the third dimension.

Also the LIGA technique, which has been developed at the Karlsruhe Nuclear Research Center [5] and is continuously being improved and expanded, has benefitted from the technological potential of microelectronics. However, compared with silicon-based microengineering, it offers some major technological advantages which will be explained in more detail below.

4.1 The Basic Concepts of the LIGA Technique

The manufacture of a LIGA microstructure is represented schematically in Figure 4-1. Unlike in silicon-based technology, the source material is a polymer layer several hundred micrometers thick which is applied on to a substrate, most frequently a metallic base plate. The X-ray sensitive plastic is directly polymerized on to the base plate. Until now, poly(methyl

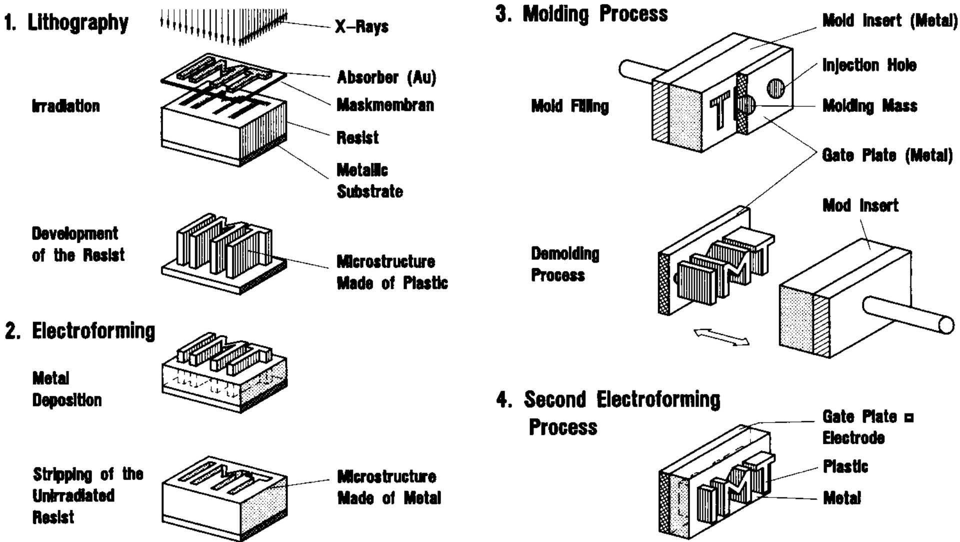


Figure 4-1. Basic process steps of the LIGA technique.

methacrylate) (PMMA) has almost exclusively been used as an X-ray resist. The pattern of a mask is transferred into the thick resist layer by means of highly parallel, high-intensity synchrotron radiation at a characteristic wavelength of 0.2–0.6 nm.

Owing to breakage in the long chains of molecules, the irradiated areas change their chemical stability and can be dissolved with a suitable developer. Techniques involving microelectrodeposition can be used to build up a complementary structure of metal, eg, copper, gold, nickel, and nickel alloys, while filling the gaps of the electrically non-conducting resist. With the metal structure so generated, almost any number of copies made of plastic can be reproduced with high accuracy of detail and at relatively little cost using injection molding, reaction resin casting, or embossing techniques [3]. For this, the metallic microstructure is used as the mold insert which is filled with the mass to be molded. After demolding these plastic structures can also be filled with metal by electrodeposition or serve as "lost molds" for the manufacture of ceramic microstructures.

These main process steps have given the technique its name, **LIGA**, a German acronym consisting of the letters **LI** (Röntgen**li**thographie, meaning: X-ray lithography), **G** (Galvanik, meaning: electrodeposition), and **A** (Abformung, meaning: molding).

4.1.1 Mask Making

Deep-etch X-ray lithography is the first and most important step in the LIGA process. Pattern transfer into thick resists calls for the use of high-energy synchrotron radiation ($\lambda_c = 0.2-0.6$ nm) and masks with a high X-ray contrast. In order to achieve a contrast of more than 200, very thick absorbers and highly transparent mask blanks are needed. During one typical lithography step the masks are exposed to X-ray doses of the order of 1 MJ/cm^2 and despite these high doses the mask has to withstand many exposures without any distortions.

For mask making, titanium is sputtered on to a metallic substrate made from Invar. The surface of this substrate must be finished mechanically with a very high accuracy. In order to obtain the desired stress in the titanium membrane, a precise annealing process must be applied. The free-standing and stretched membrane is obtained by selective etching of a window into the rear side of the substrate.

This titanium foil is spin coated with resist. The resist is patterned by electron-beam lithography or by synchrotron radiation. In a subsequent electroforming process the gold absorber structure is generated on the mask blank.

For alignment of these masks with respect to a substrate carrying structural information, alignment windows have to be provided in the titanium carrier foil of the mask [43]. No distortions could be found that were due to the opening of an alignment window if a gold rim around the window was applied to support the membrane.

4.1.2 X-Ray Lithography

In X-ray lithography, the pattern of a mask is transferred into a resist layer of up to 1 mm thickness using synchrotron radiation. Irradiation of the resist causes the long polymer chains to be cracked into short factures, i.e., a radiation-induced reduction of the molecular weight

[20]. The mechanisms leading to breakage depend on the nature and energy of the incident radiation.

By means of a suitable developer, the fractions of lower molecular weight are dissolved, leaving behind the intact molecular chains of the unirradiated areas. The developer must not swell up the unirradiated areas or attack them, which means that the dark area removal of the developer must be negligible. This is of particular importance in the development of structures with high aspect ratios such as deep trenches where the top sides of the structure are exposed to the developer much longer than the bottom of the trench.

4.1.3 Electroplating

The microstructures manufactured by means of x-ray lithography and consisting of plastic, most frequently PMMA, may be the final product in some cases, eg, microoptical components. However, in many cases metallic microstructures are needed which are manufactured by replacing the exposed and dissolved plastic with metal through electrodeposition. Using these metallic microstructures as mold inserts for a cost-effective injection molding, the LIGA technique is suitable for industrial mass production. In principle, microstructures can be manufactured by the LIGA technique which consist of any metals and alloys amenable to electrodeposition. However, as particular requirements must be made on the electrolytes used in the LIGA technique, extensive development work is necessary to supply the LIGA microstructures with a variety of pure metals and alloys.

Nickel is a standard material for electrodeposition in the LIGA technique. Nickel deposition is a well controlled process and the material exhibits high tensile strength and good resistance against corrosion. After the electroforming process the remaining polymer is stripped and the exposed plating base is removed by wet chemical etching.

Magnetic microactuators, eg, from NiFe alloys such as Permalloy (80–82 wt.-% Ni), are of particular interest if they can be manufactured by electroforming within the framework of LIGA. In this case, the NiFe alloy has to be deposited into the complementary polymer (resist) microstructure with homogeneous composition over the structural height of up to several hundred micrometers. A special difficulty is to preserve a constant alloy composition in complicated structures with different lateral geometries and aspect ratios (Figure 4-2).

Microstructural devices have been manufactured from the process described above. The coercitivity of the Permalloy microstructures obtained by electroforming was determined to be less than 5 Oe and the saturation moment was approximately 940 G. The determined grain size was of the order of 40 nm. Obviously, the magnetic properties of these Permalloy microstructures are superior to those of nickel microstructures of the same geometry manufactured by electroforming from a nickel sulfamate bath. Nickel prisms show a coercitivity of 80 Oe and a saturation moment of about 450 G [47].

4.1.4 Plastic Molding

In the preceding sections, techniques have been described that lead to the manufacture of a primary structure in PMMA or to a complementary metal structure by means of electrodeposition. The LIGA technique is of interest for commercialization on an industrial scale,

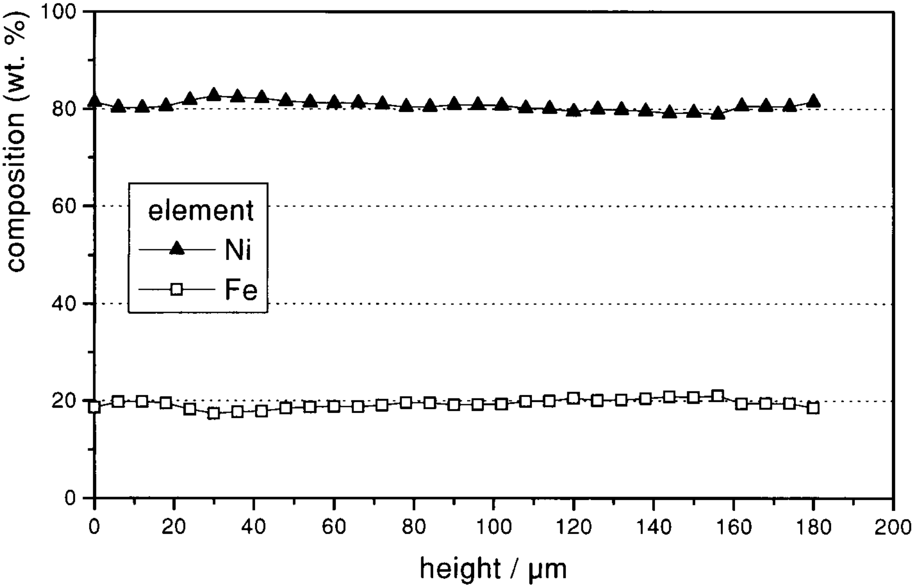


Figure 4-2. Composition gradient versus structural height of an electroformed structure of Permalloy.

especially owing to the possibility it provides of replication through injection molding, reaction casting, or embossing, ie, typical techniques lending themselves to mass production. Therefore, the molding techniques which have been developed especially with regard to the LIGA technique will be described in more detail below.

At present, modifications of the LIGA process are being developed with a view to quasi-monolithic integration of microstructures and microelectronics which will offer a competitive approach to the manufacture of microsystems. Previous investigation had made evident that building up LIGA microstructures on C-MOS circuits using x-ray lithography is difficult, since the C-MOS circuits are heavily damaged by high-energy radiation. Therefore, in order to fabricate LIGA microstructures directly on integrated circuits, molding processes seem to be a suitable method with regard to both compatibility and fabrication costs [10].

The process of hot embossing and the materials have been demonstrated to be compatible with C-MOS microelectronics. Compatibility tests were performed to prove the feasibility of this integration technique for industrial application.

As the molding technique for manufacturing movable microstructures can be combined with positioned molding, it will be possible to generate movable microstructures functionally connected with the C-MOS circuits on a processed silicon wafer without impairing the integrated circuits. This will be considerable progress in microsystem manufacture as movable microstructures (sensors and actuators) will be able to interact with the integrated circuits lying underneath.

4.2 Modifications and Extensions of the LIGA Technique

4.2.1 The Sacrificial Layer Technique

To build up a microsystem, microcomponents which can be used as sensors or actuators are essential. In many cases this requires mobile microstructures, such as acceleration sensors, microvalves, or motors.

Also with the LIGA technique mobile microstructures can be manufactured by the introduction of sacrifical layers (Figure 4-3). This has considerably extended the application of the LIGA technique because a wide range of materials and a great structural height are available for building optimum sensors and actuators and almost no limitations are imposed on the lateral shape of molding [11].

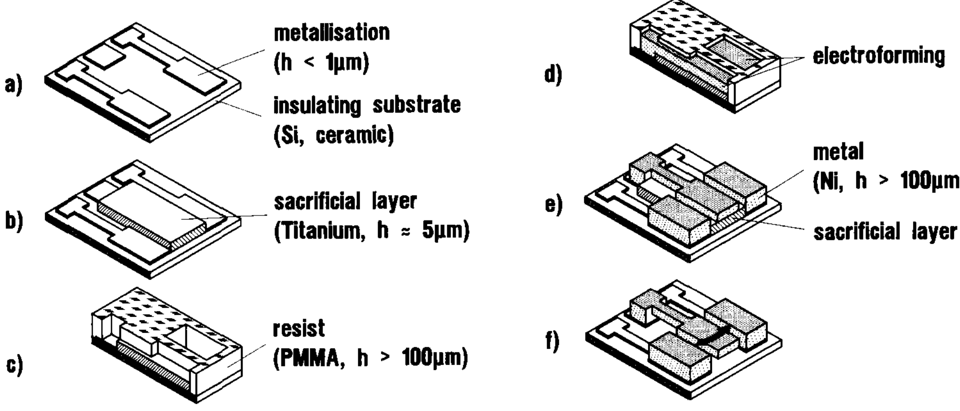


Figure 4-3. Sacrificial layer technique.

4.2.2 The Microstructure with Different Level Heights

For an increasing number of microstructural applications, more complex geometries with variably formed steps are of special interest. Within the framework of the LIGA process, our development activities are therefore aimed at making available a technology which allows the fabrication of vertical and inclined side walls and of stepped structures with a variable geometry.

On the basis of the combination of deep-etch x-ray lithography and electroforming, the process for the fabrication of two- or multi-step microstructures has been developed (Figure 4-4) [24].

4.2.3 Microstructures Combined with Membranes

For the handling of gases and liquids, rigid microstructures are needed which are bonded tightly to highly flexible functional elements. Therefore, a technique has been developed which

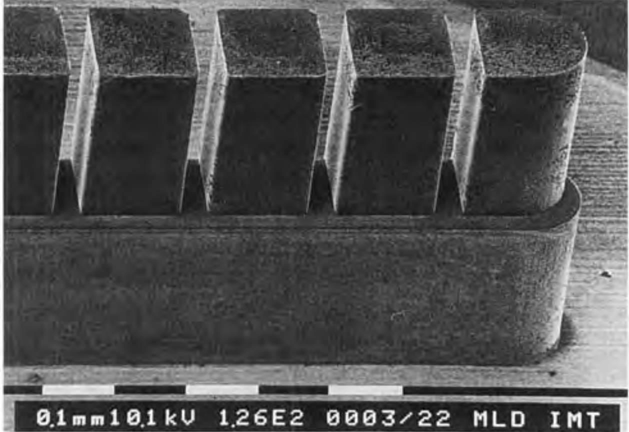


Figure 4-4. SEM of a two-stepped microstructure.

combines the advantages of membranes and microstructures manufactured by molding processes from a variety of materials [45]. With this technique micropumps and microvalves systems have been realized.

4.3 Microstructures in the LIGA Technique

4.3.1 Examples of Optical Structures in PMMA

The PMMA resist used in the LIGA technique has good optical transmission properties in the visible and near-infrared ranges. Therefore, the so-called "primary structures" manufactured from PMMA can be used in a variety of optical applications [31, 36]. This necessitates a surface roughness of the imaging planes of approximately one tenth the size of the wavelength of the light used. By phase shifting interferometry it has been demonstrated that the roughness of the lateral walls generated by x-ray lithography is of the order of 30–40 nm, which means that the major prerequisite of optical applications can be satisfied. It is also possible to manufacture microoptical components by cost-effective molding techniques. Besides PMMA, other plastics can be used which might exhibit favorable features in special applications.

In order to reduce the divergency-induced losses in the vertical direction of light-transmitting structures, a special three-layer resist system has been developed. It consists of a light-conducting core layer and two sheathing layers with a lower refractive index than the core layer. Thus, the light is guided in the core layer due to total reflection at the boundary layers. PMMA has been chosen for the core layer and a copolymer made of PMMA and tetrafluoropropyl methacrylate as the sheathing layer. Thanks to the copolymer fraction, ie, the fluorine concentration, the refractive index of the sheathing layer and hence the numerical aperture of the light-transmitting layer can be varied within broad limits and adapted to the

numerical aperture of the fibers coupled to it. When selecting the fluorinated sheathing layers the sensitivity to X-radiation has to be matched to that of the pure PMMA. Figure 4-5 shows a planar waveguide with the three-layer resist structure and a light-conducting fiber precisely positioned by means of a fiber duct simultaneously patterned.

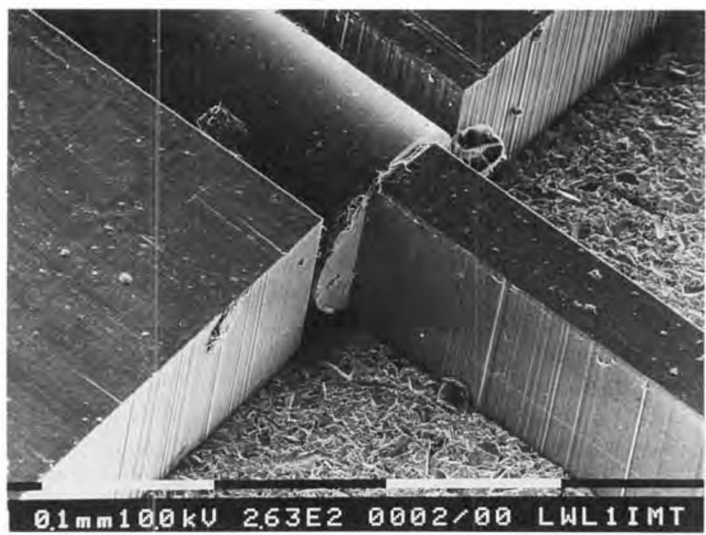


Figure 4-5. SEM of a light-guiding structure fabricated by the three-layer technique.

4.3.2 Examples of Rigid Metallic Structures

With the structure shown in Figure 4-6, band pass filters for the far-infrared region can be realized. These filters are fabricated by the LIGA technique with a precise periodic pattern of crossed "negative" dipole in a copper foil of 20–40 µm thickness. Depending on the geometry and the foil thickness, the width of the band pass and the absolute cut-off frequency can be controlled within a wide range.

Figure 4-6.
Band pass filter for far infrared radiation in the LIGA technique.

4.3.3 Examples of Flexible and Rotatable Structures

With the help of the sacrificial layer technique, microstructures which are suspended from the substrate either partly or fully can be manufactured in a single sequence of processes. These mobile microstructures have considerably extended the number of applications of the LIGA technique because this offers a means of sensor and actuator manufacture.

The first mobile LIGA structure manufacturd by means of the sacrificial layer technique was an acceleration sensor [11, 12]. This device will be discussed in detail in Section 4.5. The basic features of the acceleration sensor fabricated by this process can be seen from Figure 4-7, which shows an electroplated microstructure. A seismic mass suspended from a cantilever which was fabricated on the sacrificial layer is able to move between the two stationary electrodes firmly attached to the substrate. The change in capacitance between the seismic mass and the electrodes can be measured.

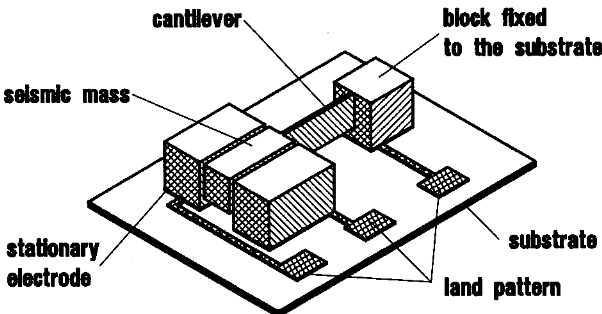


Figure 4-7.
Top: Design of an acceleration sensor in the LIGA technique. The seismic mass and the cantilever beam are suspended from the substrate by using a sacrificial layer. Bottom: SEM of a detail of the first accelerator design in the LIGA technique. The width of the cantilever beam is 10 μm and the height 100 μm. The gap between the seismic mass and counter electrode is 4 μm.

By contrast, in rotating microstructures the rotor must be fully detached from the substrate whereas the axle remains rigidly connected to the substrate. The first rotating LIGA structure to be manufactured was a microturbine which is driven by gases or liquids [30]. Figure 4-8 shows a microturbine made of nickel with a diameter of 130 μm, which is less than the height of 150 μm. The gap between the rotor and the axle in this case is only 5 μm. As conventional lubrication of the rotors would be very difficult owing to the very narrow gaps provided, air

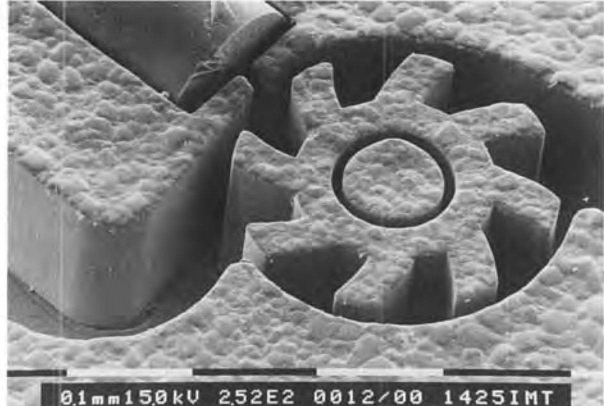


Figure 4-8.
SEM of a microturbine fabricated by the LIGA technique using a sacrificial layer. The diameter of the rotor is 130 μm and the structural height is 150 μm.

friction bearings are used. The speed of microturbines can be measured via a glass fiber laid into a prefabricated duct and illuminating the rotor. At the front faces of the rotor blades the light is reflected back into the same fiber, and by counting the light pulses the speed can be determined with high accuracy. In this configuration the microturbine could be used both as an actuator and as a flowmeter.

4.4 The Microspectrometer in the LIGA Technique

4.4.1 Introduction

Up to now, most research in micro-optics has been concentrated on the fabrication of integrated optical devices for mono-mode fiber applications. However, there is an increasing demand on micro-optical components for other applications such as microchemical analysis.

Grating spectrographs for multi-mode applications are normally fabricated by hybrid technology using, for example, a waveguide made of an accurately shaped glass to which a reflection grating must be precisely mounted [36]. The grating is usually fabricated by high-precision mechanical shaping or by a holographic technique. The light is coupled into the spectrometer by a fiber precisely adjusted to the grating. Both methods are very cost-intensive steps within the fabrication process.

By producing the grating of the spectrometer jointly with the fiber coupling trench and integrated with an array of photodiodes, it is possible to reduce the production costs and the volume of these grating spectrographs. This is possible by patterning a three-layer resist system by deep-etch X-ray lithography, which is the first step of the LIGA process [5].

4.4.2 Theoretical Aspects for the Layout of a Grating Spectrograph

The non-restrictive lateral patterning of LIGA microstructures opens up the possibility of fabricating a self-focusing grating with minimized aberrations. This requires a cylindrical

reflection grating with varying line spacing and a non-circular grating curve. The theoretical performance data have been calculated by a ray trace program [52].

When calculating the grating design, the wave propagation in a multi-mode slab waveguide has to be considered. As the wave vector can only propagate in directions defined by discrete mode angles q, the waves give rise to a self-consistent interference pattern consisting of a standing wave normal to the guiding plane with a spatial periodicity λ_p of the propagating interference wave and a propagating wave parallel to the guiding plane $\lambda_{swg} = \lambda_{vac/n}$. Theoretically, an aberration-free, self-focusing grating focuses only one effective wavelength λ_p into a diffraction-limited focus under the assumption of a point source. Therefore, a grating design independent linear dispersion or focus broadening in the case of wavelength longer than λ_{swg} is obtained.

With a concave blazed grating, a high efficiency around the determined blazed wavelength is obtained if the angles to be set over the whole grating are different. Owing to the free choice of the design in the LIGA process these blaze angles can be optimally adapted to the desired directions of reflection. The efficiency of a blazed concave grating can be further enhanced if it works near autocollimation, also called the Littrow configuration.

In Figure 4-9 the calculated intensity distribution is shown for a grating spectrograph which works in the first order and which is blazed to 740 nm. In this case the grating constant was chosen to be 2.5 µm and the step height of the grating is 0.26 µm. A glass fiber with an outer diameter of 125 µm and a core diameter of 85 µm was chosen as a detector. As can be seen, the intensity decreases with increasing distance of the wavelength to the blaze wavelength.

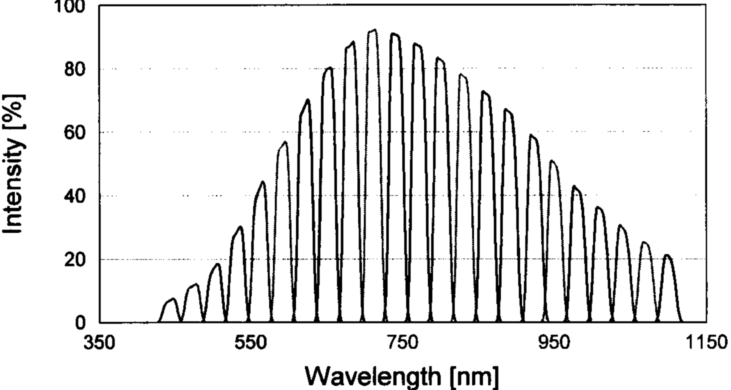


Figure 4-9. Calculated intensity distribution of a grating spectrograph blazed to 740 nm and working in the first order. A glass fiber with an outer diameter of 125 µm and a core diameter of 85 µm has been used as a detector.

Owing to technical restrictions, it is not yet possible to fabricate a grating with ideal rectangular edges. In Figure 4-10 the influence of the rounding on the efficiency is shown for a grating blazed in second order with a grating constant of 3.5 µm. As can be seen, the efficiency decreases by a factor of two for a rounding of 0.9 µm. With increasing radius of the rounding the grating approaches the shape of a sinusoidal grating, which has most of the in-

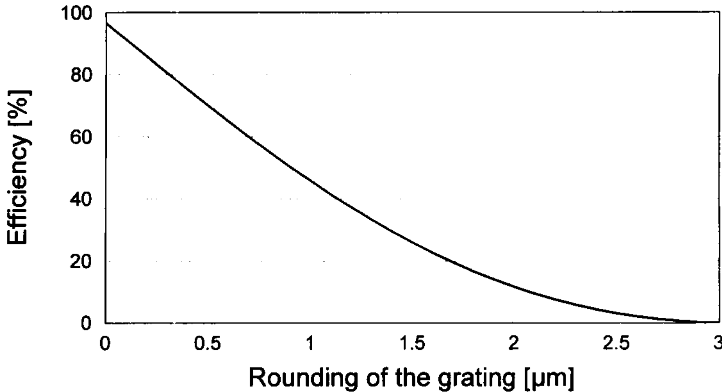


Figure 4-10. Influence of the rounding of the grating steps as a function of the rounding radius for a second-order blazed grating spectrograph with a grating constant of 3.5 µm.

tensity in first order. Therefore, it seems to be better to fabricate a blazed grating in first order. In this case the step height has to be much smaller, which increases the demands on the process technology.

4.4.3 Fabrication and Performance of a Microspectrometer Prototype

Three-Layer Resist System

For the fabrication of light-guiding components by deep-etch x-ray lithography, as required in the case of the grating spectrograph, a three-layer resist system with a step index profile has been developed [53]. PMMA or deuterated PMMA with a refractive index of 1.49 at 589 nm is used as core material. The cladding consists of a copolymer with 78 wt.-% methyl methacrylate and 22 wt.-% tetrafluoropropyl methacrylate, which results in a refractive index of 1.476 at 589 nm. The difference in the refractive index is 0.014 over the whole spectral range. Therefore, the numerical aperture of the slab waveguide fits precisely to the numerical aperture of the incoupling fiber.

The three-layer resist system is fabricated by welding polymer foils, which were polymerized from the monomer on to a polymer base plate made of an epoxy-phenol resin. Using this substrate material, which is insensitive to X-radiation, the thermal expansion coefficient of the base plate is well adapted to the resist material to reduce tensile cracking. The first layer of the three-layer resist is polymerized directly on to the chemically activated epoxy-phenol plate and machined down to a thickness which corresponds to the thickness of the cladding of the optical fibers. Subsequently a PMMA foil is welded on to this first layer and adapted by milling to the thickness of the fiber core. Finally, a cover foil is put on top of this sheet system. It serves as the top cladding layer of the light guiding assembly. During the welding process, which is performed slightly above the glass transition temperature, very little interdiffusion of the polymer chains is observed, resulting in a steep gradient index profile of the slab waveguide.

This three-layer resist system is exposed through a mask to synchrotron radiation (Figure 4-11) and subsequently treated with a special developer [54]. Thus fiber-adjustment trenches are integrated with the waveguiding component and the grating. Therefore, the optical fibers are adjusted exactly to the waveguides in the horizontal direction by means of these trenches.

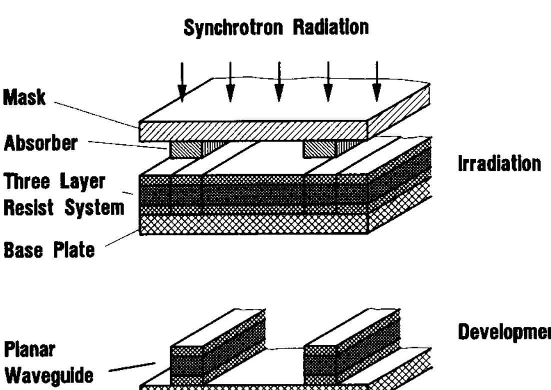


Figure 4-11. Principle of patterning of a three-layer resist system by deep-etch x-ray lithography to produce light-guiding components.

Optical Performance of Waveguides

For optical applications two major prerequisites have to be fulfilled: the attenuation within the spectral range has to be sufficiently small and the roughness of the optical surfaces should be within fractions of the wavelength to be used.

Measurements on cylindrically shaped bulk specimens made of PMMA show a mean attenuation of approximately 0.2 dB/cm in the wavelength range between 600 and 1300 nm (Figure 4-12). This value increases to 0.4 and 3 dB/cm at 900 and 1180 nm, respectively. To

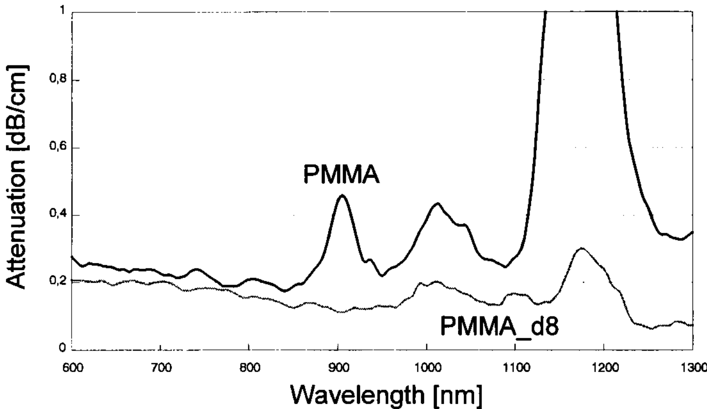
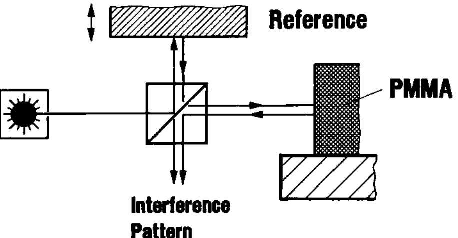


Figure 4-12. Attenuation of PMMA and deuterated PMMA in the wavelength range 600–1300 nm (cylindrical-shaped specimens).

manufacture low-loss structures for the near-IR region, deuterated PMMA (PMMA-d_8) is well suited. For this material the attenuation in the spectral range from 600 to 1300 nm is at about 0.1 dB/cm. Considering the short optical path lengths of micro-optical components, this attenuation value can be tolerated.

The roughness of the sidewall of the microstructures has been measured using phase-shifting interferometry [55]. In Figure 4-13 the surface roughness is presented. It is calculated from the interference pattern which results form superposition of the wave fronts reflected by a LIGA structure and an ideal reference mirror. As can be seen, the peak-to-peak roughness is in the range of only a few tens of nanometers, which is adequate for optical applications [56].



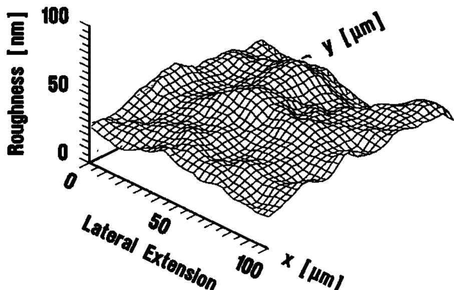


Figure 4-13.
Roughness of the side-wall of LIGA microstructures calculated from the interference pattern produced with phase-shifting interferometry.

4.4.4 Experimental Results

Principle of Layout of the Grating Spectrograph

Figure 4-14 shows the principle of the grating spectrograph with a self-focusing reflection grating. The polychromatic light, which is injected into the waveguiding component by an optical fiber (eg, 50/125) is dispersed at the grating, which is coated with gold or silver, and reflected toward the focal line. The spectrally divided light can be coupled out at the focal line by several optical fibers. It is also possible to pattern an inclined sidewall by x-ray lithography [57], where the direction of the light is changed from the horizontal to the vertical plane by internal reflection. This opens up the possibility of mounting the grating spectrograph on top of a diode array, which will be used to detect the different spectral parts of the light.

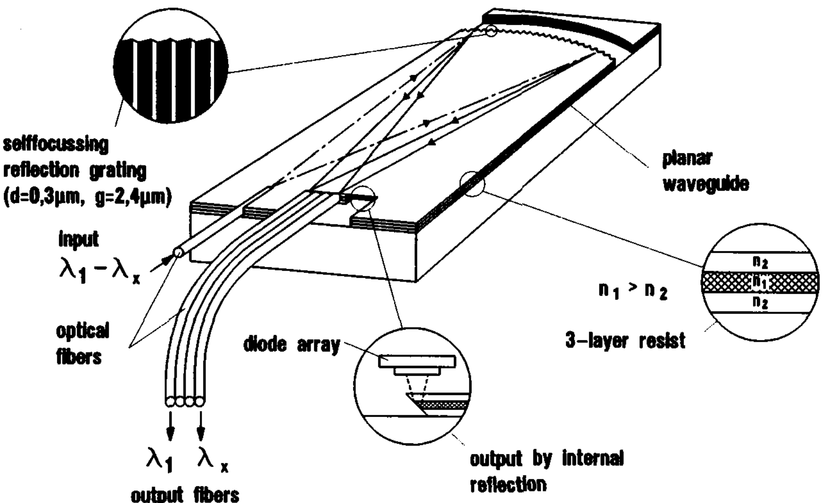


Figure 4-14. Layout of a grating spectrograph with self-focusing reflection grating fabricated by deep-etch X-ray lithography. The spectral light can be detected either by several fibers or by a diode array.

Design Data of the Grating Spectrographs

To cover the whole spectral range from about 400 to 1100 nm, different components which will work in different spectral ranges as well as in different orders have been designed. The design data are different depending on the spectral range, the required resolution, and the order in which the grating is blazed.

Mask Performance

The step height of the grating is in the range of only a few tenths of a micrometer, whereas the width of the grating may be in the range of micrometers. As was pointed out earlier, the efficiency of the grating depends considerably on the sharpness of the grating edge. Therefore, these gratings have to be fabricated with high accuracy, which gives rise to very high demands on the fabrication process of the mask and the resist pattern. Nevertheless, after optimization of the e-beam process for mask fabrication the gratings could be fabricated with the required tolerances. In Figure 4-15 a section of the absorber of a grating with a step height of only 0.256 µm is shown. The edges are very sharp and only very little rounding is to be seen.

To achieve high accuracy within the resist structure it is necessary to decrease the thermal load of the resist and the mask to avoid thermally induces distortions during the X-ray illumination. To show the high accuracy, a scanning electron micrograph of a detail of the reflection grating patterned into a three-layer resist system with a grating spacing of 3.4 µm and a step height of 0.5 µm is shown in Figure 4-16. In the picture the boundaries between the different layers can be seen from the horizontal lines in the photograph.

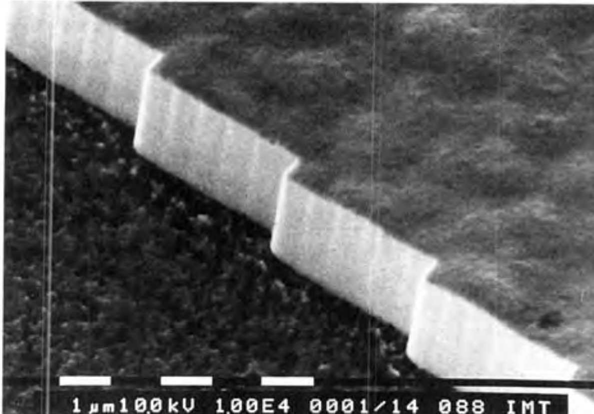


Figure 4-15.
SEM of a section of the 2.5 µm high gold absorber on the x-ray mask which is used to fabricate a grating with a step height on only 0.256 nm.

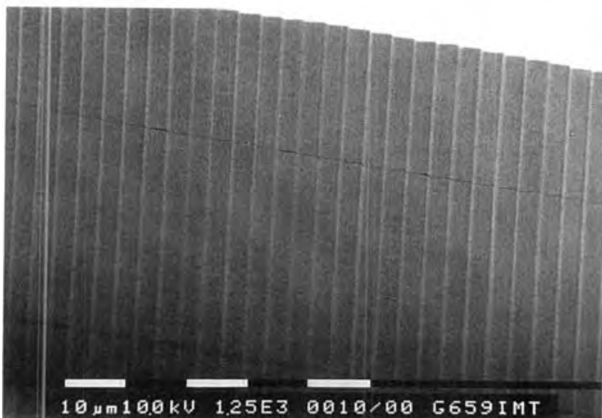


Figure 4-16.
SEM of a detail of the reflection grating patterned into a 125 µm high three-layer resist system (grating spacing, 3.4 µm; step height, 0.5 µm).

Measurements of the Grating Behavior

To determine the quality of the components, light from a monochromator was fed into the spectrometer. The outcoming intensities were measured as a function of the wavelength using a multi-mode fiber as a detector. The fiber was positioned in a trench at the focal line and the different positions of the fibers had a distance of 125 µm. The detected intensities were related to the intensities measured without the components with two butt-coupled fibers. Figure 4-17 shows the results of these measurements for the grating spectrograph which was designed to work in first order and blazed to 740 nm. The dispersion of the light and the cross-talk are excellent and in good agreement with the theoretical calculations. Also, the measured intensity which is at a maximum of 24% near the theoretical blaze wavelength, is good. It is even higher than the results for second-order grating spectrographs in spite of the smaller step height. The intensity of the wavelength far away from the blaze wavelength is also higher than the calculated values. Both facts indicate that there is a superposition of a sinusoidal grating and a blazed grating. Therefore, a slight rounding of this gratings is much more tolerable in the case of spectrographs which work in the first order.

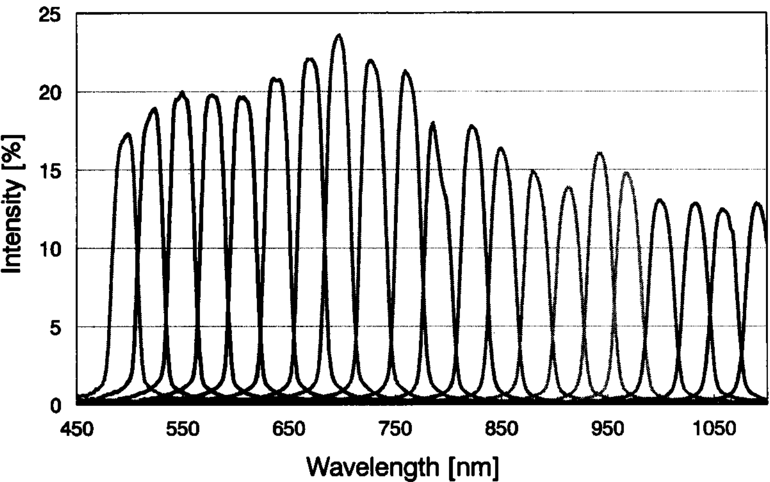


Figure 4-17. Intensities of the spectral light at the focal line detected by spectral fibers. The separation between the intensity maxima is 30 nm.

Adapting the Grating Spectrograph to a Diode Array

In Figure 4-18 a first result is shown where a grating spectrograph with an inclined sidewall at the focal line was adapted to a diode array. In this case also the dispersion of the light could be detected very well. The cross-talk, which seems to be rather high, results from the fact that the diode array which was encapsulated could not be positioned in the focal plane. This problem could be easily overcome by using a non-encapsulated diode array.

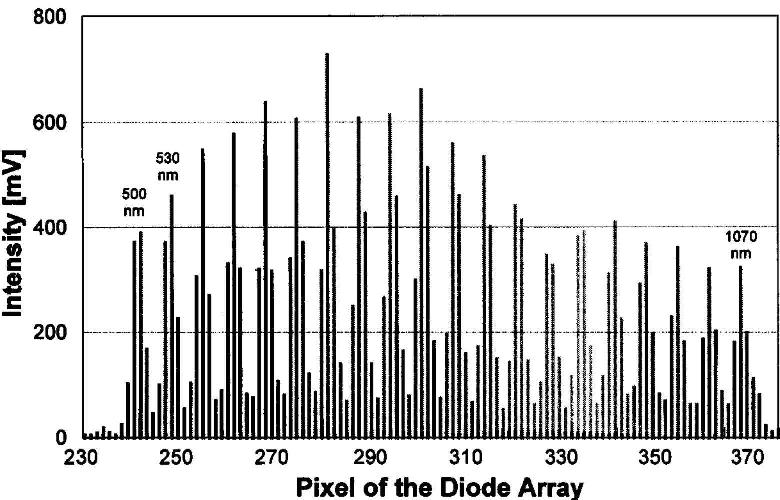


Figure 4-18. Spectral distribution measured with a diode array which was adapted to the grating spectrograph using an inclined side-wall at the position of the focal line. For the measurements discrete frequencies with a separation of 30 nm in wavelength were coupled in.

Further Improvements of Grating Spectrographs

For the grating spectrograph already measured the spectral resolution is 0.23 nm/μm, which means 2.9 nm per diode size of 12.5 μm. This value may be reduced to 1 nm per diode for a grating spacing of 1 μm. Nevertheless, the resolution depends also on the extended spot size of the incoupling fiber, which is 50 μm and which is imaged on a one to one basis on to the focal line. Therefore, at this spot size light which should be detected only by one diode is spread over at least four diodes. To avoid this, it is necessary to decrease the spot size of the incoupling fiber. This is possible on the one hand by using a taper which is patterned in the three-layer resist system behind the fiber fixing trench. On the other hand, a cylindrical lens can also be patterned. In both cases the numerical aperture of the incoupled light is increased, which increases the width of the whole element. In these two cases the whole intensity emitted by the incoupling fiber will be used. This is not the case if a slit diaphragm is patterned after the fiber fixing trench. In this case the intensity decreases whereas the dimensions of the grating spectrograph remain unchanged.

4.4.5 Application Fields: The Microspectrometer as a Component of an Analysis System

The described grating spectrographs are well suited for use in spectroscopic analysis in the visible and near-infrared regions, which lends itself to many applications. These include, eg, examinations in the enviromental area such as the detection and determiantion of different heavy metal ions in water or toxic agents in smoke. They can also be used in medicine to detect, eg, the oxygen content in blood. In Figure 4-19 the principle of an analytical system is shown which combines the spectrometer with fluid handling systems, optochemical sensors, and electronics. Thus a miniaturized intelligent, stand-alone device can be fabricated, as will be discussed in the Section 4.7.

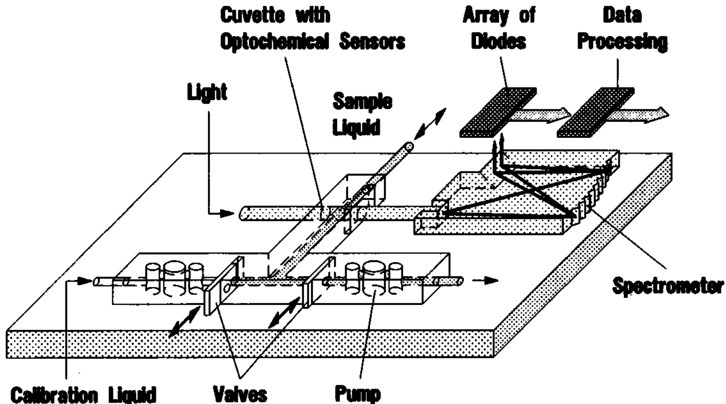


Figure 4-19. Principle of an overall analysis system which combines the spectrometer with fluid-handling components, optochemical sensors, and electronics.

4.5 The Acceleration Sensor in the LIGA Technique

4.5.1 Introduction

Combination of the LIGA process with a sacrificial layer technique offers the possibility of fabricating in one process sequence metal capacitive acceleration sensors with a movable seismic mass and stationary electrodes on top of the same substrate. The process allows micromechanical devices with a free geometry to be designed. Use has been made of this advantage to fabricate acceleration sensors with low temperature coefficients. This is achieved by designing the sensor partly with a positive and partly with a negative temperature coefficient. The sensor according to the design presented here is a 1 *g* sensor with a measured temperature coefficient of offset of 10^{-4} *g*/K in a temperature range between −10 and 100 °C.

One of the most attractive applications of mechanical sensors is in automotive engineering, where the sensors have to operate in a temperature range from −40 up to 125 °C with a minimum of thermal shift. This is the reason why temperature-resistant sensor elements are strongly needed.

Macro sensor elements with complicated layouts were fabricated in the past. As they are costly, these sensors must be replaced by batch-fabricated sensor elements based on microstructure technologies. Unfortunately, the ideas realized in the macro world are not transferable to the designs of microstructures because of the different methods of fabrication. For instance, the design of sensors fabricated by anisotropic silicon etching has its limits in crystallographic directions. Thus, for micromechanical sensors temperature effects are compensated by electronic circuits with complex signal conditioning. Frequently, this method of temperature compensation is suitable in first order only because of tolerances in the fabrication of the sensor and temperature elements.

In contrast to anisotropic silicon etching, the LIGA technology [5, 30] allows free lateral shaping. This advantage offers the possibility of transferring temperature-compensating strategies to the world of micromechanics.

Figure 4-20 shows a capacititve acceleration sensor fabricated by the LIGA process with a simple design [12, 31]. A seismic mass (SM) suspended from a cantilever (C) and fabricated on a sacrificial layer moves between two stationary electrodes (SE) firmly attached to the substrate. The seismic mass and each stationary electrode form a capacitor whose capacity depends on the acceleration of the sensor.

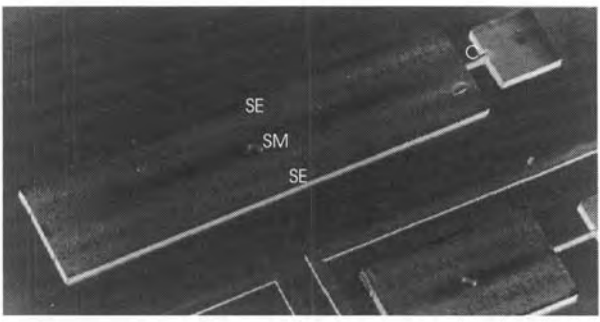


Figure 4-20.
SEM of a LIGA acceleration sensor.

The sensor element shown in Figure 4-20 has a length of 3000 μm and a width of 800 μm and is 200 μm high. The gap of the fabricated capacitors varies between 3 and 5 μm, which results in sensor capacities between 1.8 and 1.0 pF. In order to obtain a sensor with a sensitivity of 20%/g the spring constant is adapted to the width of the seismic mass which results in a resonance frequency of 550 Hz.

4.5.2 Thermal Model

For a design shown in Figure 4-20 the temperature coefficient of offset (TCO) is very high. The main reason is the difference of the thermal expansion coefficients of the ceramic substrate ($\alpha_C = 8.1 \cdot 10^{-6}$/K) and the sensor material nickel ($\alpha_{Ni} = 13.8 \cdot 10^{-6}$/K). Thus, a change in the temperature of structure results in a change in the width of the capacitor gap. This is illustrated in Figure 4-21, where the left border is the symmetry line of the whole structure.

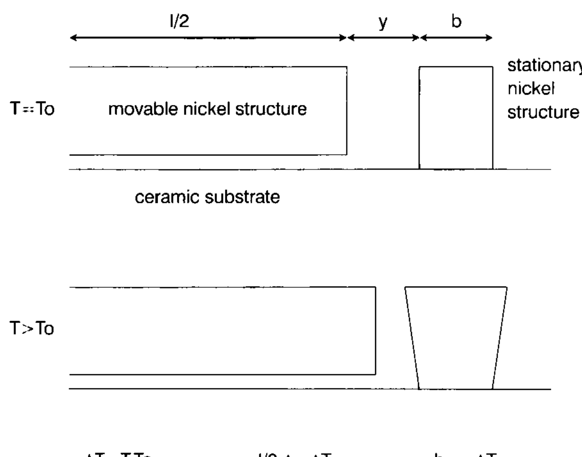


Figure 4-21.
Cross-section of a LIGA acceleration sensor at different temperatures T.

The movable parts of the sensor have no contact with the substrate and, therefore, the movable seismic mass and the substrate underneath expand freely. Owing to the different coefficients of thermal expansions, the gap y narrows by the amount of $l/2 \Delta\alpha \Delta T$ with increasing temperature. On the other hand, the stationary electrodes are fixed on the sensor substrate and cannot expand freely. The expansion at the top of the structure is larger than the expansion of the structural parts at the bottom of the substrate. On account of the complex variation of geometry the contribution to the change of the capacitor's gap width is described by an effective coefficient α_{eff}. The gap width is calculated to be

$$y(\Delta T) = y(0) - [l \cdot (\alpha_{Ni} - \alpha_C) + \alpha_{eff} b] \, \Delta T \, . \tag{4-1}$$

where

$y(\Delta T)$ = gap width in dependence of the temperature
$y(0)$ = gap width with initial conditions
l = width of the seismic mass
b = width of the stationary electrode
ΔT = deviation of temperature from initial condition.

From this equation the decrease in TCO for decreasing l and increasing $y(0)$ is obvious.

Temperature-Compensated Design

In order to reduce the TCO in a fundamental manner, one needs two capacities with opposing temperature coefficients which will be termed p and n in the following text. The sensor already discussed has a positive TCO. In order to obtain a capacity with negative TCO, the gap of the capacity has to increase with increasing temperature. The basic idea for a possible sensor design is shown in Figure 4-22.

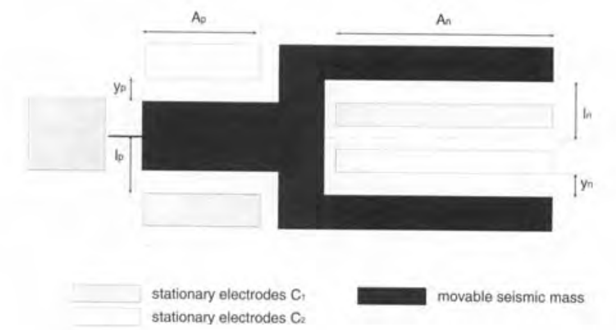


Figure 4-22.
Basic idea of a temperature-compensated sensor design of LIGA acceleration sensors.

The left part of the sensor has the same design as the acceleration sensor already discussed. For this part of the sensor the change in the gap width decreases with increasing temperature, as described in Equation (4-1). The temperature behavior of the right part of the sensor is characterized by the expansion of the open end of the seismic mass:

$$y_n(\Delta T) = y_n(0) + [l_n(\alpha_{\mathrm{Ni}} - \alpha_{\mathrm{C}}) - \alpha_{\mathrm{eff}} b]\,\Delta T\,. \tag{4-2}$$

The sensor is designed such that $l_n(\alpha_{\mathrm{Ni}} - \alpha_{\mathrm{C}})$ is dominant in comparison with $\alpha_{\mathrm{eff}} b$ and the gap increases with increasing temperature.

This results in a negative TCO of the sensor capacity.

Ideally for an absolutely temperature-compensated sensor element, the following relationships must be fulfilled:

$$\frac{\mathrm{d}}{\mathrm{d}\Delta T}(C_p + C_n) = \frac{\mathrm{d}}{\mathrm{d}\Delta T}(\varepsilon A_p/y_p + \varepsilon A_n/y_n) = 0\,. \tag{4-3}$$

where C_n = capacitance with negative TCO, C_p = capacitance with positive TCO, ε = dielectric constant, A = area of the capacitance. Equations (4-1)–(4-3) fix the ratio A_n/A_p as a function of the fixed capacitor parameters l_n, l_p, b and fixed material constants $\Delta\alpha$ and α_{eff}. The temperature compensation has to be independent of fabrication tolerances and of the acceleration to be measured. In both cases, the seismic mass is deflected away from its zero position. According to a detailed analysis, the temperature behavior is independent of the deflection of the seismic mass if Equation (4-4) holds. Calculations show that Equation (4-3) can only be satisfied for one temperature T_0. Nevertheless, the temperature dependence is minimized in a large temperature interval with T_0 in its center. Neglecting α_{eff} (Equation (4-2)), the relative temperature shift is calculated to be as in Equation (4-5):

$$l_p = l_n = l \text{ and } y_n = y_p = y \,, \tag{4-4}$$

$$1 - C(T)/C(T_0) = 1 - 1/[1 - (l\Delta\alpha\,\Delta T/y)^2] \,. \tag{4-5}$$

It is a parabolic function with a minimum at the temperature $T = T_0$. The temperature shift is minimized even for the temperature-compensated design, if the distance l is minimum and the materials differ minimally in their coefficients of thermal expansion (Table 4-1). The gap width y is fixed by the requirements of high sensitivity and high zero acceleration capacitance.

Table 4-1. Measured TCO for various widths of the seismic mass. The gap width is 4 µm.

l (µm)	TCO (10^{-3} g/K)
400	9.2
200	7.6
100	4.9
50	2.4

Experimental Results with the Temperature Compensated Design

Figure 4-23 shows a LIGA acceleration sensor with these design rules implemented. In addition to temperature compensation, the sensor offers the advantages of a small deviation from linearity, which is achieved by suspending it from the seismic mass by means of two parallel cantilevers and of a high zero acceleration capacitance produced by the interdigital structure.

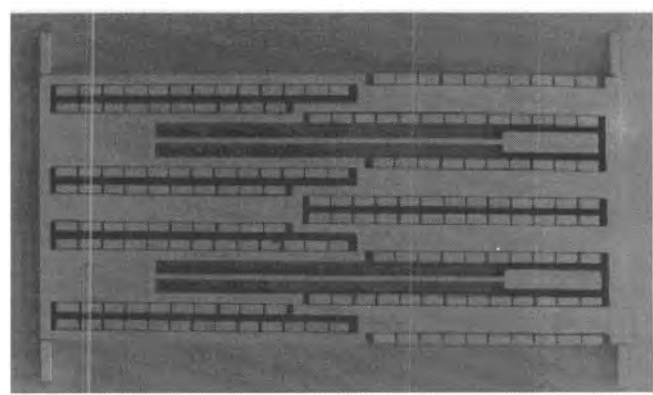


Figure 4-23. Temperature-compensated sensor design with parallel deflection and high zero acceleration capacitance.

The sensor's zero acceleration capacitance is about 4.5 pF. After a correction for a capacitance offset, which is generated by parasitic capacities, the difference between calculated and measured capacities in a temperature interval from -10 to 100 °C is only a few fF (see Figure 4-24).

For a low g sensor with a sensitivity of 20%/g the temperature shift results in a TCO of $1.02 \cdot 10^{-4}$ g/K. Compared with the non-temperature-compensated design the TCO is reduced by nearly two orders of magnitude.

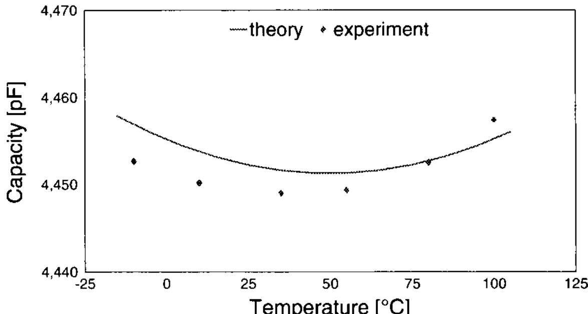


Figure 4-24. Comparison of calculated with measured data. Capacity of temperature-compensated sensor capacitor versus temperature.

4.5.3 The Concept of an "Intelligent" Sensor System

Introduction

In comparison with conventional systems, in combination with data processing elements the sensor elements are equipped with local intelligence. Microsystems allow one to realize more complex functions in the same or similar volumes. Though the cost of microelectronics has decreased dramatically, it is necessary to keep the costs of hard- and software as low as possible. The following example illustrates that high-precision sensors are needed as part of intelligent microsystems.

With increasing complexity of the correction algorithms, the system requires more and more resources. Figure 4-25 compares the hardware requirements for different stages of data improvement under the assumption that the data correction is done by a table of calibrated measurement data.

The correction for the deviation of linearity of a sensor with a resolution of $1.5 \cdot 10^{-5}$ needs a table consisting of a memory which has the capacity of 128 kbyte (16 bit address input, 16 bit data output). If a temperature compensation is additionally needed, a temperature sensor and a memory of 2 Mbyte are required. Perhaps these requirements can be reduced by using interpolation algorithms; anyway, the calibration of the system becomes complex and, therefore, expensive.

This illustrates that highly linear and thermally insensitive sensors are the basis of intelligent and fast microsystems.

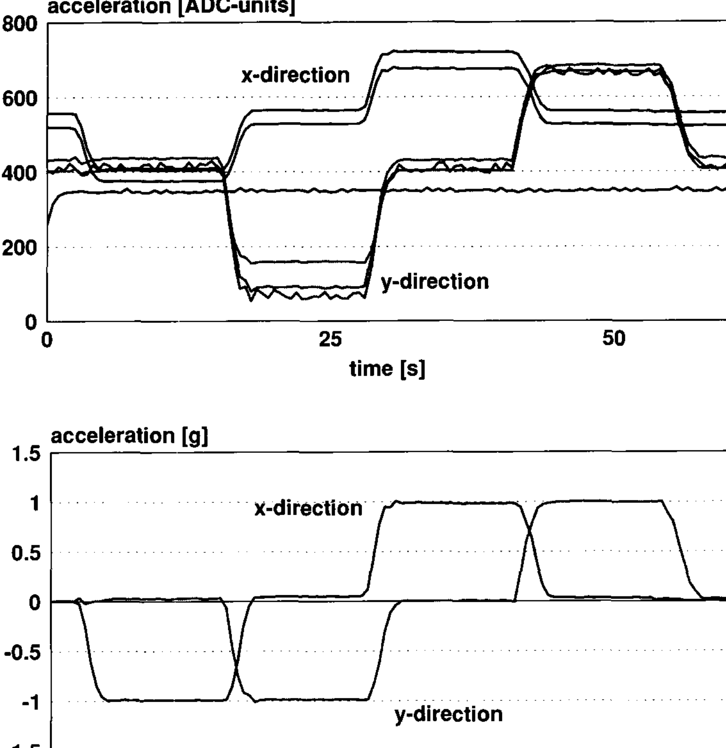


Figure 4-25. Comparison between different stages of data improvement.

Readout Techniques for Capacitive Sensors

The sensor characteristics are affected by the sensor readout technique. Because voltages applied to the electrodes of capacitive sensors result in electrostatic forces, there are two very different readout strategies. One is to avoid electrostatic forces completely, the other is to use these electrostatic forces in order to compensate mechanical forces completely. These two readout techniques are discussed in combination with the LIGA acceleration sensor in this section.

CSEM introduced an integrated readout circuit for capacitive sensors [27], which has been transferred to Bosch for the fabrication of an integrated circuit CC310 [9]. By avoiding charges on the sensors' electrodes the readout technique does not influence the deflection of the seismic mass by electrostatic forces. Additionally, if one assumes an ideal sensor without parasitic capacitance, the output signal is proportional to the parallel deflection of the seismic mass:

$$U_{\mathrm{L}} = \frac{C_1 - C_2}{C_1 + C_2} \cdot U_{\mathrm{m}} = \frac{\Delta y(a)}{y_0} \cdot U_{\mathrm{m}} \tag{4-6}$$

where

U_L = output voltage
U_m = modulating voltage
C_i = sensors capacitances
Δy = deflection of the mass
y_0 = gap width of the sensor
a = acceleration .

In reality there are parasitic capacitances parallel to the sensor capacitors, which are the main reasons for non-linearity. The relative deviation from linearity is calculated to be

$$Err_{\mathrm{lin}} = \frac{C_{\mathrm{par}}}{C_0 + C_{\mathrm{par}}} \cdot E^2_{\mathrm{CC310}}\, a^2 \tag{4-7}$$

where

Err_{lin} = relative deviation from linearity
C_0 = zero acceleration capacitance of the sensor element
C_{par} = parasitic capacitance prallel to C_0
E_{CC310} = relative sensitivity of the integrated circuit CC310 .

The deviation from linearity increases with increasing parasitic capacitance. For a parasitic capacitance of 0.5 pF and a zero acceleration capacitance of 5 pF, the relative deviation from linearity is calculated to be 0.36%.

The LIGA sensor element has been combined with a readout circuit working on this readout technique in hybrid set-up. Table 4-2 summarizes the experimental results for a combination of the LIGA sensor and the CC310 circuit.

Table 4-2. Characteristic sensor data for a combination of a 1 *g* LIGA acceleration sensor and the integrated circuit CC310.

Sensitivity	570 mV/g
Resolution	$2 \cdot 10^{-3}$ g
3 dB-frequency	150 Hz
Thermal zero shift	$1 \cdot 10^{-4}$ g/K
Thermal shift of sensitivity	$2.3 \cdot 10^{-4}$/K

A representative of the other readout technique which uses electrostatic forces in a feedback loop has been presented by KfK in combination with the LIGA sensor element [18]. It compensates the mechanical forces arising at the seismic mass, so that the mass remains in its zero acceleration position. Therefore, the output signal is not the deflection signal but the voltage used to generate the required electrostatic forces. The voltage is given by

$$\Delta U(a) = \frac{M y_0^2}{2 \varepsilon_0 A U_0} \cdot a \tag{4-8}$$

where

ΔU = output voltage
M = seismic mass
A = area of the capacitor
U_0 = constant voltage of the counter electrodes
ε_0 = dielectric constant .

The ideal linear relationship between voltage ΔU and the acceleration is disturbed by the parasitic capacitance. In contrast to the readout technique presented by CSEM, the differences in the parasitic capacitances ΔC_{par} determine the relative deviation from linearity:

$$Err_{lin}(\Delta C_{par}) = \frac{\dfrac{\varepsilon_0 A - \sqrt{\varepsilon_0^2 A^2 + 4\Delta C_{par}^2 y_0^2}}{2\Delta C_{par}}}{y_0^2 + \left(\dfrac{\varepsilon_0 A - \sqrt{\varepsilon_0^2 A^2 + 4\Delta C_{par}^2 y_0^2}}{2\Delta C_{par}}\right)^2} \cdot \frac{\Delta U}{U_0} \tag{4-9}$$

The relative deviation from linearity increases with increasing difference of the parasitic capacitances. For a difference in the parasitic capacitance of 0.05 pF and a zero acceleration capacitance of 5 pF, the relative deviation of linearity is calculated to be 0.9%.

Table 4-3 summarizes the characteristic sensor data for the feedback readout technique for a 1 *g* sensor.

Table 4-3. Characteristic sensor data for the feedback readout technique for a 1 *g* sensor.

Sensitvity	2.3V/g
Resolution (100 Hz bandwidth)	10 μg
3 dB frequency	410 Hz
Thermal zero shift	$1.2 \cdot 10^{-4}$ g/K
Thermal shift of sensitivity	$2.5 \cdot 10^{-4}$/K

A Multidimensional Acceleration Sensor System

By means of data processing which may be integrated into a microsystem together with one or more sensor elements, a further improvement of the acquired data may be achieved. Figure 4-26 shows a photograph of a two-dimensional sensor array that is the basis for such an intelligent microsystem.

The sensor array consists of two sensor groups with three acceleration sensors. They are fabricated in one process sequence and are therefore ideally aligned. This sensor array is the basis for the intelligent sensor system. A block diagram of the sensor system is shown in Figure 4-27.

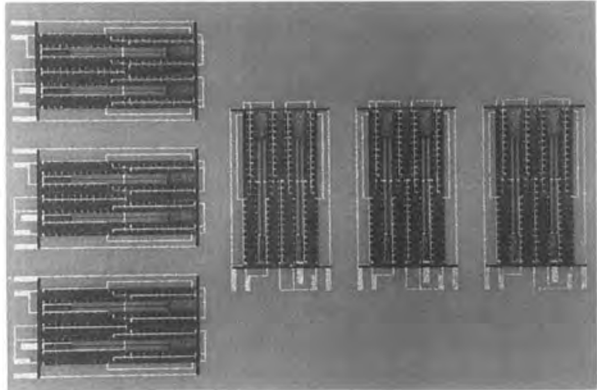


Figure 4-26. Photograph of a two-dimensional sensor array.

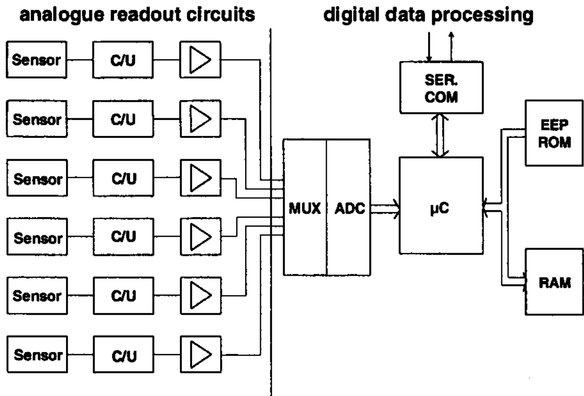


Figure 4-27. Block diagram of the two-dimensional sensor system for acceleration measurement.

The LIGA sensor elements and the integrated circuit have already been described in previous sections. The analog output signal (2.5 ± 0.5 V) is adapted to the full input range of the analog to digital converter (ADC) (2.5 ± 2 V) by an amplifier, in order to use the full resolution of the ADC. After multiplexing (MUX) and digitizing the analog data into digital values, the microcontroller (µC) is used to process the sensor raw data. This is done by a program stored in resident memory (EEPROM). The processed acceleration data are put into random access memory (RAM). The serial communication port together with the local intelligence offers the possibility of transferring as much information as needed and as little as possible to the host system.

In the first stage of data processing, the sensor raw data may be pre-processed in order to obtain more reliable and precise acceleration data than from the sensor alone. In order to obtain information about the sensor validity, different sensor tests are conceivable. A plausibility and a sensor self-test are discussed.

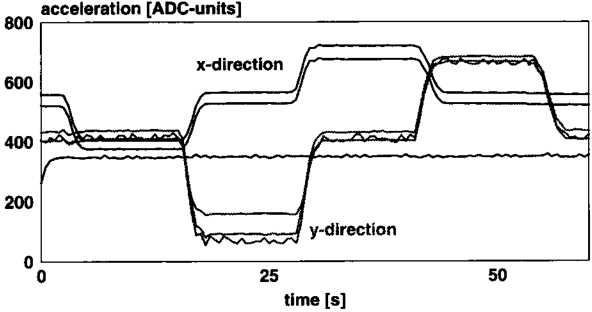
(a) Sensor Self-Test

The acceleration sensor is a spring mass system. By monitoring the mass deflection as a function of the applied forces, information about the spring mass system is calculable.

Readout circuits using electrostatic forces are well suited for this kind of self-test, because they offer the possibility of monitoring and stimulating the seismic mass by opening up the feedback loop. In this case observation of the dynamic reaction on activating pulses is possible and the damping characteristics and the time constant of the sensor element can be determined. Unfortunately, several integrated circuits are sensitive to DC voltages at the input or they are even destroyed by these voltages.

(b) Plausibility Tests

Local intelligence in combination with the redundancy of the system may be used for a plausibility test that allows one to obtain information about the validity of the sensor signals. Figure 4-28 gives an example of this kind of sensor self-test.



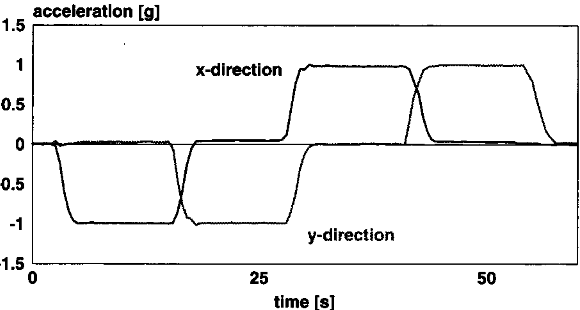


Figure 4-28.
(a) Sensor raw data of three sensors measuring in two dimensions; one sensor is not operating and another sensor signal shows large noise effects.
(b) Corrected acceleration signal after data processing.

The sensor raw data are compared for each direction. If a sensor signal differs from the average value by more than a permitted limit the sensor is not used for the determination of the acceleration value. Because of the redundancy, the system is able to operate with only two valid sensor signals.

The accuracy of the system is improved by the analysis of several sensors in a sensor array. The LIGA process allows one to fabricate sensor arrays with identical sensors and also sensor arrays with graded sensitivities.

If identical sensors with a standard deviation σ_m are used in the sensor array, accuracy increases with the number of sensors of the array. Evidently, the precision of the sensor system is affected by the precision of its components, which is represented by their standard deviations σ_m.

Making use of a sensor array consisting of sensor elements with different sensitivities, a wide acceleration range can be measured with high resolution. Figure 4-29 compares the resolution of a $4g$ sensor and a sensor array consisting of $1g$, $2g$ and $4g$ sensors under the assumption that the full-scale output is divided into $2^{10} = 1024$ parts. In the range up to $1g$ the resolution of the sensor array is four times higher than the resolution of the $4g$ sensor. The µC has to select the sensor with the highest resolution depending on the applied acceleration.

Figure 4-29. Comparison of the resolution of $4g$ sensor and a sensor array consisting of $1g$, $2g$ and $4g$ sensors.

Different applications need different data information. Therefore in a second processing stage the data are reduced to the information needed for the individual application.

The accuracy of the sensor data is dependent on the system noise, which is proportional to the square root of the system band width. In order to reduce noise, a low-pass filter algorithm with an application-specific cut-off frequency may be used.

For a flexible reaction in various measurement problems, a library of functions and procedures has been established. One application example is a control unit for machines. The acceleration sensor is used to monitor the vibrations of a machine's body by analysing the signal by a fast Fourier transform algorithm. The calculated spectrum is compared with the specified spectrum stored in the system memory. Deviations are detected and compared with their permitted limits. In the case of major differences, the sensor system sends a message to the host computer via the serial output.

Further applications in the fields of automotive applications, robotics and control engineering are under investigation. Especially for applications with field bus systems the illustrated reduction of information is needed to avoid high data rates and long transfer cycles. These are essential requirements for security of relevant data transfer.

4.6 Micropump Manufactured by Thermoplastic Molding

Microfluidic handling modules are needed in microanalysis systems to transport samples, chemicals, and reference media. These modules must consist of micropumps and microvalves, which can be controlled individually, as a function of their applications. These modules should be easy and inexpensive to make. Sufficient differential pressures and flow rates must be attainable.

A micropump was developed at the Karlsruhe Nuclear Research Center some time ago by combination of deep-etch X-ray lithography with electroplating and membrane techniques. It was driven by an external actuator [39]. Now a similar micropump has been made with a combination of thermoplastic molding and membrane techniques [13]. The new micropump includes an integrated thermopneumatic actuator and lends itself well to installation in microfluidic handling modules.

4.6.1 Micropump Design

The lateral dimensions of the micropump are 7 × 10 mm and the diameter and the depth of the pump chamber are 4 mm and 100 μm, respectively. The pump case consists of two thermoplastic parts with a single polyimide membrane mounted in between (see Figure 4-30). The fixed parts of the pump, such as the actuator chamber, pump chamber, flow channels, and valve seats, are patterned into the two parts of the pump case. The movable parts of the pump, such as the pump membrane and valve membranes, are integrated in a polyimide membrane 1 μm thick. A 250 nm thin copper wire is attached on top of the membrane. It is used for electric heating of the air in an actuator chamber located on top of the membrane and in the pump chamber. The pressure rise generated by heating deflects the membrane of the inlet valve against its seat and closes this valve, while the membrane of the outlet valve is deflected off its corresponding valve seat. The outlet valve opens and the air is forced out.

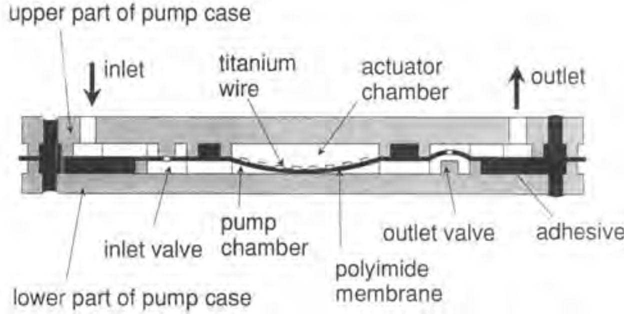
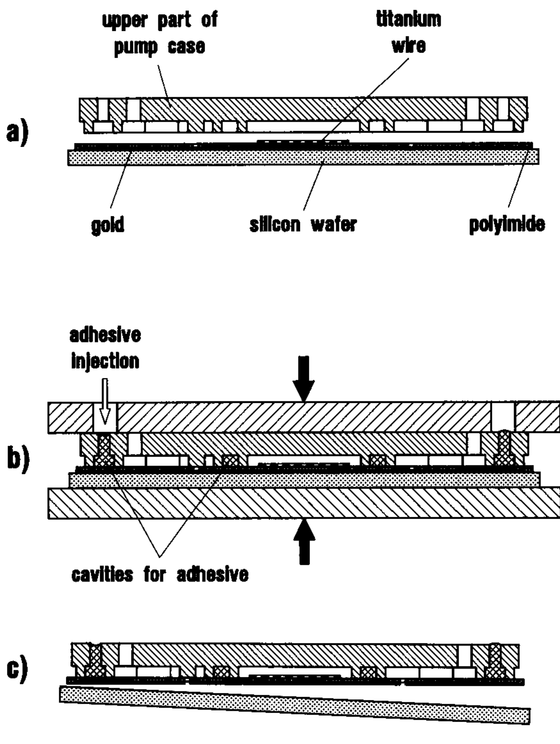


Figure 4-30.
Schematic view of a micropump consisting of two molded parts from PSU and a polyimide membrane mounted in between.

4.6.2 Manufacturing Process

In the first step, the two parts of the pump case and the polyimide membrane carrying the titanium heater were manufactured separately. In the second step, the completed and tested components were bonded by an adhesive bonding technique.

The cases of 12 pumps were molded in parallel. Cases of PVDF and PSU were produced by vacuum hot embossing and injection molding, respectively. Their high chemical resistance and temperature stability make these materials very attractive in chemical and medical applications. The corresponding mold insert was milled from a brass substrate with a cutter, 300 μm in diameter. As the critical dimensions of the micropump are not less than 100 μm, direct milling of the insert turned out to be sufficient [7]. If components of much smaller dimensions are required, the mold inserts can be made by the LIGA process.



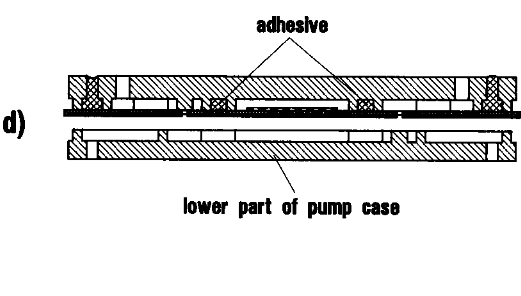


Figure 4-31.
Fabrication process of the micropump: (a) the molded upper part of the pump case is aligned with the structures on the polyimide layer; (b) the upper part of the pump case and the substrate are pressed together and adhesive is injected into the cavities; (c) the pump case and membrane are separated from the substrate by means of a low-adhesive gold layer; (d) the lower part of the pump case is mounted.

The polyimide membrane was manufactured on a silicon wafer (see Figure 4-31 a). A low-adhesive gold layer approximately 30 nm thick was sputtered on to a silicon wafer, covering the whole wafer except for a strip of 3 mm around the edge. A polymide layer 1 μm thick was spin-coated on to this substrate. Then 24 orifices, 100 μm in diameter, were patterned photolithographically into the polyimide. For making the copper heaters, a copper layer 250 nm thick was sputtered on to the polyimide and patterned by photolithography and etching with nitric acid.

The two parts of the pump case and the polyimide membrane were joined by adhesive bonding. Since the thickness of the adhesive layers is crucial to the function of the valves, a PSU part consisting of the upper parts of the 12 pump cases was aligned with the membrane and pressed on to the substrate for direct contact between structures of the PSU part and the membrane (see Figure 4-31 b). This direct contact ensures exact bonding conditions. The adhesive was injected into cavities in the PSU around the pump structures. It turned out that more than one pump could be filled with adhesive in a single injection. After the adhesive had been cured at 150 °C, the polyimide was cut around the contour of the twelve upper parts of the interconnected pump cases. While cooling, these parts were separated from the wafer together with the polyimide membrane (see Figure 4-31 c). To reduce the stress of the pump membrane still further, bonding was carried out at higher temperature: PSU structures and the polyimide structures fit together at 100 °C. As a consequence of the much higher coefficients of thermal expansion of PSU and polyimide ($55 \cdot 10^{-6}$/ °C) compared with that of the silicon substrate ($2-3 \cdot 10^{-6}$/ °C), the dimensions of the PSU structures are smaller than the structures processed on the silicon wafer under normal conditions.

The upper parts of the pump cases were bonded to the lower parts in the same way (see Figure 4-31 d). Individual micropumps were separated by sawing and contacted electrically. Figure 4-32 shows a photograph of a micropump with electrical contacts. The heater spiral can be seen in the center. The pump chamber underneath the spiral is connected to the microvalves on the sides.

4.6.3 Measurements

For realistic heat transfer conditions in a fluid-handling system, a micropump was bonded on to a silicon substrate. For measurements of the pressure and the air flow generated by the pump, unfiltered air delivered by the pump was used to force water out of a reservoir into a vertically mounted capillary. The height of the water column was measured as a function of time and the flow rate and pressure generated were derived from these data. A pulse generator delivering pulses (2 W, 1.7 ms, 30 Hz) was connected to the copper wire of the pump. The copper wire was 100 μm wide, and the pumping device in the test had an electric resistivity of 160 Ω. Figure 4-33 shows the air flow as a function of the differential pressure achieved. A maximum flow rate of 220 μl/min and a maximum pressure rise of 130 hPa were attained.

For a microfluidic handling module, several micropumps must be combined on a chip with active valves, flow sensors, and the electronics needed. It has been shown that LIGA structures can be molded [7] on top of electronic circuits [42]. Therefore, a microfluidic handling module with integrated electronics seems to be feasible.

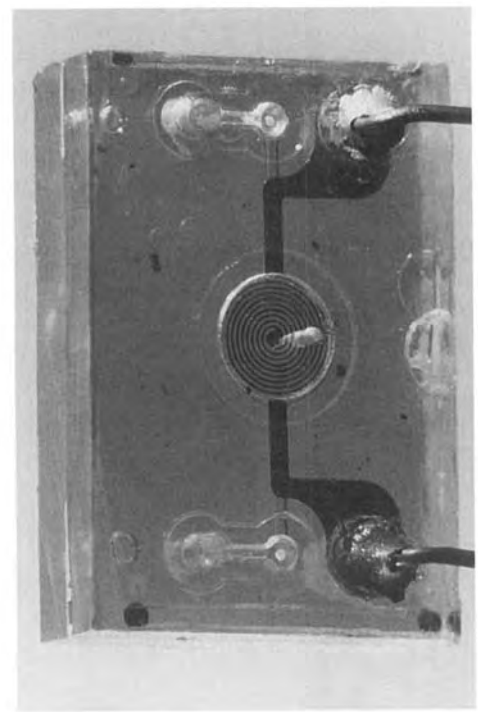


Figure 4-32.
Photograph of a micropump with electrical (in front) and fluidic (at the rear) contacts.

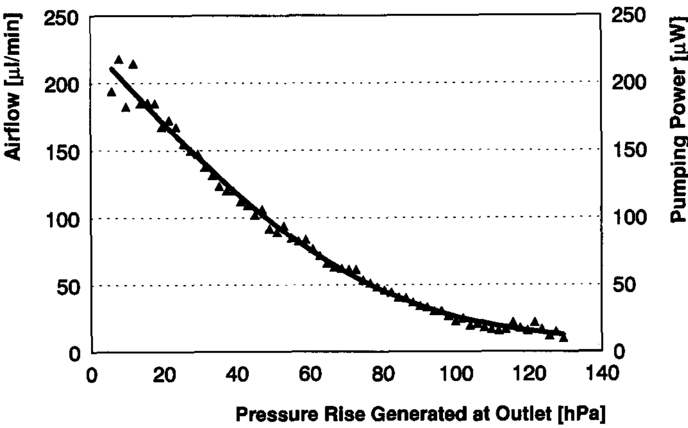


Figure 4-33. Air flow as a function of the differential pressure.

4.7 Optochemical Sensor System for Detection of Toxic Substances in Gases and Liquids

4.7.1 Introduction

There is growing interest in the replacement of conventional analytical procedures by chemical sensors. To fulfill various analytical tasks, the development and integration of

- chemical sensors (specific detection of substances)
- transducers (signal transduction)
- actuators (eg, pumps for sampling)
- processors (control, data acquisition, and evaluation)

to complete microanalytical systems would be of particular significance in many applications. Such devices allowing continuous, independent, and automatic measurements with high precision and low maintenance, would cover many analytical tasks, eg, control of various chemical processes and monitoring of hazardous parameters in environmental analysis or in medical applications, eg, diagnosis and therapy. In this context, microtechnological manufacturing techniques offer significant advantages with respect to the cost and reliability of such systems.

The key element of such systems is a specific detector for chemical substances; such chemical sensors can be fabricated by planar technologies such as are used in microelectronics. In particular, optochemical sensors can also be fabricated by modifying such processes. In this way, the concept of microtechnological mass production of sensors can be extended to include sensors based on optochemical transduction. Furthermore, continuous progress in optical (light sources, microspectrometers) and electronic instrumentation (microprocessors, microcontrollers) allows the design of small and compact microanalysers with integrated data acquisition and evaluation.

4.7.2 System Components

Optochemical Sensors

The detection of chemical components by optical sensors has particular advantages with respect to other (eg, potentiometric) sensors, in that no reference element is needed, influences of electrical interferences are minimal, and variations of the baseline can be eliminated when using multi-wavelength calibration. In principle, optochemical sensors are based on immobilized receptor molecules or dyes which show spectral changes when interacting with the analytes of interest.

For the very sensitive and selective detection of metal or heavy metal ions, chelating agents can be used. Sensing of gaseous pollutants can be achieved by measuring the changes in optical parameters that occur when the gas molecules are coordinated to immobilized dyes. A promising and chemically inert matrix for immobilization of a variety of receptor dyes is thin films of porous glass made by the sol–gel process. These films can be deposited with variable

lateral geometry on transparent substrates. Using processes similar to those used in microelectronics, the fabrication of microsensors with different functionality on the same substrate is possible.

(a) Heavy Metal Sensor

Porphyrin derivatives are a promising class of chelating agents, being well established as highly sensitive indicators for the spectrophotometric determination of metal ions. The high complexation constants of porphyrins allow very sensitive detection. Further, this class of dyes have extremely high extinction coefficients which can be used for very sensitive optical detection. In addition, the different metal complexes of porphyrins show very specific absorption bands. This class of dyes make multi-component analysis feasible, with sensors which can detect several metal ions by their specific reaction with one receptor molecule. For example, the simultaneous and accurate determination of Hg, Cd, and Pb ions is possible after deconvolution of the complex spectra using standard chemometric procedures, eg, partial least squares regression (PLS).

Figure 4-34 shows the detection principle for the detection of metal ions together with some representative spectra of metal complexes of a porphyrin derivative (5, 10, 15, 20-tetra(4-N-methylpyridyl)porphyrin, TMPyP). A sensor for cadmium and mercury was developed by electrostatic immobilization of this porphyrin on a thin film of Nafion, a strongly acidic cation exchanger. Within a measuring time of 10 min, detection limits of 5 μg/l for Cd(II) and 30 μg/l for Hg(II) were observed. However, the use of an ion exchanger as an immobilization matrix resulted in a sensor that showed saturation behavior and required regeneration after each measuring cycle [33]. Porphyrins have now been stably immobilized in thin films of porous glass and remain active with respect to their reaction with the analyte. The use of a glass matrix made by the sol-gel process allows a variable layout of the sensor and multiple sensor fabrication on the same substrate.

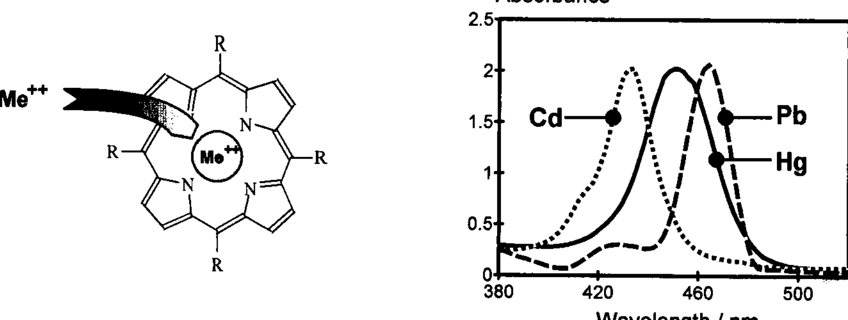


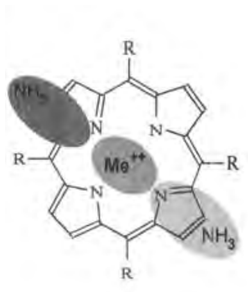
Figure 4-34. Porphyrins as receptor molecules for heavy metal ions. The complexation of porphyrins with metal or heavy metal ions leads to specific changes of the absorption spectra; different metal ions show different absorption band which can be used for specific *and* multi-component analyses. By chemometric data treatment of the spectra, different metal ions can be identified and quantitatively determined.

(b) Sensors for Toxic Gases

For optochemical gas detection, the use of metalloporphyrins as receptor molecules which coordinate the gaseous compound leading to a specific spectral change is a promising detection principle [34]. The sensors developed for the determination of NH_3 and SO_2 are based on various metalloporphyrins which are immobilized in thin films (≤ 1 µm) of porous glass prepared by the sol–gel technique. As shown in Figure 4-35, the gases can be detected by the changes in the absorbance spectra that occur when the gas molecules are coordinated to the central metal ions of the immobilized porphyrin molecules. Specific spectral changes are obtained for different gas components when coordinated to different metalloporphyrins.

The detection limits are in the range 5–10 ppm. The sensors respond reversibly within minutes and show stable responses over several months.

In principle, the specific spectral changes of the sensors allow multi-component detection by chemometric analysis of the spectra.



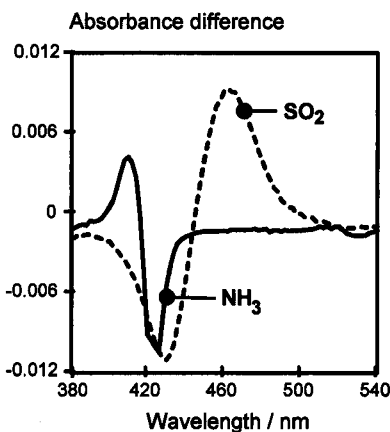


Figure 4-35. Immobilized metalloporphyrins as detectors for toxic gases. Coordination of gas molecules to the central atom of the dye leads to changes in the porphyrin spectra. The difference spectra of some metalloporphyrins in the presence of ammonia and sulfur dioxide are shown. The reactions are reversible with time constants of a few minutes; detection limits, 5–10 ppm.

LIGA Microspectrometer

To use the particular advantages of multi-component analysis with porphyrin-based optochemical sensors, the specific spectral changes of the sensors have to be measured and analyzed in the visible range between 400 and 600 nm. The spectra are recorded with an optical LIGA component, a microspectrometer, which covers the desired spectral range. Furthermore, the optical resolution of 4 nm allows sufficiently fine scanning of the spectra. Since both the heavy metal sensors and the gas sensors respond within minutes, a large number of spectra can be accumulated and averaged, so that very fine differences can be detected. The measuring beam is coupled into the flow cell in which the sensors are integrated by means of fiber optic. This modular constructing permits rapid replacement of the flow cell and the sensors in case of defects or in order to switch to the detection of other analytes.

4.7.3 Analytical Microsystem

The key component of the microsystem is an optochemical sensor either for toxic gases or for heavy metal ions. As an integrated system, it will be able to perform on-line sampling with the help of actuator components (pumps, valves), to carry out cycles of calibration and measurement, and to evaluate and display the data acquired by means of suitable algorithms (spectral deconvolution and evaluation with pattern recognition procedures). The construction is illustrated in Figure 4-36: the optochemical sensors are integrated in a flow cell which is illuminated with white light. The modular construction (via light-guide connectors) permits the use of various sensor elements according to the particular application, and allows the rapid replacement of defective components. The measuring beam is then transmitted to the LIGA microspectrometer, where the spectral changes arising from contact with the analytes are recorded, processed and output via a serial port. The microprocessor also controls the fluidic components (valves, pumps), so that variable, programmable measurement and calibration cycles can be carried out. The evaluation of the data (spectral deconvolution, calculation of concentration) and presentation of the results are currently performed by an external PC. For specific applications, the complete data evaluation can be performed by an additional processor module.

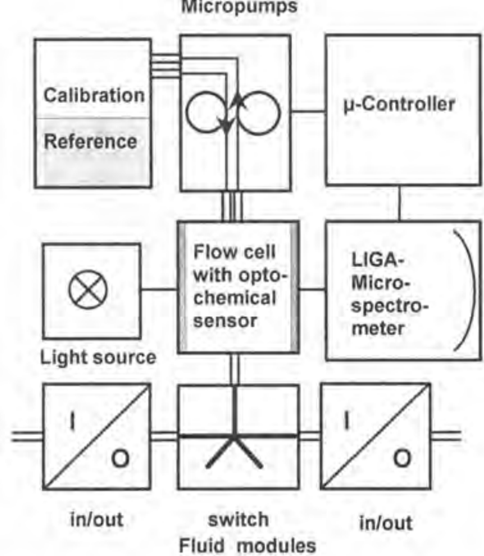


Figure 4-36. Schematic diagram of the optical microsystem. The key element is an optochemical sensor which is integrated in a flow cell. The measuring light is guided by fiber optics. The spectral changes of the sensor are detected by the LIGA microspectrometer. Sampling, calibration, and measurement cycles are performed with pumps and valves controlled by a processor which also controls the spectrometer and evaluates and displays the data.

4.8 Electrochemical Microanalytical System (ELMAS) for the Ionometry of Liquid Media

4.8.1 Introduction

The continuously rising demand for more chemical information in a broad analytical field from chemical process control to environmental protection, medical diagnostics, and health care has resulted in the development of numerous chemical microsensor devices [35, 37]. In many cases these sensors can be manufactured in small geometric dimensions – compared with the widespread conventioal macrosensors – and at low cost because of the use of mass production technologies, which are available from the microelectronics industry.

However, in spite of the evident and very promising advantages, chemical microsensors are not dominant on the market and they have had no practical significance in widely used analytical techniques so far. Problems concerning the reliability and long-term stability and, in some cases, bad selectivity of the sensor signals are the most serious drawbacks.

Therefore, attempts have been started in some advanced laboratories for integrating sensors, actuators, and microelectronic components into complete miniaturized chemical analytical systems [21, 38, 48] based on microsystem technologies. The basic idea involves self-testing and self-calibration of the system by recording signals alternatively from sample solutions and from calibration solutions, in a similar way as is done in the flow injection analysis technique [25]. Drift phenomena of the sensors can be effectively compensated and the reliability can be improved by using feedback mechanisms (Figure 4-37).

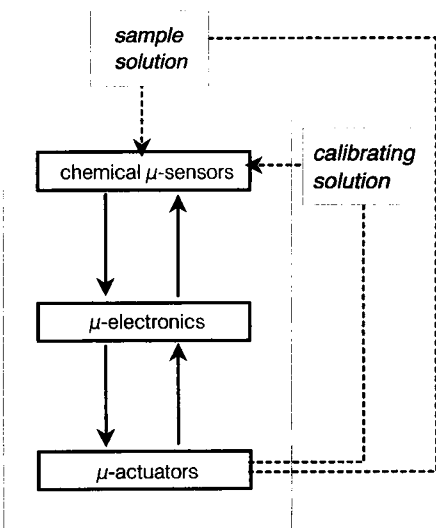


Figure 4-37.
Flow of signals (→) and materials (-->) in a chemical microanalytical system.

Conditioning of the sample to take advantage of the optimum sensor performance and cleaning and preconditioning of the sensor itself are further challenges which can be met by sensor/actuator systems. The creation of so-called "total micro-analytical systems" [50] is en-

visaged as a final goal of all these development efforts. It means that the essential functions of modern automated laboratory analyzers have to be scaled down to small-sized devices.

In such systems, electrochemical sensors are favoured. They convert directly the chemical signal (kind and concentration of electrochemically active species, eg, ions in the analyte) into an electronic signal. In the case of ion-selective electrodes (ISEs) and ion-sensitive field effect transistors (ISFETs) [6], an electrochemical potential established at the solution/sensor interface is controlled by the ion concentration of the solution. A thin ion-sensitive membrane (ISM), which is placed at this interface, has a determining influence on the basic characteristics of these potentiometric sensors [29]. Any distortion of this sensitive interface will disturb the sensor function, thus demanding self-control and self-calibration.

4.8.2 The ELMAS Concept

Using technologies inherent in or compatible with the LIGA process, the various microsystem components are integrated into an electrochemical analytical device (Figure 4-38).

ISFETs and miniaturized ISEs are used as sensor components, responsive to ion concentration changes of liquid media. The reference elements, which are necessary in a potentiometric electrochemical measurement set-up, and which must not change their signal during measuring cycles, are identical ISFETs or ISEs (sensors) in this concept. They are positioned at some distance from the "measuring" sensor, which is exposed to the analyte sample for measuring, and to a solution of known composition during calibration. In these alternating calibration

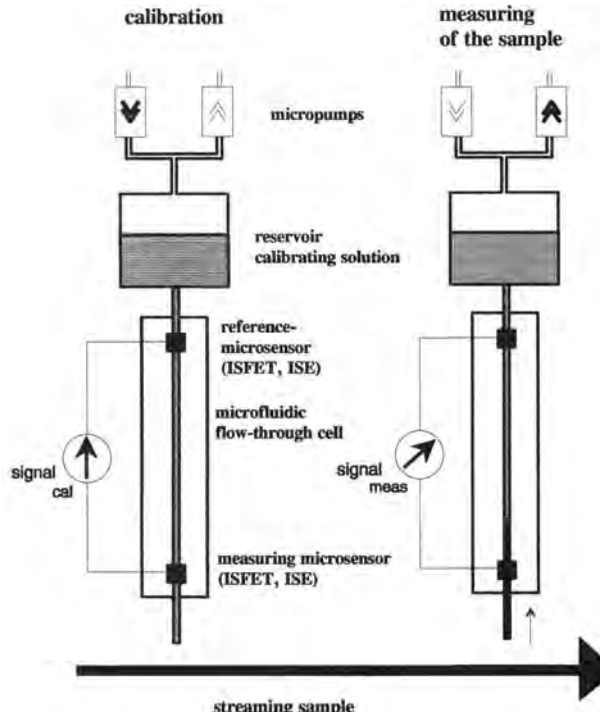


Figure 4-38.
The ELMAS concept, schematically.

and measuring cycles, the difference signal is correlated with the unknown ion concentration of the analyte. Any change of the sensor baseline, eg, resulting from drift phenomena, can be compensated by this difference measuring mode. Application of LIGA-manufactured components offers special advantages for the design of microanalysis systems.

The electrochemical sensors have to be integrated into a microfluidic flow-through cell. In ELMAS it is constructed from a substrate plate containing microgrooves and covered by another plate with openings for the microsensors. The microsensors can be connected to the microchannels by adhesives or simply by clamping. Special attention has to be paid to minimizing the dispersion, ie, the mixing of the solutions at the calibration/measuring liquid interface should be as small as possible for rapid and sharp signal changes during operation and for lowering the liquid consumption.

In contrast to micropumps manufactured with silicon technology, these pumps are able to handle gaseous media. Thus solution transport is possible indirectly with the help of an intermediate air layer over the solution. In this way all liquids, which could contaminate and corrode the microparts of the pump, or which could contain foreign solid particles that could damage the sensitive microchemical parts, can be excluded, resulting in greater reliability and a longer lifetime of the pumps. Further, no problems will arise from air bubble formation, which is often observed in direct liquid pumping and avoiding which takes a lot of care.

All ELMAS components will be realized as standardized modular elements, which can be stacked to obtain a compact microsystem unit.

Modifications of the set-up shown in Figure 4-38 are possible by changing modules.

Multiple-ion sensing is possible by integrating different sensor elements. Two-point calibrations can be realized with an additional pump and calibration reservoir. Also redundancy of all functions can be ensured by adding further elements. This might be of special importance with respect to any device failures. Thus reliability of the results may be increased in addition to the lifetime of the microanalytical device.

The ELMAS functions are driven and controlled by a microprocessor-based electronic device. The whole construction, including the housing and the battery as an internal energy source, will also be miniaturized, leading to a small complete analytical unit for comfortable handling during application, eg, patient monitoring in critical medical situations or environmental or process control, requiring high mobility at moderate cost.

4.8.3 Potentiometric Microsensors

Basic potentiometric sensor chips have been developed for ELMAS. They contain dual ISFET structures on chips suitable for redundant measurements or for the analysis of two different ion species. The sensitive areas of each ISFET are 16 × 400 μm. The basic ISFETs have a Si_3N_4 pH-sensitive membrane, 70 nm thick; they are prepared in an industrial *n*-channel CMOS process. Further, a temperature-sensitive diode is integrated for temperature measurement and compensation.

The overall size of the chips is 5 × 8 mm, intentionally not driven to the lowest possible size for easy and highly reliable chip handling during encapsulation and chip bonding to the flow-through cell. The electrical connnections to the electronic measuring device can be established simply by pressing the chip into the plug of the conductor cable. Therefore, the size of the contact pads on the chip is matched to the raster of commercially available microconnectors

with only a 0.5 mm contact spacing. In this way a multi-channel connection can be easily realized without any conventional wire bonding process (Figure 4-39). Further, the connection is detachable, an advantage during any device development process.

For the investigation of dynamic sensor performance, a simple flow-through cell was constructed for easy packaging of the sensor chip. The chips are clipped between the microcell and a support sealed by an O-ring (Figure 4-39); this mounting is removable for easy chip inspection or exchange during chip development.

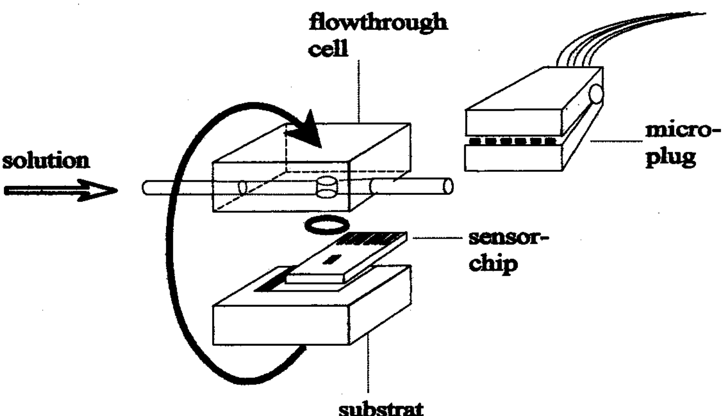


Figure 4-39. Schematic view of a clip mounted flow-through cell and a sensor flip chip which can be plugged into a microconnector.

The ISFET chips may be replaced by ion-selective electrode chips. For this purpose, chips of the same dimensions can be used alternatively. Most simply they are constructed from a metallic electrode covered by an ion-sensitive membrane.

Ion-sensitive membranes (ISMs) have to be integrated on to the ISFET or ISE chips for measuring specific ions selectively. In this way a great variability of ELMAS is possible, allowing adaption to a wide field of application.

For reasons of stability ISMs should be prepared from solid-state material. Also, fabrication with methods compatible with semiconductor technologies is desirable. The transfer of membrane materials well known in conventional macro ISEs to the chip level with the help of thin-film techniques proved to be suitable. Special attention was given to this point in the research activities at KfK.

High-vacuum evaporation of ISM material could be demonstrated as a method feasible for the preparation of iodide-sensitive AgI membranes [44]. Silver iodide layers, 1.0–2.0 μm thick, were evaporated on to semiconductor and ISFET chips. The problems of defects forming within the layers, leading to sensor performance failure, were overcome by an additional adhesive layer of chromium and by substrate heating during evaporation. Dense layers of AgI, polycrystalline and of large grain dimensions (up to 10 μm), were obtained, showing stable iodide sensitivity down to 10^{-6} M solutions and during operation for more than 3 months in electrolyte solution.

Magnetron sputtering as another thin-film fabrication method could be used succesfully for the preparation of thin glass ISMs. Copper ion-sensitive membranes were sputtered from

Cu-As-Se targets, the components of which are known from bulk macro copper ISEs [15]. Ultra-thin films of only 15–30 nm showed excellent Cu sensitivity in solutions containing down to 10^{-8} M Cu; the selectivity coefficient for most of the interfering ions was 10^{-4}–10^{-7}; only for iron was the interference of some significance.

Similarly sodium ISMs were obtained by sputtering of aluminosilicate glass, which is the ion-sensing material in conventional sodium glass electrodes [4]. Here, a cylindrical glass target geometry was used for sputtering, which avoids the effect of changing stoichiometry normally observed in repeated sputtering of oxygen-containing materials. ISFETs prepared with these layers, 150 nm thick, showed after thermal conditioning a sensitivity comparable to that of commercially available macro sodium ISEs and also the interference from pH changes and potassium was reasonable [22] (Figure 4-40), at least in the range of medical relevance.

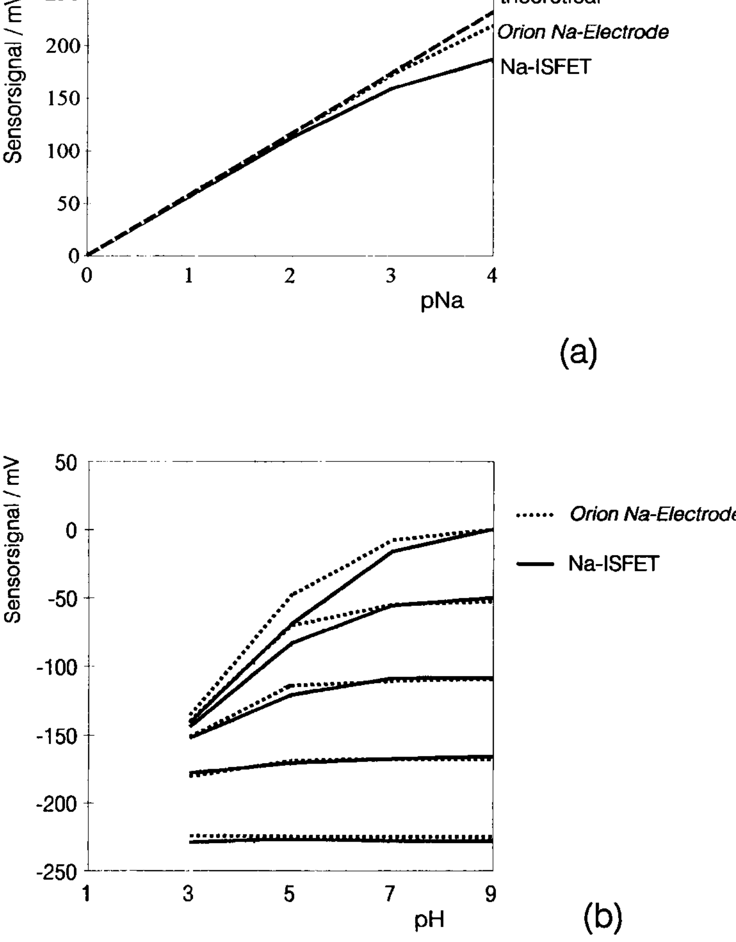


Figure 4-40. Response of a pNa-ISFET compared to a commercial available macro Na-ion selective electrode to pNa (a) and cross sensitivity to pH (b).

Solid-state ISMs exhibit a fast response to changes in analyte concentration. As an example, the pH-ISFET signal recording during a rapid change of the solution pH is given in Figure 4-41. Virtually no signal delay is observed with variation in pH; the curve exhibits distinct plateaux allowing fast measurements of alternating solutions.

Therefore, solid-state ISMs are well suited in dynamic measurement methods to be used in ELMAS. Further, it was found for the Cu-sensitive ISE that the detection limit and also the selectivity coefficient could be increased by changing from the static to the dynamic measurement mode under continuous-flow conditions [15]. This emphasizes once more the combination of sensors with actuators needed for microfluidic handling.

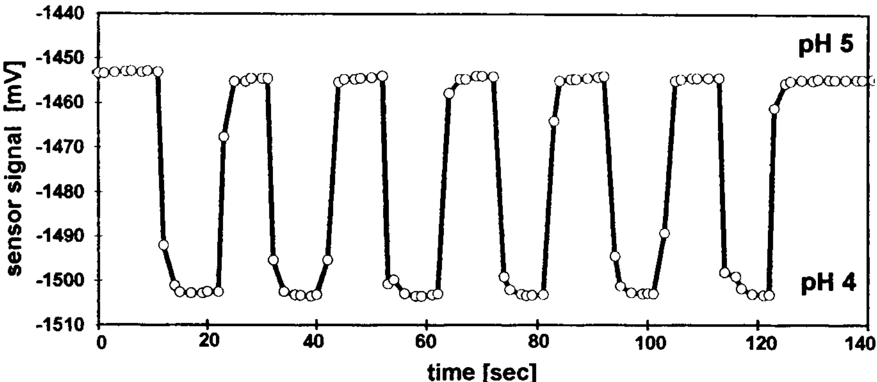


Figure 4-41. Response of a pH-ISFET during rapid change of pH buffer solutions.

4.8.4 ELMAS Functional Model

For purpose of ELMAS demonstration, microsensors of the ISFET type mounted in a flow cell as shown in Figure 4-39 and a micropump fabricated by thermoplastic molding were connected by silicon rubber tubes. The whole system, only driven by the small internal

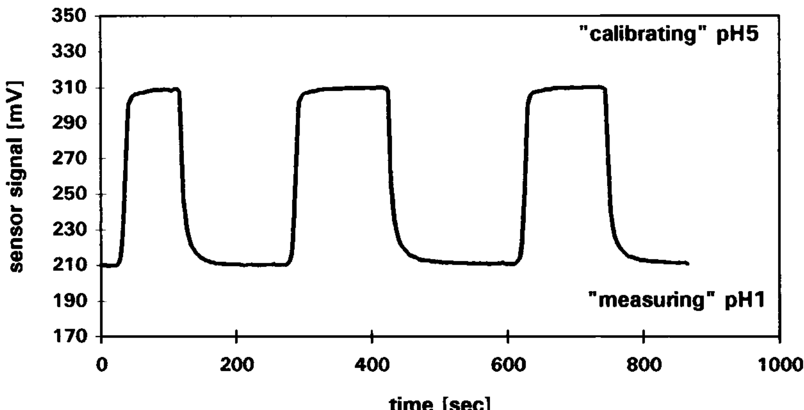


Figure 4-42. Example for alternating calibrations and measurements in an ELMAS functional model.

microfluidic actuator, reflects the ELMAS function of alternating calibration (eg, reference solution of pH 5) and measuring (eg, sample solution of pH 1) correctly (Figure 4-42). Further developments are directed primarly to the technological integration of the flow-through cell for showing the feasibility of the unique manufacture of all relevant ELMAS components and to integrated sensor/actuator signal processing and system-driving microelectronic devices.

4.9 Concepts for Miniaturized Analytical Microprobes: Integrated Optical NIR Evanescent Wave Sensor System for Chemical Analysis

The contamination of drainage and industrial waste waters with organic substances such as chlorinated hydrocarbons, aromatic substances, or mineral oils requires the developement of rapid, inexpensive and simple but nevertheless powerful chemical sensors that allow the continuous in situ monitoring of such pollutants. Opto-chemical sensors are among the most promising approaches to fulfil such demands. In this context at the Institut für Radiochemie of the Karlsruhe Nuclear Research Center (KfK), in recent years an evanescent field absorbance sensor (EFAS) built from a quartz glass fiber with polysiloxane cladding has been developed. In combination with a conventional near-infrared (NIR) spectrometer it allows the monitoring of nonpolar organic compounds in aqueous and gaseous media down to the low mg/l range. Its sensing principle is based on the enrichment of these substances in the hydrophobic polysiloxane membrane and the measurement of their evanescent wave NIR absorption spectra without spectral interferences from broad water OH absorption bands [16].

Within the framework of the PMT project at KfK, current research is now directed towards realizing a miniaturized sensor by using integrated optics technology and planar waveguide structures. The construction of planar evanescent wave absorbance sensors offers some advantages over the existing fiber-optic sensor:

- in contrast to a cylindrical fiber geometry, a planar substrate allows easy deposition of tailor-made polysiloxane superstrates, which act as a sensing layer;
- the planar structure provides much higher mechanical stability than a sensing fiber coiled up close to its minimum bend radius on a support;
- owing to the smaller waveguide dimensions, a higher number of reflections and thus an increased sensitivity per unit length and a higher fraction of light intensity in the evanescent field are obtained.

4.9.1 Integrated Optical Evanescent Wave Sensor

The basic set-up of the integrated optical evanescent field absorbance sensor (IO-EFAS) is shown in Figure 4-43. One part of the sensor is formed by a surface waveguide structure, produced by Ag^{+}/Na^{+} ion exchange in a BGG31 borosilicate glass substrate. Since the maximum refractive index (RI) of the waveguide is directly at the glass surface, and owing the smaller waveguide dimensions compared with an optical fiber, a higher fraction of the modal

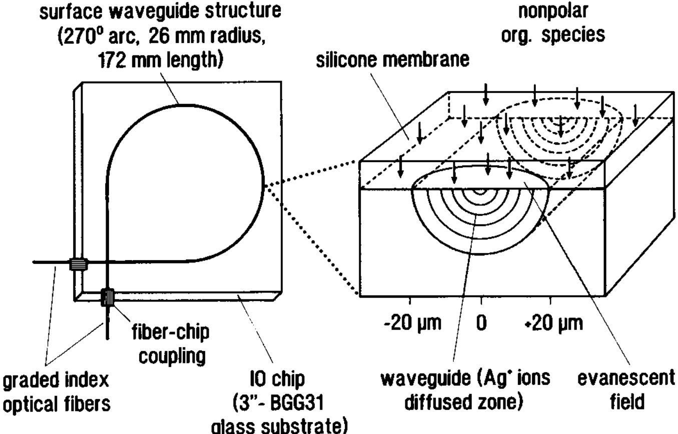


Figure 4-43. Set-up of an integrated optical evanescent wave absorption sensor for nonpolar organic species.

field of the light wave extends out of the surface. The sensing membrane that covers this waveguide structure is built from a suitable silicone polymer with lower refractive index, which acts as a hydrophobic matrix for the reversible enrichment of nonpolar organic contaminants from water or air. If light from the near-infrared range is coupled into the waveguide, the evanescent wave part of the light field penetrating into the silicone layer can interact with the extracted organic species and light intensity in the spectral region 1600–2000 nm will be absorbed at the characteristic frequencies of the corresponding CH overtone and combination band vibrations.

The waveguide structure in the BGG31 borosilicate glass substrate has been realized in cooperation with IOT (Integrierte Optik Technologie, Waghäusel), a company that mainly produces integrated optical components for telecommunication and sensor applications on a pre-production scale. The present surface waveguide structure was generated by the Ag^+/Na^+ ion-exchange process developed by IOT [41]. The procedure starts by writing the mask set for photolithography by an e-beam machine. The next steps of waveguide production in glass are carried out under clean-room conditions. After careful cleaning of the glass wafer, a thin titanium or chromium layer is evaporated on it. In the following photolithographic process (resist spinning, baking, exposure, developing, etching), the structure of the mask set for the later-formed strip waveguide is copied into the titanium layer. After removal of the resist, the wafer with the titanium mask is dipped into an $AgNO_3$ or $AgNO_3/NaNO_3$ melt. The thermal Ag^+/Na^+ ion-exchange process takes place at temperatures between 220 and 500 °C. The refractive index increase in the glass that is necessary to form a waveguide is caused by the higher polarizability of silver ions compared with sodium ions. After cooling to room temperature and removal of the titanium mask, the waveguide structure can be cut out from the wafer.

The present multi-mode surface waveguide structure design for the IO-EFA sensor (see Figure 4-43) consists of a 270° arc with 26 mm radius terminating in two straight lines that intersect in a 90° angle. The total length of the strip waveguide is 172 mm and the dimensions of the light-guiding zone are around 35 × 20 µm.

To increase the penetration depth of the evanescent wave and hence the sensitivity, it is essential that the RI of the superstrate approaches that of the waveguide as closely as possible. On the other hand, with a smaller RI difference between waveguiding zone and superstrate, the light losses increase. Hence one has to find a compromise between good sensitivity and bad signal-to-noise ratio.

At the moment the preferred sensing materials for nonpolar organic substances which are synthesized at the IRCH are polysiloxanes with methyl and phenyl side-chains, which allow one to change the refractive index of the superstrate in a fairly broad range from 1.41 (pure polydimethylsiloxane) to 1.55 (methylphenylpolysiloxane with 50% phenyl groups). Owing to this dependence of the silicone RI on the degree of phenylation, it is possible to approach the RI of the sensing membrane to the higher value of waveguiding zone, which has a graded index profile. The RI value in the waveguide zone of BGG31 glass can be adjusted in the range 1.459–1.519, depending on the ion-exchange conditions. The "tailor-made" polysiloxanes for the sensitive coating of the waveguide structure are synthesized from the corresponding dichlorosilanes by hydrolysis [51]. The reaction products are α,ω-dihydroxyalkyl(aryl)polysiloxanes, which are obtained as viscous liquids. After purification by rectification, they can be used as precursors for the polymers. After adding a curing catalyst, the α,ω-dihydroxypolyalkyl(aryl)siloxanes are dissolved in toluene and poured on to the IO-glass substrate fixed in a mold. Owing to their low surface tension they spread easily over the wafer. The solid silicone membrane is formed within 24 h of curing at room temperature.

4.9.2 Evaluation of the Sensor Performance

To optimize the sensor design and for testing the sensor response to organic pollutants, the IO-EFA sensors are adapted over a lens system to a conventional incandescent light source (100 W tungsten–halogen lamp) and a Czerny–Turner spectrograph with a 300 lines/mm planar grating. A 256-element InGaAs diode array that can be applied at wavelengths up to 1680 nm is used as a light-detection element. The measuring light is conducted from the tungsten–halogen lamp to the sensor and back to the spectrograph via 50 μm all-silica graded index fibers of 2 m length, which are glued to the waveguide structure by silicon fiber-waveguide couplers [41]. The instrumental set-up allows one to measure NIR absorbance spectra with the IO-EFA sensor over a 120 nm wavelength range.

In this way, trichloroethene (TCE) has been determined in both the aqueous and the gas phase. From the two NIR absorbance spectra of aqueous solutions of TCE shown in Figure 4-44 one can easily identify the CH first overtone vibration of this molecule at 1650 nm. The absorbance/concentration calibration values show positive deviations from linearity at higher concentrations similar to those obtained with the fiber-optic EFA sensor [16]. For measurements in aqueous solution it is necessary to solve problems of light losses that occur if the silicone membranes come into contact with water. They lead to a bad signal-to-noise ratio of the absorption signals. These losses are due to contraction of the polymer in contact with water, which leads to an RI increase and as a consequence to the coupling out of higher order modes. An improved mechanical stability of the rubber-like silicones seems to be achievable by a higher degree of cross-linking of the polymer chains.

For gas-phase measurements, where these problems are absent, the absorbance signal of the IO sensor with a waveguide length of only 172 mm is already higher than the signal of a fiber-

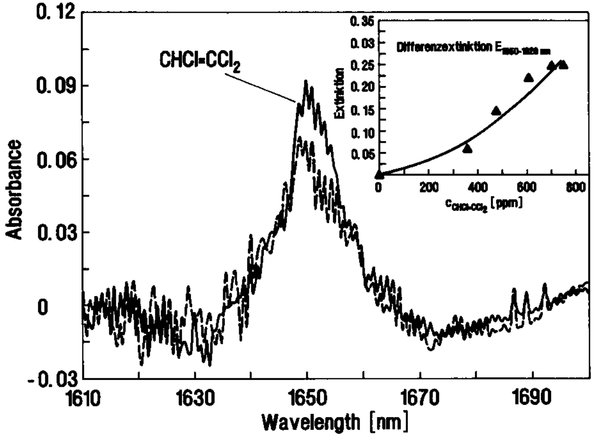


Figure 4-44. NIR spectra and calibration function of aqueous trichloroethene solutions measured with IO-EFA sensor (172 mm waveguide, silicone superstrate with an RI of 1.449) in combination with a conventional InGaAs diode-array spectrograph.

optic sensor with a length of 11 920 mm, thus showing the increased sensitivity of the IO component compared with the fiber-optic sensor. The increase in sensitivity per unit length of the waveguide is of the order of a factor of 60.

4.9.3 Spectroscopic Microsystem for Chemical Analysis

An optimized evanescent wave microsensor allows integration into an spectroscopic micro system for chemical analysis, by combining it with a miniaturized incandescent light source, an NIR microspectrometer, and a diode-array detector. The concept of this set-up is shown in Figure 4-45 [1]. A miniaturized incandescent light source, which provides a continuous emission spectrum in the near-infrared spectral range, could be formed by a thin silicon filament, which is heated to 1100 °C by restistance heating. Such a microlamp has been developed by Mastrangelo et al. [28]. In a first approach, already available miniaturized tungsten-halogen lamps, which are used in mini-pocket lamps, seem to be applicable.

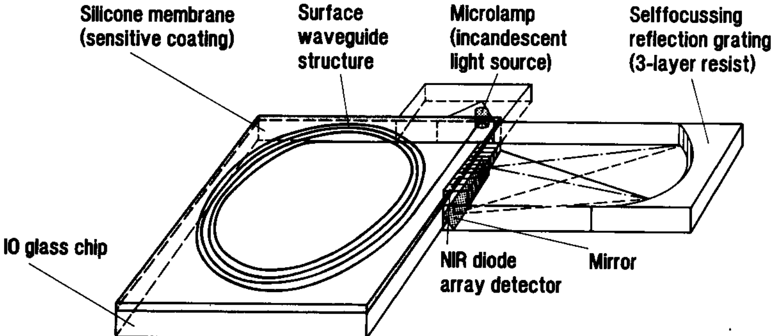


Figure 4-45. Concept of an optical micro system for chemical analysis based on an integrated optical evanescent wave sensor, a miniaturized incandescent light source, a microspectrograph, and an NIR diode-array detector.

For the light dispersion modular element, a planar grating spectrograph realized by the LIGA technique, as described in Section 4.4, could be used. For the fabrication of the spectrograph with the LIGA process poly(methyl methacrylate) is used (three-layer resist), which only has good transparency at wavelengths up to 1100 nm. Therefore, for measurements in the spectral range 1600–2000 nm the spectrograph has to be fabricated from deuterated PMMA, which is suitable at higher wavelengths.

For light detection in the NIR range, either a InGaAs photodiode-array detector or a PbS array detector could be used. InGaAs arrays, whose diodes are light sensitive up to wavelengths of 1700 nm, are applicable only to a restricted number of analytes because the first overtone CH absorptions of most organic compounds are found at wavelengths ≥1700 nm. Therefore, the use of photoconductive lead sulfide detectors, which can be applied up to wavelengths of 3 μm, seems to be an alternative.

For the realization of such a micro system, a combination of the optimized single elements by suitable adhesion and bonding techniques has to be developed. Morevoer, miniaturized data acquisition and signal processing electronics and corresponding intelligent software for calibration and evaluation purposes have to be generated. If these problems can be solved, the described concept compared with the fiber-optic approach with a conventional spectrometer allows the construction of a cheaper but nevertheless powerful system for in situ chemical analysis of organic pollutants.

4.10 Near Field Thermal Lens System

Photoacoustic and photothermal techniques, such as photoacoustic spectroscopy (PAS), thermal lensing (TL), photothermal deflection spectroscopy (PDS), and photothermal phase shift spectroscopy (PTPS), have proved in the past few years to be valuable analytical tools for measuring very small concentrations of analytes in solids, liquids, and gases [40, 2, 19, 49, 23, 8]. Laser-induced photoacoustic spectroscopy (LIPAS) seems to be currently the most frequently applied and advanced method, particularly in the speciation and trace determination of rare earths and actinides in aqueous solutions [26]. The main reason why LIPAS is specially favored over photothermal techniques in analytical measurements simply lies in the very fast and easy alignment procedure allowing sporadic use of the apparatus. However, our recent experimental studies on the performances of the four related techniques mentioned above have shown that TL and PDS are superior in the determination of environmental pollutants [46].

Generally environmental pollutants are released into the environment during manufacture, transport, handling, and application in agriculture, industry, and households. The analytical in situ determination of the concentration levels of these organic and inorganic micropollutants, present in significant ecological components, is very important in order to acquire knowledge of their distribution, transfer pathways, and impact on ecosystems. The present state of the art for the analysis of pollutants, eg, pesticides, is discontinuous, off-line analysis, in which single samples are collected in the field, transported to the laboratory, and analysed there, mostly after time-consuming chemical treatments with modern, partially very sensitive instrumental analysis techniques, eg, gas chromatography-mass spectrometry (GC-MS), high-

performance liquid chromatography (HPLC), spectrofluorimetry, and immunoassay techniques. All these methods reach limits of detection (LODs) in the low parts per billion (1 ppb = 1 μg/kg) and sub-ppb range, but the accuracy and the reproducibility of the results in the sub-ppb range are often insufficient, as clearly demonstated in an interlaboratory experiment for the detection of atrazine using HPLC, GC, and GC-MS: eleven independent laboratories provided analytical values which deviated from the true amount of atrazine in the test samples (0.1 μg/l) by up to two orders of magnitude. In view of such inaccurate analytical results, major efforts are needed to enforce EC regulations on drinking water.

Recently, photothermal methods have been developed as novel, rapid, continuous, in situ, real-time procedures that reach the required detection limits without chemical enrichment steps. However, they still have the severe limitation that many time-consuming steps are needed to align the laser beams, requiring highly skilled staff.

Especially with respect to the potential hazards to the human population, the availability of rapid and accurate assay techniques for toxic pollutants in ground and drinking water and air is imperative because of the large number of water samples or workplace samples that may have to be assayed. To give an example, the EC regulations limit the total amount of pesticides in drinking water to 0.5 μg/l and that of a single pesticide to 0.1 μg/l. In order to achieve this goal, low-cost, accurate, real-time and easy to handle sensors are needed.

For the solution of this problem, recently in our laboratory a so-called near-field thermal lens system has been developed [46], allowing the construction in the future of a microstructured, miniaturized, adjustment-free and user-friendly analytical microprobe by using LIGA techniques.

4.10.1 The Thermal Lens Effect

The thermal lens was discovered by Gordon et al. [19]. The idea underlying thermal lensing is simple: an excitation light beam (pump beam) passes through the sample of interest, the light is tuned to an absorption line of the analyte, and the optical energy is absorbed by the medium. If the collisional quenching rate in the analyte is significant higher than the radiative rate, then most of the energy appears in the rotational-translation modes of the molecules. This heating of the medium modifies the refractive index. The change in the refractive index of the medium is detected by a probe laser beam, which changes its shape (converges or diverges) when passing the excitation light-irradiated region. This effect can be monitored as a change in intensity of the probe beam passing through a pinhole (Figure 4-46). The optimal distance of the waists of the two light beams should be a confocal length. The distance z_d between the sample and the pinhole is responsible for the size of the sensor.

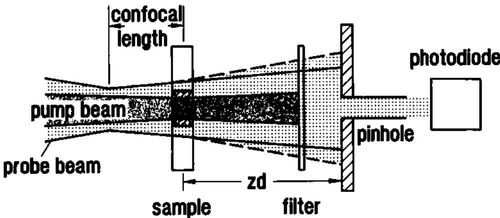


Figure 4-46.
Schematic illustration of the collinear photothermal lensing technique. The solid line shows the size of the probe beam when the pump beam is off and the dashed line that when the pump beam is on.

The thermal lens signal on the photodiode is ultimately dependent on the refractive index profile induced in the sample by the pump beam. The spatially variant refractive index is in turn due to a change in density as a consequence of energy being absorbed or, equivalently, as a result of a temperature change in the sample. Under equilibrium conditions, the temperature change is related to the refractive index through the differential

$$dn/dT = (dn/dT)_p + (dn/d\rho)(d\rho/dT)_p \tag{4-11}$$

where n is the refractive index, T the temperature, p the pressure, and ρ the density.

In practice, the thermal lens signal $S(t)$ obtained is a time-dependent function of the probe laser beam intensity. The intensity change of the probe beam in the presence of the thermal lens, $I_T(t)$, relative to its initial value, $I_T(t = 0)$, can be expressed by the following equation:

$$S(t) = \frac{I_T(t) - I_T(t = 0)}{I_T(t = 0)} . \tag{4-12}$$

4.10.2 Apparatus

First the thermal lens experiments were performed in a transverse and a collinear arrangement using the following experimental set-up. A continuous-wave UV argon ion laser (364 nm) beam, modulated between 2 and 10 Hz by a mechanical chopper, impinged into a quartz fluorescence cuvette. The radiation of an He-Ne probe laser (632.8 nm) was focused to intersect at different angles (0° or 90°) the excitation beam without offset appropriate to the photothermal method used [46]. In thermal lensing experiments a photodiode was used for the monitoring of the He-Ne probe beam. The distance between the pinhole and sample cell was varied between 400 and 4 mm, depending on a far-field or a near-field thermal lensing geometry. In the latest experiments with the NFTL system, an optical fiber (400 μm core diameter) served as the aperture and as a light guide of the probe beam to the photodiode detector (see Figure 4-47). The pre-amplified TL signals were processed by two lock-in amplifiers and for data processsing a personal computer was used.

Various comparable experiments with pesticides in aqueous solutions have been performed with a conventional spectrophotometer and different TL arrangements (transverse/collinear) and, the most important point with a conventional far field ($z_d = 400$ mm, Figure 4-46) and near field ($z_d = 4$ mm, Figure 4-46). The detection limits for the traditional spectrophoto-

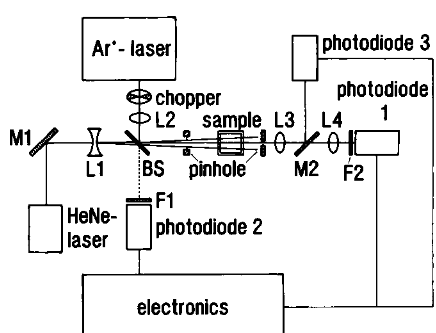


Figure 4-47.
Detailed illustration of experimental set-up of the near-field thermal lens system on a laboratrory scale [9].

meter and transverse TL were in the low parts per billion range. However, collinear near-field thermal lensing was shown to be superior in the determination of environmental pollutants (LOD = 0.1 μg/l), where detection limits in the parts per trillion range (eg, for pesticides in drinking water) are often required. It is interesting to note that irrespective of the distance between the sample and the pinhole (far field 400 mm, near field 4 mm), identical detection limits for both arrangements were obtained.

Recently, we have carried out experiments that show that the normally huge He-Ne probe laser (Figure 4-47) can be easily replaced by tiny laser diodes. Also, we succeeded in substituting the very expensive and space-consuming excitation laser (Figure 4-47) by smaller solid-state lasers or conventional light sources, which have recently become available.

This excellent performance of the near-field thermal lens (NFTL) opens up new doors to the construction of a miniaturized analytical microprobe.

4.10.3 Concept of Miniaturized NFTL System

The great potential of the thermal lens technique as a routine method for process control or environmental surveillance lies in a miniaturization of the near-field thermal lens system for the following reasons:

- quality of the light beams;
- fixed components;
- no adjustment;
- form of the sample cell.

This has at least two distinct advantages. On the one hand, a small, compact analytical apparatus allows direct real-time and on-site measurements. On the other hand, it is possible to build a robust and reliable device, in one cast and without any movable part, unaffected by external influences. With regard to reproducibility of the results of the measurements, short light paths of the excitation and probe lasers are desirable, because of the pointing stabilities of the two light beams, which has a very important impact on the signal-to-noise ratio, and thus on the accuracy of the analytical method.

Figure 4-48 shows a schematic drawing of an envisaged miniaturized prototype, which consists of a compact cast of the size of a match-box. The pump beam, transmitted by an optical fiber, converges with the probe laser beam through an edge filter and impinges through a pinhole on the sample, producing the thermal lens there, and is absorbed by an optical filter behind the sample. The laser diode probe beam defocused by the thermal lens passes through the optical filter and the second pinhole and hits the photodiode. As optical lenses, directly integrated microgradient lenses are used. The detector and the pinhole on the right side can be easily replaced by an optical multi-mode fiber (core size 400 μm), as shown in Figure 4-47. All these components have already been installed in various photothermal experiments and successfully tested.

What still remains to be done is the realization of a microstructured flow-through sample cell with a volume in the nanoliter range, fed by an integrated fluid-handling system. The remaining optical components, such as the edge filter, quartz windows, and absorption filter, will be integrated into the sensor head by using the LIGA technique.

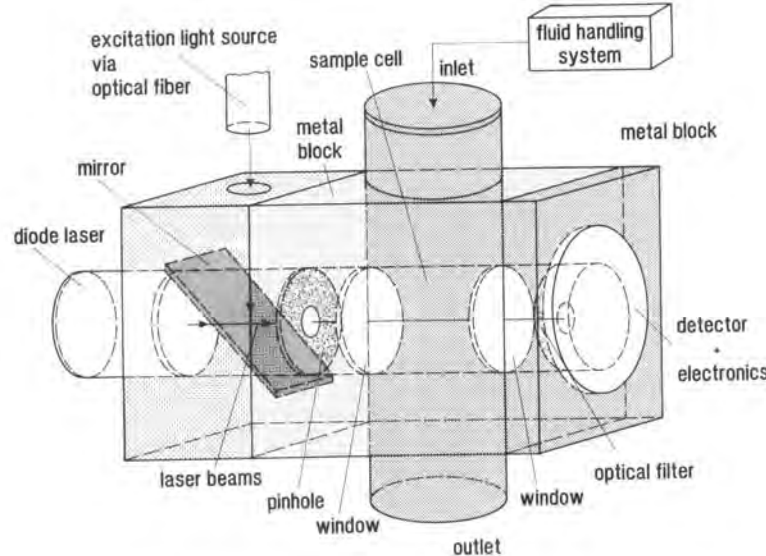


Figure 4-48. Concept of the miniaturized NFTL system.

4.11 Chemical Microanalyzer Systems: Their Role and Importance in Modern Chemical Analysis in Environmental Workplace and Process Control and Medical Diagnostics: a Resumé

Monitoring the environment for the presence of compounds that may adversely affect human health and local ecosystems is a fundamental part of the regulation, enforcement and remedial processes which will be required to maintain a habitable environment. However, emission monitoring is only one aspect which has to be considered to accomplish this goal. Equally important is the emission control of toxic compounds and, more generally, the process control of chemical processes. If a chemical process can be carried out under optimal conditions, less materials will be required, less waste will be produced, and effluent emission to the environment can be reduced to a minimum.

Although classical analytical techniques are being continuously refined and improved, these laboratory-based methods, the so-called off-line analyses, are, for the most part, relatively expensive and time consuming. With the rapidly increasing demand for more and tighter emission control in all sectors of industry and of the environment, it becomes obvious that an entirely new and less costly approach based on on-line or preferentially in-line techniques where chemical parameters are continuously and perhaps remote-controlled measured has to be developed.

It seems that the realization of such a new, innovative instrumental concept for process control and monitoring can best be achieved by microsystems technology, especially utilizing "three-dimensional microsensor technology".

4.12 References

[1] Ache, H.-J., *Chem. Ind. (Berlin)* **1** (1993) 40-43.
[2] Adelhelm, K, Faubel, W., Ache, H.-J. *Opt. Sci.* **69** (1992) 41-45.
[3] Bacher, W., Feit, K., Harmening, M., Michel, A., Mohr, J., Stark, W., Stölting, J., *KfK-Nachr.* **23** (2-3) (1991) 84-92.
[4] Becht, R., Bruns, M., Hoffmann, W., Ache, H.-J., *44th ISE Meeting, Berlin,* (1993) 599.
[5] Becker, E. W., Ehrfeld, E., Hagmann, P., Maner, A., Münchmeyer, D., *Microelectron. Eng.* **4** (1986) 35.
[6] Bergveld, P., *IEEE Trans. Biomed. Eng.* **17** (1970) 70.
[7] Bier, W., Guber, A., Linder, G., Schaller, T., Schubert, K., *KfK-Ber.* No. 5238 (1993) 132-137.
[8] Bohnert, B., Faubel, W., Ache, H.-J., Fresenius, *J. Anal. Chem.* **343** (1992) 513-517.
[9] *CC310 Zielspezifikation;* Reutlingen: Robert Bosch, 1990.
[10] Both, A., Bacher, W., Ruprecht, R., *Sensors Mater.* submitted for publication.
[11] Burbaum, C., Mohr, J., Bley, P., Ehrfeld, W., *Sensors and Actuators* **27** (1991) 559-563.
[12] Burbaum, C., Mohr, J., *KfK-Ber.* No. 4859, (1991).
[13] Büstgens, B., Bacher, W., Bier, W., Ehnes, R., Keydel, L., Maas, D., Ruprecht, R., Schomburg, W. K., *Proc. Actuator '94, Bremen, June 15-17 1994* (1994) 86-90.
[14] Büstgens, B., Bacher, W., Menz, W., Schomburg, W., *Proc. MEMS '94,* Oiso/Japan (1994) 18-21.
[15] Bychkov, E., Bruns, M., Klewe-Nebenius, H., Pfennig, G., Raptis, K., Hoffmann, W., Ache. H.-J., *Sensors and Actuators,* submitted for publication.
[16] Conzen, J.-P., Bürck, J., Ache, H.-J., *Appl. Spectrosc.* **47** (1993) 753-763.
[17] Faubel, W., Schulz, T., Seidel, B. S., Steinle, E., Ache, H.-J., *Technical Digest 8 ITMP³, Guadeloupe* (1994) 323-324.
[18] Fromhein, O., Gemmeke, H., Krömer, O., Kühner, T., Mohr, J., Strohrmann, M., *Eurosensors VIII,* Toulouse, (1994).
[19] Gordon, J. P., Leite, R. C. C., Moore, R. S., Porto, P. S., Whinnery, L. R., *J. Appl. Phys.* **36** (1964) 3-8.
[20] Harmening, M., Bacher, W., Bley, P., El-Kholi, A., Kalb, H., Kowanz, B., Menz, W., Michel, A., Mohr, J., *Proc. MEMS '92, Travemünde, Germany* (1992) 202-207.
[21] Hoffmann, W., Eggert, H., Schomburg, W., Seidel, D., *KfK-Ber.* No. 5238 (1993), 89-93.
[22] Hoffmann, W., Bruns, M., Ache, H.-J., Freywald, K.-H., *185th ECS Meeting, San Francisco* (1994) 300.
[23] Jackson, W. B., Amer, N. M., Boccara, A. C., Fournier, D., *Appl. Opt.* **20** (1981) 1333-1344.
[24] Kalb, H., *Dissertation;* Universität Karlsruhe, 1994.
[25] Karlberg, B., Pacey, G. E., *Flow Injection Analysis;* Amsterdam: Elsevier, 1989.
[26] Klenze, P., Kim, J. I., Wimmer, *Radiochim. Acta* **52/53** (1991) 97-103.
[27] Leuthold, H., Rudolf, F., *Sensors and Actuators* **A21-A23**, (1990) 278-281.
[28] Mastrangelo, C. H., Muller, R. S., Kumar, S., *Appl. Opt.* **30** (1991) 868-873.
[29] Matsuo, T., Esashi, M., *Sensors and Actuators* **1** (1981) 77.
[30] Mohr, J., Burbaum, C., Bley, P., Menz, W., Wallrabe, U., in: *Micro System Technologies 90,* Reichl, H. (ed.); Berlin: Springer, 1990, pp. 529-537.
[31] Mohr, J., et al., *Eurosensors V,* Rome. (1991).
[32] Mohr, J., Kohl, M., Menz, W., *Proc. Transducers 93, Yokohama, June 1993* (1993) 120.
[33] Morales-Bahnik, A., Czolk, R., Reichert, J., Ache, H.-J., *Sensors and Actuators B* **13** (1993) 424-426.
[34] Morales-Bahnik, A., Czolk, R., Ache, H.-J., *Sensors and Actuators B* **18-19** (1994) 493-496.
[35] Muller, R. S., Howe, R. T., Senturia, S. D., Smith, R. L., White, R. M., *Microsensors;* New York: IEEE Press, 1990.
[36] Müller, C., Mohr, J., *Interdiscip. Sci. Rev.* **18** (1992) 273-279.
[37] Murray, R. W., Dessy, R. E., Heineman, W. R., Janata, J., Seitz, W. R., *Chemical Sensors and Microinstrumentation; ACS Symposium Series, No. 403;* Washington, DC: American Chemical Society, 1989.
[38] Pham, M. T., Howitz, S., Hellfeld, T., Albrecht, J., *IGT-Fachber.* **126** (1994) 349-354.

[39] Rapp, R., Bley, P., Menz, W., Schomburg, W. K., *KfK-Ber.* No. 5251 (1993).
[40] Rosencwaig, A., *Photoacoustics and Photoacoustic Spectroscopy;* New York, Wiley, 1980.
[41] Ross, L., *Glastech. Ber.* **62** (1989) 285-297.
[42] Ruprecht, R., Both, A., Bacher, W., *MicroMechanics Europe '93, Workshop Digest* (1993) 93-96.
[43] Schomburg, W. K., Baving, H. J., Bley, P., *Microelectron. Eng.* **13** (1991) 323-326.
[44] Schöning, M., Bruns, M., Hoffmann, W., Ache, H.-J., Hoffmann, B., *Sensors and Actuators B* **15-16** (1992) 192-194.
[45] Schomburg, W. K., Vollmer, J., Büstgens, B., Fahrenberg, J., Hein, H., Menz, W., *J. Micromech. Microeng.,* 4/3 (1994).
[46] Schulz, T., Faubel, W., Ache, H.-J., *KfK-Ber.* No. 5370 (1994).
[47] Thommes, A., Stark, W., Leyendecker, K., Bacher, W., Liebscher, H., Jakob, Ch., *Proc. 3rd Int. Symp. Magnetic Materials, Processes and Devices* **94-6** (1994) 89-102.
[48] van der Schoot, B. H., Jeanneret, S., van den Berg, A., de Rooij, N. F., *Int. Meet. Chem. Chem. Sensors, Tokyo* (1992) 3B09.
[49] Vyas, R., Gupta, R., *Appl. Opt.* **27** (1988) 4701-4711.
[50] Workshop on Micro Total Analysis Systems, Twente, 1994.
[51] Zimmermann, B., Bürck, J., Ache, H.-J., *KfK-Ber.* No. 4967 (1991).
[52] Anderer, B., Ehrfeld, W., Münchmeyer, D., SPIE, Vol. 1014 (1988) 17-24.
[53] Mohr, J., Bley, P., Strohrmann, M., Wallrabe, U., *J. Michromech. Microeng.* **2** (1992) 234-241.
[54] Ghica, V., Glashauser, W., German Patent 30 39 110 (1982).
[55] Göttert, J., Mohr, J., *KfK-Ber.* No. 5153 (1993).
[56] Göttert, J., Mohr, J., in: W. Ehrfeld et al. (eds.) *Integrierte Optik und Mikrooptik mit Polymeren,* Teubner Taschenbücher der Physik (1992).
[57] Mohr, J., Bacher, W., Bley, P., Strohrmann, M., Wallrabe, U., *Proc. 1er Congrès Franco-Japonnais de Mecatronique,* Besançon, France (1992).

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

5 Acoustic Wave Devices

GERHARD FISCHERAUER, Siemens AG, Munich, Germany
A. MAUDER AND R. MÜLLER, Technische Universität, München, Germany

Contents

5.1 Surface Acoustic Wave Devices 136
5.1.1 Introduction 136
5.1.2 Wave Types 136
5.1.3 SAW Devices 139
5.1.4 Sensing Mechanisms 142
5.1.5 System Considerations 145
5.1.6 Physical Sensors 147
5.1.7 Chemosensors 152
5.1.8 Conclusions and Outlook 156

5.2 Quartz Microbalance Transducers 157
5.2.1 Introduction 157
5.2.2 Quartz Substrate 159
5.2.3 Active Coating 160
5.2.4 Acoustic Equivalent Circuit 164
5.2.5 Electric Equivalent Circuit 167
5.2.6 Oscillator Circuits 170
5.2.7 Signal Processing 171
5.2.8 Noise Equivalent Concentration 172
5.2.9 Examples of Applications 173

5.3 References 175

5.1 Surface Acoustic Wave Devices

5.1.1 Introduction

This section deals with sensors based on surface acoustic wave (SAW) propagation phenomena. The design of such sensors generally follows the same principles that have been established for classical SAW filters and that are well documented in several textbooks [1-6]. One has to cope with the additional difficulty, however, that SAW sensors cannot be sealed hermetically like conventional filters, but must necessarily be exposed to the influence of external perturbations by physical or chemical quantities.

SAW sensors have many attractive features. In addition to being small, inexpensive, and easy to fabricate, they respond to many different measurands and have a wide dynamic measurement range. Further, their rf output signal is easily post-processed.

The objective of this section is threefold: first, to survey the variety of wave types, device geometries, and sensing mechanisms that may be used to design a SAW sensor; then, to present a review of the physical and chemical sensors that have been developed in the last decade; and finally, to assess current trends and possible developments in the future.

5.1.2 Wave Types

It was Lord Rayleigh [7] who, at the end of the last century, theoretically treated the problem of acoustic wave propagation along the surface of a semi-infinite isotropic medium. The physics is mathematically described by a boundary value problem derived from the equation of motion (Newton's second law for infinitesimally small volumes), the constitutive relationship for an elastic medium (generalized Hook's law), and the boundary conditions (no normal forces on the free substrate surface). One possible solution, now called the *Rayleigh wave* after its discoverer, describes a wave strongly bound to the surface in that there is hardly any motion in the substrate at depths greater than about one acoustic wavelength λ. In fact, the particle displacement amplitude decreases exponentially with depth. The particles near the surface move along elliptical trajectories in the *sagittal plane*, ie, the plane containing the surface normal and the propagation direction (Figure 5-1 a). The phase velocity of the wave is independent of frequency (no dispersion), and smaller than that of the slowest bulk mode. Therefore, no radiation losses into the substrate occur because no phase-match condition between the Rayleigh wave and a bulk mode can be satisfied [8].

The isotropic half-space geometry does not support surface waves with a particle displacement perpendicular to the sagittal plane. Such a *shear-horizontal* (SH) polarization is only associated with a bulk wave that does not, however, decay with depth [3].

When the semi-infinite substrate consists of an anisotropic material, the governing equations admit more complex wave solutions, a particularly interesting case being that of piezoelectricity by which mechanical displacements and electric fields couple to each other. One possible wave resembles the Rayleigh mode of the isotropic geometry with its characteristics slightly modified by anisotropy. The particle motion associated with this *generalized Rayleigh wave* is not entirely confined to the sagittal plane (except for so-called pure-mode propagation directions), but has a small shear component. A different surface

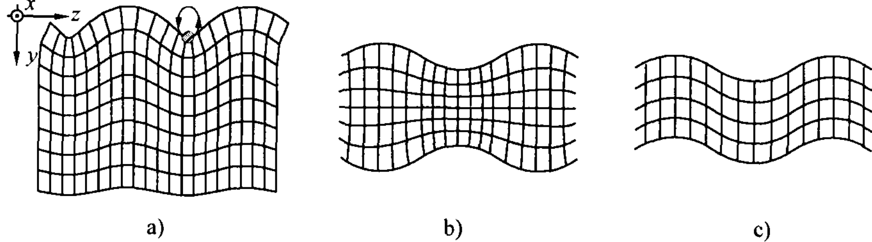


Figure 5-1. Fundamental acoustic wave types. (a) Rayleigh wave. Also shown is the elliptical trajectory of a selected particle at the surface and a reference coordinate system. The wave propagates in the $+z$-direction, the unperturbed surface normal points in the $-y$-direction, and the sagittal plane is the yz-plane. (b) Symmetric Lamb wave. (c) Antisymmetric Lamb wave.

wave, the purely shear-horizontally polarized *Bleustein-Gulyaev (BG) wave*, exists if the sagittal plane is normal to a twofold symmetry axis of a piezoelectric crystal [9, 10]. It may be viewed as the bulk SH wave of the isotropic case which piezoelectricity has caused to be bound to the surface. The penetration depth is typically a few λ for a metalized surface and ten times as much for a free surface [1, 3].

Some anisotropic substrate cuts support surface waves that are faster than the slow bulk shear wave. Such a wave can only propagate without significant losses because its polarization is orthogonal or almost orthogonal to that of the bulk wave. In the former case, one speaks of a *pseudo-surface wave*, and in the latter case, which is connected with finite but small losses, of a *leaky surface wave* [3]. When bulk waves propagate at almost grazing angle to the surface (*surface skimming bulk waves* (SSBW)), the acoustic energy remains concentrated near the surface for very long distances so that the wave may be considered as surface-bound for all practical purposes [11]. Sometimes only the boundary conditions at the surface determine if the dominant mode is an SSBW or a leaky SAW.

In practice, the assumption of a semi-infinite substrate is an approximation valid only if the wave penetration depth is much smaller than the substrate thickness. This motivates the study of the characteristic waves in plate structures, the acoustic plate modes (APM), although they may not necessarily be true surface waves. An isotropic plate supports *SH* waves and the sagittally polarized *Lamb waves*. In either case, a series of dispersive modes with different cut-off frequencies and velocities exist. As to the Lamb waves, one distinguishes between symmetric, or dilatational, modes (Figure 5-1 b) and antisymmetric, or flexural, modes (Figure 5-1 c). In thin plates, the fundamental symmetric (S_0) mode propagates without dispersion, but the phase velocity of the fundamental antisymmetric (A_0) mode is proportional to the frequency times the plate thickness (hence, for plates thin enough, the A_0 mode is slower than the sound in the surrounding medium, even if it is a liquid, so that no radiation losses can occur). A Rayleigh wave can be thought of as superposition of S_0 and A_0 Lamb waves when the plate thickness is large in terms of λ. On the top face of the plate the two Lamb modes add up to the Rayleigh wave whereas they cancel each other on the rear face. The Lamb wave character will not show up unless the plate is thinner than a few wavelengths [8].

Anisotropic plates support *generalized SH and Lamb waves* whose polarizations are almost like those of the corresponding mode in an isotropic medium. Pure modes occur only for certain crystal orientations.

Unfortunately, and in particular in the field of sensors, one rarely deals with free substrate surfaces. Instead, one often encounters additional coatings such as metal overlays and electrode structures, chemosensitive materials, passivation layers, or thin piezoelectric films on non-piezoelectric substrates. Even in the simplest case, that of an isotropic layer on top of an isotropic half-space, one finds both a set of SH modes and a set of waves with sagittal polarization. The former, so-called *Love waves*, exist only if the bulk shear wave velocity in the layer is less than that in the substrate. At very low and very high frequencies, they reduce to bulk shear waves in the semi-infinite substrate and the layer, respectively. These waves may be regarded as bulk modes which are trapped near the surface and made dispersive by the presence of a layer with low acoustic velocity. The sagittally polarized modes are termed *generalized*, or *layered*, *Rayleigh waves* [1, 3], or *generalized Lamb waves* [8]. When the shear velocities in the layer and the substrate differ appreciably from each other, these modes, too, reduce to pure half-space Rayleigh waves at very high or very low frequencies [1]. There are many anomalies, though, for special material combinations. If the layer thickness (in λ) is small but not too small, degenerated modes called *Sezawa waves* exist. If the layer and substrate shear velocities are almost equal and the layer is much thicker than λ, a *Stoneley wave* propagates along the layer–substrate interface [8]. As before, anisotropic materials complicate the situation somewhat, but the resulting waves can be classified according to their similarity with one of the waves in the isotropic case.

Table 5-1 lists the various wave types discussed so far. In practical situations, there is not always a clear distinction between them. For instance, a sagitally polarized wave in a finite substrate coated with a film overlay may be viewed as a perturbed Rayleigh wave, as a superposition of perturbed Lamb waves, or as a generalized Rayleigh (or Lamb) wave. For this

Table 5-1. Basic surface-generated wave types. All wave types propagating in anisotropic substrates have already been applied for sensing purposes.

Material	Polarization[a]	Geometry[b]		
		Semi-infinite	Plate	Layered structure
Isotropic	SP	RW	Lamb wave	Layered RW (Sezawa wave, Stonely wave)
	SH	–	SH plate mode	Love wave
Anisotropic	SP	Gen. RW Pseudo SAW Leaky SAW	Gen. Lamb wave	Gen. layered RW
	SH	SH wave BGW Leaky SAW SSBW	Gen. SH wave	Gen. Love wave

[a] Wave polarization. SP = (quasi-)sagittal polarization; SH = (quasi-)shear-horizontal polarization.
[b] Gen. = Generalized; BGW = Bleustein-Gulyaev wave; RW = Rayleigh wave; SSBW = surface skimming bulk wave.

reason, although the emphasis in this review has been laid on "true" surface waves, some attention has also been paid to Lamb and SH plate modes, because these waves can be excited by the same means as surface waves in the strict sense (see the following subsection).

5.1.3 SAW Devices

The key factor to the success of SAW filters as signal processing devices has been the development of a very efficient means for generating and receiving the acoustic waves: the *interdigital transducer* (IDT). In its simplest form, an IDT consists of a series of parallel planar metal electrodes periodically spaced on the surface of a piezoelectric substrate and alternately connected to each other (Figure 5-2). An rf voltage applied to the two interleaved electrode combs produces an electric field with spatial and temporal periodicities $2p$ and $1/f$, respectively. Owing to the piezoelectric effect, the oscillatory electric field gives rise to a periodically changing mechanical stress pattern in the substrate. This, in turn, leads to the launching of elastic surface waves which leave the transducer in both directions normal to the electrodes (the IDT is said to be *bidirectional*). By reciprocity, the detection process works the other way round: an acoustic wave impinging on an IDT is converted into an rf output voltage between the two electrode combs.

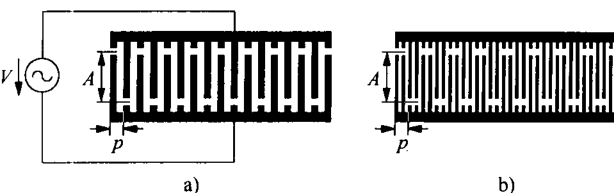


Figure 5-2. Uniform interdigital transducers (IDT) with aperture A and finger period p. (a) Single-electrode IDT with rf voltage source V. (b) Double-electrode, or split-finger, IDT.

The launching and receiving efficiency is maximum when the finger period p coincides with half of the acoustic wavelength $\lambda = v_{ph}/f$ (or odd multiples thereof), where v_{ph} is the phase velocity of the wave. Hence the fundamental resonance, or synchronous, frequency of the IDT is

$$f_0 = \frac{v_{ph}}{2p} \,. \tag{5-1}$$

For a given substrate, it depends only on the finger period p. Since the IDT is a planar structure, it can be fabricated conveniently and accurately by photolithographic methods. However, the higher the frequency is to be, the more challenging become the demands on the photolithographic resolution. Given today's limit of about 1 μm for p, and considering that the typical acoustic velocity is 3000–4000 m/s, the maximum frequency achievable with fundamental-harmonic SAW devices is about 2 GHz.

The main features of SAW filter design are appreciated by a simplified analysis which assumes that the inter-electrode gaps, or overlaps, work as delta-function sources. In this ap-

proximation, the amplitude of the launched surface wave for a uniform array of $N + 1$ fingers and for frequencies close to the fundamental resonance f_0 is given by

$$A(f) = A_0 N \frac{\sin \bar{f}}{\bar{f}} \quad \text{where } \bar{f} = N\pi \frac{f - f_0}{2f_0} \tag{5-2}$$

with some constant A_0. The 4 dB band width of this function is $B = 1/T$ with the approximate transducer time "length" $T = Np/v_{ph}$. There is clearly an intimate relationship between geometry (transducer length) and frequency characteristics (band width). In fact, it can be shown that the inverse Fourier transform of the frequency response, ie, the impulse response, of an IDT is proportional to the function describing the position-dependent wave excitation strength. For the uniform IDT considered so far, the sources are all equal in magnitude. Hence, the impulse response is a (sampled) rectangular-shaped function whose Fourier transform yields the well known $(\sin x)/x$ behavior of Equation (5-2).

The art of the filter design consists in obtaining a desired frequency response other than that given by Equation (5-2), mostly in order to suppress unwanted frequencies. This is accomplished by non-uniform transducers in which the local wave excitation strength varies along the finger array. Some methods commonly applied to this end are the entire omission of selected sources (*withdrawal weighting*, Figure 5-3 a), or the local variation of electrode lengths (*overlap weighting*, or *apodization*, Figure 5-3 b). *Chirped transducers* are characterized by a locally varying finger period (Figure 5-3 c). Owing to Equation (5-1), signals of different frequency are excited in different sections of such an IDT. As a consequence, the group delay time of a filter comprising chirped transducers becomes frequency dependent. Finally, in *modulated transducers*, a finger is not simply connected to the next but one neighbor. Instead, the basic electrical connection sequence has been multiplied by some periodic modulation function (Figure 5-3 d). In the frequency domain, this amounts to the generation of new harmonics. The procedure may be used to obtain devices that operate efficiently at some higher harmonic.

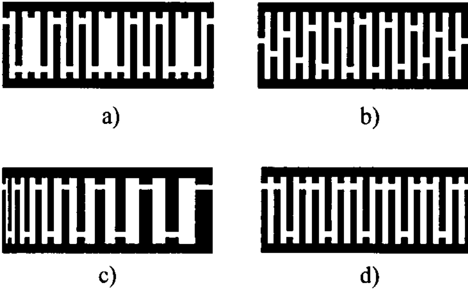


Figure 5-3.
Some IDT weighting techniques.
(a) Withdrawal weighting. (b) Apodization.
(c) Chirped IDT. (d) Modulated IDT.

Next to the IDT, the most important SAW filter element is the acoustic reflector. Any mechanical, geometrical, or electrical discontinuity at the substrate surface leads to a partial reflection because it locally changes the nature of the acoustic wave guide. Therefore, distributed reflectors can be constructed from arrays of small discontinuities such as electrodes or grooves. Arrays of $\lambda/4$-width fingers reflect most strongly at the *Bragg frequency* which happens to coincide with the synchronous frequency of an IDT given by Equation (5-1).

Especially when dealing with narrowband (long) transducers, one often attempts to avoid finger reflections in the filter passband by using split-electrode or modulated IDTs which shift the Bragg frequency with respect to the synchronous excitation frequency. On the other hand, finger reflections, and in particular the interplay of mechanical and electrical reflections, may be exploited to construct directional transducers.

In connection with sensors, two filter structures have been most widely used: the *delay line* and the *resonator*. The former essentially consists of two transducers placed some distance apart, where the center-to-center spacing determines the time delay (Figure 5-4a). A resonator is formed by one or more IDTs placed between reflector gratings which confine the acoustic energy inside a resonant cavity (Figure 5-4b, c). Several such resonators may be coupled electrically or acoustically to give a *resonator filter*. Yet another device is the *ID tag*, an electrical one-port comprising several reflectors and an IDT connected to an antenna (Figure 5-4d). When the antenna picks up an electromagnetic interrogation impulse, the IDT launches an acoustic wave that will be partially reflected by each of the reflectors. The reflected wave portions are received again by the IDT and re-radiated by the antenna. Hence the impulse response of the device consists of a series of impulses the time separation of which is determined solely by the reflector positions.

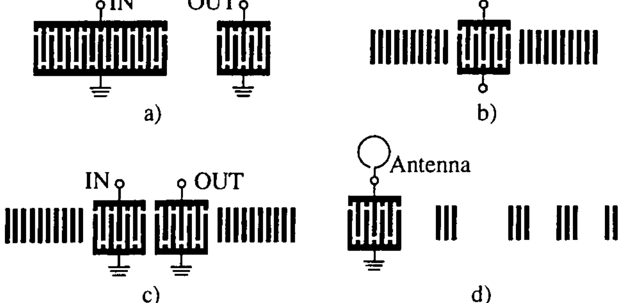


Figure 5-4. Some fundamental SAW devices. (a) Delay line. (b) One-port resonator. (c) Two-port resonator. (d) ID tag.

The most important parameters of the SAW substrate material are the wave polarization, the propagation velocity v_{ph}, the strength of the electroacoustic interaction (characterized by the coupling factor k^2), and the temperature behavior (characterized by the linear temperature coefficient of delay (TCD) for the free surface). These data have been compiled in Table 5-2 for some commonly used substrate cuts.

Table 5-2. Properties of some commonly used substrate material cuts.

Material	Orientation[a)]		Wave type	v_{ph}	k^2	TCD
	Cut	Prop.		(m/s)	(%)	(ppm/°C)
Quartz	ST	X	Gen. RW	3158	0.1	0
	37°rotY	90°rotX	SH wave	5094	~0.1	0
$LiNbO_3$	Y	Z	Pure RW	3488	4.1	94
	41°rotY	X	Leaky SH wave	4750	15.8	69
	128°rotY	X	Gen. RW	3980	5.5	75
$LiTaO_3$	36°rotY	X	Leaky SH wave	4220	~6.6	30

[a)] Cut = crystalline orientation of the substrate surface normal; Prop. = crystalline orientation of the wave propagation direction.

The fabrication of SAW sensors, especially chemosensors, offers some problems not encountered with traditional (hermetically sealed) devices. The chip mounting is very crucial because residual stresses in the chip or the outgassing of adhesives may interfere with the sensing process. In aggressive environments, additional passivation layers must protect the conventionally employed aluminum fingers from corrosion, or inert electrode materials such as gold must be used. To reduce the water cross-sensitivity, a hydrophilic substrate surface may have to be hydrophobized.

5.1.4 Sensing Mechanisms

SAW sensors are often equated with mass-sensitive devices because many of them exploit the velocity decrease caused by mass adsorption at the surface. However, many other sensing mechanisms exist which are discussed in the present subsection.

The phase response, or equivalently the delay time, of a SAW device is directly influenced by measurands that alter the propagation path length between input and output IDTs, a typical example being a mechanical force straining the substrate chip. The situation is complicated, however, by the fact that almost all mechanisms that strain the chip also influence the elastic material constants and, therefore, the wave velocity. A delay line with center-to-center transducer spacing L exhibits a delay time of $\tau = L/v_{\mathrm{ph}}$. Hence, a small variation in the measurand x results in a delay time change of

$$\frac{\Delta\tau}{\tau} = \frac{\Delta L}{L} - \frac{\Delta v_{\mathrm{ph}}}{v_{\mathrm{ph}}} = (S_x^L - S_x^{v_{\mathrm{ph}}})\,\Delta x \tag{5-3}$$

where the *sensitivity* of y to a change in z is defined by

$$S_z^y := \lim_{\Delta z \to 0} \frac{1}{y} \cdot \frac{\Delta y}{\Delta z} = \frac{1}{y} \cdot \frac{\partial y}{\partial z} \tag{5-4}$$

The term $\Delta L/L$ represents the mechanical strain generated by the measurand. It may be caused, of course, by direct application of a mechanical force, but also by other effects such as electrostriction [71, 75] or temperature changes (in that case the sensitivity $S_T^L = (1/L)\, \partial L/\partial T$ is the well known thermal expansion coefficient) [2]. The second term in Equation (5-3) involves the velocity change Δv_{ph} produced by a change in the material parameters. Depending on the geometry and substrate material, either effect can be dominant. For instance, in Y,X-$LiNbO_3$ subjected to a biasing electric field, the strain contribution exceeds the velocity change by far, whereas for lead zirconate titanate (PZT), the ratio is less than 1% [71].

Many phenomena, however, influence the acoustic wave velocity and the propagation attenuation without straining the crystal at all. In fact, most SAW sensors realized to date exploit the sensitivity of the velocity to changes in the boundary conditions at the substrate surface. Such changes may have mechanical, electrical, magnetic, or thermal causes.

One often deals with small disturbances. This justifies the use of perturbation methods to derive approximate formulas for the fractional velocity shift $\Delta v_{\mathrm{ph}}/v_{\mathrm{ph}}$ and the change in the attenuation coefficient α (in Np/m) normalized to the wavenumber β, ie, $\Delta\alpha/\beta$. Table 5-3 sum-

Table 5-3. Fractional velocity shift $\Delta v_{ph}/v_{ph}$ and normalized change in the attenuation coefficient $\Delta\alpha/\beta$ caused by different surface wave perturbations.

Origin of perturbation	$\Delta v_{ph}/v_{ph}$ [a]	$\Delta\alpha/\beta$ [a]	Equation	Reference
Thin-film overlay[b]:				
Mass loading	$-(K_x + K_y + K_z)\rho' h' f$	0	(5-5)	[8, 14]
Elastic loading	$\frac{\mu'}{v_{ph}^2}\left(K_x + 4\frac{\lambda' + \mu'}{\lambda' + 2\mu'}K_z\right)h' f$	0	(5-6)	[8]
Viscoelastic loading	$-C\omega h'\mu' \cdot \frac{(\omega\tau')^2}{1 + (\omega\tau')^2}$	$C\omega h'\mu' \cdot \frac{\omega\tau'}{1 + (\omega\tau')^2}$	(5-7)	[22]
Electrical interactions	$-\frac{k^2}{2} \cdot \frac{(\sigma' h')^2}{(\sigma' h')^2 + v_{ph}^2(\varepsilon_0 + \varepsilon_S)^2}$	$\frac{k^2}{2} \cdot \frac{v_{ph}\sigma' h'(\varepsilon_0 + \varepsilon_S)}{(\sigma' h')^2 + v_{ph}^2(\varepsilon_0 + \varepsilon_S)^2}$	(5-8)	[4, 23–25]
Operation in liquid[c]:				
Viscous loading	$-\frac{K_x}{2\pi \cdot \sqrt{2}}\sqrt{\omega\varrho'\eta'}$	$\frac{K_x}{2\pi \cdot \sqrt{2}}\sqrt{\omega\varrho'\eta'} + \frac{K_y}{2\pi}\sqrt{\varrho'\kappa' - \frac{K'^2}{v_{ph}^2}}$	(5-9)	[26, 90, 177]
Electrical interactions	$-\frac{k_L^2}{2} \cdot \frac{\sigma'^2 + \omega^2(\varepsilon_L' - \varepsilon_L)(\varepsilon_L' + \varepsilon_S)}{\sigma'^2 + \omega^2(\varepsilon_L' + \varepsilon_S)^2}$	$\frac{k_L^2}{2} \cdot \frac{\sigma'\omega(\varepsilon_L' + \varepsilon_S)}{\sigma'^2 + \omega^2(\varepsilon_L' + \varepsilon_S)^2}$	(5-10)	[26, 90, 177]

[a] The formulas express the change with respect to the free substrate surface except for the perturbation by electrical interactions with a liquid in which case the reference is given by a non-conducting liquid.

The symbols used have the following meanings (note that all primed quantities refer to the thin film or the liquid): h' = film height; η' = dynamic viscosity; κ' = volume elasticity of liquid; λ', μ' = Lamé constants; ρ' = mass density; σ' = conductivity; $\tau' = \eta'/\mu'$ = shear relaxation time; ε_0 = free-space permittivity; ε_L, ε_L' = unperturbed and perturbed liquid permittivities, respectively; $\varepsilon_S = [\varepsilon_{yy}^T \varepsilon_{zz}^T - (\varepsilon_{yz}^T)^2]^{0.5}$ = effective substrate permittivity at constant stress; k^2, k_L^2 = electroacoustic coupling factors for operation in air and in the reference liquid, respectively ($k_L^2 \approx k^2(\varepsilon_0 + \varepsilon_S)/(\varepsilon_L + \varepsilon_S)$ [21, 30]); $\omega = 2\pi f$ = angular frequency; K_x, K_y, K_z, and C are coordinate-dependent constants (the reference coordinate system is shown in Figure 7-1a).

[b] The formulas for perturbation by mass and elastic loading are valid for surface waves in general but the other two cases apply to Rayleigh waves.

[c] The formulas have been derived for leaky SH waves. The equations relating to electrical interactions also describe the case of a BG device [29] or of an SH plate mode device [21, 30–32] immersed in a liquid with constant permittivity ($\varepsilon_L' = \varepsilon_L$), but varying conductivity.

marizes the results of such calculations. The frequency-independent constants K_i ($i = x, y, z$) that appear in these formulas are functions of the substrate material parameters. They can be computed from the unperturbed particle velocity components $\dot{u}_i$ at the surface and the average power flow per unit width P by $K_i = v_{ph}|\dot{u}_i|^2/4fP$. Special cases are $K_x = 0$ (pure Rayleigh wave), $K_y = K_z = 0$ (pure SH wave), or $K_z = 0$ (quasi-SH, or leaky SH, wave).

The K_i are typically 50 ppb/(Hz kg/m^2). Hence, according to Equation (5-5) for the fractional velocity shift caused by mass loading, a measurement set-up capable of resolving velocity shifts of 0.1 ppm and based on a 1 GHz delay line with an active area of 1 mm^2 is able to detect mass changes in the subpicogram range. Further, the mass sensitivity $S_m^{v_{ph}}$ of surface wave devices coated with a given thin film is proportional to the frequency (hence the tendency to operate SAW sensors at high frequencies, typically a few hundred MHz). For Rayleigh waves in an isotropic medium of mass density ρ, the equations can be rearranged analytically in the form $\Delta v_{ph}/v_{ph} = -(K/\rho\lambda)\rho' h'$ where K is a material constant ranging from 1 to 2 for most materials [16]. Obviously the mass sensitivity is inversely proportional to the mass per unit area of a characteristic layer which is one wavelength thick and within which most of the acoustic energy is concentrated. This also holds true for Lamb waves [16, 18–20] and SH modes [21] in plates, the characteristic layer now being formed by the entire plate (hence the tendency to use very thin plates). Generally, the wave interacts the stronger with surface perturbations the more the acoustic energy is concentrated near the surface [17]. Thicker coatings [168, 191, 192] and additional layers increasing the energy trapping therefore lead to higher sensitivities [118, 176].

In contrast to mass and elastic loading, which do not introduce additional propagation losses, viscoelastic interactions affect both the velocity and the attenuation of an acoustic wave. A lightly cross-linked polymer can be modeled as Maxwellian viscoelastic medium with dynamic viscosity η', shear stiffness μ', and density ρ' [22]. It exhibits a Newtonian viscous behavior for frequencies much smaller than the inverse of the shear relaxation time $\tau' = \eta'/\mu'$, ie, for $\omega\tau' \ll 1$, and behaves as an elastic amorphous solid with shear modulus μ' for $\omega\tau' \gg 1$.

Similar comments apply to the loading by a viscoelastic liquid. Note that Equation (5-9) holds for the Newtonian case. The corresponding expressions are $\Delta v_{ph}/v_{ph} \sim (\omega\eta')^2$, $\Delta\alpha/\beta \sim \omega\rho'\eta'$ for BG waves [27] and $\Delta v_{ph}/v_{ph} \sim (\rho'\eta'/\omega)^{0.5}$, $\Delta\alpha/\beta \sim (\omega\rho' h')^{0.5}$ for SH plate modes [21, 63]. The result for the fractional velocity shift of a Lamb wave is more complicated, but the attenuation change is given by $\Delta\alpha/\beta \sim (\omega\rho'\eta')^{0.5}$ [28]. Shear relaxation effects, ie, the transition from Newtonian viscous behavior to elastic behavior, can be taken into account by multiplying the low-frequency formulas for $\Delta v_{ph}/v_{ph}$ and $\Delta\alpha/\beta$ with correction functions $F_v(\omega\tau')$ and $F_\alpha(\omega\tau')$, respectively. These functions are given by $F_v(x) = [F_\alpha(x)]^2 = 1/(1 + x^2)$ for BG waves [27] and $F_v(x) = (1 + x^2)/F_\alpha(x) = [1/(1 + x^2)^{0.5} - x/(1+x^2)]^{0.5}$ for SH plate modes [21, 63].

Perturbations caused by conductivity changes in a thin film overlay (cf, Equation (5-8)) have been shown to be the dominant effect in the detection of NO_2 or halogens with (metallo)phthalocyanine coatings [24, 128], in H_2S sensors based on metal oxide films [25, 129, 130, 132, 135, 150], and in polymer-coated humidity sensors [161]. Conductivity changes in a loading liquid (cf, Equation (5-10)) have been used in practical sensors, too [26, 90–92, 177]. Note that the wave attenuation peaks for some critical conductivity given by $v_{ph}(\varepsilon_0 + \varepsilon_S)/h'$ and $\omega(\varepsilon_0 + \varepsilon_S)$ for film and liquid loading, respectively.

Martin and co-workers [22, 133, 142, 143, 159, 164] were the first to point out that the simultaneous monitoring of velocity and attenuation leads to additional information not

available from either variable alone. In the presence of losses, wave propagation is characterized by a complex propagation constant $\gamma = \beta - j\alpha$. As the result of a perturbation, the normalized change in the propagation constant is $\Delta\gamma/\beta = \Delta\beta/\beta - j\Delta\alpha/\alpha = -\Delta v_{ph}/v_{ph} - j\Delta\alpha/\beta$. By plotting the locus of $\Delta\gamma/\beta$ in the complex plane as function of the perturbing parameter, one obtains characteristic curves which shed light on the particular mechanism at work. For instance, mass loading decreases the velocity, but does not affect the attenuation; hence the locus is a line parallel to the real axis. In contrast, $\Delta\gamma/\beta$ traces a semicircle $(\Delta v_{ph}/v_{ph} + R)^2 + (\Delta\alpha/\beta)^2 = R^2$ with $R = C\omega h'\mu'/2$ and $R = k^2/4$ for viscosity and conductivity changes in thin films, respectively, (cf, Equations (5-7) and (5-8)).

In addition to monitoring two independent variables, one can use multiple-frequency devices to augment the information content of the sensor output [164, 177]. For Rayleigh waves, the fractional velocity shift due to the mechanical, viscoelastic, and electrical perturbation by a thin film is proportional to f, $f^{3/1}$ (Newtonian/elastic case), and f^0, respectively. The normalized attenuation coefficient change is zero for elastic loading and goes as $f^{2/0}$ (Newtonian/elastic case) and f^0 in the other two cases.

A different sensing mechanism is the strong dependence of the finger reflectivity in an IDT or a reflector on the mechanical and electrical boundary conditions. This has apparently not yet been exploited in practice.

The discussion above reveals the difficulty in answering the question which wave type is the most sensitive. While, at a fixed frequency, plate modes exhibit higher mass sensitivities than surface waves proper [16, 168], one can easily operate the latter at much higher frequencies and thus increase their mass sensitivity. Further, surface wave devices may have a smaller detection limit because of their superior noise behavior [166], or they may become more sensitive than comparable plate mode sensors when mechanisms other than mass loading are exploited [26]. It is obvious that the "optimum" wave type for a given application is certainly not determined by the mass sensitivity alone but also by the nature of the sensing mechanism, the device producibility and ruggedness, the complexity of the required electronic instrumentation, the noise and aging behavior, etc.

5.1.5 System Considerations

SAW velocity changes can be detected by various methods. The most straightforward approach is to compare directly the transfer phase with some reference phase (Figure 5-5a). Another method consists in measuring the pulse rate obtained when a pulse detector at the output IDT of a delay line triggers the next pulse at the input IDT ("sing-around method", Figure 5-5b).

The most widely used arrangement, however, measures the rf frequency of an oscillator formed by inserting the SAW device in the feedback loop of an amplifier (Figure 5-5c). One condition for stable oscillation is, of course, the requirement that the round-trip phase shift in the closed loop is an integer multiple of 2π, ie, $\Phi_{SAW} + \Phi_e = -2n\pi$, where Φ_{SAW} and Φ_e are the phase responses of the SAW device and the electrical circuit, respectively, and n is a mode integer [12]. The phase response of a delay line with an effective length of L can be split into two parts. The first is due to the electroacoustic transduction process in the transducers and will be incorporated in Φ_e. The second is produced as the wave propagates down the

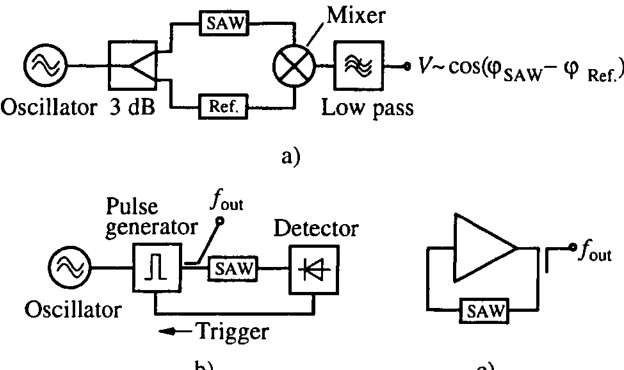


Figure 5-5. Basic electronic circuitry. (a) Direct phase measurement. (b) "Sing-around" arrangement. (c) Oscillator.

acoustic waveguide of length L and is given by $\Phi_{SAW} = -\omega L/v_{ph}$. Inserting this relationship in the phase equation and solving for the oscillation angular frequency ω yields

$$\omega = \frac{v_{ph}}{L}(2n\pi + \Phi_e)\,. \tag{5-11}$$

Now suppose that all quantities depend on a measurand x (eg, a gas concentration) and on a number of interfering quantities y_i (eg, temperature, humidity, acceleration, or even, to describe drift phenomena, time). For small deviations from the quiescent oscillation state, the frequency shift caused by the influence of both the measurand and the interfering quantities is obtained by implicitly differentiating Equation (5-11):

$$\frac{\Delta\omega}{\omega} = \frac{1}{\dfrac{v_{ph}}{v_{gr}} + \omega S_{\omega}^{L} - \dfrac{v_{ph}\,\Phi_e}{L} S_{\omega}^{\Phi_e}} \left[(S_x^{v_{ph}} - S_x^L + \alpha S_x^{\Phi_e})\,\Delta x + \sum_i (S_{y_i}^{v_{ph}} - S_{y_i}^L + \alpha S_{y_i}^{\Phi_e})\,\Delta y_i\right] \tag{5-12}$$

where the sensitivities S_z^y have been defined in Equation (5-4), v_{gr} denotes the group velocity of the acoustic wave, and $\alpha = 1/(1 + 2n\pi/\Phi_e)$. This equation has many implications, but a few fundamental consequences are easily appreciated. First, let us neglect interfering quantities and assume that the measurand itself only influences the wave velocity (not the IDT distance nor the electrical circuit, the influence of which, by the way, can be seen to vanish with increasing delay line length). Then,

$$\frac{\Delta\omega}{\omega} = \frac{v_{gr}}{v_{ph}} S_x^{v_{ph}}\,\Delta x = \frac{v_{gr}}{v_{ph}} \cdot \frac{\Delta v_{ph}(x)}{v_{ph}}\,. \tag{5-13}$$

The fractional oscillator frequency shift equals the fractional velocity shift produced by the measurand times the ratio of group to phase velocity. Hence, in the presence of dispersion, the frequency perturbation can be greater ($v_{gr} > v_{ph}$) or smaller ($v_{gr} < v_{ph}$) than the velocity perturbation. The two cases are exemplified by the A_0 Lamb wave [18] and by SH plate modes [21], respectively. The possible increase in sensitivity is unique to oscillators and does not occur in phase or sing-around measurement arrangements [18]. Equation (5-12) also shows that the chirping of transducers ($\partial L/\partial \omega \neq 0$) affects the oscillator response in the same manner as the intrinsic dispersion.

Another (expected) consequence of Equation (5-12) is that a dual-oscillator arrangement, in which the difference frequency between a sensing and a reference oscillator is evaluated, shows a residual sensitivity to interfering terms if the two oscillators behave in a different manner, however slightly. This is especially problematic for chemosensors because the reference sensor must of necessity be free of the sensing layer [160], but the problem also troubles physical sensors such as accelerometers. Joshi [75] has shown the advantage of monitoring the ratio of difference to reference frequency rather than the difference frequency alone.

The performance of an entire sensor system depends, of course, on both the sensor element and the electronic circuitry. It is of no use to maximize the sensitivity of the SAW filter if other relevant characteristics such as cross-sensitivities or the noise and aging behavior are adversely affected by this measure. A well designed sensor system results only when this interdependence of component characteristics and system performance is taken into account. For instance, the phase noise in a SAW oscillator decreases with the group delay time of the SAW filter, but increases with its insertion loss [12]. This leads to contradictory design requirements which must be resolved by some compromise. There is ample reason to believe that SAW sensor systems have not reached their performance limit by far.

5.1.6 Physical Sensors

Since the early 1970s, primarily motivated by the requirement for stable oscillator designs, much effort has been devoted to the problem of how to reduce the sensitivity of SAW devices to external perturbations such as temperature fluctuations, pressure changes, mounting stresses, or vibrations. On the other hand, it was soon realized that this very sensitivity can be exploited either to fine-tune the performance of the SAW device (eg, to vary the time delay of a delay line by applying an external electric or magnetic field) or to measure the interfering quantity.

The stress sensitivity of SAW filters can be employed to measure directly such stress-related mechanical quantities as pressure or tension, force, strain, and acceleration. These sensors share some common problems, of which the most severe are the very strong influence of the chip mounting, which can lead to sensitivity changes of some orders of magnitude [34, 37, 48, 53], and the large temperature cross-sensitivity [51–54]. The latter cannot even be completely eliminated by differential oscillator configurations because of sample variations and because the temperature coefficient of delay depends on stress (up to several ppm/(°C MPa) on quartz [53], although the influence may vanish on special material cuts such as 40.5°rotY,X-quartz [39]), which is different for the sensing and the reference elements. Quartz is most often used as substrate material because it shows a larger stress sensitivity and a lower temperature dependence than, eg, $LiNbO_3$ [54].

Hydrostatic pressure sensors have been realized by mounting SAW delay lines on thin circular diaphragms of quartz [33, 34, 37] or ZnO/Si [33], or with resonators on quartz plates [36, 39]. These devices typically exhibit a pressure sensitivity of the center frequency of 3 ppm/kPa. Much higher sensitivities of 2100 and 27 ppm/kPa, respectively, albeit with a proportionally smaller dynamic range, have been reported for A_0 [40, 41] and S_0 [42] Lamb mode delay lines on thin membranes of ZnO/Si or AlN/Si. The sensitivity of SAW force sensors strongly depends not only on the chip mounting, but also on the direction of the force. A typical value is 0.45 ppm/N, obtained with SAW delay lines on thin quartz membranes [34, 37]. Again, Lamb waves at 4 MHz are much more sensitive but also more fragile [40, 41]. Staples et al. [35] developed a force sensor based on a resonator filter which changed the amplitude response by a factor of 18 for a force of 400 N. Ishido et al. [43] used a delay line with input and output IDTs placed on separate chips in contact with each others' surfaces. An external force causes one chip to slide relative to the other, thereby changing the delay line length. The resulting sensitivity, although not linear, was about 10^4 ppm/N.

This two-chip structure also works as a linear position sensor with a sensitivity of 120.5 ppm/μm. Angular positions can be determined by mounting SAW resonators on pendulum-like geometries [44, 45]. Sensitivities of almost 3 ppm/mrad have been reported for such „spirit-levels“ [45]. Special position-sensing applications are to be found in the design of graphic tablets [46] and touch panels [47].

Mechanical strain has been measured both with resonators [38] and with delay lines [57]. The use of a PZT substrate results in a strain sensitivity of 21 ppm/10^{-6}, about 20 times that of quartz [57].

Acoustic accelerometers consist of a SAW device mounted on a thin diaphragm or a cantilever beam loaded by a proof mass of typically 10 g that exerts inertial forces on the SAW substrate when it is accelerated (Figure 5-6) [49–57]. Devices without quiescent stress yield the best repeatability of the measurement. Cantilever beams loaded in flexure cantilever beam accelerometers to compressional and transverse accelerations can be suppressed by a factor of 10^9 relative to the main response if FST-cut quartz (cut angles $\Psi = 26°$, $\Theta = -41.5°$, $\Phi = 6.34°$) is used and the IDTs are properly positioned on the beam. Recently, Sugizaki et al. [57] designed an accelerometer based on a PZT substrate with a sensitivity of 8.7 ppm/(m/s^2).

Rayleigh wave devices cannot be used as liquid viscosity or density sensors because the vertical displacement component strongly couples to compressional waves in the liquid and produces heavy radiation losses [8]. Instead, one must aim at suppressing the mode coupling by working with wave types either orthogonal to the sound wave in the liquid ((quasi-)SH waves

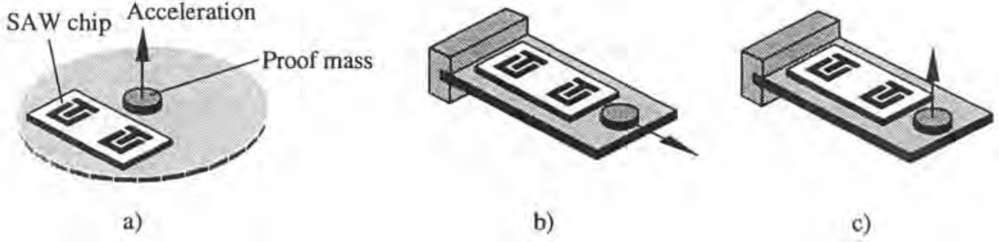


Figure 5-6. SAW accelerometer configurations. (a) Circular diaphragm. (b) Tensioned cantilever beam. (c) Flexured cantilever beam.

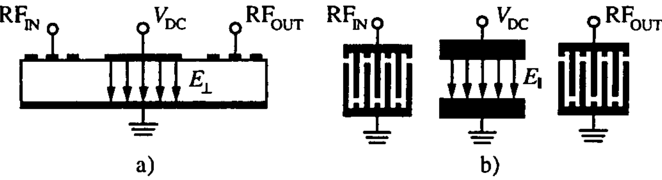
[58-60] or SH plate modes [63-64]) or propagating at a much smaller velocity (A_0 Lamb mode [28, 62]). The information on SAW viscometers published to date is more of a qualitative than a quantitative nature. In general, it seems better to evaluate the attenuation change rather than the velocity shift. Martin et al. [28], working with an A_0 Lamb mode, showed that both the fluid viscosity and density can be determined if velocity and attenuation are monitored simultaneously. In a low-viscosity water-glycerol solution, the observed density sensitivity of the frequency was about 30000 ppm/(g/cm^3). The sensitivity of the loss to the viscosity-density product was 34 dB/(g/cm^3 P)$^{0.5}$.

Gas and liquid flow-rate sensors are constructed by exploiting the strong temperature dependence of the SAW velocity in some substrates [65-70]. The crystal is heated above ambient temperature by integrated meander-line heaters or, as demonstrated by Brace et al. [67], by an absorbing epoxy layer dissipatively heated by the acoustic wave energy normally lost anyway at the chip ends. The convective heat loss increases with the flow rate and can be directly detected as velocity change. Obviously, substrates with large temperature coefficients such as $LiNbO_3$ and small devices having a small thermal mass must be used to obtain sensitive and fast responses [66]. The sensitivity also increases with heater power. The most common problems encountered are associated with the device mounting, which should guarantee a good thermal isolation, a precisely dimensioned flow channel, and a minimum perturbation of the flow, and with the large cross-sensitivity to temperature or heater power fluctuations [70]. In the case of liquid flow sensors, the measurement is also influenced by the thermal conductivity, the dynamic viscosity, and the specific heat of the fluid [68, 69].

As already mentioned, acoustic devices respond to an electric field because it both strains the piezoelectric substrate and modifies the elastic material parameters. The phenomenon has been studied for Rayleigh waves [71-85] and Lamb waves [86-88]. Voltage sensors are constructed by applying the unknown voltage to two electrodes positioned on or near the substrate surface (Figure 5-7). Their main advantages are the direct voltage-to-frequency conversion and a very large input impedance (ca TΩ) [72, 75]. The field-strength sensitivites of some SAW substrate cuts with a linear behavior have been included in Figure 5-7. Many other SAW substrates have been found to behave in a non-linear fashion. Among them are PZT [71], Y,Z-$LiTaO_3$, Y,Z-$LiNbO_3$ [72-75], and $Bi_{12}GeO_{20}$ [77]. In such cases, one can significantly improve the sensor linearity by a differential configuration [77]. "Layered" Rayleigh waves on ZnO/Si lead to linear sensor responses, but the field-strength sensitivity depends both on the layer thickness and the mode number, a typical value being 10 ppm/(V/μm) [80]. This also holds true for Lamb modes on thin plates. On Y,Z-$LiNbO_3$ the Lamb wave sensitivity is linear with a value of 122 ppm/(V/μm) [86, 87].

The best voltage sensitivities have been obtained with the normal-field configuration. Typical values are 0.18 ppm/V on 128°rotY,X-$LiNbO_3$ [75], 0.93 ppm/V on 16.5° doubly rotated $LiNbO_3$ [76, 78], or 0.83 ppm/V on Si_xN_y/ZnO/SiO_2/Si [83]. Even when the acoustic energy is concentrated in very small apertures with the help of acoustic waveguides, the in-plane sensitivity does not exceed 0.01 ppm/V [81, 84].

Acoustoelectric interactions with liquids enable one to measure the liquid conductivity and permittivity. Obviously, the highest sensitivities are obtained with strongly coupling substrates such as 36°rotY,X-$LiTaO_3$. This material has been studied extensively by Kondoh and coworkers [26, 90-92, 177]. If the conductivity is mainly determined by H^+ or OH^- ions, the sensors can be used to determine the pH value of solutions. This has been used, eg, to study enzyme reactions [177]. Dahint et al. [89] probed strong and weak electrolytes using an SH



Material	Cut	$S^f_{E\perp}$	$S^f_{E\parallel}$	Reference
$LiNbO_3$	Y, X	15		[71]
	37.86°rotY, X	20	120	[82]
	128°rotY, X	82		[72, 75]
		141	6	[76]
	131°rotY, X		4.5	[81, 84]
	16.5° doubly rotated	136	1.4	[76]
Quartz	X, 33.44°rotY	9.5		[79]
$Li_2B_4O_7$ 1)		300		[85]
ZnO 2)		14.9		[83]

1) on poled piezoelectric ceramic; 2) sandwiched between Si_xN_y and SiO_2/Si.

c)

Figure 5-7. SAW voltage sensors. (a) Normal-field configuration in cross-sectional view. (b) Transverse-field configuration. (c) Field strength sensitivity of selected materials with linear behavior to normal and transverse fields in ppm/(V/μm).

plate mode delay line on Z, X-$LiNbO_3$. By recording the alkali metal ion concentration which produces the maximum wave attenuation, they were able to calculate the ion mobility. The latter serves to identify the metal ion. Liew and co-workers [31, 32] used a similar geometry. The liquid conductivity decreased when a glass slide coated with a substance that binds metal ions was inserted into the liquid flow, and this conductivity change was detected by the APM device.

The magnetic-field sensitivity of SAW devices coated with magnetostrictive films has been exploited to design phase shifters, tunable oscillators, phase modulators, and, to a lesser extent (no doubt because of the strong competition of Hall probes [13]), sensors [93–100]. The film overlay consists of a magnetically soft metallic glass such as $Fe_{0.8}B_{0.15}$ [93], $Fe_{0.7}Si_{0.3}$ [94], or $Fe_{0.8}B_{0.15}Si_{0.05}$ [96], or of a magnetically hard rare-earth metal/iron material such as $Fe_{0.5}Tb_{0.5}$ [96], although simple nickel films have also been used [93]. Hietala and Robbins [98, 100] compared many different film–substrate combinations and found that the largest sensitivity is obtained for an Fe-B film on ST,X-quartz with 0.08 ppm/(A/m) for a 1 μm film at 120 MHz. An earlier source reported an even higher sensitivity of 0.38 ppm/(A/m) for a similar configuration [93]. All these sensors generally suffer from large propagation losses introduced by the film overlay (10 dB/mm and more [93, 96]) and from inherent non-linearities due to saturation effects. The sensor response depends strongly on the direction of the magnetic field which may therefore be determined by rotating the sensor until a maximum change in frequency is observed [99].

Temperature cross-sensitivity is a nuisance in almost all SAW devices, so the idea of developing SAW temperature sensors suggests itself. There exist a few quartz cuts with a linear temperature coefficient of delay (TCD) of up to 30 ppm/ °C [34, 37, 101], and thin AlN films on metalized glass substrates have demonstrated a sensitivity of 50 ppm/ °C [104]. However, the most widely used material is $LiNbO_3$ with a TCD of 90 ppm/ °C. Both delay lines [102, 106] and one-port [105] or two-port resonators [103] can be employed. Problems often occur due to slow drift phenomena (aging) and hysteresis effects caused by mounting stresses. Sometimes, information on other physical quantities may be inferred from the temperature response of a SAW device. This has been used to measure indirectly flow rates (see above) or thermal conductivities [151]. In the latter case, a heated SAW device is placed near a heat sink and put into a gas stream. The temperature decrease at the substrate surface is a measure for the thermal conductivity of the gas.

Although SAW devices do interact with light, for instance via optically induced electrical conductivity changes [23], the principle apparently has not been exploited for practical detectors. This is attributed to the strong competition of semiconductor sensors [13]. SAW quartz resonators also show some sensitivity to ionizing radiation which depends both on the surface treatment and on the type of radiation [107]. The highest sensitivity is obtained with non-swept, uncoated quartz. It is proportional to some power of the total dose and reaches values of 0.48 $ppm/(J/kg)^{0.5}$ for γ-rays, 0.035 $ppm/(J/kg)^{0.7}$ for β-rays, and 0.056 $ppm/(J/kg)^{0.7}$ for proton radiation.

Since the mid-1980s, SAW ID tags, which may be viewed as remotely read position sensors, have been marketed for passive identification applications [109, 110]. Bao et al. [108] developed a remotely read temperature sensor using a two-reflector ID tag on Y,Z-$LiNbO_3$. An interrogation unit sent a pulse-modulated FM signal to the tag and evaluated the phase difference between the two signals reflected by the two reflectors. This phase difference depends on the SAW velocity and, therefore, on the chip temperature. Between 20 and 140 °C, a phase shift dependence on temperature of 2.9°/ °C and a resolution of 0.35 °C was achieved. The concept is intriguing as it allows the contactless sensing (not only of temperature, but also of other physical quantities) in hazardous or inaccessible environments or on moving machine parts.

The sensitivity of acoustic waves to small surface perturbations suggests their application to the monitoring of physical properties of thin films. Different researchers used surface waves to detect the glass transition temperature of polymers [113, 126], follow the etching of a photoresist film in an oxygen plasma [111], determine the diffusion rate of gas molecules into polyimide and the porosity of silicate film overlays [112], elucidate the nature of polymer-water interactions [114], investigate quantum effects in a 2D electron gas on a semiconducting heterostructure [115], monitor the curing state, adhesion strength, and surface state of polyimide overlays [116, 121], measure the thin film conductivity of a nickel film [117], determine the elastic properties of thin metal films [119], and monitor in real time the selfassembly of organized alkanethiol monolayer films [120]. In the same sense, acoustic plate modes helped to measure the diffusion rate of salts into different gels [122], completely determine the viscoelastic properties of various polymers [123], detect liquid-to-solid phase transitions (freezing) [124], and study the wavelength-dependent photopolymerization of a photoresist [125].

Table 5-4 surveys the performance of selected physical sensors which have shown a first-rate sensitivity in their particular group.

Table 5-4 Performance of selected physical sensors.

Measurand	Device[a]	Frequency (MHz)	Substrate	Sensitivity[b] Value	Unit	Reference
Pressure	DL	105	Quartz	3.8	ppm/kPa	[34, 37]
	DL	90	AlN/Si	27	ppm/kPa	[42][e]
Force	DL	8.3	$LiNbO_3$	10.8	ppm/kN	[43]
Strain	R	140.2	Quartz	1.28	ppm/10^{-6}	[38]
	DL	10.9	PZT	21	ppm/10^{-6}	[57]
Position (linear)	DL	8.3	$LiNbO_3$	120.5	ppm/μm	[43]
Position (angular)	R	434	Quartz	2.86	ppm/mrad	[45]
Acceleration	DL	251	Quartz	45	ppm/(m/s^2)	[52]
	DL	10.9	PZT	8.7	ppm/(m/s^2)	[57]
Rotation rate[c]	DL	10.9	PZT	25.7	ppm/s^{-2}	[57]
Flow rate (gas)	DL	73	$LiNbO_3$	204	ppm/(cm^3/s)	[70]
Flow rate (liquid)	DL	68	$LiNbO_3$	105	ppm/(mm^3/s)	[68, 69]
Liquid viscosity	DL	30	$LiNbO_3$	2.7	ppm/cP	[59]
Liquid density	DL	6	ZnO/Si_xN_y	30000	ppm/(g/cm^3)	[28][f]
Electric field (normal)	DL	900	$LiNbO_3$	141	ppm/(V/μm)	[78]
	R	85	$Li_2B_4O_7$ on piezoceramic	300	ppm/(V/μm)	[85]
Electric field (transverse)	DL	1000	$LiNbO_3$	120	ppm/(V/μm)	[82]
Voltage	DL	900	$LiNbO_3$	0.93	ppm/V	[76, 78]
Liquid conductivity	DL	51	$LiTaO_3$	13400	ppm/(S/m)	[90]
Magnetic field	DL	140	Fe-B/quartz	0.38	ppm/(A/m)	[93]
Temperature	DL	43	$LiNbO_3$	92.13	ppm/°C	[102]
Radiation dose[d]	R	199	Quartz	0.48	ppm/$(J/kg)^{0.5}$	[107]
Thin-film thickness	DL	75	$LiNbO_3$	9.25	ppm/nm	[111]

[a] R = resonator; DL = delay line.
[b] Fractional frequency change of sensor oscillator per unit change in measurand.
[c] Quadratic relationship between frequency and measurand.
[d] Square-root relationship between frequency and measurand.
[e] Lamb wave device.
[f] SH plate mode device

5.1.7 Chemosensors

Chemosensors are obtained by coating a SAW device with a substance that selectively binds the measurand to be detected. The first researchers to investigate such sensors were Wohltjen and Dessy in 1979 [126]. The subsequent development until about 1989 has been thoroughly reviewed by Nieuwenhuizen and Venema in Volume 2 of this series on sensors [15]. This subsection therefore concentrates on the progress that has been made since then.

As evidenced by the data compiled in Table 5-5, which gives a fairly comprehensive overview of recently reported SAW sensors for inorganic gases, no fundamentally new results have

been published in this area since Nieuwenhuizen and Venema's review [15]. As in the past, humidity sensors exploit the hygroscopic properties of polymer films and most nitrogen dioxide sensors work with (metallo)phthalocyanines. Vetelino and coworkers continued their work on tungsten trioxide coatings for H_2S sensing. They discussed ways to optimize the sensor response by choosing an appropriate layer thickness and conductivity [25], by doping the metal oxide [132], or by subjecting it to an activation process (heat treatment) [135]. Watson et al. [151] inferred the CO_2 concentration in a helium carrier gas from the thermal conductivity measured with an uncoated resonator.

Table 5-5. SAW gas sensors for inorganic gases. All devices are delay lines except in [151], where a resonator was used.

Measurand	Frequency (MHz)	Temperature (°C)	Substrate	Chemical interface[a]	Sensitivity[b]	Reference
H_2O[c]	74.3	–	$LiNbO_3$	PI	14.7	[127]
	40	30	Quartz	PEI	1.04	[136]
	234	–	Quartz	PPA	0.21	[153]
	74	–	$LiNbO_3$	NaPss polyelectrolyte	40.54	[106]
	30	–	$LiNbO_3$	PS sodium sulfonate	507.6	[161]
CO_2	40	30–70	Quartz	PEI	–	[136]
	342	50/150[d]	$LiNbO_3$	None	0.00058	[151]
H_2S	60	130–180	–	WO_3	5	[25]
	–	250	Quartz	Doped WO_3	1.7	[132]
	41	200	Quartz	Activated WO_3	–	[135]
NO_2	98.6	22	Quartz	$ttbSiPcCl_2$	0.10	[131]
	600	40–120	Quartz	Fe-Pc	9.5	[144]
	70	150	$ZnO/SiO_2/Si$	Cu-Pc	13.14	[149]
	100	120	–	Cu-Pc	0.2	[162]

[a] Abbreviations: NaPss = a polyelectrolyte thin film; Pc = phthalocyanine; PEI = polyethyleneimine; PI = polyimide; PPA = polyphenylacetylene; PS = polystyrene; $ttbSiPcCl_2$ = tetra-4-*tert*-butylsiliconphthalocyanine-dichloride.
[b] Fractional frequency shift of sensor oscillator (eg, in ppm) per volume concentration of measurand (also in ppm).
[c] The sensitivity is expressed in ppm/% RH (relative humidity).
[d] Gas/sensor temperature.

Sensors for organic vapors have been the subject of many publications. More and more, the simple polymer films of the pioneering days are being supplemented by specialized materials leading to a fast and sensitive response. Dickert and co-workers [148, 154–158] have been particularly active in this respect. They showed the usefulness of anisotropic polymer phases containing hydrophobe molecule tails that bind long-chained volatile hydrocarbons and of materials such as carbenium ions that use donor-acceptor interactions to detect gas molecules by their functionality. Perhaps the most intriguing possibility, which can even differentiate between isomers, is molecule detection by morphology. This is accomplished with host-guest reactions, ie, the specific inclusion of gas molecules in cavities. The cavity can be part of a molecule (paracyclophanes, which respond to toxic substances like benzene, chloroform, or

tetrachloroethene [154, 156], cyclodextrins [155, 156], calix[*n*]arenes [157, 158], or cyclotriphospazene [155]), or it is formed by the space between molecules (clathrates such as triphenylsilanol) [156].

The film adhesion can be promoted by coating the quartz surface with trimethylsilane, $Si(CH_3)_3$ [156]. As an additional advantage, the procedure hydrophobizes the surface, thus reducing the water cross-sensitivity. The silylated surface itself exhibits a remarkable (cross-)sensitivity to many organic substances such as alkanes, cycloalkanes, ethers, acetone, benzene, chlorinated hydrocarbons, and heterocyclic compounds such as tetrahydrofuran. The sensitivity of the frequency to these analytes is typically 0.5 ppb per ppm volume concentration. The largest value, 3.8 ppb frequency shift per ppm volume concentration, has been measured for tetrachloroethene [156].

Table 5-6 contains detailed information on the characteristics of recently reported SAW sensors for organic vapors. The data have been classified according to the analyte detected and

Table 5-6. SAW gas sensors for organic vapors. The substrate is quartz for all devices.

Measurand[a)]	Device	Frequency (MHz)	Temperature (°C)	Chemical interface[a)]	Sensitivity[b)]	Reference
Pentane	–	7	–	PIB	–	[159]
	DL	25-200	25	PB/PS copolymer		[164]
Heptane	R	434	–	A silane	0.0016	[148]
	R	434	–	Urea	0.00069	[156]
Octane	DL	158	40	PDMS	0.0089	[163]
Methanol	R	310	–	Asolectin	0.0032	[134]
	R	310	Room	PE	0.0026	[139]
	R	310	Room	PC	–	[140, 141]
	DL	158	40	Nafion	0.0025	[163]
	DL	71.6	30	HBMC	0.60	[146]
Ethanol	R	310	–	Asolectin	0.00065	[134]
	R	310	Room	PE	0.0048	[139]
	R	310	Room	PC	–	[140, 141]
	DL	71.6	30	HBMC	0.89	[146]
Propan-1-ol	R	310	–	Asolectin	0.0019	[134]
	R	310	Room	PE	0.019	[139]
	R	310	Room	PC	–	[140, 141]
Propan-2-ol	DL	25-200	25	PB/PS copolymer	–	[164]
Butan-1-ol	R	310	–	Asolectin	0.0048	[134]
	R	310	–	PE	0.0068	[134]
	R	310	Room	PE	0.047	[139]
	R	310	Room	PC	–	[140, 141]
Acetone	DL	71.6	30	HBMC	0.58	[146]
	DL	71.6	30	PECH	0.54	[147]
Acetoin	R	310	Room	PC	0.013	[140, 141]
Acetic acid	DL	98.6	22	ω-Tricosenoic acid	0.0065	[131]
DMMP	DL	98	–	MUA with Cu^{2+}	0.15	[120]
DIMP	DL	98	–	MUA with Cu^{2+}	0.25	[120]
Benzene	DL	434	–	A paracylophane	–	[154]

Table 5-6. (continued)

Measurand[a)]	Device	Frequency (MHz)	Temperature (°C)	Chemical interface[a)]	Sensitivity[b)]	Reference
Toluene	DL	71.6	–	GC phases	0.0056	[138]
	DL	71.6	30	Ethylene-vinyl acetate copolymer	7.22	[146]
	DL	71.6	30	PECH	4.6	[147]
	DL	434	–	A paracylophane	–	[154]
	–	7	–	PIB	–	[159]
TNT	R	500	175/14.3[c)]	None	12.9[d)]	[145, 152]
Dichloro-methane	DL	71.6	30	Polycarbonate resin	0.98	[146]
	DL	71.6	40	Carbowax 550	1	[147]
Trichloro-ethene	DL	25–200	25	PB-PS copolymer	–	[164]
	DL	97	–	PIB	0.038	[142]
	DL	97	–	Polysiloxane	0.11	[143]
	–	7	–	PIB	–	[159]
Tetrachloro-ethene	R	434	–	β-Cyclodextrin	0.0035	[148]
	R	434	–	A paracylophane	0.014	[154]
	R	434	–	β-Cyclodextrin	0.0046	[155]
	R	434	–	A paracylophane	0.010	[156]
	R	434	–	Silylated calix[8]arene	0.14	[157]
	R	434	–	Calix[4]resorcinarene	0.0092	[158]
	–	7	–	PIB	–	[159]
Citral	R	310	–	Asolectin	0.11	[134]
	R	310	Room	PE	45.16	[139]
β-Ionone	R	310	–	Asolectin	0.15	[134]
Menthone	R	310	–	Asolectin	0.039	[134]
	R	310	Room	PE	3.55	[139]
	R	310	Room	PC	0.21	[140, 141]
	R	310	Room	PE	0.11	[140, 141]
	R	310	Room	Stearic acid	0.07	[140, 141]
Tetrahydro-furan	R	434	–	Triphenylmethyl	0.0023	[148]
Cocaine	R	925	120/room[c)]	None	21.62[d)]	[137]
	R	500	175/14.3[c)]	None	74.6[d)]	[145, 152]
Heroin	R	925	120/room[c)]	None	21.62[d)]	[137]
Amylacetate	R	310	–	Asolectin	0.0084	[134]
	R	310	Room	PE	0.65	[139]
	R	310	Room	PC	0.02	[140, 141]

a) Abbreviations: DIMP = diisopropyl methylphosphonate; DMMP = dimethyl methylphosphonate; HBMC = hydroxybutylmethylcellulose; MUA = mercaptoundecanoic acid; PB = polybutadiene; PC = phosphatidylcholine; PDMS = polydimethylsiloxane; PE = phosphatidylethanolamine; PECH = polyepichlorohydrin; PEI = polyethyleneimine; PIB = polyisobutylene; PS = polystyrene; $ttbSiPcCl_2$ = tetra-4-*tert*-butylsiliconphthalocyanine dichloride; TNT = trinitrotoluene.

b) Fractional frequency shift of sensor oscillator (eg, in ppm) per volume concentration of measurand (also in ppm).

c) Gas/sensor temperature.

d) ppm/ng (relative frequency shift of sensor per unit mass of sorbed material).

appear in the following order: alkanes (pentane to octane), alcohols (methanol to butanol), ketones (acetone), hydroxyketones (acetoin), carboxylic acids (acetic acid), organic phosphorus compounds (DIMP, DMMP), aromatic compounds (benzene to TNT), chlorinated hydrocarbons (dichloromethane to tetrachloroethene), terpenes (citral to menthone), heterocyclic compounds (tetrahydrofuran to heroin), and other (amyl acetate).

Fluid sensors are of particular interest in the area of biochemical sensing. Surface waves with a (quasi-)shear polarization [91, 92, 169-178], Lamb waves [19], SH plate modes [21, 31, 32, 89, 179-186], and Love waves [187-192] have all been investigated with respect to their suitability for chemosensing applications in liquid environments. Most sensors reported to date limit themselves to the detection of the presence of a given substance without quantifying the concentration. Several researchers, eg, Kondoh and co-workers [92, 177], did not employ chemosensitive coatings at all, but measured a physical quantity such as conductivity and calculated the analyte concentration from this information. Table 5-7 reviews the state-of-the-art performance of sensors for ions and biological substances.

Table 5-7. SAW fluid sensors. All devices are delay lines.

Measurand[a)]	Frequency (MHz)	Substrate	Chemical interface[a)]	Sensitivity[b)] Value	Unit	Reference
Cu^{2+}	–	Quartz	Ethylenediamine	52000	ppm/mol	[21, 179][c)]
Fe^{3+}	13	$LiNbO_3$	X	130	ppm/vol. ppm	[31, 32][c)]
K^+	13	Quartz	Y	850	ppm/pK^+	[187, 188][d)]
IgG	44	$LiNbO_3$	Goat anti-IgG	30	ppm/(μg/cm^3)	[180, 183, 186][c)]
	252	Quartz	IgG antibodies	120	ppm/(nmol/cm^3)	[173]
BSA	5.85	ZnO/Si_xN_y	None	0.11	ppm/(μg/cm^3)	[19][e)]
dA	13	$LiNbO_3$	dT	1840	ppm/μg	[181, 182, 184][c)]
CMV	50	$LiNbO_3$	Unspecif. DNA	75	ppm/μg	[185][c)]
Urea	51	$LiTaO_3$	None	1.5	ppm/wt. ppm	[91]
	51	$LiTaO_3$	Urease	0.086	ppm/$mol^{2/3}$	[92]
Fruit juices	30-100	$LiTaO_3$	None	–	–	[177]
Glucose oxidase	>100	$LiTaO_3$	Anti-glucose oxidase	<0.002	ppm/(μg/cm^3)	[178]

a) Abbreviations: BSA = bovine serum albumin; CMV = cytomegalovirus; dA = deoxyadenylic acid; dT = deoxythymidylic acid; IgG = human immunoglobulin-G; X = *N*-(3-trimethoxysilylpropyl)ethylenediamine; Y = mixture of PVC, dibutyl sebacate, tetrahydrofuran, and valinomycin.

b) Fractional frequency shift of sensor oscillator per some measure for the amount of analyte. This amount is specified as mass, volume, or molar concentration, or as the absolute value injected into a liquid cell.

c) SH plate mode device.

d) Love wave device.

e) Lamb wave device.

5.1.8 Conclusions and Outlook

The first physical sensors based on SAW devices were reported about 20 years ago, and SAW chemosensors have been studied for the past 15 years. Hence, SAW sensors cannot be regarded as a very young technology. Nevertheless, and this holds especially true for chemosensors,

there is still a large amount of active research in the field. Truly state-of-the-art sensors that can be marketed commercially will only be obtained if the common emphasis on selected aspects of the sensor design (mostly the chemical interface) gives way to an interdisciplinary approach paying equal attention to the SAW device, the electronic circuitry, and the chemical interface. Several current trends point into this direction. First, the exclusive use of mass-sensitive Rayleigh wave devices is being increasingly replaced by a judicious choice of wave type and sensing mechanism that takes into account the particular task at hand. Second, the SAW device design is becoming more sophisticated (multiple-frequency devices [164, 177], dual-output sensors [22, 133, 142, 143, 159, 164], use of directional transducers to reduce the insertion loss [183, 185], etc.). Third, specialized chemosensitive materials which are both sensitive and selective are being developed. The integration of the sensor element with the necessary electronic circuitry has been demonstrated successfully [167]. Finally, pattern recognition techniques in connection with sensor arrays allow the identification of analytes even with moderately selective individual sensors [163, 165, 177].

On the other hand, many severe problems have not yet been tackled: the long-term stability of SAW sensor systems must be improved, the resistance to aggressive environmental influences is still not good enough, and the cross-sensitivity to interfering quantities has to be minimized. These and other problems can frequently be neglected in tightly controlled laboratory experiments, but become very important when a sensor is to operate under real-life conditions where, eg, temperature and gas composition are not known a priori. It should be noted that the mounting and packaging of the SAW device play a crucial role in solving many of the difficulties.

In spite of the efforts that have been spent to develop SAW sensors in the past, there is still an untapped potential for improving the performance. The sensitivity can be further enhanced by operating at higher frequencies (up to 2 GHz) or using chirped transducers [193], and remote sensing systems promise to open up new markets [194]. A careful design that pays attention to both the SAW device and the electronic instrumentation should lead to a sensor system with a sensitivity, noise behavior, and stability much better than achieved previously.

5.2 Quartz Microbalance Transducers

5.2.1 Introduction

The use of a mechanical resonance for the detection of mechanical properties has a long tradition (eg, [195]). One of the early applications was the monitoring of the thickness of evaporated metal layers by simultaneous application of the layer to a quartz crystal and measuring the corresponding change in the resonance frequency. This frequency change is proportional to the applied mass of the layer.

Since frequencies can be measured with very high accuracy (eg, 1 part in 10^8), very small changes in the mass can be detected. The adsorption (or absorption) of a gas or a vapor in a layer on top of a mechanical resonator can thus give significant changes in the resonant frequency. This effect is used in quartz microbalance (QMB) chemical sensors [196].

Figure 5-8 shows a schematic diagram of such a chemical sensor. A properly cut quartz crystal is covered with two metal electrodes. Owing to the piezoelectric properties of quartz (eg., [197]), an applied electric voltage produces a deformation of the crystal. The mechanical resonance frequency f_0 of this mode of oscillation (thickness shear mode) is given by the mechanical properties of the crystal (phase velocity of the acoustic wave c) and its thickness d_Q (neglecting the influence of the metal electrodes) [195]:

$$f_0 = \frac{c}{2\, d_Q} \,. \tag{5-14}$$

A thickness of $d_Q = 168$ µm of an AT-cut quartz slice results, for instance, in a resonance frequency of 10 MHz. The "T" in AT-cut quartz means that this cut has been chosen so that the temperature dependence vanishes for a given temperature (usually for 25 °C). The "A" has historical reasons.

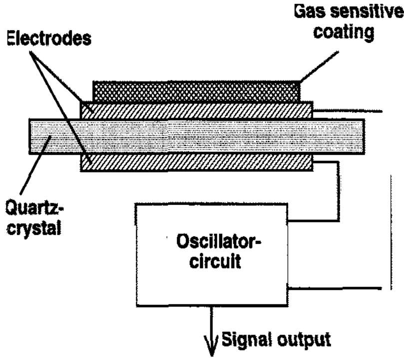


Figure 5-8.
Schematic diagram of a QMB gas sensor system.

The metal electrodes and a coating applied for the absorption of gases reduce this resonant frequency (typically by 0.1–1%). An additional change in the mass Δm (for instance by gas absorption) will result in a frequency change Δf:

$$\frac{\Delta f}{f_0} = -\frac{\Delta m}{\rho_Q\, d_Q\, A} = k \frac{\Delta m}{A} \tag{5-15}$$

where ρ_Q is the density of the quartz and A the area covered by the metal electrodes. A typical value for k is $-2.2 \cdot 10^8$ cm^2 g^{-1}, so that, assuming a frequency resolution of 1 Hz and an operating frequency of 10 MHz, a specific change of mass of about 5 ng cm^{-2} can be detected. For this consideration a thin and rigid coating has been assumed (see Section 5.2.4).

The direct application of a force to the crystal will also alter the resonance frequency, so that sensors for force, pressure, torque, and acceleration can be made with quartz crystals [197].

Generally, all these effects depend on the temperature, so that also the temperature can be measured with these sensors. In connection with chemosensors, this is done for compensation purposes. Finally, a quartz crystal can be used as transmitter or receiver for ultrasonic waves.

Table 5-8 shows a survey of the applications of quartz sensors together with the proper crystal cuts and vibrational modes [198].

In the following sections, applications as chemical sensors will be emphasized.

Table 5-8. Potential sensor applications for different quartz crystal cuts (see also Figure 6-9).

Crystal cuts / Vibrational modes / Applications	*X* longitudinal	*X* transversal	*Y* shear	Flexural vibration	AT, BT thickness shear	AC, LC thickness shear	ST surface wave
Ultrasonic transmitter/ receiver	×	×	×				
Temperature sensor				×		×	×
Sensors for mechanical properties: force	×	×	×	×			×
Sensors for mechanical properties: pressure	×	×		×			×
Sensors for mechanical properties: torque			×				×
Sensors for mechanical properties: acceleration	×	×					
Thickness monitoring					×		×
Chemical sensors: gases					×		×
Chemical sensors: vapours					×		×
Chemical sensors: liquids					×		
Chemical sensors: particles					×		

5.2.2 Quartz Substrate

Figure 5-9 shows the different cuts with respect to the crystal orientation. For many applications, especially for chemosensors, the temperature dependence of the resonance frequency should be as low as possible. A preferred cut in this respect is the AT-cut. Figure 5-10 shows that for this cut a constant resonance frequency is possible within a limited temperature range and that the cutting angle must be equal to 35 ° 10′ with very high accuracy. Even a small devia-

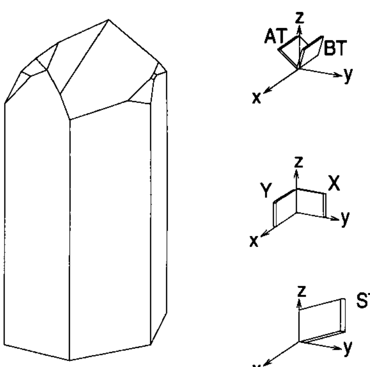


Figure 5-9.
Different crystal cuts for quartz (SiO_2).

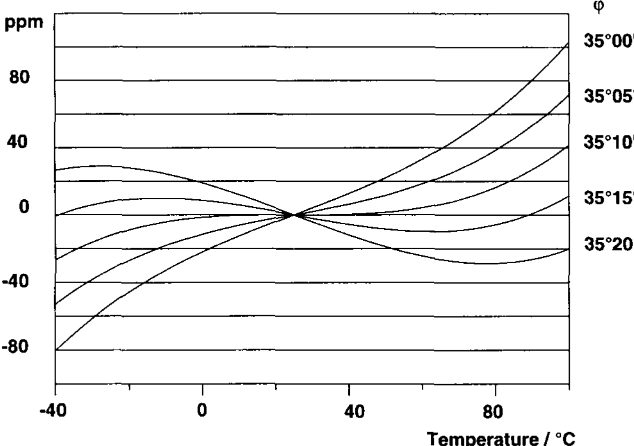


Figure 5-10. Relative change in parts per million (ppm) of the resonance frequency of an AT-cut quartz device with a change of temperature. Parameter: cutting angle φ.

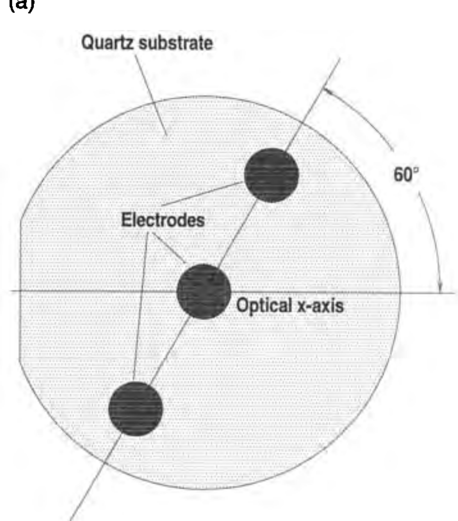
tion of this value will spoil this insensitivity to temperature. Assuming a typical fabrication tolerance of 10 min, this leads to a relative frequency change of 10^{-6} K^{-1} at 25 °C. If a pair of sensors from the same cutting process is used, the frequency fluctuations of the two sensors due to temperature fluctuations will be nearly the same (assuming good thermal coupling), so that the fluctuation of the difference frequency will be greatly reduced, for instance by one order of magnitude. The two resonators are slightly detuned by different thicknesses of the coatings (metal electrode and/or chemical selective coating) so that a finite frequency difference is obtained.

If for the purpose of pattern recognition (eg, [199]) an array of sensors is required, the different mechanical resonators must oscillate independently of each other, but the thermal coupling must be maintained. This can be achieved by arranging the active resonator parts along certain angles of low acoustic coupling on the same quartz substrate. To be more specific, the acoustic coupling must be small enough to bring the different oscillators out of their pulling range. Figure 5-11a shows a proper arrangement for a sensor array. Figure 5-11b shows the acoustic coupling as a function of the angle between coupling direction and optical axis [198].

The metal electrodes can be applied to the quartz crystal either by evaporation through a mask (Figure 5-12a) or by evaporation and a subsequent lift off process (Figure 5-12b).

5.2.3 Active Coating

A so-called active coating has to be applied generally to the quartz crystal to provide for an interaction with the gas or vapor to be detected. This coating has the same purpose as the coating on SAW sensors, that is, the more or less selective absorption or adsorption of substances so that the mass loading is increased and the resonance frequency is decreased. In contrast to SAW sensors, the electrical properties of the coating, and also their change by the substances to be measured, have no influence.



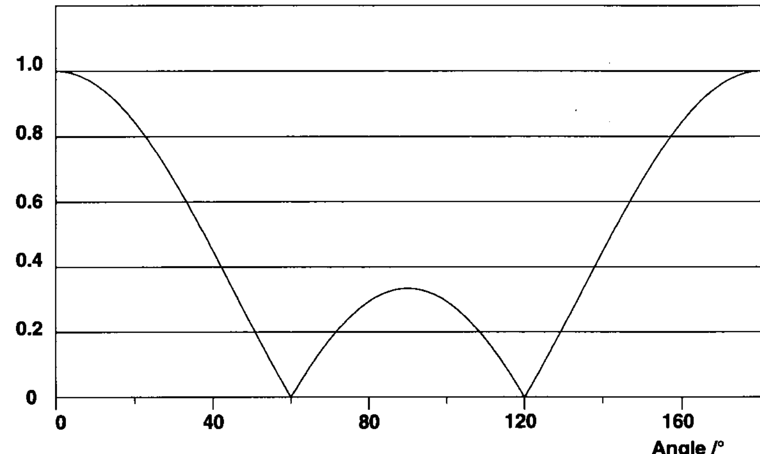


Figure 5-11. (a) Arrangement of the electrodes on an AT-cut quartz substrate for minimum acoustic coupling between the three resonators. (b) Relative acoustic coupling between two resonators located on a line as a function of the angle between the line and the optical *x*-axis.

If the coating is thin compared with the acoustic wavelength (see Section 5.2.4), it can be considered as a rigid mass. For the detection of gases or vapors these assumptions hold fairly well. In liquids a wave will be excited and only if this wave is evanescent perpendicular to the crystal surface is the mechanism of signal generation in principle the same as in gas sensors. The acoustic properties of the liquid, such as density and viscosity, however, will influence both the resonance frequency and the *Q*-factor of the resonator. Depending on the circuitry used, conductance of the liquid may influence damping and resonance frequency in the conventional way.

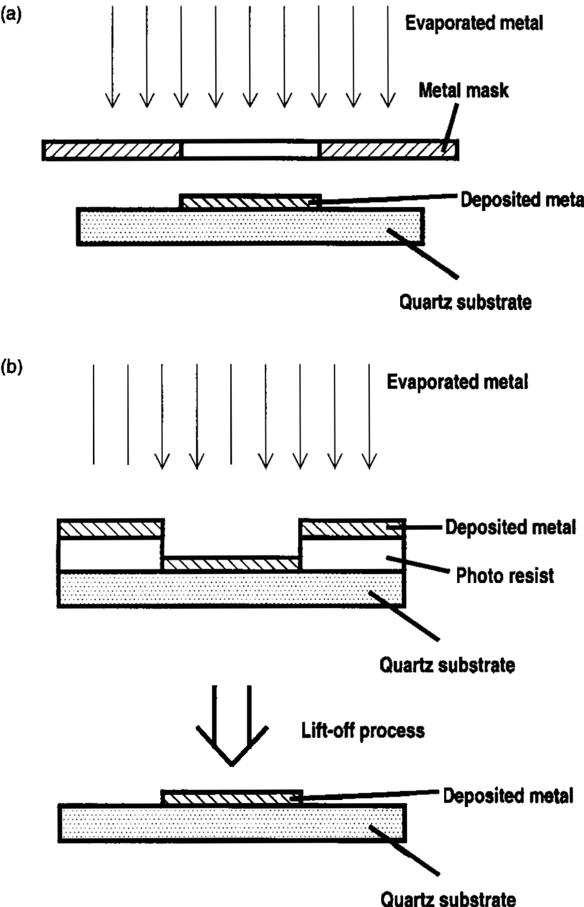


Figure 5-12. (a) Metallization by evaporation of the electrode metal through a metal mask. (b) Metallization by evaporation of the electrode metal and a subsequent lift-off process.

In rare cases, the adsorption of molecules on the electrode surface itself may result in changes of the resonant frequency, but this does not seem to be an effective way to make chemosensors.

The active coatings can be classified into three different groups:

1. GC coatings.
2. Selective biological coatings.
3. Selective host–guest coatings.

1. Coatings as used in gas chromatography (GC)

Since the purpose of the coating is both the high absorption of a gas present and the release of this substance when the gas vanishes, GC coatings which have the same purpose are well

suited as active coatings for QMB sensors. Usually they are chemically stable and they can be used to detect a great number of organic molecules [200]. They have the additional advantage that the sensitivity of the sensor can be estimated from the retention time, which is available for a wide variety of coatings and substances. An example for such a calculation is given in Ref. 201.

The disadvantage of these coatings is their poor selectivity, which either restricts the applications to cases in which the identification of the substance is of secondary importance or which demands a sensor array with an electronic selectivity [202]; an example of this technique is given in Section 5.2.9.

2. *Selective biological coatings*

These coatings are highly selective because of their gene-antigene reaction. They can be used in liquids [203].

3. *Host-guest coatings*

With these coatings, high selectivity is achieved by stereochemical fitting of the analyte into a "cavity" within the coating [204]. This "fitting" is more a fitting of the local distribution of the binding forces than a geometrical fitting and the agreement between corresponding model calculations and experimental results is fairly good [205]. The selectivity is not perfect and the mechanism is similar to that in the nose [206].

The technique for the application of the coatings will be described with respect to the GC coatings. Two techniques have proved to be simple and efficient [207]: airbrush coating and ultrasonic evaporation.

Figure 5-13 shows a cross section through a conventional airbrush as used for painting. The coating substance is dissolved in a suitable solvent (eg chloroform [201]) and sucked into a jet stream of pressurized air. Typically drops of 50–100 μm are blown on to the target with a velocity around 1 ms^{-1}. Local coating through a mask is possible. Owing to the inhomogeneous distribution of the drop size over the cross section of the beam, the reproducibi-

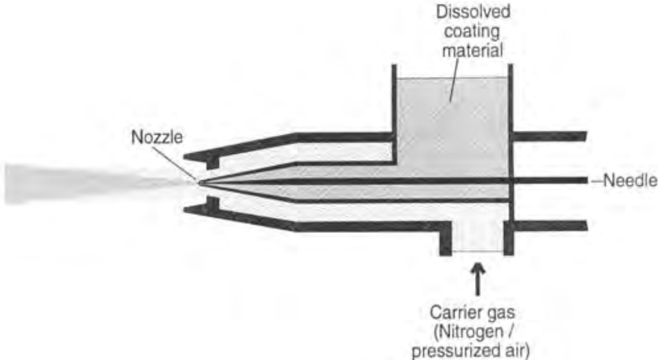


Figure 5-13. Principle of operation of an airbrush.

lity is poor. After evaporation of the solvent, the frequency shift of the quartz resonator can be measured and usually with a second, shorter coating the desired thickness can roughly be reached. Typical growth rates are around 0.5 μms^{-1}. Concerning the molecular mass of the coating, there is no restriction.

The second technique is based on an ultrasonic evaporator as used in medical laboratories. Figure 5-14 shows a schematic diagram of a typical set-up. A piezoelectric ceramic generates ultrasonic waves in the solvent liquid, which causes the build-up of an aerosol containing the solvent with the coating to be applied. The typical diameter of the aerosol drops is 1 µm. With this technique, very homogeneous coatings can be applied with growth rates of about 0.05 μms^{-1}. This technique is applicable for coating molecular masses up to 20000 u.

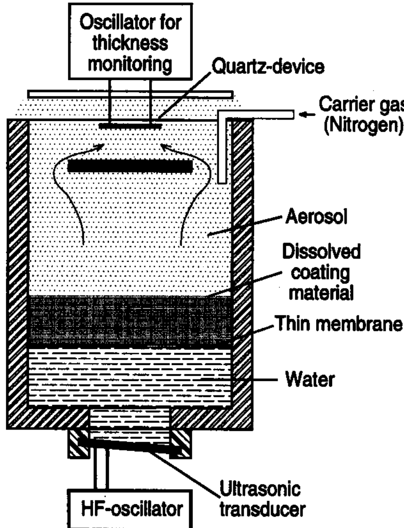


Figure 5-14.
Principle of operation of an ultrasonic evaporator for coating QMB gas sensors.

5.2.4 Acoustic Equivalent Circuit

The quartz device is in mechanical resonance during operation. Owing to losses in the bulk, the crystal's quality factor, Q, defined as $2\pi \cdot$ (stored energy) divided by dissipated energy, during a period, is finite. Typical values for Q are in the range 10000–30000 for uncoated crystals (see also Section 5.2.5).

If the quartz crystal is used for measurements in gases, the resonance frequency of the crystal is determined by the crystal and the coating only (provided that the damping has a negligible influence); the gas itself has virtually no direct influence. The indirect influence comes via the coating and constricts the desired measurement effect.

In liquids, however, an appreciable amount of energy is coupled into the liquid. In this case it is of advantage to consider the quartz crystal as a half-wavelength resonator (in the fundamental mode) [208].

This transmission line concept implies the assumption that the amplitude of the acoustic oscillation is independent of the locus on the crystal surface, so that a one-dimensional treat-

ment is allowed. The acoustic impedance of the coating and the liquid gives the loading of the crystal.

The acoustic wave impedance Z is defined as the product of the velocity of sound c and the density of a material (eg [209]):

$$Z = \rho c \,. \tag{5-16}$$

The "inductance" per unit length L' is equal to the density ρ and the "capacitance" per unit length C' is equal to $1/E$ [210]. The sound velocity (phase velocity) c in solids is given by

$$c = \sqrt{\frac{E}{\rho}} \,. \tag{5-17}$$

In liquids the coefficient of compression $K = 1/\chi$ has to be taken instead of the coefficient of elasticity E (χ = compressibility). Typical values for the material constants are given in Table 5-9. For a homogeneous solid medium with density ρ and coefficient of elasticity E, the impedance is given by

$$Z = \sqrt{\frac{L'}{C'}} = \sqrt{\rho E} \,. \tag{5-18}$$

Table 5-9. Some elastic and electric properties of relevant materials

	Density ρ [kg/m^3]	Sound velocity c [m/s]	Acoustic impedance Z [Ns/m^3]	Relative permittivity
AT-cut quartz	2650	1670	$4.43 \cdot 10^6$	4.5
Air (20 °C)	1.21	340	411	1
Silicon rubber (GC-coating)	1010	1030	$1.04 \cdot 10^6$	2.5

The unloaded quartz crystal in the thickness shear mode has an acoustic impedance Z_Q:

$$Z_Q = \sqrt{C_{66}\,\rho_Q} \tag{5-19}$$

where C_{66} is the corresponding elastic modulus and ρ_Q the density of the quartz. The phase velocity of sound is

$$c = \sqrt{\frac{C_{66}}{\rho_Q}} \tag{5-20}$$

and therefore the wavelength λ at an operational frequency f_0 is

$$\lambda = \frac{c}{f_0} = \sqrt{\frac{C_{66}}{\rho_Q}}\,\frac{1}{f_0} \,. \tag{5-21}$$

The resonance condition for a quartz crystal with thickness d_Q in the fundamental mode is

$$d_Q = \frac{\lambda}{2} = \frac{1}{2 f_0} \sqrt{\frac{C_{66}}{\rho_Q}} \,. \tag{5-22}$$

Assuming that the acoustic wavelengths in the electrodes and the coating are about the same as the acoustic wavelength in the quartz, it can be estimated with Equation (5-21) whether the electrodes and the coating of total thickness d_E or d_C, respectively, can be considered as rigid masses:

$$d_E + d_C \ll \frac{\lambda}{4} \,.$$

For a 10 MHz AT-cut quartz crystal the wavelength λ is 336 μm so that $d_E + d_C$ should be less than 8 μm = $\lambda/40$.

With f_0 as the operating frequency, L'_E the equivalent inductance per length of the electrode and ρ_C and ρ_E the density of the coating or the electrode material, respectively, such a thin solid electrode has an acoustic impedance Z_E [208] (see Figure 5-15):

$$Z_E = j\, 2\pi f_0 L'_E\, d_E = j\, 2\pi f_0\, \rho_E d_E \,, \tag{5-23}$$

and the acoustic impedance Z_C of the coating can be described as

$$Z_C = j\, 2\pi f_0\, \rho_C\, d_C \,. \tag{5-24}$$

The wave impedance of air can be taken from Table 5-9 to be 411 N s m^{-3}. Compared with the impedance of the quartz (4.43 · 10^6 N s m^{-3}), this is practically a short circuit (see Figure 5-15, left side).

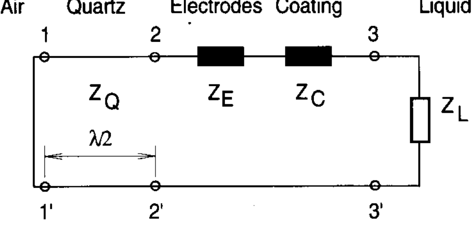


Figure 5-15.
Acoustic equivalent circuit for a QMB device. Both electrodes have been summarized on the right side of the quartz. When operated as gas sensor the impedance Z_L has to be replaced by a short circuit.

If the sensor operates in air, the terminals 3-3′ in Figure 5-15 will also be short-circuited and the loading by the electrodes and the thin coating is Z_E and Z_C. These impedances cause changes Δf_E and Δf_C of the resonant frequency f_0 according to Ref. 202:

$$\frac{\Delta f_E}{f_0} = \frac{\rho_E\, d_E}{\rho_Q\, d_Q} \quad \text{and} \quad \frac{\Delta f_C}{f_0} = \frac{\rho_C\, d_C}{\rho_Q\, d_Q} \,. \tag{5-25}$$

If the viscosity η_L of the liquid is high enough ($\eta_L \gg 1/2\pi f_0 \chi$), the damping will be strong and the wave will be evanescent. In this case the liquid can be assumed to be semi-infinite and described by a complex impedance Z_L:

$$Z_L = (1 + j)\sqrt{\pi f_0 \rho_L \eta_L} \;. \tag{5-26}$$

If the damping is not high enough, the acoustic wave in the liquid will not be evanescent and the sensor signal will depend on the reflections at the surrounding.

The impedance of the coating and the liquid together (assuming high damping) is therefore

$$Z_{22} = j \cdot 2\pi f_0 L'_C d_C + Z_L = \sqrt{\pi f_0 \rho_L \eta_L} + j\left[2\pi f_0 \rho_C d_C + \sqrt{\pi f_0 \rho_L \eta_L}\right] . \tag{5-27}$$

The quality factor Q of the liquid loaded quartz using the above definition is then given as [208] (neglecting the attenuation in the quartz and the coating)

$$Q = \frac{\pi Z_Q}{2\,\mathrm{Re}(Z_L)} = \frac{\pi}{2}\sqrt{\frac{C_{66}\rho_Q}{\pi f_0 \rho_L \eta_L}} \;. \tag{5-28}$$

With the change of the phase angle $\Delta\varphi$ of the impedance of the quartz device:

$$\tan(\Delta\varphi) \approx \Delta\varphi \approx -2\frac{\mathrm{Im}(Z_{22} + Z_E)}{Z_Q} \tag{5-29}$$

the frequency change Δf as concequence of the electrodes, the coating, and the liquid loading can be calculated:

$$\frac{\Delta f}{f_0} = \frac{\Delta\varphi}{2\pi} = -\frac{\mathrm{Im}(Z_E + Z_{22})}{\pi Z_Q} = -\frac{f_0}{\sqrt{C_{66}\rho_Q}}\left[2\rho_E d_E + 2\rho_C d_C + \sqrt{\frac{\rho_L \eta_L}{\pi f_0}}\right] . \tag{5-30}$$

In Ref. 208, examples for the evaluation of the acoustic impedance by measurements are given. For highly polished electrodes the decay length in water was estimated to be about 0.2 μm for a frequency of 9 MHz. Therefore, such acoustic sensors are very sensitive to the roughness of the quartz surface.

Measurements with a blood antibody serum [208] and different blood samples showed that for certain blood samples a decrease in the resonant frequency and the quality factor was obtained according to Equations (5-28) and (5-30). This corresponds to a change in the real part of the acoustic impedance equal to the change in the imaginary part as expected for "pure viscous liquids" ($\eta_L \gg 1/2\pi f_0 \chi$).

5.2.5 Electric Equivalent Circuit

Figure 5-16 shows the electric equivalent circuit. The device behaves as a series-resonant circuit with lumped elements which can be calculated, eg according to Ref. 210. R_Q represents

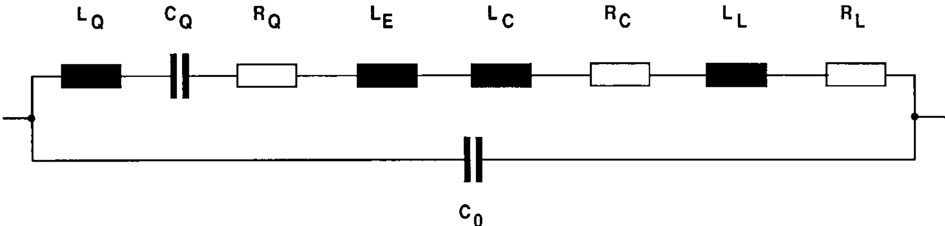


Figure 5-16. Electric equivalent circuit for a QMB device with lumped elements.

the losses in the quartz crystal itself. The dynamic capacity C_Q denotes the crystal's restoring forces:

$$C_Q = \frac{1}{K_v} \frac{1}{2\pi f_0 Z_Q} \tag{5-31}$$

where

$$K_v = \frac{\pi}{32 \varepsilon_Q \varepsilon_0 k_T^2 f_0^2 A \rho_Q} . \tag{5-32}$$

k_T^2 denotes the piezoelectric coupling coefficient (AT-quartz: $k_T^2 = 8.5 \cdot 10^{-3}$), $\varepsilon_Q \varepsilon_0$ the electric permittivity of the quartz and Z_Q is the acoustic impedance of quartz from Equation (5-19).

The dynamic inductance L_Q of the quartz can be calculated according to

$$L_Q = \frac{1}{4\pi^2 f_0^2 C_Q} = K_v \frac{Z_Q}{2\pi f_0} . \tag{5-33}$$

The inductivities L_E, L_C and L_L, representing the electrodes, the coating, and the liquid respectively, can be written as a function of the corresponding acoustic value:

$$L_E = K_v \rho_E d_E \tag{5-34a}$$

$$L_C = K_v \rho_C d_C \tag{5-34b}$$

$$L_L = \frac{K_v}{2} \sqrt{\frac{\rho_L \eta_L}{\pi f_0}} . \tag{5-34c}$$

With the definition of the quality factor in Equation (5-28) R_L, which represents the losses due to the liquid loading in the electric equivalent circuit, can be evaluated:

$$R_L = K_v \sqrt{\pi f_0 \rho_L \eta_L} . \tag{5-35}$$

The electric equivalent of the viscous losses in the quartz crystal R_Q is given by [210]

$$R_Q = K_v \frac{2\pi f_0 \eta_Q \rho_Q}{Z_Q} \tag{5-36}$$

where η_Q denotes the dynamic viscosity of the quartz. The influence of R_Q can be neglected compared with the electric equivalents of the damping in the coating R_C and in the liquid R_L. For the calculation of the losses caused by the coating [210], the coating cannot be considered as a rigid mass but must be considered as a short transmission line, yielding

$$R_C = K_v d_C \cdot 4\pi f_0^2 \frac{\rho_C}{E_C} \eta_C \, . \tag{5-37}$$

The dynamic viscosity of the coating is given by η_C and its elastic modulus by E_C. The value of R_C also is small compared with that of liquids R_L.

The element C_0 parallel to the series resonant circuit is a static, parasitic capacity due to the metal electrodes and the housing of the quartz.

Typical values for a quartz crystal (AT-cut, thickness 168 μm, 10 mm diameter; electrodes: each side 15 nm Cr and 150 nm Au with a diameter of 7 mm) coated with 0.5 μm methyl silicone (OV-101; Macherey – Nagel, Düren, Germany) are as follows:

$$\left.\begin{aligned} R_Q &= 20\ \Omega \\ C_Q &= 30\ \text{fF} \\ L_Q &= 8\ \text{mH} \\ L_E &= 120\ \mu\text{H} \end{aligned}\right\} \text{without coating: } Q = 25\,000$$

$$\left.\begin{aligned} C_0 &= 10\ \text{pF} \\ R_C &= 600\ \Omega \\ L_C &= 16\ \mu\text{H} \end{aligned}\right\} \text{coated: } Q = 1000.$$

Measurements with crystals used as gas sensors give results in good agreement with the above values. For liquid sensors the acoustic impedance of the liquid Z_L has to be taken into account as explained above.

The usual way to detect the change in resonance frequency of a quartz sensor is to use it as frequency-determining device in an oscillator circuit. A quartz sensor can be operated in series resonance f_s:

$$2\pi f_s = \sqrt{\frac{1}{(L_Q + L_E + L_C)C_Q}}$$

or in parallel resonance f_p:

$$2\pi f_p = \sqrt{\frac{1}{(L_Q + L_E + L_C)C_0}} \, .$$

Operation in series resonance is preferred because there is no influence of the stray and parasitic capacitances included in C_0.

5.2.6 Oscillator Circuits

Figure 5-17 shows, as examples of oscillators that can be used, a Pierce [211], a Driscoll, and a bridge circuit [212]. A capacitor in series with the quartz device in order to tune the oscillator frequency can be omitted in sensor applications. In this application the exact value of the oscillator frequency is of minor interest.

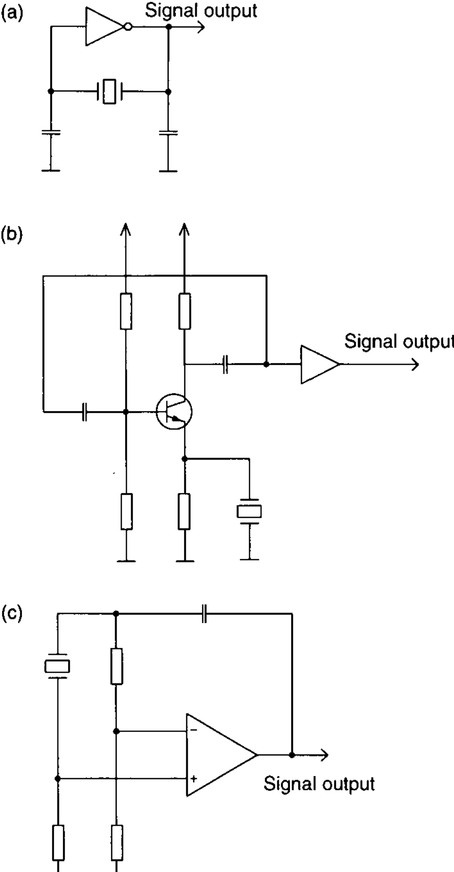


Figure 5-17.
(a) Pierce oscillator circuit. (b) Driscoll oscillator circuit. (c) Bridge oscillator circuit.

The most important requirements for these oscillators are the following:

- The circuit must be able to drive a coated sensor quartz with high losses corresponding to high values of the series resistor $R_Q + R_C(+ R_L)$ up to 100 times the value of a standard quartz device.
- The electrical power dissipated in the quartz crystal should not exceed 20 μW for high long-term stability. Power dissipation beyond 20 mW in an AT-cut quartz device can destroy the sensor [213].

- A circuit which is capable of driving sensors with one electrode grounded is preferred for measurements in liquids and with integrated sensor arrays. When used as a sensor for liquids, the electrode in contact with the liquid has to be grounded, or put on a certain potential with respect to the liquid, in order to avoid corrosion problems and currents due to electrochemical reactions. For applications with integrated sensor arrays, one grounded electrode facilitates the connections and reduces the complexity and costs of the housing.
- The quartz sensor in one arm of a bridge circuit oscillator is recommended when the sensor is used remote from the oscillator electronics. In this case parasitic capacitances and inductances can be compensated with the other elements of the bridge.
- The frequency noise and temperature drift of the oscillator electronics and the whole set-up should be as low as possible because these effects limit the sensor's detection limit and resolution (see Section 5.2.8).

5.2.7 Signal Processing

The information relevant for chemical sensing is the frequency shift as a consequence of the mass increase in the coating. In order to reduce the influence of frequency changes as a consequence of temperature changes, a second uncoated crystal is used and the frequency difference between these two oscillators is used as an improved signal.

Since the selectivity of the coatings generally is not very high, the use of sensor arrays and proper signal processing to provide electronic selectivity is of advantage [214]. Figure 5-18 shows a measurement chamber with two coated resonators and one uncoated resonator for temperature compensation. Thus a two-dimensional signal is obtained from this set-up. The dimensions can be extended to four by placing a second, similar sensor array opposite the first. Proper methods of pattern recognition [215–217] provide transformation rules from the "signal space" to the "gas space", so that (in the best case) n gas concentrations can be evaluated from the n signals of an n-dimensional signal space. In this way the gases applied can be identified and their concentrations measured.

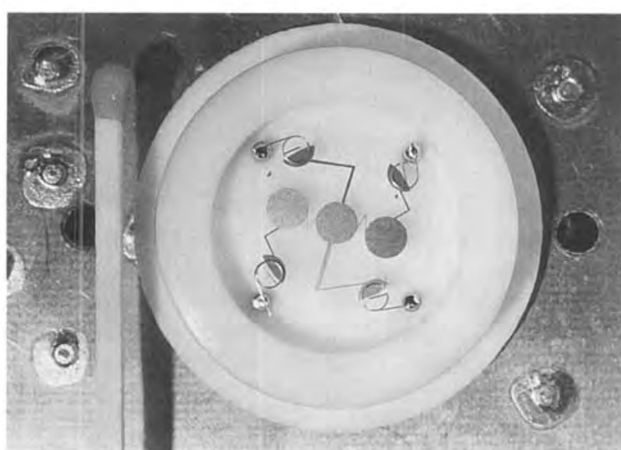


Figure 5-18.
Integrated QMB gas sensor array consisting of two coated quartz resonators and one uncoated reference.

The classical methods of pattern recognition, eg PLS (partial least-squares) [218], give very accurate and reliable results but sometimes require expert knowledge about the sensor transfer characteristics.

Neural networks have the advantage that they learn themselves the proper transformation and usually the identification is good, but the quantification (indication of the concentration) has higher errors than conventional pattern recognition methods unless again the method is matched to the specific transfer characteristic [219].

The use of fuzzy logic has an advantage when expert knowledge about cross sensitivities is to be transferred into electronic signal processing [220].

5.2.8 Noise Equivalent Concentration

The detection of chemical substances is limited by disturbances such as drift, external perturbations, eg temperature fluctuations, and statistical fluctuations in the sensor and the electronic circuitry (noise). Since the influence of drift can be greatly reduced by gas modulation techniques [221-223] and external perturbations by compensational methods, the final limit is determined by random noise. Therefore, the lowest detectable concentration can be called the noise equivalent concentration (NEC), even though in practice the frequency shift or external perturbations still limit the lowest detectable concentration.

Figure 5-19 shows schematically the sources of external perturbations and noise for a QMB sensor system. The influence of these different "noise" sources will be discussed briefly in the following.

1. The impact of gas molecules on the sensor's surface leads to a stepwise change in the oscillator's frequency. This effect represents a lower theoretical limit for the NEC value and is many orders of magnitude smaller than the technical relevant sources of perturbations.

The most relevant influential factors (see items 2-4 below) have their origin in a change of temperature.

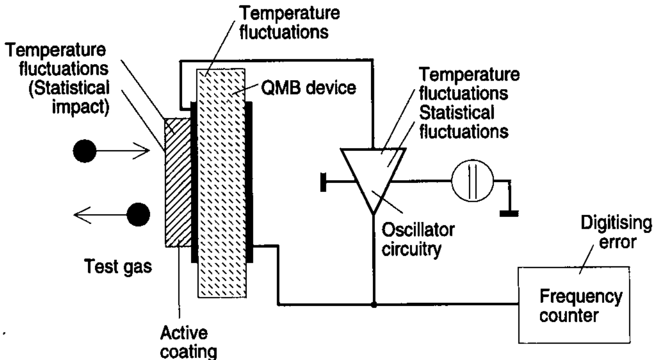


Figure 5-19. QMB gas sensor system (schematic) with "noise" sources.

2. The interaction between the test gas and the coating depends upon temperature [224]. At room temperature and for the detection of solvent vapors with GC coatings the relative signal change is 5% K^{-1} due to this effect [225]. This effect will influence the accuracy of the measurement but not the smallest detectable concentration.
3. The temperature drift of the AT-cut quartz device itself is in the vicinity of the turning-point of the temperature curve zero, but within the typical fabrication tolerance (see Figure 5-10) a relative change of 10^{-6} K^{-1} of the resonance frequency is possible.
4. The amplifier can be considered as a two-terminal device. The imaginary part of its impedance is subject to temperature variations and therefore the oscillator frequency changes. Without compensation the relative frequency change is typically of the order of 10^{-6}–10^{-5} K^{-1}. As mentioned above, a reference crystal is used in order to compensate for temperature fluctuations. If the corresponding two electronic circuits are subject to the same temperature fluctuations, this influence will also be compensated for. In this way an improvement by one to two orders of magnitude is possible.
5. The frequency noise of the oscillator electronics is typically in the order of 10^{-7} (relative), and is also affected by the Q-value of the resonator, ie by the kind and thickness of the coating [226].
6. The frequency counters used generally have a digitizing error of 0.1 or 1 Hz.

A summary of the parameters influencing the NEC value and a comparison between SAW and QMB gas sensors is given in Table 5-10. It can be concluded that SAW gas sensors can be more sensitive owing to their higher frequency of operation. The improvement is about a factor $\sqrt{f_{SAW}/f_{QMB}}$ and thus not extremely high. The value of f^2_{SAW}/f^2_{QMB} usually stated in the literature (eg Ref. 227) does not compare optimized systems. Especially QMB sensors can have coatings much thicker than SAW sensors, so that the comparison of the two devices with the same thicknesses of coatings is not meaningful.

5.2.9 Examples of Applications

QMB sensors can be used as "single-element sensors" (with an additional temperature compensation crystal) if either the coating is very selective or if the substance to be measured is known. Simple sensors of this kind can be used, for instance, for the supervision of storage rooms for chemicals (eg for organic solvents).

Sensor arrays have a wider field of applications. Figure 5-20 shows a stand-alone system with a six-dimensional QMB sensor array and signal processing electronics. The typical lowest detectable concentration (under real conditions of the environment) is of the order of 100 ppm and the typical resolution time (including pattern recognition) is about 10 s. Such sensor equipment can be used for the on-line control of chemical processes, for the supervision of the atmosphere in closed rooms, and for the quality control of food or cosmetic articles. Figure 5-21 shows the ability to distinguish different mixtures of spices of this "electronic nose" as compared with the human nose. Generally the resolution is as good as or better than that with the human nose and the sensitivity limit is of the same order of magnitude for both. The additional advantage of the electronic nose is its better reproducibility.

Table 5-10. Typical values for frequency fluctuations caused by the different "noise" sources and corresponding NEC components (temperature fluctuations of the device and the electronics are assumed to be correlated)

AT-cut quartz at 10 MHz, active coating silicone GE-SE 30 (500 nm), test gas chloroform.		
	Uncompensated	Compensated
Temperature drift, quartz/K	10 Hz 2000 ppm	0.1 Hz 20 ppm
Temperature drift, electronics/K	10 Hz 2000 ppm	0.1 Hz 20 ppm
Frequency noise, oscillator (rms)	0.1 Hz 20 ppm	0.1 Hz 20 ppm
Digitising error, counter	1 Hz 200 ppm	1 Hz 200 ppm
Total error (rms)	20 Hz 4000 ppm	1.02 Hz 205 ppm
SAW device: delay line 3 μs, 433 MHz, active coating silicone GE-SE 30 (5 nm), test gas chloroform.		
	Uncompensated	Compensated
Temperature drift, SAW/K	100 Hz 2000 ppm	1 Hz 20 ppm
Temperature drift, electronics/K	100 Hz 2000 ppm	1 Hz 20 ppm
Frequency noise, oscillator (rms)	1 Hz 20 ppm	1 Hz 20 ppm
Digitising error, counter	1 Hz 20 ppm	1 Hz 20 ppm
Total error (rms)	200 Hz 4000 ppm	2.5 Hz 50 ppm

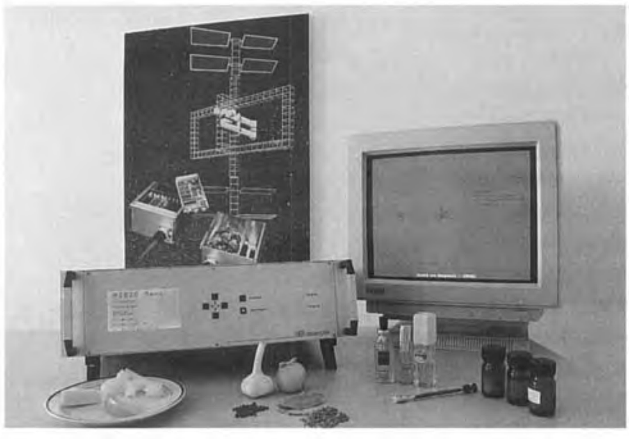


Figure 5-20.
QMB measurement system with integrated pattern recognition for use in olfactometry in the food and cosmetics industries. Photograph courtesy of HKR Sensorsysteme, Munich, Germany.

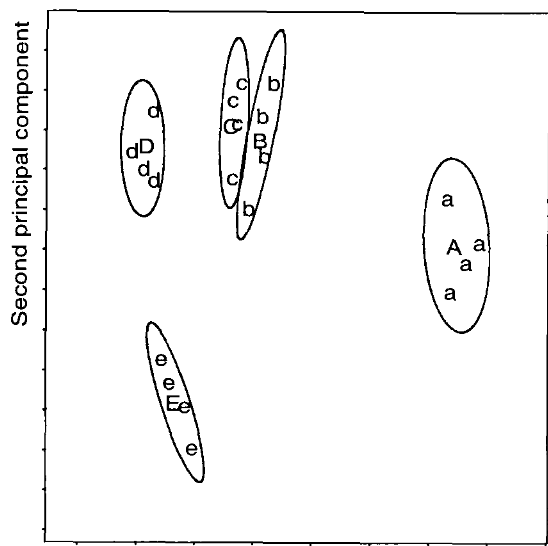


Figure 5-21.
Identification of different mixtures of spices in the feature space. C denotes the standard mixture. The samples B and D cannot be distinguished from C by test persons. Mixtures A and E can be distinguished by test persons.

5.3 References

[1] Matthews, H., (ed.), *Surface Wave Filters;* New York: Wiley, 1977.
[2] Oliner, A. A., (ed.), *Acoustic Surface Waves;* Berlin: Springer, 1978.
[3] Morgan, D. P., *Surface-Wace Devices for Signal Processing;* Amsterdam: Elsevier, 1985.
[4] Datta, S., *Surface Acoustic Wave Devices;* Englewood Cliffs, NJ: Prentice-Hall, 1985.
[5] Campbell, C., *Surface Acoustic Wave Devise and Their Signal Processing Applications;* Boston: Academic Press, 1989.
[6] Feldmann, M., Hénaff, J., *Surface Acoustic Waves for Signal Processing;* Boston: Artech House, 1989.
[7] Rayleigh, Lord, *Proc. London Math. Soc.* **17** (1885) 4-11.
[8] Auld, B. A., *Acoustic Fields and Waves in Solids,* Vol. II; New York: Wiley, 1973.
[9] Bleustein, J. L., *Appl. Phys. Lett.* **13,** (1968) 412-413.
[10] Gulyaev, Yu. V., *Sov. Phys. JETP Lett.* **9** (1969) 37-38.
[11] Lewis, M., *Proc. Ultrason. Symp.* (1977) 744-752.
[12] Parker, T. E., in: *Precision Frequency Control: Oscillators and Standards,* Vol. 2, Garber, E. A., Ballato, A. (eds.); New York: Academic Press, 1985.
[13] White, R. M., *Proc. Ultrason. Symp* (1985) 490-494.
[14] Wohltjen, H., *Sensors Actuators* **5** (1984) 307-325.
[15] Nieuwenhuizen, M. S., Venema, A., in: *Sensors. A Comprehensive Survey,* Vol. 2, Göpel, W., Hesse, J., Zemel, J. N. (Series eds.), Jones, T. A., Kleitz, M., Lundström, J., Seiyama, T. (Volume eds.); Weinheim: VCH, 1991.
[16] Wenzel, S. W., White, R. M., *Appl. Phys. Lett.* **54** (1989) 1976-1978.
[17] Anisimkin, V. I., Kotelyanskij, I. M., *Sov. Phys. Tech. Phys.* **37** (1992) 226-227.
[18] White, R. M., Wicher, P. J., Wenzel, S. W., Zellers, E. T., *IEEE Trans. UFFC* **34** (1987) 162-171.
[19] Costello, B. J., Martin, B. A., White, R. M., *Proc. Ultrason. Symp.* (1989) 977-981.
[20] Wenzel, S. W., White, R. M., *Sensors Actuators A* **21-23** (1990) 700-703.
[21] Martin, S. J., Ricco, A. J., Niemczyk, T. M., Frye, G. C., *Sensors Actuators* **20** (1989) 253-268.
[22] Martin, S. J., Frye, G. C., *Appl. Phys. Lett.* **57** (1990) 1867-1869.

[23] Ingebrigtsen, K. A., *J. Appl. Phys.* **41** (1970) 454-459.
[24] Ricco, A. J., Martin, S. J., Zipperian, T. E., *Sensors Actuators* **8** (1985) 319-333.
[25] Lec, R., Vetelino, J. F., Falconer, R. S., Xu, Z., *Proc. Ultrason. Symp.* (1988) 585-589.
[26] Kondoh, J., Shiokawa, S., *Electron. Commun. Jpn., Part 2,* **76**, No. 2 (1993) 69-82.
[27] Shana, Z., Josse, F., *Proc. Ultrason. Symp.* (1988) 549-554.
[28] Martin, B. A., Wenzel, S. W., White, R. M., *Sensors Actuators A* **21-23** (1990) 704-708.
[29] Josse, F., Shana, Z. A., *IEEE Trans. UFFC* **38** (1991) 297-304.
[30] Niemczyk, T. M., Martin, S. J., Frye, G. C., Ricco, A. J., *J. Appl. Phys.* **64** (1988) 5002-5008.
[31] Liew, S., Josse, F., Haworth, D. T., Shana, Z. A., Kelkar, U. R., Grunze, M. *Proc. Ultrason. Symp.* (1990) 285-290.
[32] Josse, F., Shana, Z. A., Haworth, D. T., Liew, S., *Sensors Actuators B* **9** (1992) 97-112.
[33] Reeder, T. M., Cullen, D. E., Gilden, M., *Proc. Ultrason. Symp.* (1975) 264-268.
[34] Hauden, D., Planat, M., Gagnepain, J.-J., *IEEE Trans. SU* **28** (1987) 342-348.
[35] Staples, E. J., Wise, J., DeWames, R. E., *Proc. Ultrason. Symp.* (1981) 155-158.
[36] Risch, M. R., *Sensors Actuators* **6** (1984) 127-133.
[37] Hauden, D., *IEEE Trans. UFFC* **34** (1987) 253-258.
[38] Zwicker, U., *Sensors Actuators* **17** (1989) 235-239.
[39] Sinha, B. K., Onodera, S., Jinzaki, Y., Obuchi, A., Groves, J., *Proc. Ultrason. Symp.* (1992) 257-262.
[40] Wenzel, S. W., White, R. M., *IEEE Trans. ED,* **35** (1988) 735-743.
[41] Wenzel, S. W., Martin, B. A., White, R. M., *Proc. Ultrason. Symp.* (1988) 563-567.
[42] Tirole, N., Choujaa, A., Hauden, D., Martin, G., Blind, P., Froelicher, M., Pommier, J. C., Cachard, A., *Proc. Ultrason. Symp.* (1993) 371-374.
[43] Ishido, M., Imaizumi, T., Toyoda, M., *IEEE Trans. IM* **36** (1987) 83-86.
[44] Motegi, R., Okajima, H., Ohuchi, H., Guan, N., *Proc. Ultrason. Symp.* (1991) 331-335.
[45] Vorwalder, G., Schrom, G., Seifert, F., *Final Report on Project P5311;* Vienna: TU, Applied Electronics Laboratory, 1991 (in German).
[46] Ishii, A., Hashimoto, S., *Proc. Ultrason. Symp.* (1981) 167-170.
[47] Adler, R., Desmares, P. J., *IEEE Trans. UFFC* **34** (1987) 195-201.
[48] Parker, T. E., Callerame, J., *Proc. Ultrason. Symp.* (1981) 129-134.
[49] Lao, B. Y., *Proc. Ultrason. Symp.* (1980) 687-691.
[50] Tiersten, H. F., Stevens, D. S., Das, P. K., *Proc. Ultrason. Symp.* (1980) 692-695.
[51] Hartemann, P., Meunier, P.-L., *Proc. Ultrason. Symp.* (1981) 152-155.
[52] Meunier, P.-L., Hartemann, P. *Proc. Ultrason. Symp.* (1982) 299-302.
[53] Hartemann, P., Meunier, P.-L., *Proc. Ultrason. Symp.* (1983) 291-294.
[54] Bonbrake, T. B., Erikson, C. A., Thoma, D., *Proc. Ultrason. Symp.* (1985) 591-594.
[55] Hauden, D., Bindler, F., Coquerel, R., *Proc. Ultrason. Symp.* (1985) 486-489.
[56] Motamedi, M. E., *IEEE Trans. UFFC* **34** (1987) 237-242.
[57] Sugizaki, G., Takenaka, T., Sakata, K., Toda, K., *Jpn. J. Appl. Phys.* **32**, Part 1, No. 9B (1993) 4237-4240.
[58] Moriizumi, T., Unno, Y., Shiokawa, S., *Proc. Ultrason. Symp.* (1987) 579-582.
[59] Nomura, T., Yasuda, T., *Proc. Ultrason. Symp.* (1990) 307-310.
[60] Furukawa, S., Furukawa, H., Nomura, T., Yasuda, T., Tamura, M., *Proc. Ultrason. Symp.* (1992) 303-306.
[61] Furukawa, S., Nomura, T., Yasuda, T., Nishimura, R., Hiramatsu, T., Tamura, M., *Proc. Ultrason. Symp.* (1993) 355-358.
[62] Rajendran, V., Koike, M., Hashimoto, K., Yamaguchi, M., *Proc. Ultrason. Symp.* (1992) 263-268.
[63] Ricco, A. J., Martin, S. J., *Appl. Phys. Lett.* **50** (1987) 474-1476.
[64] Hoummady, M., Hauden, D., Bastien, F., *Proc. Ultrason. Symp.* (1991) 345-348.
[65] Ahmad, N., *Proc. Ultrason. Symp.* (1985) 483-485.
[66] Joshi, S. G., *Proc. Ultrason. Symp.* (1988) 555-558.
[67] Brace, J. G., Sanfelippo, T. S., Joshi, S. G., *Proc. Ultrason. Symp.* (1989) 573-578.
[68] Joshi, S. G., Jin, Y., *Proc. Ultrason. Symp.* (1990) 319-322.
[69] Joshi, S. G., Jin, Y., *IEEE Trans. UFFC,* **37** (1990) 475-477.
[70] Joshi, S. G., *IEEE Trans. UFFC* **38** (1991) 148-154.
[71] Joshi, S. G., *Proc. Ultrason. Symp.* (1980) 438-441.

[72] Joshi, S. G., Dasgupta, B. B., *Proc. Ultrason. Symp.* (1981) 319-323.
[73] Inaba, R., Kasahara, Y., Wasa, K., *Proc. Ultrason. Symp.* (1982) 312-316.
[74] Inaba, R., Kasahara, Y., *IEEE Trans. SU* **SU-29** (1982) 381-385.
[75] Joshi, S. G., *J. Acoust. Soc. Am.* **72** (1982) 1872-1877.
[76] Budreau, A. J., Carr, P. H., Bertoni, H. L., *Proc. Ultrason. Symp.* (1983) 327-330.
[77] Gatti, E., Palma, A., Verona, E., *Sensors Actuators* **4** (1983) 45-54.
[78] Budreau, A. J., Scalzi, G. J., Carr, P. H., Bertoni, H. L., *IEEE Trans. SU* **31** (1984) 646-651.
[79] Joshi, S. G., *Proc. Ultrason. Symp.* (1984) 203-206.
[80] Palmieri, L., Socino, G., Verona, E., *Appl. Phys. Lett.* **49** (1986) 1581-1583.
[81] Ishido, M., Zhu, X., *Proc. Ultrason. Symp.* (1987) 637-640.
[82] Thaxter, J. B., Carr, P. H., Silva, J. H., *IEEE Trans. UFFC* **35** (1988) 525-530.
[83] Vellekoop, M. J., Visser, C. C. G., *Proc. Ultrason. Symp.* (1988) 568-575.
[84] Ishido, M., Ojima, K., Kikuta, N., *Proc. Ultrason. Symp.* (1991) 341-344.
[85] Hikita, K., Suetsugu, T., Iizuka, H., Miyayama, M., Yanagida, H., *Jpn. J. Appl. Phys.* **32**, Part 1, No. 6A (1993) 2762-2767.
[86] Palma, A., Palmieri, L., Socino, G., Verona, E., *Proc. Ultrason. Symp.* (1984) 958-961.
[87] Palma, A., Palmieri, L., Socino, G., Verona, E., *Appl. Phys. Lett.* **46** (1985) 25-27.
[88] Bogacki, F. J., *Proc. Ultrason. Symp.* (1989) 647-651.
[89] Dahint, R., Grunze, M., Josse, F., Andle, J. C., *Sensors Actuators B* **9** (1992) 155-162.
[90] Kondoh, J., Shiokawa, S., *Jpn. J. Appl. Phys.* **31** Suppl. 31-1 (1992) 82-84.
[91] Kondoh, J., Shiokawa, S., Georgiev, Z., *Sensors Actuators B* **13-14** (1993) 429-431.
[92] Kondoh, J., Matsui, Y., Shiokawa, S., *Proc. Ultrason. Symp.* (1993) 337-340.
[93] Webb, D. C., Forester, D. W., Ganguly, A. K., Vittoria, C., *IEEE Trans. Magn.* **15** (1979) 1410-1415.
[94] Simpson, E. M., Robbins, W. P., *IEEE Trans. Magn.* **16** (1980) 919-921.
[95] Yamaguchi, M., Hashimoto, K. Y., Kogo, H., Naoe, M., *IEEE Trans. Magn.* **16** (1980) 616-618.
[96] Altan, B., Robbins, W. P., *Proc. Ultrason. Symp.* (1981) 311-314.
[97] Robbins, W. P., Young, A., *IEEE Trans. SU* **32** (1985) 423-427.
[98] Hietala, A., Robbins, W. P., *Proc. Ultrason. Symp.* (1986) 239-243.
[99] Hanna, S. M., *IEEE Trans. UFFC* **34** (1987) 191-194.
[100] Robbins, W. P., Hietala, A., *IEEE Trans. UFFC* **35** (1988) 718-722.
[101] Hauden, D., Jaillet, G., Coquerel, R., *Proc. Ultrason. Symp.* (1981) 148-151.
[102] Neumeister, J., Thum, R., Lüder, E., *Sensors Actuators A* **21-23** (1990) 670-672.
[103] Viens, M., Cheeke, J. D. N., *Sensors Actuators A* **24** (1990) 209-211.
[104] Odintzov, M. A., Sushentzov, N. I., Kudryavtzev, T. L., *Sensors Actuators A* **28** (1991) 203-206.
[105] Möller, F., Kuhn, J. *Sensors Actuators A* **30** (1992) 73-75.
[106] Mingfang, L., Haiguo, L., *Sensors Actuators B* **12** (1993) 53-56.
[107] Hines, J. H., Stapor, W. J., *Proc. Ultrason. Symp.* (1990) 471-476.
[108] Bao, X. Q., Burkhard, W., Varadan, V. V., Varadan, V. K., *Proc. Ultrason. Symp.* (1987) 583-585.
[109] Reindl, L., Ruile, W., *Proc. Ultrason. Symp.* (1993) 125-130.
[110] Yamanouchi, K., Shimizu, G., Morishita, K., *Proc. Ultrason. Symp.* (1993) 1267-1270.
[111] Joshi, S. G., *Proc. Ultrason. Symp.* (1987) 601-604.
[112] Martin, S. J., Frye, G. C., Ricco, A. J., Zipperian, T. E., *Proc. Ultrason. Symp.* (1987) 563-567.
[113] Ballantine, D. S., Wohltjen, H., *Proc. Ultrason. Symp.* (1988) 559-562.
[114] Brace, J. G., Sanfelippo, T. S., Joshi, S. G., *Sensors Actuators* **14** (1988) 47-68.
[115] Wixforth, A., Scriba, J., Wassermeier, M. Kotthaus, J. P., Weimann, G., Schlapp, W., *Phys. Rev. B* **40** (1989) 7874-7887.
[116] Galipeau, D. W., Vetelino, J. F., Lec, R., Feger, C., *Sensors Actuators B* **5** (1991) 59-65.
[117] Ricco, A. J., Martin, S. J., *Thin Solid Films* **206** (1991) 94-101.
[118] Drobe, H., Leidl, A., *Micromech. Microeng.* **2** (1992) 176-178.
[119] Neubrand, A., Hess, P., *J. Appl. Phys.* **71** (1992) 227-238.
[120] Ricco, A. J., Kepley, L. J., Thomas, R. C., Sun, L., Crooks, R. M., *Technical Digest of the IEEE Sensor and Actuator Workshop, Hilton Head Island, SC, USA, 22-25 June 1992* (1992) 114-117.
[121] Galipeau, D. W., Vetelino, J. F., Feger, C., *Sensors Actuators B* **13-14** (1993) 432-436.
[122] Costello, B. J., Wenzel, S. W., Wang, A., White, R. M., *Proc. Ultrason. Symp.* (1990) 279-283.
[123] Caliendo, C., Fioretto, D., Socino, G., Verona, E., *Proc. Ultrason. Symp.* (1991) 387-391.
[124] Hughes, R. C., Martin, S. J., Frye, G. C., Ricco, A. J., *Sensors Actuators A* **21-23** (1990) 693-699.

[125] Martin, S. J., Ricco, A. J., *Sensors Actuators A* **21-23** (1990) 712-718.
[126] Wohltjen, H., Dessy, R. E., *Anal. Chem.* **51** (1979) 1458-1475.
[127] Joshi, S. G., Brace, J. G., *Proc. Ultrason. Symp.* (1985) 600-603.
[128] Snow, A. W., Barger, W. R., Klusty, M., Wohltjen, H., Jarvis, N. L., *Langmuir* **2** (1986) 513-519.
[129] Vetelino, J. F., Lade, R., Falconer, R. S., *Proc. Ultrason. Symp.* (1986) 549-554.
[130] Vetelino, J. F., Lade, R. K., Falconer, R. S., *IEEE Trans. UFFC* **34** (1987) 156-161.
[131] Holcroft, B., Roberts, G. G., *Thin Solid Films* **160** (1988) 445-452.
[132] Falconer, R. S., Lec, R., Vetelino, J. F., Xu, Z. *Proc. Ultrason. Symp.* (1989) 585-590.
[133] Martin, S. J., Ricco, J., *Proc. Ultrason. Symp.* (1989) 621-625.
[134] Chang, S., Suzuki, M., Tamiya, E., Karube, I., *Denki Kagaku* **58** (1990) 1107-1113.
[135] Falconer, R. S., Vetelino, J. F., Smith, D. J., Osborn, M. J., *Proc. Ultrason. Symp.* (1990) 315-318.
[136] Nieuwenhuizen, M. S., Nederlof, A. J., *Sensors Actuators B* **2** (1990) 97-101.
[137] Watson, G., Staples, E., *Proc. Ultrason. Symp.* (1990) 311-314.
[138] Arn, D., Blom, N., Dubler-Steudle, K., Graber, N., Widmer, H. M., *Sensors Actuators A* **25-27** (1991) 395-397.
[139] Chang, S.-M., Tamiya, E., Karube, I., *Biosensors Bioelectron.* **6** (1991) 9-14.
[140] Chang, S.-M., Ebert, B. Tamiya, E., Karube, I., *Biosensors Bioelectron.* **6** (1991) 293-298.
[141] Chang, S.-M., Tamiya, E., Karube, I., Sato, M., Masuda, Y., *Sensors Actuators B* **5** (1991) 53-58.
[142] Frye, G. C., Martin, S. J., Cernosek, R. W., Pfeifer, K. B., Anderson, J. S. *Proc. Ultrason. Symp.* (1991) 311-316.
[143] Frye, G. C., Martin, S. J., *Proc. Transducers'91, IEEE* (1991) 566-569.
[144] Rapp, M., Binz, D., Kabbe, I., von Schickfus, M., Hunklinger, S., Fuchs, H., Schrepp, W., Fleischmann, B., *Sensors Actuators B* **4** (1991) 103-108.
[145] Watson, G., Horton, W., Staples, E., *Proc. Ultrason. Symp.* (1991) 305-309.
[146] Amati, D., Arn, D., Blom, N., Ehrat, M., Saunois, J., Widmer, H. M., *Sensors Actuators B* **7** (1992) 587-591.
[147] Arn, D., Amati, D., Blom, N. Ehrat, M., Widmer, H. M., *Sensors Actuators B* **8** (1992) 27-31.
[148] Dickert, F. L., Banski, H. Bengsch, A., Haunschild, A., Hofmann, P., Mages, G., Bulst, W.-E., Knauer, U., Ihm, W., Obermeier, E., Möller, S., *VDI Ber.* No. 939 (1992) 421-426.
[149] Nieuwenhuizen, M. S., Nederlof, A. J., *Sensors Actuators B* **9** (1992) 171-176.
[150] Smith, D. J., Vetelino, J. F., Falconer, R. S., Wittman, E. L., *Technical Digest of the IEEE Sensor and Actuator Workshop, Hilton Head Island, SC, USA, 22-25 June 1992* (1992) 78-81.
[151] Watson, G., Ketchpel, R. D., Staples, E. J., *Proc. Ultrason. Symp.* (1992) 253-256.
[152] Watson, G., Horton, W., Staples, E., *Proc. Ultrason. Symp.* (1992) 269-273.
[153] Caliendo, C., Verona, E., D'Amico, A., Furlani, A., Iucci, G., Russo, M. V., *Sensors Actuators B* **15-16** (1993) 288-292.
[154] Dickert, F. L., Haunschild, A., Reif, M., Bulst, W.-E., *Adv. Mater.* **5** (1993) 277-279.
[155] Dickert, F. L., Bruckdorfer, Th., Feigl, H., Haunschild, A., Kuschow, V., Obermeier, E., Bulst, W.-E., Knauer, U., Mages, G., *Sensors Actuators B* **13-14** (1993) 297-301.
[156] Dickert, F. L., Haunschild, A., *Adv. Mater.* **5** (1993) 887-895.
[157] Dickert, F. L., Schuster, O., *Adv. Mater.* **5** (1993) 826-829.
[158] Dickert, F. L., Bäumler, U. P. A., Zwissler, G. K., *Synth. Met.* **61** (1993) 47-52.
[159] Frye, G. C., Martin, S. J., *Proc. Ultrason. Symp.* (1993) 379-383.
[160] Liron, Z., Greenblatt, J., Frishmann, G., Gratziani, N., Biran, A., *Sensors Actuators B* **12** (1993) 115-122.
[161] Nomura, T., Yasuda, T., Furukawa, S., *Proc. Ultrason. Symp.* (1993) 417-420.
[162] Rebière, D., Duchamp, G., Pistré, J., Hoummady, M., Hauden, D., Planade, R., *Sensors Actuators B* **13-14** (1993) 642-645.
[163] Reichert, J., Coerdt, W., Ache, H. J., *Sensors Actuators B* **13-14** (1993) 293-296.
[164] Ricco, A. J., Martin, S. J., *Sensors Actuators B* **10** (1993) 123-131.
[165] Zellers, E. T., Pan, T.-S., Patrash, S. J., Han, M., Batterman, S. A., *Sensors Actuators B* **12** (1993) 123-133.
[166] Baer, R. L., Costello, B. J., Wenzel, S. W., White R. M., *Proc. Ultrason. Symp.* (1991) 321-326.
[167] Vellekoop, M. J., van Rhijn, A. J., Lubking, G. W., Venema, A., *Sensors Actuators A* **25-27** (1991) 699-703.

[168] Rebière, D., Pistré, J., Hoummady, M., Hauden, D., Cunin, P., Planade, R., *Sensors Actuators B* **6** (1992) 274-278.
[169] Calabrese, G. S., Wohltjen, H., Roy, M. K., *Proc. Ultrason. Symp.* (1986) 607-610.
[170] Lec, R., Vetelino, J. F., Clarke, P., Roy, A., Turner, J., *Proc. Ultrason. Symp.* (1988) 543-548.
[171] Baer, R. L., Flory, C. A., *Proc. Ultrason. Symp.* (1991) 279-284.
[172] Josse, F., Shana, Z. A., *IEEE Trans. UFFC* **38** (1991) 297-304.
[173] Baer, R. L., Flory, C. A., Tom-Moy, M., Solomon, D. S., *Proc. Ultrason. Symp.* (1992) 293-298.
[174] Nomura, T., Yasuda, T., Furukawa, S., *Jpn. J. Appl. Phys.* **31**, Suppl 31-1 (1992) 78-81.
[175] Nomura, T., Yasuda, T., Furukawa, S., *Proc. Ultrason. Symp.* (1992) 299-302.
[176] Drobe, H., Leidl, A., Rost, M., Ruge, I., *Sensors Actuators A* **37-38** (1993) 141-148.
[177] Kondoh, J., Shiokawa, S., *Proc. Ultrason. Symp.* (1993) 421-424.
[178] Rapp, M., Moss, D. A., Reichert, J., Ache, H. J., *Proc. 7th Int. Conf. Solid-State Sensors and Actuators* (1993) 538-540.
[179] Martin, S. J., Ricco, A. J., Frye, G. C., Niemczyk, T. M., Adhihetty, I., *Proc. Ultrason. Symp.* (1988) 607-611.
[180] Andle, J. C., Vetelino, J. F., Lec, R., McAllister, D. J., *Proc. Ultrason. Symp.* (1989) 579-584.
[181] Andle, J., Vetelino, J., Lade, M., McAllister, D., *Proc. Ultrason. Symp.* (1990) 291-294.
[182] Andle, J. C., Vetelino, J. F., Josse, F., *Proc. Ultrason. Symp.* (1991) 285-288.
[183] Andle, J. C., Josse, F., Vetelino, J. F., McAllister, D. J., *Proc. Ultrason. Symp.* (1992) 287-292.
[184] Andle, J. C., Vetelino, J. F., Lade, M. W., McAllister, D. J., *Sensors Actuators B* **8** (1992) 191-198.
[185] Andle, J. C., Weaver, J. T., Vetelino, J. F., McAllister, D. J., *Proc. Ultrason. Symp.* (1993) 331-335.
[186] Andle, J. C., Weaver, J. T., McAllister, D. J., Josse, F., Vetelino, J. F., *Sensor Actuators B* **13-14** (1993) 437-442.
[187] Caliendo, C., D'Amico, A., Verardi, P., Verona, E., *Proc. Ultrason. Symp.* (1990) 383-387.
[188] Caliendo, C., Verona, E., D'Amico, A., Mascini, M., Moscone, D., *Sensors Actuators B* **7** (1992) 602-605.
[189] Gizeli, E., Stevenson, A. C., Goddard, N. J., Lowe, C. R., *IEEE Trans. UFFC* **39** (1992) 657-659.
[190] Kovacs, G., Lubking, G. W., Vellekoop, M. J., Venema, A., *Proc. Ultrason. Symp.* (1992) 281-285.
[191] Gizeli, E., Stevenson, A. C., Goddard, N. J., Lowe, C. R., *Sensors Actuators B* **13-14** (1993) 638-639.
[192] Stebenson, A. C., Gizeli, E., Goddard, N. J., Lowe, C. R., *Sensors Actuators B* **13-14** (1993) 635-637.
[193] Reindl, L., Müller, F., Ruppel, C., Bulst, W.-E., Seifert, F., *Int. Pat. Appl.* WO 93/13495, 1992.
[194] Bulst, W.-E., Ruppel, C., *Siemens Rev. (Germany),* Special Issue, Spring (1994) 2-6.
[195] Sauerbrey, G., *Z. Phys.* No. 155 (1959) 206-222.
[196] King, W. H., *Anal. Chem.* **36** (1964) 1735-1739.
[197] Kleinschmidt, P., Winter, H., in: *Sensorik,* 4th edn., Heywang, W., (ed.); Berlin: Springer, 1993, Ch. 6.
[198] Sohre, F., *Quarzsensoren, Symposium Schwingquarze '89, Frankfurt/Main, June 14, 1989,* pp. 145-169.
[199] Müller, R., Göpel, W., Hesse, J., Zemel, J. N., Grandke, T., Ko, W. H., (eds.); in: *Sensors,* Vol. 1, Weinheim: VCH, 1989, Ch. 11.
[200] *Gas-Chromatographie Catalogue;* Düren: Macherey – Nagel, 1991, pp. 24-25.
[201] Mauder, A., *Calculation of Segregation Coefficients from GC Retention Times,* Internal Report, Lehrstuhl für Technische Elektronik, 1994.
[202] Schmautz, A., *Piezoelektrisch gravimetrische Gassensoren,* Doctor Thesis, Technische Universität München, Munich, 1992.
[203] Näbauer, A., et al., *Sensors Actuators B* **1** (1990) 508-509.
[204] Dickert, F. L., et al., *Sensors Actuators B* **6** (1992) 25-28.
[205] Dickert, F. L., et al., *Sensors Actuators B* **13-14** (1993) 297-301.
[206] Hearst, J. E., *Contempory Chemistry;* San Francisco: Freeman, 1976, Ch. 11.
[207] Wagner, F.-J., *Untersuchung und Verbesserung des Beschichtungsprozesses piezoelektrischer und interdigitialer Gassensoren,* Diplomarbeit, Technische Universität München, Munich, 1994.
[208] Tessier, L., et al., *Sensors Actuators B* **18-19** (1994) 698-703.
[209] Kuchling, H., *Taschenbuch der Physik,* 5th-8th edn.; Thun: Harri Deutsch, 1986, pp. 329-332.
[210] Rosenbaum, J. F., *Bulk Acoustic Wave Theory and Devices;* Boston: Artech House, 1988, Ch. 10.

[211] Matthys, R. J., *Crystal Oscillator Circuits;* New York: Wiley, 1983, pp. 173-179.
[212] Neubig, B., *Schwingquarze in der Oszillatorschaltung, Symposium Schwingquarze '85, Frankfurt/Main, 1985,* pp. 51-108.
[213] Schulzke, L., *Elektrische und thermische Parameter von AT-Schnitt-Quarzen, Symposium Schwingquarze '89, Frankfurt/Main, June 14, 1989,* pp. 45-106.
[214] Schmautz, A., *Sensors Actuators B* **6** (1992) 38-44.
[215] Kowalsky, B. R., Illman, D. L., Sharaf, M. A., *Chemometrics (Chemical Analysis,* Vol. 2); New York: Wiley, 1986.
[216] Müller, R., Horner, G., *Siemens Forschungs- und Entwicklungsber.* No. 15 (1986) 95-100.
[217] Niebling, G., *Sensors Actuators B* **18** (1994) 259-263.
[218] Horner, G., Hierold, C., *Sensors Actuators B* **2** (1990) 173-184.
[219] Niebling, G., Schlachter, A., *Sensors Actuators* in press.
[220] Wolf, W., et al., *Sensors Acutators* in press.
[221] Kindlund, A., Sundgren, H., Lundström, I., *Sensors Actuators* **6** (1984) 1-17.
[222] Otagawa, T., Stetter, J. R., *Sensors Actuators* **11** (1987) 251-264.
[223] Auerbach, F. J., Seeger, A., *Sensors Actuators* in press.
[224] Atkins, P. W., *Physikalische Chemie;* Weinheim: VCH, 1987, Ch. 8.2.
[225] Pfefferseder, A., *Piezoelektrisches Gas-Sensor-Array, PTB, Bericht Schwerpunktprogramm "Sensorsysteme", Braunschweig, April 13-14, 1994,* pp. 148-159.
[226] Mauder, A., *Sensors Actuators* in press.
[227] Ballantine, D. S., Wohltjen, H., *Anal. Chem.* **61** (1989) 704-712.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

6 High-Temperature Microsensors

J. GERBLINGER, K. H. HAERDTL *) and H. MEIXNER, Siemens AG, München, Germany,
ROBERT AIGNER, Technische Universität, München, Germany

Contents

6.1	Gas Sensors	182
6.1.1	Introduction	182
6.1.2	Thin-Film and Thick-Film Technology	183
6.1.3	Conduction Mechanisms	185
6.1.3.1	General Considerations	185
6.1.3.2	Defect Models for $SrTiO_3$ as a Resistive Oxygen Sensor	187
6.1.3.3	TiO_2, Ga_2O_3, and CeO_2 as Materials for Resistive Oxygen Sensors	190
6.1.4	Kinetics	190
6.1.4.1	General Considerations	190
6.1.4.2	Measurement Methods	192
6.1.4.3	Measurement Results	193
6.1.5	Surface Models and Cross-Sensitivities	195
6.1.6	Selective Gas Sensors	198
6.1.7	Sensor Arrays	200
6.2	Temperature Sensors	202
6.2.1	Principle of Operation	202
6.2.2	Heating Structures	203
6.3	Applications	205
6.3.1	Selective Oxygen Sensors on the Basis of Variously Doped $SrTiO_3$ Films	205
6.3.2	Oxygen Sensors for Applications in the Automotive Field	209
6.4	Si Planar Pellistor	212
6.4.1	Introduction	212
6.4.2	Function Principle	214
6.4.3	Structure and Technology	214
6.4.4	Operating Modes	215
6.5	References	217

*) Universität Karlsruhe, Germany

6.1 Gas Sensors

6.1.1 Introduction

The large and growing consumption of energy in our modern industrial world is leading to more and more air pollution from exhaust gases that arise from combustion processes in the course of generating this energy. When the combustion is not managed in an optimum way, the exhaust gases contain the very noxious gases carbon monoxide, unburned hydrocarbons, and nitrogen oxides of every sort, in addition to harmless water and a probably unavoidable amount of carbon dioxide. It is known [1-3] that these harmful exhaust gases can be reduced and the energy can be better utilized by way of optimum combustion control in the form of, eg, a stoichiometric mixture of air and fuel. Such a mixture can be monitored and adjusted by measuring the oxygen content in the exhaust gas.

This is one of the main tasks of modern gas sensor technology. Since the oxygen concentration must be measured within the hot exhaust gas, the only gas sensors of interest here are those whose *operating temperature lies above 650 °C.* At present they are already being used in the automotive field as mass-produced devices, and in the future they will presumably also become familiar for the monitoring of household heating systems. Both fields of application require low-cost solutions using miniaturized devices in the form of high-temperature microsensors, whose dimensions should lie *in the micrometer range,* at least in *one direction.*

For some time now, high-temperature oxygen sensors based on oxygen-ion-conducting ZrO_2 have been successfully used in practice as lambda probes. They work in both the potentiometric and amperometric operational modes and have already been described extensively in the literature [3, 4]. At least up to now, their geometric dimensions have been such that they cannot be called microsensors.

Another kind of high-temperature microsensor works according to the so-called resistive principle. The materials used here are exclusively oxides. At high temperatures, the oxygen content of these oxides settles into thermodynamic equilibrium with the oxygen partial pressure of the surrounding gas atmosphere. This changes the electrical conductivity, which serves as the measuring variable of the sensor. Let us stress here that the conducting mechanism of these materials is predominantly of an electronic nature. Thus the conduction is sustained by electrons and holes, in contrast to potentiometric sensors based on ZrO_2, which, being solid-state electrolytes, have oxygen ions as charge carriers.

The oxides used, eg, TiO_2 [5, 6], $SrTiO_3$ [7-13], CeO_2 [12, 14], and Ga_2O_3 [12, 15], are stable under high temperatures to beyond 1000 °C, so that it is permissible to speak of high-temperature sensors. Since they are used mostly in thin-film or thick-film structures, here the dimensions in the micrometer range that are required for microsensors are present, at least with regard to the film thickness.

At the typical working temperatures of resistive gas sensors, from 700 to 1000 °C, the oxygen ions causing the conductivity changes permeate the entire volume of the sensor materials by diffusion. In this way they are different from the so-called Taguchi sensors made up primarily of SnO_2, whose change in electrical resistance caused by gas action already occurs at temperatures of roughly 300 °C, since here a surface reaction between the gas atmosphere and solid body takes place. Of course, this proceeds at lower temperatures than with sensors where the reaction must take place over the entire volume. Likewise for calorimetric gas sensors

(pellistors) that operate on the basis of catalytically induced reactions, the establishment of equilibrium also takes place at temperatures around 300 °C [17, 18]. Therefore, Taguchi sensors and pellistors are not high-temperature sensors in our meaning of the word and are not considered in the following observations.

Although resistive gas sensors are used above all to detect oxygen, other gases can also be detected with them [19, 20]. At temperatures above 700 °C all combustible gaseous impurities "burn" in oxygen-containing gases by way of a homogeneous reaction, and this happens all the more rapidly and completely the higher the temperature is. Moreover, the oxygen concentration is reduced by an amount dependent on the gaseous impurities. At temperatures below 700 °C, catalytically activated heterogeneous reactions between oxygen and the combustible gases occur at the surfaces of the sensor and housing. The types and rates of these heterogeneous reactions are very diverse and complex, since they depend not only on the composition of the gas mixture, but also on the kinds of surfaces on which the reactions take place.

6.1.2 Thin-Film and Thick-Film Technology

As already mentioned in the Introduction, we apply the term "micro" in microsensors to those cases where at least one dimension of the sensor element lies in the micrometer range. Since sensor elements are produced mainly in a planar technique, the technologies used for making the functional components of a primary sensor (gas-sensitive films, electrodes temperature sensors, heating structures, passivations) are those that allow film thicknesses in the micrometer range to be produced. The most common technologies for this are still the sputtering technique [21-23] and screen printing [24, 25], although in recent years more and more work has been done to develop alternative technologies as a possible replacement for the sputtering process, such as the chemical vapor deposition (CVD) [26] and ion cluster beam (ICB) methods [27-29]. However, as none of these alternatives has direct applications yet, in what follows screen printing and sputtering processes are described as typical fabrication methods for microsensors.

The screen-printing method (also known as thick-film technology) was developed in the early 1970s for manufacturing hybrid circuits, and it was used especially for making passive components. Of all the technologies for making microsensors, thick-film technology is still regarded as the most economical fabrication principle, especially on a mass-production scale [24]. In the screen-printing process, a rubber squeegee presses a screen-printing paste specifically designed for this process through the openings of a screen on to the substrate lying underneath. The screens used are generally fine-meshed steel or nylon screens, which are mounted on aluminium or steel frames. Screen-printing pastes usually contain four basic components: the functional substance, binders, resins, and organic solvents. Fine-grained powders are needed for the functional substance. The binders consist mostly of low-melting glasses, which after the firing process bring about the bonding of the functional film with the substrate. Resins are added to the screen-printing pastes in order to obtain a certain porosity in the thick film later. The solvents are needed to lower the viscosity of the paste and make it able to print.

The functional substances used in the field of high-temperature microsensors can be either fine metal powder for the later conductive paths (electrodes, heaters, temperature sensors), or fine-grained metal oxide powders (gas-sensitive layers, passivations). Now the greatest problem

in thick-film technology consists in finding a material which does not affect the later properties of the functional powder as part of the primary sensor. Above all in the field of gas-sensitive films, impurity ions can cause drastic changes in the electrical properties if they become incorporated into the base material as dopants (see Section 6.3.1). In exceptional cases it may even make sense to dispense with the use of binders altogether or, in cases where binders are necessary, to dope the functional powders in advance in such a way that in the fired condition the binders no longer exert any influence [9].

In order to obtain reproducible properties in such screen-printed structures, for each annealing process an anneal profile must be applied which is designed specifically for the paste used in that process. Basically this firing process can be divided into three stages. In an initial stage, the solvents are volatilized. Then the binders, if present, are fused and washed out of the functional components. At the same time, the resins present in the pastes are burned off, which produces a certain basic porosity in the structures. Finally, in the third stage the functional portion of the screen-printed paste is sintered together and the ultimate porosity of the screen-printed structure is set. Therefore, the temperature in the third stage has a strong influence on the size of the grains of the screen-printed structure that crystallize out. Especially with the gas-sensitive layer, the ratio of surface area to volume is the factor that essentially determines the gas-sensitive behavior.

The sputtering process has been described much more extensively in the literature than the screen printing process [21-23]. In this fabrication process, which takes place under vacuum conditions, various (mostly neutrally) charged particles (atoms or molecular fragments) are ejected from a target via collision cascades due to bombardment from noble gas ions (mostly argon) energized by a gas discharge. These particles pass through the sputter plasma and are deposited on a substrate that faces the target. The typical film thicknesses of sputtered films are in the micrometer range.

Generally, there are several ways to accelerate the noble gas ions arising in the plasma in such a way that they strike the cathode (target) at a high velocity. If the sputter target is an electrically conducting material, a direct-current voltage can be used for this. However, with insulating materials, such as metal oxides, DC sputtering is not possible because of the charge build-up on the cathode. Hence in these cases a high-frequency voltage is used both for ionizing the sputter gas and also for generating the electric field to accelerate the noble gas ions. A magnetic field is superimposed on the electric field in order to limit the spatial extension of the sputter plasma and thus to increase the growth rate of the film being produced. This is done by attaching a permanent magnet behind the target holding plate. Because of the Lorentz force of the magnetic field, the concentration of electrons in the region between the target and substrate is increased. This produces an increased degree of ionization of the plasma in front of the target, which causes the higher sputtering rate.

Studies on polycrystalline targets (eg, alloys) have shown that the bombardment of their surfaces changes the atomic composition at such surfaces [30]. This ultimately has the result that the stoichiometric composition of the growing films becomes different from that of the target [22, 31]. This process, which is called preferential sputtering, can also lead to a stoichiometry different from that of the target when the sputtering is of ternary metal oxides of the form ABO_3 (eg, $SrTiO_3$). Studies performed in recent years have shown that the total sputter pressure is mainly responsible for the A/B ratio of the growing layer [32]. However, the cation ratio in the growing films can also be affected by way of a bias voltage imposed in front of the substrate.

When sputtering gas-sensitive films made of metal oxides, it must be noted that in general reactions of oxygen atoms among one another can develop in the plasma, with the products then being aspirated off by the pumping system. However, the sub-stoichiometry of oxygen that this causes in the growing films can be corrected for by adding a certain amount of oxygen to the customary sputter gas argon [33]. However, up to now there has been hardly any work aimed at clarifying possible partial reactions during a reactive sputtering process. The difficulty is that the oxygen molecules added to the plasma can chemically react in a large number of ways both with the target surface and with the sputtered-off particles in the plasma.

In order to match the kinetic properties or the characteristic curves of gas-sensitive thin films to the particular requirements of the measurement task, the possibility arises of deliberately doping the materials. However, because of the already discussed preferential sputtering, it is advisable here to avoid sputtering from a doped target. This is because it is entirely unclear at what concentrations the dopant is being transferred to the sputtered film. In this case it is considerably more precise to use the so-called sandwich process: here separate layers of the base material and of the dopant being introduced are sputtered on top of one another [34]. A subsequent annealing process at a temperature optimized for the particular film then ensures that the dopant atoms become distributed within the base material. For the base material $SrTiO_3$, for example, it could be shown that dopant concentrations of up to 1 at.% for both acceptors and donors become distributed almost homogeneously over the entire sputtered film if after the sputtering the entire sandwich is post-annealed for 40 h at 1100 °C.

In order to use sputtered or screen-printed films in practice at high temperatures ($T > 600$ °C), in temperature-sensor or gas-sensor technology, it is generally advisable to subject these to a certain pre-aging process. Here it must be made certain in particular that the various functional parts of an elementary sensor do not influence each other. Such problems will be illustrated by way of two examples: Ga_2O_3 films sputtered on Al_2O_3 show a distinct interdiffusion of substrate and gas-sensitive film already at relatively low temperatures ($T > 600$ °C) [35]. Only a switch from Al_2O_3 to BeO substrates permits Ga_2O_3 sensors to be made that can be used at temperatures higher than 800 °C (see Section 6.1.6) [36]. In the case of $SrTiO_3$ films on Al_2O_3 substrates, the sputtered structures can be pre-aged up to a temperature of 1100 °C. Above this temperature, in both thick films and thin films a portion of the strontium in the transition region between film and substrate diffuses into the substrate [11]. Moreover, it must also be made certain in the annealing processes that no interactions (eg, oxidation processes) of possible passivations with the underlying temperature sensors or heating structures occur at the transition region that could lead to changes in the conductivity behavior of these structures. Especially in a combined system of screen-printed electrodes and screen-printed gas-sensitive films, the binders present in each of the pastes should not influence the electrical properties of the other paste.

6.1.3 Conduction Mechanisms

6.1.3.1 General Considerations

The materials used for resistive oxygen sensors are primarily oxides. So far, the literature has described in detail the conduction mechanisms in TiO_2 [37–39], in double oxides with a perovskite structure, eg, $SrTiO_3$ [40, 41], and more recently in CeO_2 [14] and Ga_2O_3 [15, 36, 42].

Common to all these substances is the fact that with increasing oxygen pressure they take up oxygen from the atmosphere until there is a super-stoichiometric concentration of oxygen anions in relation to the metal cations, whose concentration generally does not change with varying oxygen partial pressure. Conversely, a falling oxygen pressure leads to a release of oxygen from the oxidic material. This causes the anion to cation ratio to be shifted in the direction of a cationic super-stoichiometry.

A change in the stoichiometry relationship has decisive effects on the conduction properties. Whereas an anion excess causes an electronic p-type conduction, a cation excess correspondingly produces an electronic n-type conduction. Only exactly at the stoichiometry point is highly resistive, intrinsically conducting material present. These relationships are shown schematically in Figure 6-1.

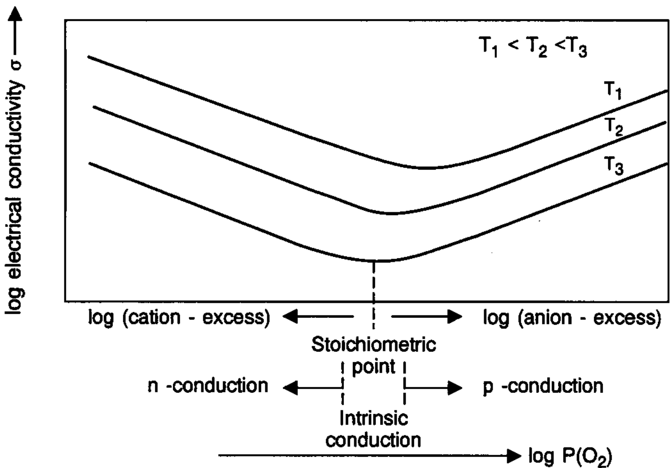


Figure 6-1. Schematic representation of the conductivity variation of a metal oxide as a function of the cation-to-anion ratio. Similar conductivity curves can be expected when the logarithm of the oxygen partial pressure (log $P(O_2)$) is plotted on the x-axis.

Whereas in the case of simple semiconductors the p- or n-type conduction is brought about by corresponding dopant additions in the fabrication process, in oxide semiconducting materials the oxygen absorption or release acts as a doping, which in fact can take place continuously in the same material specimen. Here it is presupposed that no temperature-dependent kinetic obstacles exist to the atmosphere–solid body oxygen transfer, such as are treated in Section 6.1.4.

With increasing temperature, the tendency towards oxygen release and thus towards a cation excess increases, which leads to a shift of the stoichiometric point in Figure 6-1 to higher oxygen partial pressures. At the same time the curves shift towards higher conductivity values, since with increasing temperature the generation of electron–hole pairs grows exponentially.

The type and concentration of the atomic point defects, which compensate for a departure from the exact stoichiometric distribution of an ideal crystal lattice, are specific to the material and must be worked out for each particular material in time-consuming studies. Such investigations have been carried out in greatest detail on perovskite titanates such as $BaTiO_3$

and $SrTiO_3$ [43, 44]. There have also been comprehensive studies on TiO_2 [45], CeO_2 [14], and Ga_2O_3 [15]. In what follows, the defect relationships and their effects on the electric properties that are important to sensor applications are shown by using the example of $SrTiO_3$.

6.1.3.2 Defect Models for $SrTiO_3$ as a Resistive Oxygen Sensor

Perovskite titanates are excellent materials not only for passive components (non-linear resistors, capacitors, piezoelectric ceramics), but also for resistive gas sensors. The factors responsible for their good applicability are their

- great chemical stability over a wide range of oxygen partial pressures, from 1 to 10^{-20} bar;
- temperature stability up to more than 1000 °C;
- adequate oxygen sensitivity of the electrical resistance;
- excellent reversibility of the electrical properties in connection with oxygen absorption and release.

Further, $SrTiO_3$ in particular, like perovskite materials in general, is especially suited to defect-model considerations, since here the defects are limited to just a few characteristic types. Because of its close-packed perovskite lattice, in $SrTiO_3$ no interstitial ions are present, but only vacancies in the anion or cation sublattice. The vacancies in the anion sublattice are oxygen vacancies, V_O, and those of the cation sublattice are strontium vacancies, V_{Sr}. As is customary, the Kroeger-Vink notation is used as the defect notation [46], with the charge state of the defect always being given relative to the undisturbed state. Other defects are formed by extrinsic ions that have been deliberately added as external dopants or else unintentionally added as an impurity during the production process.

Oxygen vacancies may be neutral ($V_O^{\times}$) or else arise in the singly ($V_O^{\bullet}$) or doubly ($V_O^{\bullet\bullet}$) ionized state and thus act as simply or doubly ionized donors. The corresponding situation holds for the strontium vacancies ($V_{Sr}^{\times}$, V_{Sr}', V_{Sr}''), which have an acceptor character.

Since the materials are uncharged in total, electron neutrality must prevail, namely the sum of the concentrations of all negatively and positively charged defects must be equal:

$$n + [V_{Sr}'] + 2[V_{Sr}''] = p + [V_O^{\bullet}] + 2[V_O^{\bullet\bullet}] \qquad (6\text{-}1)$$

where n and p represent the concentrations of the electrons and the holes.

The various defects are related to each other via the reaction equations shown in the first column of Table 6-1. Hence by using the law of mass action, one can derive from these the defect concentrations and their temperature dependence, as shown in the second column.

In the case of singly ionized donor-type ($D^{\bullet}$) or acceptor-type (A') extrinsic dopants, the neutrality equation expands to

$$n + [V_{Sr}'] + 2[V_{Sr}''] + [A'] = p + [V_O^{\bullet}] + 2[V_O^{\bullet\bullet}] + [D^{\bullet}] \qquad (6\text{-}9)$$

Furthermore, as in any semiconductor the relationship

$$np = K_i \qquad (6\text{-}10)$$

Table 6-1. Survey of defect-forming reactions in perovskite titanates.

Reaction equation	Law of mass action	Equation
O (lattice) $\rightleftharpoons V_O + 1/2\ O_2$ (gas)	$[V_O]\ P(O_2)^{1/2} = K_1 = N_1 \exp(-E_1/kT)$	(6-2)
$V_O \rightleftharpoons V_O^{\bullet} + e$	$[V_O^{\bullet}]\ n/[V_O] = K_2 = N_2 \exp(-E_2/kT)$	(6-3)
$V_O^{\bullet} \rightleftharpoons V_O^{\bullet\bullet} + e$	$[V_O^{\bullet\bullet}]\ n/[V_O^{\bullet}] = K_3 = N_3 \exp(-E_3/kT)$	(6-4)
$V_{Sr} \rightleftharpoons V'_{Sr} + h$	$[V'_{Sr}]\ p/[V_{Sr}] = K_4 = N_4 \exp(-E_4/kT)$	(6-5)
$V'_{Sr} \rightleftharpoons V''_{Sr} + h$	$[V''_{Sr}]\ p/[V'_{Sr}] = K_5 = N_5 \exp(-E_5/kT)$	(6-6)
$D \rightleftharpoons D^{\bullet} + e$	$[D^{\bullet}]\ n/[D] = K_D = N_D \exp(-E_D/kT)$	(6-7)
$A \rightleftharpoons A' + h$	$[A']\ p/[A] = K_A = N_A \exp(-E_A/kT)$	(6-8)

must hold, and likewise the Schottky equation:

$$[V_O^{\times}]\,[V_{Sr}^{\times}] = K_S \tag{6-11}$$

can be applied.

Thus, for the ten unknown defect concentrations n, p, $[V_O^{\times}]$, $[V_O^{\bullet}]$, $[V_O^{\bullet\bullet}]$, $[V_{Sr}^{\times}]$, $[V'_{Sr}]$, $[V''_{Sr}]$, $[D^{\bullet}]$, and $[A']$, one has the ten independent conditional Equations (6-2)-(6-11), from which the dependence of the conductivities on temperature and oxygen partial pressure can be calculated from the known mobility of the electrons and holes and from a knowledge of the mass-action constants K_S, K_i, K_D, K_A, K_1, ..., K_5.

Now the main problem lies in obtaining the mobilities of the charge carriers and the values of the mass-action constants from reliable data in the literature or from otherwise independent measurements. For the case of $SrTiO_3$ this has been accomplished to a large measure and has been described extensively [44]. Also Figures 6-2a and 6-2b are taken from that study, and show the following:

- measured curves and calculated model curves are in good agreement (Figure 6-2a);
- depending on temperature and doping, one obtains positive or negative dependences of the electrical conductivities on the oxygen partial pressure $P(O_2)$ of the form $\sigma \sim P(O_2)^{\pm m}$ with $|m|$ between 0.17 and 0.25 (Figure 6-2b);
- through a suitable selection of dopants, materials can be prepared whose dependence of conductivity on partial pressure is unequivocal and as large as possible in the range of partial pressures of interest for measurement purposes (Figure 6-2b).

Thus, for example, tantalum donors can be embedded, so that the n-conducting branch of the equilibrium characteristic line shifts to higher oxygen partial pressure values far enough that in the range of partial pressures investigated only n-conduction is still present, and a clear connection between electrical conductivity and oxygen partial pressure can be seen [47]. Conversely, in acceptor doping by iron or aluminium the equilibrium characteristic shifts to the left and the p-conducting branch of the characteristic covers a larger range of oxygen partial pressures [44] (see Figure 6-2b).

The oxygen absorption or release due to changes in partial pressure occurs in $SrTiO_3$ solely via an alteration of the vacancy concentrations in the overall volume of the material. This occurs without any changes in the number of lattice cells present, which lessens the electrode problems and gives rise to the great degree of reversibility that is found.

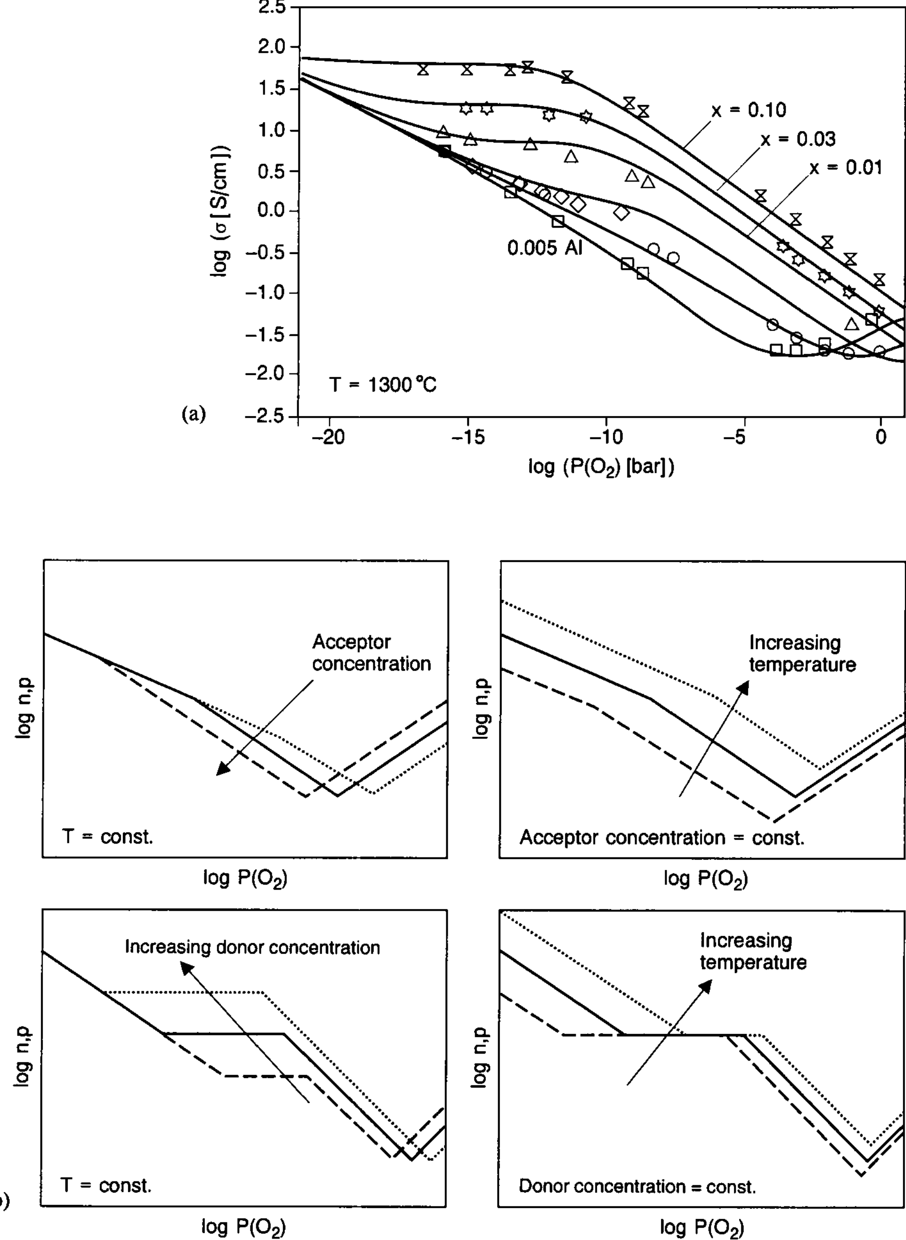


Figure 6-2. (a) Measured values of the electrical conductivity (symbols) as a function of the oxygen partial pressure ($P(O_2)$) in comparison to the calculated model curves. $T = 1300\,°C$; $Sr_{1-x}La_xTiO_3$ donor ceramic; $SrTi_{0.995}Al_{0.005}O_3$ acceptor ceramic. (b) Top: schematic pattern of the charge-carrier densities as the acceptor concentrations or the temperature are varied. Bottom: schematic pattern of the charge-carrier densities as the donor concentration or the temperature is varied.

6.1.3.3 TiO_2, Ga_2O_3, and CeO_2 as Materials for Resistive Oxygen Sensors

In the early 1980s, Logothetis [5, 6] described TiO_2 ceramics as materials for resistive oxygen sensors. Recently such sensors have become commercially available (NGK, Japan). However, our own investigations (Haerdtl) from the 1960s and 1970s showed definite long-term creep effects on the part of the electrical conductivity, the causes of which were first revealed by the detailed studies of Pennewiss [45]. Although the characteristic equilibrium lines found with TiO_2 have many similarities to those of $SrTiO_3$, the defect models of the two substances differ substantially. Whereas the perovskite lattice of $SrTiO_3$ has no interstitial ions, in the rutile lattice of TiO_2 the conductivity is determined by titanium interstitial ions in the form of quadruply ionized donors. With changes in partial pressure, TiO_2 also reacts by the absorption or release of oxygen. The oxygen vacancies arising from release of oxygen at the surface cause a titanium excess at this surface, with this dying away through diffusion of titanium interstitial ions into the interior. This is associated with a change in volume of the material and an increase in density. If oxygen is again made available to the ceramic, the restoration of the lost volume begins by way of interstitial titanium diffusing back from the interior to the surface and forming new lattice cells with oxygen molecules there. When these surfaces are provided with metal contacts, then in the decomposition and buildup of lattice cells it is possible to have changes in the contact behavior at the TiO_2-metal contact interface, which are very likely responsible for the creep behavior.

Investigations with CeO_2 have shown [14] that the defect behavior in this substance is due to ionized oxygen vacancies and thus to n-type conduction. Here we obtain an unequivocal correlation between electrical conductivity and oxygen partial pressure over a large range of partial pressures. More recent investigations with sputtered CeO_2 thin films have shown that this defect model is also applicable to thin films [48, 49].

Like CeO_2, Ga_2O_3 is also an n-type semiconductor over the entire range of oxygen partial pressures investigated, with its electrical conductivity being determined by singly ionized oxygen vacancies [15]. Initial measurements have been made with Ga_2O_3 thin films that have demonstrated the suitability of using this material in connection with the regulation of combustion processes [50].

6.1.4 Kinetics

6.1.4.1 General Considerations

Gas sensors designed for use in controlling on a cylinder-selective basis the combustion processes in motor vehicles need short response times in the range of a few milliseconds. Therefore, a knowledge of the kinetic processes occurring in the oxygen absorption or release is needed for resistive gas sensors designed for this purpose.

The kinetic processes can be represented in three substeps [51], as seen in Table 6-2:

– *Gaseous Diffusion (Step I)*

The first substep in oxygen absorption is the diffusion of oxygen molecules O_2 from within the gas space to the surface of the gas sensor, where the oxygen molecules $O_{2,S}$ are physisorbed.

Table 6-2. Schematic representation of oxygen absorption by titanates (the subscript S denotes particles on the surface).

O_2	$\rightleftharpoons$	O_2							
		$O_{2,S}$	$\rightleftharpoons$	$2O_S$					
				$2O_S$	+	$2V_{O,S}^{\bullet\bullet}$	$\rightleftharpoons$	$2O_O$ +	4h
						$V_{O,S}^{\bullet\bullet}$	$\rightleftharpoons$	$V_{O,Volume}$	
Gas diffusion				Surface reaction				Volume diffusion	

– *Transfer Reaction (Step II)*
There, in a second step, the physisorbed oxygen molecules are split up into oxygen atoms O_S and chemisorbed. These oxygen atoms O_S react at the surface with doubly ionized oxygen vacancies $V_{O,S}$, with the formation of oxygen ions at regular lattice sites O_O and four holes (h) in each case.

– *Volume Diffusion (Step III)*
The last step in the scheme depicted is the ambipolar diffusion of the oxygen vacancies V_O and the holes h within the interior of the solid. This diffusion is driven by the concentration gradient that develops during the transfer reaction.

All three steps are reversible and can proceed in either direction, depending on whether there is oxygen absorption or release. It is to be expected that in the indicated sequence of reactions one or other of the successively occurring steps is slower than the other two and thus determines the overall reaction rate [51–53].

If one disregards the case where with 1 bar of total pressure only vanishingly small amounts of oxygen or oxygen-containing gases are present in the gaseous space that can react to form surface oxygen, then step I can be viewed as rapid compared with steps II and III. When the resistive oxygen sensor is in the shape of a rectangular solid, as shown in Figure 6-3, then if the extensions in the *x*- and *y*-directions are large compared with the *z*-direction, diffusion in the first two directions can be ignored compared with that in the *z*-direction. A similar situation is true for the transfer reactions, where the small side faces play a negligible role compared with the relatively large top and bottom faces.

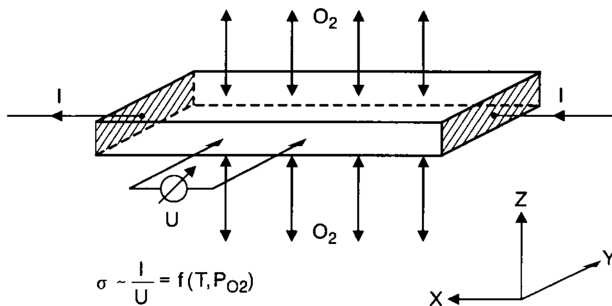


Figure 6-3.
Oxygen exchange between gaseous phase and a box-shaped solid body with dimensions in the *z*-direction small compared with those in the *x*- and *y*-directions.

The rate constants are responsible for the extent of the transfer reactions (step II) and the ambipolar-diffusion constants of the oxygen vacancies are responsible for the volume diffusion (step III). Even without an exact knowledge of these two constants, it can be concluded that for very thin samples in the z-direction the transfer and for thick samples the volume diffusion become rate determining. This is because there is a quadratic increase in time constants with sample thickness in volume diffusion compared with a linear dependence in surface-transfer reactions.

The film thickness below which kinetics controlled by volume diffusion go over into kinetics controlled by transfer reactions depends entirely on the quantitative values of the particular material used and on the measured temperature, and this must be found through measurement [7, 11, 51-54].

6.1.4.2 Measurement Methods

In recent years, two fundamentally different measurement methods have been developed for determining the response times of gas sensors. The first and currently most common measurement records the sensor conductivity as a function of time, for example, in reaction to a step-like change in oxygen partial pressure, with the total pressure of 1 bar being held constant (see top part of Figure 6-4). The response time t_{90}, within which the sensor has completed 90% of its change in conductivity, gives a measure of its speed of response. This measurement in the time domain was first undertaken by Wernicke [52] and later by others, and has been improved with respect to shorter response times [7, 8, 54-59]. This method has disadvantages at relatively short response times, since the abrupt changes in the oxygen partial pressure must be considerably shorter than the measured t_{90} response times.

By contrast, in the frequency method shown in the bottom part of Figure 6-4 the oxygen partial pressure is changed by varying the total pressure while holding constant the atmospheric composition [51, 52]. Here we are working with periodic small pressure changes, which has the advantage that the sensor kinetics can be described by a simple, linear model.

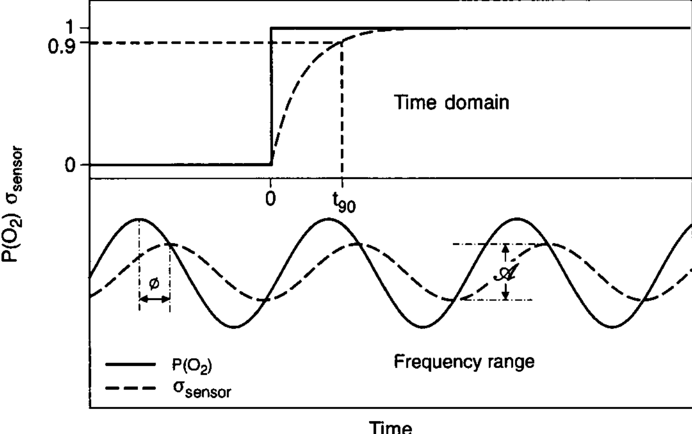


Figure 6-4. Comparison of the usual measurement in the time domain with measurement in the frequency domain.

Although in principle the frequency response provides the same information as the step response, in the former case the measuring principle can be effected in a better defined way.

If one makes a log-log plot of the amplitude of the frequency signal $|A|$ or the phase response φ against the frequency, at low frequencies the frequency is independent of $|A|$ and φ. Above a threshold frequency, in diffusion control a descending straight line with a slope of $-1/2$ is obtained, whereas in the case of reaction control a straight line with a slope of -1 is found. Above the threshold frequency the phase of the frequency signal φ drops off by the angle $-\pi/4$ diffusion control, and by $-\pi/2$ in reaction control.

Compared with information from step responses, measurements in the frequency domain provide not only information on threshold frequencies and thus on response times but also supplementary information from the amplitude and phase variations, and thus give a clearer picture about the nature of the rate-determining step in the transfer of oxygen between gaseous phase and solid body.

6.1.4.3 Measurement Results

From what has been said above, it is understandable that rapidly acting resistive oxygen sensors must have a small film thickness. Regardless of whether volume-diffusion or surface-reaction control determines the oxygen transfer, the response times decrease with decreasing film thickness, in a quadratic fashion in diffusion control and linearly in reaction control.

This predicted changeover from diffusion control to reaction control with the above-described changes in amplitude response and phase position is demonstrated in Figure 6-5 for $SrTiO_3$ single crystals of varying film thickness [53]. It can be seen clearly how crystals that are 840 μm thick show a definite diffusion control (amplitude response $-1/2$, phase position $-\pi/4$), whereas the thin crystals (230 and 90 μm) show in addition to the expected higher threshold frequencies also a definite reaction control (amplitude response -1, phase position $-\pi/2$).

If one takes as a basis the ambipolar diffusion coefficients from Ref. [7] determined for $SrTiO_3$ and presupposes pure volume-diffusion control (although this becomes less and less certain the smaller the film thicknesses are), then for decreasing film thicknesses one obtains extrapolated response times in the millisecond range for film thicknesses of a few micrometers.

$SrTiO_3$ Thick Films

The thick films used have a morphology consisting of individual grains sintered together and with an open porosity between them. Thus the film thickness relevant to the kinetics of the films is characterized by half of the grain diameter and not by the entire film thickness itself. Since the films used almost always have grain diameters of the order of 1 μm, short response times in the millisecond range can be expected here. This agrees with the values obtained by Tragut [51], who studied $SrTiO_3$ thick films doped with 0.5% Fe. At about 1000 °C and with measurements made in the frequency domain, he found threshold frequencies of 70 Hz, corresponding to a response time of 5 ms, and a definite reaction control.

$SrTiO_3$ Thin Films

$SrTiO_3$ thin films made by sputtering [11, 54] with a film thickness of 1 μm and varying accepter dopants show, in time-domain measurements, decreasing response times with increas-

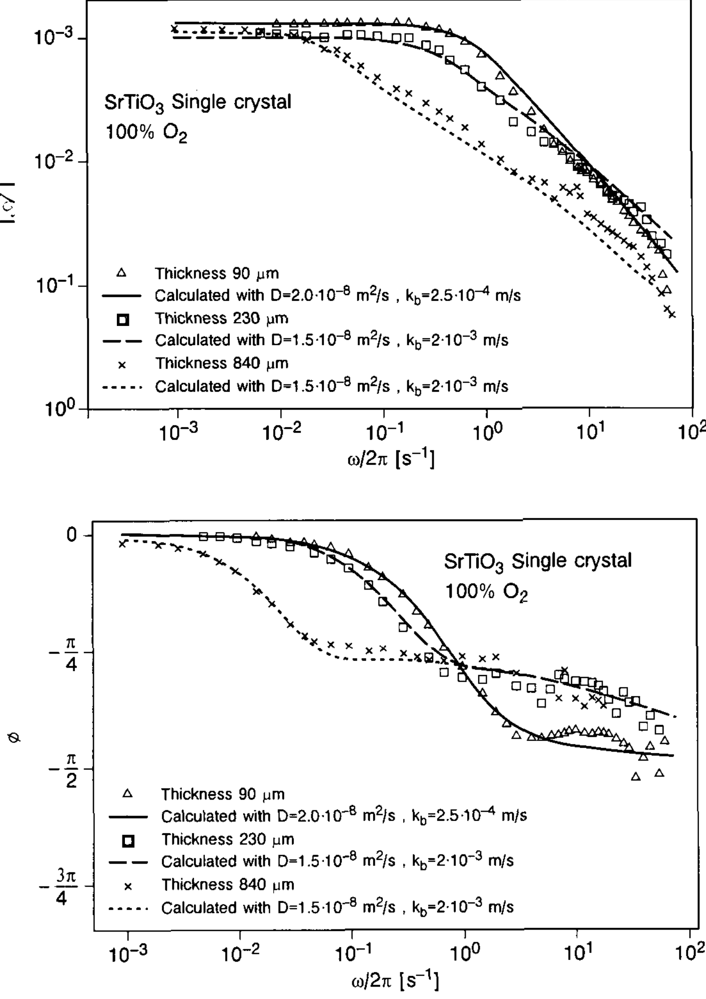


Figure 6-5. Measured and calculated response functions of $SrTiO_3$ single crystals of varying thickness in 100% O_2 at 1000 °C.

ing temperature, which at 1000 °C amount to a few milliseconds. Studies by Tragut [51] that were carried out on the same $SrTiO_3$ thin films in the frequency domain came to the same result: response times in the millisecond range at 1000 °C, and definite reaction control.

CeO_2 Thin Films

Measurements have also been made of response times in the time domain for sputtered CeO_2 thin films with film thicknesses in the range between 1 and 3 μm [48, 49, 56, 57]. The results show the expected decrease in response times with changes in temperature and film thickness. At temperatures of about 1000 °C and a film thickness of 1 μm, minimum response times of 30 ms are reached. The surface reaction is given as the rate-determining step.

6.1.5 Surface Models and Cross-Sensitivities

For gas-sensitive materials, the volume effects described in Sections 6.1.3 and 6.1.4 can be seen in the limiting case of high temperatures. At lower temperatures (in the case of $SrTiO_3$ at about $T < 700\,°C$), it is no longer possible to observe an interaction between the oxygen vacancies in the lattice and the oxygen of the environment, since the defects in the interior of the crystal can be regarded as quasi-frozen in place. In this temperature range above all, an interaction with other gases besides oxygen can clearly be seen [20]. Since the sensor material $SrTiO_3$ has definite advantages for applications in fields of oxygen-sensor technology, we use the term "cross-sensitivities" when speaking about this interaction with other gases. The mechanism for this interaction is governed by reactions at the surface (or in a layer near the surface) of the semiconducting sensor material [60]. The manner in which these reactions take place at the surface depends definitely on what potential relationships prevail there.

Since the charge states at the grain boundaries cannot be found directly, indirect methods must be used to try to determine the electrical potentials [61-64]. An initial comprehensive quantitative picture of the boundary layers in $SrTiO_3$ comes from Tragut [51]. In this model it is assumed, and which in fact was already known from other investigations [65], that the surface of titanates has an excess concentration of titanium. If the electron density is very low, as is usually the case with the titanates, then the d-orbitals of titanium that are not occupied by oxygen cannot accept any electrons and thus can no longer adequately offset the nuclear charge of the titanium. Considered in isolation, the complex seems to be positive. Owing to this positive surface charge, a concentration of electrons and negatively charged vacancies of the lattice (essentially Sr vacancies) develops in a space-charge region along the grain boundary. This leads to an n-type channel, which in thin films is also revealed by the behavior of the electrical conductivity. The partial conductivity of the electrons is distinctly raised compared with that of the positive holes. Therefore, if we assume that single crystals, polycrystalline bulk materials, and thin films all have a like enlargement of their space-charge regions with increasing surface-to-volume ratio, the conclusion is that the minimum in the electric conductivity is shifted towards a higher oxygen partial pressure.

CeO_2 also has charge states at the grain boundaries similar to those of $SrTiO_3$ [48]. Measurements of the electrical conductivity for grain structures of various sizes give rise here to the conjecture that a positive charge develops directly at the surface of sputtered structures and that this charge is balanced by a space-charge zone enriched with electrons. Like $SrTiO_3$, CeO_2 also can be characterized as a stable oxygen-sensitive material at temperatures greater than 700 °C. However, both materials also show interactions with respect to other gases, which occur at almost the same sensor temperatures and are probably due to the same reaction mechanisms [66]. In what follows, $SrTiO_3$ sensors are used by way of example to describe cross-sensitivities with respect to certain selected gases (CO, CO_2, H_2, and CH_4) In order to arrive at the most reliable information possible, we also studied samples of $SrTiO_3$ doped with 1 at.% Ta or 1 at.% Al in parallel with the undoped material. The cross-sensitivity is defined by the following equation:

$$\text{Cross-sensitivity} = \frac{R(Y\%\ \text{react. gas in}\ X\%\ O_2\ \text{and}\ 100-Y\%-X\%\ N_2)}{R(X\%\ O_2\ \text{and}\ 100-X\%\ N_2)} \tag{6-12}$$

In Figure 6-6, the cross-sensitivity of the variously doped $SrTiO_3$ films is shown vis-à-vis 1% CO or 1% CO_2 in 1% O_2 (remainder N_2). A maximum sensitivity to the two gases is found in all films within the temperature range 500–600 °C, with the sensitivity to CO always being greater. Above 600 °C the sensitivities decrease, and at 750 °C (CO_2) and 1000 °C (CO) the effects have virtually disappeared. The behavior shown in Figure 6 corresponds to what would be expected in the interaction of a reducing gas with an n- or p-type semiconductor.

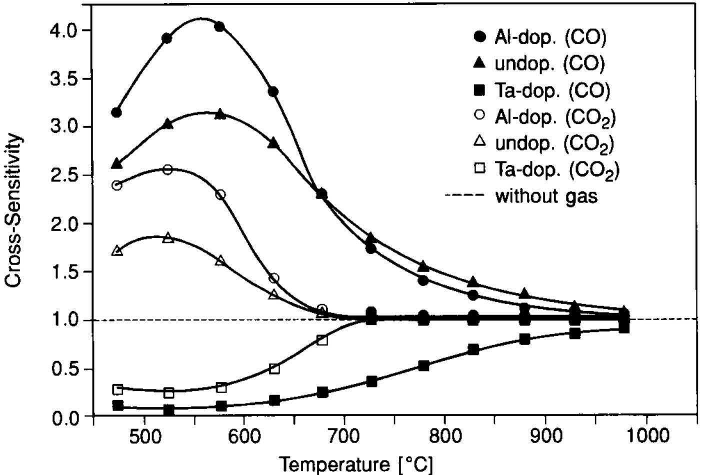


Figure 6-6. Cross-sensitivities of variously doped $SrTiO_3$ films in relation to 1% CO and 1% CO_2 in 1% O_2 (remainder N_2).

The greater change in resistance of the Al-doped $SrTiO_3$ compared with the undoped material can be explained by the greater p-type conductivity here.

The simplest way to depict the reaction with CO is by

$$CO \rightarrow CO^{+}_{ads} + e' \tag{6-13}$$

Alternatively, or in addition, also possible is the formation of a charged complex at the surface or a further oxidation to CO_2 with the generation of an ionized oxygen vacancy:

$$CO_{ads} + O^{\times}_{O} \rightarrow (OCO_O)^{\times} \rightarrow (OCO_O)^{\bullet} + e' \tag{6-14}$$

$$(OCO_O)^{\bullet} \rightarrow CO_2 + V^{\bullet\bullet}_{O} + e' \tag{6-15}$$

In the reactions (6-13), (6-14), and (6-15), one electron is liberated each time and the positive charge at the surface is increased. In the case of p-type materials the released electrons counterbalance the positive holes present. With an n-type material, the number of free charge carriers increases. In the case of CO_2, a reaction at the surface with a direct electron transition does not seem plausible. It should be assumed instead that at the surface a carbonate-like complex is formed:

$$CO_{2,ads} + O^{\times}_{O} \rightarrow (O_2CO_O)^{\times} \tag{6-16}$$

A change in electronegativity is probable due to the complexing described in Equation (6-16), so that the energy gap at the surface is changed. Thereby the existing space-charge region is increased. If for CO the reaction described by Equation (6-14) should take place, this complexing has the same consequences.

The cross-sensitivity studies of 1% H_2 in 1% O_2 (remainder N_2) show that H_2 has a reducing effect only below about 700 °C. Above about 750 °C, with all the samples a behavior can be observed that cannot be expected a priori: in the n-type material an increase in resistance and in the p-type materials a lowering of resistance arise, with the maximum effect occurring at about 770 °C. Above about 900 °C no noticeable cross-sensitivities to H_2 can be seen [20]. As with CO_2, with H_2 also the simplest explanation for the reducing effect is that the molecule is adsorbed at the surface and thereby an electron is released:

$$H_2 \rightarrow H^+_{2\,ads} + e' \tag{6-17}$$

Also possible are the formation of a corresponding surface complex:

$$H_{ads} + O_O^X \rightarrow (HO_O)^{\bullet} + e' \tag{6-18}$$

and the formation of a charged oxygen vacancy.

Figure 6-7 shows the cross-sensitivity behavior of the variously doped $SrTiO_3$ films to 0.4% CH_4 in 5% O_2 (remainder N_2). The results clearly show that at temperatures below about 900 °C, methane has a reducing effect on the sensor materials, as expected. Above about 950 °C, a reversal of the cross-sensitivity behavior takes place, just as was found in the studies on cross-sensitivity to H_2. Also here the temperature of the transition from a reducing to an "oxidizing" effect is dependent on the doping. At a sensor temperature of 1100 °C the cross-sensitivity drops to zero for all films.

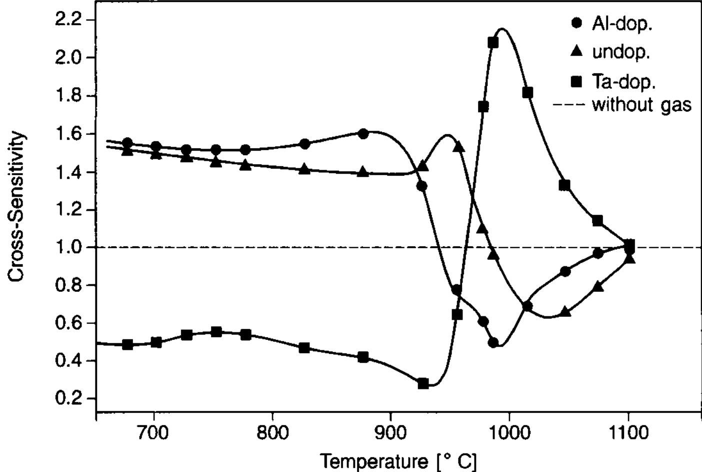


Figure 6-7. Cross-sensitivity of variously doped $SrTiO_3$ films in relation to 0.4% CH_4 in 5% O_2 (remainder N_2).

The reducing effects of methane on the variously doped $SrTiO_3$ films can be described by the following reactions:

$$CH_4 \rightarrow CH_{3,ads} + H_{ads} \tag{6-19}$$

$$2\ CH_{3,ads} + 5\ O_{ads} \rightarrow 2\ CO_{ads} + 3\ H_2O_{ads} \tag{6-20}$$

$$CH_{3,ads} + 2\ O_O \rightarrow HCOO_O + 2\ H_{ads} \tag{6-21}$$

$$HCOO_O \rightarrow CO + HO_O + V_O^{\times} \tag{6-22}$$

$$H_2O_{ads} \rightarrow H_{ads} + OH_{ads} \tag{6-23}$$

Here CO, CO_2, H_2O, H_2, and, for example, formic acid can occur as reaction products. In the case of methane, the "oxidizing" effects at temperatures greater than 950 °C may also be due to the formation of hydroxyl groups at the surface.

6.1.6 Selective Gas Sensors

In the typical materials used for high-temperature oxygen-sensor systems such as $SrTiO_3$ or CeO_2, the sensitivities with respect to other gases are relatively weak. However, there are also metal oxides that under similar conditions show much more drastic changes in resistance, a typical example of such a metal oxide being Ga_2O_3 [15, 67, 68].

Figure 6-8 shows the conductivity behavior of a sputtered Ga_2O_3 sensor in reaction to reducing and oxidizing gases in an inert carrier gas (nitrogen). When the oxygen concentration is changed from 100% to 0.1%, the oxygen sensitivity described in Section 6.1.3 is observed

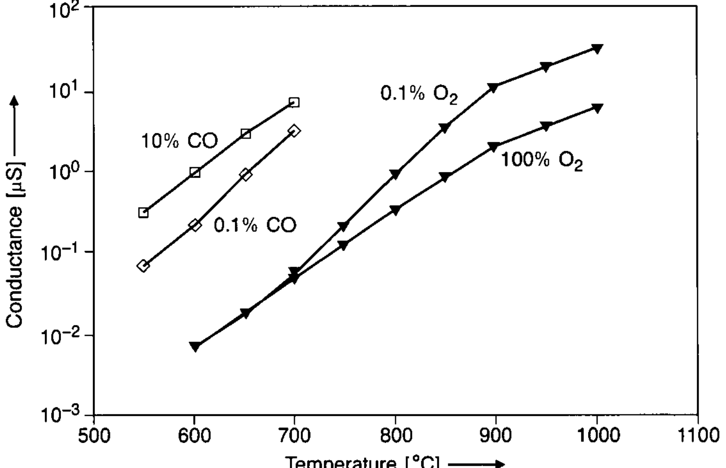


Figure 6-8. Conductivity of a 1 μm thick Ga_2O_3 thin film in the presence of reducing and oxidizing gases.

in the temperature range 850–1000 °C [15]. At lower temperatures the effect diminishes. Finally, below 700 °C it is no longer possible to detect any influence of oxygen concentration on conductivity behavior. However, in this range there is a great sensitivity to reducing gases [67]. Figure 6-8 clearly shows that in the presence of CO with concentrations in the 1% range in an inert carrier gas, the sensor resistance of sputtered Ga_2O_3 films decreases by up to two orders of magnitude.

However, the reaction of sputtered Ga_2O_3 sensors to H_2 in the temperature range 550–650 °C is considerably stronger than to CO. Figure 6-9 shows the corresponding sensor characteristic lines. For intermediate concentrations, both sensor characteristics can be described by an exponential law with an exponent of $-1/3$ (Freundlich's adsorption isotherm). So far, sensor effects have been looked at solely in inert carrier gases. However, the results of these studies can be regarded only as foundations for possible applications of sputtered Ga_2O_3 sensors. In order to test sensor effects under roughly realistic conditions, other carrier gases must be chosen.

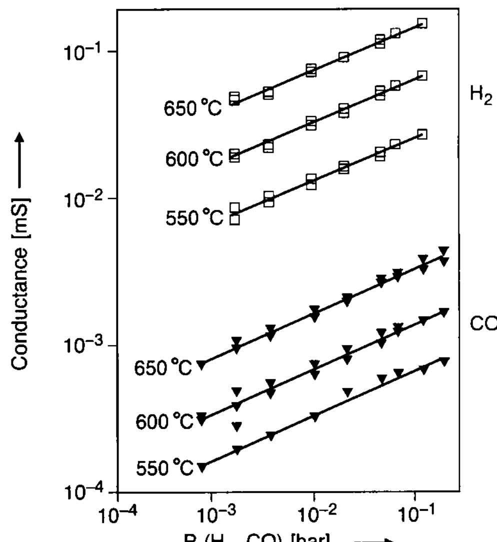


Figure 6-9.
Sensor characteristics of the Ga_2O_3 sensor for reducing gases.

Figure 6-10 shows the behavior of a Ga_2O_3 sensor with respect to CH_4, H_2, CO_2, CO (at a pressure of 5 mbar in each case), NH_3 (30 ppm) and NO (90 ppm) in a carrier gas consisting of wet synthetic air, for operational temperatures in the range 520–870 °C. Here the reciprocal of the cross-sensitivity (see Equation (6-12)) was used as a sensitivity measure for better comparison with one another and to represent these sensor effects. From the results shown it becomes clear that in this temperature range the Ga_2O_3 films have almost no sensitivity with respect to CO_2, NH_3, and NO. However, as in the case of inert gases (see Figure 6-9), here also a very distinct sensitivity with respect to H_2 can be seen. With the addition of 5 mbar of H_2 to wet air at 670 °C, the resistance of the Ga_2O_3 sensor changes by about three orders of magnitude. Hence the H_2 effect in moist air is even greater than that in the carrier gas N_2. The situation is different with the CO sensitivity. This likewise reaches its maximum at about 670 °C, but on the addition of 5 mbar of CO this leads only to a reduc-

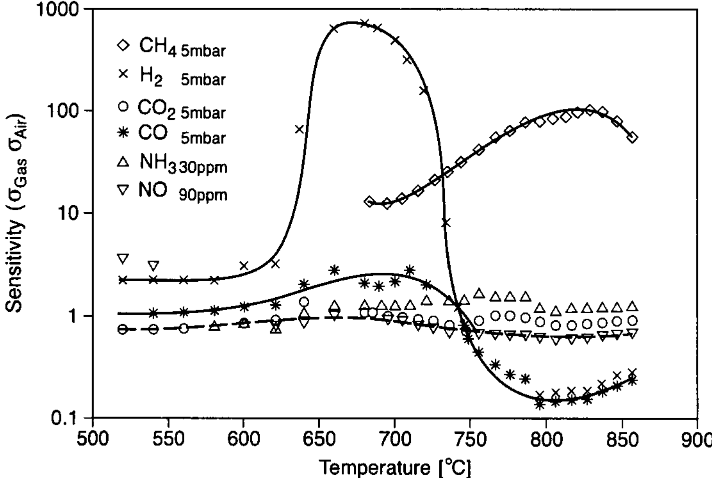


Figure 6-10. Temperature-programmed sensitivity of the Ga_2O_3 sensor for reducing gases (wet air).

tion in resistance by a factor of two, hence the sensitivity in moist air is considerably weaker than in the carrier gas N_2.

The results shown in Figure 6-10 also make it clear that at temperatures above 700 °C, the sensitivity of Ga_2O_3 sensors is higher by a factor of 10 with respect to hydrocarbons (here CH_4), which at higher temperatures (about 820 °C) even grows to a factor of 100. From a sensor temperature of 750 °C on, apart from the methane sensitivity no interaction with any other gas can still be seen, and the sensitivity with respect to CO manifests itself as an increase in resistance. Hence within this temperature range Ga_2O_3 sensors are suited for the detection of hydrocarbons. However, the important prerequisite for such a use is a constant oxygen partial pressure in the environment (see Figure 6-8).

From the results shown in Figure 6-10 it can be further concluded that Ga_2O_3 is a suitable sensor material for employment in hydrogen sensors at operational temperatures of 650–700 °C, since the effects with respect to other gases (even with respect to CH_4) are negligibly small in the carrier gas used. Here it shows a very significant advantage with regard to possible applications because of the fact that so far no detectable influence on H_2 sensitivity has been found from the oxygen concentration in the carrier gas used.

6.1.7 Sensor Arrays

In Sections 6.1.5 and 6.1.6 it became clear that a sputtered metal oxide is usually sensitive to a number of gases. However, for each sensor material there is generally a main sensitivity at a certain temperature. For this reason, in principle it would be possible to detect a number n of gas components in a gas mixture with a sensor array consisting of at least n sensors, which differ either in their sensor material or, when the sensor material is the same, in their working temperature. For no more than $n \geq 3$ different gas components, the chemical-physical effects occurring in the gaseous space, on the sensor surface, or within the interior of the sensor

become so complex that the usual mathematical or statistical analysis methods are no longer adequate to provide sufficiently meaningful results – the individual sensitivities with respect to the separate gases do not add in linear fashion. In Figure 6-11 we show as an example the sensitivity of a 1 at.% Ta-doped $SrTiO_3$ sensor (pure n-type conductor, see Section 6.3.1) at a working temperature of 850 °C with respect to 1% H_2, 1% CO, 1% CO_2, 0.4% CH_4, and all possible combinations of these in 5% O_2 (remainder N_2).

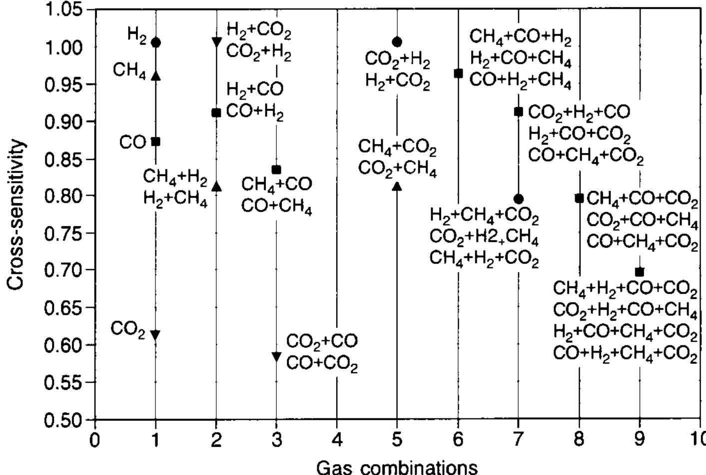


Figure 6-11. Cross-sensitivities of a 1 at.% Ta-doped $SrTiO_3$ film in relation to 1% H_2, 1% CO, 1% CO_2, 0.4% CH_4, and all possible combinations of these gases.

In order to be able generally to determine concentrations of the separate gas components for $n \geq 3$, the possibility arises of making the sensor signals of the individual sensors of an array available to a teachable system such as a neural network [69]. The individual neurons that are linked to one another in the neural network contain information consisting of non-linear functions f (decision function) and sets of parameters g_i. The activation of a neuron is prescribed in terms of an input pattern defined beforehand.

One way in which a neural network functions is for a decision to be made in an initial operational step as to whether a certain gas component is or is not present in the mixture. Then in a second operational step a rough decision is made on the basis of the signals of the separate sensors as to the range (low, intermediate, high) in which the concentrations actually lie of the gas components that are in fact present. Then only in a third operational step are the precise gas concentrations determined.

Other advantages to the use of neural networks in gas sensor systems are that despite an increase in the computational work required, the principles of pattern recognition are retained when the number of gas components increases (from n to $n + m$). In addition, with the use of neural networks a certain long-term drift of the individual sensors of the sensor array can be tolerated. Basically, neural networks can be implemented with the use of reasonably priced digital processors, permitting compact recognition systems to be constructed.

6.2 Temperature Sensors

6.2.1 Principle of Operation

In addition to the fact that the resistance of sputtered metal oxides is a function of their specific gaseous environment, at temperatures in excess of 600 °C these materials also show an exponential decrease in resistance with increasing temperature. Since as a rule the dependence on temperature of the resistance has a stronger effect than the interaction with certain gases, in gas-sensor systems that use metal oxides it is essential to eliminate this temperature dependence. This is done in most applications by using a combination of a temperature probe and a heating element. In order to establish a constant temperature in the region of the gas-sensitive film, enough heater power must always be employed so that the resistance value of the temperature sensor is maintained at a level corresponding to the desired temperature [70].

There are basically several ways to measure high temperatures up to 1000 °C. We probably can rule out in advance thermocouples implemented by thin- or thick-film techniques that could be applied to the substrate in the immediate vicinity of the gas-sensitive film. Here the kinds of problems familiar from the field of conventional thermocouples are clearly intensified once again – a low signal sensitivity per °C (about 30 μV), electromagnetic influences, cold-spot compensation, aging phenomena due to an in-diffusion of impurities, or a sensitivity to mechanical stress [71, 72]. It seems more promising to take advantage of the exponential resistance dependence of the metal oxides themselves by sealing them in gas-tight from the environment. This can generally be done by passivating a sputtered metal oxide with a gas-impermeable layer or by enclosing the sputtered film in a gas-tight reference volume [73]. However, both possibilities lead to certain problems. In the first case any possible interaction of the passivation with the metal oxide film must be excluded, since this can lead to resistance drifts, especially in long-term applications. In addition, the passivation must be as thin as possible, and thus it must be applied by either a thin- or thick-film technique, since the temperature sensors are supposed to see the same temperature as the gas-sensitive element attached in the immediate vicinity. This requirement then leads again to the problem of how to implement the passivation so that it is perfectly gas-tight, in order to freeze in a certain oxygen partial pressure within the metal oxide lying underneath. The final problem that then arises in this connection is the question of what oxygen partial pressure actually gets frozen in. The second possibility, namely to enclose the sputtered metal oxide in a reference volume, likewise leads to the problem that the temperature sensor may see a different temperature to that seen by the gas-sensitive film, especially when the gas flows are rapid. It is considerably more promising anyway to combine together on an elementary sensor different metal oxides with differing variabilities of their electrical conductivity, but with an identical dependence of their conductivities on the temperature, and to eliminate the temperature dependence by a difference measurement. This sort of temperature compensation is described in Section 6.3.1.

In the applications of high-temperature gas sensors developed so far, it has proved useful to design the temperature-sensitive sensor elements as sputtered or screen-printed metallic conducting paths. However, most metals are not suitable for high-temperature applications up to 1000 °C, since they either oxidize very easily in oxygen-containing atmospheres or have

melting points that are too low (eg, gold at 1063 °C). Thus, in the final analysis, only the metals of the platinum group (ruthenium, rhodium, palladium, iridium, osmium, and platinum) are left for this application. Of these noble metals with high melting points, platinum is one of the least prone to form oxides, and especially because of this property it is used the most frequently for making temperature sensors (eg, Pt 100) [74]. For the dependence on temperature of the resistance R_T in the range $0\,°C < T < 850\,°C$, the following relationship holds as an approximation:

$$R_T = R_0(1 + \alpha T + \beta T^2)$$

where R_0 designates the platinum resistance at 0 °C. The values of the temperature coefficients for platinum bulk material ($\alpha = 3.908 \times 10^{-3}\ K^{-1}$ and $\beta = -5.802 \times 10^{-7}\ K^{-2}$) are not reached in sputtered or screen-printed structures. For platinum temperature sensors made by thick-film methods, for example, one obtains maximum values of about $\alpha = 3.45(\pm 0.15) \times 10^{-3}\ K^{-1}$ and $\beta = -4.2(\pm 1.2) \times 10^{-7}\ K^{-2}$) [11]. Whith respect to both thin- and thick-film fabrication methods, these temperature coefficients show a strong dependence on the firing temperature or pre-aging temperature (see Section 6.1.2).

As a rule, the standard designs for a Pt-100 sensor made with thin- or thick-film technology have well defined properties only up to 850 °C, since even platinum further oxidizes to PtO_2 at the surface to some extent as early as from 600 °C on. However, this oxidation process is reversible, namely when the temperature sensor is cooled the oxide is again reduced. But above 850 °C, PtO_2 becomes volatile, which means that if the sensor is used for long periods of time above this temperature, material losses will continually take place, which ultimately entails a constant drift on the part of the temperature sensor [75]. When the temperature sensor is to be used for temperatures up to 1000 °C, it is important to inhibit these losses of platinum material. One promising possibility for doing this is to cover the platinum meander made with thin- or thick-film technology by a glass passivation layer (as thin as possible). By these means, and following prolonged seasoning for 500 h at 1000 °C, temperature sensors can be made with thin-film technology, for example, whose temperature accuracy at 1000 °C is ±15 °C [71]. These values cannot be achieved with temperature sensors of platinum that have been screen-printed and then likewise glass passivated, since it is more difficult to achieve a gas-tight seal because of the roughness of the burned-in paste.

6.2.2 Heating Structures

Basically the same problems arise for heating structures as for temperature sensors with respect to their long-term stability at high application temperatures. Therefore, it is advisable here also to use meander structures produced by a thin- or thick-film technology. As a rule, platinum is again used as the material, covered with a glass passivation to protect it against oxidation losses.

It is best to attach the heating structures for the primary sensors to the back side of the substrate used. Thereby, a more homogeneous heating of the gas-sensitive film is achieved, and in addition possible parasitic influences of the heater power on the sensor signal are suppressed. The heating meanders should if possible have a structure such that the entire gas-sensitive film is homogeneously distributed over the desired application temperature and the

temperature drops as linearly as possible from the termination of the gas-sensitive film to the contacts of the primary sensor. The latter is essential if a sensor is exposed to large temperature fluctuations in the environment, since the large changes in the heater output necessitated by this will then lead to the minimum possible mechanical stresses along the primary sensor.

In what follows, the design of a heating structure for oxygen sensors based on $SrTiO_3$ thin films is described, which can be used, for example, as a lambda probe in a motor vehicle. In these sensors, the homogeneity of the sensor temperature is of concern not only because of the dependence of the sensor characteristic line on temperature (see Section 6.1.3) but above all because of the dependence of the sensor response time to changes in the gas composition (see Section 6.1.4) [70].

In order to optimize the design of a heating structure for this type of application, the temperature distribution on an empirical starting structure was simulated using the finite-element program ANSYS [76]. In this simulation, not only the geometric and physical boundary conditions but also the specific properties of the materials making up the primary sensor (eg, thermal conductivity) are considered. The primary sensor is broken down into individual elements, and in order to determine the model data for these functional parts, not only the geometry of the individual element but also all three types of heat transfer (thermal conduction, convection, and thermal radiation) are considered. It is necessary to consider convection because the gas mixture of the environment that is heated by the heating element passes along the sensor element and is constantly being replaced by a colder gas mixture flowing in behind it. Since according to the Stefan-Boltzmann law the heat flux given off by thermal radiation is proportional to T^4, substantial temperature losses also arise especially because of thermal radiation.

Unknown determining variables, such as the thermal convection coefficient of the surface of an Al_2O_3 substrate or of a screen-printed platinum structure, were varied in a number of calculations for different heating outputs until the temperature distribution determined for the starting structure by infrared and thermography films agreed with the distribution calculated via the ANSYS model.

With the known and the empirically determined parameters, the geometry of the heating meanders was varied until the boundary conditions described at the beginning could be achieved for the desired temperature distribution. The intrinsic suitability of the model had been demonstrated only when these parameters could be applied to modified structures and

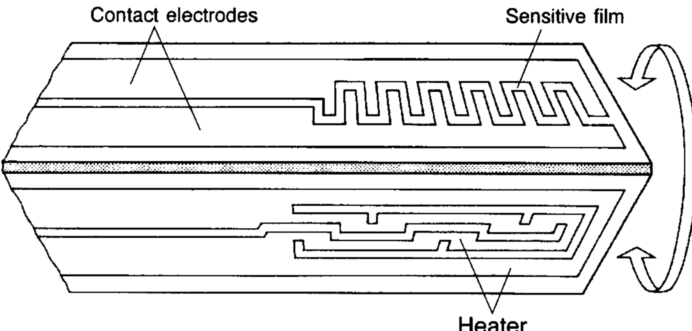


Figure 6-12. Schematic sensor structure of an oxygen sensor.

then an agreement between simulation and experiment was again found [70]. Figure 6-12 shows schematically a primary sensor of the type that can be used with gas-sensitive $SrTiO_3$ thin films. The front side shows the screen-printed comb-type electrodes (material: platinum) and on the rear side is the heating meander, likewise implemented by the silk-screen printing process.

The primary sensor shown in Figure 6-12 does not have a temperature-sensing element. It is possible to dispense with this temperature sensor when the heating meander can also take over its function. In this case, the resistance of the heating meander itself is used as a temperature sensor via a four-point method, or in a pulsed heating operation the resistance of the heater is measured in the heating pauses.

6.3 Applications

6.3.1 Selective Oxygen Sensors on the Basis of Variously Doped $SrTiO_3$ Films

As was the case for motor-vehicle exhaust gas, the oxygen partial pressure $P(O_2)$ in the exhaust gas of large- and small-scale furnace systems is likewise a measure of the stoichiometry of the combustion that has taken place. Thus the $P(O_2)$ values determined by an oxygen sensor are the necessary basis for continuing optimization of the burner setting. Typical $P(O_2)$ values in the exhaust gas of furnace systems lie in the range 0.001 bar $< P(O_2) <$ 0.1 bar. Since for this application what is important is only to determine the oxygen partial pressure as accurately as possible, cross-sensitivities of the gas-sensitive material should not be present. Because $SrTiO_3$ has proved to be a stable material in extremely inhospitable atmospheres (see its application as a lambda probe in Section 6.3.2), it is also favored for applications in the exhaust gas of furnace systems.

From the behavior of electrical conductivity as a function of oxygen partial pressure as described in Section 6.1.3, it becomes clear that an undoped sensor material could lead to ambiguous measurement signals in the above application field because of its conductivity minimum. However, by means of the sandwich method (see Section 6.1.2), it has proved feasible to embed dopants in an extremely well defined way into the base material and to distribute them homogeneously there via a subsequent annealing process (40 h at 1100 °C).

Especially suitable as possible acceptors for $SrTiO_3$ are trivalent transition metal ions whose ionic radii are of the same order as that of the Ti^{4+} ion. Aluminium oxide (Al_2O_3), gallium oxide (Ga_2O_3), and chromium oxide (Cr_2O_3) have proved to be particularly suitable for making acceptor-doped films in the sandwich method [34]. This embedding of acceptors can be verified either by analytical means (SIMS = secondary ion mass spectrometry) or via measurements of the electrical conductivity as a function of the oxygen partial pressure. Figure 6-13 shows the results of the SIMS analysis of a 0.45 at.% Ga-doped $SrTiO_3$ film (on an Si substrate) before and after the annealing process. In this representation, the sputter time that elapses during the analysis is plotted on the x-axis. In the non-annealed sample, after about half of the total sputter time (corresponding to half the film thickness) that Ga_2O_3 intermediate film is clearly discernible. The broadening of the $^{69}Ga^+$ peak towards higher

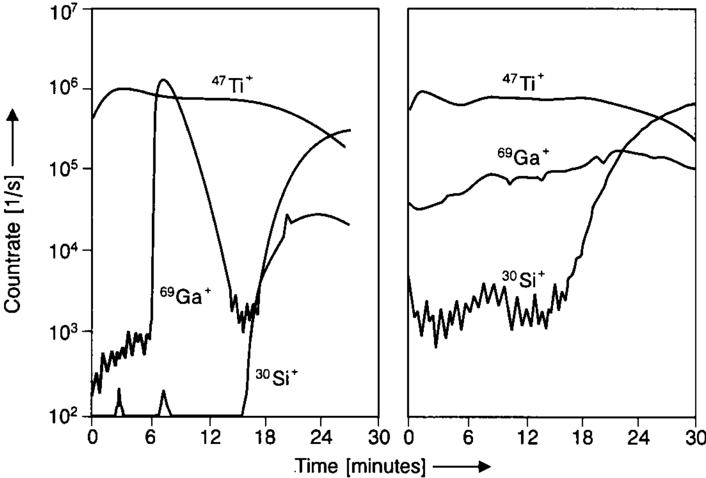


Figure 6-13. SIMS analysis of a 0.45 at.% Ga-doped $SrTiO_3$ film (before and after annealing).

sputter times is due to the sputter crater formed during the analysis. After the annealing process, the Ga doping is distributed nearly homogeneously over the entire film.

Conductivity investigations above all have made it possible to establish that in the sandwich process a maximum acceptor concentration of 1 at.% is actually embedded in the $SrTiO_3$ grains [11]. These acceptors cause the minimum in the electrical conductivity to shift to a lower $P(O_2)$ value by about one power of ten.

Ta^{5+}, Nb^{5+}, and W^{6+} ions are feasible donors for the doping of sputtered $SrTiO_3$ thin films [47]. All of these ions have radii of the same order of magnitude as the Ti^{4+} ion. The reason for the choice of the dopant oxides mentioned is that at high temperatures they are chemically stable especially in the presence of moisture, which is not the case with the oxides of the lanthanide series (La_2O_3 or Y_2O_3), whose metal could be incorporated into the lattice at an Sr site.

Analytical investigations show that at comparable concentrations, donors are not nearly as well incorporated into the $SrTiO_3$ lattice as acceptors are. The reason for this must probably be attributed to the potential relationships at the surface of the grains of the base material $SrTiO_3$ described in Section 6.1.5. However, despite the apparently smaller incorporation of donors, it is all the more astonishing that donor concentrations of just a few tenths of a percent are already sufficient (in the case of a Ta doping, 0.3 at.% is enough) to shift the minimum of the electrical conductivity to oxygen partial pressures of $P(O_2) > 1$ bar (Figure 6-14). Then, starting with a dopant concentration of about 1 at.%, given a high $P(O_2)$, an approximately linear relationship results between log σ and log $P(O_2)$, with a slope of $1/m = -1/4$.

When the doped $SrTiO_3$ thin films are made in the sandwich process, then it turns out that within the limits of measuring accuracy, the dependence of the electrical conductivity σ on temperature (characterized by the thermal activation energy E_A) is unchanged owing to the incorporation of the dopant, in contrast to the undoped material, at temperatures $T > 800$ °C. For undoped $SrTiO_3$ a thermal activation energy of $E_A = 1.4(\pm 0.1)$ eV is found for 0.01 bar $< P(O_2) < 1$ bar. Figure 6-15 shows the dependence of resistance on temperature for an

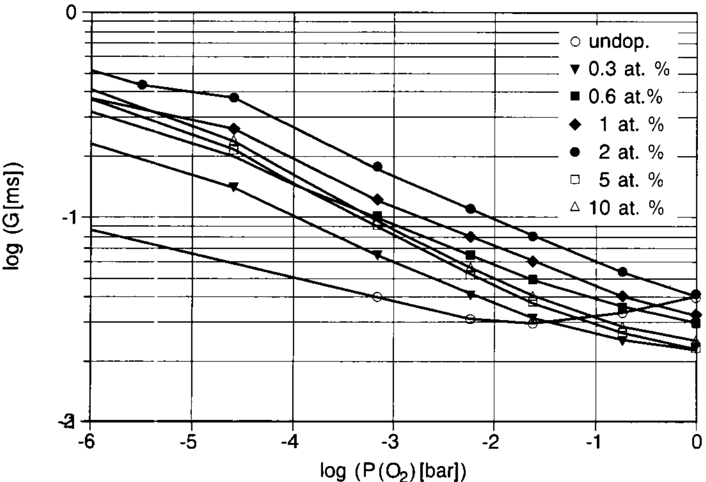


Figure 6-14. Conductivity behavior of undoped and variously Ta-doped $SrTiO_3$ films.

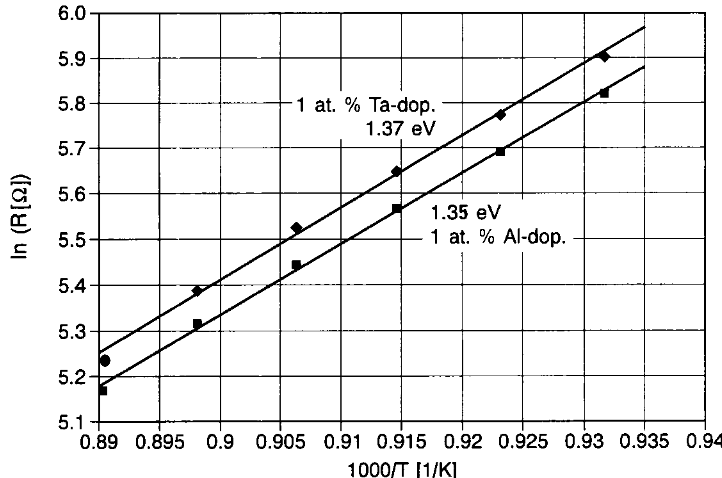


Figure 6-15. Arrhenius diagram of the resistance values of a 1 at.% Al-doped and a 1 at.% Ta-doped $SrTiO_3$ film.

$SrTiO_3$ film doped with either 1 at.% Ta or 1 at.% Al ($800\,°C < T < 850\,°C$) at $P(O_2) = 1$ bar.

If one designates with Δ the factor between the resistance (R_{nL}) of a donor-doped and the resistance (R_{pL}) of an acceptor-doped $SrTiO_3$ film ($R_{nL} = \Delta R_{pL}$) at a certain oxygen partial pressure P_L (L = air), then it can be shown that under the assumption of a like thermal activation energy also at other oxygen partial pressures P the following holds:

$$\log R_n - \log(\Delta R_p) = (\log P - \log P_L)\,(1/n_n - 1/n_p) \tag{6-24}$$

where n_n and n_p are the slopes of the curves of the donor-doped and acceptor-doped material, respectively, and R_n and R_p are the corresponding resistances at the oxygen partial pressure P. Since n_n and n_p have opposite signs, this means that by subtracting the resistances between the p-conducting and n-conducting film not only is the temperature eliminated, but in addition the measuring effect is also doubled compared with a simple sensor. Figure 6-16 shows both the theoretical signal of a difference sensor that would be expected for the optimum slopes of $+1/4$ and $-1/4$, and also the values measured for 1 at.% Al-doped and 1 at.% Ta-doped $SrTiO_3$ films. From the results shown in Figure 6-16 it becomes clear that although the theoretical slope for the difference sensor of 1/2 is not obtained, the measured value and the good linearity of the measured curve make it appear that there are good chances that the difference sensor described here can fulfil the requirements placed on oxygen sensors used in the exhaust gas of furnace systems. The deviation of the measured from the theoretical values may be due to the less steep slopes of the acceptor-doped materials in the p-conducting region compared with their theoretical values. The reason for this lies again with the space-charge zones at the grain boundaries that are seen especially in thin films (see Section 6.1.5).

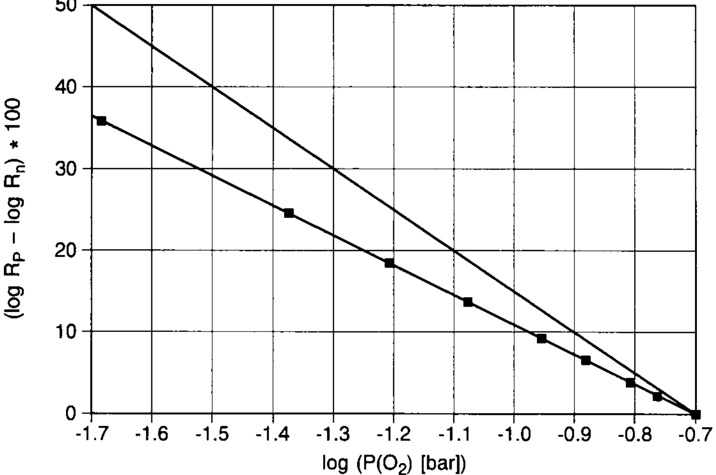


Figure 6-16. Theoretical and measured signal of a difference sensor based on n- and p-type $SrTiO_3$ films.

A simple but nevertheless extremely effective way to suppress the cross-sensitivities of oxygen sensors with respect to other gases as described in Section 6.1.5 is to surround the oxygen sensor with a membrane permeable only to oxygen, in the form of oxygen ions. Figure 6-17 shows schematically a cross section through a sensor structure that can form the basis for a measuring principle designed to eliminate cross-sensitivities.

The membrane permeable to oxygen ions is a solid electrolyte (eg, stabilized zirconium oxide (ZrO_2), whose electrodes on both sides (made of, eg, platinum) have been short-circuited. The volume formed by the solid electrolyte and the substrate is sealed on the front and back sides. Neither the solid electrolyte nor the electrodes attached on the two sides are in direct contact with parts of the oxygen sensor.

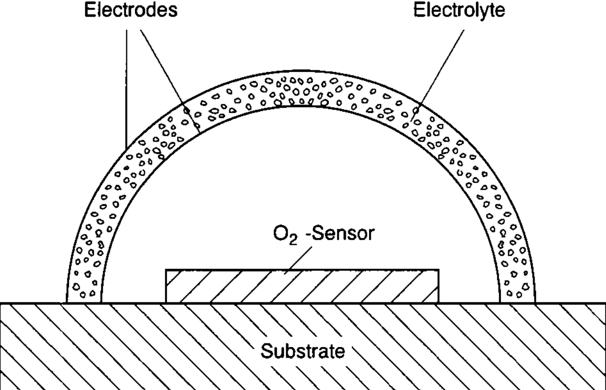


Figure 6-17.
Measurement principle for the elimination of cross-sensitivities.

Figure 6-18 shows, by the example of the cross-sensitivity to methane of the 1 at.% Al-doped component of the difference sensor, that the sensor structure shown in Figure 6-17 fulfils its purpose. As a carrier gas, a gas mixture consisting of 5% O_2 and 95% N_2 was used for the measurement shown in Figure 6-18. When 0.4% CH_4 was added, the N_2 percentage was correspondingly reduced, so that the preassigned oxygen percentage remained constant during the entire measurement. Through the results shown in Figure 6-18 it is essentially proved that cross-sensitivities can be eliminated in the field of oxygen sensors.

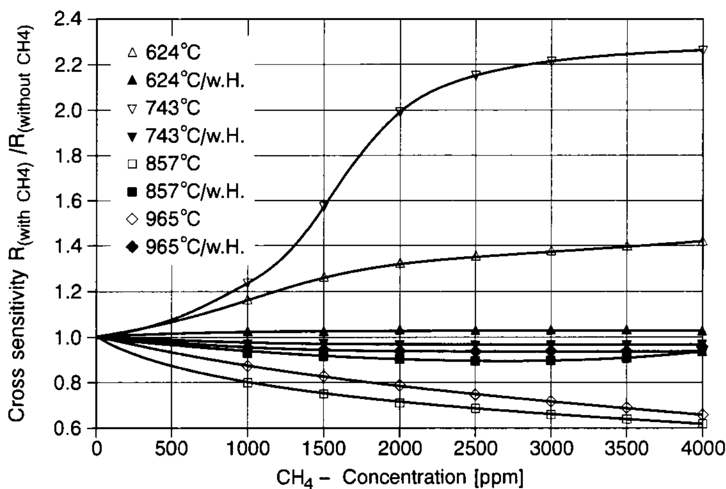


Figure 6-18. CH_4 cross sensitivity of a 1 at.% Al-doped $SrTiO_3$ film without and with a solid electrolyte ("/w. H.").

6.3.2 Oxygen Sensors for Applications in the Automotive Field

Oxygen sensors are used in motor vehicles in order to increase the efficiency of combustion and to monitor pollutant emissions [77]. So far, two types of sensors have become popular for this application: resistive oxygen sensors based on TiO_2, and ion-conducting oxygen sen-

sors based on ZrO_2 stabilized by CaO or Y_2O_3. Whereas the currently used TiO_2 sensors have already been implemented in a screen-printed form, a similar development for sensors based on solid electrolytes has just begun. The principles of the functioning of the two sensor materials have been adequately described in the literature and will not be repeated here [78-82]. However, for purposes of future engine-management systems it turns out that the abrupt change in the sensor signal at $\lambda = 1$ and the relatively slow reaction of the sensor signal to changes in the exhaust-gas composition are distinct disadvantages. This is explained below.

The most important device for reducing pollutant emissions from internal-combustion engines is the regulated three-way catalytic converter. This catalytic converter is intended to support simultaneously three different reduction and oxidation reactions:

$$CO + 1/2\ O_2 \rightleftharpoons CO_2 \tag{6-25}$$

$$C_xH_y + (x + y/4)\ O_2 \rightleftharpoons x\ CO_2 + y/2\ H_2O \tag{6-26}$$

$$2\ NO \rightleftharpoons N_2 + O_2 \tag{6-27}$$

This is possible only with a strictly stoichiometric mixture, since otherwise oxidizable or reducible components are left over. The composition of the air/fuel mixture is described by the stoichiometry number λ:

$$\lambda = (m_{Air}/m_{Fuel})_{real}/(m_{Air}/m_{Fuel})_{stoich} \tag{6-28}$$

A slight change in the λ value from the optimum value around $\Delta\lambda \gg 0.001-0.002$ already reduces the degree of conversation of the catalytic converter from about 98% to between 80% and 85%. However, the lambda probes used up to now have been designed in such a way that they have an abrupt change in their sensor signal at $\lambda = 1$. Moreover, it is assumed that the combustion in the engine is optimum when the particular lambda probe detects a lean mixture (air excess) during half of its measurement cycle and a rich mixture (fuel excess) the other half of the time. The lack of information about the extent to which the combustion happens in the rich or in the lean lambda region ultimately leads to the conclusion that the present lambda probes cannot fulfil the requirement for optimum performance by such catalytic converters. In order to fulfil future requirements with respect to reduced pollutant emissions of internal-combustion engines, sensors are needed that have response times of less than 10 ms and are able to measure the exhaust gas on a cylinder-selective basis as linearly as possible within a wide range around $\lambda = 1$.

As already becomes clear from the results shown in Section 6.1.4, sputtered $SrTiO_3$ thin films are able to detect changes in oxygen partial pressures within less than 10 ms at a certain application temperature. In addition, these structures can be used in temperatures up to 1100 °C without interactions occurring with the substrate (Al_2O_3) or electrode materials (Pt) used. The latter fact means that these sensor materials can be placed within the exhaust gas of internal-combustion motors very near the outlets of the separate cylinders. This is necessary because in the immediate vicinity of the discharge valves the exhaust-gas packets of the separate cylinders are intermixed the least.

Figure 6-19 shows the resistance behavior of an $SrTiO_3$ thin film for various engine speeds as a function of the lambda value at a temperature of 900 °C, which is held constant at all

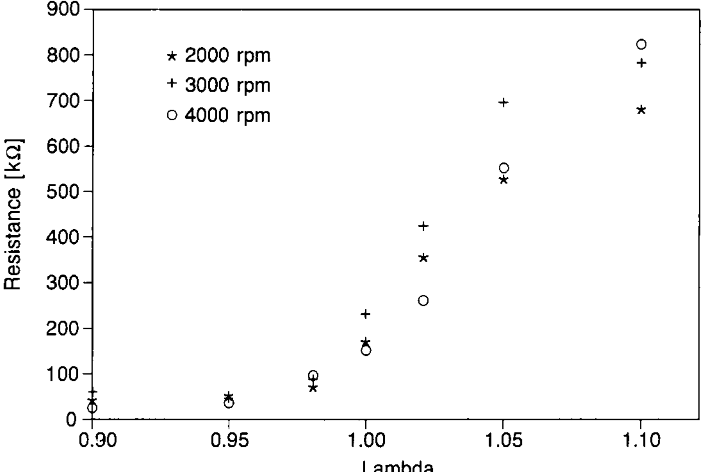


Figure 6-19. Sensor characteristics for 2000, 3000, and 4000 rpm, each at a temperature of 900 °C produced by active heating.

times by a temperature regulation mechanism consisting of a temperature sensor and heating element (see Section 6.2). The results depicted make it clear that the $SrTiO_3$ sensor material shows a definite dependence of the sensor signal on the lambda value in the range $0.9 < \lambda < 1.1$. In addition, when the temperature is held constant this signal is sufficiently linear regardless of the engine speed, at least in the range $0.98 < \lambda < 1.02$. If one compares the sensor signal shown in Figure 6-19 with the measured variations of the electrical conductivity as a function of the oxygen partial pressure (see Section 6.1.3), it becomes clear that for measurements of oxygen concentrations in the vehicle exhaust gas it is exclusively the n-type conduction region that comes into play [2, 58].

Figure 6-20 shows by way of example the variation with time of the sensor signal obtained when the $SrTiO_3$ sensor is installed in one of the two exhaust-gas lines of a BMW engine [11]. In this case the lambda probe sees the exhaust-gas packets of cylinders 1-3, where the results shown in Figure 6-20 are based on preselected values for the individual cylinders of $\lambda = 1.1$

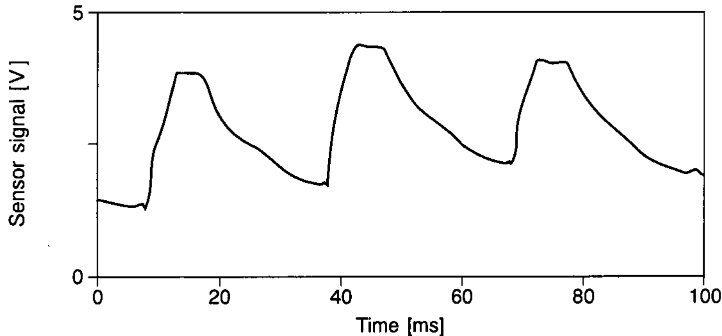


Figure 6-20. Variations in the sensor voltage for $\lambda = 1.1$ (cylinder 2) and $\lambda = 0.9$ (cylinders 1 and 3).

(cylinder 2) and $\lambda = 0.9$ (cylinders 1 and 3). At a speed of 4000 rpm, precisely ten exhaust-gas packets within 100 ms can be expected. The $SrTiO_3$ lambda probe is not only able to see this number of exhaust-gas packets, but in addition is able to differentiate clearly the lean exhaust gas of cylinder 2 from the rich exhaust-gas packets of cylinders 1 and 3. In the last analysis, this also represents a confirmation of the response times measured in Section 6.1.4.

6.4 Si Planar Pellistor

6.4.1 Introduction

The thermo-catalytic principle has been used for many years for the detection of combustible gases and vapors. These sensors, known as pellistors, detect the heat of a catalytic reaction. The sensing elements have to be operated at temperatures of 300–600 °C. Conventionally built pellistors are highly developed and already used in fields such as explosion warning, workplace safety and process control. Unfortunately, these pellistors require as much as 1 W of electrical heating power and have a slow thermal response because of the high thermal capacity of the sensing element.

Future applications will be focused, for example, on environmental monitoring and portable sensor systems. To meet the requirements of these applications, sensor systems have to become smaller in size and the power consumption must be reduced. Silicon technology is known to be capable of meeting these requirements, and silicon versions of pellistors have been proposed and developed by several groups [83–86].

Methods of silicon micromachining are used to miniaturize the sensor area; thin membranes for mechanical suspension offer excellent thermal insulation. Thus the power consumption can be reduced by a factor of 10. Batch processing reduces manufacturing tolerances and production costs. Figure 6-21 shows the cross section of a Si planar pellistor.

The Si planar pellistor shows a fast thermal response because of its small thermal time constant, in the range of a few milliseconds. This makes fast temperature modulation possible [87],

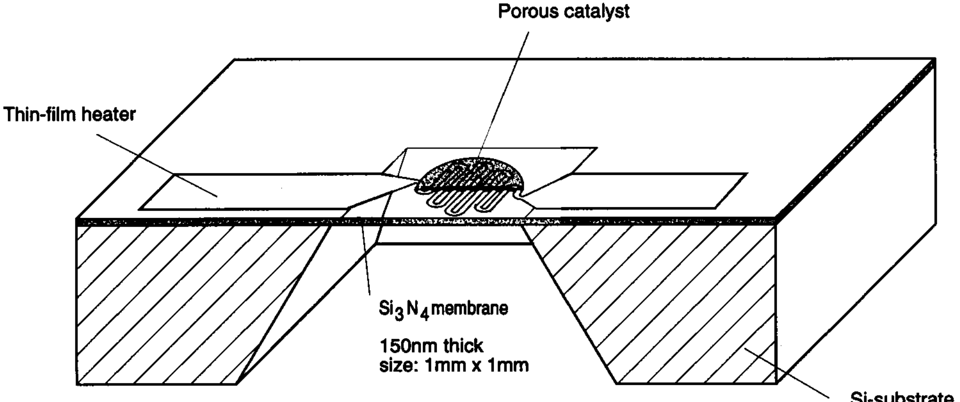


Figure 6-21. Si planar pellistor (cross-sectional view).

a new mode of operation with reduced drift effects. Figures 6-22 and 6-23 explain the different operating modes.

Single pellistors generally have a poor selectivity in detecting different gases. Selectivity can be achieved by using sensor arrays and pattern recognition methods [88]. Si planar pellistor arrays can easily be integrated on one chip whereas conventional pellistor arrays are difficult to handle.

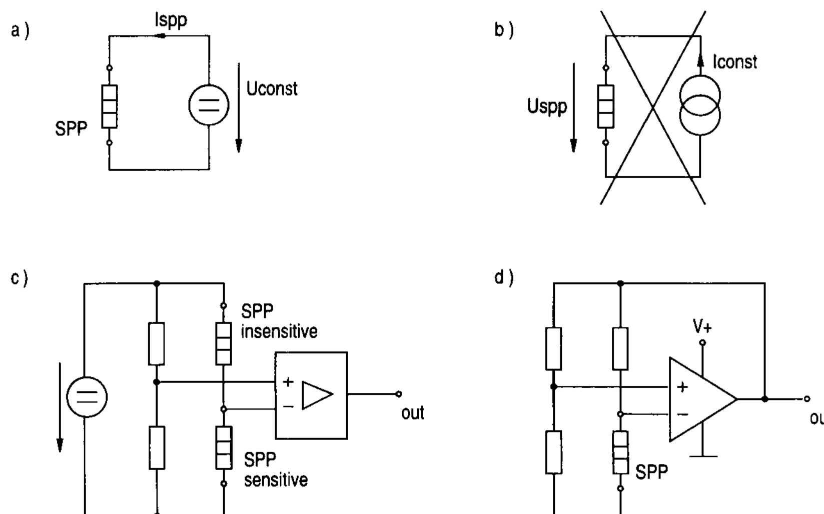


Figure 6-22. Operating circuits.

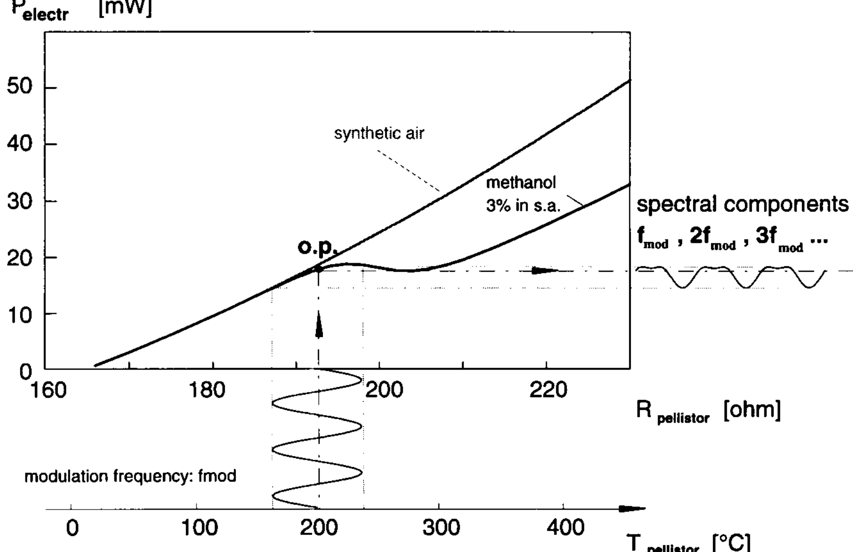


Figure 6-23. Sinusoidal temperature modulation.

6.4.2 Function Principle

An electrically heated noble metal catalyst is exposed to the gas atmosphere. Chemical power, caused by the oxidation of combustible gases on the catalytic surface, imbalances the equilibrium of electrical heating and thermal loss. Depending on the mode of operation, this leads either to an increase in temperature for a fixed electrical heating power or a reduced need in electrical heating power for a fixed temperature.

In both cases the sensor signal is achieved by measuring voltage and current of the sensor, because the heater resistance (or an additional probe resistor) is temperature dependent. The chemical power P_{chem} is defined as the difference in the electrical heating power with and without combustible gas, while keeping the temperature constant. Depending on the operating temperature of the catalyst, either the reaction rate or gas transport determines the chemical power.

Since catalytic reactions require a certain activation energy, the reaction rate (assuming a constant gas supply) follows the Arrhenius law:

$$k = k_0 e^{-\frac{E_A}{RT}}$$

where E_A = activation energy (J mol^{-1}), k = rate of reaction (ms^{-1}), R = gas constant (JK^{-1}), and T = temperature (K). At low temperatures the reaction rate is limited as given above. At higher temperatures the transport of gas to the catalyst (by diffusion) limits the reaction. Combining both effects results in the chemical power:

$$P_{chem} = A_{\mathrm{eff}} C \Delta H \frac{1}{\frac{1}{k} + \frac{1}{\beta}}$$

where C = concentration of gas (mol m^{-3}), A_{eff} = effective catalyst surface area (m^2), ΔH = enthalpy of reaction (J mol^{-1}) and $\beta = \beta T^x \; 1 \leq x \leq 2$ (ms^{-1}) = coefficient of material transport.

In Figure 6-23 the chemical power can be seen as the difference between the curves with and without gas.

6.4.3 Structure and Technology

The source material of the Si planar pellistor is a ⟨100⟩ silicon wafer with an additional 150 nm thick LPCVD Si_3N_4 (silicon nitride) layer on both sides. The Si_3N_4 on the rear side is first structured by plasma etching to form an etch mask for the anisotropic silicon etching process. This is done in 33% KOH solution at 80 °C. The etching stops at the front side Si_3N_4, forming thin membranes which are well suited for carrying the hot sensing areas. Their good lateral thermal insulation is essential for low-power operation.

The heater metalization is made of sputtered platinum 100–200 nm thick, with 10 nm of tantalum as an adhesive layer. The lift-off technique is used to transfer the heater structure. The temperature coefficient of the heater is $3 \cdot 10^{-3}$ k^{-1}. The temperature profile along the

sensing area should be flat. Inhomogeneity results in irregular reaction onset and shorted lifetime owing to hot spots. In simple designs the thin film surface of the heater is used for catalytic oxidation. As the chemical power is a function of the effective surface area, better sensitivity can be achieved by using porous catalysts.

Different methods for obtaining a thin but porous catalytic layer have been tested. The best results have been achieved by applying γ-alumina as a porous carrier first. This is done either by silk-screen printing or airbrush spraying. In a second step, noble metal is introduced into the pores in liquid form by impregnation with H_2PtCl_6 (hexachloroplatinic acid) solution. Increasing the temperature induces thermal decomposition, leaving the platinum inside the pores.

In addition to porous catalysts, advanced designs have separate heaters and temperature probes, which lie on top of each other spaced by isolating by layers such as Si_3N_4 or Al_2O_3.

6.4.4 Operating Modes

All operating modes are based on the fact that the heater resistance (or temperature probe resistance) is determined by the sensor temperature. Figure 6-22 shows the operating circuits.

a. Constant heater voltage

In this mode the Si planar pellistor is connected to a voltage source and the current is measured. Thermal stability is achieved at low concentrations. Problems occur at high concentrations when the chemical power is of the same order of magnitude as the electrical heating power; the sensor is overheated and destroyed. Drift of the heater resistance directly influences the sensor signal.

b. Constant heater current

This is not recommended because of thermal instability (positive feedback). A temperature increase causes an increase in electrical heating power ($P_{el} = I^2R$) and vice versa.

c. Bridge operation with compensating element

An electrically equal but gas-insensitive compensating element is connected in series. This is a commonly used mode for conventional pellistors. Drift due to temperature variations can be compensated, long-term drift caused by ageing processes can be reduced. Under reasonable conditions no problems with overheating have been reported.

d. Constant-resistance (temperature) mode

A feedback bridge circuit keeps the heater resistance at a fixed value, thus reducing the electrical heating power when combustible gases are present. Thermal stability is guaranteed even at high concentrations. Operation at constant temperature results in good proportionality of P_{chem} to the concentration over a wide range, whereas in modes a–c a temperature increase causes nonlinearity. Adding gas-insensitive compensating elements in a second bridge can reduce drift effects.

In the modes a–d the sensor signal represents a small change in a large value. To obtain finally a zero signal for zero gas concentration, an accurate reference value has to be subtracted. High temperatures together with high current densities cause ageing effects in the metallization, resulting in a slightly drifting heater resistance. Even that is harmful to the zero signal for zero gas relationship which is necessary for high sensitivity. The detection limits

with Si planar pellistors have been found to be as low as 50 ppm (methanol), assuming that zero adjustment is done just before each measurement. If a detection limit of 100 ppm is required, zero adjustment must be done at least every half hour, which causes problems in standalone applications. As can be seen in Figure 6-24, the period between zero adjustments determines the detection limit.

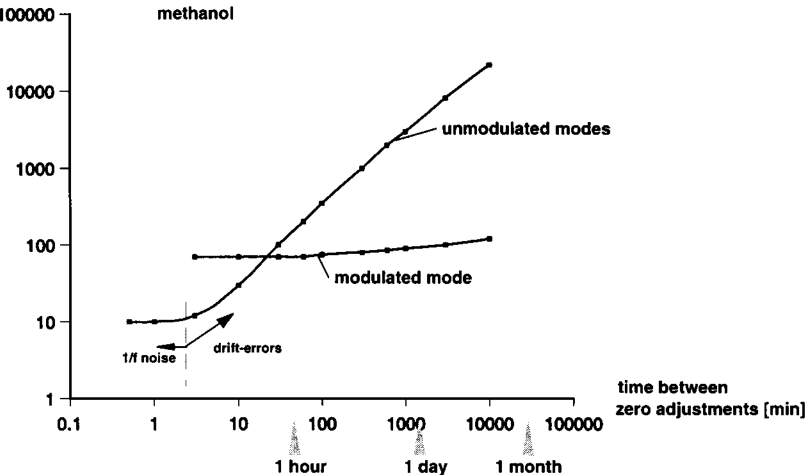


Figure 6-24. Comparison of modulated and unmodulated modes.

e. Modulated mode

A more sophisticated way to deal with drift and noise problems is to apply ac measurements [87]. The low thermal capacity of the sensor structure makes temperature modulation possible at frequencies up to f_{mod} = 100 Hz.

Figure 6-23 shows the typical behavior of a Si planar pellistor in air and in combustible gases. The most important thermal losses are linear with temperature so that (in air) the temperature (resistance) is a linear function of the electrical heating power P_{el}. For this reason, sinusoidal temperature modulation, which is done in a control circuit, is achieved by a (nearly) sinusoidal variation of the electrical heating power. In a combustible gas a sinusoidal temperature modulation will result in a strongly distorted electrical heating power signal, as the control circuit keeps track of the temperature. Signal evaluation is done by analysing the harmonic spectral components of P_{el} at $2f_{mod}$ and $3f_{mod}$.

The strongest distortions are generated when the operating point is near the onset of the catalytic reaction. This temperature is gas dependent.

The great advantage of the modulated mode is the reduction of the zero-point drift. The zero signal for zero gas relationship is fulfilled without a drifting reference value to be subtracted.

In addition to the drift errors, there is $1/f$ noise at low frequencies. This noise has been found to be influenced by the current density in the thin film metalization. At f_{mod} the $1/f$ noise densitiy is reduced.

Sensitivity is determined by the quality of modulation and the resolution of spectral analyses. Drift effects cause virtually no deterioration of sensitivity. Good results have been achieved by using digital modulation and calibration. Detection limits as low as 100 ppm (methanol) have been obtained over several days without zero adjustment. Figure 6-24 compares the modulated and unmodulated modes. Obviously the modulated mode is better suited for stand-alone applications.

6.5 References

[1] Patterson, D. J., Henein, N. A., *Emissions from Combustion Engines;* Ann Arbor, MI: Ann Arbor Science Publishers, 1972, p. 347.
[2] Meixner, H., Gerblinger, J., Fleischer, M., *Sensors Actuators B* **15-16** (1993) 45-54.
[3] *Technische Druckschrift (KH/VDT);* Stuttgart: Robert Bosch GmbH, 1980.
[4] Kleitz, M., Siebert, E., Fabry, P., Fouletier, in: *Sensors, a Comprehensive Survey,* Vol. 2, Part I, Göpel, W., Hesse, J., Zemel, J. N. (eds.); Weinheim: VCH, 1991, Chap. 8, pp. 341-428.
[5] Logothetis, E. M., *Ceram. Eng. Sci. Proc.* **1** (1980) 281.
[6] Logothetis, E. M., *Sensors Actuators* **4** (1983) 330.
[7] Müller, A., *Dissertation,* Universität Karlsruhe, VDI-Fortschrittsberichte Reihe 15, No. 54; Düsseldorf: VDI-Verlag, 1988.
[8] Müller, A., Härdtl, K. H., *Appl. Phys. A* **49** (1989) 75-82.
[9] Schönauer, U., *Tech. Messen* **56** (1989) 260-263.
[10] Gerblinger, J., Meixner, H., *J. Appl. Phys.* **67** (1990) 7453-7459.
[11] Gerblinger, J., *Dissertation,* Universität Karlsruhe, 1991.
[12] Meixner, H., Lampe, U., Gerblinger, J., in: *Proc. SENSORS '93,* Kongressband VI, pp. 23-32.
[13] Gerblinger, J., Meixner, H., *Fortschr. Dtsch. Keram. Ges.* **7** (1992) 145-153.
[14] Gnörich, A., *Dissertation,* Universität Erlangen, VDI-Fortschrittsberichte Reihe 15, No. 86; Düsseldorf: VDI-Verlag, 1988.
[15] Fleischer, M., *Dissertation,* TU München, 1992.
[16] Göpel, W., Schierbaum, K. D., in: *Sensors, a Comprehensive Survey,* Vol. 2, Part I, Göpel, W., Hesse. J., Zemel, J. N. (eds.); Weinheim: VCH, 1991, Chap. 9, pp. 429-466.
[17] Gentry, S. J., Jones, T. A., *Sensors Actuators* **10** (1986) 141-163.
[18] Riegel, J., *Dissertation,* Universität Karlsruhe, VDI-Fortschrittsberichte Reihe 8, No. 194; Düsseldorf: VDI-Verlag, 1989.
[19] Quenett, C., *Diplomarbeit,* Inst. Technologie der Elektrotechnik, Universität Karlsruhe, 1990.
[20] Gerblinger, J., Lampe, U., Meixner, H., Perczel, I. V., Gieber, J., *Sensors Actuators B* **18-19** (1994) 529-534.
[21] Westwood, W. D., in: *Physics of Thin Films,* Vol. 14, Francombe, M. H., Vossen, J. L. (eds.); Boston: Academic Press, 1989, pp. 1-79.
[22] Betz, G., Wehner, G. K., in: *Sputtering by Particle Bombardment II,* Behrisch, R. (ed.); Berlin: Springer, 1983.
[23] Harrison, Jr., D. E., *Radiat. Effects,* **70** (1983) 1-64.
[24] Reichl, H., *Hybridintegration;* Heidelberg: Hüthig, 1988.
[25] Ikegami, A., Arima, H., Iwanaga, S., Kaneyasu, M., in: *Proc. 4th European Hybrid Microelectronic Conf., Copenhagen,* 1983, pp. 211-218.
[26] Spear, K. E., Dirkx, R. R., *High Temp. Sci.* **27** (1990) 107-129.
[27] Gabovich, M. D., Pleshivtsev, N. V., Semashko, N. N., *Ion and Atomic Beams for Controlled Fusion and Technology;* New York: Plenum, 1989.
[28] Hauffe, W., in: *Sputtering by Particle Bombardment III,* Behrisch, R., Wittmaack, K. (eds.); Berlin: Springer, 1991.

[29] Chu, W. F., Fischer, D., in: *Biosensors – Applications in Medicine, Environmental Protection and Process Control,* Schmid, R. D., Scheller, F. (eds.), GBF Monographs, Vol. 13; Weinheim: VCH, 1989, pp. 179-182.
[30] Kelly, R., *Surf. Sci.* **100** (1980) 85-107.
[31] Galkute, L., Pranevivius, L., Zubauskas, G., *Nucl. Instrum. Methods Phys. Res.* **B21** (1987) 46-48.
[32] Gerblinger, J., Meixner, H., *Silicates Ind.,* No. 1-2 (1994) 31-38.
[33] Tomasshpol'skii, Y. Y., *Russ. Chem. Rev.* **45** (1976) 998-1027.
[34] Gerblinger, J., Meixner, H., *Sensors Actuators B* **6** (1992) 231-235.
[35] Fleischer, M., Hanrieder, W., Meixner, H., *Thin Solid Films* **190** (1990) 93-102.
[36] Fleischer, M., Meixner, H., *Sensors Actuators B* **5** (1991) 115-119.
[37] Thannhauser, D. S., *Solid State Commun.* **1** (1963) 223.
[38] Blumenthal, R. N., Baukus, J., Hirthe, W. M., *J. Electrochem. Soc.* **114** (1967) 172.
[39] Baumard, J. F., Tani, E., *J. Chem. Phys.* **67** (1977) 857.
[40] Balachandran, U., Eror, N. G., *J. Solid State Chem.* **39** (1981) 351-359.
[41] Chan, N. H., Sharma, R. K., Smyth, D. M., *J. Electrochem. Soc.* **128** (1981) 1762-1769.
[42] Jones, T. A., Firth, J. G., Mann, B., *Sensors Actuators* **8** (1986) 281-305.
[43] Daniels, J., Härdtl, K. H., Hennings, D., Wernicke, R., *Philips Res. Rep.* **31** (1976) 487-566.
[44] Moos, R., *Dissertation,* Universität Karlsruhe, 1994.
[45] Pennewiss, J., *Dissertation,* Universität Karlsruhe, 1986.
[46] Kröger, V. A., Vink, H. J., *Solid State Physics,* Vol. III; New York: Academic Press, 1956, p. 307.
[47] Gerblinger, J., Reitmeier, N., Meixner, H., *VDI-Ber.,* No. 939 (1992) 433-438.
[48] Lohwasser, W., Gerblinger, J., Lampe, U., Meixner, H., *J. Appl. Phys.* **75** (1994) 3991-3999.
[49] Lohwasser, W., *Diplomarbeit,* Walter-Meissner-Institut für Tieftemperaturforschung, TU München, 1992.
[50] Lampe, U., Fleischer, M., Meixner, H., *Sensors Actuators B* **17** (1994) 187-196.
[51] Tragut, C., *Dissertation,* Universität Karlsruhe, VDI Fortschrittsberichte Reihe 8, No. 291; Düsseldorf: VDI-Verlag, 1992.
[52] Wernicke, R., *Philips Res. Rep.* **31** (1976) 526-543.
[53] Tragut, C., Härdtl, K. H., *Sensors Actuators B* **4** (1991) 425-429.
[54] Schönauer, U., *Sensors Actuators B* **4** (1991) 432-436.
[55] Gerblinger, J., Meixner, H., in: *Proc. Transducers '91, San Francisco,* 1991, pp. 592-596.
[56] Lampe, U., Gerblinger, J., Meixner, H., *VDI-Ber.,* No. 939 (1992) 15-21.
[57] Lampe, U., Gerblinger, J., Meixner, H., *Sensors Actuators B* **7** (1992) 787-791.
[58] Gerblinger, J., Meixner, H., in: *Proc. EURO-Ceramics II,* Vol. 3, Ziegler, G., Hausner, H. (eds.); pp. 2463-2468.
[59] Gerblinger, J., Meixner, H., *Sensors Actuators B* **4** (1991) 99-102.
[60] Göpel, W., *Techn. Messen* **52** (1985) 92-105.
[61] Waser, R., Klee, M., *Integrated Ferroelectrics* **2** (1992) 23-40.
[62] Henrich, V. E., Dresselhaus, G., Zeiger, H. J., *Phys. Rev. B* **12** (1978) 4908-4921.
[63] Lo W. J., Somorjai, G. A., *Phys. Rev. B* **12** (1978) 4942-4950.
[64] Powell, R. A., Spicer, W. E., *Phys. Rev. B* **6** (1976) 2601-2604.
[65] Chiang, Y. M., Tagaki, T., *J. Am. Ceram. Soc.* **73** (1990) 3278-3291.
[66] Giber, J., Perczel, I. V., Gerblinger, J., Lampe, U., Fleischer, M., *Sensors Actuators B* **18-19** (1994) 113-118.
[67] Fleischer, M., Meixner, H., *Sensors Actuators B* **6** (1992) 257-261.
[68] Fleischer, M., Meixner, H., *Sensors Actuators B* **13-14** (1993) 259-263.
[69] Chemisky, E., Gerblinger, J., Meixner, H., in: *ITG-Fachbericht, 126: Sensoren – Technologie und Anwendung,* pp. 129-134.
[70] Gerblinger, J., Gabler, H., Mock, R., Meixner, H., *Sensors Actuators, B* **15-16** (1993) 396-400.
[71] Schiesser, P., *Sensor Mag.* **3** (1993) 9-13.
[72] Wienand, K. H., Sander, M., Englert, W., *Sensor Rep.* **4** (1993) 34-36.
[73] Hausner, M., *Diplomarbeit,* Walter-Meissner-Institut für Tieftemperaturforschung, München, 1992.
[74] Jehn, H., *J. Less-Common Met.* **100** (1984) 321-339.
[75] Stacey, M. H., in: *Catalyses,* Vol. 3; London: Chemical Society, 1980, pp. 98-122.
[76] *ANSYS Rev. 5.0A;* Swanson Analysis Systems, 1992.
[77] Hockel, K. G., Langen, P., Mallog, J., *Motortechn. Z.* **53** (1992) 326-338.

[78] Subbarao, E. C., *Trans. Indian Ceram. Soc.* **49**, No. 3 (1990) 47-59.
[79] Yamada, T., Hayakawa, N., Kami, Y., Kawai, T., *SAE Tech. Pap.* No. 920234 (1992) 23-36.
[80] Tien, T. Y., Stadler, H. L., Gibbons, E. F., Zacmanidis, P. J., *Ceram. Bull.* **54** (1975) 280-185.
[81] Benammer, M., Maskell, W. C., *Smart Mater. Struct.* **1** (1992) 134-138.
[82] Takeuchi, T., *Sensors Actuators* **14** (1988) 109-124.
[83] Gall, M., *Sensors Actuators B* **4** (1991) 533-538.
[84] Krebs, P., Grisel, A., *Sensors Actuators B* **13-14** (1993) 155-158.
[85] Vauchier, C., Charlot, D., Delapierre, G., Accorsi, A., *Sensors Actuators B* **5** (1991) 33-36.
[86] Visser, J. H., Zanini, M., Rima, L., Soltis, R. E., Kovalchuk, A., Hoffman, D. W., Logothesis, E. M., Bonne, U., Brewer, L. T., Bynum, O. W., Richard, M. A., *Sensors Actuators* in press.
[87] Aigner, R., Auerbach, F., Huber, P., Müller, R., Scheller, G., *Sensors Actuators B* **18-19** (1994) 143-147.
[88] Niebling, G., *Sensors Actuators B* **18** (1994) 259-263.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

7 Integrated Optical Sensors: New Developments

HELMUT TEICHMANN, Universität-GH Paderborn, Paderborn, Germany

Contents

7.1 Introduction 222

7.2 Displacement Sensors 223
7.2.1 Michelson Interferometer Scheme 223
7.2.2 Reflecting Grating Scheme 228

7.3 Vibration Sensors and Velocimeters 229

7.4 Pressure, Force, and Acceleration Sensors 232
7.4.1 Mach-Zehnder Interferometer Pressure Sensors 233
7.4.2 Microbridges and Cantilevers with Optical Waveguides 234

7.5 Rotation Sensors 236
7.5.1 Fiber-Optical Gyrometers with IO Signal preprocessing 237
7.5.2 All-Integrated Optical Gyrometers 238

7.6 Electric Field Sensors 239

7.7 Optical Spectrum Analyzers and Wavelength-Selective Devices 241
7.7.1 Acousto-Optical Tunable Wavelength Filters 241
7.7.2 Multi-Channel De/Multiplexers and On-Chip Spectrometers 242
7.7.3 Spectrum Analyzers 243

7.8 Chemical Sensors and Biosensors 244
7.8.1 Effective Refractive Index Sensing 245
7.8.2 Fluorescence Methods 247

7.9 Gas Sensors 248
7.9.1 Evanescent Field Sensing 249
7.9.2 Free Path Absorption Sensing 250

7.10 Optical Data Pick-Ups 251

7.11 Conclusions 253

7.12 Acknowledgments 254

7.13 References 254

7.1 Introduction

With the inception of "integrated optics" (IO) by Miller in the late 1960s, this new optical concept had created great expectations. However, it took several years until progress in planar material processing, namely submicron lithography, in micro-electronics fabrication, in reliable semiconductor lasers, and in low-loss silica optical fibers gave a strong impetus to the development of the new optical discipline. Today, research activities have pushed integrated optics towards the threshold of broad commercial exploitation.

Additional to the applications in the field of optical communications, integrated optics – as well as microelectronics and micro-optics – represent a key technology for the realization of miniaturized sensors and sensor–actuator systems. In the future, exploitation of the innovative potential accumulated in research and development centers is a prerequisite for a modern and competitive industry. Particularly, the broad spectrum of mature integrated optical sensor devices has been substantially enlarged during the last few years.

Focusing on this topic, it is the scope of this chapter to survey new developments during the last 4 years and the realistic potential offered by integrated optics for different sensor types. The survey is addressed to readers interested in specific sensor functions as well as to researchers concerned with integrated optics. To meet this goal, this review on sensor devices is organized according to realized sensor functions rather than to principles of measurement or fabrication technologies. The range of applications of integrated optical sensors extends from sensing of mechanical measurands (displacement, vibration, pressure, force, acceleration), measurement of electrical fields and of spectral properties, detection of (bio-)chemical substances and gases, to readout of optical storage media.

Whereas microelectronics chiefly focusses on silicon technology, integrated optics still makes use of several material systems like glass, silicon, lithium niobate ($LiNbO_3$), polymers, and III-V compound semiconductors, each with different waveguide fabrication technologies (see [2]). Although new techniques have eased the restrictions imposed by the different material properties, there is no material well suited for all function purposes. Thus, depending on application requirements, integrated optical sensor technology faces a broad variety of materials, processing, and hybridization techniques. Integrated optical structures may consist of various basic device components such as waveguides, y-junctions, directional couplers, phase and amplitude modulators, and different interferometer types. In Volume 6 of this series [1], Brandenburg, Hinkov, and Konz [2] give a systematic description of materials and principles of integrated optical sensing devices. To avoid repetition, their chapter [2] functioned as a basis and starting-point for the contents of the following sections. For further information, the reader may turn to citations in [2] and also to the comprehensive book by Nishihara et al. [3], the study of active optical integrated circuits [4], and, in particular, to the review on integrated optical sensors by Parriaux [5]. The very extensive and complete study on the potentials of integrated optics for small and medium enterprises by Sohler and Volk [6] and the report on integrated optics – current state of the art by the Integrated Optics Committee of the DGG (Deutsche Glastechnische Gesellschaft) [7] are also worthwhile reading, although they are in German.

7.2 Displacement Sensors

Contactless distance sensing with sub-micrometer precision is a domain of optical metrology. With the advent of well developed small-sized integrated optical displacement sensors, it became possible to exploit the high position resolution for active control of machining tools. Compared with bulk optical set-ups, IO interferometric distance-sensing devices show higher stabiltity. Additionally, with the use of fiber-optical interrogation, electromagnetic immunity is achieved. The strong demand for small and cheap high-resolution position sensors has led to a market niche; nowadays, the IO displacement sensing devices, together with IO components for rotation sensing, can be regarded as the most developed IO sensor systems.

IO displacement sensors are mainly based on two different principles of operation: for sensing of displacement along the propagation direction of the light the Michelson interferometer configuration (Section 7.2.1) is used; the reflecting grating scheme (Section 7.2.2) allows the detection of lateral movements.

7.2.1 Michelson Interferometer Scheme

Most devices make direct use of the wavelength λ of an electromagnetic wave as a natural length unit exploiting Michelson interferometer configuration schemes (Figure 7-1 a). Using retro-reflectors or tightly focusing reflecting mirrors, the length of interest forms the measuring arm of the interferometer; with fixed optical length of the reference arm reflector displacements Δx of the size of $\lambda/2$ in the measuring arm cause the output powers (signals P_G and P_W) of each output port to be modulated complementary (Figure 7-1 b). Output signal fringe counting or output signal interpolation then gives a measure for the longitudinal displacement along the direction of the measuring arm. With an electronic interpolation of the period output signal sub-μm resolution can be achieved.

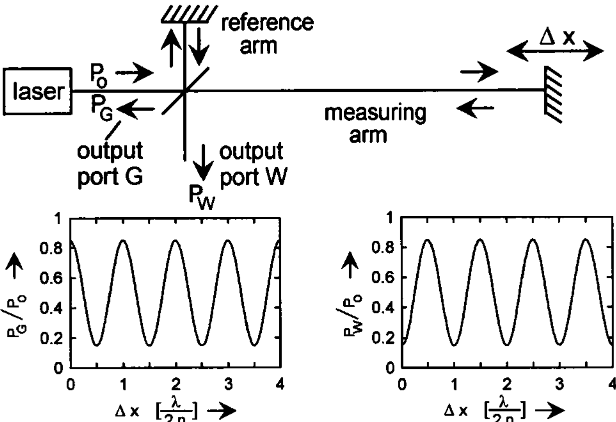


Figure 7-1. (a) Upper part: Michelson interferometer scheme for detection of displacement Δx. (b) Lower part: variation of both output signals, P_G and P_W, normalized on incident power P_0, versus displacement Δx in units of half the laser wavelength in air (refractive index n). Different optical losses in the reference and measuring arms cause a reduction of the modulation depth.

However, in the simple Michelson interferometer scheme the signal variation with displacement goes to zero if the difference in optical path length of both arms is equal to multiple integers of the wavelength itself. Additionally, no indication of the direction of movement can be derived from both signals, as they are out of phase by 180°. To avoid this ambiguity, quadrature signals (eg, two signals 90° out of phase) must be generated; IO devices allow optical signal processing either by passive phase demodulation or active phase modulation (see below).

Interferometric Sensors with Passive Phase Detection Schemes

An integrated optical interferometer chip with a 3 × 3 coupler structure (Figure 7-2 [8, 9, 10]) (IO chip manufactured by IOT, Jena, Germany) defined by an ion-exchange process is the key device of a commercially available displacement measuring system (Euchner, Leinfelden, Germany). Light from a stable, narrowband semiconductor laser (λ = 780 nm) is coupled to the central waveguide via a single-mode optical fiber. The reference arm is contained within the chip, whereas the measuring arm is formed by a GRIN (Graded Index) lens and an external reflecting mirror sphere. Deviating from the normal Michelson configuration, the returning light in both arms recombines in the 3 × 3 coupler designed to provide a passive phase demodulation: at the three output ports interferometer signals are generated which differ by 120° in phase. Any two of the port signals always exhibit a non-zero displacement sensitivity and provide information on movement direction. Together, all three returning signals allow for normalization of optical power.

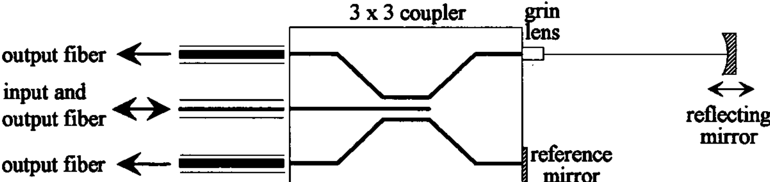


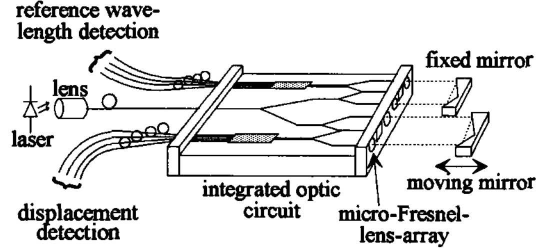
Figure 7-2. 3 × 3 coupler IO chip for displacement sensing (after [9]).

For a stable optical connection of the input and output ports with the photodetectors in the signal processing frame, one single-mode (input and output fiber in Figure 7-2) and two multi-mode optical fibers are used. The fibers are prealigned by anisotropically etched V-grooves on a silicon carrier and glued to the glass interferometer chip.

With the corresponding readout electronics, the measuring system is capable of fast movements (up to 1 m/s) over a distance of 50 cm with a displacement accuracy of $1:10^6$. Whereas the measuring increments can, in principle, be divided down to 65 nm by signal interpolation, the accuracy limit is affected by the stability of the laser wavelength and by the fluctuations of air density. Temperature stabilization and suppression of back-reflections into the light source by using an optical isolator are necessary to precisely stabilize the laser wavelength.

To compensate quantitatively for fluctuations of the laser wavelength and the refractive index of air in the measuring arm, a fixed reference distance nearby must be measured addi-

tionally. At the Swiss Centre for Electronics and Microtechnology (CSEM), an integrated optical displacement sensor with a "double interferometric circuit" has been developed (Figure 7-3, [11]). Fabricated by K^+-Na^+ ion exchange in BK7 glass, the phase demultiplexing of the interferometer signals is realized with the waveguide imaging film technique (Figure 7-4 [12-14]): the light entering a segment of a planar waveguide ("caisson") by two single-mode waveguides will laterally redistribute due to the multi-mode guiding properties of the caisson. The resulting power distribution at the four output ports strongly depends on the phase lag of the waves in the two incoming waveguides. With a well designed caisson length, four output ports with optical signals shifted by 90° have been realized. Another key issue of the CSEM sensor device is the use of an RIE-etched binary four-step Fresnel lens matrix in a silica substrate for coupling light from the waveguides to the free space interferometer arms and back to the waveguides again (bottom part of Figure 7-3).



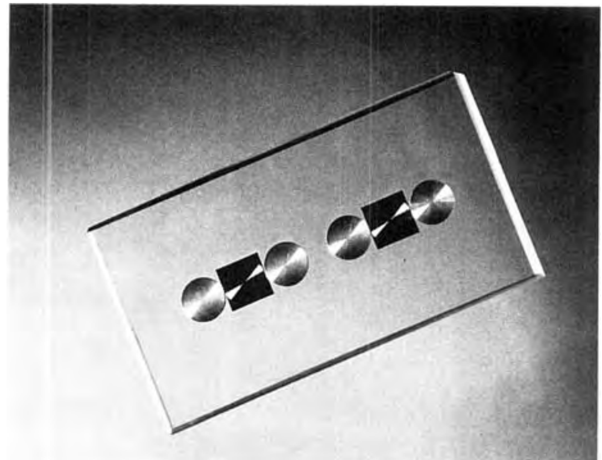


Figure 7-3.
Top: sketch of the hybrid integrated optical displacement sensor with wavelength stabilization (after [11]). Bottom: four-step micro-Fresnel lens array used for coupling prepared by RIE etching of a silica substrate [11]; photograph by courtesy of CSEM.

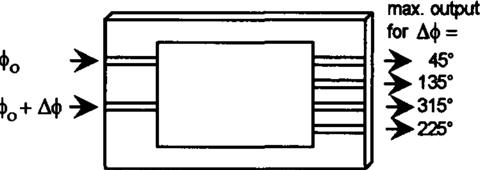


Figure 7-4.
Phase demodulation by multi-mode waveguide self-imaging (caisson technique) for detection of the phase difference $\Delta\phi$ between the two incoming guided waves.

Another way to detect the direction of movement is to use a double Michelson interferometer scheme: from two interferometers with a common measuring arm two signals 90° out of phase are obtained, if an optical path difference of $\pi/4$ is introduced between the two reference arms. This principle of operation was transferred into a commercially available (CSO, Grenoble, France) IO device developed by LETI (Grenoble, France). The device is based on a waveguide technology with SiO_2-Si_3N_4-SiO_2 layers on silicon ("IOS-1 technology") (Figure 7-5 [15]); on the chip (Figure 7-6) two lenses, a mirror, a beam splitter, and a beam divider are integrated. The passive phase shifting element, which introduces a constant phase lag between the two reference arms, is realized by an additional cladding layer on top of the waveguide, which changes the velocity of propagation of the guided optical wave. As shown in Figure 7-6, this phase shifting element is designed to laterally affect one half of the guided reference beam only, thus defining two separate reference arms. The two output signals at the photodiodes are 90° out of phase and can be used to derivate a quadrature signal.

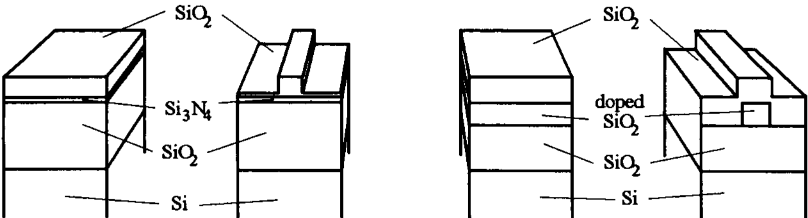


Figure 7-5. Strong and more weakly guiding waveguide structures ("IOS-1", left and "IOS-2", right, respectively) on a silicon substrate for planar and stripe waveguides (after [15]).

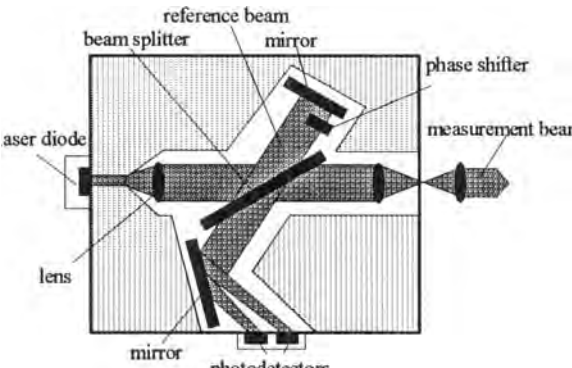


Figure 7-6. Schematic diagram of the sensor chip for displacement sensing based on IOS-1 technology developed by LETI (after [16]).

As the refractive index of Si_3N_4 waveguiding film ($n = 2.014$ at 633 nm) greatly exceeds the refractive index of silica, the coupling efficiency to the optical fiber is very low; therefore, the integrated optical chip is hybridly integrated with the laser diode, the dual photodetector chip, a preamplifying unit, a thermoelectrical stabilization element, and an outcoupling GRIN lens in a single cylindrical packaging of only 17 × 73 mm (Figure 7-7). To avoid fiber coupling losses, a different waveguide technology using doped and undoped SiO_2 ("IOS-2 technology", Figure 7-5, to our knowledge not yet commercialized by CSO) was developed by LETI:

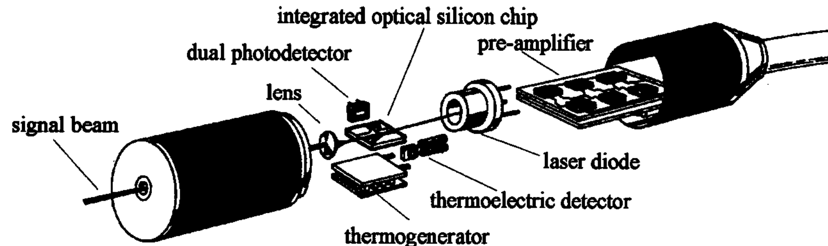


Figure 7-7. Schematic diagram of the complete displacement sensing detector head including the laser module, the IO chip, the detector, and preamplifying circuits, as developed by LETI and commercialized by CSO (Grenoble, France).

as the doping leads to a small increase in the refractive index ($\Delta n \approx 1\%$ for p-doping [69]), the coupling efficiency to silica fiber is strongly improved. The design of the displacement sensor with the IOS-2 technology is also based on a double interferometer with a phase-shifting element ([2], p. 411). New developments on IO displacement sensor heads performed at LETI take advantage of microguide architectures with a tri-coupler element [17] which provides passive phase demodulation. These activities are close to commercial exploitation.

Interferometric Sensors with Active Phase Detection Schemes

The phase of the interferometer signal can also be derived by active detection schemes like heterodyne signal processing techniques. They provide advanced sensitivity and allow to determine the direction of object movement. However, as active optical phase modulation or frequency shifting elements are necessary in this case, an adequate choice of the substrate material is required. Materials like lithium niobate ($LiNbO_3$) or III-V-semiconductors allow to realize broadband electrically driven phase shifters, whereas the exploitation of thermooptical effects in glass or in silica layers on silicon are restricted to frequencies less than 50 kHz.

Using ion exchange in glass, a fiber pigtailed distance sensing device consisting of a double interferometer scheme with thermooptical phase shifting elements was realized by Jestel et al. [19] (see also [2], p. 411). With a laser wavelength of 633 nm and a single sideband detection at modulation frequencies of 10 kHz a resolution of 1 nm was achieved, which is equal to 1° phase resolution of the interferometer signal.

At JENOPTIK (Jena, Germany) a three-beam-interferometer [18] was built by annealed proton exchange (APE) in $LiNbO_3$. Two reference arms probe a well-defined mechanical length as a master scale. With electrooptical phase modulation at 10 MHz and a heterodyne readout technique (see Section 7.3.1), a detection accuracy of 10^{-7} (ca. 10 nm) was obtained. In contrast to APE fabrication techniques waveguides prepared by titanium-indiffusion in $LiNbO_3$ support both polarizations. The sophisticated design of a $Ti:LiNbO_3$ double-heterodyne interferometer device realized at the University of Osaka [26] involves waveguide polarization optics (polarization converter, splitter, polarizer, etc.) to define one measuring arm and two reference arms. Detecting at the harmonic of the modulation frequency a resolution of ± 3 nm was achieved.

The goal of work at the Paul Scherrer Institute for Applied Solid State Physics (PSI, Zurich and Villingen, Switzerland) is a monolithical integration of a Michelson displacement in-

terferometer in GaAs, including a DBR laser, waveguides, modulators, and detectors. Passive and active elements (eg, waveguides and detectors) require different transparency conditions of the waveguiding multi-quantum well film (and adjacent cladding layers). To meet this, band gap shifting techniques such as dielectric cap (SiO_2) annealing [20] have been developed at the PSI. In future, these techniques might lead to a sensor head on GaAs which can be produced solely by batch fabrication techniques.

7.2.2 Reflecting Grating Scheme

A very unusual interferometric sensor device for the detection of lateral movements is shown in Figure 7-8. It consists of a planar glass waveguide sputtered onto a thermally oxidized silicon substrate. Light of a laser diode (λ = 0.79 μm) is coupled into the waveguide and spreads laterally while propagating. By two linear focussing grating couplers (LFGCs) two slightly convergent light beams are coupled out, which intersect in the region of the grating scale with a tilt angle of 20° (10° to the chip surface normal, each). By an adequate choice of the periodicity of the reflective grating scale (4.6 μm), two of the diffracted beams (+1 order of diffraction of the right outcoupled beam and −1 order of the left) overlap and interfere while impinging on the two on-chip photodiodes. As the phase difference between the interfering beams depends on the lateral position of the grating scale, its movement along that direction causes the interferometer signals to vary sinusoidally with a periodicity of 2.3 μm.

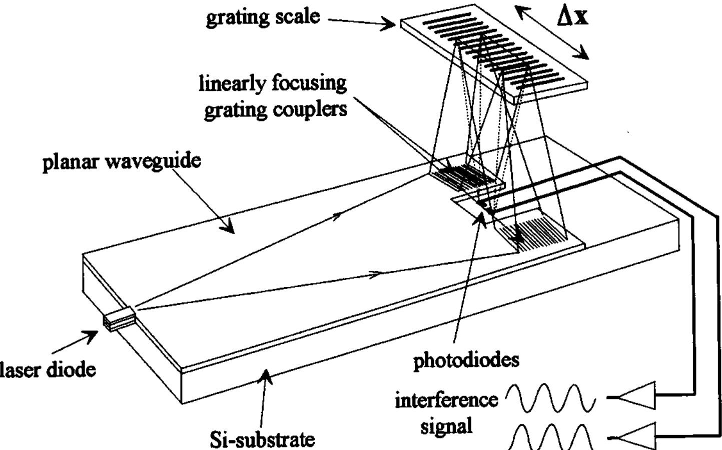


Figure 7-8. Integrated optical device for sensing lateral displacements of a grating scale (after [21]).

Proposed and realized by an Osaka University group, the displacement sensing chip (Figure 7-8) belongs to a family of IO devices (see [22] and citations therein) using chirped and curved gratings for coupling in and out focused beams from planar waveguides. Based on a thermally oxidized n-Si substrate, a PCVD silicon nitride layer is deposited in which the grating grooves are patterned by direct electron beam writing and reactive ion etching techniques. An rf sputtered glass layer (Corning 7059 glass) is used as waveguiding film. The on-

chip photodetector p-n junctions are fabricated by local boron indiffusion before the grating formation.

This fabrication technique has also been exploited to realize other integrated optical devices including a compact disk pick-up device [136], a magneto-optical read/write head [137, 138] (for both, see Section 7.10), and very recently an optical head for a confocal scanning microscope [23]. With the help of an additional focusing distributed Bragg reflector grating also another interferometric position sensor was realized [24] (see also [2], p. 411), which allows to detect the height of a reflecting surface above the IO chip with a resolution of 10 nm. However, to our knowledge, no commercial use has been made of IO components belonging to this device family up to now.

In summary, Michelson interferometer IO displacement sensors based on glass or silicon substrates have achieved a high degree of maturity. As key components in commercially available systems, they allow contactless measurements of displacements with resolutions of $1:10^6$ compared with the measuring distance which is typically restricted to the order of 0,5 m. Additionally, promising new devices using chirped and curved waveguide gratings have recently been demonstrated for sensing of lateral displacements.

7.3 Vibration Sensors and Velocimeters

In principle, all displacement sensors can be used as vibration sensors or velocimeters. Both IO sensor types normally use optical interferometry as the principle of operation. For the determination of vibration amplitudes, the direction of actual movement is insignificant and slower response times are tolerable. The signal processing schemes are therefore different in the two cases.

Any movement of the reflecting mirror leads to a Doppler frequency shift of the reflected light in the measuring arm. A periodic vibration or a constant velocity causes new frequency components to arise in the interferometer signal. Thus output signal processing in the frequency domain is more appropriate than a high bandwidth fringe counting in time domain necessary for a fast readout of displacement-sensing devices. Additionally, by introducing heterodyne detection techniques into the interferometer scheme, the sensitivity for vibration amplitudes can be greatly enhanced compared with fringe counting.

In a heterodyne interferometer the frequency of the optical signal passing the reference arm is shifted by $\Delta\nu$; without movement of the object mirror the interferometer output signal varies sinusoidally with this carrier frequency $\Delta\nu$. Any linear or periodic movement of the reflection mirror gives rise to a side-band modulation which can be detected with an electrical spectrum analyzer.

In contrary to glass and silicon, lithium niobate ($LiNbO_3$) allows to take advantage of its excellent electro-optical and acousto-optical properties. With this material, the frequency shift required for heterodyne detection can be obtained by electro-optical phase modulators using serrodyne techniques (sawtoothlike periodic phase shifts of 2π). Several integrated optical heterodyne interferometer chips have been proposed and realized on the substrate $LiNbO_3$ [25–30]. In this way, amplitude sensitivities of ± 3 nm have been achieved using a 0.633 µm He-Ne laser and frequency shifts of a few hundred kHz.

A device concept (Figure 10-9) which takes advantage of the excellent acousto-optical properties of $LiNbO_3$ and makes optimum use of the laser power budget has been proposed and realized at the University of Paderborn, Germany [31]. Several discrete elements such as frequency shifters, beam splitters, and polarization optics are monolithically combined on a $LiNbO_3$-substrate using Ti-indiffused waveguides. The frequency shifter is realized by an acousto-optical TM-TE converter (Figure 7-9). With the help of an rf signal applied to an interdigital transducer a guided surface acoustic wave is excited. Owing to the piezo-elastic and electro-optic effects, optical power is transferred from the optical TM mode to the optical TE mode, thereby causing a frequency shift in the 100 MHz range equal to that of the acoustic wave f_a. The mode coupling strength rises with the square root of the power of the acoustic wave; by adjusting the rf power a TM-TE mode coupling of 50% is achieved over the fixed interaction length. The laser diode and the mode converter together function as a two-frequency light source. Thus, this unit replaces an expensive and complex Zeeman laser normally used in conventional vibrometers which are commercially available. Passing the polarization splitter, TM and TE modes with the frequencies f_0 and $f_0 + f_a$ are directed to the reference and measuring arm, respectively. In both arms, pairs of an electro-optic TE-TM converter and a phase shifter operate as quarter-wave plates. After two passes, mode polarization states are interchanged for the returning light reflected from the end-face mirror or the external vibrating mirror. Both modes therefore leave the polarization splitter towards the output arm. With the help of an external 45° polarizer and a fast photodetector, the interferometer signal is observed with an rf spectrum analyzer.

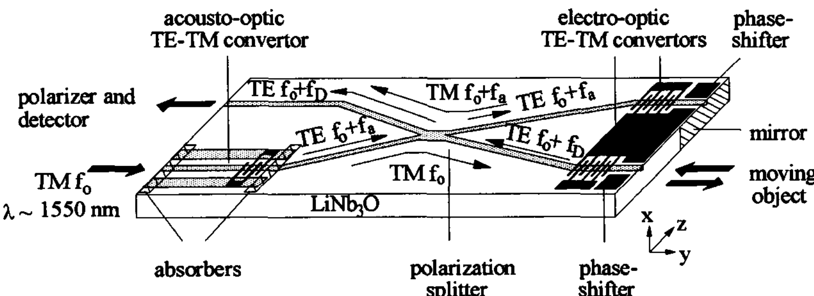
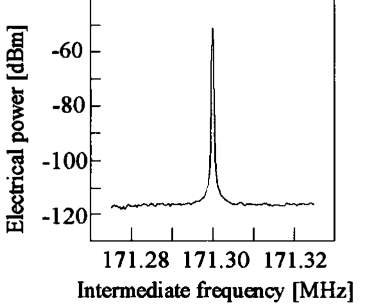


Figure 7-9. *Schematic diagram of the IO acousto-optical heterodyne interferometer (after [32]).*

The interferometer chip works with a commercial DFB semiconductor laser at 1545 nm showing a linewidth of approximately 20 MHz. Optical power of 150 µW is fed into the optical input port. Typically, the reflecting mirror is 25 cm apart from the interferometer device. Without any mirror movement only the frequency shift of the acousto-optic mode converter (here f_0 = 171.3 MHz, equal to the frequency of the surface acoustic wave) is observed in the spectrum (Figure 7-10) with a carrier-to-noise ratio (CNR) exceeding 65 dB at 300 Hz electrical bandwidth. With a sinusoidal oscillation $A \sin(2\pi f_v t)$ of the external mirror, sidebands arise in the spectrum separated from f_0 by multiples of f_v (Figure 7-10; here f_v = 5 kHz). As the height of the nth sideband is determined by the square of the Bessel function J_n with the argument $2A(2\pi/\lambda)$, the vibration amplitude A can be extracted from the sideband-to-carrier ratio in the rf spectrum. For smaller amplitudes the signal-to-noise



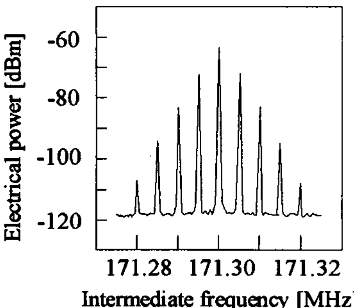


Figure 7-10. Measured spectrum of the heterodyne signal for operation with static (left) and vibrating (right) mirror terminating the measuring arm (after [32]).

ratio approaches unity if the height of the first sideband is equal to the noise floor level. In turn, a sideband-to-carrier ratio equal to the carrier-to-noise ratio of 75 dB, as in the case of a static mirror, indicates that the minimum detectable amplitude is less than 45 pm for this sensor device [32].

The high resolution for vibration amplitudes down to the sub-nm range makes this heterodyne interferometer device an ideal sensor for the detection of cantilever oscillations in atomic force microscopes (AFMs). The mechanical resonance curve (Figure 7-11) of such a cantilever, measured with the sensing device, indicates a resonance FWHM of only 1.1 kHz at a resonance frequency of 284.5 kHz [33].

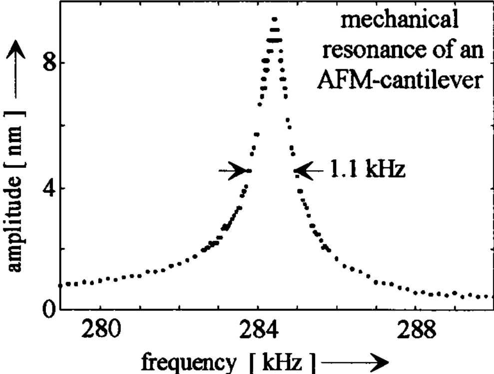


Figure 7-11.
Mechanical resonance of an AFM cantilever [33] measured with the IO heterodyne interferometer shown in Figure 7-9.

In the future, it is intended to realize more complex circuits of higher functionality; development efforts show that a monolithic integration of an erbium-doped waveguide laser (to be pumped optically) [34] as well as on-chip polarization mixing [35] are feasible on $LiNbO_3$ as substrate.

If low coherence sources as high-power laser diodes are utilized for heterodyne interferometry, modifications of the IO interferometer chip are necessary. In particular, the optical paths for the measuring and reference arms must be matched in length to preserve "white light" interference ability. This approach is followed in a laser-Doppler velocimeter device fabricated by UTP (Figure 7-12 [36]) on an $LiNbO_3$ substrate by proton exchange and an-

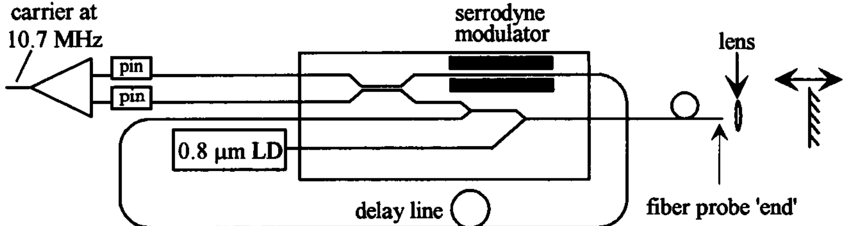


Figure 7-12. Integrated optical/fiber-optical laser-Doppler velocimeter and vibrometer for use with low coherence laser diodes (after [36]).

nealing techniques. The device consists of two 3 dB couplers, a serrodyne modulator, a directional coupler, a fiber probe, and a fiber delay line. Light from a 0.8 µm laser diode (power 7.5 nW) with a coherence length of 1 cm is coupled into the input port of the optical chip (Figure 7-12). At the end of the fiber probe, the light is partly reflected; this light in part reaches the two photodetectors after passing the fiber delay line and the serrodyne modulator. The light leaving the fiber probe enters the measuring arm; after collimation and reflection at the object mirror, the light re-enters the fiber. Part of this light takes the direct route to the detectors and interferes with the properly delayed guided reference wave. With a serrodyne frequency shift of 10.7 MHz, a mirror vibration at 5.7 kHz with an amplitude of 36.8 nm was observed with a sideband power of 40.5 dB below the carrier.

In summary, integrated optical velocimeters and vibration sensors using heterodyne detection schemes allow ultra-high sensitivity at moderate response rates. The necessity for active frequency shifting elements makes $LiNbO_3$ a favorite choice as substrate material. In addition to serrodyne phase modulation techniques, acousto-optical mode converters are very well suited as frequency shifters for the MHz range. With a highly monolithically integrated acousto-optical IO device, amplitude sensitivities of 45 pm have been demonstrated, which can be advantageously exploited in atomic force microscopes. Despite this enormous sensing potential, no commercial IO devices are available so far.

7.4 Pressure, Force, and Acceleration Sensors

Pressure, force, and acceleration are mechanical measurands. Therefore, transducing principles for these physical quantities typically involve devices with movable parts. As this is valid for IO devices also, the techniques of microfabrication (sometimes also called micromachining) are very important for this class of sensors. Owing to its anisotropic etching behavior, silicon offers the most powerful techniques for the fabrication of microstructures. In addition, optical and electronic circuits can also be integrated on an common silicon substrate.

Several micromechanical IO devices on silicon have been proposed and realized. Among them is a highly integrated, membrane-based, interferometric IO pressure sensor with a fiber-optical interconnection and an on-chip photodiode (Section 7.4.1). Microbridges and cantilevers fabricated by etching or by wafer bonding techniques are briefly discussed in Section 7.4.2.

7.4.1 Mach-Zehnder Interferometer Pressure Sensors

The IO pressure sensor, developed by the Technical University of Hamburg-Harburg, is based on a Mach-Zehnder interferometer (MZ) as basic optical structure (Figure 7-13 [37–39] and references cited therein). Light entering the device from a fiber is split up by a 3 dB coupler. Having passed identical optical paths the optical waves in both arms interfere constructively at the second coupler; in this case the light is fed into the output waveguide and is led to the detector. If any influences detune the MZ interferometer in such a way that the optical path lengths differ by a phase shift of $\pi/2$, the recombining waves interfere destructively and the related optical power is radiated into the substrate.

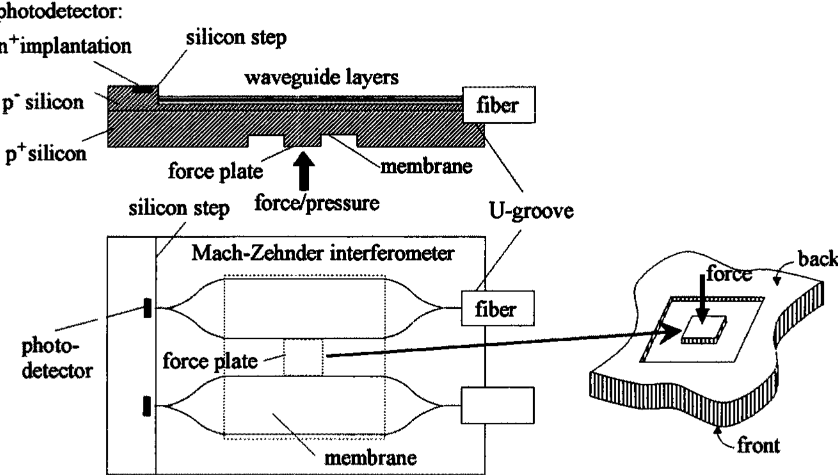


Figure 7-13. Micromechanical IO pressure sensor on silicon; schematic cross view and top view (left) and membrane and force plate geometry (right) (after [38, 39]).

Owing to the elasto-optical effect, application of mechanical strain leads to unequal optical paths. In order to enhance the pressure sensitivity of the device, the silicon substrate is etched down to a thickness of 50–100 μm in a membrane window unsymmetrically placed over the two MZ interferometer arms. If the membrane (or the central force plate, see Figure 7-13) becomes loaded with an increasing force, the interferometer signal has been shown to vary sinusoidally.

The device works with 0.633 nm light from a He-Ne laser. For the realization of strip-loaded waveguides, SiO_2-SiON-SiO_2 sandwich structures are fabricated on silicon. Preparation of U-grooves by plasma etching (PE) allows a self-aligning fiber-chip coupling. This PE-based U-groove etching technique can be applied to all silicon substrate orientations whereas anisotropic etching of V-grooves is less flexible because the tilted V-faces are always formed by {111} planes. A main feature of this waveguide device is the monolithic integration of broadband photodetectors: by dopant implantation either vertical or horizontal PIN junctions are formed next to a waveguide step; optimized endfire coupling from the waveguide to the detector is achieved by a silicon nitride antireflection layer.

Comparable to this structure, a membrane-based interferometric pressure sensor with oxynitride rib waveguides and an integrated a-Si : H photodiode has been realized by DASA (Munich, Germany) [40]. Both IO devices are highly integrated, rugged, fast sensors for mechanical measurands such as pressure or force.

7.4.2 Microbridges and Cantilevers with Optical Waveguides

With isotropic and anisotropic etching techniques, the silicon substrate can also be removed completely under an SiO_2 layer-based waveguiding stripe or finger. This leads to the formation of self-supporting microbridges or cantilever structures with waveguiding properties (Figures 7-14 and 7-15; see reviews [41, 42]). If the microbridge waveguide is bent by any applied force, additional bending losses decrease the optical transmission. For a cantilever structure, the light coupled out from the lever endface must re-enter the opposite waveguide; the lever deflection δ strongly determines the coupling efficiency (Figure 7-15). Therefore, both structures offer a very sensitive transducer scheme for force related measurands, encoding the information into an optically interrogated intensity variation.

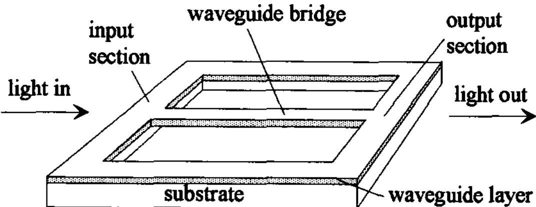


Figure 7-14. Optical waveguiding microbridge.

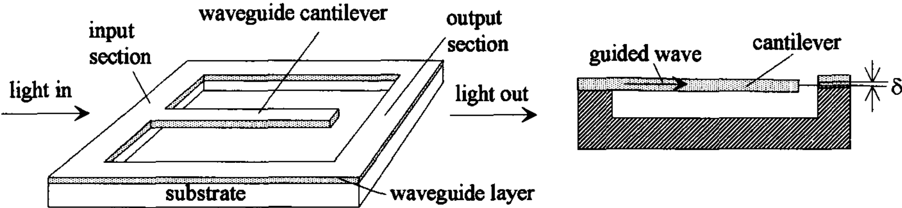
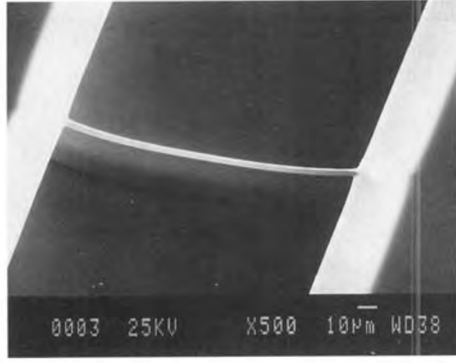


Figure 7-15. Left: optical waveguiding cantilever structure. Right: optical coupling gives the transducer function for cantilever deflections δ.

Waveguiding cantilever and microbridge structures have been realized by different groups using either planar SiO_2-Al_2O_3-SiO_2 waveguides ([43–45], see Figure 7-16) or strip-loaded SiO_2-SiON-SiO_2 waveguides [46, 47]. Wu and Frankena have shown that their cantilever and microbridge structures (Figure 7-16) may be used as very sensitive and highly linear acoustic signal detectors [43, 45]. Bezzaoui and Hoffmann [48] have combined a passive 1×8 waveguide coupler with an array of eight cantilevers chirped in lengths and resonance frequencies; the device is designed for use as a vibration sensor.



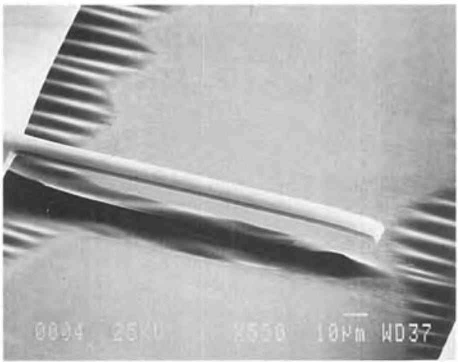


Figure 7-16. Left: SEM of a 100 µm long, 30 µm wide, and 2.5 µm thick microbridge [43]. Right: SEM of a cantilever of the same dimensions and a gap of 25 µm between the lever end and the planar output waveguide [43] (both photographs by courtesy of Delft University of Technology).

For vibration and acceleration sensing an additional mass load – a bridge center mass and a cantilever foot mass – greatly enhances the sensitivity due to higher deflections [43, 49]. As another feasible application, the exploitation of optically interrogated cantilever/stylus combinations in scanning force microscopes (SFMs) will in principle allow the detection of sub-Å stylus displacements and forces of 10^{-10}–10^{-11} N, as discussed by Wu ([43], p. 91).

In contrary to the intensity encoding structures mentioned above, waveguide Mach-Zehnder interferometers with short self-supporting waveguide segments in one interferometer arm have also been realized as pressure sensors [50, 51]. To accumulate a sufficient phase change, long self-supporting sensing segments are required for small pressure resolution; the mechanical integrity of these special interferometric devices is therefore a major drawback [50, 51]. Nevertheless, a very interesting aspect is that this device works with a newly developed, so-called anti-resonant reflecting optical waveguide (ARROW): a very thin, high refractive index, undoped polycrystalline silicon layer is deposited by LPCVD between two layers of SiO_2 (left part of Figure 7-17). Thus the waveguide fabrication has the advantage of full compatibility with silicon IC technology. As the modal field distribution (right part of Figure 7-17) vanishes in the polysilicon layer, the contributions of optical absorption in this layer to waveguide losses are negligible (for the modal characteristics of ARROW structures, see [52] and references cited therein). Integration of ARROW structures and photodetectors is feasible; as

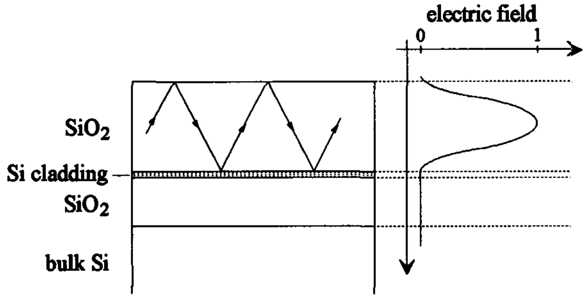


Figure 7-17. Layer structure (left) and schematic modal field distribution (right, here for a TE mode) of an ARROW waveguide (after [50, 51]).

ARROW structures can be made leaky by variation of the polysilicon layer thickness, coupling of guided light to an integrated detector is achieved. Thus, the ARROW waveguide technique is very promising for silicon-based IO devices, whether micromachined or not.

Wafer bonding techniques have also been employed for the formation of optical microstructures sensitive to pressure. However, in this case bridges or cantilevers are spanned over the waveguide: if the evanescent field of the guided wave reaches out considerably to the overlayed material, a change in separation will either cause the waveguide losses to increase owing to outcoupling (as discussed in the context of an ARROW waveguide-based silicon pressure sensor by Young et al. [56]) or result in a shift of the effective refractive index of the guided mode [53, 54]. If the latter is the dominant effect, a force-induced mechanical deflection of the so-called "refractive index shifting element" (Figure 7-18) can be interrogated optically by interferometric methods. In 1991, Lukosz and Pliska [55] demonstrated a highly sensitive, low-frequency acoustic pressure sensor using a 1 mm thick glass slide as a force plate mounted 2 µm above a planar waveguide prepared by TiO_2-SiO_2 dip coating of an oxided silicon wafer. For readout of the device, a polarization interferometer termed a "difference interferometer" (right part of Figure 7-18) was used; in most cases the effective refractive index of the TE and TM polarized modes changes differently with the separation d; on exciting both polarizations the interference pattern of the two waves, as observed with a Wollaston prism tilted by 45° behind the planar waveguide, is modulated with changes of the separation d.

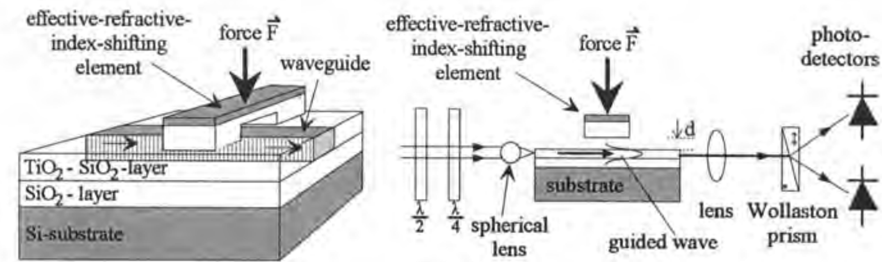


Figure 7-18. Left: cross section of a waveguide with an overlayed bridge structure; the effective refractive index of the guided optical mode depends on the distance d (after [54]). Right: polarimetric "difference interferometer" for optical interrogation (after [55]).

As pointed out by Lukosz [54, 57], the application of nanomechanical devices is not restricted to sensor applications; making use of electrical actuation of the bridges and cantilevers bonded to the wafer over the waveguide, phase modulators, intensity modulators, interferometric switches, tunable frequency filters, and planar slab waveguide deflecting and (de)focusing elements have also been proposed and demonstrated for the most part [58, 59].

7.5 Rotation Sensors

Optical rotation sensors (gyrometers) are based on the Sagnac effect [60]: in an optical loop, two counterpropagating waves experience different round-trip times, if the loop itself is rotated. In a simplified view, the round-trip length in a rotating coordinate system is shortened

(lengthened) for the counter (non-counter) propagating wave, whereas the velocity of light remains unchanged. The difference between the two round-trip times is proportional to the angular rotation rate.

Rotation-sensing devices find application in avionics, sea and land navigation, antenna position control, robotics, guided and guideless vehicle systems, and so on. After the invention of ring laser gyroscopes [60], striving for lower cost devices has led to commercially available, mature fiber-optical gyrometers using integrated optical signal preprocessing (see Section 7.5.1). In the future, less expensive, all-integrated optical devices (see Section 7.5.2) are likely to open up additional application fields with reduced sensing requirements.

7.5.1 Fiber-Optical Gyrometers with IO Signal Preprocessing

Fiber-optical gyrometer (FOG) configurations and signal-processing schemes have extensively been discussed in the literature ([60] and references cited therein). For the minimum configuration (Figure 7-19) a fiber coil, a laser module, a polarizer, a splitter, and a phase modulator are necessary. The last three elements can be integrated on an $LiNbO_3$ substrate using either titanium-indiffused or annealed proton-exchanged (APE) waveguides. According to [61], IO chips fabricated by APE techniques on $LiTaO_3$ show a similar performance to their APE $LiNbO_3$ counterparts.

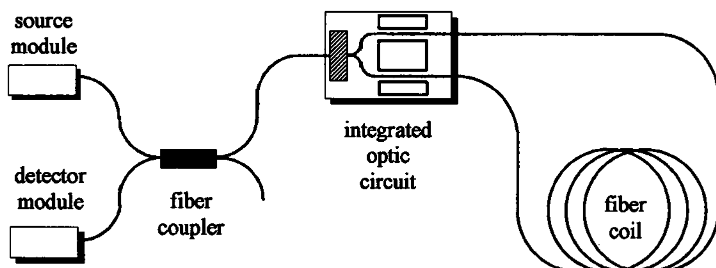


Figure 7-19. Minimum configuration for a integrated optical/fiber-optical gyroscope.

Today, fiber gyrometers using active integrated optical signal-processing devices are a well developed, reliable commercial product (Alcatel SEL, LITEF, UTP and several Japanese companies). Although work concerning the IO signal-processing chips is still being carried out in several laboratories (research on chip reliability and aging [62], choice of substrate materials [61], etc.), the main industrial efforts are focused on improvements for FOG system applications. The signal-processing chip fabricated at LITEF by APE on $LiNbO_3$ is shown in Figure 7-20. At Alcatel SEL, a similar chip is produced using the Ti-$LiNbO_3$ waveguide technique [63]. Both companies have established facilities for commercially manufacturing FOG units with a high drift stability of the order of 1°/h. In the Alcatel SEL device, analog and digital signal processing are performed by ASIC-based electronics. According to [64], progress has been made in combining three IO chips and fiber coils with a single laser diode source and a single detector by using time division multiplexing techniques. Such FOG triads, combined with three accelerometer sensors each, are the basis of inertial measurement units

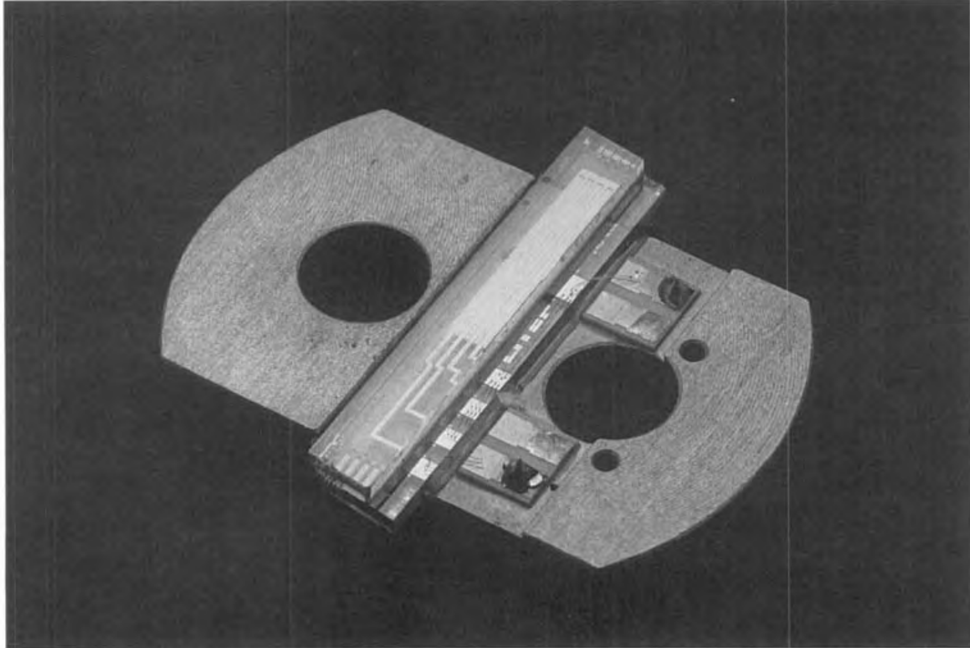


Figure 7-20. Integrated optical signal processing chip with holder, fabricated by annealed proton exchange in $LiNbO_3$, for fiber-optical gyrometer application (photograph by courtesy of LITEF, Freiburg, Germany).

(IMUs) as shown in Figure 7-21. In principle, IMUs provide all information on acceleration to determine the actual position. However, having started from a reference point, position errors increase with time. Therefore, enterprises like Alcatel SEL or LITEF have developed inertial navigation systems, which additionally make use of the Global Positioning System (GPS), a system of 21 satellites operated by the U.S. Department of Defense. Today, such GPS-aided inertial navigation systems based on fiberoptical gyrometers are commercially available.

7.5.2 All-Integrated Optical Gyrometers

To an increasing degree, low-cost gyrometers find sensing applications in robotics, in navigation systems for automobiles ([65] and references cited therein), and in guiding systems for autonomous vehicles. For these fields of application, however, key issues are different: frequent recalibration possibilities render long-term drift and random walk stability less important. Instead, high accuracy of the scale factor and of linearity are required: exact control of U-turns must be possible without recalibration. Up to now, devices for those applications have mostly been realized as all-fiber-optical gyrometers (see, eg, [66]). Under the stringent demand of low cost, only all-integrated optical gyrometers, based on glass or silicon substrates, may meet the mentioned requirements.

Based on research efforts on low-loss ring resonators with silica-on-silicon waveguide technology [67, 68], a first solid-state, integrated optical gyrometer has been demonstrated at

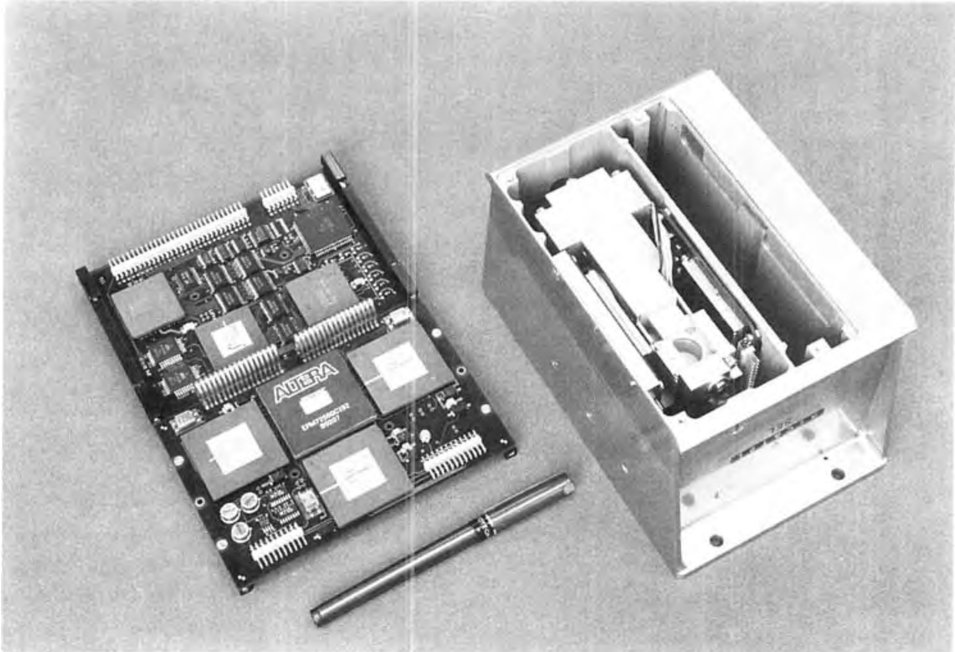


Figure 7-21. Inertial measurement unit with FOG triad integrated into a packaging of 16 × 12 × 12 cm; the two digital electronic boards shown left are housed in the slot next to the optical parts (photo by courtesy of Alcatel SEL, Stuttgart, Germany). Optimizing towards smaller size packaging is under way at Alcatel SEL.

LETI [69]. The gyro loop was formed by a spiral waveguide coil with seven turns and a diameter of 3 cm. Waveguide losses as low as 0.03–0.05 dB/cm have been realized, which total only 3.3 dB for the spiral waveguide coil. Additional losses of 0.18 dB, caused by incorporated 2N-2 X-junctions, are negligible. A 3 × 3 coupler (coupler scheme in principle comparable to Figure 7-2) is used as an input splitter and as a passive phase demodulator. The signal difference of the two non-central output ports is utilized, thus allowing a strong common mode rejection of low-frequency laser noise. With a coupler structure not optimized for the trade-off between output power and phase shift rotation, sensitivities of a few tens of degrees per second have been achieved, which are still far from the calculated achievable limit of about 0.5 °/s. The simple passive readout scheme, the possibility of batch fabrication processes, and a future hybridization of sources and detectors might make this type of sensor device a less sensitive, low-cost alternative to the fiber-optical gyrometer for applications with reduced requirements.

7.6 Electric Field Sensors

Electric field sensors are required as electromagnetic signal receivers (antennas), for electromagnetic compatibility (EMC) studies, and for other applications, such as medical purposes. Integrated optical sensors for electrical fields make use of electro-optical modulators

by directly driving the modulator electrodes with the voltage signal received from an antenna. Commonly, (asymmetric) Mach-Zehnder interferometers on $LiNbO_3$ are used as modulators, which show low driving voltages owing to the excellent electro-optical properties of $LiNbO_3$. Optical interrogation and fiber-optical transmission of the received antenna signal has several advantages: it avoids the necessity for matching the antenna to a low-impedance electrical link and it provides electrical isolation and immunity. No perturbations of the field distribution are introduced by the sensor as, with the exception of the antenna itself, it is free of metal parts. Moreover, the integration of the antenna and the electrode structure on a common substrate offers additional design options.

Several "extremely" low-frequency (ELF, $f < 3$ kHz) electric field sensors have been proposed and demonstrated (eg, [70, 71]), which use integrated optical Mach-Zehnder interferometers on $LiNbO_3$ with antenna structures exceeding the IO chip surface. With an on-chip combined antenna and electrode structure, Kuwabara et al. [72] have realized a wideband electrical field sensor for the frequency range from 100 Hz up to 2.5 GHz (Figure 7-22); with a fiber-optical interrogation based on a laser diode-pumped Nd:YAG laser, they achieved a frequency-independent sensitivity of 0.1 V/cm at 7.5 kHz band width.

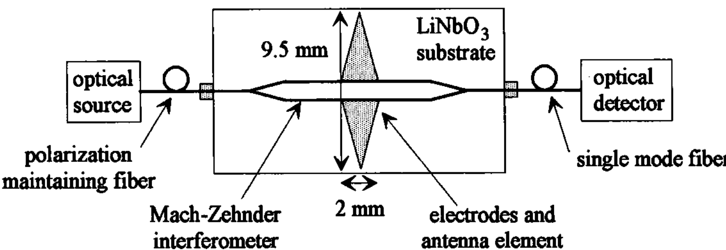


Figure 7-22. Mach-Zehnder interferometer as wideband electrical field sensor with fiber-optical readout (after [72]).

Petermann and co-workers [73, 74] have pointed out that optimization of the modulator efficiency by increasing the electrode lengths and the antenna size is not the only way to achieve high electric field sensitivity. Instead, in a segmented structure (Figure 7-23), each antenna segment can be impedance-matched to a corresponding short electrode. With an overall modulation length of 30 mm segmented into 12 sections, they have demonstrated a sensitivity of 1 mV/(m$\sqrt{\text{Hz}}$). Moreover, for an application in clinical hyperthermia, the electrode segmentation can also be designed to minimize the influence caused by changes in the permitivity of the surroundings. Allowing a reduction in sensitivity, the performance of the sensor for medical applications can thus be optimized.

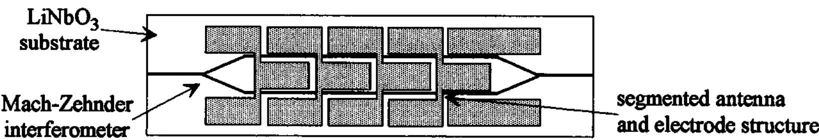


Figure 7-23. Mach-Zehnder interferometer with segmented antenna and electrode structure for enhanced sensitivity (after [73, 74]).

7.7 Optical Spectrum Analyzers and Wavelength-Selective Devices

Integrated optics provides a great variety of different operational principles for wavelength-dependent devices: among them are (resonant) coupler structures, electro-optical or acousto-optical mode converting filters, grating-based or acousto-optical Bragg-type devices, and waveguide resonators. Wavelength division multiplexing capabilities of fiber-optical communication systems have strongly pushed the development of devices such as wavelength channel selectors, routers, add-drop filters, and de/multiplexers. In the following sections, some of these devices and their applications for sensor systems are briefly discussed.

7.7.1 Tunable Wavelength Filters

Although tunable filter devices are not sensors themselves, they potentially can provide integrated optical signal (pre-)processing functions in sensor systems. For example, Xu et al. [75] have pointed out that acousto-optical tunable filters (AOTFs) may find use as key components for active wavelength demodulation systems (AWDS) which allow wavelength interrogation of fiber-optical strain sensor systems. As strain applied to a fiber grating will cause a shift in the wavelength of (Bragg) reflection, wavelength demodulation of the light returning from a fiber grating provides information on actual fiber grating microstrain. Integrated optical versions of tunable filters are ideal devices for such applications, as they are matched to fiber optics.

Exploiting the excellent acousto-optical properties of $LiNbO_3$, integrated optical, acoustically tunable wavelength filters make use of surface acoustic wave (SAW)-induced mode conversion. With an rf signal of frequency f_a applied to a transducer electrode structure (left part of Figure 7-24), a guided surface acoustic wave is excited. Optical TE/TM or TM/TE mode conversion is achieved if the phase-matching condition $\beta_{TM} - \beta_{TE} \pm \beta_a (f_a) = 0$ is fulfilled: the SAW propagation constant β_a must compensate for the difference $\beta_{TM} - \beta_{TE}$ in the propagation constants of the two optical modes to be converted. According to this condition, the wavelength of conversion can be selected by the choice of the SAW frequency f_a. Adjusting the applied acoustic power, complete conversion can be obtained. Filtering is achieved by blocking all light but the converted mode.

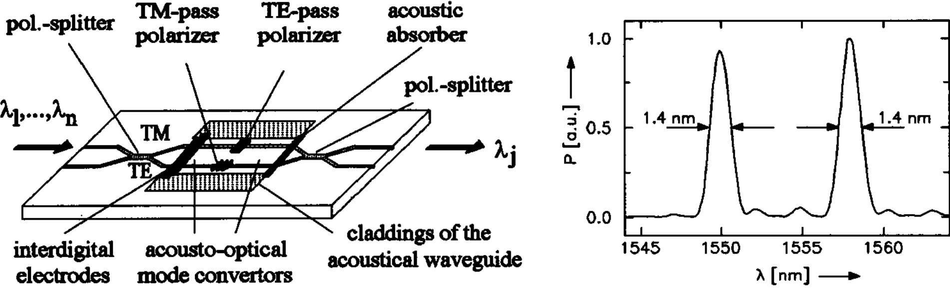


Figure 7-24. Left: polarization-independent integrated optical, acoustically tunable double-stage wavelength filter in $LiNbO_3$ (after [77]). Right: simultaneous two-wavelength filtering at 8 nm channel spacing is achieved by exciting two acoustic frequencies (173.40 and 174.38 MHz).

At the University of Paderborn, single- and double-stage acoustically tunable filter devices have been developed and fabricated, using on-chip polarizers for the selection of the converted mode. In a polarization-dependent, single-stage configuration, a filter band width of 2.8 nm has been achieved. The center wavelength was broadly tunable from 1.45 to 1.57 μm with a tuning slope $\partial\lambda/\partial f_a$ of 8 nm/MHz [76]. A very low electrical rf driving power of 90 mW (ca. 9 mW acoustic power) was required. Inherent in the acousto-optical mode conversion, there is an optical frequency shift of the converted optical mode equal to the SAW frequency f_a. This shift is in the range of a few 100 MHz and can be exploited for heterodyne measurements, as discussed for the frequency shifting device in Section 7.3.1. For filter applications, the optical frequency shift can be compensated by a second conversion stage. This scheme was used to realize a polarization-independent, double-stage, tunable filter on $LiNbO_3$ [77] (left part of Figure 7-24) with zero net frequency shift. Following the principles of polarization diversity, the incoming optical signal is split into two channels according to its polarization components TE and TM. These are fed into the double stage polarization-dependent TE and TM pass filters [78], respectively. Finally, the two waves are recombined. With this device, multi-wavelength filtering was demonstrated with a band width of 1.4 nm (right part of Figure 7-24); fiber-to-fiber insertion losses of the packaged and pigtailed device did not exceed 4.8 dB for both polarizations.

Acousto-optic mode converter devices on $LiNbO_3$ can also be configured as wavelength-selective switches [79], as frequency shifters (as mentioned above) and as "effective index" refractometers for chemical sensing applications ([130]; see also Section 7.9.1). Hence this family of devices shows potential for a wide range of applications in integrated optical and fiber-optical sensing systems.

7.7.2 Multi-Channel De/Multiplexers and On-Chip Spectrometers

Apart from tunable integrated optical add/drop filters multi-channel de/multiplexers and on-chip spectrometers are also of interest for optical communication and for fiber sensing. Such devices have been realized either with the help of an interferometric scheme using a bent bundle of waveguides with tapered optical length or by incorporating a diffractive grating into a planar waveguide. In the first approach (left part of Figure 7-25), light fed into one input

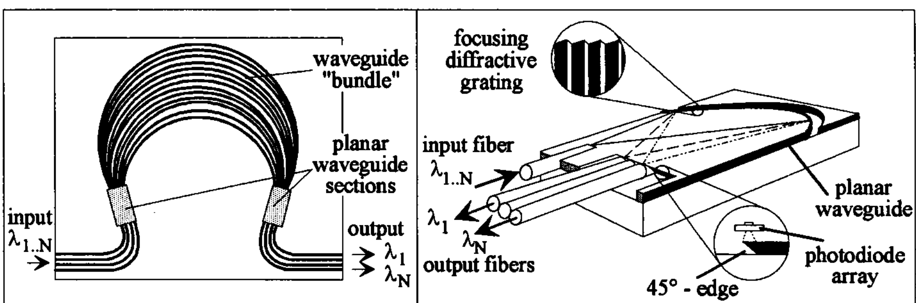


Figure 7-25. Two types of on-chip spectrometer: left, "waveguide grating" type (after [80, 81]) consisting of a bundle of waveguides tapered in length; right, diffracting and focusing grating in a planar waveguide, configured either with output fibers or with a hybridly fixed photodiode array (after [82]).

port is distributed to the waveguide bundle (also called "waveguide grating") via a planar waveguide section. In the second planar waveguide section, the light is recombined. Owing to the tapered waveguide length in the bundle, the phase front of the recombined optical wave depends strongly on the wavelength; light of different λ is therefore coupled to different output ports, which are formed by waveguides again. Based on this configuration, Zirngibl et al. at AT&T have developed a seven-channel WDM receiver [83] and a 12-wavelength optical multi-channel controller [84] on InGaAsP/InP. Using silica-on-silicon waveguide technology, NTT has achieved polarization insensitivity of a 16-channel multiplexer by hybridly introducing a TE/TM mode converting waveplate [85]. Very good results have been achieved by Philips and Delft University of Technology; their low insertion loss, low cross-talk, polarization-independent de/multiplexer device called PHASAR [80, 81] is based on InGaAsP/InP raised strip waveguides, and has a channel band width (spacing) of 0.9 μm (2 μm) and a cross-talk of better than −23 dB.

The second type of de/multiplexers make use of grating structures incorporated into a planar waveguide. Such on-chip spectrometers (right part of Figure 7-25) have been realized in the InGaAsP/InP material system at BNR Europe [86], at Bellcore [87], and at Siemens (Munich) [88]; also at Siemens, a "flat-field spectrograph" device [89] has been demonstrated, which is based on SiO_2-Si waveguides fabricated by flame hydrolysis. All de/multiplexer devices mentioned above have in common that they operate around $\lambda = 1.55$ μm; according to the intended use in optical communiations, their channel spacing/wavelength resolving power is matched to the needs of non-coherent, medium, and dense WDM systems, which range from a few nm to 0.5 nm per channel.

To our knowledge, only the development of a miniaturized waveguide spectrometer at KfK Karlsruhe is directed towards sensing applications [90]. Fabricated in PMMA by the LIGA technique (X-ray lithography, electroplating and moulding) [91], a metal-coated focusing diffraction grating is vertically incorporated into a planar three-layer polymer waveguide. Incoupling is achieved by a multi-mode fiber-optical link. Light diffracted by the grating is imaged via a totally reflecting 45° waveguide edge towards the photodiode array hybridly fixed face down on top of the waveguide. For the spectral range 400–1100 nm the device provides high transmission figures between 0.13 and 0.24 and a resolving power of 5 nm per array pixel [82]. As one possible application, the spectrometer chip was used for spectrally resolved interrogation of opto-chemical layers sensitive to ammonia. In various other applications such as fluorescence sensing, flame analysis, and reflection spectroscopy, the device has the potential to replace bulky spectrometer set-ups and will thus make small-sized spectral sensing systems feasible. What is more, the device is the central component of a complex microsensing system recently under development at the KfK Karlsruhe (for a detailed description of the system, see Chapter 4 of this volume).

7.7.3 Spectrum Analyzers

Waveguide Fabry-Perot resonators combined with electro-optical phase modulators can be used as spectrum analyzers (Figure 7-26). Their resolving power exceeds that of the integrated optical, acoustically tunable filters mentioned above ($\Delta\lambda/\lambda < 10^{-3}$), but it does not reach the extremely high resolution of bulk optical piezo-mechanically tuned Fabry-Perot resonators owing to waveguide losses. Nevertheless, interesting features of IO spectrum analyzers are rug-

gedness, freedom from moving parts and the high scanning speed achievable. Based on a low-loss Ti-$LiNbO_3$ stripe waveguide resonator with lumped electrodes, a pigtailed spectrum analyzer for $\lambda = 1.2-1.3$ μm with a resolution of 150 MHz was demonstrated [92] (Figure 7-26), which could be tuned over a free spectral range of 9.8 GHz within 300 ns. The device is particularly well suited for real-time spectral characterization of pulsed lasers or laser-like sources (such as the integrated optical parametric oscillator (see Section 7.9.2)). To compensate for the waveguide losses, resonator structures with gain for $\lambda \approx 1.55$ μm have also been fabricated by Er indiffusion in $LiNbO_3$ [93]. Increasing the optical gain towards the lasing threshold spectrum analyzer, resolution and fan-out are significantly enhanced [94].

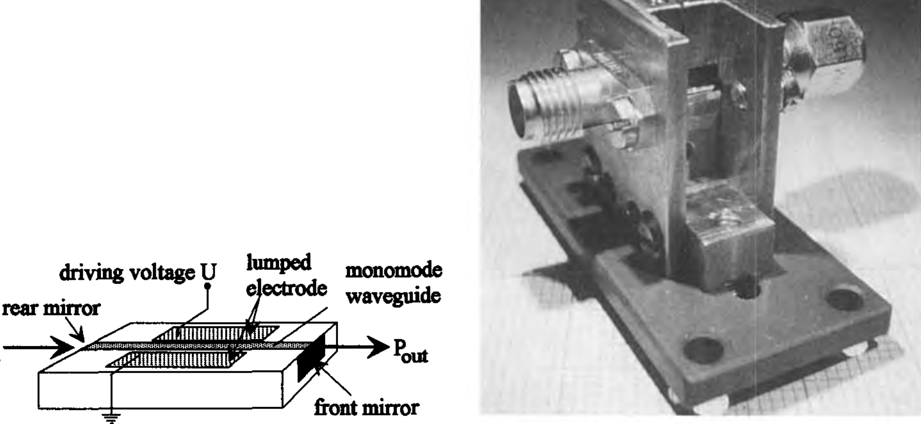


Figure 7-26. Left: integrated optical spectrum analyzer as used in [92]. Right: photograph of the device before fiber pigtailing.

7.8 Chemical Sensors and Biosensors

Chemical sensors and biosensors make use of sensitive layers, which selectively change their properties in the presence of substances to be measured. For example, organically modified polymers or immobilized dye molecules can be used as chemically sensitive layers for hydrocarbons [127, 128] or ammonia gas [123, 124], respectively (see Section 7.9.1). Biosensing of antigens, hormones, drugs, etc., can be performed with very high selectivity if the sensitive layers consist of immobilized biological molecules such as antibodies or receptors which are capable of recognition and binding (see, eg, [95, 96] for different immunoassay formats). These surface-related interactions cause changes in properties (refractive index, optical absorption, adsorption layer thickness, etc.) only in a small region near the interface. Therefore, detection schemes with enhanced sensitivities in the interface region are required.

By using guided optical waves, the evanescent field – the exponential decay of the electric field strength into the superstrate – provides an appropriate spatial discrimination. This

"evanescence field sensing" scheme can be exploited with different optical techniques: in fluorescence measurements, either the exciting wave or the fluorescent wave, or both, may be guided. Optical absorption of the evanescent field gives rise to increased waveguide loss, which can be detected by intensity-encoded measurements. Refractive index changes in the probe region, caused by the response of sensitive layers or an adsorption layer, will detune the propagation velocity (ie, the effective refractive index) of the guided optical mode, which can be interrogated with several measurement techniques (see Section 7.8.1). Luminescent quenching is another evanescent sensing method, which was used, eg, for the realization of a glucose sensor [2, 97]. As Lukosz [95] has pointed out, the method of surface plasmon resonance (SPR), widely used for biosensing [98], is also based on the evanescent field of an optically excited plasmon as a collective oscillation of electrons. A classification of transduction and readout principles has been given by Lambeck [99].

It is common to all evanescent field methods that their sensitivity is strongly affected by the fraction of power contained in the evanescent field or, for interface-related sensing, by the field strength at the interface. General design and optimization rules have been given [100, 101] for the widespread integrated optical sensor configuration with planar step index waveguides.

Today, research and development efforts on chemical sensing seem to be focused on biosensor devices, particularly on immunosensors, which meet market needs such as low cost, ease of use, and picomolar sensitivity, as discussed in [102]. As there is an extensive literature on biosensing techniques explaining the chemistry of sensitive layers, optical detection and readout schemes are emphasized in the following.

7.8.1 Effective Refractive Index Sensing

Grating Coupler Refractometer

Using very thin, planar, high refractive index TiO_2-SiO_2 waveguides with a large evanescent field, Lukosz and co-workers first observed the evanescent field sensor effect [103]: property changes in the evanescent field region such as a change Δn_c of the cover index affect the effective refractive index of the guided optical mode. Owing to the pioneering work of Lukosz and co-workers on waveguide materials fabricated by the sol-gel process and on embossing techniques for surface relief grating as coupling elements, disposable, single-use waveguide grating chips with different sensitive coatings for chemical and biosensing applications are commercially available (Artificial Sensing Instruments (ASI), Zurich, Switzerland), in addition to a grating coupler system for use as an "effective index refractometer".

Using waveguide gratings for optical coupling of light into a planar waveguide, the coupling angle depends strongly on the effective refractive index n_{eff} of the guided optical mode. By tracking the optimum coupling angle either cover index changes of a (homogeneous) superstrate down to 5×10^{-5} [104] can be resolved. Or in the case of an adsorption process, the adlayer thickness can be monitored in real time with a surface coverage resolution of a few pg/mm^2, as demonstrated for a monomolecular avidin adlayer [104]. To calculate the cover index or adlayer thickness changes from the variation of the coupling angle, the waveguide parameters are required. Thickness and refractive index of the waveguiding film are determined first by measuring the absolute (input or output) coupling angle for both TE and TM polarized modes.

The advantages of different coupling schemes as input and/or output coupling have been extensively discussed in the literature (see, eg, [95]). A new coupling scheme without moving parts [105] makes use of convergent or divergent illumination of a waveguide grating; for the angle fulfilling the coupling condition, a "dark line" results in the reflected spectrum, which can be tracked with a precision of $\Delta n_{eff} = 5 \times 10^{-4}$.

A very interesting device called an "integrated optical light pointer" (Figure 7-27) has been realized by Kunz and co-workers [106, 107]; they succeeded in fabricating a broad, uniform grating into a Ta_2O_5 waveguide film laterally tapered in thickness. If the device is illuminated under a well defined angle, the incoupling condition is fulfilled only for one lateral position. Detected via a second output grating, a lateral shift of the coupling position quantitatively indicates the refractive index of the cover material. In Figure 7-27, shifts are caused by coverage with either water or ethanol. This conversion scheme provides a resolution for the cover refractive index of 10^{-4}.

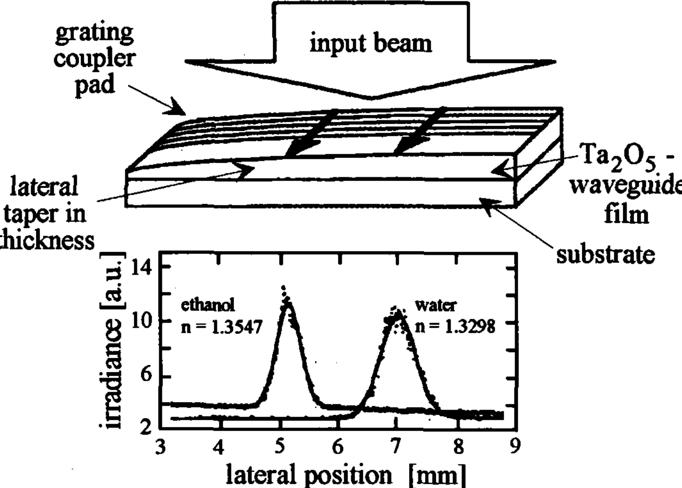


Figure 7-27. Integrated optical light pointer as realized at PSI: a uniform grating coupler in a planar waveguide with tapered thickness converts the value of the cover index into a lateral beam position (after [106]).

Very high sensitivity was achieved by Fattinger and co-workers [108, 109]; they succeeded in preparing a bidiffractive grating in glass which then was coated with a compact amorphous 140 nm thick TiO_2 layer as waveguide. Owing to the two grating periodicities (350 and 450 nm), two first-order output coupling beam directions are generated, one of which is used as an angle reference. With this measurement scheme an interfacial mass load sensitivity down to pg/mm^2 is obtained.

Difference Interferometer Scheme

Another readout scheme, which does not make use of grating coupling, is the difference interferometer or polarimetric interferometer [110-112] already mentioned in Section 7.4.2 (set-

up as shown in Figure 7-18 without the "refractive index shifting element"). Recently, this scheme was utilized by several groups [111, 113] for monitoring antibody-antigen immunoreactions. For the interaction of anti-human IgG and human IgG, Schlatter et al. [113] observed a broad dynamic concentration range extending over six orders of magnitude and a detectable surface concentration of 19 pg/mm^2, which is equal to 4/1000 of a single monolayer. The high sensitivity is limited only by the refractive index stability of the waveguiding layer. Using substrate materials such as fused silica or oxidizing silicon wafers, Stamm and Lukosz [111] have demonstrated that higher firing temperatures (800–900 °C) for the second step of the sol-gel process (transition from gel to amorphous layer) drastically reduce the microporosity of the TiO_2-SiO_2 waveguides and thus help to increase stability.

Waveguide Interferometers

Instead of using planar waveguides, integrated optical Mach-Zehnder interferometers have also been used for chemical sensing [114, 129] (Figure 7-28). The device surface is coated for passivation with the exception of a sensing window over one interferometer arm. Applications such as glucose sensors [115], hydrocarbon gas sensors [127, 128] (see Section 7.9.1) and biosensors [116, 117] have been demonstrated.

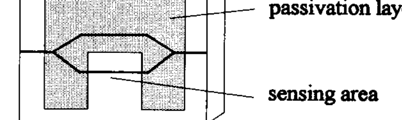


Figure 7-28.
Mach-Zehnder interferometer for chemical sensing.

Depending on the substrate material, these devices are typically fabricated by ion-exchange, diffusion or doping techniques leading to graded index waveguides with a small evanescent field strength. The dependence of the effective refractive index of the guided optical mode on the cover index is typically one order of magnitude lower than with thin high-Δn step index waveguides. This disadvantage of graded index Mach-Zehnder interferometers is compensated, however, by the long feasible interaction length of a few centimeters, which must be compared with typical interaction lengths of ca. 1 mm for grating couplers [95]. Thus, integrated optical Mach-Zehnder interferometers in principle provide the potential to realize small-sized, high-sensitivity chemical sensor and biosensor devices.

7.8.2 Fluorescence Methods

Fluorescence detection is another optical sensing scheme, which exploits the spatial discrimination of the evanescent field. For example, Zhou and co-workers [118, 119] have realized a fluorescence immunoassay sensor in which fluorescence is excited by an evanescent wave. On a patterned ion exchanged glass waveguide six different antigens are immobilized in separate window areas. Observing the fluorescence signal of each window through the device substrate, a multi-component analysis of the fluorescently labelled analyte was achieved; sensitivities of sub-µg/ml were reported, which are limited by the background fluorescence.

A very interesting commercially driven development is the fluorescent capillary fill device (FCFD) [120]. It consists of two sheets of glass spaced 100 μm apart. For immunosensing, one sheet of glass is covered with an immobilized capture antibody. After completion of the immune reaction, the labelled antibody-antigen complex is optically excited and fluoresence radiation from the interface region is collected as a guided wave and is detected at the sheet end-face. As the cell fills "automatically" by capillarity, a separate measurement of volume is not required. PSA (prostate-specific antigen) detection, required for prostate tumor diagnostics, has been demonstrated with ng/ml sensitivity [121].

7.9 Gas Sensors

For various gases such as H_2O, H_2, NH_3, CO_2, and hydrocarbons, integrated optical gas sensors have been proposed and realized. The majority of the devices are chemical sensors making use of the evanescent field detection scheme (Section 7.9.1). Other types of devices directly measure free-path optical absorptions to detect gas concentrations (Section 7.9.2).

7.9.1 Evanescent Field Sensing

As the evanescent field contains only a small fraction of the guided optical power, chemical transducer layers are required for sensitive gas sensing with short sensor structures. Induced by the presence of a special gas species, changes in the absorption or refractive index of the sensitive layer can be taken as an indicator of the gas concentration of interest.

Nishizawa et al. [122] has demonstrated a hydrogen sensor with a sensitivity of 20 ppm ([2], p. 406). A WO_3 film is deposited on a waveguide on $LiNbO_3$, which acts as sensitive layer and changes its color and absorption with variation in hydrogen concentration. For selectivity towards H_2, an additional Pd film enhances the performance as it combines the function of a molecular diffusion filter for H_2 and a catalytic converter. To provide an intensity reference, light entering the device is split by a 3 dB coupler and also guided to an unaffected second waveguide arm.

With a similar waveguide scheme on a silicon substrate (Figure 7-29), an ammonia sensor for the range 5-80 ppm was realized by Schipper et al. [123]. On top of the sensing waveguide (SiON-on-Si technology), bromocresol purple dye in a polyurethane matrix is used as a sensitive layer which changes absorption with variation in NH_3 concentration. Other main features of the device are V-grooves and on-chip spot-size transformers for stable, efficient fiber-chip coupling.

Instead of using two waveguide arms, referencing can also be done by a measurement at a second wavelength. Using 600 nm as sensing and 700 nm as referencing wavelength, Klein and Voges [124] employed a two-wavelength detection scheme for an NH_3 sensor based on a multi-mode glass stripe waveguide with multi-mode fiber-optical interrogation. Using modified sol-gel techniques, bromocresol purple dye was immobilized as an ammonia-sensitive layer. Outstanding features of this device are the high sensitivity of less than 1 ppm NH_3, very good reversibility, high selectivity against gases such as SO_2, CO, CO_2, NO, and

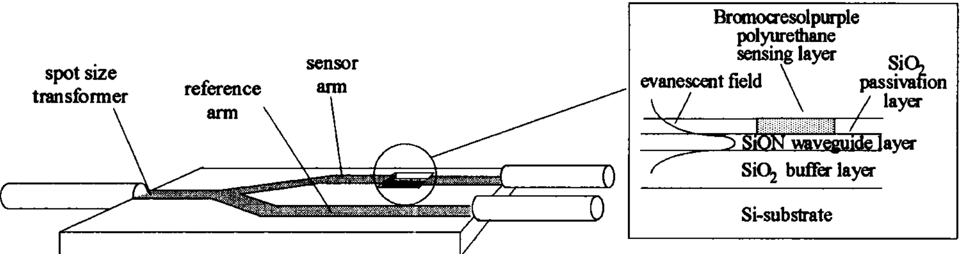


Figure 7-29. NH_3 sensor on silicon (after [123]). The y-branch device structure with sensing and reference arm (left part) is optimized for fiber-chip coupling by V-grooves and spot-size transformers (not shown). The inset shows the waveguide structure SiON-on-Si waveguide structure with a patterned sensitive layer on top.

N_2O, and short response times. In principle, this system meets all needs for continuous monitoring of NH_3 concentrations in cold-storage depots. Owing to this widespread application, sensing of ammonia potentially is a volume market for integrated optical gas sensors.

Interferometric guided wave structures such as Fabry-Perot resonators (FPRs) and Mach-Zehnder interferometers (MZI) have also been used for gas sensing. In this case, sensitive layers are utilized, which change their refractive index in the presence of a gas to be detected. Studies on CO_2- and SO_2-sensitive films fabricated from organically modified silicates on thermally K^+-Na^+-exchanged waveguides in BK7 glass have led to the development of a CO_2 sensor [125] which is capable of detecting CO_2 concentrations well below 2.5% within response times of a few minutes. However, poisoning of sensitive layers by NO_2 or NH_3 and also selectivity are still pending questions.

Based on a Mach-Zehnder interferometer chip fabricated by ion exchange in glass by IOT [126], Gauglitz and co-workers [127–129] investigated polymer layers, more precisely polysiloxanes, as a superstrate sensitive to hydrocarbons. Their results indicate that, as intended, sensor response is given for certain functional groups instead of single specimens of molecules. Having observed the Mach-Zehnder interferometer in the spectral region 760–810 nm, sensitivity limits of 2–20 ppm were evaluated for several types of halogenated and non-halogenated hydrocarbons, providing the feasibility of a hydrocarbon gas sensor system.

A completely different approach uses acousto-optical mode conversion as readout mechanism for refractive index changes of the sensitive layer [130]. Fabricating a waveguide by combination of Ti indiffusion and proton exchange on X-cut $LiNbO_3$, the evanescent field for the TE mode is much larger than that for the TM mode. Thus, mainly the effective refractive index of the TE mode responds to changes of the sensitive layer, whereas the properties of the TM mode remain almost unchanged and can be used as a reference. Due to the acousto-optical properties of $LiNbO_3$, TE-TM mode conversion can be achieved by means of a surface acoustic wave (SAW) if a phase-matching condition for the propagation constants β is fulfilled (see Section 7.7.1). Owing to this condition, the acoustic surface wave frequency f_a required for optimum conversion efficiency is very sensitive to changes in propagation constants, ie, the effective refractive index, of the two orthogonal polarized optical modes. As a feasibility study, detection of water vapor has been demonstrated with the acousto-optic refractometer using a thin DuPont Pyralin film as a sensitive layer.

Work on humidity sensors was also done by Lukosz and Stamm [131], using a thin dip-coated SiO_2-TiO_2 waveguide on an oxidized Si substrate and the difference interferometer already explained in Sections 7.8.1 and 7.4.2 (see Figure 7-18).

Kunz et al. [132] have reported on reactive low-voltage ion-plated microporous Ta_2O_5 films which simultaneously act as a waveguiding layer and as a sensitive transducer for water vapor. Compared with an evanescent field approach, combining waveguiding and sensing functions in a single layer provides enhanced sensitivity (assuming the same chemical response of the layer) as the whole optical mode experiences the property changes of the sensing layer (see Figure 7-30). In the sensor configuration, coupling light into the waveguide is achieved via a grating in the Ta_2O_5 film; owing to adsorption of water in the porous film, effective refractive index changes of the excited optical mode as small as $\Delta n = 4 \times 10^{-6}$ can be resolved by tracking the optimum in-coupling angle. With a flow gas cell above the waveguide, an excellent relative humidity resolution of $2 \times 10^{-3}\%$ has been demonstrated with very short response times at 25 °C.

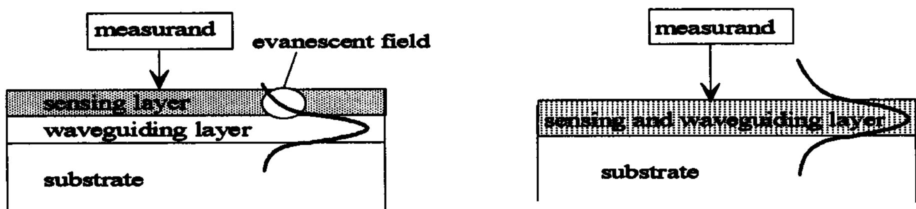


Figure 7-30. Left: evanescent field sensor with sensitive layer. Right: sensor with combined sensing and waveguiding layer (after [132]).

7.9.2 Free Path Absorption Sensing

Gas sensors using evanescent field methods achieve their small size by the strong transducing function of sensitive layers. Gas analyzers based on wavelength-selective or non-selective optical absorption, however, typically require absorption cells that are large in size compared with integrated optical sensor devices. To circumvent this restriction, Cerberus (Männendorf, Switzerland) and CSEM (Neuchatel, Switzerland) have realized an IO smoke detector (Figure 7-31) with 25 mm of free air-path integrated in the optical chip. Using buried multimode waveguides and splitters, light from two LEDs is detected on two photodetectors. By means of modulation techniques the optical power level transmitted by each of the four optical paths is determined independently. To measure the optical extinction of the free air-path caused by smoke particles, the transmission signal $A(f_1)$is normalized to the signal $B(f_1)$ of the reference channel. The other two signals ($A(f_1)$ and $B(f_2)$ allow to compensate for temperature effects like changes in LED output power and detector responsivity. Using LED power of 1 μW, a transmission measurement stability of 5×10^{-4} has been achieved for 25–40 °C with this device. Scaled to the short free optical path length, the stability figure is about half of the extinction threshold for a smoke alarm (4%/m). Owing to the multi-mode characteristics of the waveguides, the coupling tolerances of GRIN lenses, LEDs, and detectors are less demanding and low hybridization costs are feasible.

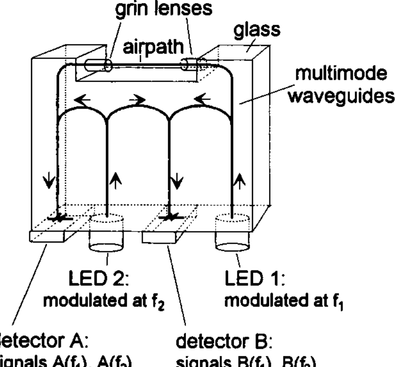


Figure 7-31.
Integrated optical smoke detector using buried multi-mode waveguides in glass (after [133]).

If macroscopic absorption cells are acceptable, integrated optical light sources may be used for the wavelength-selective detection of optical absorption. Based on non-linear optical conversion, the integrated optical parametric oscillator (IOPO) waveguide device on $LiNbO_3$ as realized by Suche and Sohler [134] exhibits a low pump threshold of only 25 mW and provides a broadly tunable, coherent, pulsed, near-infrared output compatible with the silica optical fiber mode. Using a computer-based transmission- and wavelength-measurement scheme, fiber-optical remote gas sensing of hydrogen fluoride (HF) gas was demonstrated [135] around $\lambda = 1.312$ μm, where the first molecular vibration overtone causes weak optical absorption. A sensitivity $\times$ absorption length product of 12 ppm (HF) $\times$ m was demonstrated for single-pass cells. Measured with a multi-pass White cell, the smallest HF concentration achieved was 300 ppb. This figure is equal to one tenth of the maximum allowable concentration (MAK) by (German) law for workplaces in industry. It was shown that the selectivity is excellent, owing to the high spectral resolution of $\Delta\lambda/\lambda = 6 \times 10^{-6}$ provided by the IO parametric oscillator. However, due to the lack of adequate solid-state semiconductor pump light sources, miniaturization of the optical set-up towards a sensor system is not yet feasible.

7.10 Optical Data Pick-Ups

Optical storage media such as the compact optical disk (CD) and read-and-write magneto-optical disk (MO) have become standard devices of computer equipment. Today, CD pick-up devices and MO read/write heads are fabricated by means of micro-optics using miniaturized actuators for tracking, focus control, etc. Whereas CD information bits are encoded in local reflectivity changes, in an MO disk system data are recorded magnetically on a magneto-optical film (TbFeCo, Bi-IG or Ce-IG) at optically induced elevated temperatures. Optical readout is achieved by the Faraday effect, ie, by the polarization rotation of a beam, reflected at the disk. However, as the rotation angle typically does not exceed 1 °, sensitive polarization detection schemes are required.

Based on grating couplers in glass waveguides, Osaka University has realized integrated optical heads for both storage systems. The CD pick-up device [136], as discussed [2] (p. 406),

provides a readout signal and tracking and focusing error information. The MO read/write head [137, 138] (Figure 7-32) is designed as a transmission device to be placed between an illuminating laser and the MO disk. The main idea is to make use of waveguide birefringence: the periodicities of a grating coupler are slightly different for coupling s- or p-polarized light into the TE or TM mode, respectively. An array of three focusing grating sections is fabricated on the glass waveguide; the central section couples p-polarized light to the TM mode and the two outer sections s-polarized light into TE modes. Hence this trifocal focusing grating coupler (TFGC) acts as a polarization splitter. TE and TM mode powers are equalized by appropriate choice of the incident polarization (45 °) of the illuminating beam; subtracting these signals, Faraday effect-induced polarization rotation can be detected with high sensitivity. Focusing and tracking error signals are provided by processing signals of the outer two section photodiodes.

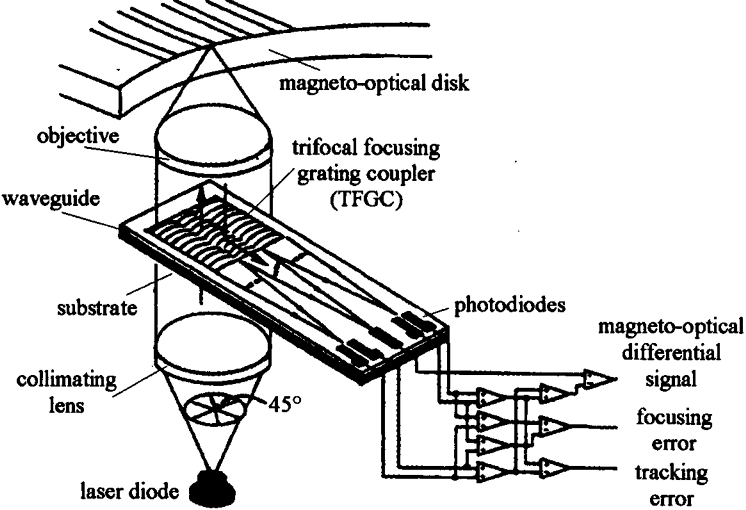


Figure 7-32. Magneto-optical disk readout device, realized by a polarization-dependent grating coupler array (after [138]).

An alternative concept was proposed by LETI [139]. The readout device was developed as a flying head, mounted on a slider at the end of a lever. As optical connects, one input and two output fibers are used. The basic approach also involves measurement and comparison of optical powers in the TE and TM modes. The sophisticated chip design, realized by SiO_2-Si_3N_4-SiO_2 waveguides on silicon (see Figure 7-5), contains a polarizer, a TE/TM polarization splitter, a TE/TM mode converter, a TM splitter and several curved mirrors. With slight modifications of the chip, readout and focusing functions have been demonstrated independently with several test circuits [140]. However, the flying head readout scheme is incompatible with removable magneto-optical disk storage devices, and thus no longer meets today's market requirements. At LETI, the integrated optical readout circuit, mentioned above, was therefore modified for a lens projection scheme, using separate IO units for tracking and autofocus control [17]. With these alterations, suitable performance was achieved [17].

7.11 Conclusions

Research and development activities in recent years have significantly enhanced the variety of different IO sensor devices and IO signal-processing devices for sensor systems. Reliable and efficiently working integrated optical sensor devices have been demonstrated for a wide field of applications including the detection of mechanical measurands such as displacement, vibration, acceleration, pressure, and rotation, chemical and biochemical sensing, measurement of electrical field strengths, and optical data pick-ups.

However, neither the maturity of planar and non-planar batch fabrication technologies, partly adapted from microelectronics, nor the outstanding performance features proved for a number of devices have led to overwhelming commercial interest in integrated optical sensors. Compared with the growing commercialization of integrated optical components for optical communications, there are only a few enterprises fabricating and dealing with IO sensors. On the other hand, in Europe there are well equipped, highly qualified, specialized research institutes and university groups which have successfully focused on the device developments with high industrial and commercial relevance. This discrepancy has two main reasons. First, integrated optics is based on a number of different substrate materials and fabrication technologies. Owing to this diversity, transfer from R&D status towards industrial manufacture is very costly. In the production process the variety of materials to be combined leads to high hybridization and packaging costs, which still account for a very large part of the overall manufacturing expenditure. Such limitations do not exist with more homogeneous silicon microelectronics. Second, the sensor market provided a large turnover of 12.8 billion DM for western Europe alone in 1993, which is expected to double by 2003 [141]; however, the market fields are strongly divided into various small segments, each dealing with different sensor functions and device types. Therefore, the major advantage of all batch fabrication technologies – the ability to produce larger numbers of devices at low cost – is of limited value.

In the future, a change of ideas and strategies may be helpful. Up to now, single fields of application have been served by individually developed sensing systems, involving various devices and basic technologies. As future sensor systems may be based on one powerful central high-tech component, the development of a broad range of application fields for such systems is demanded. For example, integrated optical interferometric displacement sensors can also be used for weighting, sensing of strain, measurement of thickness, etc. Further, integrated optical/fiber-optical gyrometers will find application in car-guiding systems as a potential volume market. Another helpful starting point is the development of integrated optical sensor platforms [142] compatible with microelectronics, which provide high application flexibility owing to the easy modular reconfiguration possibilities.

As the description of new developments in integrated optical sensors in this chapter has shown, powerful fabrication technologies and new device concepts are available, waiting for transfer to industry. The exploitation of this innovation potential for miniaturized sensors and smart sensor-actuator systems will be a prerequisite for a modern and competitive industry in the future.

7.12 Acknowledgments

The author thanks Professor Dr. W. Sohler, head of the applied physics/integrated optics group at the University of Paderborn, for many valuable discussions, Dipl. Phys. K. Schäfer for preparing most of the figures, and Mrs. I. Zimmermann for conscientiously correcting the manuscript.

7.13 References

[1] Göpel, W., Hesse, J., Zemel, J. N., (eds.), *Sensors – A Comprehensive Survey,* Vol. 6, *Optical Sensors,* Göpel, W., Dändliker, R., Spenner, K., (eds.); Weinheim: VCH, 1992.
[2] Brandenburg, A., Hinkov, V., Konz, W., in: [1], pp. 399-420.
[3] Nishihara, H., Haruna, M., Suhara, T., *Optical Integrated Circuits;* New York: McGraw-Hill, 1989.
[4] International Television Computer, *Study: Active Optical Integrated Circuits;* Boston: IGI Consulting, 1993.
[5] Parriaux, O., in: *Advances in Integrated Optics,* Martellucci, S., (ed.); New York: Plenum, 1994.
[6] Sohler, W., Volk, R., *Integrierte Optik – Potential für mittelständische Hersteller und Anwender von Mikrosystemen;* veröffentlichte Studie, erstellt im Auftrag des VDI/VDE-Technologiezentrums-Informationstechnik GmbH Berlin, August 1990.
[7] Kersten, R. Th., (ed.), *Fachausschußbericht,* Vol. 74; Frankfurt: Deutsche Glastechnische Gesellschaft, 1992.
[8] Fuest, R., Fabricius, N., Hollenbach, U., Wolf, B., *Proc. SPIE* **1794** (1992) 352-365.
[9] Grübel, H., Nitsch, G., *F&M Feinwerktechnik Meßtechnik* **100** (1992) p. 441-444.
[10] Fuest, R., *Technisches Messen TM* **58** (1991) 152-157.
[11] Voirin, G., Falco, L., Boillat, O., Zogmal, O., Regnault, P., Parriaux, O., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993,* (1993) 12-28.
[12] Parriaux, O., *Proc. SPIE* **1506** (1991) 111-119.
[13] Roth, R., *Proc. SPIE* **1141** (1989) 169-173.
[14] Ulrich, R., Ankele, G., *Appl. Phys. Lett.* **27** (1975) 337-339.
[15] Valette, S., Renard, S., Jadot, J. P., Gidon, P., Erbeia, C., *Sensors Actuators A* **23** (1990) 1087-1091.
[16] Ulbers, G., *Measurement* **9** (1991) 13-16.
[17] Valette, S., personal communication.
[18] Bauer, J., Dammann, E., Fritsch, E., Rasch, A., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 12-25.
[19] Jestel, D., Baus, A., Voges, E., *Electron. Lett* **26** 15 (1990) 1144-1145.
[20] Hofstetter, D., Zappe, H. P., *Paul Scherrer Intstitute Annual Report 1993* Annex III B (1993) 15.
[21] Ura, S., Shinohara, M., Suhara, T., Nishihara, H., *IEEE Photon. Technol. Lett.* **6** (1994) 239-241.
[22] Nishihara, H., *Anritsu News (Reprint)* (1989) 6-13.
[23] Sheard, St., Suhara, T., Nishihara, H., *IEEE J. Lightwave Technol.* **LT-11** (1993) 1400-1403.
[24] Ura, S., Suhara, T., Nishihara, H., *IEEE J. Lightwave Technol.* **LT-7** (1989) 270-273.
[25] Nishihara, H., *1990 IEEE Region 10 Conference on Computer and Communication Systems, Hong Kong, 24-27 September 1990, IEEE TENCON'90* No. 1 (1990) 99-103.
[26] Toda, H., Haruna, M., Nishihara, H., *IEEE J. Lightwave Technol.* **LT-9** (1991) 683-686.
[27] Toda, H., Haruna, M., Nishihara, H., *Electron. Lett.* **22** (1989) 982.
[28] Toda, H., Kasazumi, K., Haruna, M., Nishihara, H., *IEEE J. Lightwave Technol.* **LT-7** (1989) 364.
[29] Toda, H., Haruna, M., Nishihara, H., *IEEE J. Lightwave Technol.* **LT-5** (1987) 901.
[30] Nishihara, H., Koyama, J., Hoki, N., Kajiya, F., Hironaga, M., Kano, M., *Appl. Opt.* **21** (1982) 1785.

[31] Sohler, W., *Eur. Pat.* 90105787.7, 1990, see also: Tian, F., Ricken, R., Sohler, W., *Proc. Optical Fiber Sensor Conf. OFS'93, May 1993, Florence* (1993), paper W3.6, 263-266.
[32] Tian, F., Ricken, R., Schmid, St., Sohler, W., in: *Proc. Congress Laser '93, München,* Waidlich, W. (ed.) Heidelberg, Berlin, New York: Spinger 1993, pp. 725-728.
[33] Nowag, M., Janzen, G., personal communication.
[34] Becker, P., Brinkmann, R., Dinand, M., Sohler, W., Suche, H., *Appl. Phys. Lett.* **61** (1992) 1257-1259.
[35] Tian, F., personal communication.
[36] Suchoski, P. G., Waters, J. P., Fernald, M. R., *Proc. SPIE* **1177** (1989) 13-23.
[37] Fischer, K., Müller, J., Hoffmann, R., Wasse, F., Salle, D., *IEEE J. Lightwave Technol.* **12** (1994) 163-169.
[38] Fischer, K., Zurhelle, D., Hoffmann, R., Wasse, F., Müller, J., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 12-7.
[39] Fischer, K., Müller, J., *Sensors Actuators B* **9** (1992) 209-213.
[40] Wagner, C., Frankenberger, J., Deimel, P. P., *IEEE Photon. Technol. Lett.* **5** (1993) 1257-1259; see also *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 12-10.
[41] Deimel, P. P., *Micromech. Microeng.* **1** (1991) 199-222.
[42] Deimel, P. P., "Principles and Structures using Micromachining Process in Microoptics and Optoelectronics", to be publ. in *Semiconductor Micromachining, Vol. I and II,* New York: Wiley.
[43] Wu, S., *Ph.D Thesis; Delft University of Technology,* 1992.
[44] Wu, S., Frankena, H. J., *OSA Techn. Dig. Ser.* **10** (1992) paper TuD2-1.
[45] Wu, S., Frankena, H. J., *Proc. SPIE* **1793** (1992) 83-89.
[46] Bezzaoui, H., Voges, E., *Sensors Actuators A* **29** (1991) 219-223.
[47] Voges, E., Bezzaoui, H., Hoffmann, M., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 12-4.
[48] Bezzaoui, H., Hoffmann, M., *Jahresbericht 1992/93, Lehrstuhl für Hochfrequenztechnik, Prof. Voges, Universität Dortmund* (1994) 73-77.
[49] Burcham, K. E., De Brabander, G. W., Boyd, J. T., *Proc. SPIE* **1793** (1992) 13-18.
[50] Vadekar, A., Nathan, A., Huang, W. P., *Microelectron. J.* **23** (1992) 471-477.
[51] Nathan, A., Bhatnagar, Y., Vadekar, A., Huang, W. P., *Proc. SPIE* **1793** (1992) 19-26.
[52] Huang, W., Shubair, R. M., Nathan, A., Chow, Y. L., *IEEE J. Lightwave Technol.* **10** (1992) 1015-1022.
[53] Lukosz, W., Pliska, P., *Sensors Mater.* **3** (1992) 261-280.
[54] Pliska, P., Lukosz, W., *Proc. SPIE* **1793** (1992) 259-272.
[55] Lukosz, W., Pliska, P., *Sensors Actuators A* **25-27** (1991) 337-340.
[56] Young, A. M., Xu, C., Huang, W., Senturia, St. D., *Proc. SPIE* **1793** (1992) 42-53.
[57] Lukosz, W., *Proc. SPIE* **1793** (1992) 214-234.
[58] Gabathuler, W., Lukosz, W., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 4-18.
[59] Pliska, P., Dangel, R., Lukosz, W., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 14-30.
[60] Böhm, K., Rodloff, R., in: [1], pp. 421-465.
[61] Killian, K., Maltenfort, A., *Proc. SPIE* **1795** (1992) 87-92.
[62] Suchoski, P. G., Jr., Boivin, G. R., *Proc. SPIE* **1795** (1992) 38-47.
[63] Dorn, R., Baums, D., Kersten, P., Regener, R., *Adv. Mater.* **4** (1992) 464-473.
[64] Auch, W., *Proc. SPIE* **2070** (1993) 1-9.
[65] Hotate, K., *Proc. Optical Fiber Sensor Conf. OFS'93, May 1993, Florence* (1993) paper Tu4.1, 89-95.
[66] Nishi, Y., Iwashita, T., Nishiura, Y., Washimi, K., Okamoto, K., Ooka, A., *Proc. Optical Fiber Sensor Conf. OFS'93, May 1993, Florence* (1993) paper Tu4.5, 109-112.
[67] Bismuth, J., Gidon, P., Revol, F., Valette, S., *Electron. Lett.* **29** (1991) 722-723.
[68] Valette, S., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 12-1.
[69] Brummack, H., Poteau, P., *IEEE J. Lightwave Technol.* accepted for publication.
[70] Choi, Y.-K., Sanagi, M., Nakajima, M., *IEE Proc. J.* **140** (1993) 137-140.

[71] Bonek, E., Hornbachner, D., Riedl-Bratengeyer, E., Hadrian, W., Jobst, R., *Electron. Lett.* **28** (1992) 1994-1995.
[72] Kuwabara, N., Ideguchi, T., Kobayashi, R., *Digest of IEEE Antennas and Propagation Society International Symposium, Chicago, IL, USA, 18-25 July 1992,* Vol. 2 (1992) 725-728.
[73] Meier, Th., Schüppert, B., Petermann, K., Kostrzewa, C., *Proc. Optical Fiber Sensor Conf. OFS'93, May 1993, Florence* (1993) paper Th3.3, 467-470.
[74] Meier, Th., Kostrzcewa, C., Schüppert, B., Petermann, K., *Electron. Lett.* **28** (1992) 1327-1329.
[75] Xu, M. G., Geiger, H., Archambault, J. L., Reekie, L., Dakin, J. P., *Electron. Lett.* **29** (1993) 1510-1511.
[76] Frangen, J., Herrmann, H., Ricken, R., Seibert, H., Sohler, W., Strake, E., *Electron. Lett.* **25** (1989) 1583-1584.
[77] Tian, F., Harizi, Ch., Herrmann, H., Reimann, V., Ricken, R., Rust, U., Sohler, W., Wehrmann, F., Westenhöfer, S., *IEEE J. Lightwave Technol.* **12** (1994) 1192-1197.
[78] Herrmann, H., Müller-Reich, P., Reimann, V., Ricken, R., Seibert, H., Seibert, Sohler, W., *Electron. Lett.* **28** (1992) 642-643.
[79] Herrmann, H., Smith, D. A., Sohler, W., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 10-1.
[80] Amersfoort, M. R., de Boer, C. R., Verbeek, B. H., Demester, P., Loyen, A., van der Tol, J. J. G. M., *IEEE Photon. Technol. Lett.* **6** (1994) 62-64.
[81] Amersfoort, M. R., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) Post-Deadline Paper No. 1.
[82] Müller, C., Mohr, J., *Interdisc. Sci. Rev.* **18** (1993) 273.
[83] Zirngibl, M., Chien, M. D., Koren, U., Joyner, C. H., Miller, B. I., Young, M. G., Presby, H. M., Meester, J., Stulz, L. W., *OSA Tech. Dig. Ser.* **3** (1994) paper FC4-1.
[84] Zirngibl, M., Joyner, C. H., *Integrated Photonics Research Topical Meeting IPR'94, San Francisco, CA;* OSA Post-Deadline Paper PD2-1.
[85] Inoue, Y., Ohmori, Y., Kawachi, K., Ando, S., Sawada, T., Takahashi, H., *OSA Tech. Dig. Ser.* **3** (1994) paper FC4-1.
[86] Thompson, G. H. B., Ojha, S. M., Clements, S., White, I. H., Asghari, M., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 2-12.
[87] Soole, J. B. D., Scherer, A., Leblanc, H. P., Andreadakis, N. C., Bhat, R., Koza, P. A., *Electron. Lett. 27* (1991) 132-134.
[88] Cremer, C., Schier, M., Ebbinghaus, G., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 2-10.
[89] Clemens, P. C., März, R., Reichelt, A., Schneider, H. W., *IEEE Photon. Technol. Lett.* **4** (1992) 886-887.
[90] Mohr, J., Anderer, B., Ehrfeld, W., *Sensors Actuators A* **27** (1991) 571-575.
[91] Rogner, A., Eicher, J., Münchmeyer, D., Peters, R.-P., Mohr, J., *J. Micromech. Microeng.* **2** (1992) 133-140.
[92] Suche, H., Sohler, W., Teichmann, H., *1st Eur. Conf. Quantum Electronics, EQEC'88, Hannover, Digest of Technical Papers* (1988) paper WeBB3.
[93] Baumann, I., Brinkmann, R., Buchal, Ch., Dinand, M., Fleuster, M., Holzbrecher, H., Sohler, W., Suche, H., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 3-14.
[94] Hiller, D., Suche, H., Sohler, W., to be published.
[95] Lukosz, W., *Biosensors Bioelectron.* **6** (1991) 215-225.
[96] Gauglitz, G., Brecht, A., Ingenhoff, J., Kraus, G., *Spektrum Wiss.* January (1994) 92-97.
[97] Hesselink, G. L. J., Kreuwel, H. J. M., Lambeck, P. V., Bovenkamp, H. J. v. d., Engbersen, J. F., Reinhold, D. N., Popma, Th. J. A., *Sensors Actuators A* **7** (1992) 363-366.
[98] Robinson, G. A., in: *Methods of Immunological Analysis, Vol. 1, Fundamentals,* Masseyeff, R. F., Albert, W. H., Staines, N. A., (eds.); Weinheim: VCH, 1993, p. 372.
[99] Lambeck, P. V., *Sensors Actuators B* **8** (1992) 103-116.
[100] Parriaux, O., Dierauer, P., *Opt. Lett.* **19** (1994) 508-510.
[101] Parriaux, O., Sixt, P., *Opt. Lett.* accepted for publication.
[102] Robinson, G. A., *Biosensors Bioelectron.* **8** (1993) xxxvii-xxxx.

[103] Lukosz, W., Tiefenthaler, K., *Second European Conference on Integrated Optics 1983, Florence;* IEE Conf. Publ. No. 227, London: IEE, 1983, pp. 152-155.
[104] Clerk, D., Lukosz, W., *Sensors Actuators B* **11** (1993) 461-465.
[105] Brandenburg, A., Gombert, A., *Sensors Actuators B* **17** (1993) 35-40.
[106] Kunz, R. E., Gale, M. T., Kempen, L. U., Appassito, C., Curtis, B. J., Kuhn, M., Stutz, R., Edlinger, J., Rudigier, H., Sixt, P., Mayor, J.-M., Freiner, D., Spichinger, U. E., de Rooij, N. F., *Paul Scherrer Institute Annual Report 1993* Annex III B (1993) 22.
[107] Kempen, L. U., Kunz, R. E., Appassito, C., Curtis, B. J., Schütz, H., Edlinger, J., Rudigier, H., *Paul Scherrer Institute Annual Report 1993* Annex III B (1993) 25.
[108] Fattinger, Ch., Gale, M. T., Curtis, B. J., Schütz, H., Heming, M., Otto, J., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 4-12.
[109] Fattinger, Ch., Mangold, C., Vögelin, D., Wehrli, P., Schlatter, D., *2nd Eur. Conf. Optical Chemical Sensors and Biosensors EUROPT(R)ODE II, Florence, April 1994* (1994) paper Tu1A.6.
[110] Lukosz, W., Stamm, Ch., *Sensors Actuators A* **25-27** (1991) 185-8.
[111] Stamm, Ch., Lukosz, W., *Sensors Actuators B* **11** (1993) 177-181.
[112] Fattinger, Ch., Koller, H., Schlatter, D., Wehrli, P., *Biosensors Bioelectron.* **8** (1993) 99-107.
[113] Schlatter, D., Barner, R., Fattinger, Ch., Huber, H., Hübscher, J., Hurst, J., Koller, H., Mangold, C., Müller, F., *Biosensors Bioelectron.* **8** (1993) 109-116.
[114] Boiarski, A. A., Ridgway, R. W., Miller, L. S., Bhullar, B. S., *Proc. SPIE* **1368** (1990) 264.
[115] Liu, Y., Hering, P., Scully, M. O., *Appl. Phys. B* **54** (1992) 18-23.
[116] Boiarski, A. A., Busch, J. R., Bhullar, B. S., Ridgway, R. W., Wood, V. E., *Proc. SPIE* **1793** (1992) 199-211.
[117] Drapp, B., Ingenhoff, J., Gauglitz, G., Hollenbach, U., *2nd Eur. Conf. Optical Chemical Sensors and Biosensors EUROPT(R)ODE II, Florence, April 1994* (1994) paper P1.5.
[118] Zhou, Y., Laybourn, P. J. R., Magill, J. V., De La Rue, R. M., *Biosensors Biolectron.* **6** (1991) 595-607.
[119] Zhou, Y., Magill, J. V., De La Rue, R. M., Laybourn, P. J. R., Cushley, W., *Sensors Actuators B* **11** (1993) 245-250.
[120] Robinson, G., *2nd Eur. Conf. Optical Chemical Sensors and Biosensors EUROPT(R)ODE II, Florence, April 1994* (1994) invited paper PL2.4.
[121] Fletcher, J. E., O'Neill, P. M., Stafford, C. G., Daniels, P. B., Bacarese-Hamilton, T., Cookson, A. D., Robinson, G. A., *Tumor Marker Update* **5** (1993) 99-101.
[122] Nishizawa, K., Sudo, E., Yoshida, M., Yamasaki, T., *Proc. Optical Fiber Sensor Conf. 1986, Tokyo,* (1986) 131-134.
[123] Schipper, E. F., Wijbrans, R. A., Lambeck, P. V., in: *Proceedings of Dutch Conference on Sensor Technology, University of Twente, Enschede, February 1994,* Lambeck, P. V., (ed.); 1994, pp. 237-24.
[124] Klein, R., Voges, E., *Sensors Actuators B* **11** (1993) 221-225.
[125] Brandenburg, A., Edelhauser, R., Hutter, F., *Sensors Actuators B* **11** (1993) 361-374.
[126] Ross, L., *Glastech. Ber.* **62** (1989) 285.
[127] Gauglitz, G., Ingenhoff, J., *Sensors Actuators B* **11** (1993) 207-212.
[128] Ingenhoff, J., Gauglitz, G., Fabricius, N., *Proc. SPIE* **1796** (1992) 51-61.
[129] Fabricius, N., Gauglitz, G., Ingenhoff, J., *Sensor Actuators B* **7** (1992) 672-676.
[130] Hinkov, I., Hinkov, V., *Sensors Actuators B* **11** (1993) 227-231; see also *IEEE J. Lightwave Technol.* **LT-11** (1993) 554-559.
[131] Lukosz, W., Stamm, Ch., *Sensors Actuators A* **25-27** (1991) 185-188.
[132] Kunz, R. E., Du, C. L., Edlinger, J., Pulker, H. K., Seifert, M., *Sensors Actuators A* **25-27** (1991) 155-159.
[133] Tompkin, C., Ryser, P., Roth, P., *7th International Conference on Solid State Sensors and Actuators, Yokohama, Japan, June 1993, Abstract of Late News Papers* **D12-09** (1993) 38.
[134] Suche, H., Sohler, W., *Optoelectron. Devices Technol.* **4** (1989) 1-20.
[135] Teichmann, H., Sohler, W., in: *OSA Proceedings on Advanced Solid-State Lasers, 1994, Vol. 20,* Fan, T. Y., and Chai, B. H. T. (eds.), Washington: Optical Society of America 1994, pp. 447-450.
[136] Ura, S., Suhara, T., Nishihara, H., Koyama, J., *IEEE J. Lightwave Technol.* **LT-4** (1986) 913-918.
[137] Sunagawa, H., Ura, S., Suhara, T., Nishihara, H., *Int. Symp. Optical Memory, Tokyo, September 16-18, 1987, Techn. Digest* (1987) paper FA-7.

[138] Suhara, T., Ishimaru, H., Ura, S., Nishihara, H., *IOOC'89, Kobe, Japan, Techn. Digest* 4 (1989) 80.
[139] Valette, S., Jadot, J. P., Gidon, P., Kevorkian, A., *Proc. SPIE* **864** (1987) 125-133.
[140] Lapras, V., Labeye, P., Gidon, P., *Proc. 6th Eur. Conf. Integrated Optics ECIO'93, Neuchatel, Switzerland, April 1993* (1993) 12-34.
[141] Pahlke, W., Pötter, H., *Tech. Messen* **60** (1993) 339-346.
[142] Joppe, J. L., Bekman, H. H. P. Th., Krijger, A. J. T., Lambeck, P. V., Albers, H., Chalmers, J., Holleman, J., Ikkink, T., van Kranenburg, H., Zhou, M. J., *Proceedings of Dutch Conference on Sensor Technology, University of Twente, Enschede, February 1994,* Lambeck, P. V., (ed.) 1994, pp. 189-193.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

8 Optical Microsensors

H. Bartelt, Institut für Physikalische Hochtechnologie e.V., Jena, Germany

Contents

8.1 Trends in Optical Sensing 260

8.2 Advanced Technologies 261

8.3 Application Trends 264
8.3.1 Examples of Microsensors 264
8.3.2 Possibilities for Optical Actuators 271

8.4 Outlook 272

8.5 References 273

8.1 Trends in Optical Sensing

Sensors are increasingly inportant for surveying technical and biological systems and also in medical diagnostics. Competition between different measurement principles for a specific variable is very intense. Optical or photonic sensors offer specific advantages, such as [1-4]

- contactless measurement;
- immunity against electromagnetic interference;
- intrinsic electrical insulation;
- improved explosion safety (no electrical sparks);
- small size;
- usuable in hazardous environments (high temperatures, corrosive);
- high accuracy (interferometry);
- possibility of distributed sensing.

Some typical application examples of successful optical sensors are

- light detection (flames, absorption analysis, spectroscopy, image detection in cameras);
- light barriers;
- position detection (with optical decoders);
- distance or surface measurements (triangulation, interferometry, focusing);
- speed measurement (laser Doppler anemometry);
- rotation measurements (optical gyroscopes);
- optical temperature sensing.

Not all of these optical sensors can be called optical microsensors. In the present context, optical microsensors are characterized as sensors using microelements or microcomponents with dimensions in the micrometer range. Such elements or components are optical fibers, integrated optics, micro-optical elements such as lenses, filters or gratings with micrometer structures, micromechanical elements and specific optoelectronic elements. Especially modern techniques for microstructuring offer the potential to make complex miniaturized elements with high precision. Some examples of such microsensors that have already found successful applications include a miniaturized pressure sensor for invasive medical applications [5], temperature sensing based on fluorescence decay time [6-9], optical fiber gyroscopes [10], an integrated optical interferometer for distance measurements [11, 12], and an optical humidity sensor [13]. The development of such complex optical sensors is often based on modified technologies and components already developed for other applications, especially in the field of electronics, telecommunications, or optical memory technology. Actual optical microsensors today are directed to specialized applications and are rarely found in mass markets owing to their high price.

Some general trends in the development of modern optical sensors can be identified as

- more efficient sensors (high accuracy at reduced price, long lifetime without service, immunity against electromagnetic fields);
- small and light sensors (miniaturization, integration);

- flexible sensors (smart, intelligent sensors with signal processing);
- sensor systems (optical sensor networks, optical sensors connected by an optical bus, multisensors, integration in microsystems, distributed sensors, embedded sensors, smart structures).

In addition there is a market for speciality sensors with new or extreme properties. Considering the strong research activities in recent years, new microsensor products can be expected in the future for example in

- high voltage applications: current sensors, voltage sensors;
- distributed sensors: sensors in electrical cables, embedded stress sensors;
- biosensors: measuring optical properties of biological materials.

Materials used in optical microsensors cover a great variety besides optical materials, from electronic solid-state materials, ceramic materials, crystals, and magnetic materials to piezoelectric, ferroelectric, pyroelectric, electrochemical and even biological materials. The success of optical sensors will increasingly depend on an optimized material combination and on the knowledge of systems technologies for the efficient production of such sensors.

Before discussing application examples, some relevant technologies will be mentioned in the following section.

8.2 Advanced Technologies

New technologies, especially those developed for telecommunication applications and microelectronics, had a strong impact on optical sensors. In this section some new developments relevant for optical sensing will be mentioned, namely integrated optics, fiber optics, microstructuring technologies and microelectronics.

Integrated optics have been realized in different material classes such as lithium niobate, glass, silicon, solid-state materials and organic polymers [12]. With lithium niobate, electro-optical elements such as switches and modulators can be easily integrated together with passive optical components. Integrated optics in glass have advantages because of the same material basis as fibers, so that efficient light coupling is possible. They have been used until now mostly for passive optical components and sensor applications, for example, as integrated optical interferometers. Integrated optics in solid-state materials such as silicon or gallium arsenate are under development and would allow simple integration of optical and electronic functions. Absorption is relatively high but acceptable for short path lengths in an integrated optical sensors. Technologies of integrated optics and integrated optical sensors are discussed in more detail in Chapter 7.

In the field of fiber technology, several special developments have been realized for sensor applications, including active fibers for fiber lasers and fiber amplifiers, fibers with specific non-linear properties such as Brillouin scattering and polarization-maintaining fibers [14-17]. Polarization-maintaining fibers are designed with a specific non-circular symmetric core or by using stress-inducing elements within the fiber that cause strong linear birefringence. Typical

representatives of polarization-maintaining fibers are known as "panda" or "bow tie" fibers because of their typical cross section (Figure 8-1). These fibers permit the propagation of polarized states without a change of polarization. Typical values of birefringence that have been achieved are in the range of 10^{-4} for the difference of refractive index. A highly birefringent fiber with different attenuation for both linear absorption states can be used as a polarizing fiber. The attenuation is caused by bending the fiber with a suitable radius of curvature or by using asymmetric metal absorbers close to the fiber core. Specific fiber properties can also be achieved by spinning the fiber during the production process. For low-birefringent fibers ("spun low bi fibers") this spinning process reduces further the residual birefringent effects. With spun highly birefringent fibers ("spun hi bi fiber") the effects of external stress or vibration in sensors have been reduced by a factor of 50 over conventional fibers [18, 19]. Local fiber properties are changed for sensor applications with fiber gratings. UV radiation illuminates a fiber portion with an interferometric grating pattern and causes a permanent index modulation within the fiber core. The reflection and diffraction properties may be than influenced by external effects such as stress or pressure [20–22].

Microstructuring technologies have been successfully applied for optical structures such as gratings, microlenses and synthetic optical elements [23]. Computer-controlled writing procedures allow the fabrication of optical elements that may combine several features such as

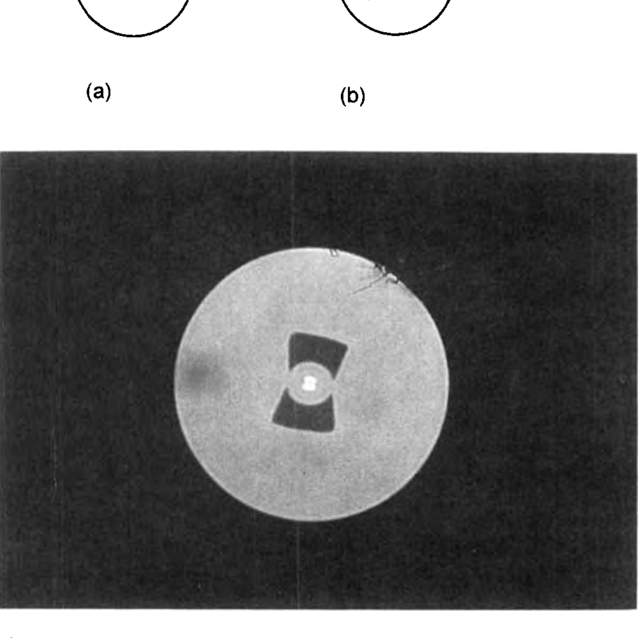


Figure 8-1. Polarization-maintaining fibers with stress inducing areas: (a) cross section of panda fiber; (b) cross section of bow tie fiber; (c) cross section photograph of bow tie fiber.

beam splitting, focusing, or spectral dispersion. Masks or structures may be written using laser scanning (micrometer resolution), synchroton radiation (LIGA process [24], sub-micrometer resolution), and e-beam lithography (nanometer resolution). Basic methods for structuring include anisotropic and isotropic etching in silicon, etching in quartz glass, recording in photoresist or in photopolymers and molding optical structures with metallic masters in plastic materials (eg, polymethyl methacrylate) (PMMA)) for mass replication. Technologies for microstructuring of silicon have reached a very high performance level because of the long experience of applications in microelectronics. With isotropic etching procedures in silicon, the etching rates are independent of crystal direction. These methods permit the fabrication of arbitrary patterns, but with restrictions on accuracy. With anisotropic etching the etching rates differ considerably for different crystal directions. The etching patterns achievable are usually defined by the very stable (111) crystal planes of silicon. Typical structures with their dependence on the wafer material orientation are shown in Figure 8-2. Anisotropic etching is limited in flexibility but offers extremely high mechanical precision. Doped layers can be used to stop the etching process so that diaphragms with well defined thickness are produced. Additional layers of different materials (eg, silicon dioxide, silicon nitride, metals) are used to produce underetched elements such as bridges or paddles.

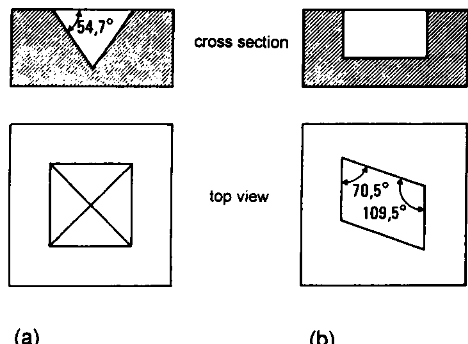


Figure 8-2.
Typical anisotropic etching patterns of silicon: (a) with (100) wafer; (b) with (110) wafer.

An extremely high aspect ratio in micromechanical components (ratio of depth to width of about 100) has been achieved with the LIGA process. Structural heights of several hundred micrometers and lateral dimensions down to 0.1 µm are possible. The process itself is based on X-ray lithography with synchrotron radiation, electroplating, and a molding technique for cost-efficient replication. The material for recording and for molding is mostly PMMA. A three-layer resist offers the possibility of realizing structured light-guiding components with a step index profile [24a].

Microstructuring technologies also make possible the realization of miniaturized mechanical structures, namely bridges, beams, or diaphragms. Elements with V-shaped channels are used for fixing optical microlenses and fibers. Such structures are also a basis for optical microplatforms, where different elements are mechanically positioned and fixed with high precision.

In the field of optoelectronics, cheap and efficient laser diodes are the basis for many optical sensors. Available power levels have increased considerably up to several watts for laser diode arrays, which is important for laser pumping and optical energy transmission. Of great interest for sensor applications are also research activities on tunable lasers pumped with

solid-state light sources that will cover a wide range of optical wavelengths, ie, from 600 to 1000 nm [25]. Especially for spectroscopic sensors such tunable light sources will increase considerably the flexibility and the range of applications. Photoelements are not only of interest as general light detectors but also as optoelectric energy transformers for use with electronic sensor elements. Photoelement arrays are able to deliver electrical power levels from an optical input at voltages as required in electronics (>3 V) and they are a central component in hybrid sensor systems.

8.3 Application Trends

8.3.1 Examples of Optical Microsensors

In this section we shall discuss some specific examples of modern optical sensors that illustrate current trends. These examples will include the concept of hybrid sensors, fiber-optic sensors for current, a spectral optical micro system, frequency-modulated sensors and a multisensor concept.

Hybrid Sensors

The concept of a hybrid sensor combines the advantages of electrical and optical systems [26–28]. It is based on optical energy transmission and optical data transmission for an electrical sensor element (Figure 8-3). In this way many commercial electrical or electronic sensor elements can be made to achieve essential properties of optical sensors such as electrical potential separation or immunity against high electromagnetic fields. Since optical data transmission is a well established technology, the important new aspect is therefore the possibility of energy transmission via fibers. Efficient light sources such as laser diode arrays are available up to power levels of several watts of optical power. For high coupling efficiency, fibers with relatively large core diameters (>100 μm) are preferable and coupling efficiencies of about 80% are feasible. Transmission losses of such fibers are typically in the range of about

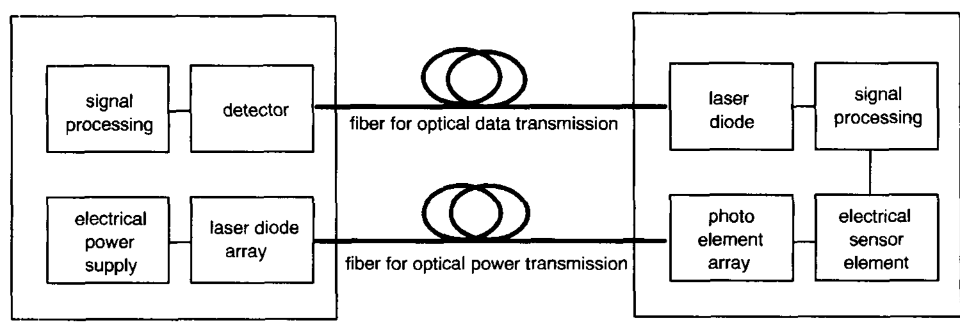


Figure 8-3. Principle of hybrid sensors.

2-4 dB km^{-1}. At the output end of the fiber an opto-electric coupler is required such as a photoelement. In order to achieve voltage levels as required for electronics, a combination of several elements in an array is an efficient and simple solution (Figure 8-4). Depending on the number of elements, different voltage levels can be synthesized. The efficiency of such photoelements today reaches up to 45%, depending on intensity. Therefore, a typical energy transmission efficiency of about 25% (optical output of light source to electrical input of sensor head) is possible. This means that with a 400 mW laser diode more than 100 mW of electrical power can be supplied with a single fiber. In principle, higher power levels are possible, but with the requirement of forced cooling of the light source. With increasing energy density within the fiber, care must be taken to avoid absorbing defects in the fiber or on a fiber end. Signal transmission is usually done by a signal-to-frequency conversion, but also serial digital transmission using a low-power A/D converter is possible. An alternative to photoelement arrays would be single photoelements with electrical transformers that can also be built in integrated form.

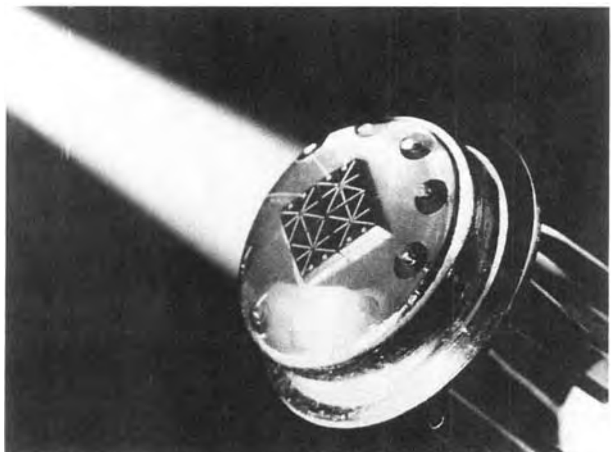


Figure 8-4.
Photoelement array (source: Siemens).

Fiber-Optic Current Sensor

Especially in high-voltage applications, optical sensors will be of great importance owing to the intrinsic electrical insulation. For magnetic field detection (or current measurement), sensors based on the Faraday effect are being developed and tested [29-32]. In a similar way, the Pockels effect (linear electro-optic effect) is under investigation for measuring electrical fields or voltages. Discussion will be concentrated here on the fiber-based Faraday sensor because this type of sensor is influenced strongly by new technologies in fiber optics. The basic Faraday effect can be described in the following way: linearly polarized light is turned in its polarization direction under the influence of magnetic field in materials such as glass (Figure 8-5). The rotation angle is proportional to the line integral of the magnetic field. If the light path is turned around an electric wire, the rotation angle α is then proportional to the electric current I and the number of turns N with the Verdet constant V as proportionality constant:

$$\alpha = VNI$$

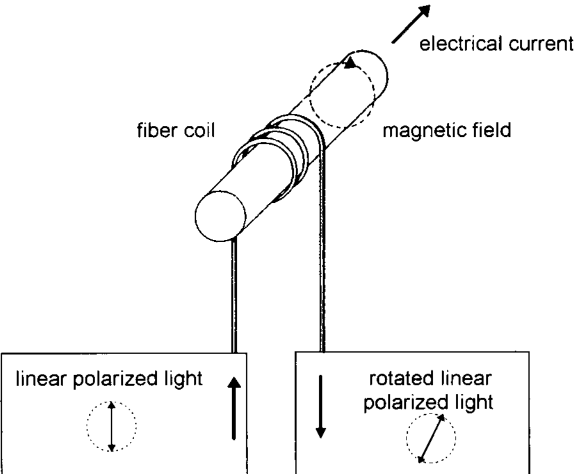


Figure 8-5.
Principle of a magneto-optic current transformer with fiber coil.

The main problem for the accuracy of such a sensor is the temperature and vibrational sensitivity. Unfortunately, the polarization state may be changed from linear to elliptical owing to internal linear birefringence. This effect changes the characteristics of the sensor and limits the accuracy. The approache most often applied is to use a glass block for building a glass ring around the electric conductor. The massive glass block is relatively stable in its mechanical properties, but shows temperature effects which limit the accuracy today to about 1%. The flexibility is limited because the realization of several light turns around the electrical conductor is not very practicable. With a fiber-optic coil the number of turns can be easily adapted to the desired current range. Unfortunately, the sensitivity of conventional fibers to temperature and vibration is fairly high. In the case of an alternating current it is possible to reduce the measurement errors of the AC part by using the low-pass filtered DC signal for normalization.

The temperature sensitivity of fiber coils is strongly reduced by using annealed fiber coils. To this end the fiber coil is heated to about 850 °C and is cooled to room temperature in a controlled way. This procedure removes residual stress and minimizes internal linear birefringence [27, 32]. The temperature sensitivity decreases by a factor of about 20 with this pro-

Figure 8-6.
Annealed fiber coil (source: Siemens)

cedure. Additionally, it is possible to decrease the temperature effect further by using the different effects at the two polarization output ends in a transmissive setup. By measuring the output polarization for zero current in one output end (which is a measure of the ambient temperature), the current sensitivity can be corrected for the second output end. Accuracies of better than 1% have been achieved. However, this correction method does not remove the vibrational sensitivity (Figure 8-6).

The vibrational sensitivity can be reduced by using spun fibers. In this case the system has to be used in the reflective mode to compensate for the effect of the turns within the fiber. The spinning can be done with a finished fiber or during the production process. If spinning is done in production with the fluid glass material, fairly high spinning rates are possible. The resulting strong circular birefringence suppresses the vibrational effects of linear birefringence. The vibration noise can be reduced by about a factor of 50 [18, 19]. In this arrangement, however, the temperature sensitivity still remains.

Microspectrometer

The LIGA process has been used to pattern light-guiding polymer layers for a miniaturized grating spectrograph [34, 35]. Several functions such as light guiding, imaging, dispersion, and deflection of the light on a diode array are performed with a single micro-optical component. The light from a wideband light source is coupled from a multi-mode fiber to a three-layer photoresist light guide. This light is directed to a self-focusing reflection grating and imaged on an output fiber array or a diode array (Figure 8-7). The resist layers (PMMA) and the con-

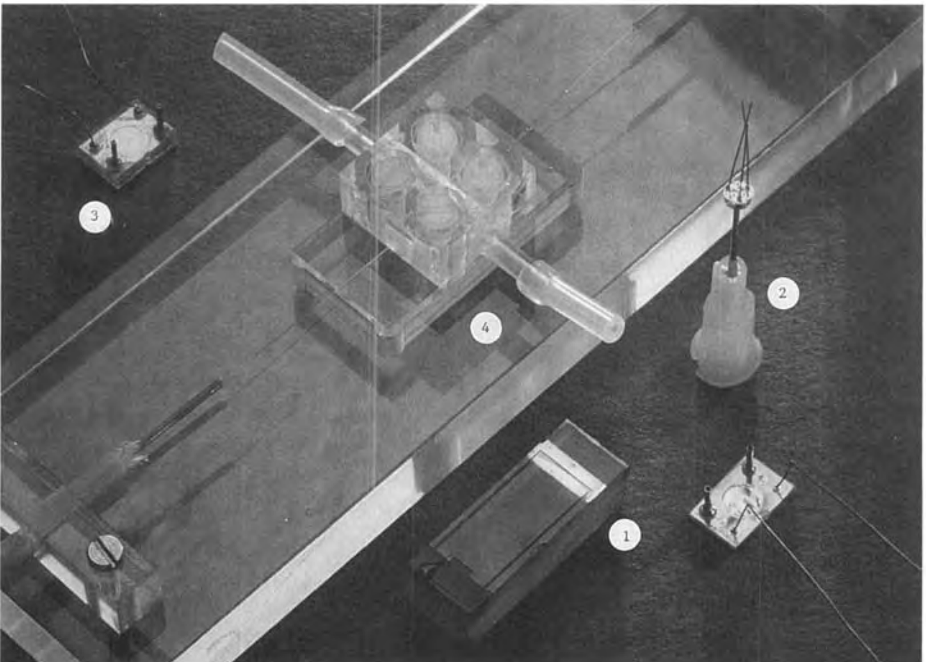


Figure 8-7. Components of a microanalysis system with microspectrometer (1), microvalve system (2), micropump (3), measurement volume (4) (source: KfK).

ical reflection grating are patterned by deep etch X-ray lithography. Refractive indices of 1.47 for the core and of 1.49 for the cladding (at 589 nm) are chosen to achieve the same numerical aperture as for the used fibers. A blazed grating structure with a grating constant of about 2.5 μm is coated with gold or silver for efficient light reflection. Several different grating structures have been produced to cover spectral ranges from 500 to 1100 nm with measured diffraction efficiencies of up to 24% and a spectral resolution of 5 nm. The grating spectrograph with a size of typically 17 mm by 6 mm demonstrates the possibilities of integration in micro-optical systems and is well suited for spectroscopic analysis in the visible and infrared regions.

Frequency-Modulated Sensors

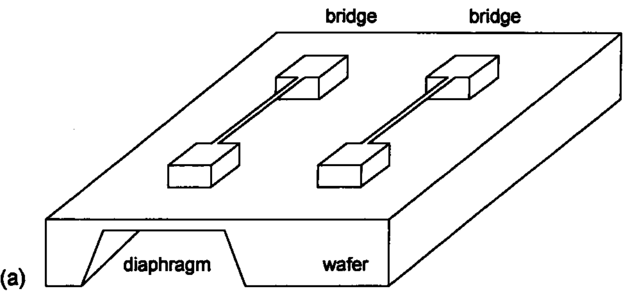
A severe problem for the accuracy of many intensity-modulated fiber sensors is attenuation (eg, because of bending losses, aging of components, and losses at couplers) in the transmission line. An approach avoiding such problems is to use frequency-encoding sensor schemes. For this purpose vibrational elements can be used, eg, based on quartz crystals or micromachined beams and bridges in silicon. The physical parameter to be detected changes the resonant frequency and this frequency is measured by modulating a light beam. The transmitted signal is then, to a first approximation, independent of attenuation effects in transmission.

In Figure 8-8a a micromachined example in silicon of a bridge structure is shown. The bridge structure is made of silicon nitride and built on a diaphragm with a typical length of 500 μm, a width of 6 μm and a height of 2.8 μm. A pressure applied to the diaphragm changes the forces on the bridge and therefore shifts the resonance frequency, which in this case is in the range of 150 kHz. The small size of the bridge structure permits optical activation based on photothermal effects with an optical power of about 0.25 mW. The vibration is measured using an interferometer owing to the small vibrational amplitudes of about 40 nm. In the case of resonators with larger amplitudes (eg, quartz fork), direct amplitude modulation of the reflected light can also be used. Figure 8-9 shows the resonance curve and the pressure dependence of the resonance frequency. The temperature dependence was then corrected down to a residual error of about 1% by measuring the temperature with an identical bridge structure on a solid wafer without a diaphragm. Similar bridge structures for temperature and pressure sensors have also been realized in silicon oxide [36–40].

Another frequency-encoding scheme uses the electric field dependence of the resonance frequency of quartz. In order to obtain this field dependence a specific cut of the quartz (X-cut) has to be used. A frequency change of about 1 Hz per 10 V has been achieved with a typical resonance frequency in the range 1–5 MHz. For measurement of the extremely small vibrational amplitudes of less than 10 nm in a complex mode pattern, a beam deflection method has been used [41].

Multi-Sensor Array

There are several possibilities for the multiplexed use of several sensors on a single fiber line: wavelength multiplexing, frequency multiplexing, polarization multiplexing, coherence multiplexing, or time multiplexing [42–45]. Logical addressing is only possible today with a hybrid concept and electrical logical processing at the sensor head.



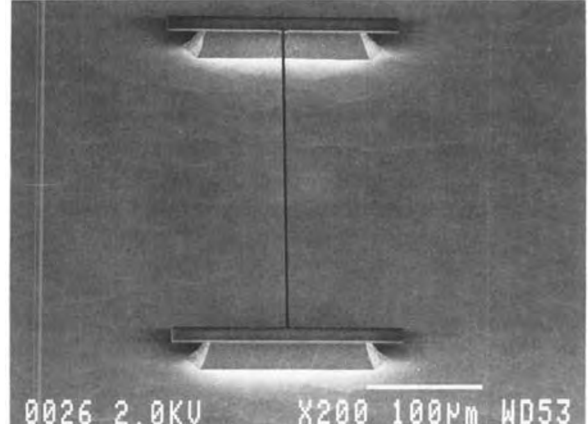


Figure 8-8.
Bridge structure for a frequency-modulated pressure sensor: (a) principle of setup; (b) picture of a silicon nitride microbridge (source: Siemens)

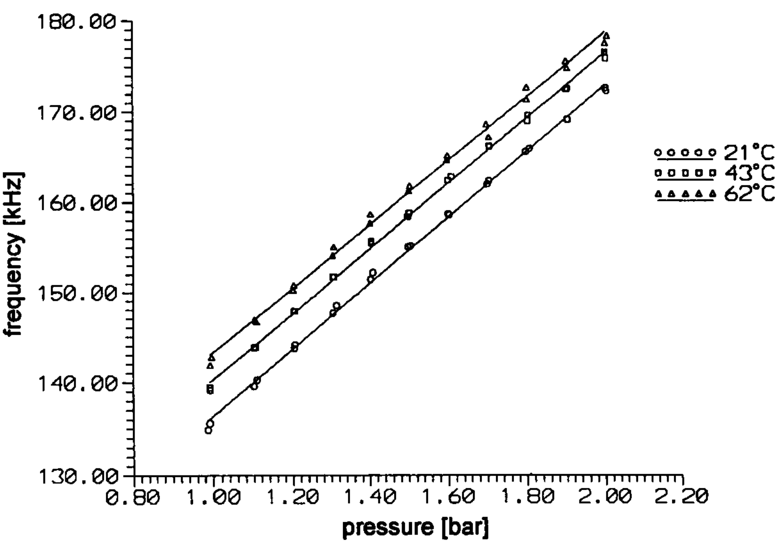


Figure 8-9. Frequency dependence of a silicon nitride microbridge.

Time multiplexing has been applied successfully in optical time domain reflectometry (OTDR) systems for distributed temperature sensing. A generalized concept using single optical sensor elements connected to a common fiber line is shown in Figure 8-10. The different

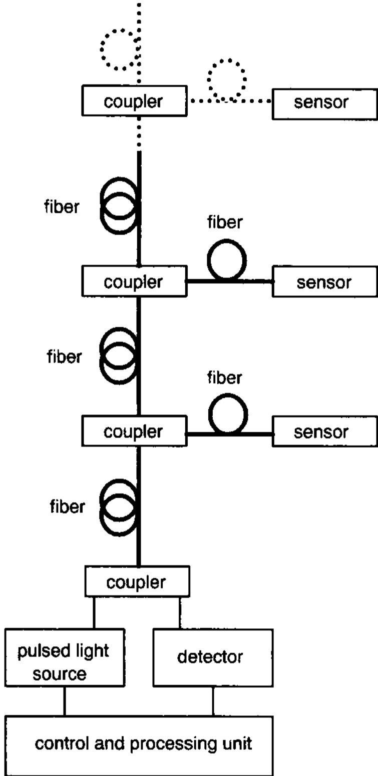


Figure 8-10. Concept of a multi-sensor array.

sensor elements can be distinguished using an optical pulse and measuring the travelling time of the reflected pulse. Each sensor element is characterized by its specific delay time Δt. The position resolution δx is given by the optical pulse length δt:

$$\delta x = (c/2n)\delta t$$

where c = speed of light in a vacuum and n = refractive index. For typical values (δt = 10 ns, n = 1.5), a spatial resolution of about 1 m is achieved. The length of the fiber line can be in the range of many kilometers. The number of addressable sensor heads is limited by the detectable intensity levels. Since each sensor element is connected by a coupler element, the intensity available for the following sensor elements decreases. In the case of identical couplers, the intensity levels will decay exponentially. By adjusting the coupling ratio depending on position, a constant intensity level at every sensor element can be achieved. In this case, the maximum number of sensor elements is given by the coupling ratio. The time resolution of the sensor signals is defined by the possible repetition rate ν of the optical pulses and hence by the length of the fiber line, L. With some typical values (L = 10 km, ν = 20 kHz), one can

achieve a sampling rate for the sensor signal that would allow the detection of time-dependent variables of up to 10 kHz. In Figure 8-11 an example is shown for two different sensors measuring a vibrational surface (5 Hz) and an electrical current (50 Hz) at the same time. The vibration sensor is based on distance-dependent light coupling into a fiber and the current sensor is based on mechanical movement of the fiber with a magnet due to magnetic forces of the current.

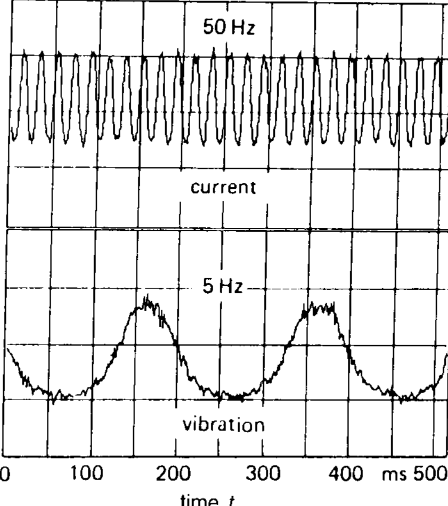


Figure 8-11.
Example of two different sensor signals from a multi-sensor array.

8.3.2 Possibilities for Optical Actuators

Optical actuators are closely related to mechanical sensors and may be considered as inverse sensors that transform an optical signal into a mechanical movement [47–49]. Such actuators may be combined with optical sensors, eg, for switching purposes or as a compensating mechanism for mechanical sensors. Referring to different classes of optical sensors, we can again distinguish between a hybrid and a pure optical concept. Optical energy supply as described for hybrid optical sensors is applicable to an optical actuator. Power levels are sufficient to switch relays or to control a piezoelectric transducer. The possibilities of direct optical to mechanical actuation are limited in practice because of the relatively low energy levels and conversion efficiencies of typical optical signals. However, micromechanical technologies permit the fabrication of systems that are miniaturized enough to run with the power levels obtained from optical signals. The most often used principle is the photothermal effect that produces a temperature-dependent expansion of solid, fluid, or gaseous materials by heating with the help of highly absorbing materials. Another physical principle based on light pressure also permits the application of mechanical force but it is used only for very low masses in optical levitation systems. This principle will therefore not be discussed further in this context.

The main advantages of optical actuators are similar to those of optical sensors, namely electrical insulation and immunity against electromagnetic interference combined with the small size of micromechanics and optical fibers.

An example for optomechanical actuation by using a thermopneumatic force is shown in Figure 8-12a. A micromachined membrane of bimetallic structure and carbon-wool as absorbing material is used. This membrane is part of a light-driven micropump. Light from a fiber heats a working fluid and moves the membrane. A low light power of about 11 mW is sufficient to deflect the membrane by about 30 μm. This setup was used in an array structure as a light-driven micropump to move fluids from one cell to the next and to achieve a pumping operation [47].

A similar setup permits the operation of a valve as in Figure 8-12b. In this case a flexible membrane opens or closes a micromachined valve. Although this system was powered by electrical heating, the required power levels of about 25 mW could be also delivered by optical means [48].

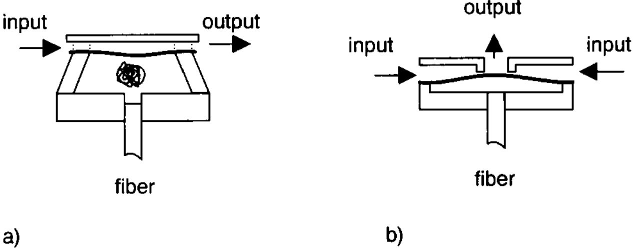


Figure 8-12. Optical actuators: (a) optical pumping element; (b) optical valve for gas or fluid.

A completely different concept for direct optical-to-mechanical actuation uses optothermal deflection of a fluid jet. Depending on the deflection, a differential pressure at two output ports is modulated [49]. Such an analog change of differential pressure was achieved using up to 100 mW of optical power. Since light is only used for changing the direction of the fluid jet, the actual mechanical force is achieved by the fluid pressure and can reach higher levels than with purely optical systems.

8.4 Outlook

Optical microsensors often take advantage of technologies developed in optoelectronics and telecommunications. Commercial applications are mostly directed to high-precision interferometric sensors or to applications where non-metallic or electrical insulation properties are required. Major problems still have to be solved to make possible sensor families with the possibility of combining them in sensor networks. Another problem is that no flexible optical signal processing is available today. Hence the development of smart optical sensors is restricted and so far requires optoelectrical signal conversion. The availability, for example, of optical switches for optical addressing of different optical sensor elements, of optical look-up tables for signal processing, or of an optical analog-to-digital conversion would help to improve the flexibility of optical sensors considerably. Component developments for digital optical signal processing have already started in the field of optical telecommunications and will hopefully give a further impetus to optical microsensors.

8.5 References

[1] Dakin, J., Culshaw, B. (eds.); *Optical Fiber Sensors: Principles and Components,* Boston and London: Artech House, 1988.
[2] Moore, E., DePaula, R., in: *Optical Fibers and Integrated Optics in Sensors,* Vol. 1, Grandke, T., Ko, W. H. (eds.); Weinheim: VCH, 1989.
[3] Kohn, D. A., *Fiber Optic Sensors;* Instrument Society of America, 1988.
[4] Adrian, P., *Sensors* (September 1991) 23-45.
[5] Trimble, B., in: *Proc. 9th Optical Fiber Sensors Conference, Florence, Italy;* 1993, p. 457.
[6] Hartog, A. H., Leach, A. P., *Electron. Lett.* **21** (1985) 1061-1062.
[7] Ogawa, K., Ozawa, Y., Kawakami, H., Tsutsui, T., Yamamoto, S., *Springer Proc. Phys.* **44** (1989) 544-551.
[8] Bosselmann, T., Reule, A., Schröder, J., *Proc. SPIE* **514** (1984) 151-154.
[9] Burgos, A. R., Allison, S. W., *Sensors* (March 1990) 17-25.
[10] Smith, R. B. (ed.); *Selected Papers on Fiber Optic Gyroscopes,* Vol. MS 8; SPIE, 1989.
[11] Fabricius, N., Ganglitz, G., Ingenhoff, I., *Sensors Actuators B* **7** (1992) 672-676.
[12] Kersten, Th., in: *Optical Fiber Sensors: Principles and Components,* Dakin, J., Culshaw, B. (eds.); Boston and London: Artech House, 1988.
[13] Schwotzer, G., Ecke, W., Hoffmann, K.-J., Hoppe, L., Reuter, Th., in: *Proc. 9th Optical Fiber Sensors Conference, Florence, Italy;* Post-deadline paper PD-8.
[14] Gambling, W. A., *Ann. Telecommun.* **41** (1986) 541-546.
[15] Kurashima, T., Horiguchi, T., Izumita, H., Furukawa, S., Koyamada, Y., *IEICE Trans. Commun.* **E76-B** (1993) 382-390.
[16] Bao, X., Webb, D. J., Jackson, D. A., *SPIE Proc.* **2071** (1993).
[17] Noda, J., Okamoto, K., Sasaki, Y., *J. Lightwave Technol.* **LT-4** (1986) 1071.
[18] Laming, R. I., Payne, D. N., *J. Lightwave Technol.* **7** (1989) 2084-2095.
[19] Clarke, I. G., Bassett, I. M., Poole, S. B., in: *Proc. ACOFT 1992, Hobart/Tasmania (Australia).*
[20] Morey, W. W., Dunphy, J. R., Multz, G., *Fiber Integrated Opt.* **10** (1991) 351-360.
[21] Hill, K. O., Malo, B., Bilodeau, F., Johnson, D. C., *Annu. Rev. Mater. Sci.* **23** (1993) 125-157.
[22] Xu, M. G., Reekie, L., Chow, Y. T., Dakin, J. P., in: *Proc. 9th Optical Fiber Sensors Conference, Florence, Italy;* 1993, pp. 27-30.
[23] Gal, G., in: *Diffractive and Miniaturized Optics,* Lee, S. H. (ed.), Vol. CR 49; SPIE, 1993, pp. 329-359.
[24] Göttert, J., Mohr, J., Müller, C., in: *Teubner Taschenbücher der Physik,* Vol. 27; Stuttgart, 1993.
[24a] Forrest, G. T., *Laser Focus World* (April 1992) 153-154.
[25] Lenz, J., Bjork, P., *Proc. SPIE* **961** (1988) 8-25.
[26] Spooncer, R. C., Jones, B. E., Philp, G. S., *J. Inst. Electron. Radio Eng.* **58**, No. 5 (Supplement) (1988) S85-S91.
[27] Gross, W., in: *Proc. Sensor Systems,* Vol. 2; London, 1989.
[28] Donati, S., Annovazzi-Lodi, V., Tambosso, T., *IEE Proc.* **135** (1988) 372-382.
[29] Tang, D., Rose, A. H., Day, G. W., Etzel, S. M., *J. Lightwave Technol* **9** (1991) 1031-1037.
[30] Bosselmann, T., in: *Proc. 9th Optical Fiber Sensors Conference, Florence, Italy;* 1993, pp. 297-301.
[31] Menke, P., Rosenbaum, D., Bartelt, H., Bosselmann, T., Süsse, R., Schulz-Dubois, E. O., in: *Proc. Opto 1994, Leipzig, Germany;* 1994, pp. 131-138.
[33] Müller, C., Mohr, J., Reichert, J., in: *Proc. Opto 94, Leipzig, Germany;* 1994, pp. 55-64.
[34] Müller, C., Mohr, J., *Interdiscip. Sci. Rev.* **18** (1993) 273-279.
[35] Uttamachandani, D., Thornton, K. E. B., Culshaw, B., *Electron. Lett.* **24** (1988) 573-579.
[36] Fatah, R. M. A., *Sensors Actuators A* **33** (1992) 229-236.
[37] Elwenspoek, M., *Journal A (Belgium)* **32** (1991) 15-22.
[38] Unzeitig, H., Bartelt, H., *Electron. Lett.* **28** (1992) 400-401.
[39] Bartelt, H., Unzeitig, H., *Sensors Actuators A* **37-38** (1993) 167-170.
[40] Bauerschmidt, P., Bartelt, H., Krämmer, P., Lerch, R., *ITG-Fachbericht 126;* Berlin: VDE-Verlag, 1994, pp. 259-264.
[41] Rogers, A. J., *SPIE* **1511** (1991) 2-24.

[42] Cowle, G. J., Dakin, J. P., Morkel, P. R., Newson, T. P., Pannel, C. N., Payne, D. N., Townsend, J. E., *SPIE* **1586** (1991) 130-145.
[43] Dakin, J. P., in: *Proc. Advanced Materials for Lightweight Structures,* ESA SP-336, Noordwijk: ESTEC, 1992, pp. 385-392.
[44] Culshaw, B., *SPIE* **1511** (1991) 168-178.
[45] Niewisch, J., Bartelt, H., *Sensors Actuators A* **41-42** (1994) 562-566.
[46] Jones, B. E., McKenzie, J. S., *Sensors Actuators A* **37-38** (1993) 202-207.
[47] Mizoguchi, H., Ando, M., Mizuno, T., in: Proc. Micro Electro Mechanical Systems, Travemünde, Germany; 1992, pp. 31-36.
[48] Zdeblick, M. J., *Sensors* (February 1993) 26-33.
[49] Hockaday, B. D., Waters, J. P., *Appl. Opt.* **29** (1990) 4629-4632.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

9 Materials in Nanotechnology

HELMUT SCHMIDT, Institut für Neue Materialien, Saarbrücken, Germany

Contents

9.1 Introduction 276

9.2 Sol-Gel Processing for Nanostructured Materials and Films 278
9.2.1 Materials Synthesis 278

9.3 Formation of Films and Bulk Materials 283

9.4 Potential for Sensors and Conclusion 291

9.5 References 292

9.1 Introduction

Nanostructured materials have gained increasing importance in research during the last decade. The reason for this interest is based on a variety of very special properties related to the small size of the phase dimension and, as a consequence, to a relatively high volume fraction of grain boundary phases. The basic phenomena, such as high diffusion coefficients, structure and other physical properties of these grain boundary phases, were described by Gleiter [1]. Interfaces of almost any kind of material possess structures different from those of the bulk. This is well known, for example, for polymer structures in adhesives, surface structures of metals, and others. If appropriate combinations of materials are chosen, interfacial structures can be used for specific alignments such as in Langmuir-Blodgett films or in micelle-type structures. Special properties can be expected in all cases where interfacial phases gain a significant volume fraction.

Another important factor is the large surface-to-volume ratio related to nanoscale particles. This leads to an enhancement of all types of reactions related to the surface of a system, such as catalysis. In the field of catalysis, high surface areas are required in order to obtain high overall reaction rates. However, in general, this type of material is based on porous carriers doped with active components such as noble metals. While these noble metal particles have diameters in the lower nanometer range, the processing of these materials can be considered as a true nanotechnology.

A basic process for synthesizing systems containing small disperse particles is doping a porous substrate with precursors which, through a chemical reaction (crystallization, nucleation, reduction, growth), form small particles on the support, in general, by surface diffusion processes on the carrier surface. However, these techniques do not lead to nanostructured materials in the sense that a porous or dense microstructure consisting of crystallites or amorphous particles with diameters in the lower nanometer range with a substantial volume fraction of an interfacial phase exists.

There is a series of problems related to nanotechnologies, starting from the fabrication of nanoscale particles, the control of agglomeration, processing them to green bodies, layers or coatings, and densifying them to bulk materials or dense films. Different techniques have been investigated in order to fabricate nanoscale materials. Chemical vapor condensation is widely used for metals in fundamental research [1], but has already gained significance for laboratory-scale preparation. Plasma-enhanced technologies have been developed by Trusov et al. [2], and powders are already being produced on a pilot scale. This is also valid for a chemical vapor reaction (CVR) process [3], where titanium nitride has been prepared very successfully. Other processes are high energetic milling and precipitation reactions. The latter, which involve wet chemistry, basically follow simple precipitation processes, which take place in supersaturated solutions and which follow the rules of nucleation and growth. In these systems, the critical nucleation radius has to be exceeded as a prerequisite for further growth. The growth reaction then continues following the thermodynamic drive since the precipitated system shows a lower free energy value than the dissolved system. This means, however, that, in general, the growth reaction only stops if the stock of precipitating ions is exhausted, which means if their concentration falls below the solubility product. In these cases the particle size of the precipitate can be tailored by adjusting concentration, solubility, and temperature if nucleation can be controlled. However, in cases of extremely low solubilities, as is the case with

the majority of interesting oxides (oxides for structural and functional ceramics) or metals, this route leads to very diluted systems in which it is difficult to separate powders from the solution and difficult to handle the process for industrial technologies, as shown by Matijevic [4]. In concentrated systems, however, agglomeration is unavoidable, leading to hard agglomerates.

The situation requires methods to control growth reaction, to stop it at a desired level, and to avoid agglomeration in order to keep the particles isolated for suitable further processing. In other words, one has to carry out precipitation reactions in such a way that stable colloids are produced in the first step in a sufficient concentration.

Stabilization of colloids can be achieved in various ways. The "simplest" way is to charge them electrically by the selective absorption of ions (for example, protons or hydroxyl ions), generating a surface potential and, as a consequence, repulsive forces between the equally charged particles. The theoretical background of electrically stabilized colloids has been treated extensively by Stern [5], who showed that actually an electric double layer is formed, preventing the colloids from agglomeration. Owing to the range of the electric forces (which can reach from several to 20 nm), the solid content of colloidal solution, in general, is low (from several up to 20% by volume for special cases). This type of low solid content colloidal solution (sols) can be transformed into gels by destabilization, for example by adjusting the pH to a range where the surface charges are neutralized (point of zero charge). In this case the van der Waals forces are strong enough to agglomerate the systems to fairly stable solids (gels), which, in general, show a disordered structure with respect to the particle arrangement which is responsible for the low solid content of these gels (<20% by volume). Processing of large parts is difficult owing to the large shrinkage concomitant with densification. If the sols are used for coatings, densification can be achieved fairly easily if the layer thickness does not exceed the critical size. This phenomenon has been investigated by Lange [6], who determined the critical thickness t_c obtained for crack-free coatings after firing according to Equation (9-1):

$$t \geq t_c = \frac{E \cdot G_c}{A \cdot \sigma^2} \tag{9-1}$$

where G_c = energy required to form cracks at the surface, t = thickness, A = constant, E = strain energy per unit volume and σ = triaxial tensile stress. This relation is valid for particulate systems with high particle–particle interaction, as it is the case in most oxidic systems.

Based on these considerations, two basic requirements for nanoprocessing of small particles have to be formulated. First, it is necessary to control the nucleation and growth process in such a way that the process can be stopped at a desired particle size and the particles can be stabilized at this level, and second, it is necessary to control the particle–particle interaction during "gelation" (or green part, layer or film formation) in such a way that high package densities can be obtained (for example, over 50% by volume for ceramic nanoparticle processing). A third option would be to treat the nanoscale particle for obtaining desired solid-state properties (for example, crystallinity, defect structure, or phases in mulitcomponent systems) in order to obtain desired properties in the solid material (for example, porous systems with the desired crystallinity after sintering while maintaining the nanoparticulate structure, eg, for sensor application).

In the following section, the advantages and disadvantages of the sol–gel process for obtaining nanocrystalline systems will be discussed.

9.2 Sol–Gel Processing for Nanostructured Materials and Films

9.2.1 Materials Synthesis

The synthesis of nanoparticulate systems is of great importance since it is the basis for processing these systems to solid materials such as films or bulks. In addition to the properties associated with the high surface-to-volume ratio, a variety of other properties of small particles and their processing to solid materials are of interest. These are quantum size effects such as quantum confinement and properties related to it such as high third-order non-linear susceptibilities if semiconductor or metal particles are considered, especially in combination with transparent matrices for embedding these particles. In these cases, so-called "nanocomposites" with high optical transparency become possible, unifying material properties related to the solid state of the particles and processing properties of, for example, a polymeric matrix. For this reason, the particle size has to be kept below the range at which Rayleigh scattering becomes disturbing, which is, in general, in the range of several nanometers.

Another interesting aspect is related to the surface defect structure of small particles, which, as already pointed out, is fairly high with respect to its volume fraction. These surface defect structures should lead to a considerable decrease in the sintering temperature if processed to ceramics. Independent of the type of mechanism to be assumed, as shown in Equations (9-2) and (9-3), the particle diameter affects the sintering rates considerably. Depending on a Nabarro–Herring or a Coble mechanism [7, 8], the particle diameter is efficient in a second- or third-order way to increase the sintering rate with decreasing size.

$$S_V = \frac{a \cdot \delta \cdot D_{GB} \cdot V_V \cdot \Delta\sigma}{d^3 \cdot k \cdot T} \tag{9-2}$$

$$S_{GB} = \frac{b \cdot D_V \cdot V_V \cdot \Delta\sigma}{d^2 \cdot k \cdot T} \tag{9-3}$$

where S = sintering rate (subscripts V and GB = volume and grain boundary diffusion), a and b = constants, T = temperature, D_V = volume diffusion constant, D_{GB} = grain boundary diffusion constant, δ = thickness of grain boundary, k = Boltzmann constant, $\Delta\sigma$ = pressure difference, V_V = defect volume (voids) and d = grain diameter.

Another interesting point is the use of nanosized particles to coat sub-micrometer particles and to build up two-component systems with engineered grain boundaries, for example zinc oxide-based varistor ceramics. For this reason, it is necessary to precipitate colloids, for example, to establish a zeta potential on the colloids opposite to that of the particles to be coated and to precipitate them on the surface. This could be another interesting way of generating very special surface properties for catalysts or sensors.

For the synthesis of nanoscale particulate systems by wet chemistry, it is essential to control growth and agglomeration, as mentioned above. In general, a precipitation reaction starts from homogeneous solution by a nucleation process. If the critical radius for nucleation is exceeded, the particle starts to grow, driven by thermodynamics. A simple thermodynamic consideration, as shown in Figure 9-1, shows that after nucleation particle growth is favored from thermodynamics. The growth reaction can be stopped by adding surface-active com-

ponents, leading to a metastable thermodynamic sink. In this case, the binding free energy for the surface-active components has to be lower than the free energy for further growth. This basis concept allows one to tailor particle size and particle distribution in a straightforward way, since chemistry provides several means of binding a variety of molecules to surfaces. For oxides, carboxylic acids, or β-diketones, amines or chelating molecules can be chosen, and for semiconductor or metals, electron-donating or -accepting groupings. Additionally, the concept allows one to control growth, to stabilize colloidal particles, and to "fit" them with special functions at the same time. If the size of the surface-modifying molecules is small, high packing densities can be expected in the case of compaction as long as these molecules do not interfere with each other too strongly. A schematic diagram of this concept is shown in Figure 9-2.

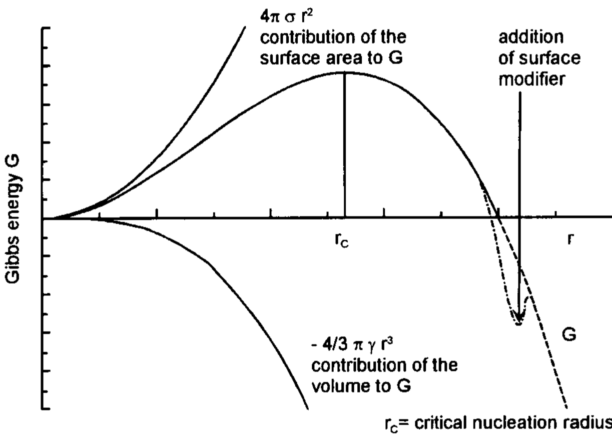
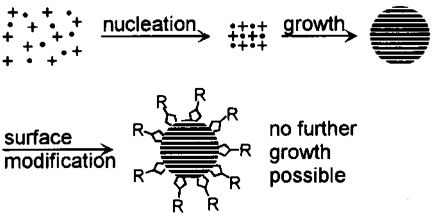


Figure 9-1.
Development of the free energy of the precipitating system after nucleation.



R: a: inert groupings like alkyl or aryl
b: fuctional groupings like acids or
c: reactive groupings like
or "sol-gel" groupings, e.g. $Si(OR)_3$

Figure 9-2.
Schematic diagram of the preparation of surface-stabilized colloids.

A prerequisite for a successful application of this concept is a controlled surface-to-particle bond. As shown elsewhere [9], carboxylic acids, for example, are bonded to boehmite particles very strongly in a mainly ionic form, as is typical for aluminium salts from organic acids ("colloidal salts"). If the binding strength of the surface-modifying compound is sufficient, the growth reaction can continue and the surface compound "moves" to the surface of the growing particle without leaving it. In this case, the surface-modifying component can be already present from the beginning of the reaction, as is shown schematically in Figure 9-3.

growth
continues

surface modifying grouping, e. g. ionic tensides or carboxylic acids

Figure 9-3.
Principles of a growth reaction in the presence of surface-modifying molecules.

In this type of reaction, the surface modifier acts as a "coating", only preventing agglomeration of nanoscale particles. It should be mentioned that stabilization of nanoscale particles can also be carried out by electric charges, of course, for example, by using appropriate pH values. However, this type of stabilization is very sensitive to the thermodynamic parameters of the system, especially concentration: as a rule, if the concentration exceeds a certain level, the electric shielding breaks down and leads to agglomeration. For example, as shown elsewhere [10], electrically stabilized boehmite sols cannot be concentrated to concentrations of more than 15–20% by volume without gelation. The gels are rigid and show no strength. If the electric stabilization is substituted by a carboxylic acid bonded to the surface (for example propionic acid), the solid content can be increased up to 45% by volume without gelation. These "sols", of course, show increased viscosities, but are still plastically deformable, in contrast to the other gels.

Precipitations also can be carried out in so-called "mini-reactors", ie, within small droplets with already nanosized diameters. For this purpose, thermodynamically stable microemulsions have to be prepared. For the stabilization of a second liquid phase within a liquid system both liquids have to be immiscible and the interfacial free energy has to be reduced in a way that the ΔG value of the system is lower than it would be in the separated state. This can also be carried out with an emulsifier, as is well known [11–14]. These microemulsions can be prepared from aqueous solutions of sols, for example, and oxidic particles can be precipitated within the droplets leading to a microsuspension. By appropriate choice of the emulsifiers, the basic structure of the system can be maintained, and the particles are protected from agglomeration by the adsorbed emulsifiers. One of the major advantages of these systems is the easy way to prepare multicomponent systems even with very different concentrations of the single components. An alternative to this route is to add the emulsifier to the homogenous solution of sols and to precipitate by pH change [15, 16]. In this case, an aqueous dispersion of emulsifiers in which a micelle type of structure exists is added to the reacting system. In Figure 9-4, the basic principle of both these routes is given with yttria-doped zirconia as an example.

The suspensions obtained have to be heat treated up to 400 °C to obtain the crystalline phases. Owing to the coating of the particles, only a weak agglomeration is observed, which is responsible for full redispersability with, for example, amines as dispersants. In Figure 9-5 a high-resolution transmission electron micrograph of the ZrO_2 nanoparticles treated at 400 °C and subsequently redispersed is shown. As can clearly be seen, the particle size ranges from 5 to 10 nm, and the particles are well crystallized.

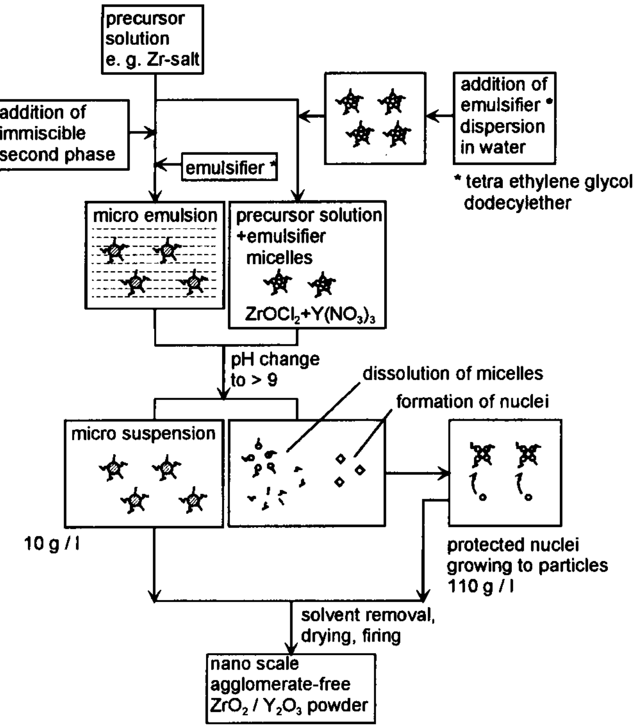


Figure 9-4.
Synthesis of Y_2O_3-doped ZrO_2 by microemulsion and by controlled growth routes.

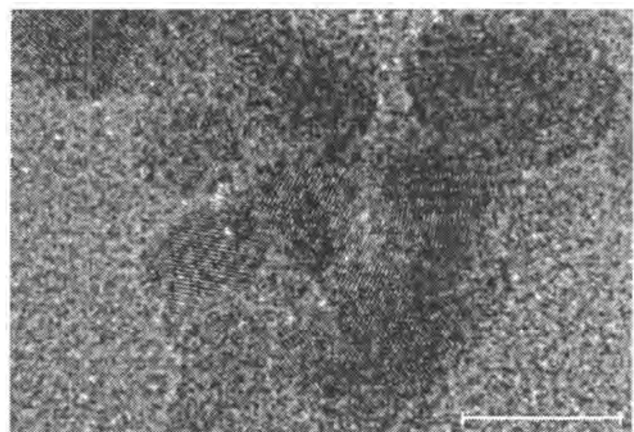


Figure 9-5.
High-resolution transmission electron micrograph of nanocrystalline ZrO_2 obtained by microemulsion or controlled growth techniques. The bar indicates 10 nm.

Another example is the synthesis of zirconia colloids from alkoxides in the presence of methacrylic acid. In such systems particle diameters between 5 and 30 nm can be obtained, depending on the surface-modifier concentration. Methacrylic acid also can be used for surface modification; this type has been used for the polymerization of particles into organic polymers, and in these systems crystalline (monoclinic) zirconia particles can be obtained at 120 °C [17].

These principles can be extended to a variety of systems such as cadmium sulfide [18, 19], zinc oxide, titanium oxide, silver sulfide, and silver halides, and metal colloids can also be prepared. In the case of metals, a reduction step has to be included, and surface-modifying ligands have to be chosen which remain bonded to the metal particle surface after reaction. The principles of these reactions are shown in Figure 9-6. As can be seen, there is a variety of different routes, which all have been successfully used for metal colloid-containing glasses or polymer composites [20–23].

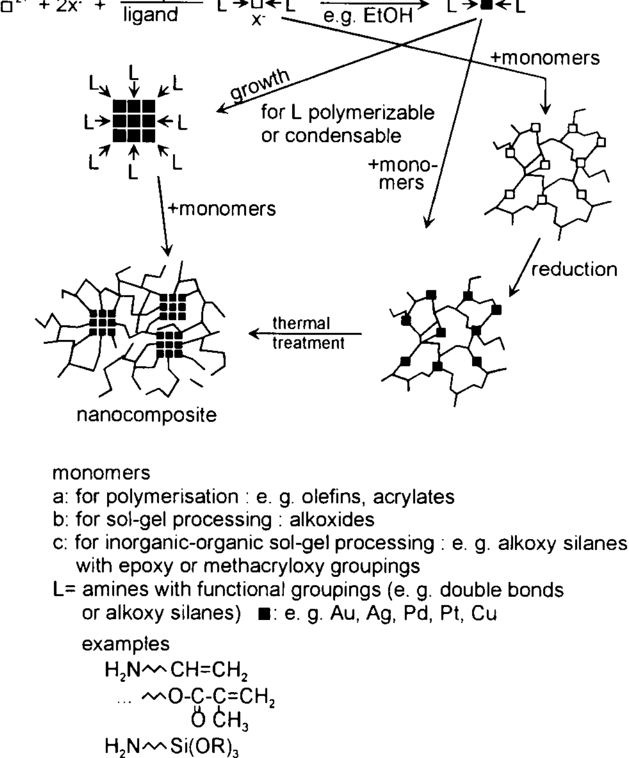


Figure 9-6. Principles of the reactions for the synthesis of metal colloids by wet chemistry, including the reaction steps and colloid formation.

The results show that it is possible to synthesize metal colloids in desired sizes and to incorporate them into tailored matrices such as glasses and ceramics (porous and nonporous) and also into a more polymeric type of matrix. The interesting point is that these metals, in general, show quantum size effects with high third-order nonlinearities (ca 10^{-6} esu in the case of gold) and also catalytic properties, if they are accessible, for example, in the systems described by Mennig et al. [20]. Summarizing, one can say that the synthesis of small nanosized particles by wet chemistry offers high potential for tailored, cost-effective preparation techniques.

9.3 Formation of Films and Bulk Materials

In previous sections, the basic principles for the fabrication of nanoscale systems have been considered. To exploit these properties for various applications, it is necessary to process them into solidified materials such as films, coatings, or bulk materials. As is well known from sol-gel techniques, film formation can be achieved by dip coating of substrates into sols and drying and firing them to give materials of suitable density. However, as indicated by Equation (9-1), the film thickness that can be obtained is restricted to layers between several hundred nanometers and 1 μm owing to the limited relaxation of the heavily agglomerated systems. It is very difficult to control porosity and crystallinity at the same time because, based on the rules of thermodynamics, crystallization of sol particles at the same time reduces the porosity substantially. This process strongly depends on the type of system, so, for example, titania tends to crystallize very easily, whereas α-alumina is much more difficult to obtain by this process. For this reason, an alternative route, solving at least some of the described problems, would be to synthesize the nanoscale particles independently of the sol formation and to "arrange them" in the form of films with a controlled package density. For this reason, it is necessary to control the particle-particle interaction and to use the principle of surface modification to obtain the appropriate surface properties.

This can be shown with colloidal silica as an example, used in form of aqueous or alcoholic suspensions. While coatings prepared from these sols can only be obtained with thicknesses of about 0.5-10 μm by a one-step dip-coating process, the surface modification of aqueous suspensions of colloidal silica with alkyl- or aryl-substituted silanes leads to the formation of films of thicknesses up to 10 μm. These films can be densified to silica by a heat treatment to almost 100% density without crack formation. Before densification, the films can be embossed on optical gratings, which is useful for optical technologies, eg, for incouplers. The systems can be doped and coated with functional groups for sensor applications. In Figure 9-7 an example of the embossing of these layers is shown.

Figure 9-7 clearly demonstrates that the shape of the structure is maintained during the firing process, which is very unusual for glass-like systems. The surprising fact is that the densification to almost fully dense SiO_2 takes place at temperatures as low as 500 °C. This means that SiO_2 layers can be prepared on a variety of substrates at four different medium temperature ranges. Since T never approaches T_g, the shape of the grating is maintained precisely.

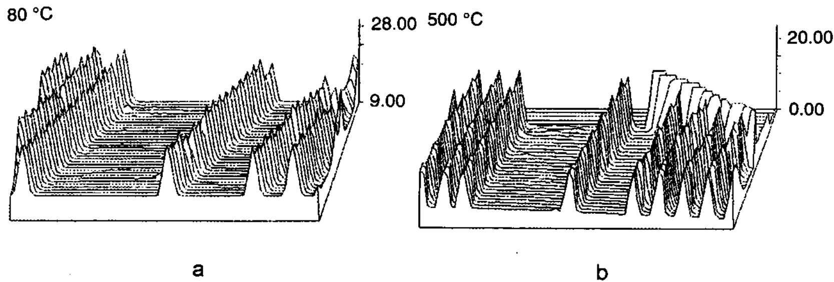


Figure 9-7. Comparison of an embossed SiO_2 film (a) before heat treatment and (b) after 500 °C heat treatment (after [24]).

The porosity of films derived from particulate sols can be tailored by the particle size. In Figure 9-8 the pore size distribution of a zirconia green body with a particles size of about 50 nm is shown [25]. The zirconia is a commercially available product from Degussa with an average particle diameter of 50 nm, and the pore size is in the range 7–8 nm. The pore sizes that can be obtained depend strongly on the particle shape and the particle size distribution. If spherical monodispersed particles are present, the residual pore sizes are in the range of about one third of the particle diameter.

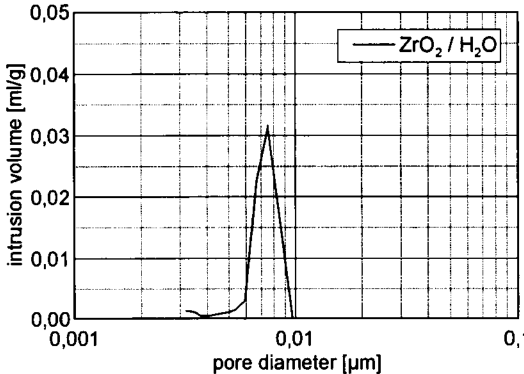


Figure 9-8.
Pore size distribution of an extruded zirconia green body.

A very important finding with nanosized systems is that during sintering, the particle size does not increase, in general, except with densely packed nanostructured green bodies. That means that tailored pore sizes can be obtained as a function of particle diameter and temperature treatment, and further means that high specific surface areas are possible. Another example is ZnO films, derived from a controlled precipitation process from zinc acetate [26]. In Figure 9-9 the pore diameter of zinc oxide densified and crystallized at 200 °C in form of a transparent film is shown. In this case, the particle diameter is 20 nm and the pore sizes are also in the range of several nanometers.

X-ray patterns of the zinc oxide show that even with these small particle sizes, crystallization is well developed in the correct (modified wurzite) crystal structure.

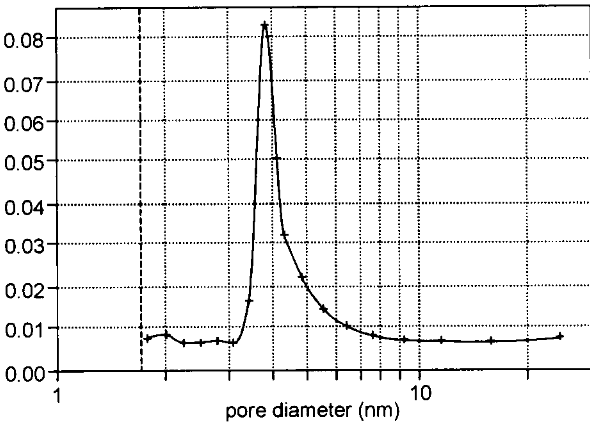


Figure 9-9.
Pore size distribution of ZnO film densified at 200 °C.

A prerequisite for all these types of nanoprocessing is the surface modification as described above. In the case of zirconia, complex formers such as β-diketones have been used and for zinc oxide organic bases such as amines.

In another example, α-alumina has been fabricated by a nanotechnology route [27, 28]. In this case, aluminium alkoxides were used and complexed with ethyl acetoacetate. The reaction is shown in Equation 9-4.

$$\text{RO, RO, Al, OR} \quad + \quad \text{O=C, H, C, O=C} \longrightarrow \text{RO, RO, Al, O–C, C—, O–C} \quad + \quad \text{HOR}$$

The particle size of these systems can be adjusted between 2 and 10 nm, depending on the type of β-diketone used. From these sols, films can be prepared by dip-coating, and these films are transformed into α-alumina above 950 °C without forming any other alumina crystalline phases such as γ- or θ-Al_2O_3. If the sols are seeded by α-nuclei, crystallization to α-alumina starts at 800 °C. These films are completely transparent and show a high porosity at these temperatures. From high-resolution scanning electron micrographs, the particle size can be estimated and is between 10 and 20 nm.

In Figure 9-10 the crystallization behavior in unseeded gels as a function of temperature and in Figure 9-11 the pore size distribution are shown. Atomic force microscopy, however, shows additional pores in the range of about 50 nm, as shown in Figure 9-12 (the formation mechanism is not clear). It can be clearly seen that crystallization starts at 950 °C without any signs of phases other than α-alumina. The pore size distribution is narrow and in the range of about 4 nm.

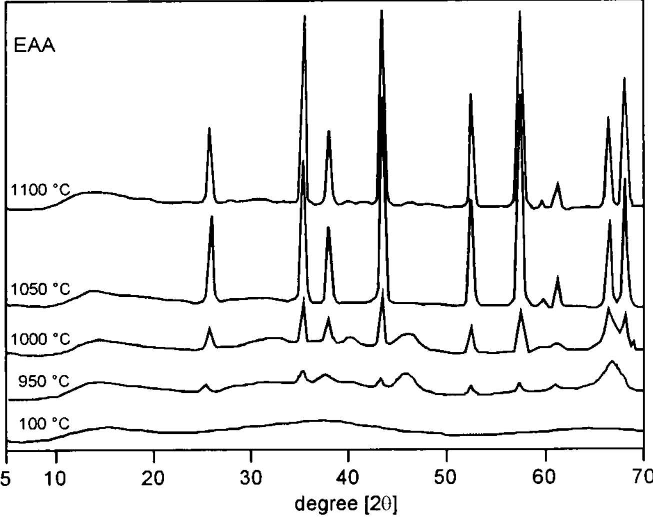


Figure 9-10. X-ray patterns of α-alumina formation from ethyl acetoacetate-modified alumina sols.

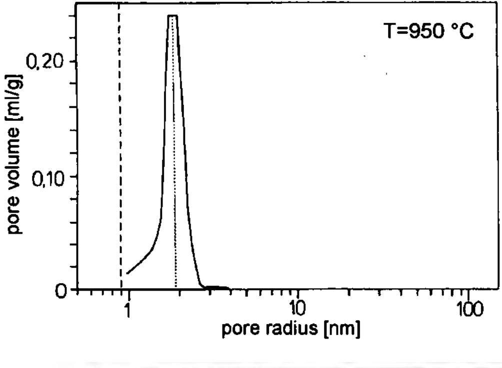


Figure 9-11.
Pore size distribution of nanocrystalline α-Al_2O_3.

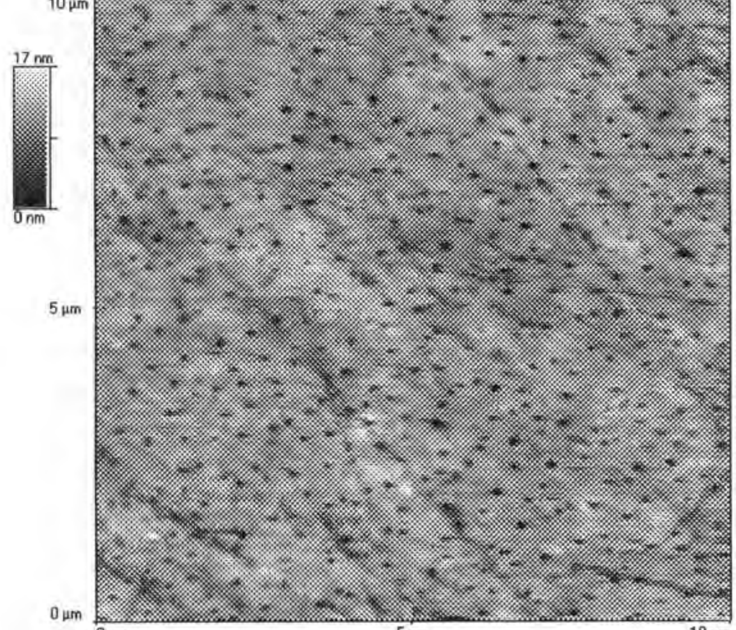


Figure 9-12. Atomic force micrograph of the α-Al_2O_3 surface (transparent layer) of a layer treated at 950 °C.

As in γ-alumina, in these α-alumina films electrical conductivity has been measured as a function of the partial pressure of water vapor in the atmosphere. Films deposited on fused-silica substrates have been electrically contacted and tested for their sensor behavior for testing the dependence on water vapor pressure [29]. The results are shown in Figure 9-13.

As can be seen, there are two regimes with linear dependences of the water vapor pressure with respect to the impedance of the film. The phenomena to which this behavior can be attributed is not clear. One explanation could be that nanopores present in these films may be "filled" with increasing p_{H_2O} in the high-sensitivity regime. The experiments show that with nanoscale systems sensitivites with α-Al_2O_3 are obtained otherwise only observed with

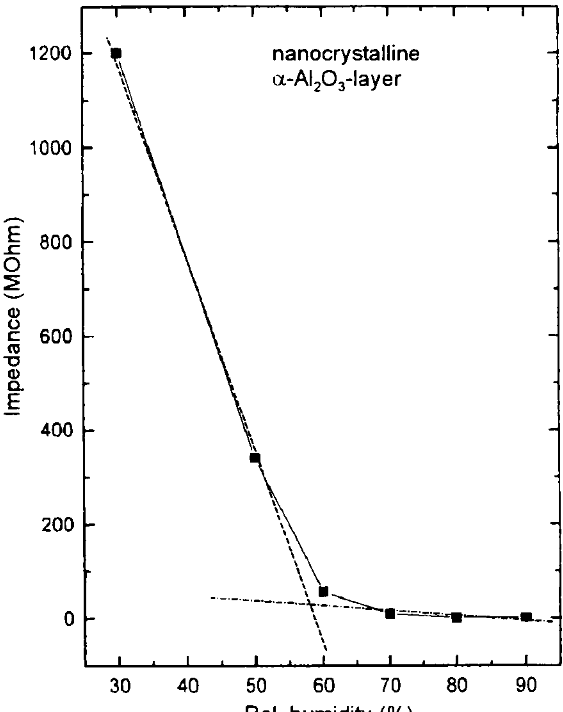


Figure 9-13.
Impedance versus relative humidity plot for a nanocrystalline α-Al_2O_3 layer (after [29]).

γ-Al_2O_3. It is expected that these α-alumina films will show high long-therm stability since α-alumina is the stable phase. The system did not show any changes in properties after several years of storage. The significance of these results is that without using the nanotechnological approach, α-Al_2O_3 is only obtained at temperatures above 1200 °C and, of even greater importance, with much larger particle sizes (sub-μm or μm range). This type of system, however, does not show sufficient sensitivity owing to the small surface area.

Similar experiments have been carried out with tungsten oxide [29], and it could also be shown that a high sensitivity from surface conductivity as a function of water vapor pressure is possible with the nanostructured oxide. High sensitivity to water vapor pressure is only obtained when WO_3 in nanoscale dimensions is present. The nanoscale phase is only present with heat treatment up to about 120 °C. In Figure 9-14 the surface of WO_3 treated at 60 °C is shown.

The primary particle size was determined from the X-ray line width according to Scherrer's equation to be about 5 nm. With increasing crystallite size as a function of temperature, the sensitivity decreases, as shown in Figure 9-15 for an Li^+-doped system. At 500 °C the sensitivity is close to zero. By doping with Li^+, the sensitivity regime can be changed by three orders of magnitude. The layers have thicknesses of about 2–400 nm and are completely transparent. The dependence of impedance on water vapor pressure is shown in Figure 9-16.

The synthesis and processing of nanoscale systems using colloidal techniques including surface tailoring is a generalizable concept developed first of all for novel ceramic processing and for the preparation of optical nanocomposites. For this reason, it was necessary to prove that

nanoscale slurries can be processed to dense sintered bodies, which was not at all clear. The difficulties in conventional sol-gel processing are considerable. One of the most interesting results was recently obtained by Nass et al. [30] with nanoscale TiN. In these experiments several significant results were obtained. First, it was possible to disperse 30 nm diameter TiN prepared by chemical reaction (CVR) [3] to its genuine particle size by using a chelating surface modifier (guanidine propionate) and to prepare green bodies up to 60% by volume green density. It is assumed that the diamine grouping acts as strong chelating ligands for surface

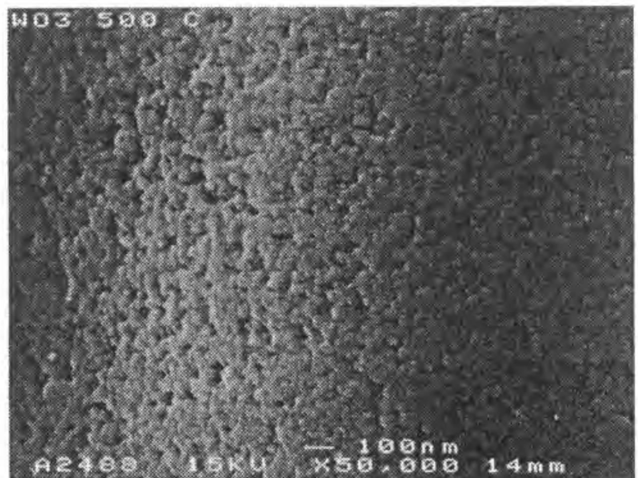


Figure 9-14. Surface of a transparent WO_3 layer treated at 60 °C.

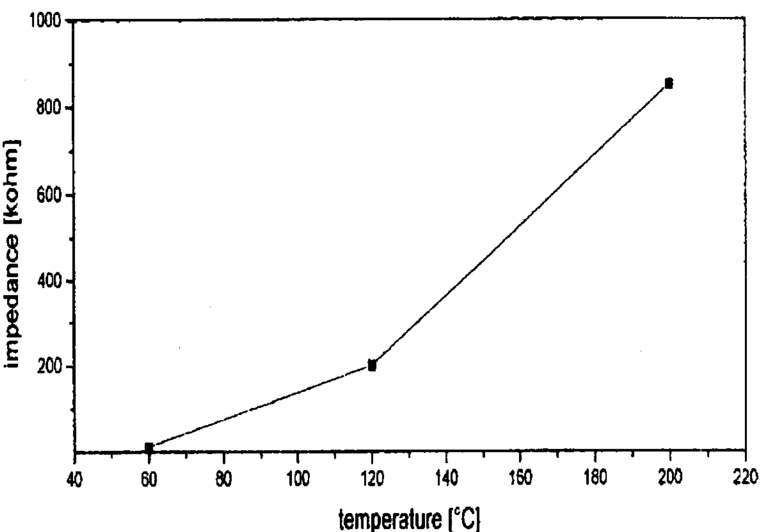


Figure 9-15. Decrease in sensitivity with temperature treatment (1 h at the indicated temperature) on a WO_3 layer on SiO_2 doped with 2.5 wt.% $LiClO_4$.

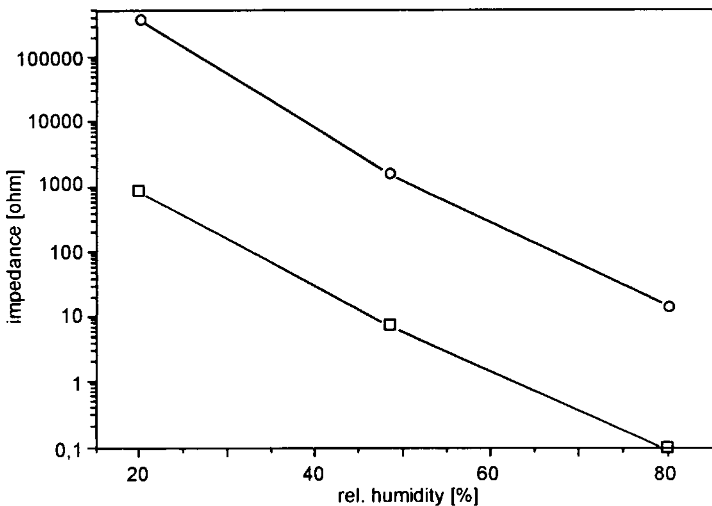


Figure 9-16. p_{H_2O} dependence of the sensitivity of (○) a doped and (□) an undoped WO_3 layer on SiO_2.

Ti^{4+} (oxidize layer). It is well known that conventional gels without surface modifiers do not show solid contents in this range (only up to 20% by volume) [31] with these high solid contents, and sintering temperatures of TiN films are as low as 1150 °C compared with >2000 °C with micrometer-scale technology. At this temperature no substantial grain growth is observed and the nanoscale structure is preserved. Further, it could be shown that TiN can be plastically deformed at 1150 °C by pressure. The fact that no grain growth takes place even at high temperatures shows that stable nanoscale phases can be obtained with an interesting potential for sensor application. Similar results with respect to lower sintering temperatures have been obtained with ZrO_2 [32].

The concepts and the results obtained on the basis of these concepts show that by use of the chemical approach nano-scale systems can be prepared in a variety of systems, as already shown for ceramics, semiconductors [33], and metals. For processing them to solid materials, films, or coatings, very thorough control of the particle-particle interaction is necessary. Therefore, the surface of the colloids or nanoscale particles has to be modified with preferably small molecules attached to the surface by selected chemical bonds. This is a prerequisite for processing these systems to composites or ceramics whilst maintaining their nanostructured state.

As already mentioned above, the major advantage of nanotechnologies for film formation is the fact that tailored microstructures with well crystallized phases and a high specific surface area can be obtained, leading to both high stability and high sensitivity to all types of surface reactions.

Another type of nanosized structures with an interesting potential for sensors is related to the fact that with small particles dispersed in transparent matrices, optically transparent materials can be prepared if the particle size is below the Rayleigh scattering regime. Rayleigh scattering (also depending on the differences in the refractive indices between the particle and the matrix and the dielectric properties of the interface) depends on the particle diameter to the power of 3. This concept leads to a variety of interesting perspectives for optical sensors,

because solid-state properties of the inorganic phases in inorganic-organic composites can be used, and at the same time processing techniques typical for organic polymers can be employed. A variety of different solid-state properties of nanoscale particles are also of interest. These are passive properties such as refractive index and Abbé number to be used for property tailoring in coatings for optical sensors; chemical resistivity of organic polymers, eg, swelling in the presence of organic solvents, can be achieved by inorganic particle reinforcement. Quantum size effects are of interest in the quantum confinement regime of nanoscale semiconductor particles as well as plasmon resonance vibration in metals, and there are first results indicating that the optical properties can be changed by the type of ligands interfering with the particle surface.

Silver halide nanoscale particles dispersed in polymeric matrices show a photochromic behavior that can be established reversibly and irreversibly [34].

If these matrices are polymer-like, microengineering techniques can be employed, as shown by Schmidt [16]. In Figure 9-17, some optical properties obtained from nanocomposite materials are shown.

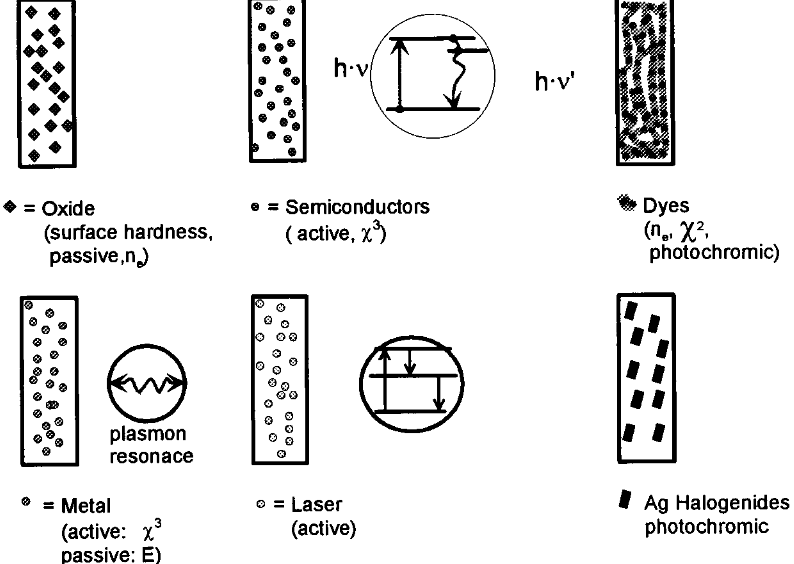


Figure 9-17. Some optical properties related to solid-state or electronic properties of small particles and/or molecules.

In combination with microengineering techniques such as embossing, photolithography, holographic techniques, or micropatterning by laser writing, components for optical sensors can be fabricated, and changes in structures of these particles should be applicable for sensors. Owing to their large interfaces, these systems should change optical properties by changing the interface in both the linear and nonlinear ranges. Figure 9-18 shows how these systems can be used for microengineering [35-37]. This is important for the fabrication of both optical components and sensing components.

Micropatterning techniques with inorganic–organic composite systems can also be used to fabricate pure inorganic systems after firing out the organic part.

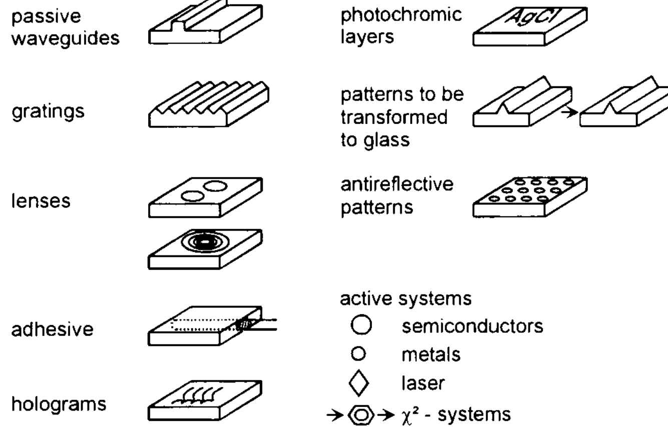


Figure 9-18. Possibilities for microengineering of different nanocomposite systems.

9.4 Potential for Sensors and Conclusion

The use of nanoscale systems has very interesting potential for sensor applications. First, as already shown with a few examples with nanostructured oxides, owing to the small particle size, thermodynamically stable systems can be used for sensor applications, which otherwise show a very low sensitivity. This basic principle can be extended to a variety of materials (for example, electroceramics) provided that nanoscale particulate systems are available. Availability is still the limiting step in the application of these systems. Processing to films or other devices may be difficult but can be solved by using the particle–particle interaction control approach. Chemical synthesis allows the generation of multicomponent systems, for example, using the microemulsion route, since the introduction of dopants is very convenient by using this technique. Another possibility to obtain nanoscale films and to establish dopants in the grain boundaries is to use infiltration techniques. The appropriate phases can then be obtained by an additional firing. Owing to the relatively low-temperature sintering techniques with nanoscale ceramics, a variety of systems otherwise very difficult to process owing to their high sintering temperatures should be accessible for building up sensor systems. So far, the potentional of these techniques has not been exploited sufficiently.

Another interesting system is the nanocomposite system, since it unifies optical transparency, solid-state properties, and properties related to the small particle sizes (eg, quantum confinement in semiconductors or plasmon resonances in metals) with large surface or interfacial areas. These properties provide interesting opportunities for various types of sensors. Summarizing, nanostructured materials open up a wide range for novel sensor applications owing to their large interfacial areas and the possibility of changing their optical or electronic properties with molecules contacting their surfaces.

In conclusion, one can say that the use of nanoparticles for sensors has a very interesting potiential, but is still in its infancy.

9.5 References

[1] Gleiter, H., *Nanocrystalline Materials;* Oxford: Pergamon Press, 1989.
[2] Trusov, L. I., Lapowok, V. N., Novikov, V. I., in: *Properties and Application of Disperse Powders;* Kiev, Naukova Dumka, 1986, p. 98.
[3] Winter, H., König, T. H. C. Starck Co., personal communication.
[3] Matijevic, E., in: *Ultrastructure Processing of Advanced Ceramics,* Mackenzie, J. D., Ulrich, D. R. (eds.); New York: Wiley, 1988, p. 429.
[5] Stern, O., *Z. Electrochem.,* (1924) 508.
[6] Lange, F., in: *Proc. International Symposium on Molecular Level Designing of Ceramics, Nagoya, March 1991;* Nagoya: NEDO International Joint Research Project, 1991, p. 14.
[7] Nabarro, F. R., in: *Report of Conference on Strength of Solids;* London: Physical Society, 1948, p. 75.
[8] Coble, R. L., *J. Appl. Phys.,* **34** (1963) 1679.
[9] Schmidt, H., Nass, R., Aslan, M., Schmitt, K.-P., Benthien, T., Albayrak, S., *J. Phys. (Paris), Colloq. C7,* **3** (1993) 1251.
[10] Benthien, K., *Master's Thesis,* University of Saarland, Saarbrücken, 1993.
[11] Osseo-Asare, K., Arriagada, F. J., in: *Ceramic Transactions, Vol. 12, Ceramic Powder Science III,* Messing, G. L., Hirano, S., Hausner, H. (eds.); Westerville, OH: American Ceramic Society, 1990, p. 3.
[12] Ramamurthi, S. D., Xu, Z., Payne, D. A., *J. Am. Ceram., Soc.,* **73** (1990) 2760.
[13] Nass, R., Burgard, D., Schmidt, H., in: *Proceedings of the 2nd European Conference on Sol-Gel Technology,* Nass, R., Schmidt, H., Vilminot, S. (eds.); Amsterdam: North-Holland, 1992, p. 243.
[14] Burgard, D., *Master's Thesis,* University of Saarland, Saarbrücken, 1992.
[15] Burgard, D., Kropf, C., Nass, R., Schmidt, H., in: *Proceedings of 1994 MRS Spring Meeting, Symposium "Better Ceramics Through Chemistry IV", April 1994, San Francisco; Mat. Res. Soc. Symp. Proc.,* in press.
[16] Schmidt, H., in: *Proceedings of NATO ARV on Applications of Organometallic Chemistry in the Preparation and Processing of Advanced Materials, Cap d'Agde, France, September 1994,* in press.
[17] Schmidt, H., Kasemann, R., Burkhart, T., Wagner, G., Arpac, E., Geiter, E., in: *ACS Symposium Series on Hybrid Organic-Inorganic Composites,* Mark, J. E., Bianconi, P. A., Lee, C. Y.-C. (eds.); Washington, DC: American Chemical Society, in press.
[18] Schmidt, H., *J. Sol-Gel Sci. Technol.,* **1** (1994) 217.
[19] Spanhel, L., Arpac, E., Schmidt, H., *J. Non-Cryst. Solids,* **147-148** (1992) 657.
[20] Mennig, M., Schmitt, M., Burkhart, T., Becker, U., Jung, G., Schmidt, H., in: *SPIE Proceedings 2288, "Sol-Gel Optics III, July 1994, San Diego,* Mackenzie, J. D. (ed.), (1994) 130.
[21] Mennig, M., Schmitt, M., Kutsch, B., Schmidt, H., in: *SPIE Proceedings 2288 "Sol-Gel Optics III", July 1994, San Diego,* Mackenzie, J. D. (ed.), (1994) 120.
[22] Burkhart, T., Mennig, M., Schmidt, H., Licciulli, A., in: *Proceedings of 1994 MRS Spring Meeting, Symposium "Better Ceramics Through Chemistry IV", April 1994, San Francisco; Mat. Res. Soc. Symp. Proc.,* in press.
[23] Mennig, M., Schmidt, H., Fink-Straube, C., *SPIE Proc.,* **1590** (1991) 152.
[24] Mennig, M., Krug, H., Fink-Straube, C., Oliveira, P. W., Schmidt, H., *SPIE Proc.* **1758** (1992) 387.
[25] Nass, R., personal communication, to be published.
[26] Nass, R., Fischer, R., personal communication, to be published.
[27] Nass, R., Schmidt, H., in: *Ceramic Powder Processing Science,* Hausner, H., Messing, G. L., Hirano, S. (eds.); Cologne: Deutsche Keramische Gesellschaft, Köln, 1989, p. 69.
[28] Schmidt, H., in: *Ceramic Transactions, Vol. 22: Ceramic Powder Science IV,* Hirano, S., Messing, G. L., Hausner, H. (eds.) Westerville, OH: American Ceramic Society, 1991, p. 3.
[29] Schmidt, H., in: *Extended Abstracts, Eurosensors VII, Budapest, September 1993,* Technical University of Budapest, p. 327.
[30] Nass, R., Albayrak, S., Aslan, M., Schmidt, H., in: *Proceedings of the International Conference "Ceramic Processing Science and Technology", September 11-14, 1994, Friedrichshafen/Bodensee, Germany,* in press.

[31] Brinker, C. J., Scherer, G. W., *Sol-Gel Science;* Boston: Academic Press, 1990.
[32] Schmidt, H., Nass, R., in: *Proc. Austceram '94,* Sorrell, C. C., Ruys, A. J. (eds.); *Int. Ceramic Monographs,* Vol. 1, No. 2, Sydney: Australasian Ceramic Society, 1994, p. 1065.
[33] Henglein, A., *Top. Curr. Chem.,* **143** (1988) 115.
[34] Fink-Straube, C., *PhD Thesis,* University of Saarland, Saarbrücken, 1994.
[35] Oliveira, P. W., Krug, H., Künstle, H., Schmidt, H., in: *SPIE Proceedings 2288, "Sol-Gel Optics III", July 1994, San Diego,* Mackenzie, J. D. (eds.), (1994) 554.
[36] Schmidt, H., Krug, H., Kasemann, R., Tiefensee, F., *SPIE Proc.,* **1590** (1991) 36.
[37] Krug, H., Zeitz, B., Oliveira, P. W., Schmidt, H., in: *Proceedings 8th CIMTEC Congress, June 1994, Florence, Italy,* in press.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

10 Sensors and "Smart" Molecular Nanostructures: Components for Future Information Technologies

WOLFGANG GÖPEL, Institute of Physical and Theoretical Chemistry and Center of Interface Analysis and Sensors, University of Tübingen, D-72076 Tübingen, Germany

Contents

10.1 Introduction 296

10.2 Scanning Probe Microscopy (SPM) to Identify and Manipulate Nanostructures 300
10.2.1 Identification of Chemical Species 300
10.2.2 Manipulation of Nanostructures 306

10.3 Chemical Sensing and Molecular Recognition 307
10.3.1 Transducers 307
10.3.2 Recognition by Matching Molecular Orbitals of Electronic Conductors . . . 309
10.3.3 Recognition by Matching the Doping of Mixed Conductors 312
10.3.4 Recognition by Matching Electrochemical Potentials of Ion Conductors . . 315
10.3.5 Recognition by Key-Lock Interactions with Static Molecular Cages 318
10.3.6 Recognition by Induced Fit with Dynamic Molecular Cages 320
10.3.7 Biosensors Based upon Key-Lock Interactions (Bioaffinity Sensors) 322
10.3.8 Biosensors Based upon Catalytic Reactions (Enzyme Sensors) 323
10.3.9 Biosensors Based upon Membrane Functions 323
10.3.10 Biosensors Based upon Whole Cell Functions 325

10.4 Sensor Arrays and Neural Networks 326

10.5 Molecular Electronics 326

10.6 Bioelectronics 331

10.7 Outlook 333
10.7.1 Synthesis and Patterning of New Materials 333
10.7.2 Compatibility with IC Technology 333
10.7.3 Systems 334

10.8 References 335

10.1 Introduction

A major driving force in the research and development (R&D) of new materials for future information technologies aims at the miniaturization of devices down to the ultimate limits as they are determined by basic physics and quantum mechanical principles (Figure 10-1). Another driving force results from trying to match in future devices the different performances that are currently achieved separately in biological and in technical systems. An often considered example concerns the *human brain* as compared with the *man-made computer.* Technologically unmatched performances of the brain concern high information density, low power consumption, high flexibility, association memory, etc. Biologically unmatched performances of the computer concern quantitative information processing, high reproducibility, etc.

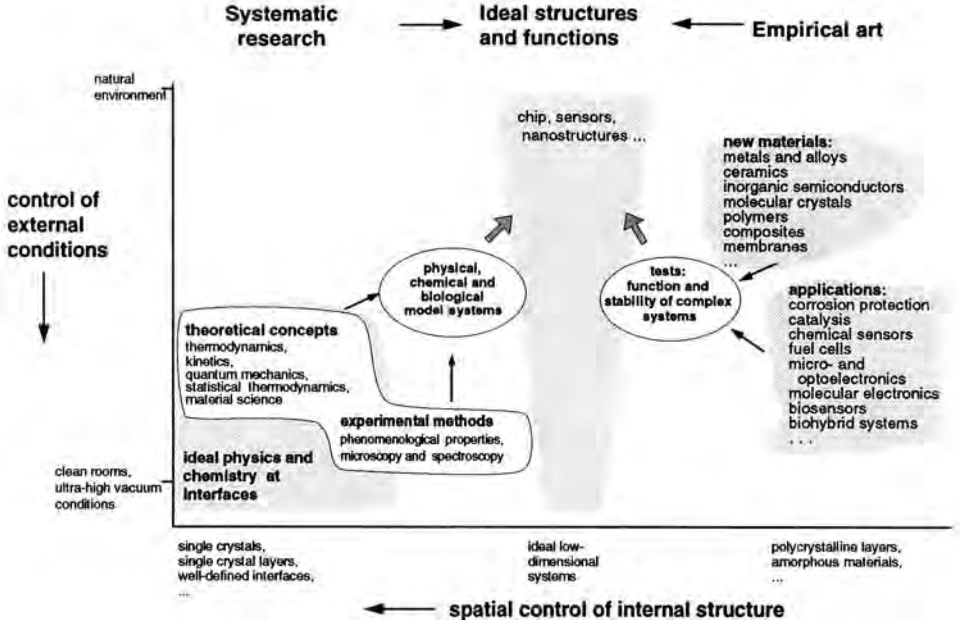


Figure 10-1. Schematic illustration of systematic research and empirical art in the design of new materials for sensors and smart molecular nanostructures [1].

With increasing complexity and demands for future information technologies, a trend is to be seen towards the design of "smart" nanostructures which will be interfaced to silicon or other substrates (Figure 10-2). These structures may consist either of *chemically synthesized* units such as molecules, supramolecules, and biologically active recognition centers, or of natural and hence very *complex biomolecular function units* with high molecular weight which may be extracted from biological systems [1-6].

Today's and future applications of such nanostructures include devices for chemical and biochemical sensing, electronically controlled drug release in the human body, molecular electronics, or bioelectronics. In this context, the field of bioelectronics is of increasing interest

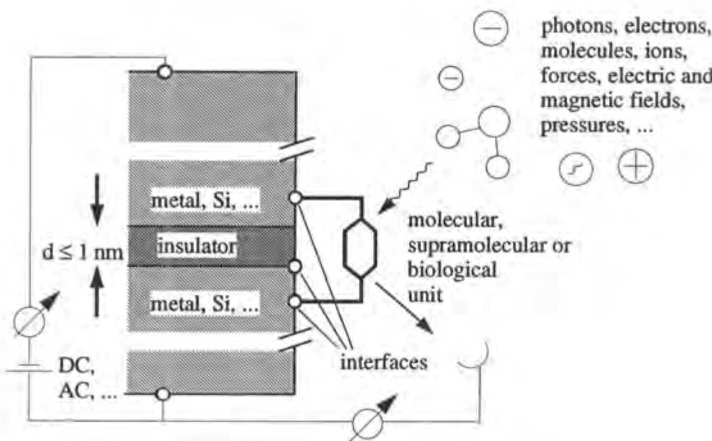


Figure 10-2. Schematic presentation of a (bio)molecular sensor device. Coupling between microelectronic structures ("top-down" approach, left) and chemically synthesized or biological structures ("bottom-up" approach, right). The molecular, supramolecular, or biological units may be modified in different environments as indicated in the upper right part of the figure. This effect may be utilized in future nanosensors by monitoring the corresponding response signals either directly in the environment (lower right side) or indirectly via signal transduction across the interfaces through the transducer device (lower left side). An example of the latter is to monitor changes in the direct DC or alternating AC current [1-3].

with its aim of transferring signals from biological function units to "conventional" electronic devices and vice versa.

In this development of existing and future information technologies, several different components have to be optimized. In a complete system these include components for perfect information sensing, information processing, information storage, and actuating. An example may illustrate the functional similarities and technological differences in the design of device components which may either be based upon artificial ("man-made") or natural ("biological) structures: Figure 10-3 shows schematically the odor analysis with components of the *human nose.* In contrast, Figure 10-4 shows schematically the chemical or odor analysis with components of an *electronic nose.* Both noses consist of completely different materials. Nevertheless, similar functions formally characterize both systems. Odorant molecules produce ionic or electronic signals through the action of individual receptors (or sensor/transducer units). The signals determine the components of a "feature vector" and are subsequently evaluated by means of pattern recognition either in the brain or in the computer. Current research in bioelectronics aims at interfacing both worlds with different interconnections and degrees of sophistication.

To illustrate the current different R&D strategies in the field of nanostructures, Figure 10-5 outlines different aspects of these approaches. As already indicated in Figure 10-2, nanolithographic patterning ("top-down approach") or molecular engineering ("bottom-up approach") offer complementary design concepts which are expected to lead to fascinating new structures, theoretical questions, experimental approaches, and applications.

"Smart" molecular materials, in particular, make it possible to tailor structures on the atomic scale with controlled electrical, dielectric, optical, magnetic, mechanical, chemical, or

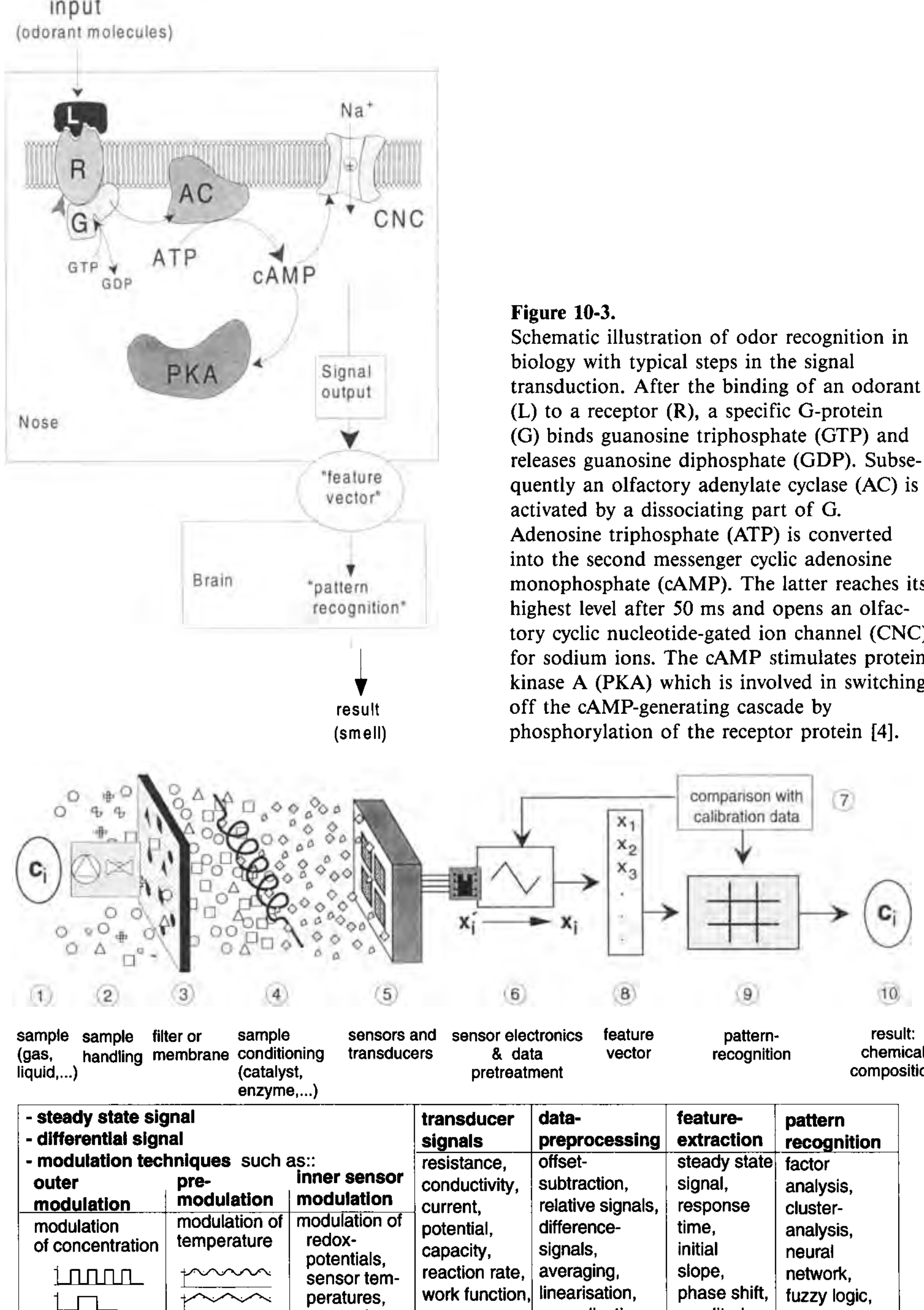


Figure 10-3.
Schematic illustration of odor recognition in biology with typical steps in the signal transduction. After the binding of an odorant (L) to a receptor (R), a specific G-protein (G) binds guanosine triphosphate (GTP) and releases guanosine diphosphate (GDP). Subsequently an olfactory adenylate cyclase (AC) is activated by a dissociating part of G. Adenosine triphosphate (ATP) is converted into the second messenger cyclic adenosine monophosphate (cAMP). The latter reaches its highest level after 50 ms and opens an olfactory cyclic nucleotide-gated ion channel (CNC) for sodium ions. The cAMP stimulates protein kinase A (PKA) which is involved in switching off the cAMP-generating cascade by phosphorylation of the receptor protein [4].

- **steady state signal** - **differential signal** - **modulation techniques** such as::			**transducer signals**	**data-preprocessing**	**feature-extraction**	**pattern recognition**
outer modulation	**pre-modulation**	**inner sensor modulation**				
modulation of concentration	modulation of temperature	modulation of redox-potentials, sensor temperatures, measuring voltages, . . .	resistance, conductivity, current, potential, capacity, reaction rate, work function, . . .	offset-subtraction, relative signals, difference-signals, averaging, linearisation, normalisation, . . .	steady state signal, response time, initial slope, phase shift, amplitude, . . .	factor analysis, cluster-analysis, neural network, fuzzy logic, . . .

Figure 10-4. Ten components of a complete technical sensor system [1].

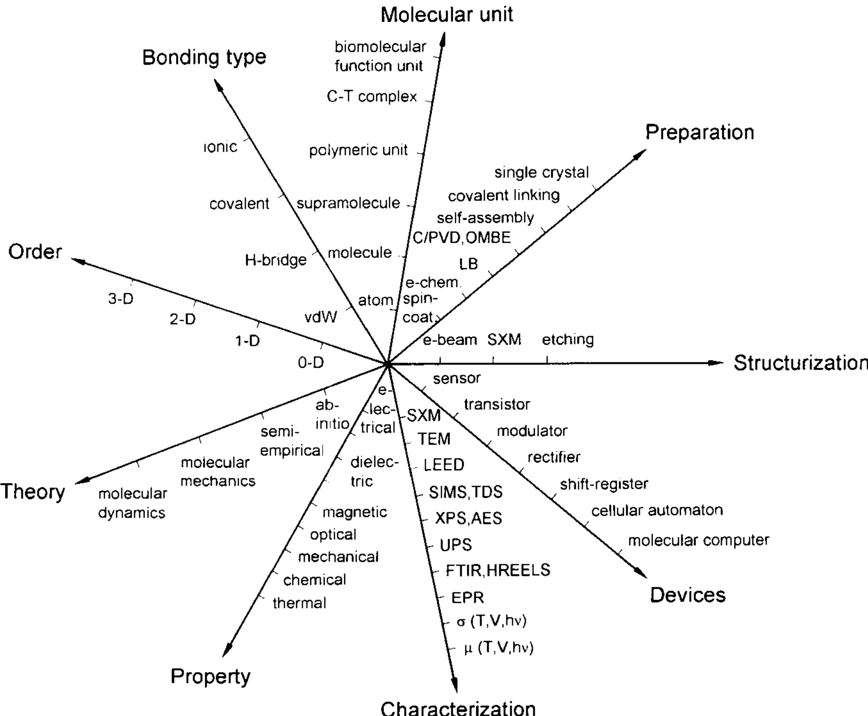


Figure 10-5. Different aspects of nanostructures based on molecular materials [7]. The different abbreviations describing the characterization techniques have been explained in Volume 2, Chapter 3, of this series [8]. For further details, see, eg, Ref. [9]. Compare also Table 10-3.

thermal properties for future potential applications in a variety of electronic, photonic, magnetic, mechanical, and/or molecular recognition devices [10]. The preparation technologies, the experimental analysis of their structures and properties, and the theoretical understanding of such molecular systems are improved rapidly in a truly interdisciplinary R&D effort. Structural control down to the nanometer scale becomes possible by using ultraclean materials and new preparation technologies and by applying analytical tools to investigate organic/organic and organic/inorganic interfaces even on the atomic scale. For analyzing and patterning these structures, scanning probe techniques play an increasing role.

As a result of this effort, basic principles and key components of existing and future information technologies in general and of chemical sensors, molecular electronics, and bioelectronics in particular are now being investigated and elaborated on a broader and more solid basis than it was possible in the past [10]. As an example, about 12 years ago mere enthusiasm to design molecular computers was the only, but nevertheless a strong, driving force of R&D efforts in this field with at that time a completely unrealistic judgement of the materials science problems involved [11]. Today, we do not (yet) have molecular computers available, but we do have a clear picture about possibilities and limitations and about alternative spin-offs that will emerge from R&D of new materials in this field.

The optimistic view envisioned today by experts in the field of nanostructured materials is basically a result of the rapid development of nanotechnologies in general, of intense materials

research effort, and of the important role which nanostructured materials already play in various practical applications. The latter include protective overlayers, composites, high-performance ceramics, controlled catalysts and polymers, macromolecular and supramolecular structures, membranes, inorganic structures for single electron devices, new materials for molecular micro- and optoelectronics, chemical and biochemical sensors, and biocompatible materials.

In the following, only the main trends in the R&D of sensors and "smart" molecular nanostructures can be characterized. The term *sensor* is used here in two different ways:

- first, in a general context (as in Chapter 11) to describe the experimental set-up of scanning probe techniques to identify and manipulate nanostructures (Section 10.2), and
- second, in a specific context to describe devices which monitor chemical and biochemical species and which make use of a smart nanostructure to convert chemical information into electronic information by means of a specific molecular recognition (Section 10.3).

A short outline will then be given of sensor arrays and neural networks (Section 10.4). Specific aspects of molecular electronics and bioelectronics will be discussed briefly (Sections 10.5 and 10.6), followed by a short outlook (Section 10.7).

10.2 Scanning Probe Microscopy (SPM) to Identify and Manipulate Nanostructures

In Chapter 11, the basic principles of scanning probe techniques are summarized and a variety of new proximal probe sensors are introduced to measure *physical quantities*. We here concentrate on those sensors which in principle allow the identification of *chemical compositions* and the controlled *manipulation of nanostructures*. They are mainly based on scanning tunnelling microscopy (STM), scanning electrochemical microscopy (SECM), scanning force microscopy (SFM), and scanning near-field optical microscopy (SNOM) [12–16].

10.2.1 Identification of Chemical Species

(a) In *vacuum*, STM is used to characterize surfaces before and after chemical treatment. For this purpose, a sharp metallic tip is kept at a distance less than 1 nm above a conductive sample. It serves as a sensor to monitor electrons which tunnel from the tip to the sample or vice versa. The tunnelling current depends exponentially on the distance between sample and tip (see Chapter 11, (Equation (11-1)). The simplest measurement mode in STM therefore uses a servo system to keep the tunnelling current and hence, to a first approximation, the distance at a constant pre-set value. This leads to topographic information. On the other hand, the decay rate and the density of "tunnelling channels" also influence the current flow. The decay rate is mainly determined by the tunnelling barrier height, which reflects the mean work function of the system. The density of "tunnelling channels" is increased by an increase in the number of unoccupied levels that are involved on the positively biased side and, to a minor

extent, the number of occupied energy levels that are involved in the negatively biased material. As a result, the scanning tunnelling image also contains chemical information which may be deduced from these occupied and empty electronic states.

This can be made use of in specific operation modes of the STM. The most important is the voltage-dependent measurement mode which allows an identification of the electronic structure of the system.

As a simple example, two STM images may be taken in the same area, the first with a positively and the other with a negatively biased sample. In the first image, electrons flow from the occupied levels of the sample to the unoccupied orbitals of the tip and thereby determine the contrast in the image. In the second image, the opposite situation is utilized. There are several examples known today of binary semiconductors in which the highest occupied valence band levels result exclusively from orbitals of one element (the anion) and the lowest the unoccupied conduction band levels from the other element (the cation). The most prominent material is GaAs, where the detection of As ions occurs at a negative and the detection of Ga at a positive bias (Figure 10-6).

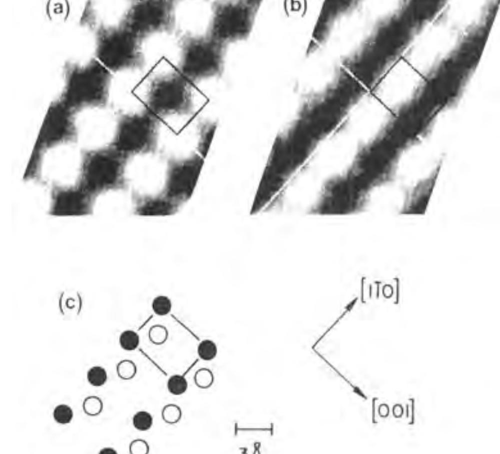


Figure 10-6.
Identification of anions (As) and cations (Ga) by STM imaging of GaAs with opposite bias: (a) $V = +1.9$ V; (b) $V = -1.9$ V; (c) top view of the surface atoms. As atoms are represented by open and Ga atoms by closed circles. The rectangle indicates the unit cell, whose position is the same in all three parts [17].

In more sophisticated experiments, the servo loop is disconnected at each measurement point after adjusting the constant height by a constant current at a given voltage. Subsequently the bias voltage is ramped and the resulting currrent is measured. For an ideal (s-type orbital) tip, the tunnelling conductance dI/dV reflects the density of states near the Fermi level of the sample. This may lead to atomically resolved density of states plots and the identification of adsorption processes on the atomic scale (Figure 10-7).

The third type of experiment uses the possibility of modulating the tip-sample distance s at a certain frequency. If the modulation is faster than the response of the servo loop, the DC image still reflects the topographic image. The signal dI/ds, on the other hand, may be identified as the tunnelling barrier heigt, which is mainly determined by the mean work function of the system. Although a quantitative analysis is still difficult, the change in the work function may be used to identify defects, adsorbates, or different crystal orientations, all of which do not show up in the topographic images or which may be mixed up with the topographic information.

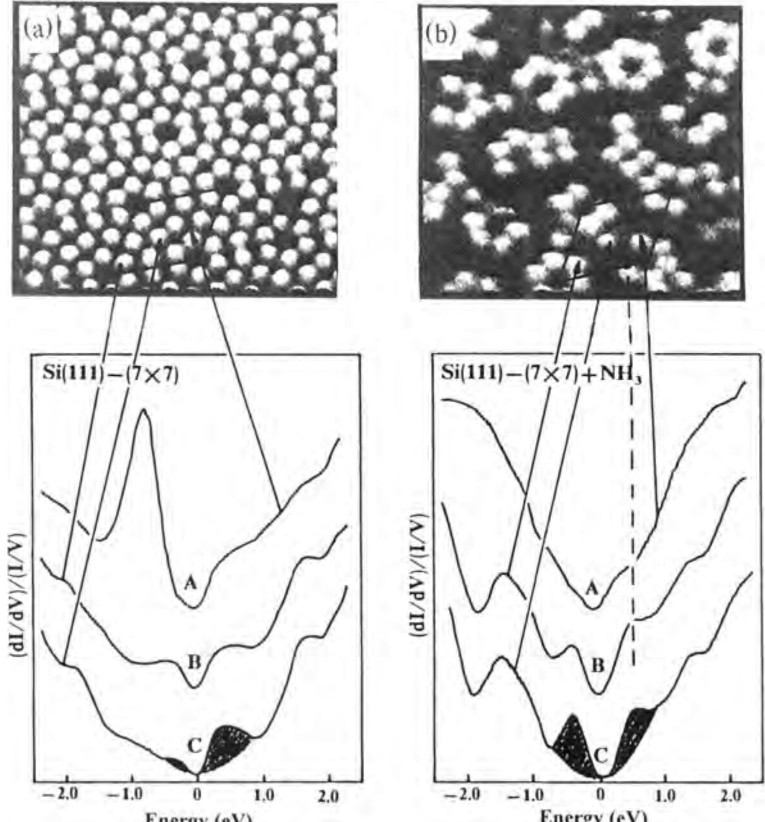


Figure 10-7. STS of clean (left) and NH_3-coated (right) Si(111) surfaces. Peaks in the differential changes in the current with voltage as a function of energy (in eV) correspond to maxima in the densities of states which were monitored at the different spots A, B and C [18].

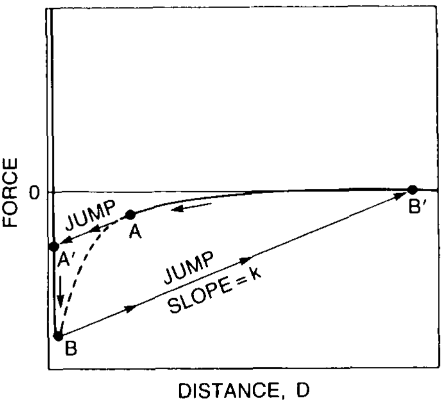
(b) In *electrolyte solutions,* STM may be used to detect electrochemical reactions on the nanometer scale.

In a first type of experiment, STM is used to monitor the change in morphology during electrochemical reactions, such as metal deposition from electrolyte solution or corrosion. For this purpose, the electrochemical treatment should be performed independently of the STM experiment. Therefore, the STM has to be operated with an electrochemically inert electrode and a bias voltage has to be applied which does not induce electrochemical reactions at the STM tip electrode.

In the second type of experiment, the STM tip serves as one of three electrodes, ie, as the counter electrode, in an electrochemical experiment (scanning electrochemical microscopy, SECM). Here, the STM tip is chosen as a small disk electrode at which electrochemical reactions may occur. The distance dependence of the current depends sensitively on the nature of the substrate: for an insulator, the current drops when the tip approaches the substrate, because the volume is reduced through which ions diffuse to the tip; for a metal, the current increases, because the reaction of the redox species between tip and substrate is then confined to a smaller volume.

(c) *Force sensors* used in SFM detect topographic features also at non-conducting samples. The distance dependence of the interaction between force sensor and sample may also be used to extract "chemical" information from the images.

As an example, Figure 10-8 shows the force–distance plot of an "ideal" van der Waals-interacting system and the "real" response of an SFM. The instability at position 3 results



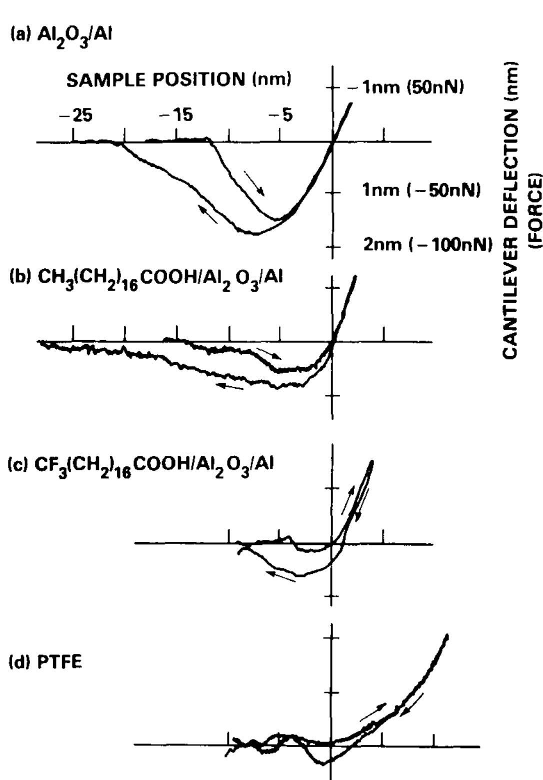


Figure 10-8.
Top: Force–distance plot of an ideal van der Waals interacting system (dashed line) and corresponding real responses monitored with an SFM. Bottom: Quantitative examples on force–distance results obtained with an SFM and characteristic hysteresis effects due to local surface tensions and intermolecular forces [19]. The surface tension decreases from (a) to (d).

from adhesion forces which retract the tip to the sample surface. These adhesion forces depend critically on the surface tension as shown in the lower part of Figure 10-8. Provided that detailed theories will be elaborated in the future, this information may be used to determine quantitatively the local surface tensions and intermolecular forces at the surfaces.

In more sopisticated experiments, the SFM tip is modified chemically by attaching molecules. This may be done by silanization reactions on SiO_2 or Si_3N_4 tips or by thiol coupling to Au-coated tips. Besides these most common strategies of providing a strong covalent coupling, other couplings of molecular recognition centers to other substrates may also be used. In any case the force-distance curve reflects directly the interaction and adhesion forces betweeen the immobilized molecules and the atoms at the sample surface. In the specific case of biomolecules, this principle may be used to monitor key-lock interactions on the molecular scale (Figure 10-9).

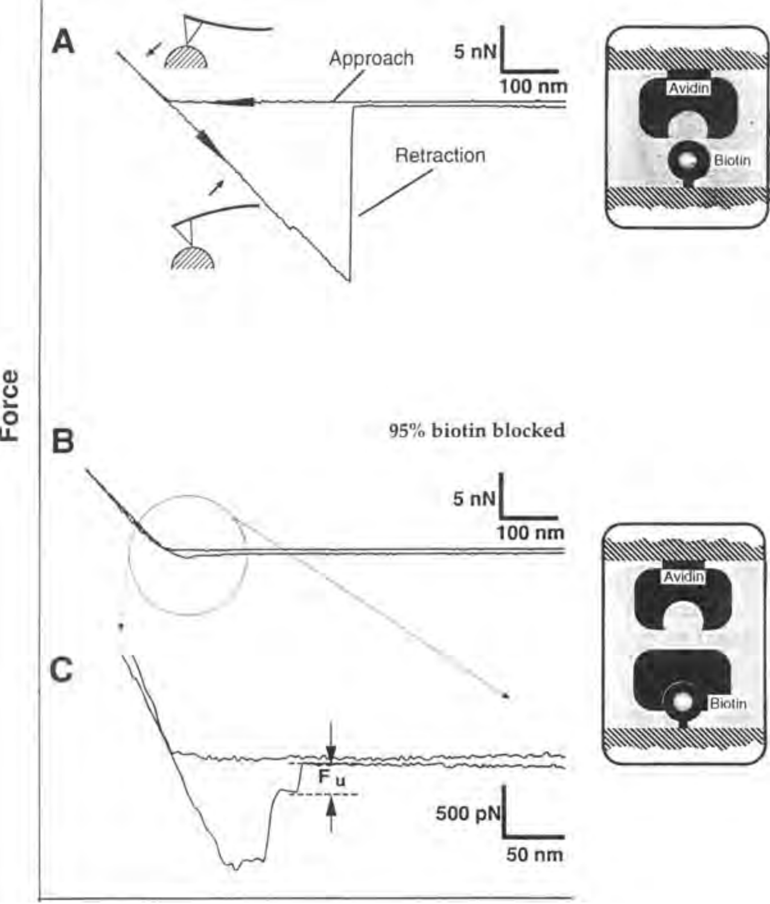


Figure 10-9. Typical force-distance curves between an avidin SFM tip and a biotin-pretreated surface. In reality several of such individual key-lock interactions are monitored and ruptured simultaneously. In recent experiments even the individual steps F_u in the force of breaking individual bonds could be monitored (lower part, note the changed force scale) [20].

(d) The *scanning near-field optical microscope* (SNOM) provides a fourth approach to extracting chemical information by using scanning probe microscopies. Although at the moment only a few experiments are known in which scanning near-field optical spectroscopy (SNOS) was applied successfully, this field will certainly develop rapidly in the near future. One major drawback today results from the intense laser light required, namely the lack of small and cheap tunable lasers and the lack of optical fibers which ideally one would like to use over a broad range of wavelengths. Therefore, monochromatic excitation in Raman and fluorescence experiments are used exclusively today. Another drawback is the relatively poor spatial resolution parallel to the surface, which is determined by the diameter of the optical fiber. The long-term goal is the development of a SNOM operating in the SNOS mode continuously between the IR and the UV/VIS region and utilizing the high spatial resolution of SPM combined with the high spectral resolution of fluorescence spectroscopy. Particularly in the IR region, the latter is a well known technique in analytical chemistry with no doubt about its extremely high selectivity to identify chemical species through their characteristic vibrational modes.

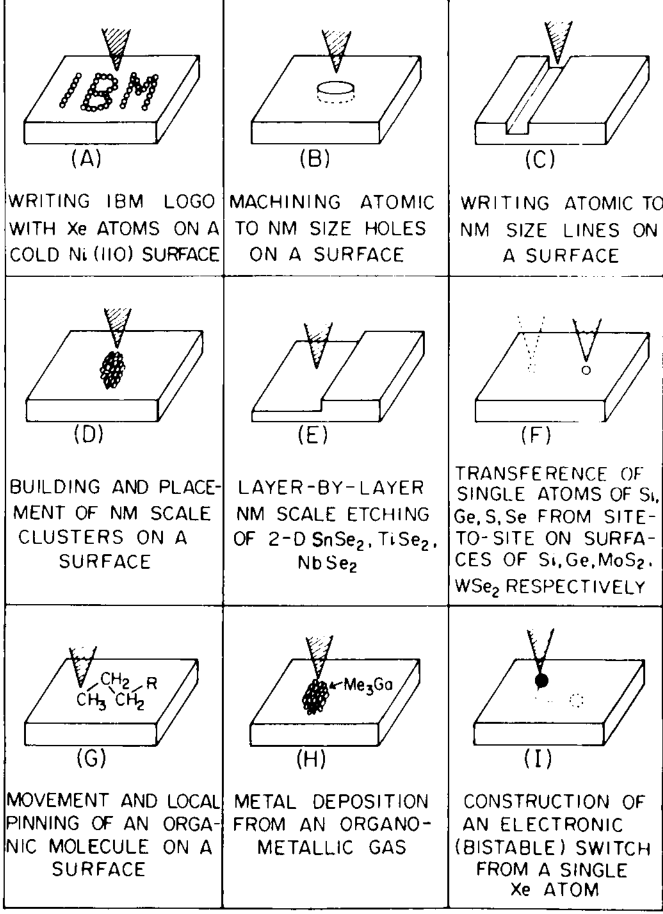


Figure 10-10. Typical manipulations to produce nanostructures with SPM technology [21].

10.2.2 Manipulation of Nanostructures

In an overview, Figure 10-10 summarizes a few of the possible manipulations which may be performed on the atomic scale. Table 10-1 lists typical forces between atoms and molecules if compared with forces arising in STM and SFM experiments summarized in Table 10-2. The data illustrate that these forces may in principle be used to transport, deposit, or modify molecules and atoms, provided that a suitable adjustment is chosen of adhesion and repulsion forces. In future applications, the controlled deposition of molecules or atoms is an important goal. For future information technologies in particular, the preparation of nanostructured

Table 10-1. Typical interaction forces between various types of atoms (calculated for typical experimental conditions determined by distances d, charges q, dielectric constants ε, ionisation energies $h\nu$, dipole moments μ and polarisabilities α, $1\ \text{Å} \triangleq 10^{-10}$ m) [12, 22].

Bond Type	Maximum Force	Conditions
Primary structure:		
Covalent bonds	$\geq 10^{-9}$ N	H_2 molecule
Secondary structure:		
Hydrogen bonds	$\leq 10^{-8}$ N	O–H..O
Coulomb, point charge	3×10^{-9} N	$q = e$, $d = 3$ Å, $\varepsilon = 1$
Charge-fixed dipole	4×10^{-10} N	$q = e$, $d = 3$ Å, $\varepsilon = 1$, $\mu = 1.85$ D
Dipole-fixed dipole	10^{-10} N	$d = 3$ Å, $\varepsilon = 1$
Bond rotation	10^{-10} N	CH_3-CH_3 rotation
Charge, non-polar	5×10^{-10} N	$q = e$, $d = 3$ Å, $\varepsilon = 1$, $\alpha = 3\,(4\,\pi\varepsilon_0)$ Å^3
Dipole, non-polar	10^{11} N	$d = 3$ Å, $\varepsilon = 1$, $\mu = 1.85$ D, $\alpha = 3\,(4\pi\varepsilon_0)$ Å^3
London dispersion	3×10^{10} N	$d = 3$ Å, $\alpha = 3\,(4\pi\varepsilon_0)$ Å^3, $h\nu = 2 \times 10^{-18}$ J
Tertiary structure:		
Charge-free dipole	4×10^{-12} N	$q = e$, $d = 10$ Å, $\mu = 1.85$ D, $T = 300$ K
Free dipole-free dipole	2×10^{-11} N	$d = 3$ Å, $\mu = 1.85$ D, $T = 300$ K
Free dipole, non-polar	5×10^{-11} N	$d = 2.3$ Å, $\mu = 1.85$ D, $\alpha = 3\,(4\pi\varepsilon_0)$ Å^3
Many body/correlations	$\leq 10^{-11}$ N	

Table 10-2. Typical interaction forces in SFM (calculated for a spherical tip and a flat surface, separated by 1 nm).

	Interaction	Maximum force (estimated)	Conditions
AFM	Van der Waals	10^{-10} N	Vacuum
		$10^{-10}-10^{-12}$ N	Solution
	Coulomb	$10^{-8}-10^{-9}$ N	Vacuum
		large variations	Solution
	Capillary	$\leq 10^{-6}$ N	Air
	Hydration	$\leq 10^{-9}$ N	Hydrophobic surface
			In electrolyte
SFM		10^{-7} N	Vacuum, air
		(measured value)	

electrodes and the control of signal transduction on the molecular scale will determine the key technologies. In the first promising experiments, metallic and even conducting polymeric structures with line widths of several tens of nanometers were produced by using an STM tip either as an electrode for electrochemical deposition or as a field-emitting electrode in chemical vapor deposition experiments which were performed in the high electrical field between the tip and the substrate.

10.3 Chemical Sensing and Molecular Recognition

The general task of molecular or elemental analysis to identify chemical or biochemical species in different environments is usually achieved without using nanostructured materials by applying the common instruments of analytical chemistry. Chemical and biochemical sensors, however, are becoming of increasing interest and importance because of their low cost, small size, microelectronic compatibility, etc. Progress in this development is basically determined by progress in the understanding of materials for sensors and sensor systems in general and for materials of the individual sensor elements with their chemically sensitive materials for molecular recognition and their optimized transducers for signal conversion in a complete sensor system, as was shown schematically above in Figure 10-4 [8]. As in a chain, the component with the weakest performance determines the overall performance of the sensor system. For most systems, the weakest component is the sensor element. Current sensor research is therefore concentrated on three approaches to reducing this bottleneck, by

- empirical optimization of sensor materials and of transducers in systematic tests under realistic measuring conditions,
- systematic studies of elementary steps of chemical sensing under thermodynamically or kinetically controlled conditions by means of microscopies and spectroscopies, and
- theoretical calculations of elementary steps of selective "key-lock" interactions.

10.3.1 Transducers

Commonly used transducers monitor chemical compositions by monitoring the phenomenological properties listed in the lower part of Table 10-3. Current microsystem technologies lead to new miniaturized transducers such as interdigital structures (for complex impedance measurements), thermopiles (for temperature measurements), piezoelectric oscillators based upon bulk, surface, or plate waves (for mass measurements), integrated optics components (for optical measurements), multi-electrode arrays (for electrochemical measurements, including recent developments for electrical connections to nerve systems), or arrays of sensor elements with integrated transducers and data processing units (for integrated "electronic noses") [8].

The trend towards improved performance, lower size and lower price per sensor element requires the systematic optimization of interface properties by means of the common tools of microscopy and spectroscopy listed in the upper part of Table 10-3. This requires in particular

Table 10-3. Survey of typical tools of surface and interface analysis which are commonly used in R&D of semiconductor and optoelectronic devices and which are now also applied in studies of molecular nanostructures with particular emphasis on chemical sensors and biomolecular devices. Measurements of phenomenological properties are optimized in practical (bio)chemical sensors by a careful choice of the transducer and its chemically sensitive coating [5, 8–10].

1. Microscopy and Spectroscopy

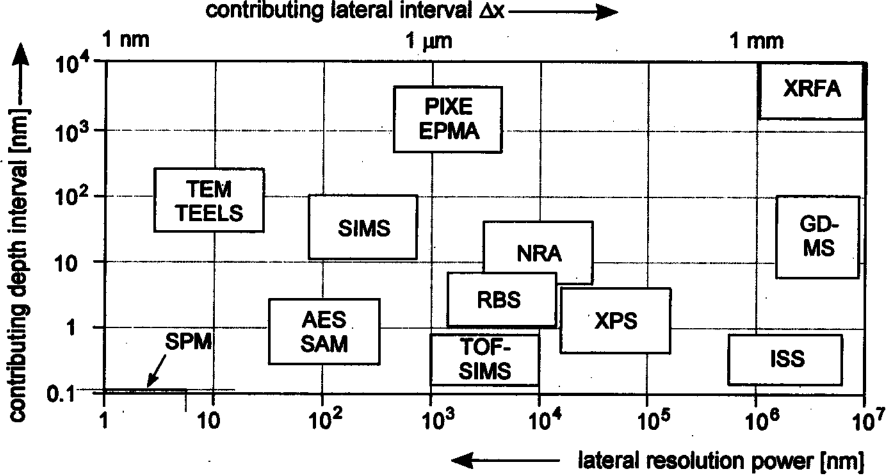
Geometric arrangement of atoms	Elemental composition	Electronic structure	Dynamic structure	Coverage of adsorbed particles, sticking coefficients		
●					SEM	Scanning electron microscopy
●					(SPA-)LEED	(Spot-profile-analysis) low electron energy diffraction
●	●				SAM	Scanning Auger microscopy
●	●				(S)SIMS	(Scanning) ion mass spectrometry
●	●				(S)EDX	(Scanning) energy dispersive X-ray analysis
●	●				ISS	Ion scattering spectroscopy
●		●			STM	Scanning tunnelling microscopy
●					SFM	Scanning force microscopy
	●	●		●	AES	Auger electron spectroscopy
	●	●		●	XPS	X-ray photoemission spectroscopy
	●			●	TDS	Thermal desorption spectroscopy
		●			UPS	Ultraviolet photoemission spectroscopy
		●			EELS	Electron energy loss spectroscopy
		●			UV/VIS	Optical spectroscopy
		●		●	(S)ELL	(Spectral) ellipsometry
		●		●	(S)SPR	(Spectral) surface plasmon resonance
		●			FS	Fluorescence spectroscopy
		●	●		HREELS	High resolution electron energy loss spectroscopy
			●		(FT)IR	(Fourier transform) infrared spectroscopy
			●		IR-ATR	Infrared spectroscopy of attenuated total reflection
			●		RAM	Raman spectroscopy

σ	Measurements of voltage-, temperature- and frequency dependent conductivities, complex impedances and photoconductivities
ϕ	Measurements of work functions
C	Measurements of capacitances
ΔE, $I(\nu)$	Measurements of electrochemical potential differences and currents as function of frequency ν
$\varepsilon(\nu)$	Optical absorption and reflexion

completely reversible bonding of the detectable particle to the sensor-active coating and completely irreversible chemical bonding of this sensor-active coating to the transducer. In current and future nanotechnologies, high spatial resolution of analytical instruments is required (Table 10-4).

This will be illustrated now for typical sensor materials which show characteristic "key-lock" interactions to detect and identify particles. Particular emphasis will be placed on the discussion of ultimate limits in the miniaturization of sensor elements as a prerequisite to developing nanosensors operating at the physical limits of miniaturization.

Table 10-4. In current and future nanotechnologies high spatial resolution of analytical instruments is required for the analysis of structures within thin layers (indicated by the lateral resolution power) and perpendicular to the layers (indicated by the contribution depth inverval in this survey) [6]. For an explanation of the abbreviations, see Table 10-3 and references given in Figure 10-5.



10.3.2 Recognition by Matching Molecular Orbitals of Electronic Conductors

Electron conductors are commonly used as chemical sensors to monitor oxidizing and reducing gases such as O_2, H_2, CO, NO_2, or hydrocarbons. The operating temperature determines the predominant interaction step: with increasing temperature, physisorption, chemisorption, surface defect reactions, catalytic reactions, and bulk reactions can be utilized for chemical sensing, provided that the interaction step leads to a change in the DC resistivity or, more general, in the AC impedance (see Figure 10-11).

Particularly suitable for miniaturizing electron-conducting devices is their operation in the chemisorption sensor mode. Depending on the position of the highest occupied (HOMO) or lowest unoccupied molecular orbital (LUMO) of the chemisorption complex relative to the Fermi level (of usually n-type conducting oxides such as SnO_2), a donor or acceptor type of interaction is monitored by changes in surface conductivities $\Delta\sigma$ or work functions $\Delta\phi$ (Figure 10-12).

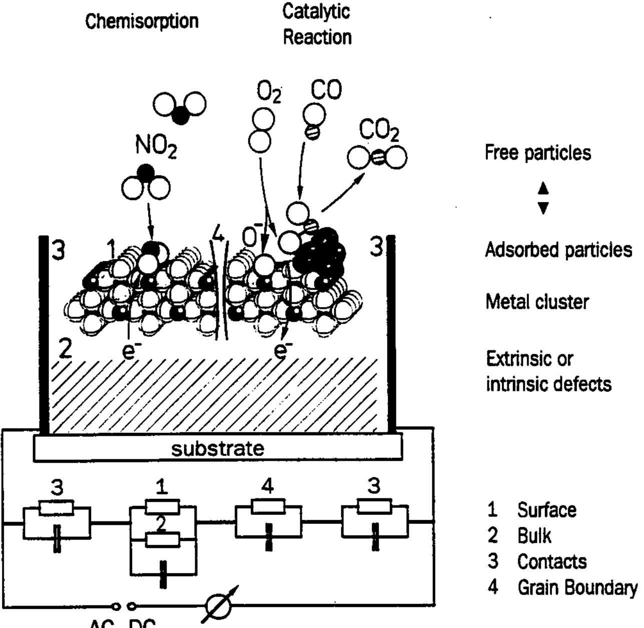


Figure 10-11. Electronic conductance sensors: Schematic presentation of different sensor-gas reactions and corresponding components of equivalent circuits formally describing the frequency-dependent (AC or DC) behavior which is influenced by reactions between the gas phase and an electronically conducting oxide. Indicated is the chemisorption of NO_2 (left) and the catalytic oxidation of CO to form CO_2 (right). For details, see text and Refs. [1-3] and [24].

Useful parameters for both the quantitative calibration of chemisorption sensors in applications and their theoretical atomistic understanding are the following:

- partial charges δ (determined from the ratio of transferred electrons per adsorbed surface complex in an evaluation of conductivity and thermal desorption experiments.);
- dipole moments μ, formally attributed to the chemisorption complex (determined from electron affinity changes $\Delta\chi$ as deduced from shifts in the ultraviolet photoemission spectra (UPS) of emitted electrons from the valence band range; see Figure 10-12, upper right section);
- isosteric heats of adsorption Q_{st} (deduced from pressure-dependent coverages, ie, adsorption isotherms at different temperatures);
- activation energies of desorption E_{act} (determined from thermal desorption experiments); and
- sticking coefficients S_0 (determined from initial rates of conductivity or work function changes upon gas exposure).

Theoretically, these parameters may be determined from quantum chemical cluster calculations by describing the atomic configurations at the adsorption site before and after formation of the chemisorption complex [5].

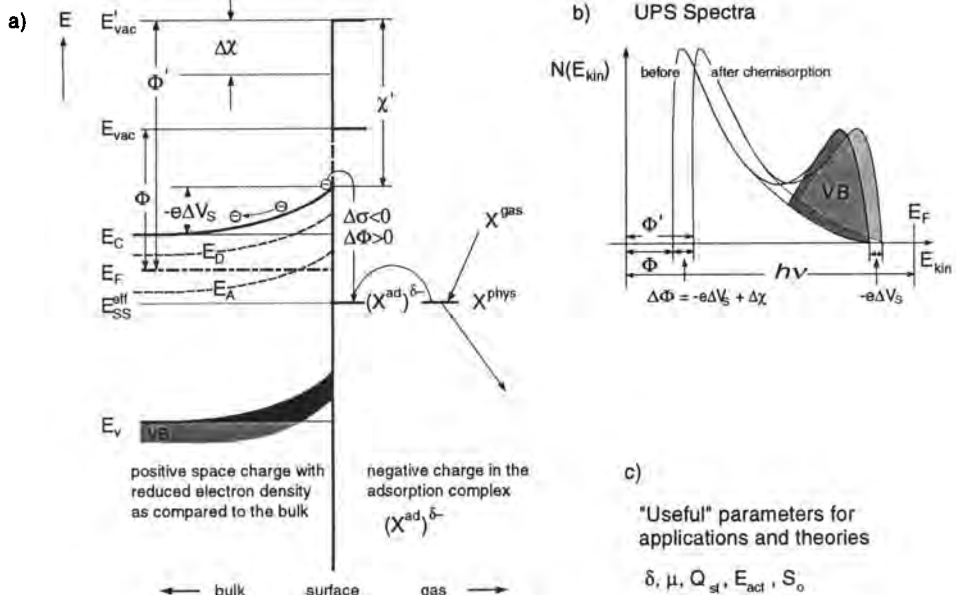


Figure 10-12. (a) Schematic presentation of chemisorption and charge transfer at a semiconductor surface in the electronic band scheme of the surface. The gas-phase molecule X^{gas} forms the precursor physisorption state X^{phys}. Also indicated is the work function ϕ, electron affinity χ, and band bending $e\Delta V_s$. Chemisorption may be measured by monitoring changes in the work functions $\Delta\phi = \phi' - \phi$ and changes in surface conductivities $\Delta\sigma$ on changes in occupation of effective surface states E^{eff} by forming $(X^{ad})^{\delta-}$ (example of acceptor-type$_{ss}$ chemisorption). The band edges are denoted by E_V (valence band) and E_C (conduction band). The position of the Fermi level E_F in the bulk is determined by bulk donors E_D and acceptors E_A. (b) UPS spectrum (intensity N of photoemitted electrons at a given kinetic energy E_{kin}) before and after chemisorption with characteristic variations in the zero energy cut-off (determining $\Delta\phi$) and in the valence band onset (determining $-e\Delta V_s$) [1, 25]. (c) The link between practical sensor applications and theory is given by determining in both cases the partial charges δ, dipole moments μ, isosteric heats of adsorption Q_{st}, activation energies of desorption of chemisorbed particles E_{act} and initial sticking coefficients S_0. For further details, see text and Refs. [26] and [28].

Conductivity measurements are particularly sensitive in nanocrystalline thin films if the Debye length L_D is larger than the semiconductor particle radius $r = \ell/2$ (Figure 10-13). Under these conditions, charge transfer across grain boundaries in the film instead of charge transfer parallel to the surface of an ideal continuous thin film is modified during chemical sensing and equilibrium coverages of less than 10^{-5} monolayers of donor or acceptor molecules can easily be detected.

Current R&D of electron conducting sensors is aimed at optimizing contact geometries, film thicknesses, particle size distributions, doping, operation temperatures, and operation frequencies. Usually, film thicknesses are chosen in the micrometer range with a recent trend to stabilize identical diameters of nanosized SnO_2 particles in these films. In principle, the charge transfer of individual electrons to or from the chemisorption complexes (see Fig-

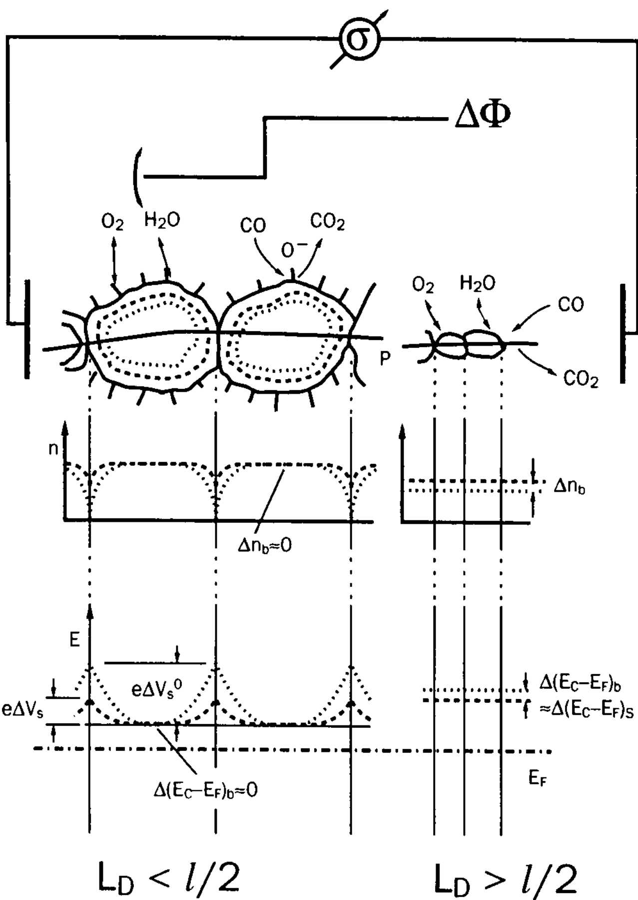


Figure 10-13. Schematic presentation of an experimental set-up to monitor conductivities σ and work function changes $\Delta\phi$ of nanocrystalline semiconducting particles (n-type) with Debye lengths L_D smaller (left) and larger (right) than the particle radius $r = l/2$. Charge carrier concentrations of electrons (n) and electron energies (E) in the band scheme are shown in the lower part. The subscript b denotes bulk and s surface values. For details, see Figure 10-12 and text.

ures 10-12 and 10-13 and for $L_D \gg r = l/2$) should be detectable by steps in scanning tunnelling spectroscopy (STS) spectra of nanosized SnO_2 particles in an arrangement similar to that chosen in recent single-electron tunnelling experiments with silicon-based devices.

10.3.3 Recognition by Matching the Doping of Mixed Conductors

Electron and ion conduction of mixed conductors is usually monitored at elevated temperatures ($T > 700K$) in gas sensors based on oxides such as TiO_2, Ga_2O_3, or $SrTiO_2$.

Platinum is often used to improve the selectivity of gas detection by adjusting the catalytic activity of the surface. The latter results from energetic shifts of the bulk Fermi level or from specific chemisorption sites at the surface. Pt is also often used to contact the sensor material electronically. The Pt contacts on TiO_2 are of particular interest because recent experiments show a switching phenomenon in the current/voltage behaviour across this interface:

- At low temperatures ($T < 700$ K) the electronic conductivity can be utilized in sensitive Schottky-barrier sensors to monitor gases such as O_2 or CO (see typical results in Figure 10-14). Under these conditions, the Pt atoms at the surface act as acceptors.
- At higher temperatures and higher oxygen partial pressures, Pt^{2+} ions are formed at the surface and subsequently move to subsurface positons at which they are stabilized as Pt^{4+} ions. The Pt^{4+} ions are significantly smaller than the Pt^0 atoms and act as subsurface ("bulk") donors between the first and second TiO_2 layers (Figure 10-15). As a result, a drastic change occurs in the conductivity across the Pt/TiO_2 interface from Schottky barrier to ohmic contact behavior (see Figure 10-14). Under the latter conditions, the point defect thermodynamics of oxygen vacancies in the TiO_2 bulk (see reaction in the lower right part of Figure 10-16) can be utilized to monitor oxygen over a large pressure range by the mixed (electron and oxygen vacancy $V_O{}^{2+}$ or alternatively oxygen ion O^{2-}) conduction mode.

The two different operation conditions of Pt/TiO_2 interfaces and their charge-generating steps as surface state controlled chemisorption and bulk controlled defect formation both depending on p_{O_2} and T may be illustrated schematically in a band scheme (Figure 10-16).

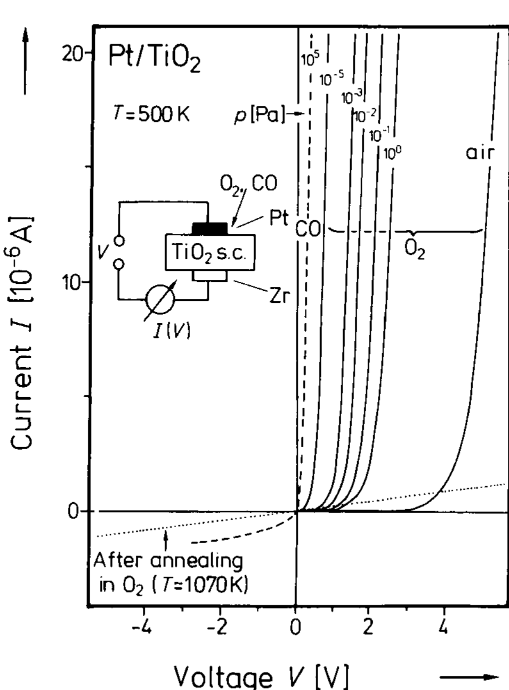


Figure 10-14.
Current-voltage (I-V) characteristics for Schottky barrier and ohmic contact sensors based on Pt/TiO_2 in an arrangement as illustrated in the upper left part of the figure. The Zr/TiO_2 interface shows ohmic contact behavior at the rearside. Similar changes in the sensor response are observed for single-crystal (sc) or thin-film samples of TiO_2. For details, see text and Ref. [27].

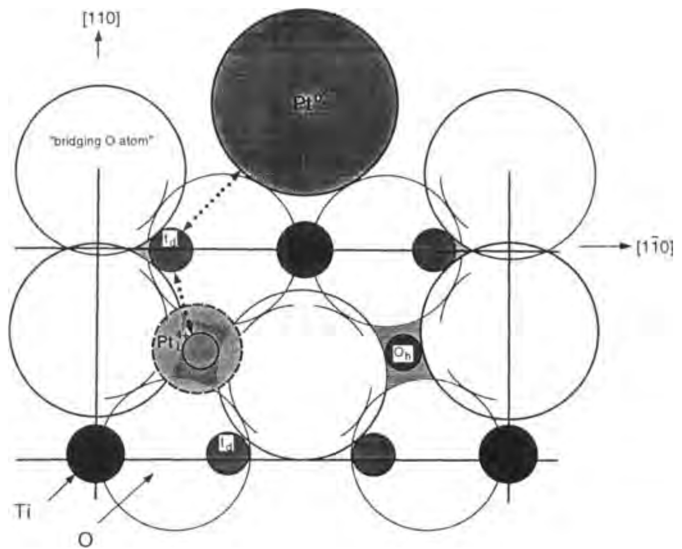
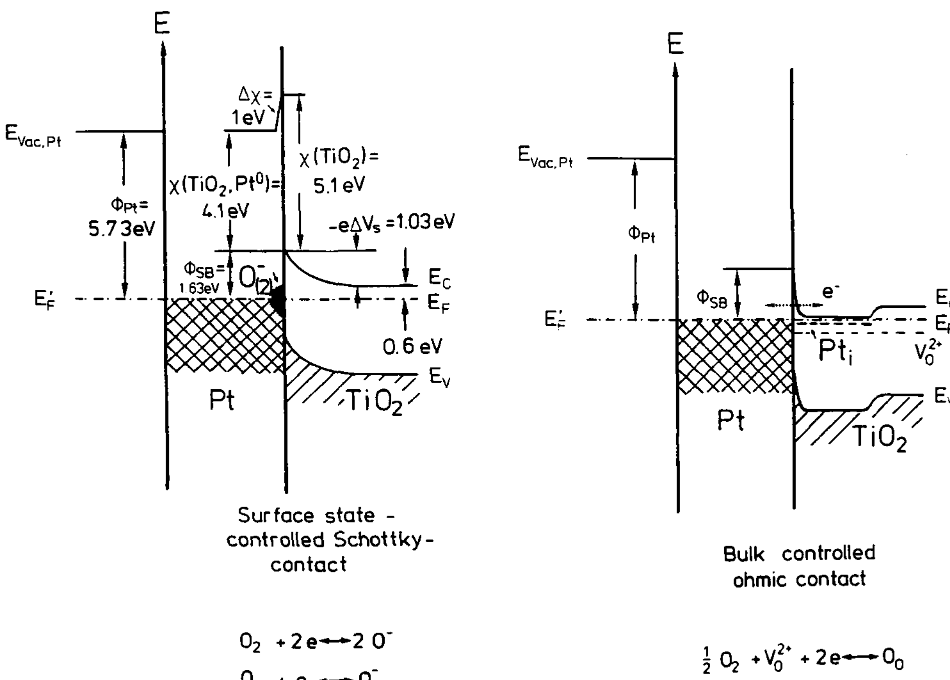


Figure 10-15. Surface and subsurface geometry of metallic Pt^0 atoms, ionic Pt^{4+} ions, tetrahedral interstitial sites t_d, octahedral interstitial sites O_h at $TiO_2(110)$ surfaces and the diffusion path of Pt between the surface and the subsurface site (dotted line) [27].



$O_2 + 2e \rightleftharpoons 2O^-$

$O_2 + e \rightleftharpoons O_2^-$

$\frac{1}{2}O_2 + V_O^{2+} + 2e \rightleftharpoons O_O$

Figure 10-16. Energy diagram and characteristic reactions leading to changes in the Schottky barrier and ohmic contact behavior of $Pt/TiO_2(110)$. E_F is the Fermi level, ϕ_{SB} the Schottky barrier height, $e\Delta V_s$ the band bending, $\Delta\chi$ the change in electron affinity on evaporation of Pt atoms onto TiO_2, E_{vac} the vacuum level, and E_V and E_C the valence and conduction band edge, respectively. For further details, see text and Ref. [27].

Recent investigations are aimed at miniaturizing these devices also down to the atomic scale. The two different sensor operation modes may be adjusted intentionally at medium temperatures by externally applying a positive or negative DC voltage to the Pt electrode and by thereby switching the effective platinum charge between neutral Pt^0 and ionic Pt^{4+} species and vice versa. This also switches the sensor and the catalytic properties of the Pt/TiO_2 interface. The next step is to perform this switching with individual Pt atoms in an STS set-up and to modify the current-voltage curves taken at individual surface or subsurface Pt sites by charge-transfer reactions as they occur during chemical sensing of donor- or acceptor-type molecules.

10.3.4 Recognition by Matching Electrochemical Potentials of Ion Conductors

ZrO_2 is widely used as a chemical sensor to monitor O_2 by utilizing its high-temperature ion (O^{2-}) conduction in the bulk. In *potentiometric devices,* the cell voltage is determined by the pressure difference between two Pt contacts of ZrO_2 samples. In *amperometric devices,* the current is monitored at a given voltage. Of key importance for both applications is the stability of the three-phase boundary at which oxygen molecules from the gas phase are converted to ions in the solid (Figure 10-17). A detailed understanding of potentiometric devices is deduced from an experimental determination of the energetic position of the Fermi level E_F under various operating conditions (Figure 10-18).

By externally applying a voltage, the value of E_F at the electrode and thereby the concentrations and chemical reactivities of different oxygen species at the surface may be shifted intentionally with reference to a constant electrochemical potential of an oxygen reference electrode (ie, a suitable metal-metal oxide system such as Ni-NiO or Pd-PdO). This leads to variations in the kinetics of different electrocatalytic reactions, some of which are illustrated in Figure 10-19. As a result, different educts from the gas phase react predominantly at different voltages. The charge-transfer rate monitored in this amperometric operation mode at

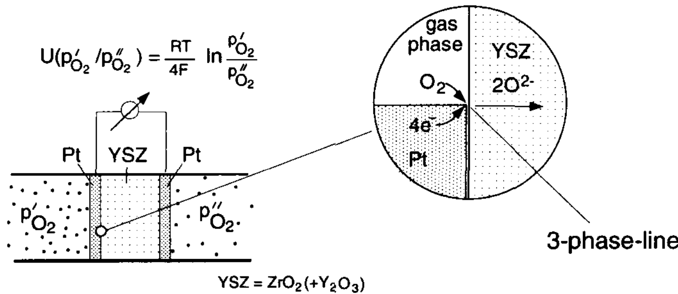


Figure 10-17. *Ionic conductance sensors:* Schematic presentation of the three-phase boundary in Y-stabilized ZrO_2 (YSZ) sensors to monitor differences in oxygen partial pressures in the gas phase. The Nernst equation in the upper left part describes the potentiometric sensor response (voltage U) to differences in partial pressures p_{O_2}, where F is the Faraday constant and the number 4 represents the number of electrons transferred per O_2 molecule to form O^{2-} ions in the bulk of ZrO_2.

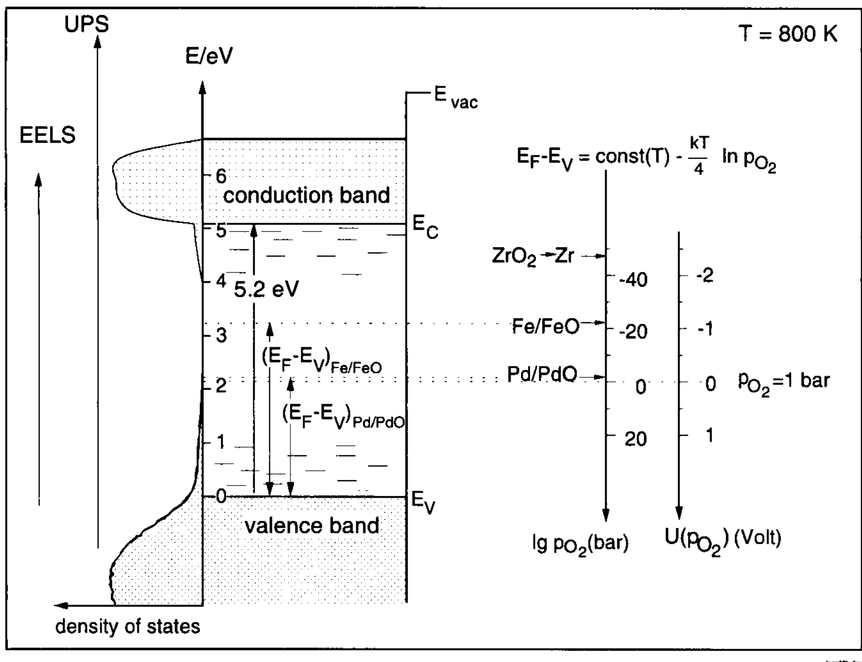


Figure 10-18. Electron energies in a band scheme of YSZ (compare notation in Figure 10-12). Experimental data from in situ electron energy loss (EELS) and ultraviolet photoemission (UPS) spectroscopic results were obtained for different partial pressures of oxygen in measurements using reference contacts based on metal–metal oxide systems. At constant temperature ($T = 800$ K was chosen here) the reference contacts (Fe–FeO and Pd–PdO) determine the chemical potential of oxygen at the energetic positions as indicated in this figure. The YSZ sample decomposes for E_F positions above the ZrO_2–Zr equilibrium. Voltages are referred to an oxygen partial pressure of 1 bar (E_F–E_V with the reference voltage $U = 0$ V). Applied voltages shift the Fermi level relative to the reference contacts chosen in the experiment. This leads to the nominal drastic variations in the oxygen partial pressures also indicated in this figure. For details, see Figure 10-10 and Refs. [1], [10], [28] and [29].

constant voltage is determined by the concentration of educts in the gas phase which compete with O_2 to form a mixed potential at the surface. Recent developments of new electrode materials (mostly perovskites) make possible the amperometric detection of CO, NO, and hydrocarbons. Independent of this sensor aspect, they are of practical importance in completely different fields of application because they make it possible to develop very efficient catalysts, fuel cells, or electro-catalysts.

The miniaturization of such devices is limited by several fundamental aspects of physical chemistry concerning the required properties of the bulk and electrode materials. The first aspect concerns the *electron-conducting* electrodes. Because of large differences in the surface free energies between metallic Pt and oxidic ZrO_2, relatively large three-dimensional Pt clusters are required to form a spatially coherent net of an electron-conducting contact which allows the O_2 to access the three-phase gas/electrode/solid electrolyte boundary (left part of Figure 10-19). For lower cluster concentrations, percolation effects become crucial. Minia-

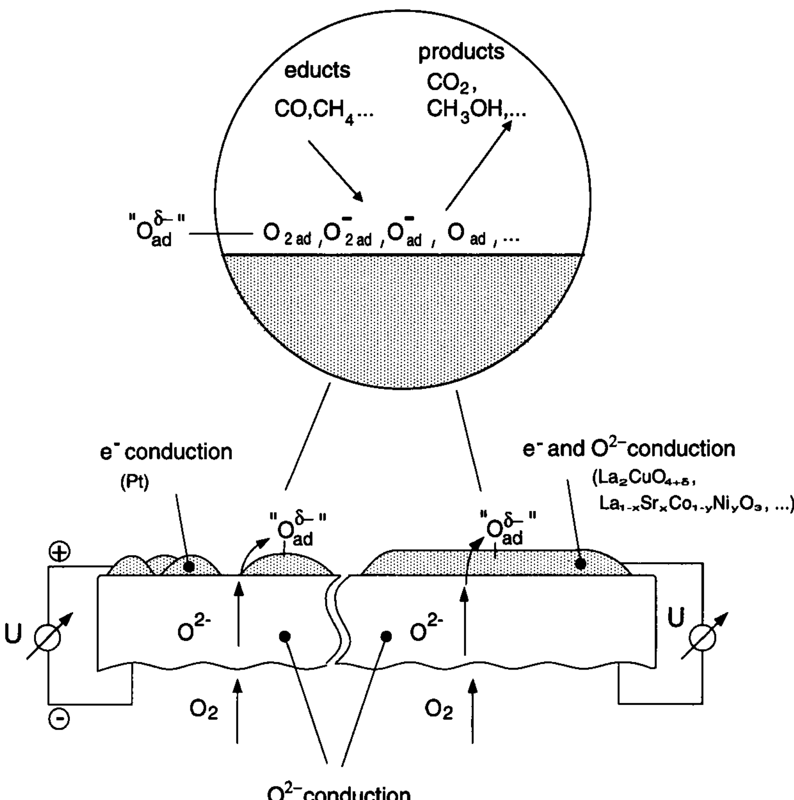


Figure 10-19. Schematic illustrations of typical oxygen surface species, electrocatalyst reactions converting educts to products, sensor parameters, and fuel cell properties of ion conductors which change by applying an external voltage. The electrode forms either a three-phase boundary (left side, eg, Pt-YSZ-gas) or a two-phase boundary (right side, perovskite-gas). For details, see text and Refs. [1, 28, 29].

turization of such contacts is therefore more promising if *mixed conducting* perovskites with surface free energies similar to that of ZrO_2 are used which form continuous overlayers (right part of Figure 10-19). The ultimate limit to reducing the thickness of such mixed conducting layers and of *ion-conducting* ZrO_2 thin films (which have to be impermeable for neutral O_2 molecules) in these sandwich devices cannot be estimated yet because corresponding epitaxial thin-film structures have not been prepared, and as a result, conductivity phenomena in monolayer thin films have not yet been investigated. Evidently, the commonly used scaling laws in the size reduction of microelectronic devices cannot be applied under these conditions and local bonding and quantum size effects must be taken into account for adjusting the chemical and electronic properties of ultra-thin films.

Typical surface and interface properties to be controlled in such devices are summarized in Figure 10-20.

Because of the experimental problems in preparing ideal epitaxial thin films with structural control in the monolayer range, the different transport phenomena of charged or neutral par-

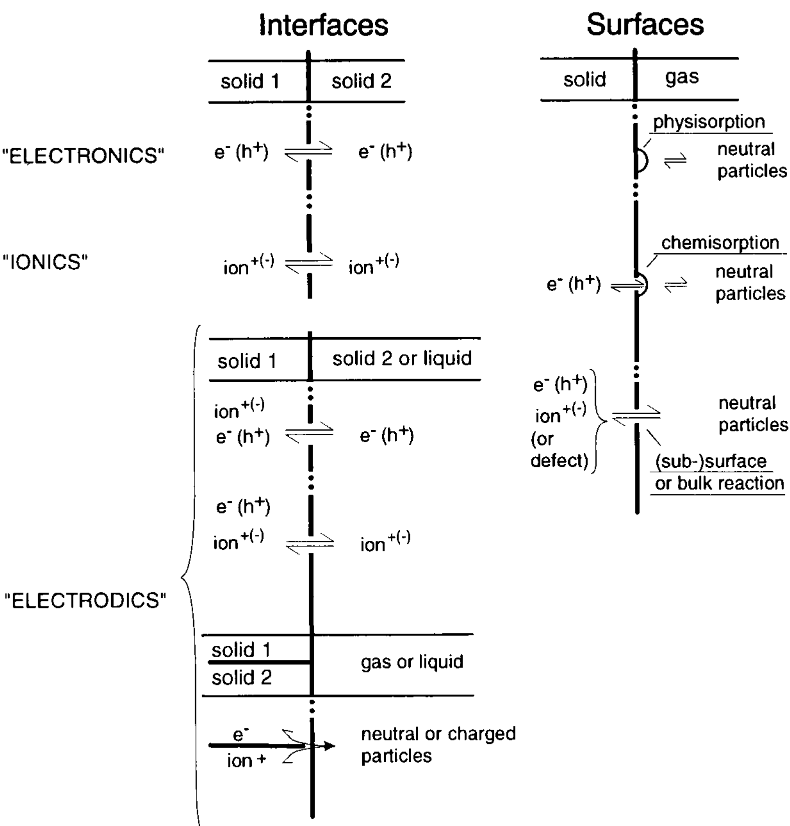
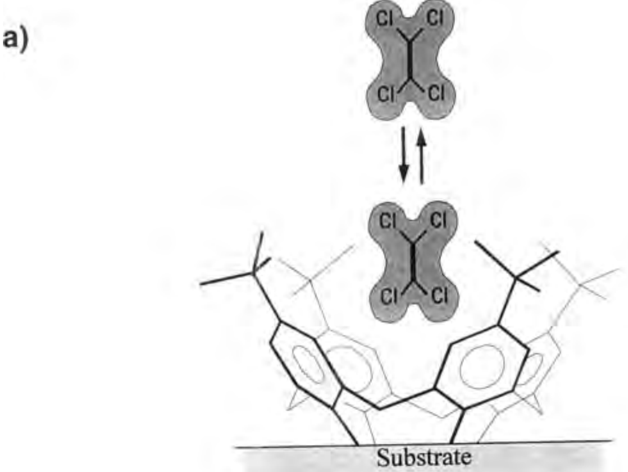


Figure 10-20. Survey on interfaces and surfaces of electronic or ionic devices including chemical sensors. For details, see text and Ref. [1].

ticles in these films and across their interfaces are not yet understood on the atomic scale. Consequently, ultimate limits for the miniaturization of these chemical sensors cannot be estimated precisely. Particular problems arise from the fact that the long-term stability and reversibility are the most important properties for any application of future "nanosensors" to identify chemical species.

10.3.5 Recognition by Key-Lock Interactions with Static Molecular Cages

A completely different optimization is required for chemical sensors which make use of supramolecular recognition structures. Typical molecules to utilize the geometric "key-lock" principle are calixarenes, which incorporate small organic molecules (Figure 10-21 a). Signal transduction of these sensors is possible by monitoring *changes in mass, capacitance or*



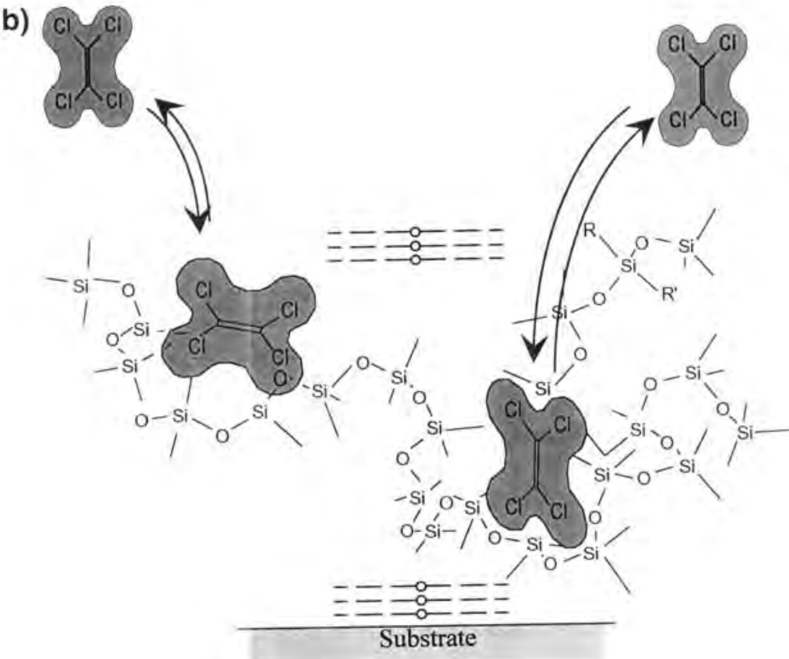


Figure 10-21. (a) Interaction of a typical volatile organic compound (VOC, here perchloroethylene, C_2Cl_4) with a simple supramolecular structure (here *tert*-butylcalix(4)arene) with its surface interaction site facing towards the gas phase. Because of the rigid recognition structure, the simple key-lock principle describes the interaction mechanism in this situation [30]. (b) Dissolution of VOCs into the bulk of polymeric thin films (here C_2Cl_4 in polydimethylsiloxane, PDMS). Because of the large fluctuations in the time-dependent structure, induced fits describe the interaction mechanism in this situation [30, 31].

temperature on incorporation of organic molecules. The geometry of calixarenes may be modified by substituting atoms in the ring or by using different numbers of repeat units in the ring. The molecule-ring interaction energies can be calculated with sufficient accuracy by static force field approaches. Larger molecules require the theoretical calculation of dynamically adjusted recognition centers ("induced fits") by molecular dynamic concepts (see Section 10.3.6). Such theoretical molecular design concepts are becoming of increasing interest also for developing new pharmaceuticals ("drug design").

To achieve a controlled key-lock geometry, the covalent coupling of cage compounds to Au(111) via sulfur bridge bonds may be used to advantage in chemical sensors. An example is the structure shown in the upper part of Figure 10-21 a, but used as an extended two-dimensional layer of resorcinarenes on Au(111) as coating of a mass-sensitive transducer (Figure 10-22).

The ultimate step towards miniaturization of these sensors is to manipulate individual molecules (the "keys") with an STM tip which has to be modified by a cage compound (the "lock", as mentioned in Section 10.2.2). This approach is particularly promising if the tip voltage causes sufficient changes in the interaction energies.

10.3.6 Recognition by Induced Fit with Dynamic Molecular Cages

In contrast to the host-guest interaction treated in the rigid lattice approach to describe the key-lock interaction in the last section (Figure 10-21 a), a huge variety of materials are commonly used for chemical sensing in which the host-guest interaction is treated as an embedding effect in a flexible lattice (induced fit in the liquid state treated by molecular dynamics calculations, see schematic presentation in Figure 10-21 b). Typical materials chosen in this context which usually show specific absorption phenomena include the following:

- liquid crystals of side-chain modified polymers of siloxanes which monitor, eg, volatile organic compounds (VOCs) such as benzene by the key-lock interaction via changes in the mean capacitance due to the disturbed nematic order;
- molecular liquids with dissolved dye molecules which show color changes upon association of gas molecules because of electronic charge-transfer interactions;
- extra- and/or intramolecular cages of large supramolecular structures which show mass changes upon host-guest interactions;
- polymers modified by characteristic chemical functional groups or intermolecular cages. The latter combine two advantages, ie, the fast bulk diffusion coefficients of gas molecules in the open network of the polymers and the usually higher selectivity of rigid supramolecular cages. A huge variety of different polymers shows excellent sensing properties to, eg, monitor VOCs in the gas, but also liquid phase (Figure 10-21 b; see also examples in Section 10.4). Recent trends in the miniaturization focus at the synthesis and preparation of long-chain oligomeric instead of polymeric structures with well controlled local variations in the chemical composition of the active recognition center.

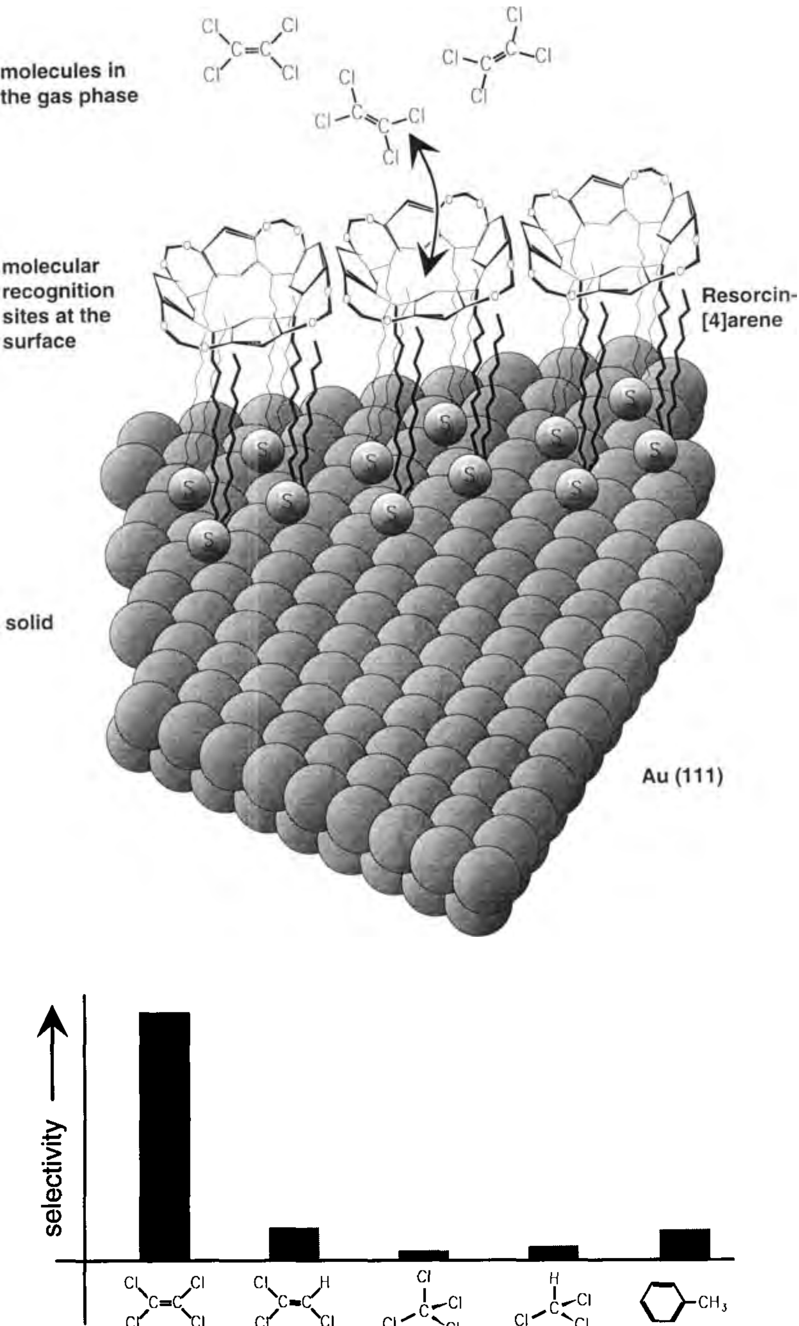


Figure 10-22. Schematic illustration of covalent bondings of a cage compound to a metallic Au(111) substrate via sulfur binding (S atoms). Molecules C_2Cl_4 detected from the gas phase with these cages by a mass-sensitive transducer (quartz crystal oscillator) are illustrated in the upper part [32].

10.3.7 Biosensors Based upon Key-Lock Interactions (Bioaffinity Sensors)

A variety of biosensors are used to monitor high molecular weight molecules in the environment, in body fluids, etc. Usually four different principles of biosensors are distinguished, which are shown schematically in Figure 10-23. The first category, bioaffinity sensors, function like the recognition centers described in Section 10.3.6 by induced fit (dynamic) key-lock interactions in molecular cages. Induced fit occurs here in a natural receptor. A variety of receptor-based interactions may be used for highly selective biomolecular recognition. The controlled preparation of thin films with embedded receptor structures of high molecular weight is a difficult task. One approach is therefore to mimic the biological function by fully synthesizing structures of lower molecular weight which represent a copy of the recognition site of the biological function unit. This "biomimetic" approach had been chosen in a typical example, illustrated in Figure 10-24. Layers of synthetic peptide epitopes with two different alkanethiols are covalently coupled to gold substrates by using a mixed composition of small (unmodified) and large (epitope-modified) molecules. Low surface coverages of the latter are required for thin-film devices to monitor selectively large biomolecules. In this case the antibodies of the hoof-and-mouth disease virus have been detected after their attachment to the epitopes by monitoring capacity variations of the sandwich structure. The signal transduction across the interface determines the sensitivity in these devices and limits their minimum lateral dimensions.

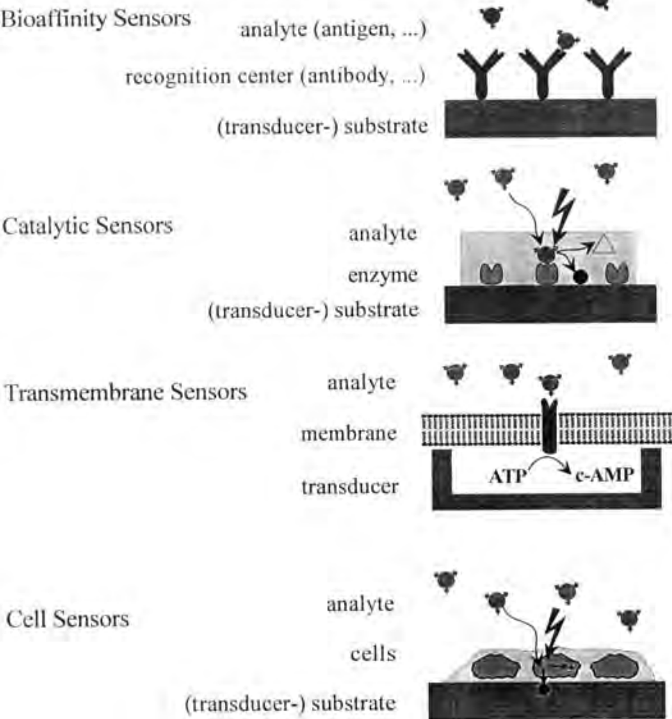


Figure 10-23. The four different principles of biosensing in a schematic overview [6].

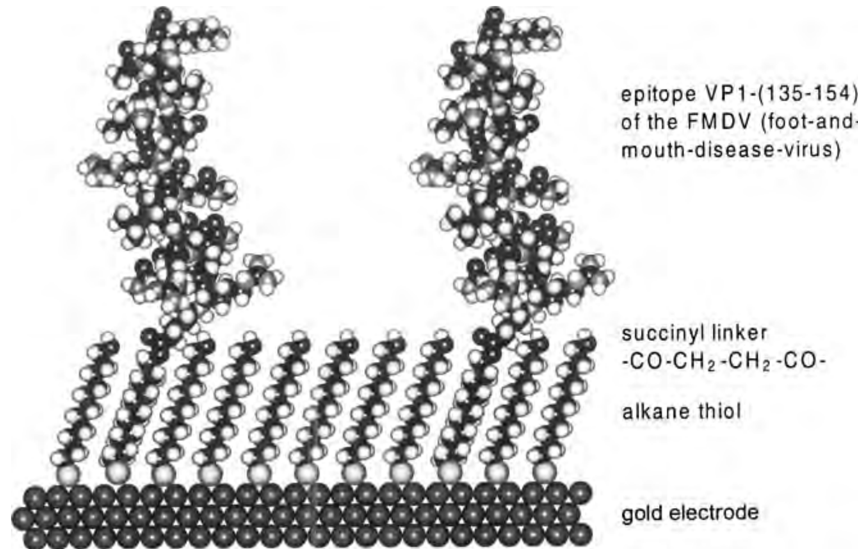


Figure 10-24. Schematic presentation of the covalent coupling of two different alkanethiol units via sulfur atoms to Au(111). The longer component contains an epitope which selectively interacts with the antibody (hoof and mouth disease virus) [1], [33].

10.3.8 Biosensors Based upon Catalytic Reactions (Enzyme Sensors)

Enzymes are particularly suitable for biomolecular recognition. The catalytic activity of their active sites may be monitored sensitively by variations in the electrochemical behavior at the electrode. This, however, requires a "molecular wiring" which makes possible an ionic or electronic charge transfer between the active site and the electrode. A few different concepts are indicated in Figure 10-25. None of these molecular wiring concepts, however, is understood yet on the molecular scale, even for the best understood enzyme glucose oxidase with a molecular weight of 186000 atomic mass units. Further details of molecular wiring are mentioned briefly in Section 10.5 in view of their importance for future molecular electronic devices.

10.3.9 Biosensors Based upon Membrane Functions

Because of their great importance in the living cell, a variety of such biomolecular function units in the membrane have been identified and may in the next step be utilized as components of future "intelligent" biomolecular devices, provided that they can be contacted or interfaced to appropriate transducers. To achieve stability, they must be incorporated in natural or synthetic bilayers. A recent example is the co-transporter molecule lactose permease (LP), which shows a 1 : 1 active transport of protons and lactose molecules (Figure 10-26). In an ion-sensitive field effect transistor (ISFET) arrangement, the protons make it possible to monitor variations in the lactose concentrations in the outer compartment by monitoring pH changes

in the inner ISFET compartment. Perfect spreading of the modified bilayer and stabilizing the thickness of the inner compartment between the bilayer and the transducer are the main issues of current R&D in this field, which limit their miniaturization and perfection.

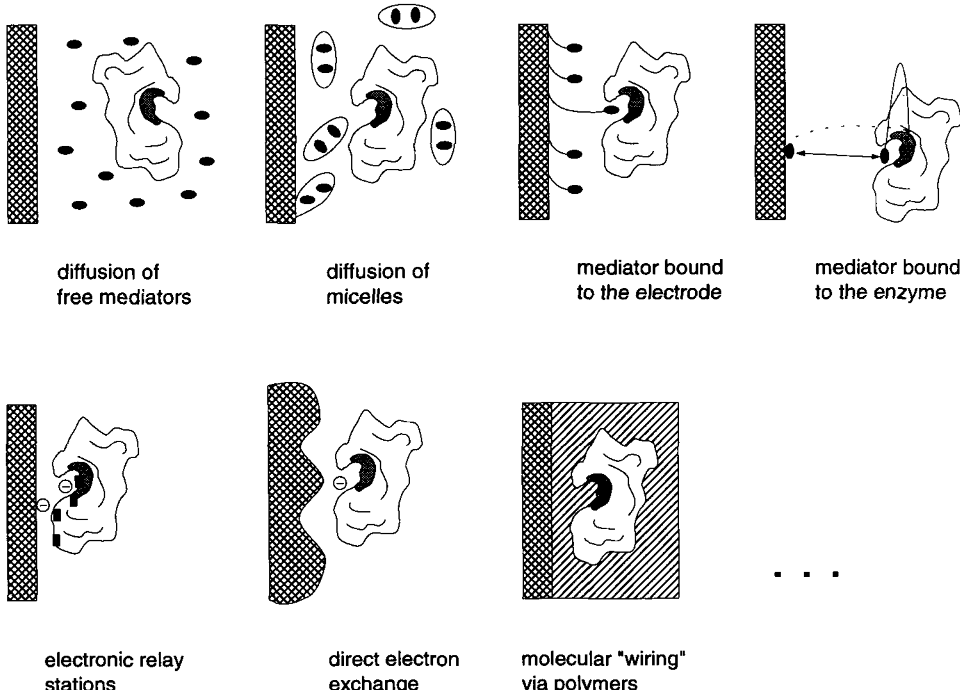


Figure 10-25. Different concepts for mediators linking the information about the electrochemical status at the nanosized catalytical active center ("inner electrode" at the pocket site) of the enzyme with the outer macroscopic electrode [5], [34].

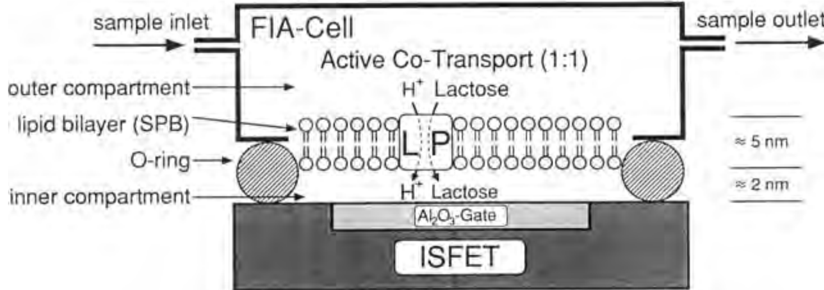


Figure 10-26. Schematic presentation of a biosensor based on an active co-transport using lactose permease (LP) which transports protons and lactose in a strictly 1 : 1 ratio in a flow injection analysis (FIA cell) arrangement. Lactose may be monitored by monitoring pH changes with an ion-sensitive field effect transistor ISFET [35].

10.3.10 Biosensors Based upon Whole Cell Functions

Because of their highly specialized biochemical properties and metabolic mechanisms, the use of whole cells and even of tissues becomes of increasing interest for the recognition and identification of high molecular weight species such as toxic molecules, drugs, and pharmaceuticals. The design problem to make use of such very complex biomolecular function units is the stability of the interface between these units and the substrate or the electrode. A variety of different concepts have been developed to deduce electrical signals from whole cells or tissues. Some typical set-ups for monitoring cell behavior are shown schematically in Figure 10-27. High spatial resolution of outer point contacts to monitor ionic or electronic current is desirable to understand, for example, the function of individual transmembrane proteins. Using metallically conducting substrates, electrical extracellular signals may be monitored. Although details about signal transduction in this arrangement are still not understood, the signals may formally be described by electrical equivalent circuits, correlated empirically with the metabolism in the cell, and hence, for example, may be used to investigate the influence of drugs on the cell metabolism. Evaluation of such signals from an array of many cells improves the statistics and hence the empirical correlation with metabolic states. Arrays may also be used to investigate cell-cell interactions, learning processes in neural networks, etc. In all of these applications, however, comparative patch-clamp and extracellular signals obtained in controlled geometric arrangements of the electrodes are highly desirable but still not yet available.

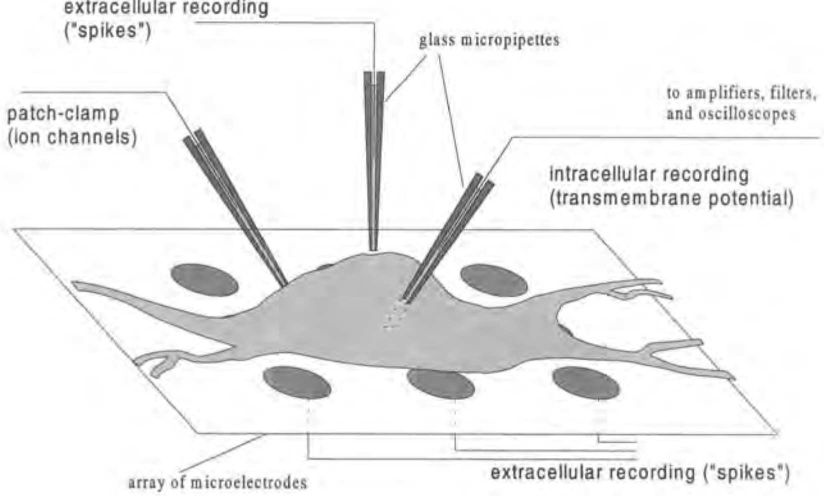


Figure 10-27. Schematic presentation of different electrophysiological recording techniques to investigate cells or neurons in a network by either intracellular or extracellular recording [36].

10.4 Sensor Arrays and Neural Networks

As already mentioned in the Introduction in the context of the development of an electronic nose, sensor arrays in which different stable sensor elements are used that show high selectivities and sensitivities towards the detection of different molecules are becoming of increasing importance. The electronic nose approach with data analysis from sensor arrays and subsequent pattern recognition will make it possible to characterize gas mixtures quantitatively or odors qualitatively. Typical examples of results from an electronic nose based on a sensor array which utilize modified polymers are shown in Figures 10-28 to 10-30. For further details, see Refs. [37–40] and references cited therein.

The alternative approach of using different specialized cells with their different individual specific receptors in an array, thereby imitating more directly the function of the human nose, cannot be realized yet because of the inherent problems of interfacing cell membranes with electrodes (see Section 10.3).

Another approach is currently being pursued in which nerve cells and cell arrays of connected neurons are monitored by their extracellular signals which average over certain cells or events. The data analysis of such complex systems, however, is not yet possible at the level of understanding the natural neural net. Therefore, a huge gap exists between neural network analysis developed from man-made sensor systems and the neural net in the brain which operates with a completely unknown algorithm.

10.5 Molecular Electronics

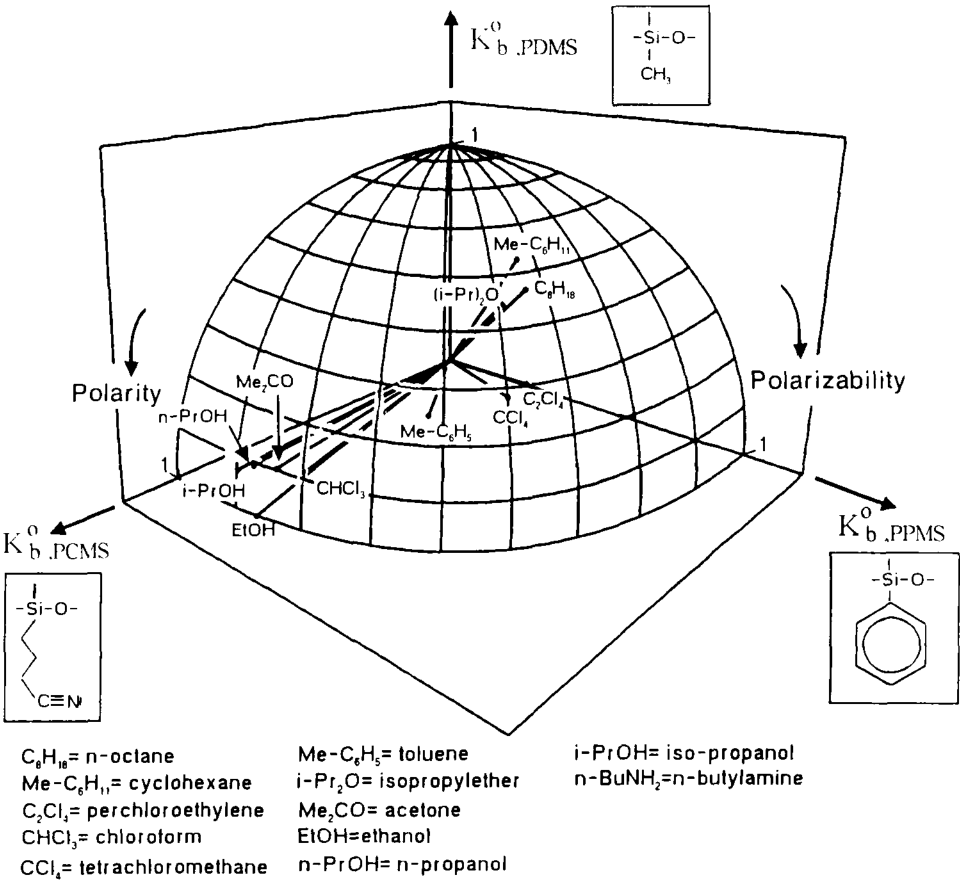
For complete nanosystems, the signal transduction, data evaluation, and the generation of the output signal have to be performed on the nanometer scale [7, 41, 42]. Besides the inorganic "top-down" approach to structure semiconductors to smaller and smaller dimensions, the alternative concept of using small assemblies of molecules has also been followed in the last few years. In this context, two definitions for molecular electronics are used:

- (conventional) electronics with molecular materials; and
- electronics on a molecular scale.

In the first approach, molecular semiconductors such as conducting polymers and oligomers, eg, polyacetylene, polypyrrole, and oligothiophenes, are used in conventional designs of field effect transistors or light-emitting diodes. These materials have some advantages over inorganic semiconductors such as mechanical flexibility and the advantage of easy tailoring of their properties by chemical synthesis. On the other hand, the major drawback is the lack of long-term stability in these devices. Therefore, "conventional" pure organic or hybrid electronics will only play an intermediate role in electronics, if any.

The second approach makes use of the specific molecular properties of organic materials. First ideas [11] were aimed at the utilization of a single conducting polymer chain as a molecular wire and of bistable molecules as molecular switches, some of which are illustrated

in Figure 10-31. Today we know that, according to Heisenberg's uncertainty principle, electronics based on the information stored in only one molecule will in principle not be possible. Therefore, small ensembles of 10^4 or more molecules have to be manipulated and controlled. Although no idea for a complete concept of a molecular computer exists at the moment, there is still some enthusiasm that such a concept will be viable in the future. Current R&D is aimed at solving basic problems of synthesizing, manipulating, structuring, and contacting prototype molecular units.



$$K^0_{b,PDMS} = K_{b,PDMS} / [K_{b,PDMS}^2 + K_{b,PPMS}^2 + K_{b,PCMS}^2]^{1/2}$$

Figure 10-28. Pattern recognition with sensor arrays: Selectivities are indicated of three modified polysiloxanes (PDMS, PCMS, and PPMS, for details, see Figure 10-30) to monitor the different VOCs listed in the lower part. The three different partition coefficients K_b obtained for one gas with the three different sensors were normalized to K_b^0 in order to make direct comparisons of orientations possible in one figure. Polarizabilities and polarities of modifying side-groups in these polymers change the selectivities in a characteristic manner. This is a prerequisite for the development of different sensors to be used in sensor arrays for an analysis of gas mixtures [1, 37–40].

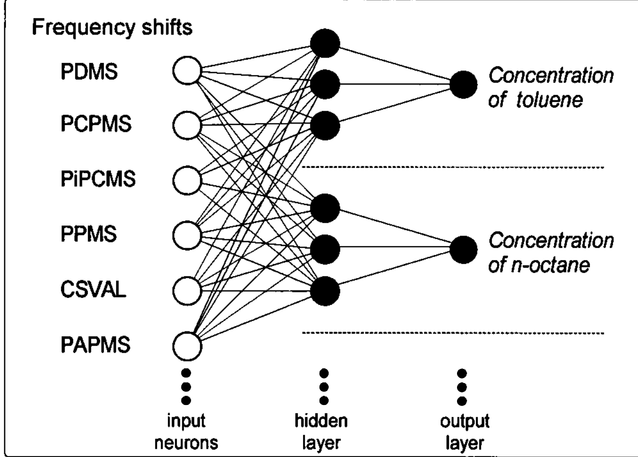


Figure 10-29. Example of pattern recognition with sensor arrays by means of neural networks: Schematic of the network architecture for the correlation of sensor signals using the polymers as selective coatings including those shown in Figure 10-28 with different VOC analyte concentrations.

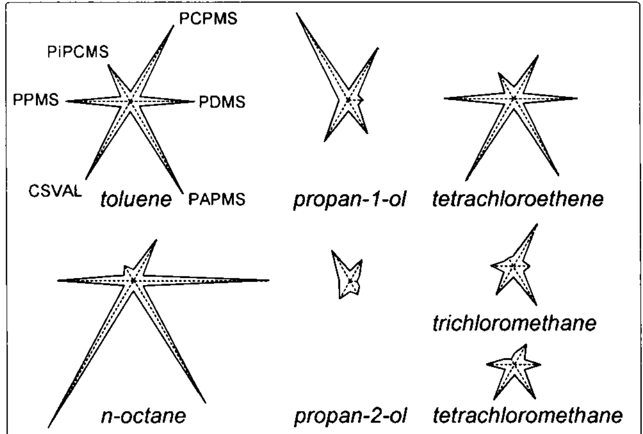
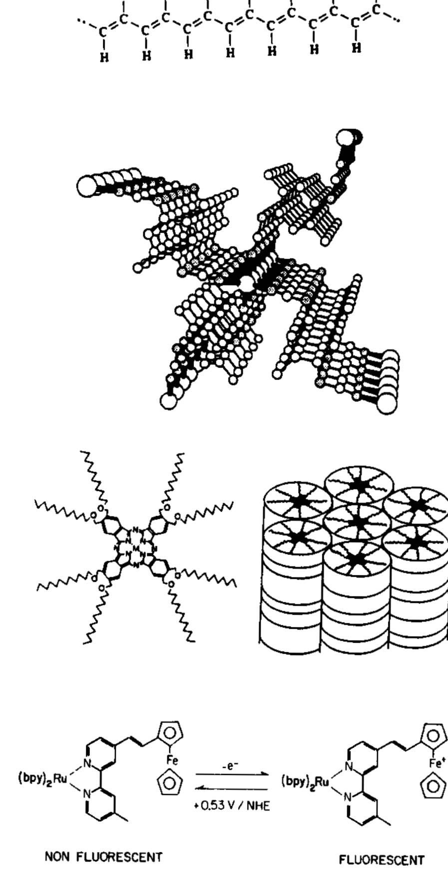


Figure 10-30. Star symbol plot for the partial sensitivities of the sensor array in Figure 10-29 to different organic solvent vapours. The results obtained from the polymer coatings PDMS (polydimethylsiloxane), PCPMS (poly(cyanopropyl)methylsiloxane), PiPCMS (poly(isopropylcarboxylic acid)methylsiloxane), PPMS (polyphenylmethylsiloxane), CSVAL (Chirasil-Val, a siloxane with chiral valine groups), and PAPMS (poly(aminopropyl)methylsiloxane) are separated by angles of 60 ° (dashed lines). The points of the stars mark the partial sensitivities of these materials to 800 ppm of toluene, n-octane, propan-1-ol, propan-2-ol, tetrachloroethene, trichloromethane and tetrachloromethane. The star symbol for tetrachloroethenes is scaled to one fifth of its real size. For PPMS, the partial sensitivity to propan-1-ol and propan-2-ol was below detection limit [1, 39].

a) Wires

b) Switches



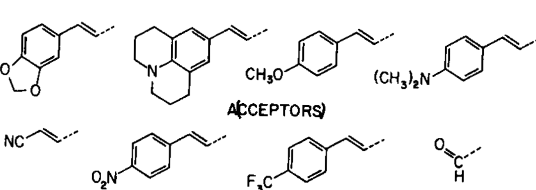


Figure 10-31. Schematic presentations of wires and switches for their potential use in future molecular electronic devices.

Information processing with chemical or biochemical structures usually exhibits an additional handicap. Many of the switches and storage media developed so far operate only at very low temperatures. Although this is not a major problem for a molecular computer, it will be a problem for on-chip processing of chemical sensor information because the sensors operate reversibly only at moderate or even high temperatures.

Because of the inherent materials problems in general, and of the interface stabilities in particular, an alternative development in this field is aimed at devices for integrated optics and optoelectronics. One advantage is that no electrical contacts have to be realized to write and read the information. In the field of nonlinear optics, for example, new functionalized multilayers show promising properties and make possible a simple technology by self-assembly to synthesize chemically and assemble layer structures (Figure 10-32).

An alternative approach makes use of mechanical rather than electronic properties of specifically designed molecules [43]. This also has the advantage that electrical contacts can be avoided and that room temperature operation is in principle possible. However, up to now only a few of the proposed molecular structures could be synthesized, but current research on fullerenes and related compounds with onion-like structures will certainly have a great impact.

Most workers agree that some kind of self-repair mechanism may have to be implemented in such molecular electronics in order to overcome the problem of the short-term stability of the systems. Other concepts are aimed at using neural nets (see Section 10.4) in which some redundancy of information storage is built in, as in the brain. These concepts are considered in the field of bioelectronics.

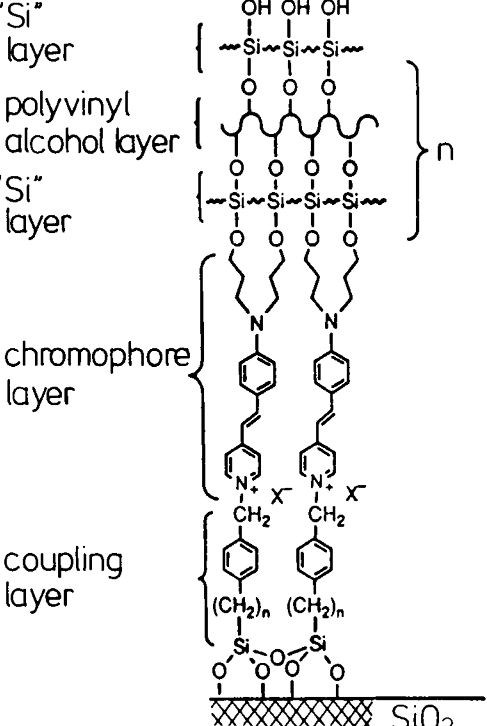


Figure 10-32.
Functionalized multilayer for an aligned charged chromophor which may be used in nonlinear optics devices [7].

10.6 Bioelectronics

Bioelectronics aims at the direct coupling of biomolecular function units of high molecular weight and extremely complicated molecular structure with electronic or optical transducer devices. Examples of biomolecular function units include whole cells, tissues, and cell arrays, which have already been mentioned above. Another example is the photosynthetic reaction center shown schematically in Figure 10-33. The design problems which have been solved in a perfect way in nature become evident from the parameters indicated in this figure. Gradients in the electrochemical potential of the order of 1 eV occur across distances on the order of 1 nm and lead to electron-transfer rates between the different functional groups of the order of 1 ps. In addition, the natural compound exhibits self-repair functions.

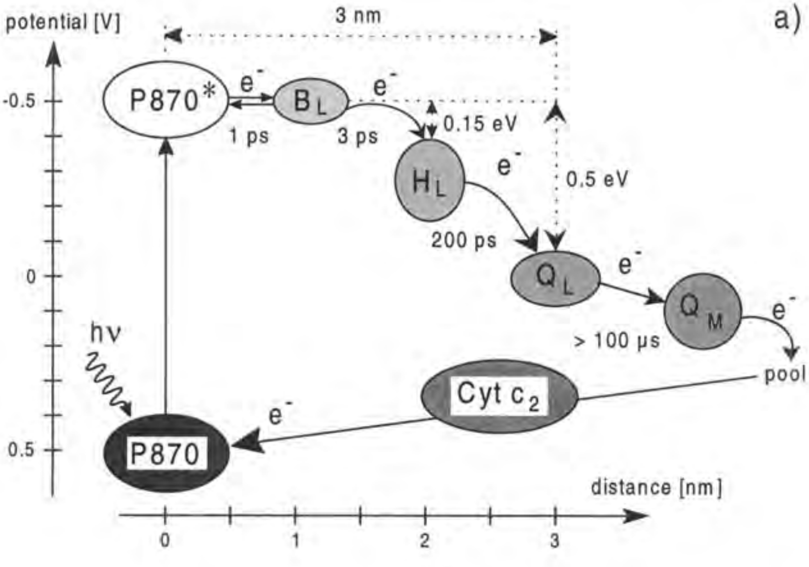
Such performances are far beyond any performance of a man-made device. As a result, significant efforts are currently being made to utilize such biomolecular function units and couple their signals into conventional electronic devices.

One example of recent trends is the signal transduction from nerve bundles to obtain characteristic signal patterns, which in subsequent pattern recognition approaches may be used to identify the information transferred across a nerve bundle. Materials problems in interfacing nerve cells with electrodes, in understanding the axon-axon interactions in the nerve bundles and the pulse code are completely unsolved today.

However, as mentioned in the Introduction, the underlaying driving force for R&D in this field is directed towards interfacing human senses including the brain with man-made electronics. The particular example mentioned in the Introduction was the nose: if *immediate success* is required to monitor certain molecules in the environment with a reliable instrument, the straightforward engineering approach must be taken, as is illustrated schematically in Figure 10-4. We have to use those components of a complete system which are available today and which contain as the most important part a suitable (bio-)chemically active coating of a carefully chosen transducer. Such coatings are usually thermodynamically stable materials of low molecular weight. A huge variety of signal evaluation procedures and pattern recognition methods are available to evaluate data from sensor arrays for the determination of odors or molecular compositions. The first few results in analyzing mixtures of molecules or odors with sensor arrays look promising (see Section 10.4). In contrast to this first-order approach, which represents a predominantly "technical" solution, alternative long-term R&D efforts are aimed at understanding the "real nose" or, more generally, the biological recognition and processing of chemical information. The latter is given by the chemical activities of certain species. "Biological systems" of this kind exist in a wide variety, ranging from the vertebrate nose to the chemotactic system of certain bacteria.

The tasks of signal generation, transduction, and data processing are performed in vertebrates by involving receptors, neurons, and cells in the brain. Unfortunately, only a few aspects of these "biological systems" are understood at present. These aspects concern specific receptors and part of their signal transduction steps between receptor sites in the nose and the brain.

In bacteria, the complete system is understood better with respect to all components of the sensor-actuator system, including integrated information processing and, to some extent, even information storage. The bacterial system is characterized in many aspects down to the molecular level. For their use in bioelectronic devices, however, function units of higher



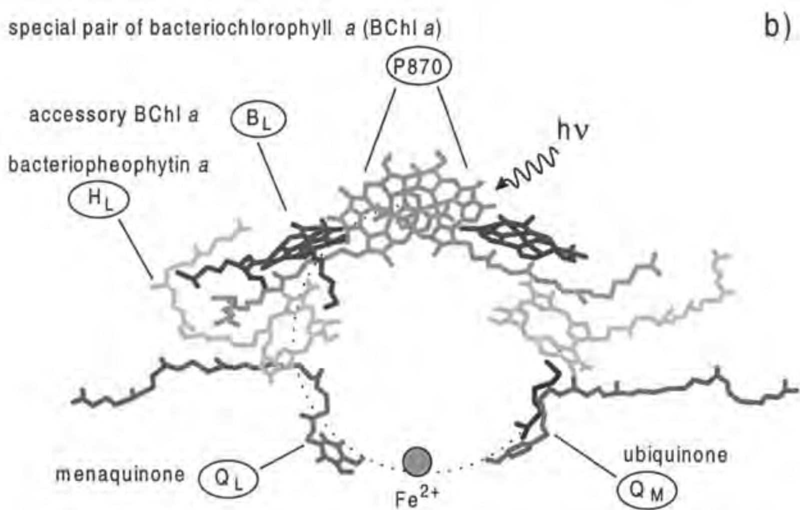


Figure 10-33. (a) Energy scheme of the electron transport chain in the bacterial photosynthetic reaction center. The photo-excited electron is removed from P870* via the bacteriochlorophyll B_L and the bacteriopheophytin H_L to the menaquinone Q_L where it is stored for about 100 µs before being transferred to Q_M and carried to the pool. Since B_L^- has the same energy level as P870*, electron oscillation occurs. The spatial distance between P870* and Q_L^- must be large enough to avoid recombination by electron tunneling (3 nm), and the energy level of Q_L^- must be sufficiently lower than that of P870* in order to avoid recombination by thermal activation (0.5 eV). (b) Scheme of the optimum spatial arrangement of the components of the reaction center. The dotted line indicates the pathway of the electron. For details, see Ref. [4] and references cited therein.

organisms seem more promising, as they operate less cooperatively and hence specialized parts of the complete biological system may be isolated and handled separately by interfacing them with technical devices. Most probably, the technical use of such biological systems will start by utilizing only certain biological units in "hybrid devices". A possible first step is to interface well-known receptors with microelectronic or microoptic transducers. Another possible step is to deduce electrical signals at different levels in data evaluation from the different components of a "real nose" of a vertebrate in its natural environment, and subsequently to apply artificial pattern recognition approaches.

Mimicking or adapting functions used by bacteria to perform more complex tasks requires an investigation of membrane-confined closed systems ("minicells"). Their interfacing is more complex than that of the planar sandwich structures described above. A basic "minicell" should at least consist of an energy source (such as a light-driven proton pump) and one sensor or actuator as a transmembrane function unit.

10.7 Outlook

10.7.1 Synthesis and Patterning of New Materials

The ideas discussed in Sections 10.3–10.6 and illustrated in Figure 10-2 point out clearly that the development of materials for chemical and biochemical sensors as well as for molecular and bioelectronics overlaps significantly with the development of materials for miniaturized physical sensors. In all cases there is general interest in the elucidation of the physical limits to miniaturizing sensors that monitor photons, electrons, molecules, ions, forces, electric and magnetic fields, or pressures.

Only a few examples could be presented here to illustrate some typical trends in the development of materials for chemical sensing in particular and for future information technologies in general. In all cases, the interfaces between chemically sensitive materials and transducers must be controlled on the atomic scale.

Typical materials are listed in Table 10-5 with increasing degree of complexity. The materials are prepared by a huge variety of different synthetic routes of inorganic or organic chemistry or extracted from biological systems and then combined with their transducers with a huge variety of preparation technology and a particular emphasis on thin-film technology.

10.7.2 Compatibility with IC Technology

Because of the detailed basic scientific understanding and wide range of sophisticated technologies available in current integrated circuit technology based on Si, the most rapid progress is expected in those areas in which Si compatibility, ie, integration to existing device technology, is possible. As just one example, Figure 10-34 shows the integration of a capacitive recognition sensor with data preprocessing and amplification of one chip. Similarly, other physical and chemical sensors may be integrated in devices for future information technologies.

Table 10-5. Typical materials used in chemical sensor research and development. The order in this table is determined by increasing complexity of the geometric structure [1].

Metals	Pt, Pd, Ni, Ag, Au, Sb, Rh, ...
Semiconductors	Si, GaAs, InP, ...
Ionic compounds	Electronic conductors (SnO_2, TiO_2, Ta_2O_5, IrO_x, ...)
	Mixed conductors ($SrTiO_3$, $La_{1-x}Sr_xCo_{1-y}Ni_yO_3$ or more general: perovskites, Ga_2O_3 ...)
	Ionic conductors (ZrO_2, LaF_3, CeO_2, CaF_2, Na_2CO_3, β-alumina, nasicon, ...)
Molecular crystals	Phthalocyanines (PbPc, $LuPc_2$, LiPc, $(PcAIF)_n$, $(PcGaF)_n$, ...)
Langmuir-Blodgett films	Phthalocyanines, polydiacetylenes, Cd-arachidate, ...
Cage compounds	Zeolites, calixarenes, cyclodextrins, crown ethers, cyclophanes, ...
Polymers	Polyethers, polyurethanes, polysiloxanes, polpyrroles, polythiophenes, polyfluorocarbons, polyolefines, PTFE, Nafion
Components of biomolecular functional systems	*Synthetic:* phospholipids, HMD- and AIDS-virus-epitopes, ...
	Natural: glucose-oxidase, lactose-permease, bacterial cellulose, E. coli-cell membranes, more general: enzymes, receptors, transport-proteins, membranes, cells, ...

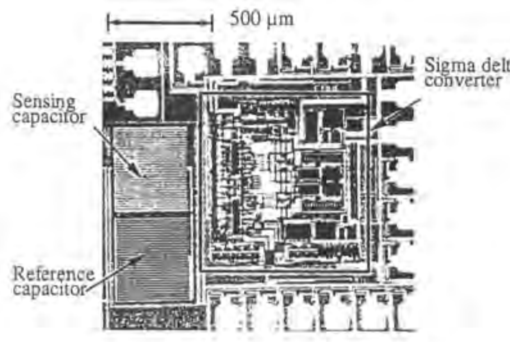


Figure 10-34.
Example of a capacitive chemical sensor in CMOS technology with polymer coating, integrated signal processing, and amplification by means of a capacitance/frequency converter [44].

10.7.3 Systems

The ultimate goal in developing nanostructured components for information technology is their integration in a system that performs information sensing, processing, storage, and actuation. A comparison of some existing and possible future (macroscopic and molecular) components is given in Table 10-6 just to stimulate the reader by outlining the drastic differences in materials and their technologies.

Possible first applicatons of such systems are expected for chemical microanalysis in different environments. Examples for chemical and biological sensor systems include the controlled release of glucose for diabetes patients from an implanted device or the controlled drug release in the brain of patients with chronic pain diseases. In both applications one might consider a telemetric operation or at least telemetric power supply from outside the human body first and in the next step an operation of the micro- or nanosystem which is based exclusively upon recognition and energy generation by body fluids.

Table 10-6. Comparison of macroscopic and molecular components in systems for general information processing.

Technology	Function	Molecular Examples
Struts, beams, casings	Transmit force, hold positions	Cell walls, microtubules
Cables	Transmit tension	Collagen, silk
Fasteners, glue	Connect parts	Intermolecular forces
Solenoids, actuators	Move things	Muscle actin, myosin
Motors	Turn shafts	Flagellar motor
Drive shafts	Transmit torque	Bacterial flagella
Bearings	Support moving parts	Single bonds
Clamps	Hold workpieces	Enzymatic binding sites
Tools	Modify workpieces	Enzymes, reactive molecules
Production lines	Construct devices	Enzyme systems, ribosomes
Numerical control systems	Store and read programs	Genetic system

Adapted from K.E. Drexler, Proceedings of the National Academy of Sciences, Vol. 78 (1981) pp. 5275–78.

10.8 References

[1] Göpel, W., Plenary Lecture, Eurosensors VII Conf., Budapest, September 1993; *Sensors Actuators B* **18–19** (1994) 1.
[2] Göpel, W., Plenary Lecture, Eurosensors IV Conf., Karlsruhe; *Sensors Actuators B* **4** (1991) 201.
[3] Göpel, W., presented at Conf. NATO ARW, Cambridge, April 1994 and Kluwer Adademic Press, in press.
[4] Göpel, W., Heiduschka, P., presented at 2nd European Workshop on Bioelectronics, Frankfurt, November 1993; *Biosensors Bioelectron.* **9** (1994) iii.
[5] Göpel, W., presented at 2nd European Workshop on Bioelectronics, Frankfurt, November 1993; *Biosensors Bioelectron.* **10** (1995) 35.
[6] Göpel, W., Plenary Lecture, Biosensors 94, New Orleans, June 1994; *Biosensors Bioelectron.* in press.
[7] Göpel, W., Ziegler, Ch. (eds.), *Nanostructures Based on Molecular Materials;* Weinheim: VCH, 1992.
[8] Göpel, W., Hesse, J., Zemel, J. N. (eds), *Sensors: A Comprehensive Survey,* Vols. 1–8, and in particular Vols. 2 and 3, Chemical and Biochemical Sensors; Weinheim: VCH, 1991 and 1992.
[9] Henzler, M., Göpel, W., *Oberflächenphysik;* Stuttgart: Teubner, 1991.
[10] Göpel, W., *State and Perspectives of Research on Surfaces and Interfaces, Report for DG XII;* Luxemburg; Commission of the European Community, Report EUR 13108 EN, 1990.
[11] Carter, F. L. (ed.), *Molecular Electronic Devices;* New York: Dekker, 1982.
[12] Wiesendanger, R., *Scanning Probe Microscopy and Spectroscopy;* Cambridge: Cambridge University Press, 1994;
Wiesendanger, R., Güntherodt, H.-J. (eds.), *Scanning Tunnelling Microscopy I,* Springer Series in Surface Sciences 20; Berlin: Springer, 1992;
Wiesendanger, R., Güntherodt, H.-J. (eds.), *Scanning Tunnelling Microscopy II,* Springer Series in Surface Sciences 28; Berlin: Springer, 1992;
Wiesendanger, R., Günterodt, H.-J. (eds.), *Scanning Tunnelling Microscopy III,* Springer Series in Surface Sciences 29; Berlin: Springer, 1993;
Bonnell, D. (ed.), *Scanning Tunnelling Microscopy and Spectroscopy;* Weinheim: VCH, 1993;
Chen, J., *Introduction to Scanning Tunnelling Microscopy;* Oxford: Oxford University Press, 1993.
[13] Chen, C. J., *Phys. Rev. B* **42** (1990) 8841;
Chen, C. J., *Scanning Microsc.* **7** (1993) 793.
[14] Hansma, P. K., Tersoff, J., *J. Appl. Phys.* **61** (1987) R1;
Tersoff, J., Hamann, D., *Phys. Rev. Lett.* **50** (1983) 1998;
Tersoff, J., Hamann, D., *Phys. Rev. B* **31** (1985) 805.

[15] Lang, N. D., *Comments Cond. Mater. Phys.* **14** (1989) 253.
[16] Sarid, D., *Scanning Force Microscopy,* Oxford Series in Optical and Imaging Sciences; Oxford: Oxford University Press, 1991.
[17] Feenstra, R. M., Stroscio, J. A., Tersoff, J., Fein, A. P., *Phys. Rev. Lett.* **58** (1987) 1192.
[18] Wolkow, R., Avouris, P., *Phys. Rev. Lett.* **60** (1988) 1049.
[19] Burnham, N. A., Dominguez, D. D., Mowery, R. L, Colton, R. J., *Phys. Rev. Lett.* **64** (1990) 1931.
[20] Florin, E. L., Rief, M., Ludwig, M., Dornmaier, C., Moy, V. T., Gaub, H. E., presented at Conf. Artificial Biosensing Interfaces, Athens 1994; *Biosensors Bioelectron.* in press.
[21] Ozin, G. A., *Adv. Mater.* **4** (1992) 612.
[22] Israelachvili, J., *Intermolecular and Surface Forces,* edn.; London: Academic Press, 1992.
[23] Göpel, W., Plenary Lecture, Eurosensors VIII, and *Sensors Actuators,* in press.
[24] Schierbaum, K.-D., Geiger, J., Weimar, U., Göpel, W., presented at 4th Int. Meeting on Chemical Sensors, Tokyo, September 1992; *Sensors Actuators B* **13-14** (1993) 143.
[25] Göpel, W., *Sensors Actuators* **16** (1989) 167.
[26] Göpel, W., *Prog. Surf. Sci.* **20** (1985) 9.
[27] Schierbaum, K.-D., Wei-Xing, X., Göpel, W., presented at Conf. In-Situ Investigations of Physicochemical Processes at Interfaces, Bunsen Discussion Meeting, Lahnstein, October 1992; *Ber. Bunsenges. Phys. Chem.* **97** (1993) 363;
Schierbaum, K.-D., Göpel, W., in: Freund, H.-J., Umbach, E. (eds.); *Adsorption on Ordered Surfaces of Ionic Solids and Thin Films,* Berlin: Springer, 1994, p. 268.
[28] Göpel, W., in: *Gas Sensors,* Sberveglieri, G. (ed.); Dordrecht: Kluwer, 1992, p. 365.
[29] Wiemhöfer, H.-D., Vohrer, U., Göpel, W., in: *Systems with Fast Ionic Transport,* Schuster, G., Künstler, K., Ullmann, H. (eds.), Materials Science Forum, Vol. 76; Zürich: Trans. Tech. Publications, 1991, pp. 265-268;
Wiemhöfer, H.-D., *Ber. Bunsenges. Phys. Chem.* **97** (1993) 461.
[30] Dominik, A., Roth, H. J., Schierbaum, K. D., Göpel, W., *Supramol. Sci.* **1** (1994) 11.
[31] Göpel, W., Plenary lecture, 5th Int. Meeting on Chemical Sensors, Rome, July 1994; *Sensors Actuators* in press.
[32] Schierbaum, K. D., Weiss, T., van Velzen, U. Th., Reinhoudt, B. N., Göpel, W., *Science* **265** (1994) 1413.
[33] Rickert, J., Heiduschka, P., Beck, W., Jung, G., Göpel, W., presented at 3rd World Congress on Biosensors, New Orleans, 1994; *Biosensors Bioelectron.* in press.
[34] Schuhmann, W., Wohlschläger, H., Lammert, R., Schmidt, H.-L., Löffler, U., Wiemhöfer, H.-D., Göpel, W., presented at Eurosensors III, Montreux; *Sensors Actuators B* **1** (1990) 571-575;
Löffler, U., Wiemhöfer, H.-D., Göpel, W., presented at Biosensors '90, Singapore; *Biosensors Bioelectron.* **6** (1991) 343-352.
[35] Ottenbacher, D., Jähnig, F., Göpel, W., presented at 4th Int. Meeting on Chemical Sensors, Tokyo, September 1992; *Sensors Actuators B* **13-14** (1993) 173-175.
[36] Mohr, A., *PhD Thesis,* University of Tübingen, 1995, and A. Mohr, W. Nisch and W. Göpel, Conf. Proc. Transducers 1995, Stockholm.
[37] Schierbaum, K.-D., Haug, M., Nahm, W., Gauglitz, G., Göpel, W., presented at 1st Eur. Conf. Opt. Chem. Sensors and Biosensors, Graz, April 1992; *Sensors and Actuators B* **11** (1993) 383.
[38] Schierbaum, K.-D., Göpel, W., presented at E-MRS Spring Meeting, 1993, Strasburg; *Synth. Met.* **61** (1993) 37.
[39] Hierlemann, A., Weimar, U., Kraus, G., Schweizer-Berberich, M., Göpel, W., presented at Eurosensors VIII, Toulouse; *Sensors Actuators* in press.
[40] Hierlemann, A., Weimar, U., Kraus, G., Gauglitz, G., Göpel, W., *Sensors Mater.* **7** (No 3) (1995), 179.
[41] Aviram, A., *Adv. Mater.* **1** (1989) 124, and references cited therein.
[42] Sienicki, K., *Molecular Electronics and Molecular Electronic Devices;* Boca Raton, FL: CRC Press, 1993.
[43] Drexler, K. E., *Nanosystems;* New York: Wiley, 1992.
[44] Cornila, C., Lenggenhager, R., Azerredo Leme, C., Malcovati, P., Baltes, H., Hierlemann, A., Noetzel, G., Weimar, U., Göpel, W.. presented at 5th Int. Meeting on Chemical Sensors, Rome, 1994; *Sensors Actuators* in press.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

11 Future Nanosensors

R. WIESENDANGER, Institute of Applied Physics and Center for Microstructure Research, University of Hamburg, Germany

Contents

11.1 Introduction . . . 338

11.2 Nanometer and Picometer Displacement Sensors . . . 339
11.2.1 General Principles . . . 339
11.2.2 Applications . . . 340

11.3 Force Nanosensors . . . 343
11.3.1 General Principles . . . 343
11.3.2 Detection Methods . . . 344

11.4 Thermal Nansensors . . . 346
11.4.1 Heat-Flux Nanosensor . . . 346
11.4.2 Micro-Thermocouple Sensor . . . 348
11.4.3 Tunnelling Thermometer . . . 348

11.5 Optical Nanosensors . . . 349
11.5.1 Evanescent Field Nanosensors . . . 349
11.5.2 Optical Sensors Based on Nanometer-Sized Apertures . . . 349
11.5.3 Fiber-Optic Nanosensors . . . 349

11.6 Magnetic Nanosensors . . . 350
11.6.1 Submicron Hall Probes . . . 350
11.6.2 Micro-SQUID Sensors . . . 351
11.6.3 Magnetic Stray Field Nanosensors . . . 352
11.6.4 Magnetic Spin Sensors . . . 353

11.7 Conclusions and Outlook . . . 353

11.8 Acknowledgments . . . 355

11.9 References . . . 355

11.1 Introduction

The expression "nanosensor" is not commonly used and therefore we first have to define its meaning. In the most general way, a nanosensor may be defined as a sensor fulfilling at least one of the three following requirements:

1. The sensitivity of the sensor is on the nano-scale (eg, sensors for displacements on the nanometer scale, force sensors with sensitivity of the order of nanonewtons, sensors for currents of the order of nanoampères, sensors for magnetic fields of the order of nanotesla, and power sensors with sensitivity of the order of nanowatts).
2. The spatial confinement of the interaction of the sensor with the object is on the nanometer scale. Combined with scanning technology this directly leads to a spatially resolved sensor technology with a spatial resolution on the nanometer length scale.
3. The size of the sensor is on the nanometer scale.

The trend in sensor technology towards the nano-scale has been pushed forward by progress in the field of silicon-based microfabrication [1] and by recent developments in the field of scanning probe technology based on the invention of the scanning tunneling microscope by Binnig and Rohrer [2, 3]. A scanning probe instrument (Figure 11-1) can be regarded as a

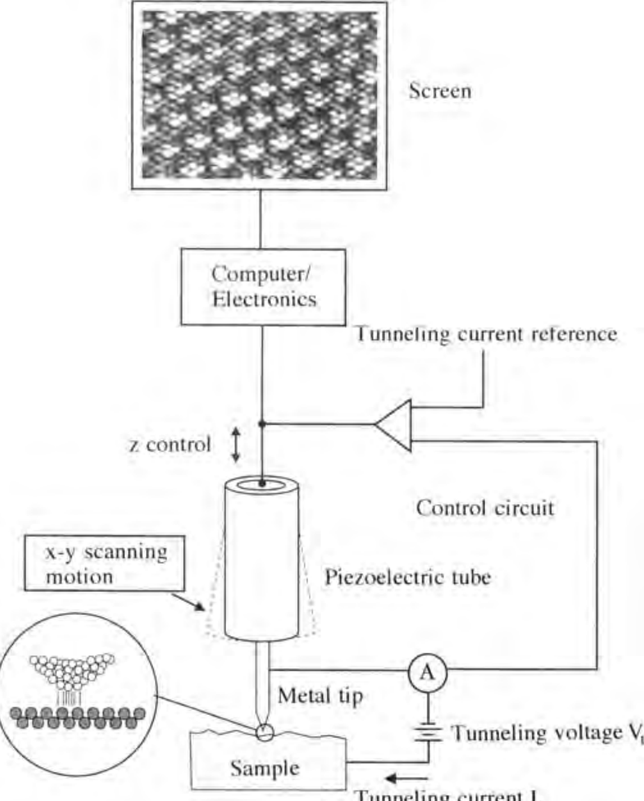


Figure 11-1. General concept of a scanning probe instrument as a microsystem consisting of a nanosensor, a signal processing system and a piezoelectric actuator.

microsystem consisting of a nanosensor (tip or microfabricated sensor with integrated tip), a signal processing system (signal amplifiers, feedback system, high-voltage amplifiers, etc.), and actuators consisting of piezoelectric drives for positioning of the nanosensor in three orthogonal directions [4, 5]. The sensitivity of the nanosensor in a scanning probe instrument is at least on the nano-scale or even beyond, as discussed below. The spatial confinement of the interaction of the nanosensor with its environment is achieved by the tip shape of the sensor combined with the distance-dependent decay of the interaction strength. By combination with silicon-based microfabrication techniques, the size of the sensor of a scanning probe instrument can be reduced to the micrometer length scale but may be miniaturized even further by applying nanofabrication methods [6].

In the first 10 years since its invention, scanning tunnelling microscopy and related scanning probe techniques have mainly been applied as imaging, measuring, and modification tools [7-9]. However, their impact on sensor technology is becoming more and more evident and an increasing number of nanosensors derived from scanning probe technology are currently at the development stage.

11.2 Nanometer and Picometer Displacement Sensors

11.2.1 General Principles

Scanning tunnelling microscopy (STM) is based on the detection of an electron tunnelling current I flowing between a sharp, electrically conducting probe tip and an electrically conducting object held at a distance s of only a few angstroms (1 Å = 0,1 nm = 10^{-10} m) with an applied bias voltage U. The strong dependence of the tunneling current I on the gap spacing s provides the key for the atomic-resolution capability of STM:

$$I(s) \propto \exp(-2\kappa s) \tag{11-1}$$

where κ is a decay rate. On the other hand, Equation (11-1) also provides the basis for nano- and picometer displacement sensors based on electron tunnelling detection because the current varies as much as an order of magnitude for each angstrom unit change in electrode separation. Displacement sensitivity of the order of 0.01 nm is nowadays routinely achieved in STM and can even be improved down to the 1 pm level [5].

Another more recently developed scanning probe instrument, the atomic force microscope (AFM) [10], extends the capabilities of STM to electrically insulating objects and sensor tips. The AFM is based on the detection of the repulsive forces F acting between a sharp probe tip in contact with an object. Since these contact forces exhibit a strong distance dependence similarly to the tunneling current in STM, atomic resolution can be achieved with an AFM as well. On the other hand, the AFM can be used as a nano- or picometer displacement sensor even with electrically insulating sensor parts.

Generally, all scanning probe instruments developed so far [6, 7-9] could be used as displacement sensors. However, the sensitivity which can be achieved depends critically on the

distance dependence of the interaction detected in a particular type of scanning probe instrument. STM and AFM are outstanding for their exponential distance dependence of the interaction to be probed and therefore easily achieve subnanometer displacement sensitivity. The major advantage of using STM or AFM as displacement sensors compared with, for instance, optical, capacitive, or piezoresistive sensors is that the interaction of the sensor with the object in STM or AFM can be spatially confined to an extremely small scale, ultimately down to the atomic level. Consequently, the objects to be probed can be arbitrarily small or, if they are of macroscopic size, arbitrarily small parts of the objects can be probed.

11.2.2 Applications

Physical sensors which measure quantities such as position, acceleration, pressure, temperature, or radiation usually rely on the measurement of a change in separation between two components. Therefore, highly sensitive displacement measurement concepts can be applied directly in sensor technology. In the following, various prototypes of physical sensors based on tunnelling detection will be described. Their actual size is determined by the capability of silicon microfabrication technology rather than by the sensor concept.

Figure 11-2 shows a schematic drawing of a prototype accelerometer based on electron tunneling detection [11]. In contrast to conventional tunneling devices, the relative position of the electrodes is controlled by electrostatic forces acting between two electrodes on application of a deflection voltage, rather than by piezoelectric drives. By using electrostatic actuators, insensitivity to thermal drift and piezoelectric creep can be achieved. The silicon sensor tip is integrated in a spring-type electrode and the active surfaces of all electrodes are coated by 300 nm thick gold films (black lines in Figure 11-2). The application of a deflection voltage leads to an attractive force between the two electrodes which deflects the spring. At close proximity (a few angstroms) between tip and tunnel counterelectrode, a tunnel current on the order of nanoampères is established which is kept constant by means of an electronic feedback circuit similarly as for STM. The displacement responsivity is typically of the order of 1 nA/Å at an average current of 1 nA, while the sensitivity to variations in the tip-counterelectrode separation was found to be $2 \cdot 10^{-4}$ Å/Hz$^{1/2}$ at a frequency of 1 kHz [11]. The responsivity of the sensor to acceleration was calculated to be $1 \cdot 10^{-7}$ g/Hz$^{1/2}$ at 10 Hz, being limited by

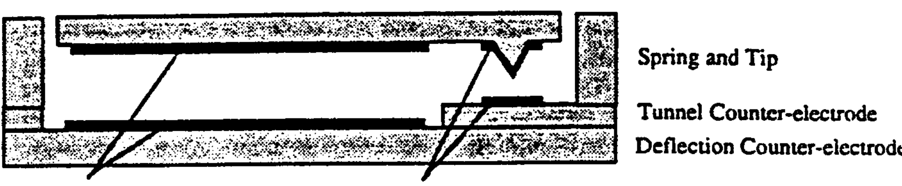


Figure 11-2. Conceptual drawing of a prototype tunnelling accelerometer. The upper plate is suspended by a micromachined spring and is deflected downwards by an electrostatic force. When the tip is within 1 nm of the tunnel counter electrode, a tunnelling current is detected. A feedback loop controls the deflection voltage so as to keep the upper plate in the same relative position. During acceleration of the sensor, the feedback force balances the inertial force. The deflection voltages are recorded as signal (from [11]).

noise originating from accelerations in the laboratory. This sensitivity is several orders of magnitude better than with conventional miniature accelerometers, whereas conventional accelerometers with comparable sensitivity are several orders of magnitude larger and heavier. Further improvements of the tunnelling accelerometer may lead to a sensitivity of 10^{-8} g/Hz$^{1/2}$ over a band width from 5 Hz to 1 kHz [11].

Another prototype of physical sensor based on tunnelling detection is the infrared sensor as shown in Figure (11-3) [12]. This device consists of a small cavity filled with gas formed by a pair of silicon wafers, one of which has been etched through to a 0.5 μm thick SiO_xN_y membrane. The outer surface of the membrane is coated with gold to serve as a counter-electrode for tunnelling detection. The device is operated by applying a voltage between the deflection electrodes and the membrane leading to an attractive electrostratic force which pulls the membrane to within tunneling distance of the tip. An electronic feedback circuit as used in STM controls the voltage applied to the deflection electrodes so as to keep the tunnelling current constant. On absorption of infrared radiation, the pressure in the trapped gas changes, affecting the force applied to the membrane by the gas. The electronic feedback circuit responds by adjusting the deflection voltage to create a rebalance force so as to keep the membrane in the same position and therefore the tunneling current at the preset value. The deflection voltage signal is therefore a direct measure for the amount of infrared absorption. The noise equivalent power of the prototype infrared sensor was measured to be $8 \cdot 10^{-10}$ W/Hz$^{1/2}$, which is already competitive with the best commercial pyroelectric sensors [12].

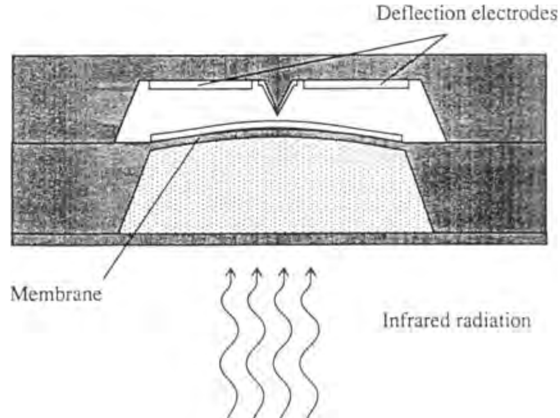


Figure 11-3.
Diagram of an infrared sensor based on tunnelling detection. An electrostatic force pulls the membrane into contact with the tip. With the absorption of infrared radiation, the pressure in the trapped gas changes. The feedback circuit responds by modifying the deflection voltage (from [12]).

Prototypes of magnetic sensors based on tunnelling detection have also been proposed [13, 14]. Figure 11-4 shows a set-up of a magnetic field sensor which is based on measuring dimensional changes of an amorphous magnetostrictive ribbon exposed to dc and ac magnetic fields by means of a tunnelling tip. A response of the sensor to ac magnetic fields down to 60 μOe was obtained with a ribbon response of 1.7 μm/Oe (at a bias field of 0.85 Oe), which produced a signal response of 30 V/Oe [14].

A proposed design of a vibrating reed magnetometer based on tunnelling detection is shown in Figure 11-5 [15]. An alternating field gradient produces a periodic force on a magnetized sample at the end of a cantilever beam, leading to an alternating bending moment of the can-

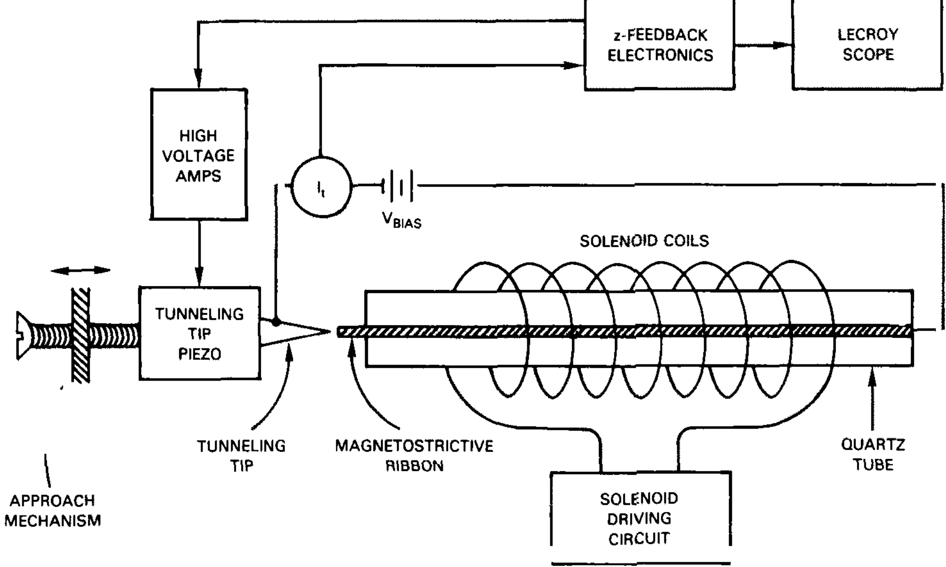


Figure 11-4. Set-up of a magnetic field sensor consisting of a magnetostrictive ribbon, solenoid coils, and a probe tip. The ribbon is 10 cm long (from [14]).

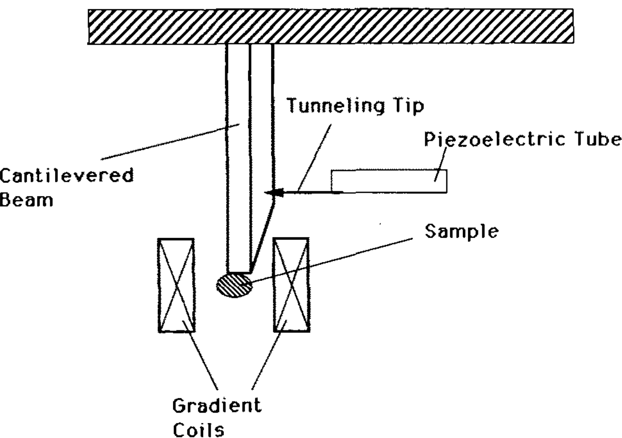


Figure 11-5. Scheme of a magnetometer design based on tunnelling detection (from [15]).

tilever. The amplitude of vibration is proportional to the driving force and hence to the magnetization of the sample. The sensitivity of measuring the vibrational amplitude therefore determines the overall sensitivity of the vibrating reed magnetometer.

Another interesting application of displacement measurements by electron tunnelling can be gravitational wave sensors based on the transformation of the motion from a large antenna to a much smaller mass [16, 17]. Tunnelling appears to be highly promising for this type of sensor because it allows one to measure the displacements of extremely small particles.

Electron tunnelling has also been proposed for the acoustic detection of particles [18]. The corresponding sensor is based on the acoustic signal induced in a solid body in which the energy of incoming particles is released and converted into a localized and adiabatic heating. The resulting amplitude of displacement of the target can then be detected by a tunnelling tip.

Many further possible applications of nano- and picometer displacement sensors based on electron tunnelling detection exist. In all cases the advantages of using tunnelling detection are given by the high sensitivity which can be achieved, in addition to the small volume and mass which can be probed. On the other hand, possible instabilities of the tunnelling current signal, eg, due to the migration of atoms in the tunnelling area, often limit the performance of the sensors. Avoiding such instabilities by choosing appropriate tunnelling electrodes and an appropriate environment constitutes the major challenge for the future development of tunnelling-based nanosensors.

11.3 Force Nanosensors

11.3.1 General Principles

Forces are usually detected by their action on a cantilever beam or a thin membrane which becomes deflected. The magnitude of the deflection, Δz, is proportional to the force F within the elastic regime (Hooke's law):

$$F = c\,(\Delta z) \tag{11-2}$$

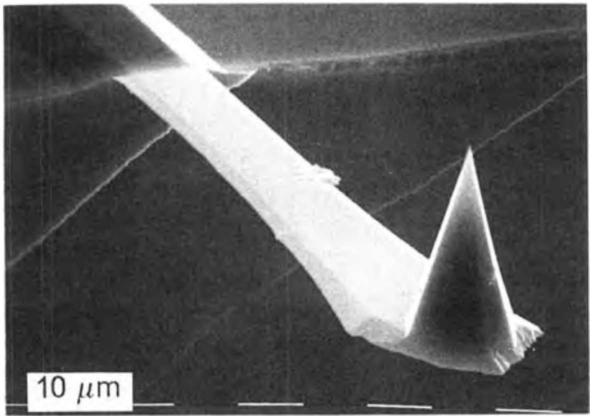
where c denotes the spring constant. The deflection Δz can be determined by a displacement sensor. To be able to measure extremely small forces, highly sensitive displacement sensors and reasonably small spring constants have to be chosen.

On the other hand, insensitivity to vibration requires that the resonance frequency

$$\omega = (c/m)^{1/2} \tag{11-3}$$

where m is the mass, should be as high as possible. Both requirements can only be fulfilled simultaneously if m and therefore the geometrical dimensions of the force sensor are as small as possible. This consideration leads directly to the demand for using microfabrication technology.

By integrating a tip in the cantilever beam or membrane (Figure 11-6), forces between arbitrarily small objects can be probed. This idea is realized in the atomic force microscope (AFM) invented by Binnig et al. in 1986 [10]. A theoretical limit of the force sensitivity of the AFM as small as 10^{-18} N has been estimated [10]. Experimentally, forces of 10^{-9} N can routinely be measured in the contact regime and down to 10^{-12} N in the non-contact regime. Recently, a force sensitivity as small as 10^{-14} N was achieved with a force microscope operated at room temperature [20]. Further improvements are expected for operation of the AFM at low temperatures.



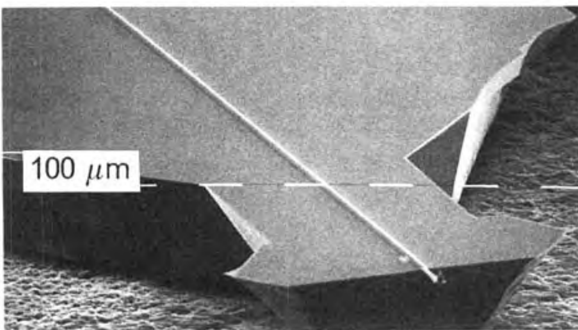


Figure 11-6.
Scanning electron micrograph of a microfabricated silicon force sensor consisting of a cantilever beam with integrated probe tip (from [19]).

11.3.2 Detection Methods

Different types of detection schemes for measuring the small cantilever defletions on a sub-angstrom scale can be applied. For instance, a displacement sensor based on tunnelling detection (Section 11.2.1) easily provides a sensitivity of typically 0.01 nm and has been used in the experimental set-up of the first AFM (Figure 11-7). To achieve stable tunnelling conditions, the rear side of the cantilever beam has to be coated with a thin gold film.

Alternatively, the cantilever displacement can be detected by the deflection of a laser beam which is reflected off the rear side of the cantilever (Figure 11-8). The direction of the reflected laser beam is measured by a position-sensitive detector, typically a bicell, which consists of two photoactive (eg, Si) segments (anodes) that are separated by about 10 μm and have a common cathode. Sub-angstrom sensitivity is routinely achieved with this method [21].

Another optical detection scheme is offered by laser interferometry. Several different types of interferometer systems have been developed to reach the required sensitivity, based on either homodyne or heterodyne detection methods. Advances in the homodyne detection method have been achieved by using a fiber-optic technique that places a reference reflector within a distance of micrometers from the cantilever beam [22]. The cantilever deflection measurement is then based on the optical interference occuring in the micrometer-sized cavity formed between the cleaved end of a single-mode optical fiber and the cantilever beam (Figure 11-9).

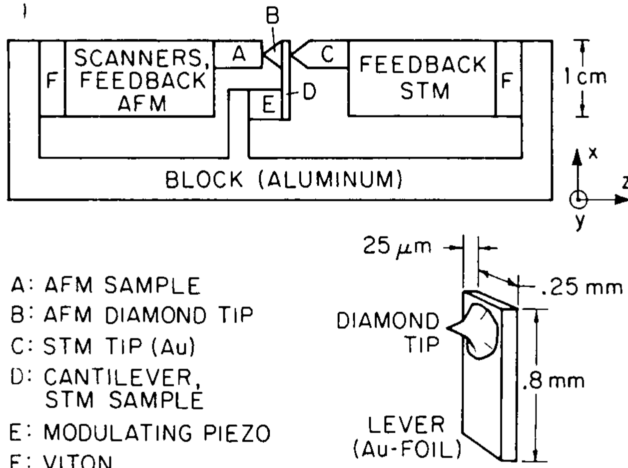


Figure 11-7. Schematic diagram of the experimental set-up for the first atomic force microscope (AFM). The cantilever beam with integrated tip is not to scale (from [10]).

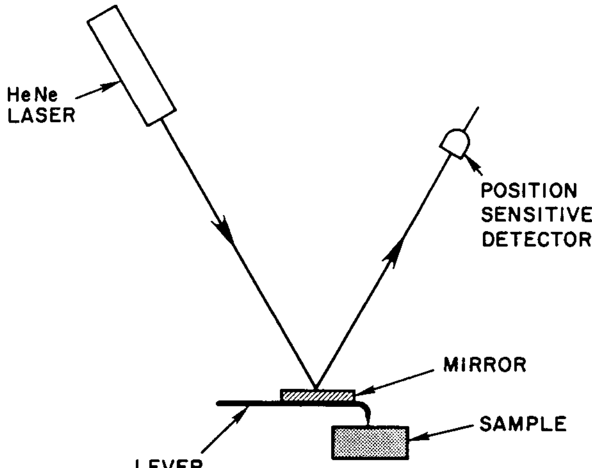


Figure 11-8.
Force detection scheme based on laser beam deflection (from [21]).

An alternative approach to measure the small cantilever deflections is based on measurements of the capacitance formed by a counter-electrode opposite the rear side of the cantilever beam. By using a capacitance transformer bridge for small capacitances, a displacement sensitivity at the sub-angstrom level can be achieved, which requires capacitance noise levels of the order of 10^{-18} F [23, 24]. Since the rate of change in capacitance with distance, and therefore the detection sensitivity, increase as the distance between the two electrodes decreases, it is desirable to operate with a small electrode separation. On the other hand, the attractive electrostatic force between two electrodes also becomes large at small distances. Therefore, the geometry must be chosen to achieve maximum sensitivity without causing the two electrodes to snap together, which would occur if the gradient of the attractive force exceeds the spring constant of the cantilever.

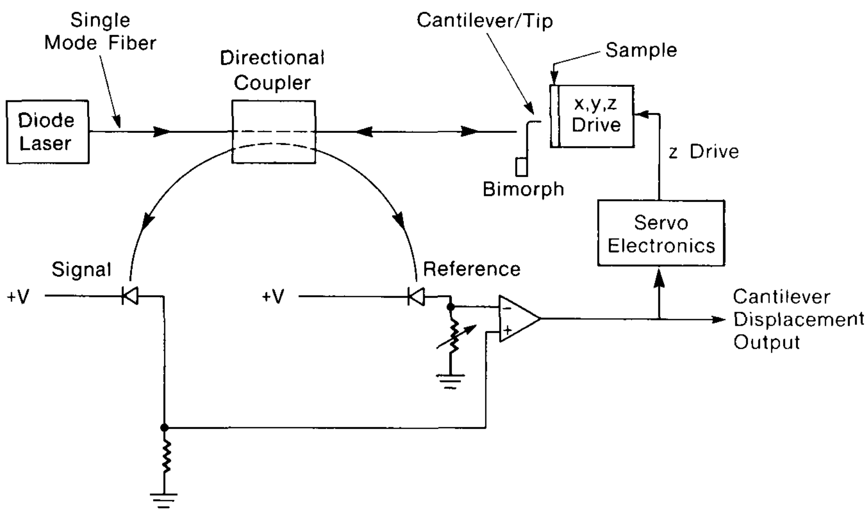


Figure 11-9. Schematic diagram of a force microscope based on a fiber-optic interferometer (from [22]).

More recently, piezoresistive detection of the cantilever deflections has been applied in AFM [25]. However, the sensitivity achieved cannot yet compete with optical or tunneling detection schemes.

In summary, recent developments in force measurements have led to a force sensitivity down to 10^{-14} N. The tip-based geometry as known from AFM in principle allows the force interaction between two objects to be spatially confined down to the atomic level. The operation of force sensors at low temperatures will lead to further improvements in force sensitivity by several orders of magnitude.

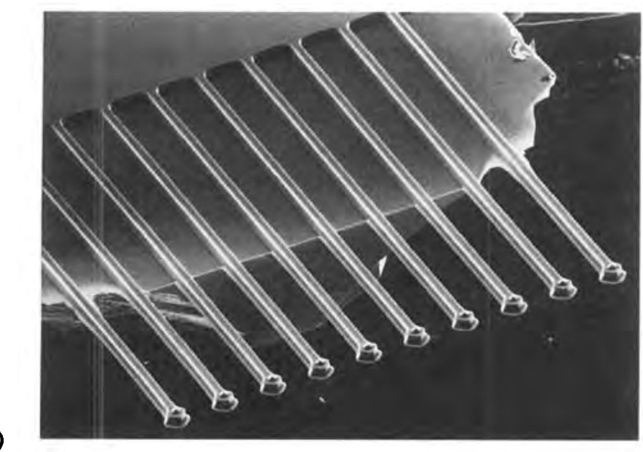
11.4 Thermal Nanosensors

Thermal nanosensors are expected to achieve nano-scale sensitivity in the detection of thermal properties and/or to allow the probing of thermal properties on a nanometer scale. In the following, three different examples of prototypes of thermal nanosensors are described which will indicate both the opportunities and challenges of future thermal nanosensor technology.

11.4.1 Heat-Flux Nanosensor

A heat-flux nanosensor can be based on a micromechanical silicon cantilever coated with a thick layer of aluminium and a thin, chemically reactive surface layer as shown in Fig-

ure 11-10 [26]. Heat fluxes are detected by measuring the cantilever deflection induced by the differential thermal expansion of the cantilever (bimetallic effect) using the laser beam deflection method described in Section 11.3.2 (Figure 11-8). The prototype device having an active area of only $1.4 \cdot 10^{-8}$ m^2 and a volume of $2 \cdot 10^{-14}$ m^3 achieved a sensitivity of 0.1 nm per absorbed nanowatt power. The limit of sensitivity to local temperature changes was found to be 10^{-5} K at room temperature. Using this device, the catalytic conversion of H_2 and O_2 to form H_2O over a thin platinum overlayer has been studied [26]. This example demonstrates nicely that silicon-based microfabrication techniques can lead to advances in both thermal and micromechanical sensor technology.


(a)

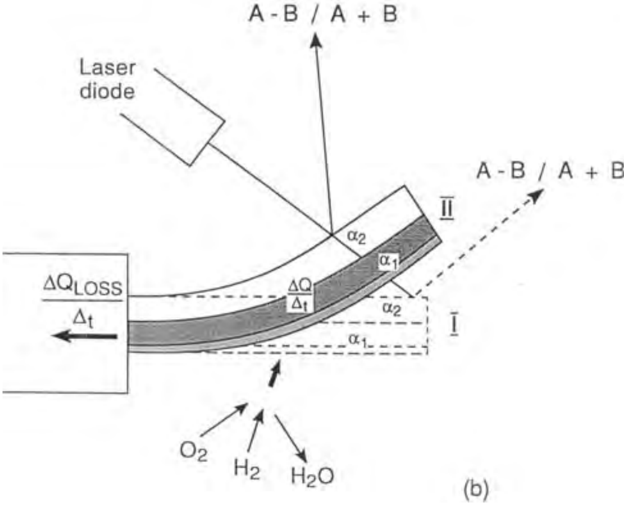


Figure 11-10. (a) Scanning electron micrograph of an array of ten microfabricated silicon levers, each protruding 0.4 mm beyond the edge of a silicon wafer. These microstructures can be used as heat flux nanosensors as explained in (b). (b) Schematic diagram of the device (from [26]).

11.4.2 Micro-Thermocouple Sensor

To probe thermal properties on the nanometer length scale, a miniaturized thermocouple sensor at the end of a sharp tip has been introduced as shown in Figure 11-11 [27, 28]. The thermal probe tips have dimensions of the order of 100 nm. The minimum detectable change in tip temperature was found to be less than 10^{-4} K, which corresponds to a heat flow of only a few nanowatts between the thermal sensor and the object being probed.

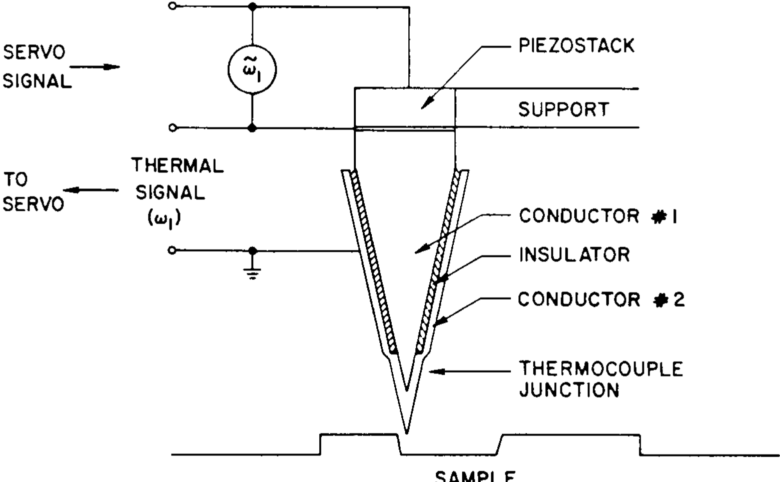


Figure 11-11. Schematic diagram of a thermocouple probe mounted on a piezoelectric drive for adjustment of the tip-to-object distance (from [27]).

11.4.3 Tunnelling Thermometer

As an alternative experimental set-up, a thermocouple can be formed by approaching an electrically conducting STM tip towards an electrically conducting object of a different material until a tunnelling current is detected. To measure the temperature, the electronic feedback circuit of the STM has to be switched off and the externally applied bias voltage has to be removed. The open-circuit potential developed across the tip and object then provides a measure of the local thermoelectric potential [29, 30]. Since the tunnelling current flow in STM is localized to atomic-scale dimensions, the tunnelling thermometer in principle allows one to probe thermal properties down to the atomic level.

11.5 Optical Nanosensors

11.5.1 Evanescent Field Nanosensors

Electromagnetic waves interacting with an object are always diffracted into two components:

1. propagating waves with low spatial frequencies ($<2/\lambda$), where λ is the optical wavelength; and
2. evanescent waves with high spatial frequencies ($>2/\lambda$).

Classical optics is concerned with the far-field regime where only the propagating waves survive, whereas the evanescent waves are confined to sub-wavelength distances from the object corresponding to the near-field regime. Information about the high spatial frequency components of the diffracted waves is lost in the far-field regime and therefore information about sub-wavelength features of the object cannot be retrieved. On the other hand, the recently developed scanning near-field optical microscope (SNOM) [31, 32], which is based on similar principles to those for STM, surpasses the diffraction limit by operating an optical probe at close proximity ($s < \lambda$) with respect to the object. Future optical nanosensors are also expected to operate in the near-field regime, where high sensitivity combined with spatial confinement of the optical interaction can be achieved. Optical sensing even down to the molecular level has already been demonstrated [33]. Additionally, optical displacement sensors operated in the near-field regime appear to be promising because the field intensity decreases strongly with a fourth power dependence on the distance in the near-field regime, whereas in the far-field regime the field intensity decreases only quadratically with the distance, corresponding to a constant flux per steradian.

11.5.2 Optical Sensors Based on Nanometer-Sized Apertures

The fabrication of nanometer-sized apertures is essential to confine spatially the optical interaction to small dimensions which, however, is gained at the expense of signal intensity. Optical nanoprobes can be obtained by a sharpened glass or fiber tip coated with a thin metallic layer, leaving a sub-micrometer aperture at the apex (Figure 11-12a). As an alternative to glass or fiber tips, sharpened micropipettes coated with a thin metallic layer can also serve as optical nanoprobes (Figure 11-12b). The micropipette apertures can either be illuminated from the rear side by a laser beam, acting as a light source, or they can be used as collectors for radiation from a small area of the sample which itself is illuminated as a whole. To enhance the transmission of electromagnetic energy through the narrow micropipette tube, it has been proposed to fill the micropipette with a fluorescent dye embedded in a plastic matrix (Figure 11-12c).

11.5.3 Fiber-Optic Nanosensors

The strong distance dependence of the evanescent field is the primary mechanism for having an effective sharp nanotip capable of optically probing nanometer-sized objects. Optical

nanoprobes can therefore alternatively be obtained just by chemically etching one end of a quartz optical fiber in hydrofluoric acid (Figure 11-12d). The resulting tip radii are typically of the order of 100 nm. However, although the use of submicrometer apertures is not necessarily required to confine spatially the near-field optical interaction to the sub-wavelength regime, it is certainly advantageous for further improvements with regard to optical probing of extremely small objects down to the molecular level.

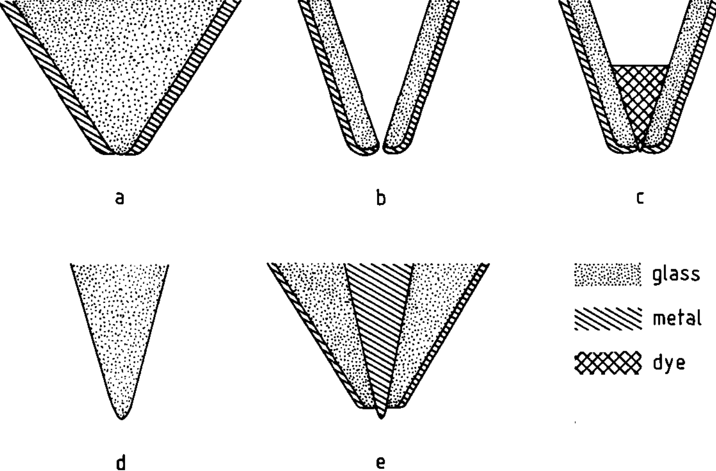


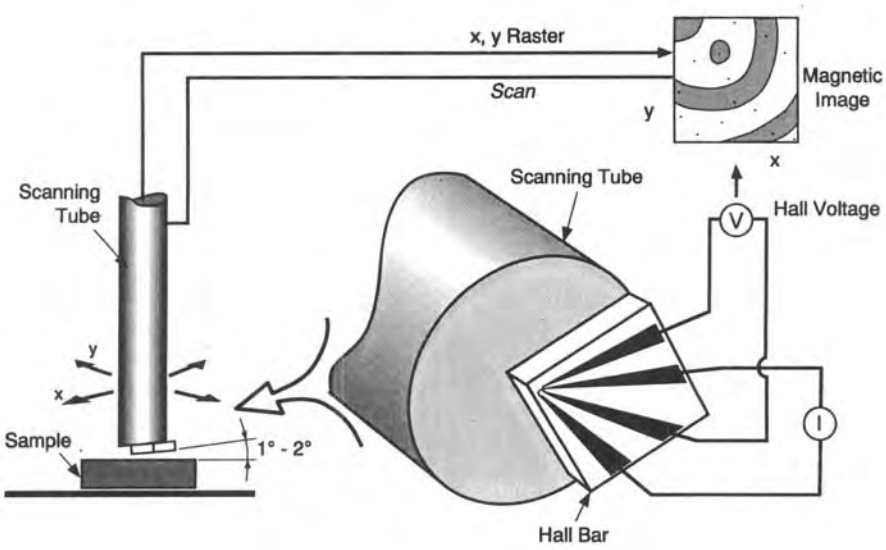
Figure 11-12. Schematic diagram of near-field optical nanosensors: (a) sub-micron aperture formed by a sharpened fiber or glass tip coated with a thin metal film; (b) sub-micron aperture formed by a micropipette coated with a thin metal film; (c) miniature fluorescent dye matrix in a micropipette; (d) sharpened optical fiber suitable as a leaky wave sensor; (e) miniature optical coaxial line (from [34]).

11.6 Magnetic Nanosensors

Several prototypes of magnetic nanosensors have been developed in recent years which are sensitive to different magnetic properties on a variety of length scales. In the following, some selected examples of magnetic nanosensors will be presented which allow one to probe magnetic properties from the submicron down to the atomic scale.

11.6.1 Submicron Hall Probes

Submicron Hall probes can be obtained by using conventional semiconductor microfabrication technology as shown in Figure 11-13. In this case, a Hall bar was patterned within 3.5 µm of one corner of a $GaAs/Al_{0.3}Ga_{0.7}As$ heterostructure. A magnetic field sensitivity of 0.1 G combined with a spatial resolution of 0.35 µm to detect magnetic stray fields at close proximity (0.2 µm) to an object has been achieved [35].



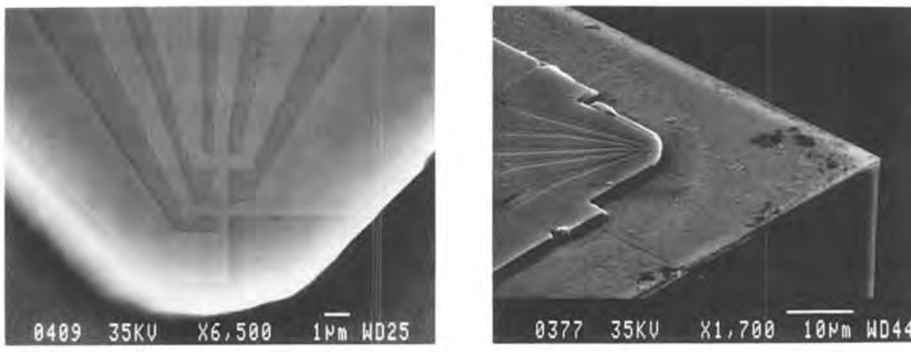


Figure 11-13. Top: Schematic diagram of the scanning Hall probe microscope. Bottom: Electron micrographs of the Hall bar showing the Hall junctions and the rounded corner which serves as the tunnelling tip (from [35]).

11.6.2 Micro-SQUID Sensors

Alternatively, a magnetic nanosensor can be based on a microfabricated dc SQUID detector as shown in Figure 11-14. A magnetic flux sensitivity of about 10^{-4} ϕ_0 (ϕ_0 = elementary flux quantum) and a spatial resolution of the order of 10 µm have been achieved [36]. By the application of nanofabriction methods [6], further miniaturization appears to be possible.

11.6.3 Magnetic Stray Field Nanosensors

Magnetic stray field nanosensors can be based on the magnetic force interaction between a miniaturized magnetic probe and the magnetic stray field of an object (Figure 11-15). The magnetic force microscope (MFM), developed in 1987 [37–39], makes use of this principle. The MFM allows one to probe the magnetic stray field down to the scale of 10 nm. Since magnetic forces as small as 10^{-12} N can routinely be detected by methods described in Section 11.3, a correspondingly high magnetic stray field sensitivity is achieved.

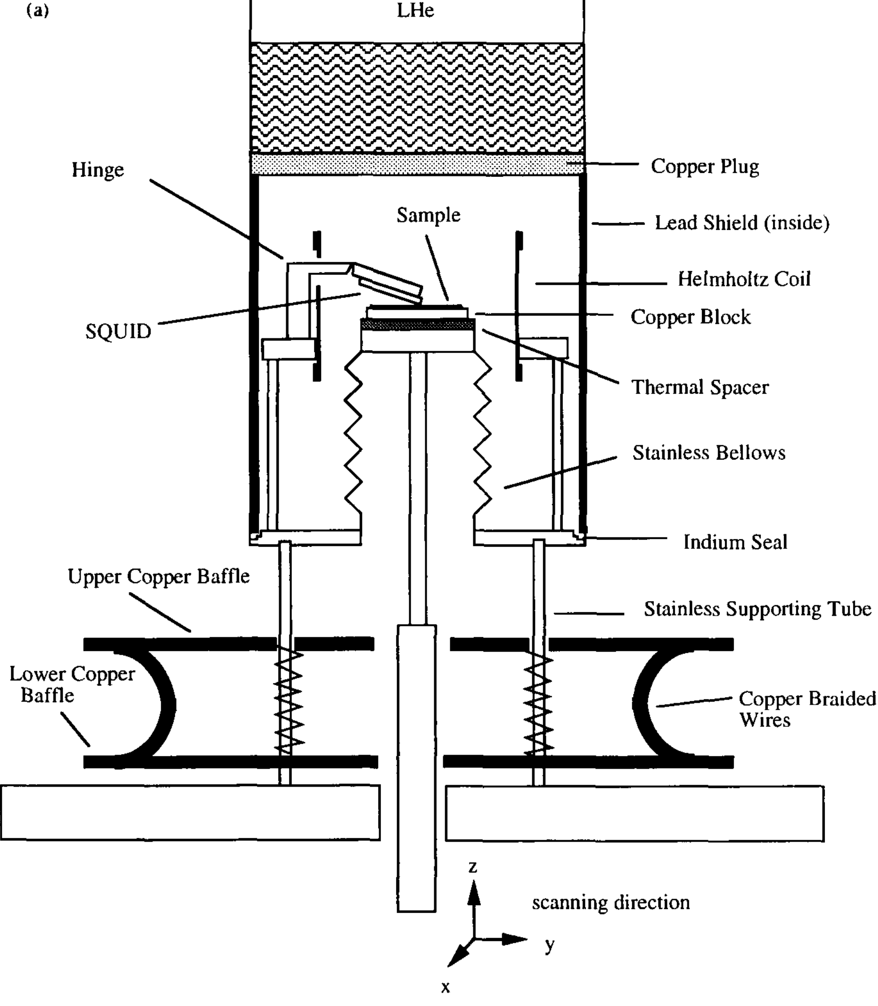


Figure 11-14a

(b)

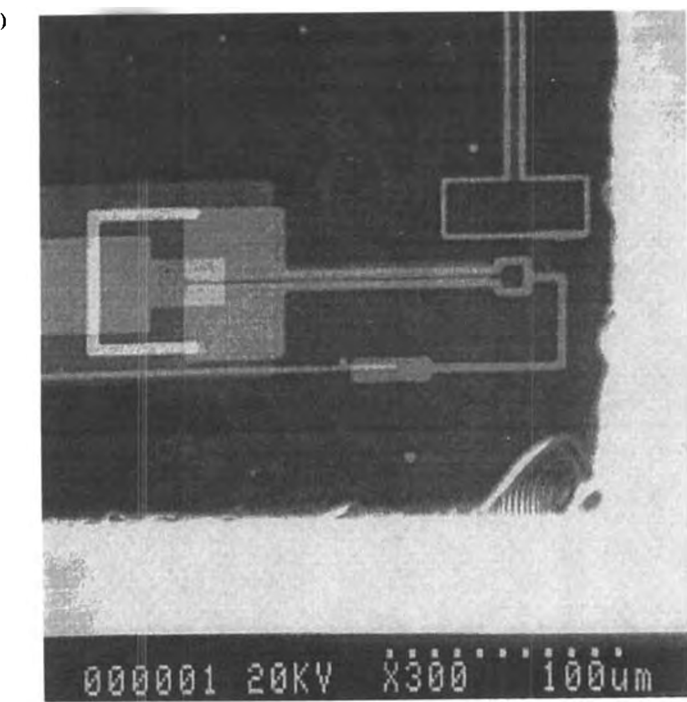


Figure 11-14. (a) Design of the scanning SQUID microscope, showing the scanning mechanism and the hinge assembly used to maintain close proximity of the SQUID to the object. (b) Electron micrograph of the dc SQUID chip (from [36]).

11.6.4 Magnetic Spin Sensors

Magnetic sensor technology may even be extended down to the atomic level by exploiting magnetic-dependent interaction mechanisms which are strongly confined in space. For instance, spin-dependent electron tunnelling between a magnetic nanoprobe and a magnetic object can be exploited to discriminate between different magnetic ions at the atomic length scale (Figure 11-16) [40]. By using a ferromagnetic iron nanosensor, contrast between Fe^{2+} and Fe^{3+} sites on an Fe_3O_4 (001) surface has already been demonstrated [41–43]. Alternatively, spin resonance techniques combined either with electron tunnelling [44] or force [20, 45] detection can be exploited to develop magnetic sensors on the nano-scale, offering a sensitivity which is far superior compared with conventional electron spin resonance techniques.

11.7 Conclusions and Outlook

Nanosensors are expected to evolve rapidly as one of the major consequences of recent advances in the field of nanotechnology. They usually offer a sensitivity at least as high as their conventional counterparts. On the other hand, since nanosensors are very much smaller and,

if batch-fabricated in large quantities, also very much cheaper, many new possible applications open up. Ultimately, new types of physical phenomena, which occur only on a mesoscopic length scale, may be exploited to develop new generations of future nanosensors.

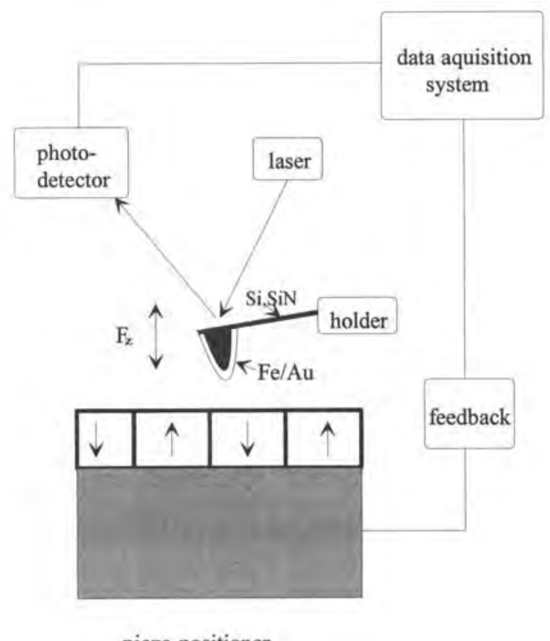


Figure 11-15. Schematic diagram of a magnetic stray field nanosensor.

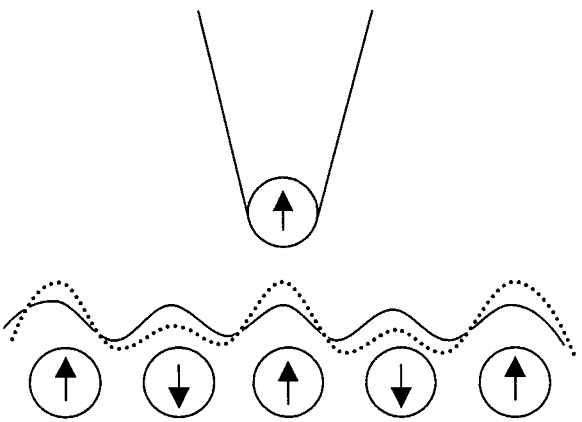


Figure 11-16. Schematic diagram of a magnetic spin sensor (from [40]).

11.8 Acknowledgments

The author would like to acknowledge all colleagues who have provided illustrations of their work for this chapter.

11.9 References

[1] Greenwood, J. C., *J. Phys.* **E21**, (1988) 1114.
[2] Binnig, G., Rohrer, H., *Helv. Phys. Acta* **55**, (1982) 726.
[3] Binnig, G., Rohrer, H., *Rev. Mod. Phys.* **59**, (1987) 615.
[4] Chen, J., *Introduction to Scanning Tunnelling Microscopy;* Oxford: Oxford University Press, 1993.
[5] Wiesendanger, R., *Scanning Probe Microscopy and Spectroscopy: Methods and Applications;* Cambridge: Cambridge University Press, 1994.
[6] Marrian, C. R. K. (ed.), *Technology of Proximal Probe Lithography;* Bellingham, SPIE Press, 1993.
[7] Güntherodt, H.-J., Wiesendanger, R. (eds.), *Scanning Tunnelling Microscopy I;* Berlin: Springer, 1992.
[8] Wiesendanger, R., Güntherodt, H.-J. (eds.), *Scanning Tunnelling Microscopy II;* Berlin: Springer 1992.
[9] Wiesendanger, R., Güntherodt, H.-J. (eds.); *Scanning Tunnelling Microscopy III;* Berlin: Springer 1993.
[10] Binnig, G., Quate, C. F., Gerber, Ch., *Phys. Rev. Lett.* **56**, (1986) 930.
[11] Kenny, T. W., Kaiser, W. J., Podosek, J. A., Rockstad, H. K., Reynolds, J. K., Vote, E. C., *J. Vac. Sci. Technol.* **All**, (1993) 797.
[12] Kenny, T. W., Kaiser, W. J., Reynolds, J. K., Podosek, J. A., Rockstadt, H. K., Vote, E. C., Waltman, S. B., *J. Vac. Sci. Technol.* **A10**, (1992) 2114.
[13] Wandass, J. H., Murday, J. S., Colton, R. J., *Sensors Actuators* **19**, (1989) 211.
[14] Brizzolara, R. A., Colton, R. J., Wun-Fogle, M., Savage, H. T., *Sensors Actuators* **20**, (1989) 199.
[15] Witek, A., Onn, D. G., *J. Vac. Sci. Technol.* **B9** (1991) 639.
[16] Niksch, M., Binnig, G., *J. Vac. Sci. Technol.* **A6**, (1988) 470.
[17] Bocko, M. F., Stephenson, K. A., Koch, R. H., *Phys. Rev. Lett.* **61**, (1988) 726.
[18] Carugno, G., Conti, E., Onofrio, R., *Europhys. Lett.* **24**, (1993) 713.
[19] Wolter, O., Bayer, Th., Greschner, J., *J. Vac. Sci. Technol.* **B9**, (1991) 1353.
[20] Rugar, D., Yannoni, C. S., Sidles, J. A., *Nature* **360**, (1992) 563.
[21] Meyer, G., Amer, N. M., *Appl. Phys. Lett.* **53**, (1988) 1045.
[22] Rugar, D., Mamin, H. J., Guethner, P., *Appl. Phys. Lett.* **55**, (1989) 2588.
[23] Göddenhenrich, T., Lemke, H., Hartmann, U., Heiden, C., *J. Vac. Sci. Technol.* **A8**, (1990) 383.
[24] Neubauer, G., Cohen, S. R., McClelland, G. M., Horne, D., Mate, C. M., *Rev. Sci. Instrum.* **61**, (1990) 2296.
[25] Tortonese, M. Barrett, R. C., Quate, C. F., *Appl. Phys. Lett.* **62**, (1993) 834.
[26] Gimzewski, J. K. Gerber, Ch., Meyer, E., Schlittler, R. R., *Chem. Phys. Lett.* **217**, (1994) 589.
[27] Williams, C. C., Wickramasinghe, H. K., *Appl. Phys. Lett.* **49**, (1986) 1587.
[28] Williams, C. C., Wickramasinghe, H. K., *Microelectron. Eng.* **5**, (1986) 509.
[29] Weaver, J. M. R., Walpita, L. M., Wickramasinghe, H. K., *Nature* **342**, (1989) 783.
[30] Williams, C. C., Wickramasinghe, H. K., *Nature* **344**, (1990) 317.
[31] Pohl, D. W., *Adv. Opt. Electron Microsc.* **12**, (1991) 243.
[32] Pohl, D. W., in: *Scanning Tunnelling Microscopy II,* Wiesendanger, R., Güntherodt, H.-J., (eds.), Berlin: Springer, 1992, p. 233.
[33] Betzig, E., Chichester, R. J., *Science* **262**, (1993) 1422.
[34] van Hulst N. F., Segerink, F. B., *Eur. Microsc. Anal.,* Jan. (1992) 13.

[35] Chang, A. M., Hallen, H. D., Harriott, L., Hess, H. F., Kao, H. L., Kwo, J., Miller, R. E., Wolfe, R., van der Ziel, J., Chang, T. Y., *Appl. Phys. Lett.* **61**, (1992) 1974.
[36] Vu, L. N., Wistrom, M. S., Van Harlingen, D. J., *Appl. Phys. Lett.* **63**, (1993) 1693.
[37] Martin, Y., Wickramasinghe, H. K., *Appl. Phys. Lett.* **50**, (1987) 1455.
[38] Saenz, J. J., Garcia, N., Grütter, P., Meyer, E., Heinzelmann, H., Wiesendanger, R., Rosenthaler, L., Hidber, H. R., Güntherodt, H.-J., *J. Appl. Phys.* **62**, (1987) 4293.
[39] Grütter, P., Mamin, H. J. Rugar, D., in: *Scanning Tunnelling Microscopy II,* Wiesendanger, R., Güntherodt, H.-J. (eds.): Berlin: Springer, 1992, p. 151.
[40] Wiesendanger, R., *J. Magn. Soc. Jpn.* **18**, (1994) 4.
[41] Wiesendanger, R., Shvets, I. V., Bürgler, D., Tarrach, G., Güntherodt, H,-J., Coey, J. M. D., *Z. Phys.* **B86**, (1992) 1.
[42] Wiesendanger, R., Shvets, I. V., Bürgler, D., Tarrach, G., Güntherodt, H.-J., Coey, J. M. D., *Europhys. Lett* **19**, (1992) 141.
[43] Wiesendanger, R., Shvets, I. V., Bürgler, D., Tarrach, G., Güntherodt, H.-J., Coey, J. M. D., Gräser, S., *Science* **255**, (1992) 583.
[44] Manassen, Y., Hamers, R. J., Demuth, J. E., Castellano, A. J., Jr., *Phys. Rev. Lett.* **62**, (1989) 2531
[45] Züger, O., Rugar, D., *Appl. Phys. Lett.* **63**, (1993) 2496.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

12 Trends in Sensor Technologies and Markets

ROBERT JONES, Cambridge Consultants Limited, Cambridge, UK

Contents

12.1 Introduction 358
12.2 The Role of Microfabrication and Advanced Electronic Design 359
12.3 A Review of Markets and Requirements 360
12.4 Concluding Comments 362
12.5 References 363

12.1 Introduction

The number of sensors required and the diversity of their application can be expected to increase significantly over the next five to ten years. The object of this section is to help the reader to understand the markets, applications and commmercial significance of this expansion*. To this end the following chapters analyze the complex interactions that exist between the requirements of the user, economic constraints and technology that are shaping the current and emerging directions. It is recognized that these interactions will depend on the industry and application and for this reason six key areas are considered:

- process sensing and control,
- manufacturing and quality assurance,
- automotive,
- aerospace,
- medical and healthcare,
- environmental.

In each of the above the author defines the basic measurement requirements and the technical and commercial constraints specific to the industry. This information is complemented by examples of current market size which serve to highlight the commercial importance of existing products. Emerging requirements and potential new markets are then identified and reviewed. It will be seen that these tend to be evolutionary in the more mature sensor markets such as process control and manufacturing quality assurance (QA) whereas the scope for entirely new products exist in, for example, the medical area where there is considerable opportunity for the application of new sensing techniques.

The development of commercially successful sensors is expensive and technically difficult. It requires sound commercial judgement and sophisticated market knowledge. It is by no means the prerogative of large companies but it is necessary that the appropriate combination of scientist, engineer, business manager and marketeer work as a well coordinated and stable team. At this level the prerquisite understanding of the technologies and markets calls for a combination of extensive industry knowledge and practical experience of which the following chapters provide a small element. There do exist, however, some central technical and commercial trends from which a broad perspective of the emerging industry may be acquired. These are now reviewed with reference to the content of the following chapters.

---

* The scale of this expansion can be assessed approximately from the compound annual growth of 7.4% per annum ($1.4 B in 1992 to $2.2 B in 1997) predicted by Frost and Sullivan for the European position sensors market since this is generic to a wide range of different sensor developments and applications.

12.2 The Role of Microfabrication and Advanced Electronic Design

The problems associated with the design and manufacture of high performance sensors that are low cost, robust, small, reliable and compatible with high volume manufacture can be mitigated by two techniques: micromachining and fabrication and the use of advanced electronic design and integrated circuits. These are both employed at the device level and complemented by the use of low cost, high performance microprocessors or PC based digital processors at the system level. (The latter topic is discussed briefly in Section 12.3 below with reference to process sensing.)

Although the importance of micromachining and fabrication to the sensor industry has been recognized increasingly over the last five years and has resulted in major research efforts, [1-3], it is important to note that it is already used in relatively mature products such as the precision, silicon micromachined pressure transducer shown in Figure 12-1. Important microfabrication techniques include the precision masking, anisotropic etch and bonding of silicon to make precise, mechanically strong, three dimensional structures [4] and precision moulding of materials using, for example, the LIGA process [5] or precision electro nickel formed masters [6]. The expansion of the demand for silicon sensors is expected to be led by the requirements of the automotive industry moving from \$64.7 M in 1992 to \$256.3 M in 1997 as part of a total increase of \$312.8 M to \$672.7 M over the same period [7]. In addition it is anticipated that a combination of microfabrication techniques will enable fully integrated optical sensors (see also Chapter 8 of this volume) to be produced at low cost and lead to the greater commercial exploitation of this class of device. (The world wide pressure sensor market [8] as led by silicon micromachined and fiber optic sensors is expected to grow from \$1.5 B in 1993 to \$2.3 B in 1998). Finally it is important to note that a wide range of microfabrication techniques are now well understood and in a number of cases available from commercial organizations. In order to exploit this situation it is necessary for manufacturers of sensors

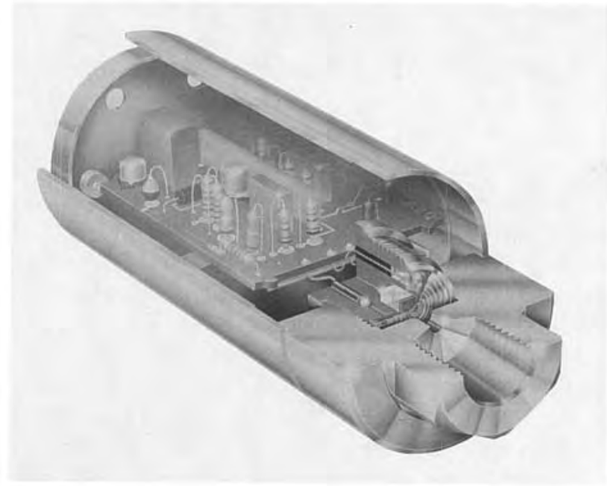


Figure 12-1.
Precision pressure transducer based on silicon micromachined diaphragm and integral strain gauge. Developed and manufactured by Druck.

to establish a good knowledge of micro system design and materials and thereby match device to the appropriate manufacturing technology.

The use of electronics to enhance sensor performance is evolving from the linear process in which the electronics is used to compensate for sensor error (for example, nonlinearity) to one in which the processor operates in feed-back with the analogue function of the transducer. To achieve this it is necessary that the transducer physics and design be combined closely with that of the electronics. This procedure is dominated by advanced design techniques and relies on conventional processes to produce ICs that enhance sensor performance whilst reducing size, cost and power consumption. Figure 12-2 shows, for example, the design of an ASIC (application specific integrated circuit) developed for incorporation in a consumer healthcare sensor based on an optical strip reader. This complex chip has been custom designed for the application and incorporates both digital and analogue functions which combine to normalise the output against degradation of the illuminating source and detector as well as transforming the reading to a reliable yes/no display. The complete system is expected to sell for less than £15, a price that could not be achieved without the use of this technology.

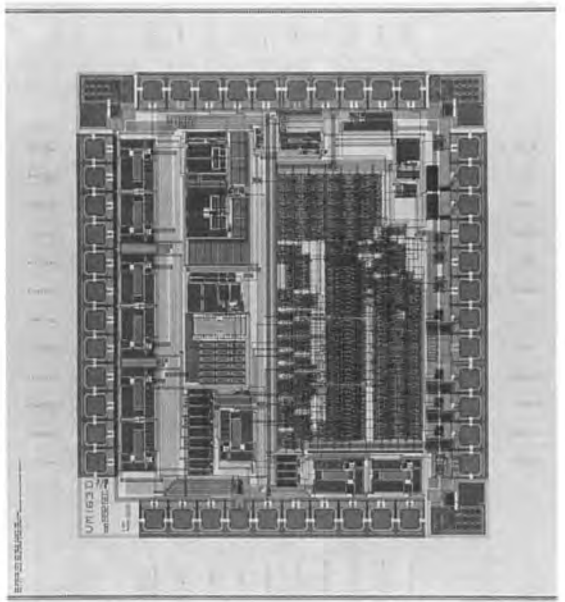


Figure 12-2.
The ASIC for a high volume, consumer healthcare sensor designed and developed by Cambridge Consultants Limited.

12.3 A Review of Markets and Requirements

In this section, a summary is presented of trends in sensor development specific to the six market areas noted in Section 12.1.

The need to optimize complex processes in order to improve efficiency means that the sensing systems for this application are increasingly required to generate detailed process information as well as provide advanced control functions. In response to this requirement systems

are now emerging in which neural networks based on low cost PCs are used to predict critical process parameters from the output of an ensemble of sensors or 'virtual' sensor (see Chapter 14). The range of process variables measured is also being extended by the use of surrogate measurements. In this technique the value of a process parameter that cannot be measured is derived indirectly from that of a closely related but measureable parameter.

Commercial and technical trends in metrology QA (quality assurance) will be driven by the increasingly intense competition that exists in large scale manufacturing. A current objective is to improve efficiency and reduce waste by closing the loope between the design and manufacturing process. For this purpose one strategy is to measure manufactured components and compare the real product with that demanded by CAD (computer aided design) data. The error can then be used to correct the production process in real-time. For this process to be effective the measurements have to be performed extremely rapidly at high accuracy, either at, or preferably on-line. This introduces a number of technical challenges centred on the need to make the measurement compact, high speed and accurate whilst operating, in the case of on-line measurement, under potentially harsh conditions. It is envisaged that these requirements will have a major influence on the design of sensors for advanced gauging systems and if satisfied will generate new markets as well as extending those for existing systems.

Sensing technology is an extremely important element of the automotive industry since it provides the means by which parameters essential to the current and emerging requirements for fuel economy, emission level control and safety are measured. Despite this key role, automotive sensors also need to be extremely low cost since their primary function is to enhance a consumer product. At the same time they must be capable of operating in a harsh environment, often at high temperatures in the presence of electro-magnetic interference. Standardization, extremely large market volumes and as noted earlier, fabrication in silicon provide the means by which this low-cost (typically $10 a sensor) may be achieved. Furthermore, smart sensing techniques in which local digital processing is used to compensate for the effect of temperature variation, nonlinearity etc. and provide the means by which the high performance may be achieved economically. Silicon piezo resisitive sensors with integral compensation are, for example, used routinely to measure manifold pressure for ignition and fuelling systems (Twenty million automotive pressure sensors are used annually throughout the world). It is also anticipated that silicon accelerometers will be used in future dynamics systems for active braking, suspension, steering and safety (e.g. air bag) systems. The need to improve safety on increasingly congested roads has also stimulated developments that aid the perception of road markings by using UV car lamps to illuminate cooperative surfaces and IR methods to aid driving in the dark and fog. Low-cost optical and microwave radars are also under development to aid collision avoidance. Here the funding of *Prometheus* (Programm for a European Traffic with Highest Efficiency and Unprecedented Safety) of £703 m over the past eight years from which these and related systems have emerged, is indicative of the importance of this relatively new aspect of automotive sensing.

The market for sensors in the aerospace industry is unique in terms of the extremely wide range of requirements that it embodies. This is a reflection of the need for control systems for aircraft, through to space and aircraft platform and payloads and ground infrastructures (for example air traffic control, ATC). The situation is complicated further by the division between the civil and military sectors each of which have specific and different needs. We see in addition an industry that is sensitive to macroscopic factors such as government spending policies, world politics (e.g. the end of the Cold War which has reduced military budgets) and

conflict (e. g. the Gulf War) which led to a major recession in the civil sector. The sensors also have to exhibit extremely high levels of technical performance and reliability over a wide range of environmental conditions within market volumes that vary from the small, 'one-off' for scientific satellite sensors incorporated in payload scientific experiments to the large (for example control surface position encoders). A specific market sector in which there is expected to be a significant growth is associated with the ground infrastructure where a rapid expansion of ATC systems to accommodate the increase in passenger traffic is expected. In parallel it is anticipated there will be sound growth rates in all of the other segments albeit at a somewhat lower level.

Emerging requirements for sensors in the medical and healthcare sector are being driven by the movement towards the continuous (and ideally non to minimally invasive) monitoring of patients. Under these conditions feedback on the state of the patient and the results of therapy can be obtained without the delays associated with conventional intermittent measurement and the use of central laboratories for chemical and biochemical analysis. Recently developed noninvasive neo-natal blood oxygenation monitors and consumer fertility hormone sensors based respectively on near IR spectroscopy and an optical biochemical assay strip reader are examples of sensors developed in response to these requirements. They are also representative of the increasing use of optoelectronic techniques in this area (see, for example, Bio-Photonics [9]). Developments in biosensing are also extremely important in healthcare in which the market value in 1992 was assessed to be $40 m but was dominated primarily by the blood glucose electrode. Against this background it is reasonable to assume significant scope for the successful commercial exploitation of innovative products although the risks and cost associated with rigorous clinical trials and the need to satisfy legislation should be taken into account.

The market for environmental sensors is expanding and diversifying as a result of rapid world wide industrialisation and population growth and the need to monitor and control the resultant pollution. Stringent legislative requirements combined with the influence of individual pressure groups are driving this expansion which is expected to stimulate the need for new sensors to detect a diversity of air, water and ground pollutants. An important technical trend is towards continual, in-situ monitoring. Problems associated with sensor ageing, fouling and general degradation will therefor need to be overcome by a combination of careful design and the development of new sensing techniques. One potentially large market is for sensors that will be used by the individual to check their personal environment, for example, local air and drinking water quality. Here the challenge is the manufacture of sensors that are by necessity simple to use and low cost.

12.4 Concluding Comments

The following chapters present a snap-shot of a range of industries and markets, all of which are in a state of change, some more rapid than others. The views presented of the future are invitably subjective and influenced by the background and experience of the individual authors. It is to be hoped, however, that they will inform as well as provoke and stimulate the reader.

12.5 References

[1] *Micro System Technologies '90, Proc. 1st International Conference in Micro Systems, Berlin, September 1990*; Springer-Verlag, 1990.

[2] *Micro System Technologies '92. Proc. 2nd International Conference on Micro Systems, Berlin, October 1992*; Springer-Verlag, 1992.

[3] *Proc. IEEE "Micro Electro Mechanical Systems", Amsterdam, January/February 1995*; IEEE Catalog number 95CH35754.

[4] Greenwood, J. C., *Instrument and Science Technology* (1988) 1114-1127.

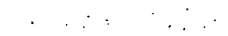
[5] Rugner, A., et al., *J. Micromech. Microeng. 2* (1992) 133-140.

[6] Harvey, T. G., et al., *Proc. CLEO/EUROPE '94*, p. 292.

[7] McEntee, J., *Physics World*, December 1993, p. 33-34.

[8] *Electro Optics*, Sept/October 1994, p. 9.

[9] *Premier Issue 1994*; Bio-Photonics, published by Laurin.



Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

13 Aerospace Sensors

REINHARD H. CZICHY, European Space Agency, ESA/ESTEC,
Instrument Technology Division, Noordwijk, The Netherlands

Contents

13.1 Introduction 366

13.2 Market Structure 366

13.3 General Requirements for Aerospace Systems 368

13.4 Sensor Types and Technological Trends 373
13.4.1 Sensors for Aviation Systems 373
13.4.2 Sensors for Space Systems 385
13.4.2.1 Spacecraft and Platform Sensors 387
13.4.2.2 Payload Sensor Instrumentation 392
13.4.3 Sensors for Ground Infrastructure 403

13.5 Market Trends 404
13.5.1 Aviation Systems 405
13.5.2 Space Systems 406
13.5.3 Ground Infrastructure 408

13.6 Summary 408

13.7 Acknowledgments 409

13.8 References 409

Disclaimer: The views expressed in this chapter are those of the author and do not necessarily represent those of the European Space Agency.

13.1 Introduction

The aerospace market has three major segments, aviation systems, space systems, and ground infrastructure, and can be characterized as a medium to small volume market for products at high technical performance. These have to comply with a number of demanding requirements which derive from a combination of severe environmental loads and operational constraints.

The influence of these requirements is the background against which the trends in the various aerospace market segments will be discussed and the technological directions identified.

Aerospace sensors can be considered today as key elements of modern avionics and spacecraft control systems. The term "sensor" here covers a diversity of devices; according to the ANSI definition, referred to in the introductory chapter of Volume 1 of this series, a sensor is defined as a "single-parameter measuring instrument which transduces a physical parameter into a corresponding electrical signal with significant fidelity" [1]. This definition applies not only to simple components but also to the highly complex systems the description of which would go far beyond the present purpose. However, in order to identify the full spectrum of the aerospace sensor market, certain more complex units will also have to be covered in this discussion.

The term "unit" describes, in this context, a self-contained piece of equipment which consists of primary detector elements and associated components, assemblies, and sub-assemblies designed to perform specific measurement tasks and to process the acquired data and present it to the next higher hardware level, via an interface. Typical examples of such units are the sensors of spacecraft attitude and orbit control systems (AOCS) that embody sun sensors, star sensors, earth sensors, gyroscopes, or the sensor units which are incorporated in an aircraft avionics system.

It is common practice to distinguish, for aerospace systems, between the payload and the platform which supports the payload in execution of a mission. Such platforms can be aircraft, launch vehicles, balloons, or satellite service modules. Payloads might be cargo, satellite communication terminals, or the weapon system of a military aircraft. The discussion of the different sensor types will be correlated with their application and function, since an understanding of payload sensors allows interesting conclusions to be drawn concerning the market trends in the general field of aerospace.

13.2 Market Structure

The aerospace market segments, each with a civil and a military subsection, are shown in Table 13-1 together with the most important product groups associated with each. For the product group a distinction is made between platform and payload products. Examples of aerospace payloads with special sensor instrumentation are listed in Table 13-2.

The market structure shown in Table 13-1 indicates also the main customers for aerospace products; in the civil sector those for aviation systems are predominantly airlines and air cargo

Table 13-1. The aerospace market structure with related typical product groups.

Related product groups / Market segment	Civil branch	Military branch
Aviation systems	• Platforms – Fixed-wing planes – Helicopters – Lighter-than-air systems • Payloads	• Platforms – Fixed-wing planes – Helicopters – Remotely piloted vehicles – Missiles • Payloads
Space systems	• Platforms – Satellites – Space stations – Launch vehicles • Payloads	• Platforms – Satellites – Missiles • Payloads
Ground infrastructure	Air traffic control (ATC) systems Simulators and trainers Maintenance systems Communication ground stations Hand-held communication terminals GPS terminals	Air traffic control (ATC) systems Simulators and trainers Maintenance systems Communication ground stations Hand-held communication terminals GPS terminals

Table 13-2. Examples of aerospace payloads with special sensor instrumentation.

Market segment	Payload types with special sensor instrumentation
Aviation systems	Remote sensing instrumentation Meteorological instrumentation Communication systems Air space surveillance systems Weapon systems
Space systems	Communication systems Space science instrumentation Earth observation systems Microgravity research instrumentation Weapon systems Satellites (as payloads of launch vehicles) Manned spacecraft (as payloads of launch vehicles)

carriers; space systems are procured by national/international government agencies and by the communications industry. The customers for aviation-related ground infrastructure are airport operators and, in the area of training and maintenance equipment airlines/air cargo carriers and their related maintenance organizations. Space system oriented ground infrastructure is procured mainly by space agencies and the communications industry. Small antenna systems

and hand-held terminals are purchased increasingly by the private consumer (see also Section 13.5.3).

The military branch of the market is entirely dependant on government contracts, which are placed via the various ordonance system procurement authorities.

The various sensor types which are specific to the above-mentioned products will be identified in Section 13.4, in which the related development directions and technological trends are also highlighted. The market trends for the various products with respect to product types and associated financial volumes are discussed in Section 13.5.

Prior to the above, the general requirements and design constraints for aerospace systems will be summarized, in order to provide the background necessary for an assessment of future market trends and development.

13.3 General Requirements for Aerospace Systems

Aerospace systems have to comply with a number of requirements which result from the combination of application-specific constraints and the environmental loads imposed during operation.

Table 13-3 shows some examples of constraints and requirements typical of aerospace systems and indicates the consequences for the development, design, and manufacture of the equipment which arise from them. The environmental loads to be taken into account for the different market segments and applications are summarized in Table 13-4.

Aerospace equipment has to be designed to withstand these environmental loads during operation without degradation of the performance beyond the specified limits. This imposes stringent requirements in particular on space systems which have to fulfil their specified per-

Table 13-3. Typical constraints and requirements for aerospace equipment and potential consequences for development, design and manufacture.

Constraints and requirements*	Design consequences*
Manned operation of the system	• Design has to comply with the safety regulations applicable for manned systems
Long lifetime[1]	• Analysis of mission scenario is required to identify potential degradation effects and to prepare for implementation of sufficient design margins
High reliability	• Analysis of design in frame of a reliability programme[5] • Design optimization based upon results of reliability analyses
Low technological risk[2]	• Only stable and reproducible technologies must be used • Parts, materials and processes with good heritage from other applications with similar loads will be preferred
High availability[3]	• Minimization of maintenance needs • Definition and implementation of preventive maintenance programme • Minimization of down times by improved maintainability

Table 13-3. (continued)

Constraints and requirements*	Design consequences*
Good maintainability[3]	• Modular design • Use of standardized components and interfaces • Good accessibility • Implementation of automated system check-out and failure diagnostic facilities
Good operational flexibility[3]	• Standardization of interfaces, units, and components
Good survivability[4]	• Design for superior flight performance • Radiation hardening and EMP protection • Implementation of countermeasure/electronic warfare systems • Implementation of efficient friend-foe-identification system
Good environmental compatibility	• Low fuel consumption • Low-noise engines • Clean combustion with minimized pollution • In case of space system: minimization of space debris production
Certifiability[3]	• Design must comply with the applicable international/national standards
Qualifiability	• The design must be such that all specified properties can be verified by test, analysis, or a combination of both • Traceability throughout all phases of design, manufacture, assembly, and testing is required
Low life cycle cost[3,4]	• Minimization of maintenance requirements, consumption values, and down times
Low costs of ownership[3]	• Maximization of payload exploitation potential
Medium to small production series	• Implementation of efficient PA programme in order to compensate for limited learning effects due to comparably small production lot size • Selection of manufacturing processes with high reproducibility • Use of well characterized and technologically stable materials only • Consistent documentation to provide traceability throughout all phases of design, manufacture, assembly, and testing
High system efficiency	• Design for specified technical performance, lifetime and reliability with – low mass – small dimensions – minimum power consumption – minimum maintenance requirements – low procurement costs and low operational costs

* Notes: 1. For space systems long operational lifetime without maintenance is required.
2. Mainly applicable to space projects.
3. Applicable to aviation systems.
4. Applicable to military systems.
5. For definition of a typical reliability program for space applications, see Ref. [2].

Table 13-4. Environmental load for aerospace systems.

Market segment	Environmental load
Aviation systems	• Mission specific loads – Thermal stress – Mechanical loads (static and dynamic) – Meteorological conditions – Radiation loads – Reduced atmospheric pressure – Contamination • On-ground loads – Ambient atmospheric conditions – Contamination – Ground specific handling loads
Space systems	• Mission specific loads – Thermal stress – Mechanical loads (static and dynamic) – Radiation loads – Vacuum conditions – Atomic oxygen – Micrometeorites – Spacecraft generated contamination • On-ground loads – Ambient atmospheric conditions – Contamination – Ground specific handling loads
Ground infrastructure	• On-ground loads – Ambient atmospheric conditions – Contamination – Ground specific handling loads

formance over the entire in-orbit lifetime without the possibility of being serviced. The operational lifetimes of space systems, depending on the particular mission, may range from a few days (space shuttle missions, for instance) to more than 10 years. Earth observation satellites in low earth orbits might have lifetimes of 1-3 years, depending on the orbital parameters, whereas commercial communication satellites in geostationary orbit have to operate in orbit for 8-10 years. The specified lifetimes for some scientific satellites dedicated to interplanetary missions might be much longer than 10 years.

In order to fulfil the above requirements, the design of space systems has to be kept inherently insensitive to technical problems by allowing for sufficient margins, by avoiding single-point failure sources, by selection of well characterized and appropriate materials, by minimization of interfaces, and by application of only highly reproducible and controllable manufacturing processes.

Aviation systems have also to be designed for long lifetimes (20-25 years) in comparison with other commercial products. However, the life cycles of such systems are characterized by sequences of airborne operational periods and maintenance periods on-ground, and therefore

the operational periods which have to be kept free of any malfunction are considerably shorter than with most space systems.

The necessity for highly reliable operations is also reflected by the stringent requirements for product assurance (PA) as they usually apply for aerospace equipment; these PA activities start in design and are performed throughout the development and production phases up to the point of operation. They comprise the elements design assurance, parts control, materials control, process and quality assurance. The design assurance, performed during the design phase, supports the analysis of the design with respect to the behavior of the system in operation and the effect of faults on other systems, and predicts the reliability of the system over its lifetime. Design Assurance hence has to be understood as a design tool for optimization of the product performance. The manufacturing phase is accompanied by the specific PA activities to be executed for safeguarding that the design is properly converted into hardware, by parts, materials, and process control, and quality assurance. Parts, materials and process control support the definition of materials, the procurement, and incoming inspections, in addition to the definition of critical processes to be applied for production of the equipment. The quality assurance activities executed during manufacture, assembly and test, have to ensure that the parts are manufactured and assembled as specified and that the tests for qualification and acceptance of the product are correctly performed.

The PA activities to be executed during the life of a product are usually defined in PA requirement specifications as illustrated in Figure 13-1 by the specification tree which shows the PA requirement specifications applicable for space projects of the European Space Agency (ESA).

In the field of aviation, the PA activities executed under manufacturers' responsibility are supplemented by the certification of the air-worthiness of the system. This certification, which has to take place prior to commencement of service, is usually performed or supervised by government authorities such as the Federal Aviation Authority (FAA) in the USA, the Civil Aviation Authority (CAA) in the UK, the Direction Générale de l'Aviation Civile (DGAC) in France and the Luftfahrt Bundesamt (LBA) in Germany.

Aviation systems also have to comply for their certification with a number of mandatory regulations and requirements concerning air traffic safety and operational reliability. These include

- operational safety and adequate performance;
- sufficient structural strength with respect to static and dynamic loads;
- adequate design and construction;
- adequate design emergency cases and implementation of suitable safety precautions;
- compliance with state-of-the-art environmental compatibility requirements.

The air-worthiness of military aviation systems is usually verified by the corresponding procurement service according to the technical specifications and standards applicable to the equipment. Specific product requirements are usually complemented by national/international standards which are made applicable for the design, manufacture and procurement of aerospace equipment.

An aspect of growing importance for the procurement of aviation or space systems is that of the "system efficiency"; this term expresses the relationship between the resources necessary to support the execution of a mission and the results of a mission, as is illustrated

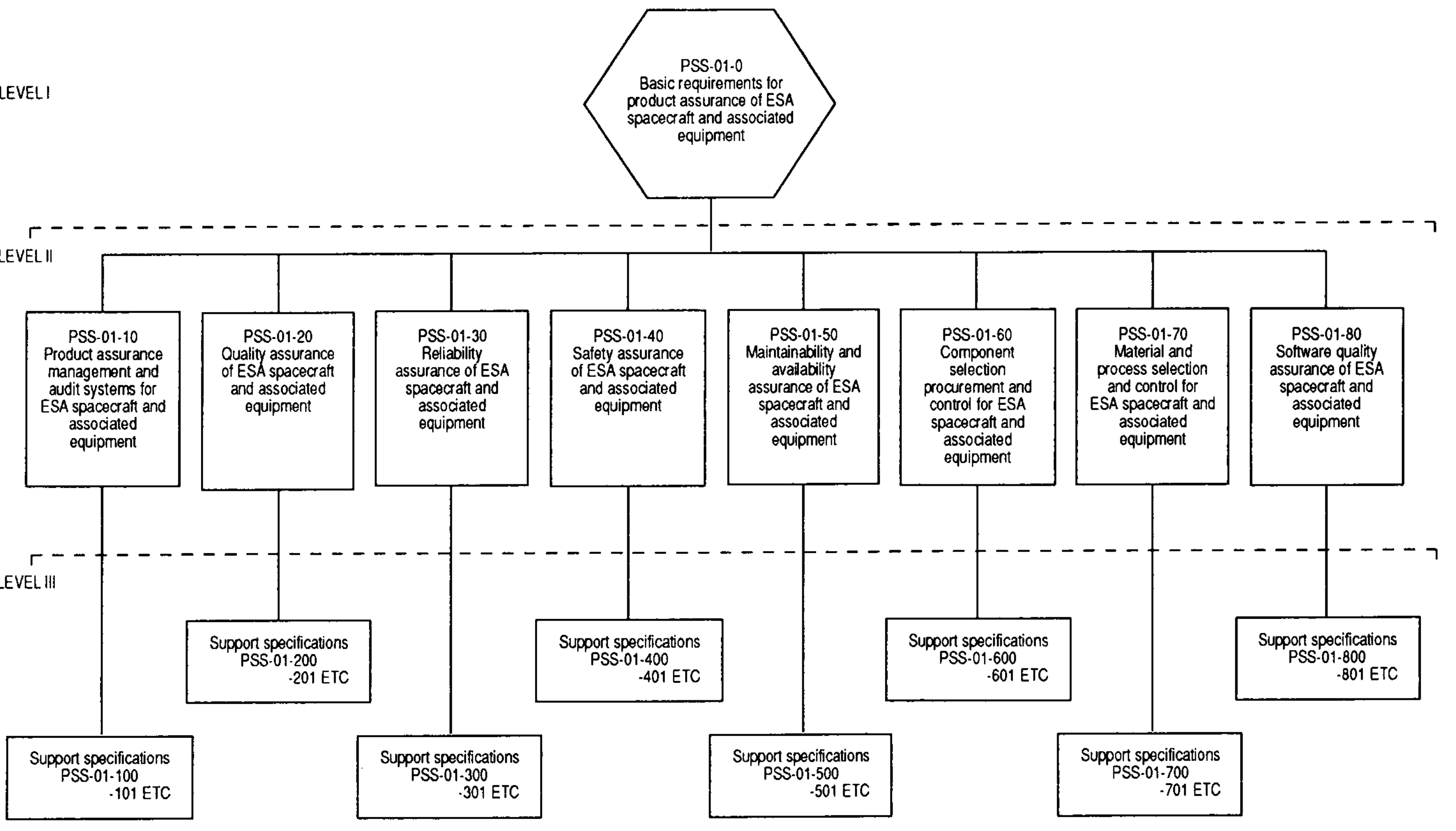


Figure 13-1. ESA product assurance specification tree [3].

by the following example. The operational lifetime of a communication satellite is determined partially by the amount of propellant available for station keeping maneuvers. The system becomes more efficient if the satellite service module can, for a given payload performance, fixed launch mass and system price, be made to carry more fuel. Under these conditions the user can operate the system for longer periods and thereby generate greater returns on investment.

It is anticipated that the combination of commercialization of space systems, particularly the area of communication, remote sensing, and navigation, and the economic constraints to which civil and military aviation are subject, will increase the importance of system efficiency in the future design of aerospace systems and the associated sensors. This aspect must therefore be considered an additional design factor and constraint for advanced aerospace systems.

13.4 Sensor Types and Technological Trends

The sensors used in aviation and in space systems are either part of a platform (eg, aircraft, satellite service module) or part of the payload which is carried and supported by the platform (see also Section 13.1). The following discussion will distinguish between these two areas of sensor applications.

The classification of sensor types used here will follow that proposed in Volume 1, ie the classification scheme will be based on "the input signal of the first transduction principle" [1]. Aaerospace sensors can hence be classified as follows:

- mechanical sensors,
- thermal sensors;
- sensors for electrical parameters;
- magnetic sensors;
- optical sensors;
- sensors for ionizing radiation;
- chemical and biological sensors.

13.4.1 Sensors for Aviation Systems

The basic types of sensors used in aviation systems will be identified taking an aircraft as an example. Table 13-5 shows its major subsystems, the associated functional units, and the sensor types forming part of it. The sensor output data is either used in unit internal control systems or transmitted, in raw or preprocessed form, to the aircraft avionics system for evaluation and/or display. (The term "avionics" is understood here as a generic term for all electronics, electrical, and electromechanical systems installed in an aircraft and used for control of the aircraft functions and its flight [4, 5].)

Table 13-6 shows the main elements of an aircraft avionics system with its functional units and the associated sensors. Reference is made here to the sensors identified in Table 13-5. This

Table 13-5. Major aircraft subsystems, its subsystem specific units and typical associated sensors.

Aircraft subsystem	Units	Associated sensors*
Aircraft structure and mechanisms	Fuselage	Smoke/fire detectors (30) Position sensors (31) Air conditioning system sensors (24, 25, 26) System status sensors (32–34) External temperature sensors (38) Flight monitoring/control system sensors (44–50) Navigation system sensors (39–43) Antennas (51–54) Strain gauges (37)
	Wings	Position sensors (15, 16, 20) Flight monitoring/control system sensors (46) Strain gauges (37)
	Empennage	Position sensors (17, 20) Antennas (43, 51) Strain gauges (37)
	Landing gear	Antennas (42) Position sensors (20) Load cells (36) System status sensors (35)
Engine	Engines	Tachometer generator (1) Temperature sensors (2–6) Pressure transducers (7–9) Flow rate sensors (10) Vibration sensors (14) Smoke/fire detectors (30) Position sensors (20, 31) Strain gauges (37)
	Tanks (fuel, oil, water)	Fluid level sensors (11–13) Flow rate sensors (10)
Electrical power supply	Generators (main/auxiliary generator)	Tachometer generators (22) Voltage/current sensors (23)
	Batteries	Voltage/current sensors (23)
Hydraulic power supply	Compressors	Tachometer generators (18) Pressure transducer (19)
	Hydraulics control assembly	Fluid level sensor (21) Pressure transducer (19) Position sensors (20)
Environmental control	Air conditioning and pressurization	Fluid level sensor (28) Thermistors (25) Humidity sensors (26) Air flow sensors (27)

Table 13-5. (continued)

Aircraft subsystem	Units	Associated sensors*
		Smoke/fire detectors (30) Pressure transducers (24)
	Fire alarm/extinguisher unit	Smoke/fire detectors (see also corresponding units (30)) Position sensors (31)
	Ice warning/de-icing unit	Ice sensors (29)
Avionics	See Table 13-6	See Table 13-6
Payload	Payload specific units	Payload specific sensors

* Numbers in parentheses indicate the association with the aircraft avionics and the sensors used (see Table 13-6).

Table 13-6. Elements of an avionics system, its units and associated sensors.

Avionics system elements	Units	Associated sensors	No.
Aircraft monitoring system			
● Engine monitoring system	Tachometer	Tachometer generator	1
	Temperature indicators for		
	– Exhaust gas temperatures	Thermoelements, platinum resistance thermometers (PTR)	2
	– Cylinder head temperatures	Thermistors	3
	– Oil temperatures	Thermistors	4
	– Water temperatures	Thermistors	5
	– Carburettor temperatures	Thermistors	6
	Pressure indicators for		
	– Oil pressure	Piezoresistive transducers, strain gauges	7
	– Fuel pressure	Piezoresistive transducers, strain gauges	8
	– Manifold pressure	Piezoresistive transducers, strain gauges	9
	Flow rate indicators for		
	– Fuel flow rate	Impellers, vortex-precession sensors	10
	Quantity indicators for		
	– Fuel quantity	Resistive, capacitive, or glass fiber sensors	11
	– Oil quantity	Resistive, capacitive, or glass fiber sensors	12
	– Water quantity	Resistive, capacitive, or glass fiber sensors	13
	Vibration indicators	Accelerometers	14

Table 13-6. (continued)

Avionics system elements	Units	Associated sensors	No.
● System monitoring instrumentation	Position status indicators for		
	– Ailerons, flaps	Limit switches, potentiometers, optical encoders	15
	– Speed brakes	Limit switches, potentiometers, optical encoders	16
	– Elevator, rudder	Limit switches, potentiometers, optical encoders	17
	Hydraulic system status indicators for		
	– Tachometer	Tachometer generator	18
	– Pressure	Piezoresistive transducers, strain gauges	19
	– Valve positions	Limit switches	20
	Hydraulic fluid quantity	Resistive, capacitive, or glass fiber sensors	21
	Electrical generators status indicators for		
	– Tachometer	Tachometer generator	22
	– Voltage/current	Ammeters, voltmeters	23
	Air conditioning system status indicators		
	– Cabin pressure	Aneroid with electrical transducers (strain gauges, piezoresistive transducers)	24
	– Temperature	Thermistors	25
	– Humidity	Capacitive sensors	26
	– Air flow rate	Anemometer	27
	– Water quantity	Resistive, capacitive, or glass fiber sensors	28
	Ice warning/de-icing unit	Ultrasonic sensor, thermistor/humidity sensor, microwave sensors	29
	Fire alarm/extinguisher unit	Smoke/fire detectors	30
		Limit switches	31
	System status indicators for		
	– Flight control computers	Control software	
	– Door status	Limit switches	32
	– Illumination systems	Relays	33
	– Cabin signals	Relays	34
	– Landing gear status	Limit switches	35
	Aircraft weight and C.O.G. display	Load cells	36
	Airborne integrated data system	Strain gauges	37

Table 13-6. (continued)

Avionics system elements	Units	Associated sensors	No.
		Registration of all relevant sensor outputs	
	External temperature display	Thermistors	38
Flight information system			
• Navigation system	Magnetic compass	Magnet	39
	Gyro compass	Gyroscope	40
	Inertial navigation unit (INU)	Gyroscope package, accelerometers and clock	41
	Radio direction finders (eg radio magnetic indicator (RMI), automatic direction finder (ADF)	Antennas	42
	Area navigation systems (eg LORAN C, OMEGA, ONTRACK, VOR Systems, TACAN, GPS)	Antennas	43
• Flight monitoring and control system	Altimeter	Pitot static tube with aneroid and electrical transducer (eg strain gauge, piezoresistive transducer), radar antennas	44
	Vertical speed indicator (VSI)	Pitot static tube with externally balanced aneroid and electrical transducer (eg, strain gauge, piezoresistive transducer)	45
	Airspeed indicator	Prandtl tube with aneroid and electrical transducer (eg strain gauge, piezoresistive transducer)	46
	Angle of attack indicator	Pivoted vane with angular encoder	47
	Turn and bank indicator	Gyroscope with force vector indicator	48
	Gyro horizon	Gyroscope	49
	Attitude direction indicator (ADI)	Sensor combination for attitude and vertical navigation	
	Horizontal situation indicator (HSI)	Sensor combination for horizontal navigation and flight management	
	Autopilot	Sensor data from various units of the navigation system and the flight monitoring/control system	
	Instrument landing system (ILS)	Sensor data from various units of the navigation system, the	

Table 13-6. (continued)

Avionics system elements	Units	Associated sensors	No.
		flight monitoring/control system and the radar	
	Microwave landing system (MLS)	Sensor data from various units of the navigation system, the flight monitoring/control system and the radar	
	Traffic alert and collision avoidance system (TCAS)	Sensor data from radar system	
	Ground proximity warning system (GPWS)	Sensor data from altimeter and navigation system	
	Low level wind shear alert system (LLWAS)	Antennas	50
• Flight management system	Flight management system computer	Interfacing to the units/sensors of the aircraft monitoring and the flight information systems	
Communication system	VHF/UHF transceiver	Antennas	51
Radar	Doppler radar	Dish antenna	52
	Weather radar	Dish antenna, waveguide array antenna	53
	Radar transponders (mode A, C, S)	Omnidirectional antennas	54

list is not exhaustive owing to the large variety of aircraft types and their associated instrumentation. However, the typical elements of aircraft avionics systems are identified. Military aviation systems are not covered specifically in either table.

The types of sensors used on-board civil aircraft comprise, with the exception of chemical and biological sensors, the full variety of sensor types identified above but with a clear emphasis on sensors for electrical, mechanical, and thermal parameters.

The evolution of new sensor concepts and the related technologies derive from trends in the field of advanced aircraft and avionic system developments, such as

- the introduction bus-oriented avionics systems with modular architecture [6];
- the use of smart sensors and actuators to reduce the load of the central avionics control system and to provide higher system fault tolerance [7];
- the use of optical fiber bus concepts for command and data acquistition links. These reduce weight and improve protection against electromagnetic interferences (EMI) and electromagnetic pulses (EMP) [7, 8];
- the application of advanced monitoring methods for engine operating parameters in order to allow for optimum engine operations, thus improving economy and minimizing emissions [9];
- the improvement of aircraft diagnostics in order to allow for early failure detection and more efficient preventive maintenance;

- the implementation of electronic library systems to support flight management [10];
- the improvement of navigation tools in terms of accuracy and reliability in order to allow for higher air traffic density [8];
- the installation of next-generation approach/landing aids in order to allow for higher precision, all-weather runway approaches [10];
- the improvement of system reliability, in particular in view of advanced flight control system software [6];
- the installation of airborne data communication systems to allow for optimum flight management and passenger data links [8];
- the improvement of aircraft environmental compatibility in particular by reduction of fuel consumption and noise emission [11].

These trends in the civil market are complemented by developments in the field of military aviation, where, as a result of budgetary restriction, most of the activities are currently aimed at improving aircraft tactical efficiency. They comprise, for instance [10, 12],

- the development and implementation of integrated airborne electronic warfare suits;
- improvement of night fighting capability in particular for close-to-ground operations;
- implementation of more efficient friend-foe-identification systems;
- improved airborne battlefield surveillance systems with on-board processing and quick-look data display capability;
- improvement of airborne communications capability.

The consequences of these trends in system development in the field of sensor technology are illustrated by the following examples.

Mechanical sensors are used for the detection of parameters such as position, acceleration, pressure, stress, and strain [1]. On-board aircraft, the simplest mechanical position sensors are limit switches such as those used on valves or doors which detect whether a given motion limit is reached. Other position sensors used to monitor the status of aileron, rudder and stabilator have to detect the position within a certain motion range.

In order to improve system reliability and lifetime, there are increasing tendencies to use non-contacting sensors which transform the position directly into an electrical signal without the need for an intermediate moving mechanical element [7]. Examples of such sensors are optical, capacitive, or inductive limit switches or optical encoders for angular or linear position measurement. Such optical encoders can also be designed to provide the position output as a digital signal without the need for an A/D converter, interfacing directly with the digital avionics bus or control electronics which are part of the corresponding actuator. These are often referred to as "smart actuators" [7]. Smart actuators offer potential for the enhancement of avionics systems; the flight control computer is offloaded from time critical control loop calculations, the amount of data to be exchanged with the corresponding actuators is minimized, and the number of electrical connections can be significantly reduced, thus, contributing to weight reductions. Moreover, such decentralized control systems offer the advantage of a better failure tolerance in the overall system. Advanced optical position sensors, which directly provide high-resolution digital position output, will hence be of interest in the development of the next-generation of avionic systems.

More complex mechanical sensors used within the navigation instrumentation are, eg, the gyro compass or the inertial navigation unit (INU) [13]. A gyro compass usually contains a mechanical (spinning-wheel) gyroscope which is either spun by a vacuum system or is electrically driven. Figure 13-2 shows a selection of such gyroscopes. INUs are based on attitude sensors in a gimballed or in a fixed configuration such as strap-down systems. The former has the accelerometers mounted in a fully gimballed configuration together with gyroscopes, whereas the latter have the accelerometers and gyroscopes mounted in a fixed configuration on the platform to be controlled.

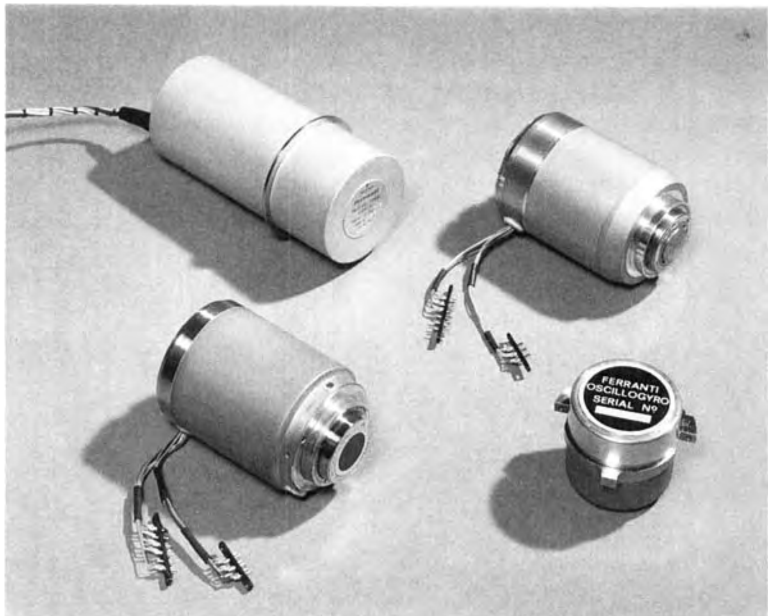


Figure 13-2. Selection of mechanical gyroscopes. Clockwise from lower left: (a) Type 125 inertial quality floated rate integrating gyroscope; life test on 12 samples continuing after 150000 h without failure; (b) Type 121 floated rate integrating gyroscope for high rate applications; (c) Type 122 inertial quality floated rate integrating gyroscope; (d) Type 142 oscillogyro dynamically tuned gyroscope. (Photograph courtesy of GEC-Marconi).

Strap-down systems which have to comply with high rotation rates ($\geq 400\ ^{\circ}s^{-1}$) are usually designed using optical ring laser gyros (RLGs), which can cope with extremely high rotation rates without losing their drift performance [14, 15]. Figure 13-3 shows an RLG-based inertial sensor assembly. Another alternative for mechanical gyros is fiber-optic gyros (FOG), which can be designed to have small dimensions by using fiber/integrated optics technology [16].

Development trends in the field of navigation sensors and systems aim at providing higher overall accuracy over extended lifetimes and full compatibility with other navigation systems, eg satellite navigation systems (GPS) or radio navigation systems [17, 18]. The objective is to prepare for the development of advanced integrated navigation concepts, in which advanced optical gyroscope types, RLGs and FOGs, will play a major role. An example of a GPS-supported RLG inertial navigation system is shown in Figure 13-4.

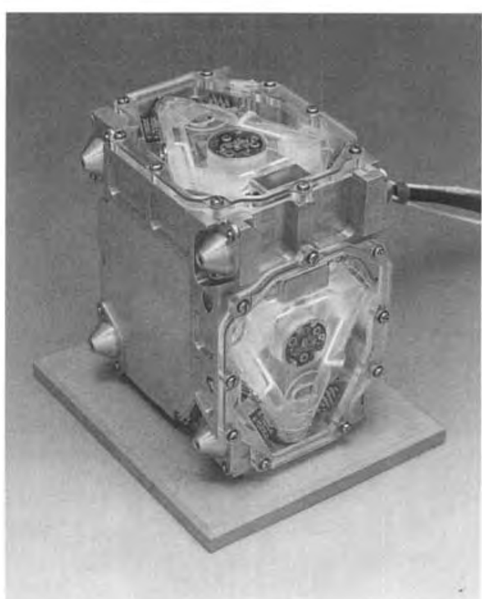


Figure 13-3.
FIN 3110G RLG inertial sensor assembly using Type 164 24 cm path length instruments. (Photograph courtesy of GEC-Marconi).

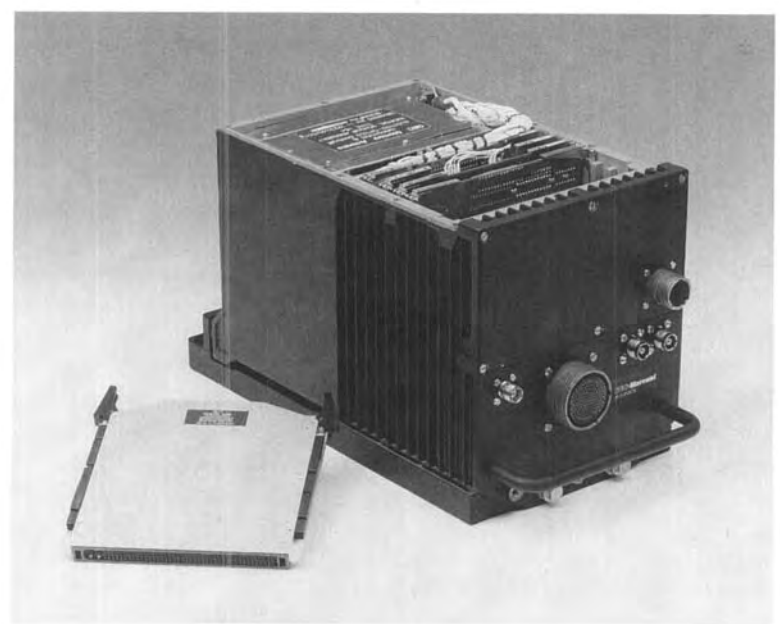


Figure 13-4. FIN 3110G ring laser gyro inertial navigation system with embedded GPS. Stand-alone performance <1 nm/h CEP. Weight 10 kg; dimensions $18 \times 18 \times 28$ cm. (Photograph courtesy of GEC-Marconi).

Another important field of application of mechanical sensors is that of structure monitoring; the concept of preventive maintenance requires monitoring of all parts of an aircraft which can fail due to fatigue, degradation, or parameter drifts. The airframe and other structural parts of the aircraft have to be equipped for this purpose with strain sensors at critical positions. For composite structures the use of fiber-optic sensor systems which are laminated into the structure and which allow for continuous monitoring of strain or even the development of cracks within the structure seems promising. Such sensors can be adapted to different shapes of composite structures and do not require the precautions which are necessary for other crack detection concepts, eg the use of pressurized cells within the structure and detection of crack-induced pressure loss.

Sensor systems of this type will be of increasing importance for structural components subject to extreme load cases and their development for airborne applications will therefore be intensified [19].

Developments in the field of thermal sensors for aerospace applications are strongly influenced by the need for more accurate measurement of temperatures of critical engine components in order to allow operation as close as possible to the optimum level [9]. The measurement of turbine blade temperatures is one example which seems to be of particular importance. Such measurements cannot be performed by contacting methods but only by remote sensing of temperature radiation using optical radiometers. The development of self-calibrating radiometers, compact and rugged enough to be used in the harsh environment of an aircraft engine, is an important requirement.

Aviation systems consist of a large number of sensors for the measurement of electrical parameters associated with communication, radar and navigation systems and also the simple sensor devices necessary for monitoring electrical operating parameters (see also Tables 13-5 and 13-6).

In particular, new developments can be expected in the fields of sensor instrumentation for navigation systems, airborne data communication links [8], and sensor systems for military applications.

The increasing air traffic density makes it necessary to improve the accuracy of navigation equipment and in particular the approach/landing support instrumentation. The existing instrument landing system (ILS) is proposed to be replaced by a microwave landing system (MLS) which offers significantly improved accuracy and would therefore allow the handling of higher traffic densities. This presupposes an upgrade of the ground-based air traffic control (ATC) installations to accomodate MLS.

Another potential approach is the improvement of the aircraft on-board navigation systems to give higher accuracy in combination with increased autonomy during all flight phases (including approach and landing). The progress made in satellite supported navigation systems, radar, and optical instrumentation supports this strategy. Such instrumentation could combine GPS, millimeter-wave radar and optical infrared sensor systems similar to those used for military night operations. In particular, the independence from ground-based ATC instrumentation makes this concept attractive for the user. However, in the field of electrical sensors development has to be devoted to airborne millimeter-wave radar with phased array antennas and to the improvement of operational reliability of GPS navigation systems in terms of coverage and guaranteed accuracy. The development of airborne digital communication systems for flight management and passenger communications (voice, data) will also require the upgrade of existing RF installations [8, 10].

The use of optical sensors in civil aircraft instrumentation will also be of importance; examples are the application of infrared sensors for advanced autonomous approach/landing systems (as mentioned above), the use of TV cameras for various monitoring functions (runway observation, cockpit monitoring, etc.), and the optical sensors associated with airborne fiber-optics bus concepts. Not all these applications will require fundamental developments but the adaptation of existing ground-based equipment or military aviation instrumentation to the needs of civil aviation. The aspects of commericalization (affordability) will play an important role in particular for equipment derived from military systems [20].

Payload Sensor Instrumentation

The most important aviation payload types with special sensor instrumentation are listed in Table 13-2. Of these the remote sensing instrumentation will be discussed as an example of the trends in airborne sensor development.

Remote sensing instrumentation has a wide range of applications in the military and in the civil field for land, ocean, and atmospheric monitoring [21-24]. It is of growing importance in particular in view of the needs in the area of environmental monitoring and for surveillance task.

There are two types of remote sensing instruments: active, which transmit signals towards a target and analyse the signal returned from it, and passive, which register only the signals emitted or reflected by the target. Typical examples of these are:

- active:
  - radar;
  - LIDAR;
- passive:
  - radiometer;
  - scanner;
  - metric cameras.

In the range of active remote sensing instrumentation, radar systems such as synthetic aperture radar (SAR), side-looking airborne radar (SLAR), or the radar altimeter are complemented increasingly by LIDAR (light detection and ranging). LIDAR uses short laser-light pulses to obtain information on the composition and physical state of the atmosphere with an unprecedented vertical/horizontal resolution, which is of particular importance for the analysis of aerosol layers, humidity profiles, and wind field structures. Important application areas of airborne LIDAR systems include the analysis of atmospheric constituents, the detection of pollutants, and the identification of their sources.

Depending on the nature of the optical interaction one can distinguish backscatter, differential absorption (DIAL), Doppler wind, fluorescence and Raman LIDAR. The systems operate with lasers from the mid-infrared wavelength range (CO_2 laser 10.2 μm) to the UV-VIS range using frequency-multiplied Nd:YAG and Ti:sapphire lasers [25]. Photomultipliers (VIS), Si-avalanche photodiodes (VIS to NIR) or HgCdTe PIN diodes (IR) are used in LIDAR systems, either as single-element detectors or as arrays for compensation of speckle effects.

LIDAR, in combination with infrared radiometers and supplementary meteorological measurement equipment, is a key instrument for the advanced environmental monitoring of the atmosphere and its constituents. Future developments will aim for small size and high operational flexibility combined with high detector sensitivity and tunable laser sources to enable a wide range of measurement tasks.

Passive remote sensing instruments are used to detect radiation which is either emitted or reflected from the atmosphere or from the ground. Radiometer instruments probe the target area pointwise and measure the received radiation in well defined wavelength ranges.

Scanners cover larger field of views with a defined spatial resolution and use either optomechanical scan units or optoelectronic devices such as linear detector arrays (push-broom scanner). Push-broom scanners use the linewise detection of the image combined with the platform flight speed to obtain the along-track image dimension [26]. The field of view and spatial resolution requirements of the focal detector plane which may consist of a single line detector array (up to 4096 pixels are commercially available today) or may be composed of several arrays which are optically combined. These scanners operate either panchromatically or over well defined wavelength ranges using spectral filters in front of the optical system; typical spectral bandwidths for such channels are of the order of 40–100 nm [27].

Another scanner type is the imaging spectrometer, which registers simultaneously an image in multiple spectral bands with high spectral resolution. Such instruments, which also use the push-broom principle, employ a post-focal spectrometer to disperse the spectral information of the image. The spectrum of the image line is registered on a two-dimensional array in which the cross-track dimension provides the spatial information of the ground track and the second dimension the spectral information of the same area. Again, the along-track dimension is gained by the flight velocity of the aircraft.

The detectors used for such imaging spectrometers are two-dimensional CCD arrays. Depending on the spectral and spatial resolution requirements, a wide range of detector dimensions are used from 256×256 to more than 1024×1024 pixels. In some cases a combination of elements which provide effective focal planes with more than 4500 pixels in the spatial domain and 576 pixels in the spectrum are employed, as for example in ESA's MERIS payload to be flown in the frame of the ENVISAT-1 space mission. The MERIS image can be formed from 15 freely selectable spectral bands from 400 to 1050 nm with spectral resolutions between 2.5 and 25 nm [26, 28, 29].

Imaging spectrometers have an interesting potential for airborne remote sensing systems that derives from their operational flexibility. Developments will be performed in connection with this instrument type in order to enhance the sensitivity in the UV region by using thinned back-side illuminated large area CCDs and to provide more powerful data evaluation systems.

The third type of passive airborne remote sensing instrument is metric cameras for photogrammetry. These cameras, which are used today on stabilized platforms with forward motion compensation, use photographic film as the recording medium. Apart from their traditional application for aerial photogrammetry, they are widely used for high-speed, low-altitude reconnaissance, a task which is difficult to perform using push-broom scanner-type instruments because at a given spatial resolution, the integration times of the linear detector arrays are too short at high flight velocities to accumulate sufficient energy per pixel [30].

13.4.2 Sensors for Space Systems

The basic sensor types used on-board a space platform are summarized in Table 13-7 for the case of a satellite, the table shows the various satellite subsystems with its units and the associated sensors.

Table 13-7. Major satellite subsystems, units and typical associated sensors.

Satellite subsystems	Units	Associated sensors*
Spacecraft structure and mechanisms	Structure assemblies	Strain gauges[1] Micro-accelerometers[1]
	Hold-down/deployment mechanisms	Position sensors Voltage/current sensors
Telemetry, tracking, and command	Antennas	RF sensors Thermistors
	Receiver and transmitter	RF sensors Housekeeping sensors[2]
	Data handling unit	Housekeeping sensors
	Telemetry unit	Housekeeping sensors
Attitude and orbit control (AOCS)	Control electronics	Housekeeping sensors
	Attitude sensors	Earth sensors Sun sensors Star sensors Magnetometers Gyroscope package Accelerometers Housekeeping sensors
	Actuators	Tachometers Position sensors Pressure transducers Thermistors Housekeeping sensors
Propulsion	Control electronics	Housekeeping sensors
	Tanks	Fuel level sensors Position sensors Thermistors Pressure transducers
	Propellant control assembly	Position sensors Pressure transducers
	Apogee engine and auxiliary attitude control thrusters (see also AOCS actuators)	Position sensors Pressure transducers Thermistors
Solar array	Solar array wings	Current/voltage sensors Thermistors

Table 13-7. (continued)

Satellite subsystems	Units	Associated sensors*
	Solar array drive mechanisms	Position sensors Thermistors
Electrical power supply	Power conditioning/control electronics	Housekeeping sensors
	Batteries	Current/voltage sensors Thermistors
Thermal control	Surface coatings and insulations	Thermistors
	Heat pipes	Thermistors
	Heaters	Current/voltage sensors Thermistors Thermostats
	Radiators	Thermistors Thermostats
Payload	Payload specific units	Payload specific sensors

* 1. Non-standard equipment, in experimental stage.
2. Housekeeping sensors are used in electronics boxes to monitor the basic electrical parameters (eg supply voltages, currents) and the operating temperatures.

There is a wide variety of spacecraft types and design concepts, each with specific conceptual and technological aspects that depend on the needs of the mission [31]. However, certain general development trends can be identified which influence the design of sensors and their technologies for this application. Such general trends are, eg,

- improvement of space platform pointing accuracy and stability;
- enhancement of operational autonomy and flexibility;
- improvement of platform-to-payload mass ratio and power consumption ratio;
- extension of in-orbit lifetime;
- reduction of procurement/operational costs.

Development trends specific to manned space missions, rendezvous-docking operations, and large space structures include:

- improvements of spaceborne navigation and guidance systems;
- development of active control systems for large structures in space;
- development of space robotics for extra- and intravehicular activities (EVA, IVA) and planetary explorations;
- development of environmental control and life support systems for long-duration manned space missions.

For the development of commercial communication satellites one aims for:

- extension of platform lifetime towards 15 years in geostationary orbit;
- reduction of platform mass in favor of an increased payload complement;

- modular design concept with as far as possible standarized building blocks;
- increase system flexibility and operational autonomy by implementation of advanced digital attitude and orbit control systems (AOCS) (adaptive control systems, multivariable and robust control, adaptability to various mission scenarios);
- improved platform pointing accuracy and stability for better beam pointing capability;
- affordability.

The influences of these trends on specific fields of sensor and sensor system development are discussed below (see also Ref. 32).

13.4.2.1 Spacecraft and Platform Sensors

(a) Attitude Measurement Sensors

The two basic methods for spacecraft attitude stabilization are spin stabilization and three-axes stabilization, both of which have a number of technical variants [31]. The latter concept is most widely used today by virtue of its flexible payload accomodation capability, its maneuverability and the achievable pointing accuracy/stability. The requirements for higher pointing accuracies and stability are stimulated by the development of increasingly high-resolution remote sensing payloads, free space laser communication systems, and advanced scientific payloads. ESA satellites launched in the early 1970s were, for example, designed for accuracies of about 1 °, whereas the requirements for today's scientific satellites are in a range below 0.01 °. The trends in the pointing accuracy requirements for NASA spacecraft are even more stringent and reach well below 0.001 ° [32].

The advanced attitude and orbit control systems currently developed for this purpose are microprocessor based and use "smart" sensors and actuators allowing offloading of the central processor. The systems are designed to perform autonomous failure detection and error checking, have high fault tolerance and also allow implementation of adaptive control concepts.

The sensors which are used as building blocks of such systems are strapped-down inertial navigation units (INUs) which comprise gyroscopes and accelerometers, optical sensors and, for certain spacecraft types, magnetometers. The gyroscopes currently used in strap-down INUs are mostly mechanical gyros. There are, however, tendencies to qualify ring laser gyros and fiber-optic gyros which will offer lower drift rates and, probably, longer in-orbit lifetime [33] (see also Section 13.4.1). Figure 13-5 shows the RLG-based guidance system used on the European launcher Ariane 4. A rate sensor system often used for communication satellites and scientific missions is shown in Figure 13-6.

The INU attitude information is often updated with data obtained from optical sensors, eg sun sensors, earth sensors, or star sensors (see Figures 13-7 to 13-9). The optical sensors which operate in the visible wavelength range use CCD matrix array detectors to provide two-axis attitude information per sensor head. The sensors are microprocessor controlled and provide digital output data to the AOCS control electronics.

Star sensors usually have multiple star acquisition/tracking capability and can therefore be used, in connection with a digital star catalogue, for fast orientation in space by comparing the detected star patterns (relative positions and magnitudes) with the catalogued data. In a simplified configuration, star sensors can also be used for spacecraft attitude recovery in cases

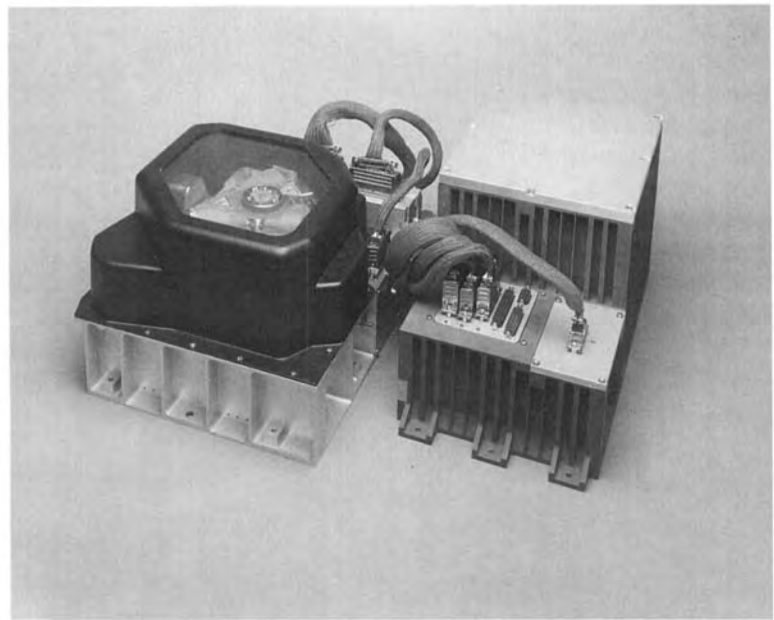


Figure 13-5. Ring laser gyro guidance system for Ariane 4 launcher. Inertial measurement unit and electronics unit. First flown on Ariane V69 on 31 October 1994. (Photograph courtesy of GEC-Marconi).

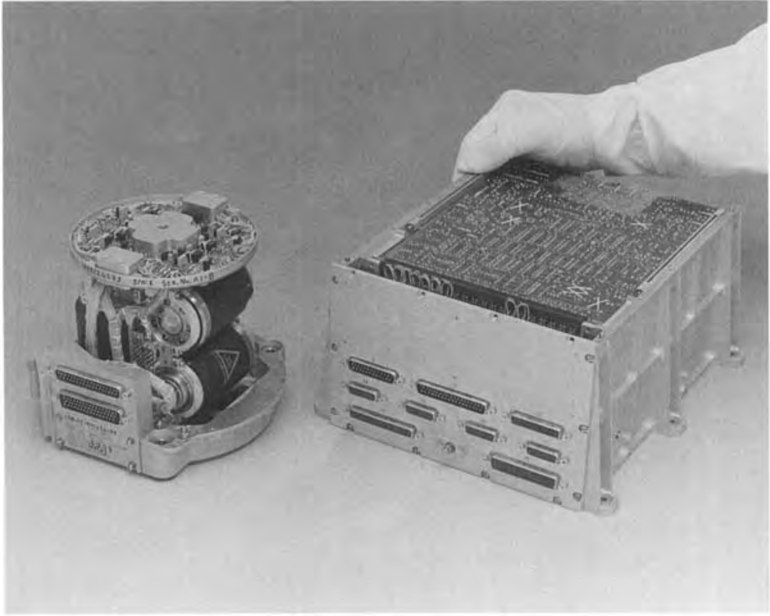


Figure 13-6. Rate sensor system, for communication and scientific satellites, using four Type 125 floated rate integrating gyroscopes. (Photograph courtesy of GEC-Marconi).

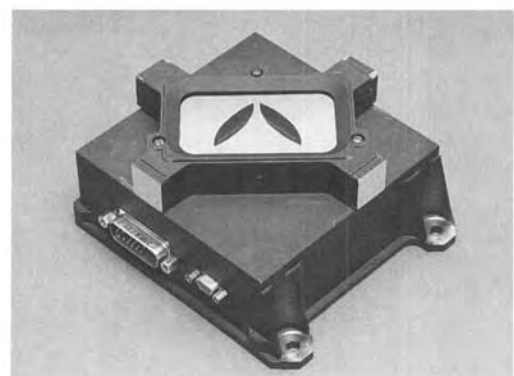


Figure 13-7.
Sensor head of second-generation precision sun sensor. (Photograph courtesy of DASA Satellite Systems/Jena Optronik).

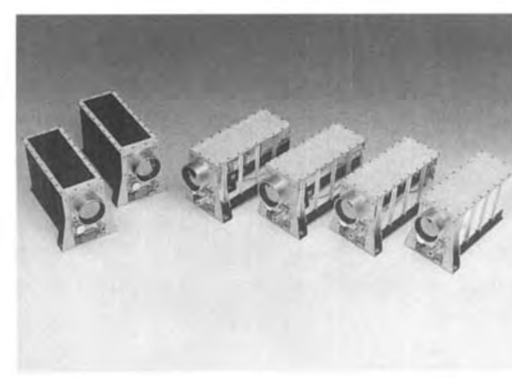


Figure 13-8.
Infrared Earth sensors. (Photograph courtesy of DASA Satellite Systems).

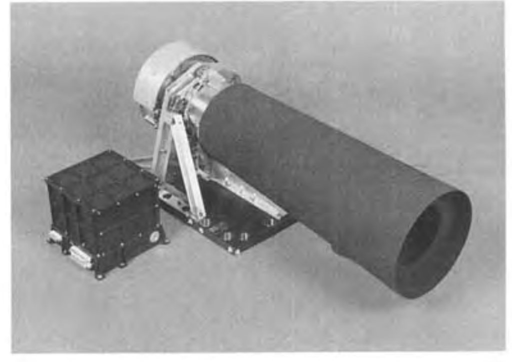


Figure 13-9.
Star sensor optical head with stray light baffle and electronics box. (Photograph courtesy of DASA Satellite Systems/Sira Ltd.).

of attitude loss. The sensors used for this purpose are in principle wide-field star sensors with moderate accuracy designed to detect bright stars passing their field of view. The measured target object positions are used for reorientation within typically less than 1 h without the need to rely on proper gyro operation, compared with at least a 12 h reorientation time if conventional schemes are used [34]. Such a "Fast Recovery Sensor" is flying successfully since July 1988 on ESA's ECS 5 communication satellite (Figure 13-10).

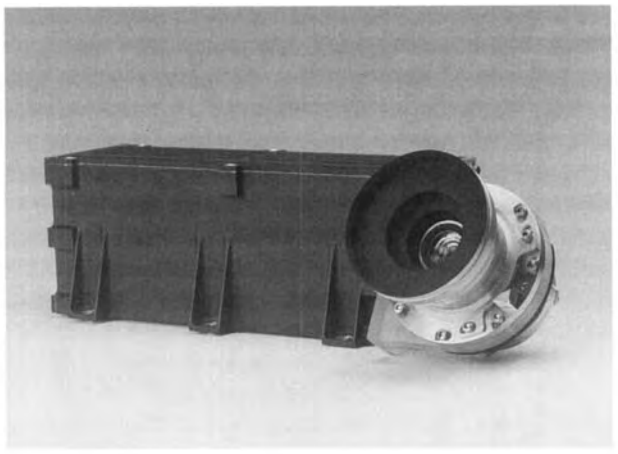


Figure 13-10.
Fast recovery sensor for ECS-5. (Photograph: ESA).

Current developments in optical AOCS sensors are aimed in general at miniaturization combined with high accuracy at moderate cost. A concept allowing for significant miniaturization of optical AOCS sensors is the use of CCD matrix arrays in connection with a special type of short focal length array optics [35]. Miniature sensors of this type can be designed for a large field of view and can provide digital two-axis attitude information with good accuracy.

(b) Navigation Sensors

Some space missions require, in addition to high-accuracy attitude control, precise trajectory control and maneuverability (six degrees of freedom) for execution of rendezvous-and-docking operations (see below). For this purpose, advanced guidance, navigation, and control concepts have been developed which combine INUs with optical sensors, satellite-supported navigation systems (eg GPS/differential GPS, GLONASS), and trajectory propagation. The sensors used for these navigation systems are basically the same as described above for the AOCS (gyros, accelerometers, magnetometers, optical sensors). In the future they will, however, be complemented increasingly by the RF sensors which are part of satellite-supported navigation systems.

(c) Rendezvous–Docking (RVD) Sensor Systems

Rendezvous–docking sensors/sensor systems are important for the in-orbit assembly and operation of space stations. These sensors have to detect the partner spacecraft within a given

uncertainty cone and provide information on its position, orientation, range, and velocity vectors. A combination of optical and radar sensors are used in this application. Within the former category laser range finders, laser radar, or image-processing video systems for both cooperative and non-cooperative targets are exploited. Radar sensors operate over longer distances mainly with non-cooperative targets and at wavelengths down to millimeter waves.

(d) Sensors for Stabilized Platforms

The need to accommodate several payloads on one spacecraft and the necessity to minimize, in certain cases, their mutual influence in terms of induced vibrations requires the development of stabilized platforms which provide active vibration damping over a wide bandwidth. This is of particular importance for earth observation systems, free space laser communication systems and some scientific payloads. Sources of disturbances include reaction wheels, microwave switches, antenna pointing mechanisms, solar array drive motors, and low-frequency structural deformations.

The sensors required for such stabilized platforms have to measure the precise attitude of the platform in space and the vibrations induced via the structural coupling with the spacecraft. The sensors used for this purpose are in most cases star sensor packages, gyroscopes, and accelerometers. The sensor information is used to derive the control signals for the actuators which provide support and damping of the platform.

(e) Sensors for Smart Structures

The following applications of smart structures in space can be identified [32]:

- structural response control;
- active shape control;
- in-orbit health monitoring;
- composite cure control.

The stabilization and control of spacecraft with large flexible structures or large antennas requires the development of active control systems for, respectively, structural vibration damping or shape control. Various sensing principles can be used for such purposes; smart sensor/actuator packages can be taken, for instance, which are distributed over the platform to measure its vibrations and to perform active damping. The sensor element may consist of accelerometers, optical sensors, or optical fiber sensors for strain measurements [36].

For the shape control of large antennas, for instance, sensor systems can be used which are based on the principles of triangulation, precision laser ranging, or image processing. Alternatively, sensors which are embedded in the structure, such as optical fiber sensors for strain measurement or displacement sensors may be used.

Another area is that of structural health-monitoring, which is of particular importance for manned systems. Structural damage and degradation can be detected in these cases using optical fiber sensors.

The cure control of composites is also an application area of optical fiber sensors [36] which are used in situ to control the curing status. Such sensors can enable critical process parameters to be optimized.

(f) Life-Support Sensor Technology

Part of the platform sensor instrumentation for manned space flight is the sensors of the life support system on-board the spacecraft. Sensors are used here as part of the environmental control and atmosphere management system. This consists of gas sensors, temperature sensors, humidity sensors, radiation monitors, pressure sensors, and systems for trace gas analysis, eg gas chromatographs, Fourier transform spectrometers or mass spectrometers. Such sensors can be designed today in the form of highly integrated microsensor devices using emerging microsystem technologies [37, 38].

In addition to the above applications, sensors are used as part of space platforms for the detection of space debris and micrometeorite impacts; such sensors consist, for instance, of microaccelerometers mounted on impact shields which measure the shocks caused by impacting particles [39]. Figure 13-11 shows a solid-state microaccelerometer which can be used for these purposes.

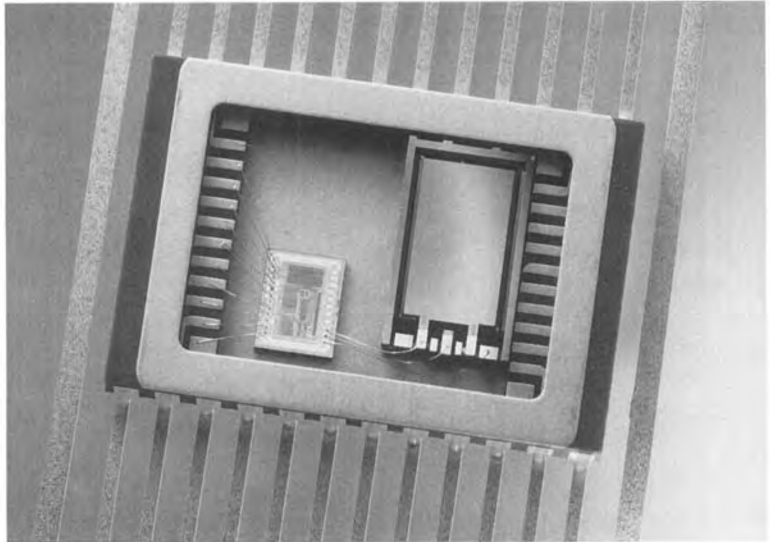


Figure 13-11. Solid-state microaccelerometer with CMOS ASIC interface circuit, mounted in a flatpack. (Photograph courtesy of CSEM).

13.4.2.2 Payload Sensor Instrumentation

Four distinct types of civil applications of space payload instrumentation can be identified:

- communication;
- space science;
- Earth observation;
- microgravity instrumentation.

Launch vehicles, manned space stations, and transportation systems in their function as carriers and platforms support the payloads in execution of their operational tasks. The sensor

types used within these systems are therefore covered as part of the platform sensor technology discussed above.

Space communication systems (see Figure 13-12) provide a variety of services such as links between fixed ground stations via one or more satellites (fixed satellite service), or mobile ground terminals via one or more satellites (mobile satellite service), audio/video broadcasting transmitters for direct/community reception (broadcasting satellite service), data transmitters for payload-generated data or telemetry, tracking and command (TT&C) data (payload satellite service), transmitters for satellite-based radio-navigation systems (radio-determination satellite service), and intersatellite communication systems (inter-satellite service) [40].

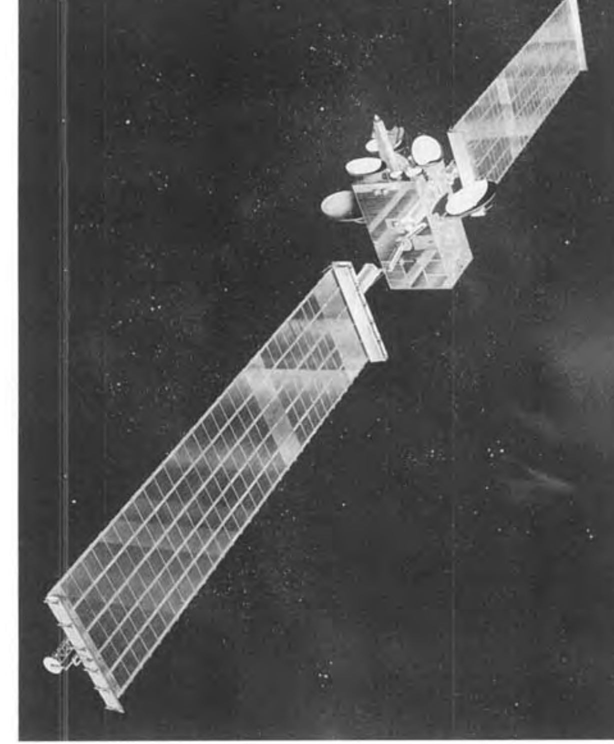


Figure 13-12. Artist's impression of European multi-purpose telecommunications satellite Olympus. One of the major purposes of this ESA satellite is direct-to-home television broadcasting. Other applications include European specialized services, intercontinental and European trunk services, integrated mobile communications and integrated services for developing countries. OLYMPUS was launched with the European Ariane 4 on 12 July 1989. (Photograph: ESA).

Antennas with associated detectors for electromagnetic radiation act as sensors that receive RF communication signals. In the case of free space laser communication the sensors consist of telescopes with optoelectronic detector elements, such as avalanche photodiodes (APDs). Current developments in the RF field include large antenna structures in which the shape of the reflector is actively controlled, various types of array antennas, large multibeam antennas for hand-held mobile communication systems, and antennas with frequency- and/or polarization-selective surfaces [32].

An important future area will be that of global navigation satellite systems (GNSS) and the related RF sensor technology.

Optical communication systems for space applications operate in the near-IR wavelength range (eg 830 nm, 1.064 µm) and are mainly used for inter-satellite (GEO-GEO, LEO-LEO) and inter-orbit (LEO-GEO) communication. Current studies of system configuration are focused on small terminals with telescope diameters in the range 3-10 cm. In addition to the communication signal sensors (telescopes with APD), optical payloads also have sensors for acquisition and tracking of the counter terminal during communication. Acquisition sensors are usually designed using CCD matrix arrays (eg, pixel numbers of 288 × 356), in order to provide a field of view (FOV) sufficiently large for reliable detection of the counter terminal. Tracking sensors are required for active beam stabilization. In order to provide a sufficiently high read-out frequency, they use either small-format CCD matrix arrays (eg, 14 × 14 pixels), or four-quadrant APDs with simultaneous analogue read-out. Examples of acquisition and tracking sensors developed for use on-board ESAs free space laser communication system SILEX are shown in Figure 13-13 [41, 42].

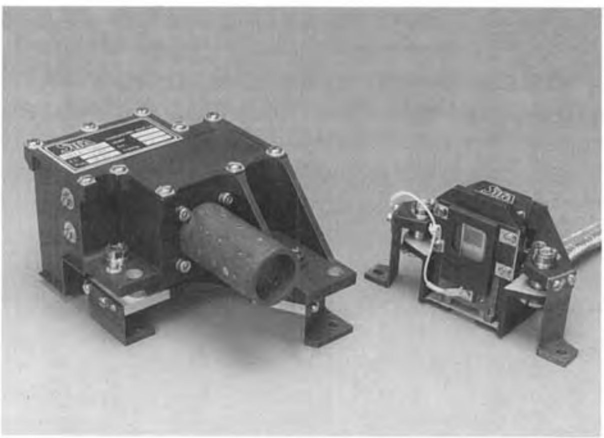


Figure 13-13.
Acquisition sensor detector unit (right) and tracking sensor detector unit (left) developed for the European free space laser communication system SILEX. (Photograph courtesy of Sira Ltd.).

Sensor systems for space science payloads encompass a wide range of types and technologies covering the entire electromagnetic spectrum from X-ray to the radiofrequency range. Depending on the mission tasks, also sensors for the detection of ionizing particles and plasma, and the measurement of magnetic fields together with the sensor systems of autonomous planetary surface landing probes may constitute part of scientific payloads.

The payloads of scientific satellites often dominate the entire spacecraft design; Figure 13-14 shows the Infrared Space Observatory (ISO), a scientific satellite of the European Space Agency devoted to infrared astronomy in the wavelength range 2.5-200 µm. The scientific payload, an infrared telescope with associated focal plane instrumentation, is housed in a cryostat which is cooled with liquid helium.

The ESA Scientific Programme Plan "Horizon 2000", presented in 1985 and running to the year 2007, defines the scientific mission elements currently planned at ESA [43]. This plan gives an indication of the current developments for scientific payload sensors and instrumentation. Horizon 2000 includes, besides low/medium-cost missions, four cornerstone missions:

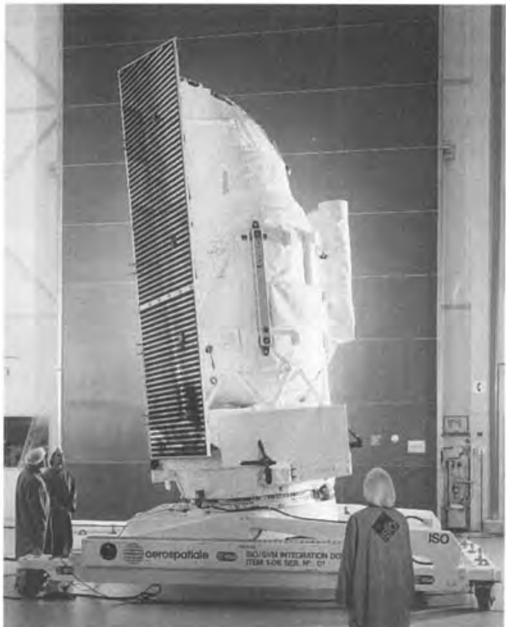


Figure 13-14.
The European Infrared Space Observatory (ISO) during test preparation at the European Space Agency (ESA) test facilities. Solar panels (left), cryostat with telescope and scientific instrumentation (center), star sensors (right), all attached to the service module at the lower end of the spacecraft. The spacecraft is mounted on an integration dolly for test preparation. The launch date is 19 September 1995. (Photograph: ESA).

1. Solar and terrestrial physics: SOHO, CLUSTER.
2. High-throughput X-ray astrophysics: XMM.
3. Sub-millimeter heterodyne spectroscopy: FIRST.
4. Comet nucleus sample return: ROSETTA.

Table 13-8 shows the sensor types used within the various payloads of the ESA space science program cornerstone missions.

Earth observation payloads are used for monitoring of the Earth and its environment [32, 49]. They include payloads for

- land applications;
- ocean and ice applications;
- atmosphere applications;
- solid Earth applications.

Land applications are of interest for agriculture, forestry, topographic mapping, land use/inventory, geology, hydrology, and monitoring of vegetation. The instrumentation used for this purpose combines medium to high resolution imaging/spectral imaging systems operating in the visible to infrared regions and synthetic aperture radar (SAR) to provide enhanced all-weather capability (see Figure 13-15).

Ocean and ice applications include the monitoring of wind and wave fields, measurement of sea surface temperatures and currents, ocean color monitoring, and the detection and classification of ice surfaces. Optical imagers and spectral imagers with low to medium spatial

Table 13-8. Payload instrumentation and sensors used in ESA Horizon 2000 cornerstone missions.

Cornerstone mission	Payload instruments/experiments	Sensor types and technologies
SOHO [44]	Global oscillations at low frequencies (GOLF)	Photomultiplier
	Variability of solar irradiance (VIRGO)	Active cavity radiometer Si photodiodes Thermistors
	Michelson Doppler imager/solar oscillations monitor (MDI/SOI)	CCD camera
	Solar ultraviolet measurements of emitted radiation (SUMER)	Multi-anode microchannel array (MAMA) detector
	Coronal diagnostic spectrometer (CDS)	Microchannel plate detectors
	Extreme-ultraviolet telescope (EIT)	CCD thinned back illuminated
	Ultraviolet coronograph spectrometer (UVCS)	MAMA detector Photomultiplier
	White light and spectrometric coronograph (LASCO)	CCD
	Solar wind anisotropies (SWAN)	Microchannel plate detectors
	Charge, element, and isotope analysis system (CELIAS)	Microchannel plate detectors Solid-state ion detectors
	Comprehensive suprathermal and energetic particle analyzer (COSTEP)	Low-energy detector Medium-energy ion composition analyzer High-energy detector Low-energy ion and electron instrument Electron-proton helium instrument
	Energetic particle analyzer (ERNE)	Low-energy detector High-energy detector
CLUSTER [45]	Fluxgate magnetometer (FGM)	Fluxgate sensors
	Spatio-temporal analysis of field fluctuations (STAFF)	Search-coil magnetometers
	Electric fields and waves (EFW)	Electric field probes
	Digital wave processor (DWP)	Uses sensors of WHISPER
	Waves of high frequency and sounder for probing of electron density by relaxation (WHISPER)	Probes for detection of plasma resonances
	Wide band data (WBD)	Uses sensors of EFW
	Electron drift instrument (EDI)	Electron detectors
	Plasma electron and current analyzer (PEACE)	Electron detectors
	Cluster ion spectroscopy (CIS)	Ion detectors
	Research with adaptive particle imaging detectors (RAPID)	Position sensitive solid-state detectors

Table 13-8. (continued)

Cornerstone mission	Payload instruments/experiments	Sensor types and technologies
XMM [46]	European photon imaging camera (EPIC)	X-ray CCD detectors
	Reflection grating spectrometer (RGS)	X-ray CCD detectors
	Optical monitor (OM)	Microchannel plate intensifier with CCD detectors CCD detectors
FIRST [47]	Multi-frequency heterodyne receiver (MFH)	Schottky diodes
	Far-infrared instrument (FIR)	Ge:Ga BIB photoconductive detectors Bolometric detectors
ROSETTA [48]	In situ imaging system	CCD cameras
	Infrared mapper	Tantalum pyroelectric detector
	Mass spectrometers	Photodetectors
	Thermal monitors	Thermistors
	Borehole stratigraphy instrument	IR detectors CCD cameras
	IR spectral mapper	HgCdTe or InSb detectors InSb CCD arrays
	Radar sounder	Antenna
	Remote imaging system	CCD cameras
	Radar altimeter	Antenna
	Dust counter	Piezoelectric detectors Electrostatic charge sensors Quartz crystal balance
	Test penetrator	Thermistor Accelerometer
	Laser scanner and range finder	Photodetector

resolution that operate in the visible to infrared regions are combined here with radar systems and IR/microwave radiometers.

The atmospheric applications of space-borne Earth observation payloads combine the needs of meteorology and climatology. Infrared radiometer and visible/infrared imaging systems for global Earth monitoring used currently in meteorological satellites will be complemented in the future by payloads with improved spectral/spatial resolution and instrumentation for monitoring of aerosols, the atmospheric boundary layer, cloud distribution, and global wind field monitoring. In addition to SAR, atmospheric LIDAR payloads and sounding systems will be developed for these tasks. Figure 13-16 shows the meteorological satellite MOP-2, which is equipped with a radiometer for observation of the Earth in three spectral bands: 0.5–0.9 μm (visible band), 5.7–7.1 μm (infrared water vapor absorption band) and

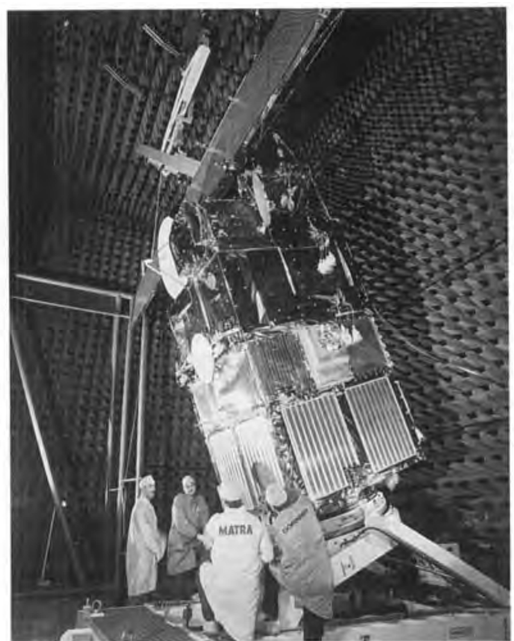


Figure 13-15.
The ERS-1 flight model during tests with its microwave antennas fully deployed. ERS-1 is a remote sensing satellite with various microwave payloads (microwave wind scatterometer, synthetic aperture radar, radar altimeter, microwave sounder) which provide all-weather Earth remote sensing capabilities. The spacecraft was launched on 17 July 1991 on an Ariane 4 launcher and has operated successfully in orbit since then. (Photograph: ESA).

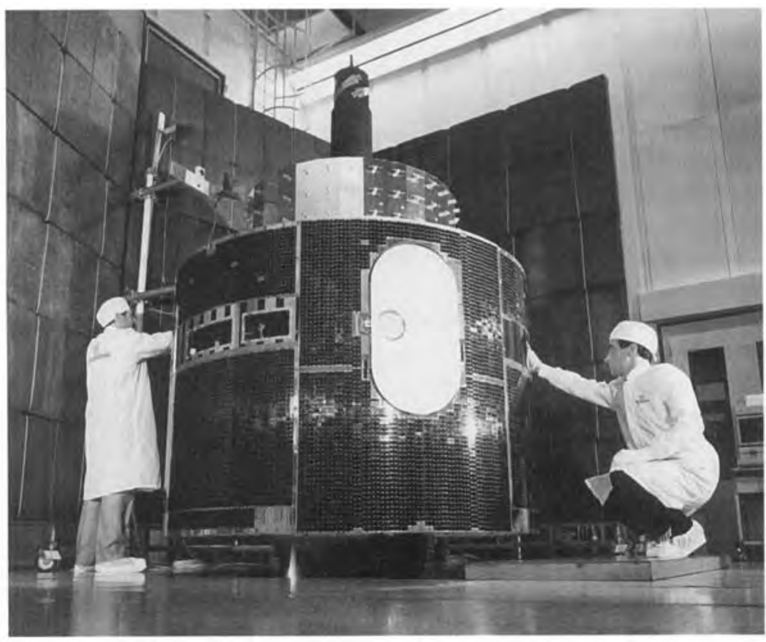


Figure 13-16. Flight unit of the European meteorological satellite MOP-2 during integration. The white oval door in the cylindrical body covers the opening aperture of the radiometer. This second spacecraft of the Meteosat operational system was launched on 1 March 1991 by an Ariane 4 launcher. (Photograph: ESA).

10.5- 12.5 μm (thermal infrared window band). The spatial resolution is 5 km in the infrared and 2.5 km in the visible band.

The payload instrumentation for solid Earth applications includes systems for measurement of the Earth's gravitational field, the geoid, the magnetic field, or the tectonics. The instruments used for this purpose are precision radar altimeters, gradiometers, LIDAR systems, GPS/differential GPS and very long baseline RF interferometry systems.

Table 13-9 summarizes the various parameters which are monitored with space-borne Earth observation systems and identifies the instruments used for this purpose [32].

The payload developments currently underway at the European Space Agency in this field and the associated sensors/detectors are summarized in Table 13-10.

Payload instrumentation dedicated to microgravity experimentation can be classified into the following two groups:

- life science:
  - general biology;
  - applied biology;
  - human physiology;
- material science:
  - high-temperature material science;
  - low-temperature material science;
  - fluid science.

Life science investigates the influences of the space environment on living organisms with the primary aim of gaining engineering data sufficient for the definition of the life support systems required for manned space missions.

Payloads devoted to life science experiments are designed for use on-board space stations or recoverable satellites. The sensor instrumentation within such payloads has to monitor biological and physiological functions, such as temperature, humidity and atmospheric gas composition, fluid composition, mass changes, and complex biochemical reactions (see also Figure 13-17). In the field of human physiology the sensors have to perform biochemical analyses of body fluids, measurement of body mass/fluid distribution, cardiac and pulmonary system monitoring and neurophysiological measurements. A wide range of novel microsensors in Si technology are under development for this purpose which combine the sensing elements with on-chip signal conditioning and evaluation [37, 38].

Automated work stations are being developed in order to relieve the crew of a space station from time-consuming routine operations associated with biological experiments and to prepare for the use of unmanned platforms. Such automated work stations have, for example, to prepare and preserve specimen, embed and section, harvest and reseed, sample and initiate cultures, separate insect generations, cultivate aquatic animals, and visualize and respond to developmental cues in experimental samples. The sensors required for the handling and analysis systems of such work stations will consist of, in addition to the above-mentioned sensors for biological sample monitoring, a wide range of sensors for control of the handling system, eg video cameras with image-processing systems for pattern recognition and motion control, mechanical sensors, and temperature sensors. Figure 13-18 shows a microaccelerometer which can be used for monitoring of accelerations during microgravity experimentation.

Table 13-9. Observation parameters and related payload instrumentation [32].

Observation parameter	Payload instrument types
Land and ocean surface	Microwave radiometer
	Synthetic aperture radar (SAR)
	Thermal imager
	Imaging spectrometer
	Metric camera
Clouds	Microwave radiometer
	Cloud radar
	Thermal imager
	Imaging spectrometer
	Backscatter LIDAR
Water vapor	Microwave radiometer
	Thermal imager
	Differential absorption LIDAR
Radiation	Imaging spectrometer
	Thermal imager
	Backscatter LIDAR (for atmospheric correction [50])
	Limb sounder
Waves	Microwave scatterometer
	Synthetic aperture radar (SAR)
Ice	Synthetic aperture radar (SAR)
	Microwave radiometer
	Laser altimeter
Atmospheric constituents	Differential absorption LIDAR
	Limb sounder
	High resolution spectrometer
Aerosols	Backscatter LIDAR
Temperature	Microwave radiometer
	Thermal imager
	Backscatter LIDAR (for atmospheric correction [50])
Wind	Doppler wind LIDAR
	Microwave scatterometer
Gravity field	Gradiometer
Ocean geoid	Radar altimeter
	Laser altimeter
Magnetic field	Magnetometer
Tectonic motions	Global positioning system (GPS)
	Very long baseline interferometry (VLBI)
	Laser altimeter
	Laser range finder
Rainfall	Microwave radiometer
	Rain radar

Table 13-10. Earth observation payload instrumentation and sensors used.

Payload instruments	Related sensor types and technologies
Microwave radiometer	Dish antenna (nadir looking)
Microwave/sub-millimeter limb sounder	Shaped reflector antenna with heterodyne receiver (Schottky diode)
GPS [17]	Antenna
Radar	
– Synthetic aperture radar (SAR)	Active antenna (slant angle looking)
– Rain radar [51]	Antenna (eg semi-active, slant angle looking)
– Cloud radar	Antenna
– Scatterometer	Antenna (slant angle looking in three dimensions)
– Radar altimeter	Dish antenna (nadir looking)
LIDAR [52, 53]	
– Backscatter LIDAR (diode-pumped Nd:YAG laser) [54, 55]	Telescope with Si-APD or PMT (incoherent detection)
– Doppler wind LIDAR (CO_2 laser) [56]	Telescope with CdTe PIN diode array (coherent detection)
– Differential absorption LIDAR (Tm:Ho:host laser)	Telescope with InGaAs heterodyne receiver
Laser altimeter	Telescope with Si-APD (nadir looking)
Laser range finder	Telescope with Si-APD (field scanning)
Thermal infrared imager	IR CCD arrays
Imaging spectrometer [26, 27]	CCD arrays
Metric camera	Photographic film
Magnetometer	Magnetometer coil
Very long baseline interferometer (VLBI)	Dish antennas
Gradiometer	Accelerometer package Superconductive SQUIDs

Material science investigations are focused on the physical and chemical properties of materials under microgravity environment and material processing methods suitable for applications in space.

The sensors required for material science payloads consists of sensors for process control, in situ diagnostics, and characterization. The conventional sensors used for these purposes are complemented increasingly by microsensors similar to those for life science applications mentioned above. In addition, sensors for handling systems similar to those described for life science instrumentation are required. Such sensors include temperature sensors for temperature ranges from cryogenic temperatures to more than 2000 °C (eg thermistors, thermocouples, radiation temperature sensors, sapphire optical fiber thermometers). Optical monitoring systems such as interferometers, scatterometers, and spectrometers, and optical instruments for 3D shape monitoring are used for the characterization of fluids, melts, gases, and vapors. The detectors for most of these instruments are CCD cameras combined with image processing packages.

In addition to these payload instruments, there exists also the equipment required to perform and to support extra- and intra-vehicular activities (EVA, IVA) such as the sensors which

Figure 13-17. The German/European SPACELAB D-1 mission (30 October–6 November 1985), which was under operational management of the German Space Research Institute DLR, was the fourth flight of the ESA-developed SPACELAB with the Space Shuttle. The photograph shows the scientist astronaut Wubbo Ockels (ESA) working with the fluid physics module designed for experimentation with open fluid columns. (Photograph: DLR/ESA).

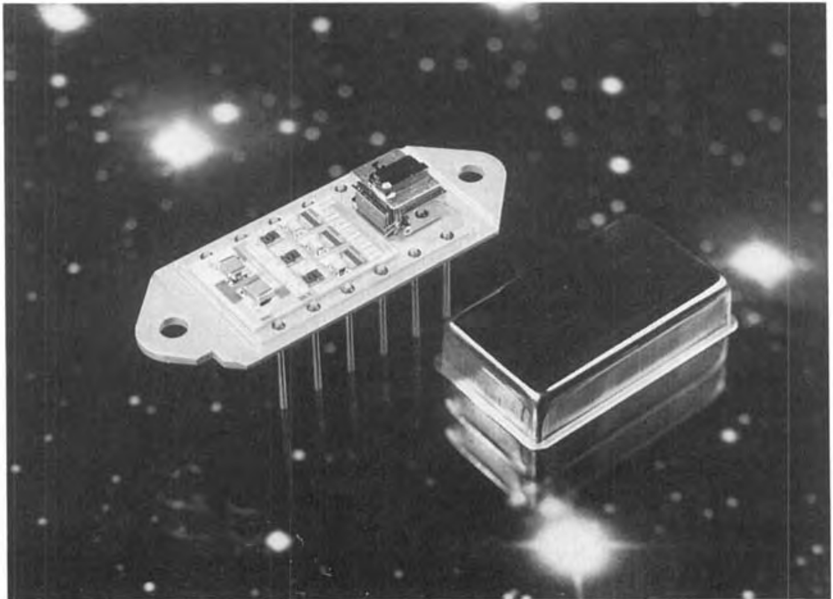


Figure 13-18. Solid-state microaccelerometer package. Three accelerometers are mounted on a cube to allow three-dimensional measurements. The associated electronics are built as a hybrid circuit and integrated in the same package together with the accelerometers. (Photograph courtesy of CSEM).

are part of the astronauts' suits (life support system, orientation system) and robotic systems. Microsensors and microsensor systems will also be used for monitoring the astronauts body functions and for control of space suits' life support systems.

13.4.3 Sensors for Ground Infrastructure

The operations-oriented ground infrastructure in the aviation field consists of systems for airport and air traffic control, simulation and training together with equipment for aircraft maintenance. The ground infrastructure required to support space systems launch and operations comprises launch support infrastructure, spacecraft telemetry, tracking and command (TT&C). The payload specific infrastructure consists mainly of RF ground stations for satellite payload TT&C or stations dedicated to transmission/reception of communication traffic. Examples of sensors associated with the respective ground infrastructure segments are shown in Table 13-11.

Table 13-11. Examples of aerospace ground infrastructure and typical associated sensors.

Market segment	Ground infrastructure element	Associated sensors
Aviation systems	Primary ATC radar	Radar antennas
	Secondary surveillance radar	Radar antennas
	ATC communication ground stations	Antennas
	Meteorological stations	Thermistors Humidity sensors (eg capacitive) Barometer Anemometer
	Wind shear alert systems	Radar antennas
	Doppler wind LIDAR stations	Telescope with electro-optic receiver
	Lightning detection systems	Optical sensors (CCD cameras)
	Runway visual range measurement equipment	Optical sensors
	Data communication ground stations	Antennas
	Optical tracking systems	Cinetheodolites with TV cameras
	Training and simulation systems	TV cameras Complete range of control loop sensors
	Service and maintenance equipment	Optical metrology equipment Endoscopes with TV cameras IR cameras Speckle interferometers with associated cameras
Space systems	Radar tracking stations	Radar antennas
	TT&C stations	Antennas
	Communication ground station	Antennas
	Hand-held terminals	Antennas

Table 13-11. (continued)

Market segment	Ground infrastructure element	Associated sensors
	Launch pad surveillance sensors	TV cameras Thermistors Strain gauges Ice sensors
	Meteorological stations	Thermistors Humidity sensors (eg capacitive) Barometer Anemometer
	Wind shear alert systems	Radar antennas
	Doppler wind LIDAR stations	Telescope with electro-optic receiver
	Lightning detection systems	Optical sensor (CCD cameras)
	Optical tracking systems	Cinetheodolites with TV cameras
	Optical ground stations	Telescopes with optical tracking and metrology equipment
	Training and simulation systems	TV cameras Complete range of control loop sensors
	Service and maintenance equipment	Optical metrology equipment Endoscopes with TV cameras IR cameras Speckle interferometers with associated cameras

13.5 Market Trends

The major customer groups in the aerospace market are airlines and air transport carriers, private aviation, airport and air traffic control authorities, the military, commercial satellite and launch system operators, national and international space agencies, and the growing group of private users of satellite communication services.

Large shares of the market are controlled either directly or indirectly by government; these include the military market, the non-commercial space market, and – to some extent – the airport and air traffic control sector. This dependence and the influence of global economic conditions on air traffic sales make the aerospace market sensitive to changes in the actual economic and political situation. For example, the Gulf War and the economic crises in eastern European countries caused in 1991 a decrease in air traffic of about 2%, the first decrease since the Second World War [57]. This combined with the overcapacities built up during previous years forced the airlines to undergo strict reorganizations. One consequence of the latter was that the airlines placed fewer orders for new aircraft and in some cases cancelled existing orders. This led in 1992 to significant problems in the aircraft industry. (At Boeing the orders in hand decreased from 1991 to 1992 by more than 10%, and Airbus Industry had to announce in this situation a reduction of its monthly production from 22 to 15 aircraft starting from 1993.)

In addition, the aircraft industry had to overcome the problems arising from restructuring the military aerospace product field and the end of the political confrontation between the East and West, which has led to a strong reduction in military development contracts. This has resulted, for instance, in a reduction of about 13% in the number of employees in the German aerospace industry and a drop in the military aircraft sales in the USA of more than 15% between 1991 and 1994.

The aerospace sensor market is coupled to the economic development of the aerospace system market, consequently it is subject to the current recession in the traditional product segments. This is likely to be balanced by the trend towards advanced aircraft design concepts: aircraft which have to comply with more stringent environmental requirements at lower operational and maintenance cost, the use of in-flight data recording systems for support of preventive maintenance schemes, and the requirement of the two-person cockpit crew for more sophisticated and reliable on-board aircraft monitoring and control systems. The need for the development of innovative sensor technology to satisfy these requirements should provide the basis for the sound growth of the aerospace sensor industry.

13.5.1 Aviation Systems

The development of sensors for civil aviation systems will be influenced in the medium term, ie over the next 2–4 years, by the necessity to improve airline productivity, the introduction of advanced navigation systems which can satisfy the requirements of an increasing air traffic density, the implementation of advanced communication systems, and by retrofit or replacement of older aircraft which do not comply with the emerging stricter noise limitations which will be in force in the USA and Europe from January 2000.

In addition, the growth of passenger numbers will require an increase in the fleets and the replacement of Stage 2 aircraft, in particular in eastern European countries. This will stimulate aircraft production rates and help to overcome the crises of recent years.

The annual growth in air traffic over the next 20 years is predicted by Deutsche Aerospace (DASA) to be on average about 4.7% and for commuter/regional routes about 5.5% [58]. Over the same period, the Asia-Pacific market will have a steady annual growth of more than 7%, thus being the fastest growing market world-wide. The European market will experience an annual growth rate of 4.5%, for North America the growth rate is predicted to be about 4% on average, for Latin America about 5.5%, for the Middle East 7% and for Africa about 5% [59].

More detailed forecasts for the US market predict for the civil aviation sector an increase of about 6% for 1995, and that the avionics market will show no variations during 1994–95 and will then slowly increase at an annual rate of about 5%. Despite these slow growth rates in the avionics market, certain product ranges should perform above average. This is expected to apply in such areas as advanced navigation equipment using GPS, airborne communication equipment for data links and for passenger office equipment and retrofits for advanced surveillance equipment.

The sector of business aviation in the USA, taken as an indicator of the world-wide market development in this field, will start to recover again slowly in 1995 with growth rates of about 5% per annum, after having experienced a strong recession in the early 1980s (about 50% sales losses) [60].

The trends in the military aviation market are strongly influenced by the political changes in the East; a time of great changes, however, is always "one of considerable instability and danger. There is no doubt of the remaining threat to security in Europe but, unlike the days of the Cold War, it is now virutally impossible to define (and hence to counter). Instead of calculated aggression, the risk is seen to come from serious economic, social, and political difficulties in the countries in central and eastern Europe. Conscious of that, NATO is searching for a new role and has dropped most of its existing strategies developed during the Cold War, replacing them with a new set of political and military principles" (© Jane's Information Group. David Brinkman: Foreword to Jane's Avionics 1992–93 [10]). This reorientation resulted in a severe reduction of defence budgets; the Stockholm International Peace Research Institute estimated for 1990 a reduction of world-wide defense expenses by 4%, with a contribution of the USA of 6% and of the CIS of about 10%. The export of defense products dropped from 1990 to 1992 by 21.7% for the USA and by almost 79% for the CIS [57].

In the USA, the procurement budget for new weapon systems was reduced by over 50% in the period 1990–95; only after 1995 can a slow increase be expected when older systems have to be replaced. In the case of military aircraft priority has been given to modifications and upgrading of existing aircraft by new weapon systems, enhanced electronic warfare suits, new radar systems, improved command and control functions, and the implementation of advanced battlefield surveillance systems in all of which advanced sensor systems are an important element of upgrade. This trend is supported by an increase in fundamental research and applied R&D and on this basis an average growth in the military aviation sensor market of about 4% per year from 1995 can be expected. These rates will be significantly higher for sensor of high strategic importance.

13.5.2 Space Systems

The space market is a good example of a successfully government promoted industrial development. Since the first artificial satellite in space (Sputnik 1 in 1957), government organizations have financed numerous space missions devoted to basic space research and supported the related technological development activities, thus establishing the foundations for a new industrial sector. The space industry with its commercial areas such as telecommunications, remote sensing, and launch services, can now be considered a profitable industrial branch with good future development potential; the commercial space market in the USA has grown from about $ US 1.8 billion in 1988 to about $ US 5 billion in 1992, of which 52% was related to communication satellites/services, 10% to launch services, 4% to remote sensing, and 34% to satellite ground equipment (see also Section 13.5.3) [61]. Figure 13-19 shows the European space expenditures in million accounting units (MAU), excluding commercial and military expenditures.

For a long time the military space market in Europe was not significant, but is now of growing importance with such projects as the HELIOS program (France, Italy, Spain), the UK programs, and the joint initiative of some West-European Union (WEU) member states for the study of a space-based military communication network (EUMILSATCOM) attracting large budgets. The budget involved in these military programs is not known precisely but can be estimated to be about 600 MAU per year, maintaining a constant level since 1991 [61].

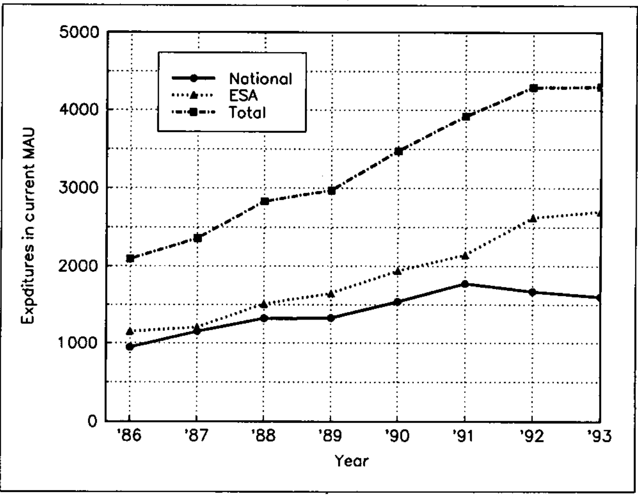


Figure 13-19. European space expenditures in million accounting units (MAU), excluding military and commercial projects, at current (1994) levels [61].

Figure 13-20 shows the European space budget in comparison with the expenditures of other countries (USA, the former USSR, Japan and Canada). Only the figures for the USA and the former USSR include military expenditure, the commercial space market sales are excluded in all figures presented.

The space market was significantly less affected by the recession than the aviation market (see Section 13.5.1). Its future prospects in the fields of civil and military communications, navigation, and space-based remote sensing are excellent and promise high growth rates not

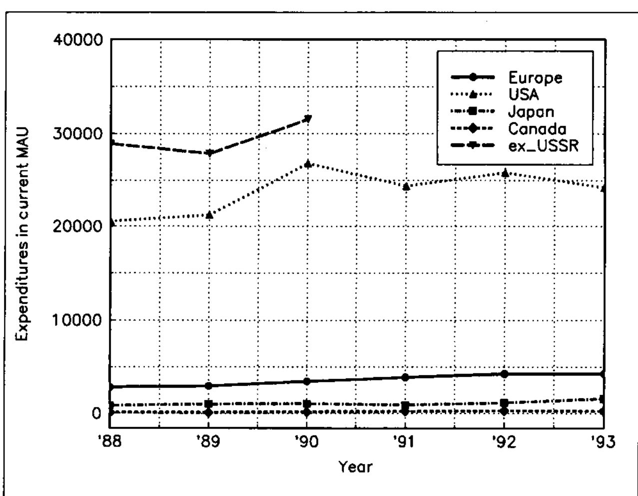


Figure 13-20. Space budgets of the USA, the former Soviet Union, Japan, Canada and Europe. Commercial expenditures are not included; military expenditures are included for the USA and the former Soviet Union. Current (1994) levels of the accounting unit are quoted [61].

only for the space segment but even more for the related on-ground equipment and services (see Section 13.5.3). The new concepts for satellite communications that incorporate hand-held terminals (eg, Motorola's Iridium, Hughes's Spaceway, Teledesic), the various positioning/data communication services, satellite-supported navigation systems for civil aviation, the new military combat scenarios which rely on space-based networks, and the increase in launch services resulting from these activities are only some examples that illustrate the future of space system applications [33, 62].

The market for space sensor systems is closely linked to current developments in this field and can therefore be expected to have excellent future perspectives for those products that are innovative and have attractive performance-to-cost ratios.

13.5.3 Ground Infrastructure

The aerospace ground infrastructure is one of the fastest growing aerospace market segments; it includes for aviation applications, among others (see Section 13.4.3), the full range of ATC equipment necessary to cope with an increase in passenger traffic of about 30% by the end of the 1990s. Of particular interest are here the markets in the Asia-Pacific region, where a strong demand for upgrades of ATC installations exist [59, 63].

Another growing field for aviation-related ground infrastructure is that of simulation and training for aircraft crews in both civil and in military aviation. In the future such facilities will be particularly required in military aviation in order to reduce the environmental impact of low-altitude training flights. For civil aviation the installation of simulators has to keep pace with the growing number of crew to be trained and the growth rates for this equipment can be derived from the growth of air traffic.

In the field of space-related ground infrastructure, high-volume markets are developing for satellite communication ground stations and hand-held wireless communication equipment. According to conservative predictions an overall sales volume for ground station equipment of $ US 65–83 billion between 1992 and 2004 will be achieved, which will be more than double the combined commercial satellite and launcher market [64]. These figures exclude the sales of hand-held personal satellite communication systems and the sales of GPS-based equipment, both of which are expected to increase also significantly during the next few years.

These market segments will greatly exceed the sales figures for space segment equipment. The sensors used in such systems will be mainly antennas with RF receiver front-ends and the associated housekeeping sensor instrumentation necessary for performance monitoring.

13.6 Summary

Sensors are key elements of modern aerospace systems for a wide range of applications in both platform and payload design. The wide variety of physical parameters to be detected, in particular for the design of space-borne payloads, requires the use of different sensor types and technologies, ranging from fairly simple mechanical sensor devices to highly complex sensor systems, such as the sensors used for the detection of variations in the Earth's gravitational field.

With the exception of some ground-based sensor instrumentation, aerospace sensors have to comply with stringent quality and reliability requirements, which is reflected to some extent in the standards applicable for design, manufacture, and performance verification. The commercial trends in all segments of the aerospace market indicate sound growth rates, with some products of the ground infrastructure segment already experiencing a steeply increasing demand.

13.7 Acknowledgments

The author thanks the European Space Agency (ESA) for permission to publish this chapter and for making available various photographs and illustrations. The author also acknowledges and thanks Marcel Toussaint, Director of Studies, Eurospace, and Sevig Press, publisher of the European Space Directory, and also Jane's Information Group, publisher of Jane's Avionics, which granted the right to cite various material from their publications.

Further, the author is indebted to the Centre Suisse d'Electronique et de Microtechnique (CSEM), Neuchâtel, Switzerland; GEC-Marconi Avionics, Edinburgh, UK; Sira Ltd., Chislehurst, UK; and Deutsche Aerospace (DASA), Munich, Germany; for many of the photographs provided for this publication.

The author gratefully acknowledges the suggestions and numerous contributions from colleagues at the European Space Agency (ESA): K.-H. David, A. Menardi, H. Lutz, H. Lechte, F. van Ingen, E. Armandillo, G. Graf, J. Steinz, D. Doyle and B. Guldimann. Their support and constructive criticism were extremely helpful in the completion of this work.

13.8 References

[1] Göpel, W., Hesse, J., Zemel, J. N., Grandke, T., in: *Sensors*, Vol. 1, (eds.); Weinheim: VCH, 1989, Ch. 1.
[2] *Reliability Assurance of ESA Spacecraft and Associated Equipment;* ESA PSS-01-30, Noordwijk: European Space Agency, 1992.
[3] *Basic Requirements for Product Assurance of ESA Spacecraft and Associated Equipment;* ESA PSS-01-0, Noordwijk: European Space Agency, 1981.
[4] Kendal, B., *Manual of Avionics;* Oxford: BSP Professional Books, 1987.
[5] Helfrick, A., *Modern Aviation Electronics;* Englewood Cliffs, NJ: Prentice Hall, 1984.
[6] Spitzer, C. R., *Digital Avionics Systems;* Englewood Cliffs, NJ: Prentice Hall, 1987.
[7] Dornheim, M. A., *Aviation Week & Space Technol.* **141**, No. 10, (1994) 89.
[8] Nordwall, B. D., *Aviation Week & Space Technol.* **140**, No. 11 (1994) 94.
[9] Warshawsky, I., in: *Sensors*, Vol. 1, Göpel, W., Hesse, J., Zemel, J. N., (eds.); Weinheim: VCH, 1989, Ch. 22.
[10] Brinkman, D., (ed.), *Jane's Avionics 1992-93;* Surrey: Jane's Information Group, 1992.
[11] Dornheim, M. A., *Aviation Week & Space Technol.* **140**, No. 11 (1994) 72.
[12] Morrocco, J. D., *Aviation Week & Space Technol.* **140**, No. 11 (1994) 48.
[13] von Fabeck, W., *Kreiselgeräte*, Würzburg: Vogel, 1980.
[14] Nutall, J. D., *Electron. & Power* November/December (1987) 703.

[15] Burchardt, W., Jungbluth, W., Rodloff, R., presented at the Symposium Gyro Technology 1984, Universität Stuttgart und Deutsche Gesellschaft für Ortung und Navigation.
[16] Mackintosh, J., Giles, I. P., Culshaw, B., presented at the Symposium Gyro Technology 1985, Universität Stuttgart und Deutsche Gesellschaft für Ortung und Navigation, 24-25 September 1985.
[17] Sonnenberg, G. J., *Radar and Electronic Navigation;* London: Butterworth, 1988.
[18] Lin, C.-F., *Modern Navigation, Guidance, and Control Processing;* Englewood Cliffs, NJ: Prentice Hall, 1991.
[19] Gusmeroli, V., Martinelli, V., Escobar Rojo, P., *Proc. SPIE* **1797** (1992) 132.
[20] Hughes, D., *Aviation Week & Space Technol.* **141**, No. 12 (1994) 43.
[21] Morain, S. A., Thome, P. G., *America's Earth Observing Industry: Perspectives on Commercial Remote Sensing:* HongKong: Geocarto International Centre, 1990.
[22] Szekielda, K.-H., *Satellite Remote Sensing for Resources Development,* United Nations Department of Technical Co-operation for Development, London: Graham and Trotman, 1986.
[23] Hord, R. M., *Remote Sensing: Methods and Applications;* Chichester: Wiley, 1986.
[24] Carter, D. J., *The Remote Sensing Sourcebook;* London: Kogan Page, 1986.
[25] Killinger, D. K., Menyuk, N., *Science* **235** (1987) 37-45.
[26] Rast, M., *Imaging Spectroscopy and its Application in Spaceborne Systems:* ESA SP-1144 Noordwijk: European Space Agency, 1991.
[27] Bodechtel, J., Hoffmann, O., Kunkel, B., Meissner, D., *Acta Astronaut.* **8** (1981) 767.
[28] Staenz, K., *Can. J. Remote Sensing* **18** (1992) 187.
[29] Gower, J. F. R., Borstad, G. A., Anger, C. D., Edel, H. R., *Can. J. Remote Sensing* **18** (1992) 199.
[30] Dreyer, G., *Opt. Eng.* **25** (1986) 1253.
[31] Fortescue, P., Stark, J., (eds.), *Spacecraft Systems Engineering;* Chichester: Wiley, 1992.
[32] *The ESA Technological Research and Development Programme 1994-1996;* TD (94) 1, Noordwijk: European Space Agency, 1993.
[33] Asker, J. R., *Aviation Week & Space Technol.* **141**, No. 10 (1994) 119.
[34] van Holtz, L., Burton, M., Rogers, R. C., presented at the 12th Annual AAS Guidance and Control Conference, Keystone, Colorado, February 1989.
[35] Czichy, R. H., *Hybrid Optics for Space Applications;* ESA SP-1158, Noordwijk: European Space Agency, 1993.
[36] Xu, M. G., Dakin, J. P., *Optical Fibre Sensors for Space Applications;* Final Report for ESTEC Study Contract 7096/87/NL/PP, 1988.
[37] Cefai, J. J., Barrow, D. A., *A Multi-User Microsensor Array for Life Sciences Research in Space;* Consultancy Report, ESA/ESTEC Contract No. 120917b, Noordwijk: European Space Agency, 1992.
[38] Cefai, J. J., Barrow, D. A., *Database on Associated Sub-System Components Required by Integrated Microsystems for Life Sciences Research in Space;* Consultancy Report, ESA Contract P. O. 131195, Noordwijk: European Space Agency, 1994.
[39] Genequand, P. M., Droz, F., in: *Proc. First European Conference on Space Debris, Darmstadt, April 1993;* ESA SD-1, Darmstadt: European Space Agency, 1993.
[40] Maral, G., Bousquet, M., *Satellite Communications Systems;* Chichester: Wiley, 1987.
[41] Oppenhäuser, G., Wittig, M., Popescu, A., *Proc. SPIE* **1522**, Optical Space Communication II, (1991) 3.
[42] Duchmann, O., Planche, G., *Proc. SPIE* **1417**, Free-Space Laser Communication Technologies III (1991) 31.
[43] *Space Science Horizon 2000;* ESA SP-1070, Noordwijk: European Space Agency, 1984.
[44] Domingo, V., (ed.): *The SOHO mission; Scientific and Technical Aspects of the Instruments;* ESA SP-1104, Noordwijk: European Space Agency 1988.
[45] Chicarro, A., Grard, R., (eds.): *ESA's Report to the 29th COSPAR Meeting, Washington DC, USA, September 1992:* ESA SP-1154, Noordwijk: European Space Agency 1992.
[46] Briel, U., Brinkman, A. C., Ellwood, J. M., Fraser, G. W., de Korte, P. A. J., Lemmon, J., Lumb, D. H., Manzo, G., Peacock, A., Pfeffermann, E., Rocchia, R., Westergaard, N. J., *The High-Throughput X-Ray Spectroscopy Mission;* ESA SP-1092, Noordwijk: European Space Agency, 1987.
[47] Beckwith, S., Cornelisse, J., van Dishoeck, E., Encrenaz, P., Genzel, R., Griffin, M., Harris, A., Hills, R., de Jong, T., Kollberg, E., Kreysa, E., Lamarre, J.-M., Lellouch, E., Martin-Pintado, J., Natale, V., Pilbratt, G., Poglitsch, A., Puget, J.-L., Rowan-Robinson, M., Stutzki, J., Volonté, S.,

Whyborn, N., *FIRST Far Infrared and Submillimetre Space Telescope;* ESA SCI (93) 6; Noordwijk: European Space Agency, 1993.

[48] Schwehm, G. H., Langevin, Y., *Rosetta Comet-Nucleus Sample Return; Mission and System Definition Document;* ESA SP-1125, Noordwijk: European Space Agency, 1991.

[49] Johnson, S. P., Corcelle, G., *The Environmental Policy of the European Communities;* London: Graham and Trotman, 1989.

[50] Kondratyev, K. Y., Kozoderov, V. V., Smotky, O. I., *Remote Sensing of the Earth from Space: Atmospheric Conditions;* Berlin: Springer, 1992.

[51] Report of the Consultancy Group, *Rain Radar;* ESA SP-1119, Noordwijk: European Space Agency, 1990.

[52] Flamant, P. H., Loth, C., Megie, G., Pelon, J., *ESA J.* **9** (1985) 449.

[53] Lutz, H., Armandillo, E., *ESA Bull.* No. 66 (1991) 73.

[54] Hueber, M. F., presented at the International Symposium on Satellite Remote Sensing, Rome, September 1994.

[55] Report of the ATLID Consultancy Group, *Backscatter Lidar;* ESA SP-1121, Noordwijk: European Space Agency, 1990.

[56] Norrie, C., Marini, A., Armandillo, E., presented at the International Symposium on Satellite Remote Sensing, Rome, September 1994.

[57] *Fischer Welt Almanach 1994*; Frankfurt: Fischer Taschenbuch 1993.

[58] Sparaco, P., *Aviation Week & Space Technol.* **141**, No. 14 (1994) 39.

[59] Kandebo, S. W., *Aviation Week & Space Technol.* **140**, No. 11 (1994) 71.

[60] Growth Trends: US Business Flying, *Aviation Week & Space Technol.* **140**, No. 11 (1994) 85.

[61] Toussaint, M., in: *European Space Directory 1994;* Paris: Sevig Press, 1994.

[62] Scott, W. B., *Aviation Week & Space Technol.* **140**, No. 11 (1994) 101.

[63] Mecham, M., *Aviation Week & Space Technol.* **140**, No. 22 (1994) 65.

[64] *World Space Market Survey: 10-Year Outlook;* Paris, Euroconsult: 1994.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

14 Process Sensing and Control

Michael J. Scott, Cambridge, UK

Contents

14.1 Introduction 414

14.2 Measurement Requirements in Modern Process Control 415

14.3 World Market Sizes for Measurement and Control Systems 416

14.4 Current and Emerging Sensor Technologies and Markets 417
14.4.1 Flow 417
14.4.2 Pressure 420
14.4.3 Level 421
14.4.4 Mass 422
14.4.5 Temperature 422
14.4.6 Turbidity and Suspended Solids 423

14.5 Emerging Trends in Process Control and Sensing 424
14.5.1 Virtual Sensors and Neural Networks 424
14.5.2 Smart Sensors 426
14.5.3 Surrogate Measurements 427
14.5.4 New Sensors and Applications 428
14.5.4.1 Color 428
14.5.4.2 Fugitive Volatile Organic Compounds 429
14.5.4.3 Odor 429
14.5.4.4 Wet Chemical Analyzers 429

14.6. Conclusion 430

14.7 References 430

14.1 Introduction

In this chapter an introduction to sensors for process control is presented and some emerging process measurement requirements highlighted to illustrate the key issues associated with measurement and control. The word "sensor" is commonly used for many different purposes, but for this chapter it is taken to mean an instrument which generates a signal useful for a control system. Other terms, such as "gauge" and "transmitter", are used where these terms are commonly associated with the particular technique or product. (A comprehensive treatment of the basic principles of process measurement instrumentation is given by [1]).

No decision to initiate control can be made without first making a measurement; in the process industries manual intervention by an experienced operator is now rare. Continuous reduction in process cost and improvement in quality is achieved by measuring and managing dynamic performance. Modern open industrial control systems consequently need more process information in a shorter time and traditional single parameter based measurements are being supplemented by dynamic performance measurements. Until recently a process sensor, such as a flow meter or temperature transmitter, was almost exclusively an instrument that provided information for control using familiar single, cascade or feed forward control; process control now requires a much broader range of process measurement techniques.

Battelle has claimed that there are five hundred thousand sensors available in Europe to measure more than one hundred parameters. Fisher Rosemount, in a recent advertisement claimed that their pressure transmitter range has 15 million variations and that there are 3 million of these transmitters installed on user sites. Fredonia Group Study #561 claims more than 1000 manufacturers of level switches and sensors in the USA. This general market information indicates the variety of process control sensors in use. Some idea of the market variety can be gained by considering a list of measurand against a list of measurement application areas, as shown in Table 14-1. Here each list is randomly ordered and not claimed to be complete. The fact that all topics can be cross-referred illustrates the complexity of the market.

The process measurement instrumentation market is also characterized by rapid change and a very large number of niche opportunities. Many new opportunities come from exploiting old ideas in new applications in new industries. For example, physiological measurement is included because the presentation of process data and plant performance information to an operator and its subsequent interpretation is one of the constraints in process control. On-line measurement of an operator's performance might therefore enhance safety and prevent costly plant failures. The term "Commerce" is included since enterprise wide control systems are being adopted by large corporations in order to reduce inventory and improve customer responsiveness. This approach to business demands new real time measurement for the packing, pricing, movement control and storage of products.

Process measurement for control purposes can be devided into four types:

- *Direct:* these include temperature, pressure and flow where the information of interest is obtained directly from the sensor.
- *Indirect:* for example, mass or volume which are often computed from flow, density, temperature etc.
- *On-or at-line analysis:* these give discontinuous information. (In some industries the laboratory analysis of samples is still widely used for control of a process, but it is being

increasingly recognized that this is not genuine process control since the time constant of the measurement process are usually orders of magnitude greater than those of the process.)

- *Surrogate:* this used when the parameter of interest cannot be measured directly and an alternative measurement is made from which the required parameter may be determined.

There is a continued increase in the precision and amount of information required from process measurement instruments in order to provide the safe and economic operation of a process or business. This is being satisfied by a sharp increase in the capability and complexity of sensors measurement and the migration of functionality from the control system. Temperature compensation, diagnostics and control algorithms are being moved from computer control systems and embedded in the transmitter. At the same time various technologies are being used to make this rapidly increasing complexity trasparent to the user. However, the increased functionality means that each measurement and control decision must be considered within the context of the total process control system and implementation now requires a combination of multi-disciplinary skills.

Table 14-1. A summary of measurand and measurement applications.

Measurement technology	Measurement application
Flow (solids, gas, liquid)	Chemical process industry
Pressure	Water & Waste Treatment
Temperature	Electricity generation
Chemical analysis	Nuclear
Level	Petrochemical/Oil
Mass	Metals
Speed	Food
Imaging (thermal, optical)	Medical
Radiation	Aerospace
Shape	Biotechnology
Physiological	Medical
Density	Commerce
Humidity/moisture	Transport
Force	Gas
Chemical	

14.2 Measurement Requirements in Modern Process Control

The control of a process and interdependent plants requires a complex combination of skills. Furthermore the trend to control systems that operate over an entire enterprise is making new demands on the acquisition and use of process performance information. Process

plant is also designed with smaller "footprints" and there is more recycling and the extensive use of heat integration. Control systems are consequently becoming more complex and more extensive in their influence. There is also a need to operate the process against economic and competitive pressures over a wide range of plant throughput. It is generally agreed that the hardware for modern control systems has, in response to these conditions, reached commodity status and open systems, where elements from different manufacturers have to be interfaced, are already beginning to dominate the market. For example, leading control system suppliers such as ABB, Honeywell, Foxboro and Bailey advertise their ability to integrate third party software and hardware within an installation.

Functionality and control capability is necessarily migrating towards the process and this trend is expected to continue. A number of pressure transmitters and process control valves, for example, already include three term controllers and pH transmitters often incorporate sophisticated self diagnostic and calibration facilities. Smart transmitters already take more than 50% of the process control market, and it is expected that there will be a continued increase in measurement and control capability at the point of measuement. Remote reconfiguration and diagnostic capabilities will subsequently become a standard feature of process measuring instruments. This rapid change in the structure of process control systems is made possible by the availability of low cost, field mountable, advanced computing. Instrument manufacturers are thereby able to provide significantly greater capabilities for the user while also reducing his workload.

Some commentators believe that the future architecture of control systems at the field level will constitute a distribution of intelligent transmitters, valves, digital input/outputs and workstations. The blurring of the roles of process sensors and the first level of control is already evident and can be expected to continue.

14.3 World Market Sizes for Measurement and Control Systems

Statistics for the process measurement and control market are notoriously difficult to obtain since the market is highly fragmented by country of origin, product type and industry usage. Most multi-client studies such as Frost and Sullivan, MIRC and Venture Development Corporation have the tendency to aggregate instrument groups too coarsely to be useful for analyzing specific product groups.

GAMBICA estimates the global market for measurement and control at 80 billion dollars with Western Europe at 23.5 billion dollars, North America at 28 billion dollars and the rest of the world at 28 billion dollars. MIRC has claimed that the Western European sensor market was 2.5 billion dollars in 1989 with a forecast of 5 billion dollars by 1995. Frost and Sullivan, in a 1994 Report puts the Western European Process Control Instrumentation market at 6098 million dollars and sensors at 2460 million dollars for 1993; 40.6% of the total PCI market. Frost & Sullivan claim 5.2% growth that would give a 3510 million dollars market in the year 2000.

Markets for particular process sensors range from a few hundreds for a particular on-line chemical measurement in the European market to a world market of more than a million per annum for pressure transmitters. It is important to note also that publicly available market

sector statistics are often very unreliable since the product and target market definitions used influence strongly the values obtained. The generation of commercially useful data for specific market sectors therefore requires carefully focused and specialist expertise.

14.4 Current and Emerging Sensor Technologies and Markets

The purpose of this section is to correlate the techniques with the market sizes for the major process sensor.

14.4.1 Flow

Flow is the most widely used industrial measurement with a large number of flow techniques available ranging from the speed of a soap bubble in a glass tube for the measurement of gases at 0.01 mL/min through to three meter diameter magnetic flow meters measuring water at the rate of 10^5 L/s. The goal of researchers and developers has for decades been to establish a noninvasive universal flowmeter that can be fitted to existing pipelines regardless of the flow conditions. Although in recent years, ultrasonic flowmeters have demonstrated considerable flexibility this goal still seems distant.

With such a diverse range of technologies there are many suppliers. An infosheet published in the January '94 issue of the magazine *Processing* gave ninety eight suppliers in the UK with twenty four classes of instrument and the five application areas: liquid, slurry, gas, steam, two phase. This understates the position since there are many small companies who trade in regional or industrial niches which are not included. The picture is same in the European and North American markets.

Table 14-2 gives an indication of the European flow meter market structure, and Table 14-3 shows the distribution of flowmeters by end user industry in Europe. Frost & Sullivan (Moun-

Table 14-2. European flowmeter market share.

Instrument	1993 (million dollars)	% growth
Electronic dp	21.7	4.2
Positive displacement	16.4	3.8
Electromagnetic	14.1	5.5
Turbine & propeller	10.3	5.0
Primary devices	9.2	2.7
Variable area	9.2	4.6
Ultrasonic	5.7	6.1
Vortex	4.0	7.3
Mass	4.0	7.5
Pneumatic dp	2.4	−6.2
Direct connected	1.6	−1.9
Other	1.4	2.8

Table 14-3. European flowmeter distribution by end user industry.

Industry	1993 (million dollars)	% growth
Oil & gas	49.9	3.4
Petroleum refining	106.0	3.8
Chemical	265.4	4.5
Power generation	122.8	4.5
Primary metals	82.0	3.5
Mineral processing	35.2	4.6
Pulp & paper	57.1	3.5
Water & effluent	88.0	5.5
Food & beverage	54.7	4.6
Other	63.8	4.4

tain View California) has claimed a 1992 world market of 2.6 billion dollars for flowmeters and Automation Research Corporation (Medfield Mass.) claim a 180 million dollars market in the USA for smart flowmeters.

The continued dominance of the orifice plate and the differential pressure transmitter, with both price reduction and increase in functionality, is an important feature of the flow measurement market. (This is despite the inherent disadvantages of unrecovered head loss and impulse tubes that can block). Rosemount, probably the world's leading pressure and differential pressure transmitter manufacturer, claim to have sold more than 3.5 million pressure transmitters worldwide with sales of more than 100000 units per annum in Europe. However, not all these transmitters are for flow measurement and it has been estimated that the "pressure transmitter" family of applications for a leading process control supplier, such as Fisher Rosemount, can be evenly split between flow, pressure and level.

Smart pressure transmitters already take more than 50% of the world market and there seems little doubt that the orifice plate and differential pressure transmitter will continue to dominate the flowmeter market for at least another decade; particularly since smart differential pressure transmitters extend the conventionally available 3 : 1 turndown to typically 100 : 1. (Turndown is the term commonly used to describe the ratio of the minimum measurement to the maximum measurement). Although the environmental industry is exhibiting strong growth, the Hydrocarbons Processing Industry continues to provide the major market for pressure and differential pressure transmitters and therefore strongly influences the technology.

The accuracy of an orifice plate and differential pressure transmitter flowmeter (as stated by suppliers) is typically 1% and BS 1042 is often quoted to substantiate their claim. However, it is only very carefully designed and maintained installations that will come even close to 1% and a figure of nearer 5% is more likely for most installations. Accuracies of 1% to 2% are sufficient to satisfy 80% of the flowmeter market for process control. Other flow techniques need to be used when greater accuracies are required, ie, for such applications as custody transfer.

Magnetic flowmeters depend on Faraday's law of magnetic induction and were launched commercially in the mid 1950s by Altometer of Holland, supported by much publicity concerning their nonobstructive mode of operation and ability to measure bidirectionally. It was

a decade, however, before the sensor became more than a specialist instrument and only in the last few years have prices dropped significantly (30% in 1993). Magnetic flowmeters are consequently approaching the commodity status of orifice plates and differentail pressure transmitters, both able to operate over the same, virtually unlimited, range of pipe sizes. In addition, dramatic extensions to the linear range, reduction in power consumption, availability of intrinsic safety and ability to operate with fluids with very low conductivity have been achieved. A recent magnetic flowmeter introduced by ABB Kent gives a 1000 : 1 turndown and 0.2% accuracy at a competitive price. ABB Kent claims that around twenty eight manufacturers compete for a world market of more than one hundred thousand units per annum with a faster growth rate than any other flowmeter. (In part this is due to the modernisation of the water and waste treatment industry). The recent launch by Fischer & Porter, of a magnetic flowmeter that can measure flow in partially filled pipes is a major contribution to reducing the cost of installation of large bore meters. (Until this launch all closed pipe flow meters needed to be mounted so that the measuring section was full of the fluid being measured, which could require expensive pipe work). Endress & Hauser claims that they have sold 80000 magnetic flowmeters worldwide since 1984. The recent rapid growth rates of the magnetic flowmeter market suggests that the current sales rate could be as high as fifteen to twenty thousand units per annum.

Ultrasonic flowmeters, using either time-of-flight or doppler techniques, were launched nearly two decades ago with great promise because of the nonobstructive and noninvasive mode of operation. Time-of-flight add-on kits are now widely used for large diameter pipes and doppler ultrasonic flowmeters have established a small niche for making flow measurement where no other technique is effective. They do not, however, have a significant market share.

The coriolis mass flowmeter, which depends on the coriolis affect on a vibrating tube, has made a significant market impact since Micromotion launched the first commercial unit in the mid 1970s. Some estimates put the current world market at around twenty five thousand units per annum of which Micromotion appears to have retained as much as three quarters. The coriolis meter has the advantages that it is bidirectional, has no Reynolds number restrictions, has no velocity profile problems and can provide signals for mass rate, solids rate, percent solids, temperature and viscosity; making it the most versatile of all the commercially available meters. All of the early meters depended on complex pipe configurations with consequent high pressure drop but new instruments from Danfoss and Krohne have a straight through, unobstructed, tube. (The Krohne design is claimed to be the result of complex mathematical modelling and the use of patented magnetic inductive coupling.) The coriolis mass flowmeter has, however, a high unit cost, has been relatively bulky and is restricted to pipe sizes under six inches. The mechanical sophistication of the coriolis meter and the tightness of the patents has also restricted the number of manufacturers to around twelve compared with, for example, more than forty for the less complex turbine flowmeters. Nonetheless the coriolis mass flowmeter is highly competitive where the control of high cost fluids with variable density is required and it seems clear that the continued application of finite element design techniques will extend and simplify application as well as reduce cost, hence enabling strong growth for the future.

The vortex meter, which depends on the measurement of the eddies shed by a bluff body, has no moving parts, and with the possibility of nonwetted sensors, seemed to promise a great deal when it was launched twenty years ago. It has, however, only found niche markets, such

as steam flow, and has sales around one third of those of the coriolis meters. Although the turndowns of 38 : 1 for gases and steam, claimed by Fisher Rosemount, are better than conventional orifice plate and differential pressure combinations, vortex meters are inflexible once specified, are limited to eight inch pipes and are restricted to fluids of viscosity less than 30 cPa ($30 \cdot 10^2$ Pa). Despite these limitations, manufacturers, such as Foxboro have, for some years, provided users with cost of ownership calculations that show that vortex meters have clear advantages over orifice plates and believe that they are now seeing the start of a strong market growth.

Leading suppliers such as ABB Kent and Danfoss have built multi-million pound high performance NAMAS approved flow laboratories to support their performance claims and this is an important contribution to the improvement of on-line flow measurement performance. Although some flow meters, such as magnetic and coriolis mass flow meters, are relatively insensitive to flowing conditions, a NAMAS approved certificate of calibration is an important contribution to the user's understanding of uncertainty and the contribution of installation conditions to the overall performance. There is no doubt however, that a great deal of work is still needed and that it is necessary for the supplier to educate most users regarding installation and the appropriate process flow conditions. (This is not unique to flow measurement and is a pacing factor in the general improvement of process information to the control system.)

14.4.2 Pressure

Note has already been made of the difficulties of separating differential pressure transmitter market statistics from purely pressure measurements when considering the product families of the leading manufacturers such as Rosemount, Honeywell, Siemens etc. Generally these transmitters rely on the detection of the movement of metal diaphragms using, for example, capacitative measurement and can withstand considerable overpressure. The world market for this class of pressure transmitter is of the order of 350000 units per annum which suggests that the pressure instrument market is around 100000 per annum. Another class of pressure transmitter uses strain gauge technology to measure the strain on the diaphragm and operate at pressures up to 690 bar (10,00 psig). Applications range between standard process control and aerospace. Druck, one of the world's leading manufacturers of high precision transmitters uses a diffused silicon semiconductor gauge of piezo resistive elements to generate a Wheatstone resistive strain bridge on a silicon substrate. Although the investment in the silicon micromachining equipment necessary to produce around eighty thousand sensors per annum is high, the almost perfect elasticity of the silicon provides unrivalled linearity without hysteresis.

Pressure is rarely a primary process control parameter but is usually used to define the status of a process or as a safety indicator. Although care must be used to ensure that the transmitter sees the correct part of the process and that blockages or encrustations are avoided, it is in practice a relatively straightforward measurement to perform.

14.4.3 Level

The market for level instruments, like that for flow, provides an interesting insight into the complexities of the market for process control sensors. The Western European level sensing market is in the region of 400 million dollars. The market is highly fragmented by technology, supplier, application, cost and industry. Sensors range from float switches costing only a few pounds to sophisticated microwave instruments costing many thousands of pounds. In the vast majority of applications the solution to a problem at the lowest cost is the major preoccupation of the user. The technology used is, by comparison, unimportant.

In terms of units in operation the level measurement market has been dominated, perhaps to as much as 90%, by point level detection. This was traditionally by float operated switch; even for high integrity protective measurements on an oil rig. In the last decade or so, ultrasonic versions have made a strong penetration of the float switch market. This is because the increasing capability and sophistication of process control systems has shown that point detection is no longer adequate, even for alarm purposes. The new processes and attendant control systems are continuous and cannot tolerate intermediate holding chambers, controlled only between maximum and minimum levels. For alarm purposes, furthermore, the rate of change of a level can be more important than its absolute value. There is therefore a strong move towards the use of continuous level measurement.

The traditional continuous level measuring instrument in the oil and chemical industries has been the Fisher Leveltrol and equivalent which use displacement or buoyancy techniques. Differential pressure transmitters are also used and their role in level measurement is frequently underestimated by market studies. Both these techniques need pressure tappings in the vessel and impulse lines to the instrument with attendant wetleg/dryleg considerations and problems of impulse line blocking. Foxboro have, for some years, shown that pressure transmitters are now sufficiently accurate that signals from individual transmitters on an oil tank in a tank farm can be summed to provide a measurement precision compatible with that generated by specialist contents metering.

The parallel trend is towards nonintrusive and ideally to noninvasive techniques. A simple count of the frequency of mention in a small survey investigating the prospects for radar techniques for level measurement, in 1993, showed ultrasonic as the leading technique at 28.7% but with industrial weighing at 20.8%. Capacitance, RF and conductivity share a 14.9% visibility and radar (steered by the interview) had 9.9%. No other technique had a visibility greater than 10%.

Ultrasonic level manufacturers such as Milltronics and Vega have developed very sophisticated signal handling software, for continuous level measurements, which can recognize, and eliminate, false echoes from foam, disturbed surfaces, agitators, incoming streams of fluid and fixed intrusion such as ladders. However ultrasonics signals are affected by temperature and temperature stratification and are intolerant of fouling of the transducer. Radar based sensors do not have these limitations but, until recently, cost has confined their use to high accuracy tank gauging. Saab, the world leader in such applications, claims 20,000 radar level gauges in use and a worldwide market of 22 million units, and are now breaking into the process control market. Radar level gauges can be mounted in Zone O safety areas and use advanced signal processing to make measurements in very demanding process conditions; often under conditions where measurement was previously impossible. Radar level gauges are claimed to have an mtbf of over 60 years with no need for maintenance or

recalibration and are easy to install. However prices are still some thousands of dollars which is high for a market that is very sensitive to price. The prospect is that strong growth will result from an appreciation of the life time ownership cost. The consequent increased sales volumes will also help reduce prices. It seems likely, however, that the radar gauge will always be more expensive than ultrasonic.

14.4.4 Mass

Mass is the parameter most often required for process control even though flow and level are the most frequently made measurements. (The emergence of a strong market for mass flowmeters has already been discussed.) Compaction, bridging, rat-holing and changes in the angle of repose all cause difficulties for level measurement but are easily handled by industrial weighing. Until comparatively recently industrial weighing has not been seen as an on-line process control measurement but there is now an increasing penetration of the level measurement markets. Where there is a need for accuracy all silos, except those made of concrete, are exclusively mounted on load cells, with the measurements used to control the flow of product to the process and to measure the contents. The European market for load cells is probably more than 100 million dollars with industrial weighing measurement and control packages having a market of perhaps 500 million dollars. Until recently the major part of this market was for inventory and batch control with less than 20% used for closed loop control. Industrial weighing is seen as expensive to purchase but it is noninvasive, has low cost of ownership, is easy to calibrate and has unrivalled accuracy.

Measurement of solid or liquid materials to and from storage tanks or silos is beginning to be used for process as well as inventory control with silo or vessel weight the preferred measurement technique. K-Tron market a load cell, using vibrating wire sensing, which has a resolution of 1 in 1000000 thus making the incrementing or decrementing of process materials in a storage tower as accurate as flowmetering.

For some years there has been a technology push towards load cells with on-board digital processing and serial communication. This is because more information can be obtained than by the conventional practice of transferring the signals from the load cells on one weighing application into a single analog amplifier. Disturbances such as compaction, loading and unloading forces and the effect of wind forces can be monitored and compensated for and plant vibration measured and used for condition monitoring. Mettler Toledo has pioneered digital load cells on industrial weigh bridges and other manufacturers are introducing products. Progress in the process industries has however been slow despite the considerable advantages to be gained. Industrial weighing is nonetheless being brought into general process control and it is expected that enterprise wide control will force control engineers to consider bulk storage as a dynamic parameter rather than a quasistatic inventory. Industrial weight sensors should perhaps be considered as a noninvasive flowmeters but to realize the benefits from such an approach will require a radical rethink on process instrumentation.

14.4.5 Temperature

There is a wide range of temperature measurement techniques available including filled bulbs, thermistors, diodes, bimetallic dial indicators, glass thermometers and noncontacting IR

devices. For process control measurements the choice usually rests between thermocouples or resistance temperature devices (RTDs) due to reasons of cost, accuracy and stability. Both thermocouples and RTDs are generally used with a signal conditioner, often mounted in the thermowell head, which converts the basic low level signal to a high level analog current signal proportional to the process temperature. This provides the major advantage of a robust signal for transmission to remote receivers.

Although the temperature ranges covered by these sensors overlap, they cannot be used interchangeably. Differences in construction methods, materials and operating principles result in each having distinct advantages and disadvantages as summarized in Table 14-4. Choosing the best sensor for a particular application requires an understanding of these differences and the needs of the particular measurement.

Table 14-4. Device parameters for different temperature sensors.

Sensor type / Parameter	RDT	Thermocouple	Thermistor
Accuracy	±0.005–0.05 °C	0.5–5 °C	0.05–0.5 °C
Stability	<0.1% in 5 years	0.1 °C per year	0.1 °C per year
Sensitivity	0.1–10 Ω/°C (moderate)	2.5–25 μV/°C	25–250 Ω/°C (best)
Linearity	Best	Moderate	Poor
Range	−200–850 °C	−190–1821 °C	−102–850 °C
Overall	Highest accuracy over a wide range; best stability	High temperatures; faster response; best economy	Best sensitivity

RTDs which operate on the principle that temperature changes in the sensing element wire cause a small change in electrical resistance, offer the highest accuracy, resolution, linearity and stability, but care is needed in compensating for temperature effects on the signal wires. In Europe platinum wire with a 100 ohm base resistance (pt100) is the dominant RTD configuration and some estimates put worldwide unit sales of RTDs in the region eight to ten million per annum.

Thermocouples rely on the fact that when two dissimilar metals are in contact a small thermoelectric voltage is generated. This technique offers the widest temperature range. However, there are many metal combinations which provide a useful temperature signal, each with advantages and disadvantages, and the leading suppliers provide detailed support literature and regularly publish technical articles to ensure selection of the most appropriate thermocouple for the task. (See, eg, [2], [3].)

14.4.6 Turbidity and Suspended Solids

Turbidity and suspended solids measurements are important inputs for process control in the water, waste treatment, pharmaceutical, food and beverage industries.

Undissolved matter in liquids, causes light scattering and absorption and account has to be taken of interacting variables such as particle size and distribution, particle concentration,

color changes in the liquid and changes in the refractive indices of the fluid particulates. All these factors combined with the optical geometry of a particular instrument cause measured values to vary between different instruments. Further complications emerge due to fouling of the instrument probe (see below). These are compounded by the need to calibrate against empirically devised standard that do not deal with all of the variables inherent in the measurement. International Standard, ISO 7027, and the Water Industry Specifications WIS 7-00-00 [4, 5] attempt to reduce the complexity and provide a common basis for making a measurement and interpreting results.

Fouling of the probe can be severe but a good knowledge of the process and careful siting can significantly contribute to the reliability of the measurement. For example, a probe with low flow across horizontal detectors will suffer from sedimentation while a flow of 1 m/s across a flat faced probe generates self cleaning. To clean optical windows a wiper may be programmed to operate at intervals, on demand or can be triggered to operate when process conditions reach some preset state. Useful information concerning the validity of the measurement is obtained by monitoring the effects of such cleaning cycles.

Turbidity or solids content is frequently a surrogate measurement for particle size distribution, which is being recognized as an important process control parameter, and a number of laser based instruments are entering the market. The on-line instruments make measurements on part of the flow only and even the particle size distribution measuring instruments which can be inserted directly into the medium are affected by the normal stratification problems.

Some American States legally require particle size monitoring in waste outlet streams and it is expected that this idea will be widely adopted. Prices are in the range of 30000 pounds to 40000 pounds at present, but important loops such as flocculation control and filter backwash cycles can be controlled on performance rather than timing. This saves money and helps to compensate for the high cost of the instrumentation (further reading recommendations: [6, 7]).

14.5 Emerging Trends in Process Control and Sensing

There is a considerable volume of diverse research and development in new process sensors and in this section some important trends are discussed. Often it is the case that the enabling technology for new instruments comes from other industries and market sectors. At a general level the microprocessor permits systematic corrections and complex real-time calculations to be performed at modest cost. The low cost of personal computers is also having a dramatic effect on the architecture of process control systems.

14.5.1 Virtual Sensors and Neural Networks

It has been common practice for many years to derive the values of a given measurand from a number of other measurements, and in some cases these derived measurements are used for closed loop control. With the increased needs of plant and enterprise wide control systems this approach is becoming more prominent with, for example, Fisher Rosemount advertising "intelligent sensors" and Foxboro promoting patented "dynamic performance measures".

Fisher Rosemount claim to use a neural-network based "virtual sensor" technology to predict key process variables that cannot be measured in real-time, thereby allowing the continuous control of critical process parameters that have until now defied cost-effective on-line measurement. (These key parameters might have been plagued by excessive noise, severe nonlinearities or multivariable interactions.)

The Foxboro approach to this measurement concept retains the individual sensor but derives measurements which provide dynamic feedback on key indicators of plant performance such as efficiency, quality, cost, emissions and margins. The ability to obtain derived measurements in real-time allows the user to consider the measurement as if it came from a single sensor; despite the fact that the calculations may require a complex combination of object management and a distributed, real-time relational data base. Considerable benefits can be gained from these new closed loop controls but, perhaps more importantly, the process operator and the process manager have a clear image of the action being taken.

Neural networks are now being used to generate virtual sensors that perform process measurement in circumstances where no other cost effective, on-line measurement is available and it has already been demonstrated that they can:

- eliminate the need for complex, inferential control strategies,
- reduce many multivariable problems to simple solutions,
- replace analyzers that need a high level of continued maintenance caused by difficult process conditions,
- predict a controlled variable in seconds rather than the minutes usually required by analyzers.

Neural networks are inspired by biological systems that can be "trained" to take ordinary "process" measurements and predict variables that are difficult to measure. Also they can learn relationships that cannot be modelled statistically or from first principles as is illustrated by the examples summarized below. Density transmitters, developed for the oil and gas industry, have a resolution which is easily able to detect changes in water quality. A number of workers in the water and waste treatment industry have noted the possibility of using such density transmitters as a toxicity detector on influents and effluents. Correlating the density changes with each of the random and systematic variables is a challenge but commercially valuable direct signals are expected to result from the careful application of neural nets to an holistic understanding of the contributing variables. Alternatively, accumulated process unit or plant data is used to determine relationships between known variables and the variable of interest. A neural network can then be used to predict current values or the time remaining until a particular value is reached.

Another example of a virtual sensor is of the *N*eotronics *O*lfactory *S*ensing *E*quipment (NOSE) which resulted from a DTI LINK Project between Neotronics, Bass Brewing and Warwick University. The response of an array of, typically, twelve conducting polymer sensors is processed by a neural network and has been shown to discriminate between types of wines, beers, perfumes etc. Although the polymer sensors can be manufactured repeatably, the responses of individual elements cannot yet be controlled. Different combinations of the polymer sensors make the instrument application specific whilst demanding large training sets for the neural network. The technique does, however, show considerable potential for quality control.

The use of neural networks for on-line process measurement is relatively new but a large amount of work is under way. Very large data sets are required, however, to train even the simplest neural network. Virtual or derived sensors have attractive possibilities and a concentration on the actual measurements needed for profitable and safe process control will be of great benefit to the process measurement industry. New multi-discipline skills and an ability to take an holistic view of the opportunities is required if profitable commercial results are to be obtained.

14.5.2 Smart Sensors

The first smart process transmitters were introduced in 1983 by Honeywell and recent reports indicate that smart transmitters have taken more than 50% of the process transmitter market in the USA in 1993 with continued growth expected. While this statistic is dominated by the contribution from pressure and differential pressure transmitters it is clear that the "smart approach" to process measurement is now established and, with little or no price premium, has become the de facto standard.

Virtually all smart instruments provide core functions, such as: control of range, zero-span adjustments, diagnostics to verify functionality, memory to store configuration and status, information such as unit identity and process materials used in its construction, plus user entered data such as tag number and a log of maintenance operations. There is an associated sharp increase in the amount of information being obtained from the process in parallel with a growth in the information and instructions being sent to the on-line sensor. Two way information flow between the process sensor and the control system made possible by smart sensors is developing rapidly with the scope for improving operational efficiency and quality control in all areas of the processing industry. Considerable economic benefits have already resulted from the use of these techniques.

Two way information flow has to be managed and the HART (Highway Addressable Remote Transducer) protocol, pioneered by Rosemount, is compatible with the huge world wide installed base of 4–20 mA transmitters [8, 9]. Now, independent of Rosemount and serviced by an open club, HART has become the standard for digital communication between smart transmitters and the software packages available to manage both the communication and the data. These low cost software packages hold configuration and measurement data, provide trends, maintenance information and audit trails and can significantly reduce the cost of ownership of measurement instrumentation.

Although HART can claim to be the current standard for the process control industry and to have strong commercial support, it does not meet all the future needs for distributed control systems. There is little doubt that the sensor communication standard for the process control industry will emerge from the various versions of the fieldbus. The 1993 ISA show in Chicago was dominated by fieldbus lectures, seminars and demonstrations with the leading groups claiming a migration path to the eventual International Standard (IEC) intended to arise from the ISA SP50 Standard Committee. Manufacturers and users are confused and irritated by the length of time it has taken for the ISA SP50 Committee to develop a fieldbus standard and some observers suspect that a standard will emerge from one of the low cost data highways already available.

The increase in complexity of process sensors is expected to continue and this presents a considerable challenge to the large number of small entrepreneurial companies in the process measurement industry. A EUREKA project JAMIE (Joint Analogue Microsystems Initiative of Europe) is aimed at developing an ASIC capable of interfacing with a wide range of sensors and signal characteristics and will incorporate such functions as analog-to-digital conversion, signal processing, self-calibration plus a variety of output formats. The initiative will provide enabling technology for small to medium manufacturers of sensors who cannot justify the development costs of mixed mode ASICs. In addition it will also meet the needs of larger volume users and if claimed that over five hundred companies have expressed interest. First prototypes are due in the second half of 1994.

14.5.3 Surrogate Measurements

Most processes have a key variable that cannot be measured on-line in real time but which, if included in the control strategy, would make a significant difference to process performance and safety. Measuring another parameter which closely correlates with the required parameter is an alternative approach and is known as surrogate measurement.

An example of a surrogate measurement is one in which the equilibrium moisture content of a package biscuit is the required variable (Fig. 14-1). In practice the surface moisture of the biscuit is measured at the exit of the oven using an on-line infrared moisture meter and the strong correlation between this signal and the in-packet equilibrium moisture exploited in order to make the required measurement. Another example is the measurement of the solids content of thickened activated sludge as it is transferred to sewage digesters using consistency meters. The latter measure viscous forces in order to infer solids content and have been widely used in the pulp and paper industry. This approach has now been adopted to generate valuable control strategies in the sewage treatment process. Surrogate measurement is also used to measure organic materials in water. This is difficult to do directly due to the difficulties correlating the many measurement techniques, problems of obtaining representative samples and the absence of reference standards. Surrogate measurements such as total organic carbon (TOC), total oxygen demand (TOD), chemical oxygen demand (COD) and respirometry are used on-line to determine the biochemical oxygen demand (BOD) of waste water and industrial

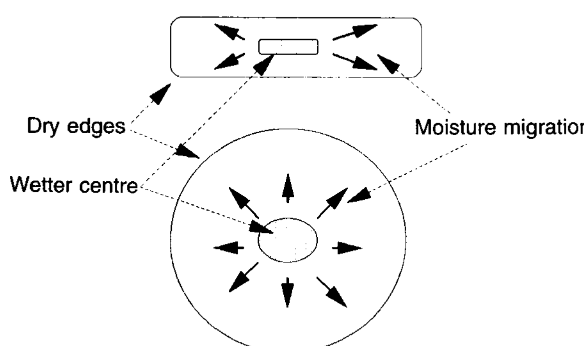


Figure 14-1.
Moisture distribution in freshly baked biscuits.

effluents. The real-time measurement of BOD allows feed-forward control from the inlet and zonal control in the digestion process with claimed pumping cost reductions of more than 10%.

Respirometers, which measure the oxygen demand of a process stream by monitoring the digestion rate of process samples by known volumes of activated sludge, are becoming popular on-line instruments in the control of and monitoring of waste water streams. Response times are between five and thirty minutes and are claimed to reflect accurately the biological activity of the sampled stream. Several manufacturers of respirometers have shown that the effect of toxicity in the sample stream on the biological activity of the respirometer's activated slude can be used as a process measurement. Toxicity detection provides a valuable feed-forward and protection measurement and has been the primary use, to date, of the respirometer. However the ability to measure the respiratory activity of activated sludge and return activated sludge has enabled the development of new control strategies to control pumps, aerators and process efficiencies.

Measurement of the oxygen demand of a digestion process has been shown to enable significant cost reductions and permit process plant to be driven harder. A recent small market study by the author suggests, however, that there is a widespread misunderstanding about the relationship between TOC, TOD, BOD, COD and respirometry measurements and BODs and this may delay the growth of an important niche market [11, 11, 12].

14.5.4 New Sensors and Applications

The increase in the number and type of measurements being made for process control has been noted in earlier sections and this trend is expected to continue. Many measurements, which are currently made by taking samples to a laboratory or locally mounted analyzer are beginning to be made by industrially hardened, unattended, robust instruments. A few examples are discussed below.

14.5.4.1 Color

Color measurement is complex and an important element of product quality control. The vast majority of industrial color measurement are at present carried out in the laboratory or at-line. Genuine on-line measurement is comparatively rare but demands for improved repeatability, deskilling and the reduced manning of production processes is forcing manufacturers to consider continuous on-line color measurement. The world wide color measurement industry was estimated in 1988 by the Core Consulting Group at close to 6 billion dollars. More recently Cachet Market Services estimated the European market to be 1.7 billion dollars and in the wider European area 2.5 billion dollars.

The major suppliers of portable color measuring instruments, Minolta, X-Rite and Gretag all now offer low cost spectrometers with the ability to select illuminant and observer angle, with a very large range of color coordinates and error definitions. The food and beverage industries also use to advantage on-line color measurement, but prices are too high at present to allow routine on-line use.

14.5.4.2 Fugitive Volatile Organic Compounds

The release of volatile organic compounds (Fugitive VOC) into either the atmosphere or water is subject to increasing environmental control. Despite measurement difficulties 100s a year of mass spectrometers with a price of around 100000 dollars are already used by the hydrocarbons industry and this number is expected to grow.

Sparging of the water and chromatography on a head space sample can provide a sensitive indication of an influent or effluent stream to levels of ppb. The National Rivers Authority (NRA) in the UK is interested in the use of this technique to monitor pesticides and pharmaceuticals. These methods are widely used in Canada and the USA. Fourier transform infrared spectrometry (FTIR) plus sparging (as demonstrated at ISA exhibition in Chicago) is claimed to be much faster than the equivalent gas chromatography technique and enables up to ten constituents to be deconvolved. A micro gas chromatograph using a micromachined silicon sample valve was shown at the same exhibition and, at a cost of around 16000 dollars, claimed to be the lowest cost, smallest and fastest gas chromatographs in the world.

14.5.4.3 Odor

Odor measurement for both process and environmental measurements is now becoming commercially available. Current instruments are based on neural networks processing the information from an array of doped conducting polymer sensors (see Section 14.5.1) and although introduction prices are in the region of 20000 dollars a substantial market is forecast. Applications range from the control of the smell of beverages and food to the monitoring of fugitive emissions on industrial sites and sewage treatment works. Since smell is an important sense for both pleasure and annoyance, the possibility for blending to achieve a specific result should provide significant market opportunities for on-line odor measurement.

14.5.4.4 Wet Chemical Analyzers

A number of on-line measurements are made by performing titrations and hence rely on the controlled addition of chemicals; so called wet analyzers. The adoption of smart technology and built-in intelligence can sharply reduce operating costs. Skalar and Bran & Luebbe, for example, claim that they added modems for remote interrogation of their ammonia, nitrate and phosphate analyzers to meet the needs of the German market. Newly emerging "wet chemical" analyzers have comprehensive microprocessor controlled displays, self diagnostic techniques and "self optimizing endpoint determination analysis" which is said to reject disturbances to the accuracy of the measurement. The need for resupply of chemicals and routine maintenance is claimed to take less than 20 minutes per month. Considerable efforts are, however, being made to remove the need for wet chemicals from on-line process measurement and the steep fall in the prices of GC/MS and spectrophotometers means that these techniques could provide a viable alternative.,

14.6 Conclusion

The number of sensing techniques that generate measurements for process control are expanding in parallel with the number of new parameters it is required to measure. The functionality of sensors will continue to increase and the development of serial communication techniques is already changing the architecture of process control systems.

The profitable design, development and application of process measurement sensors require multi-disciplinary skills and the complexity of the commercial instruments is increasing rapidly. There is at the same time a clear trend to reduce the life cycle costs and make the technology transparent to the user.

14.7 References

[1] Finkelstein, L., Grattan, K. T. V. (ed.), *Concise Encyclopedia of Measurement and Instrumentation,* New York: Pergamon Press.
[2] *Rules of Thumb for RTDs.* Available from Moore Industries.
[3] *Temperature; Issues and Answers.* Available from Moore Industries.
[4] *ISO 70927-1984.*
[5] *Water Industry Specifications. WIS 7-00-00,* Medmenham, Bucks: WRc.
[6] Spair, J. A., *The A-B-C's of NTU's, a Primer on Turbidity,* Redwood City, CA: Advanced Polymer Systems.
[7] Hill, B., *A Turbidity Primer.* BTG (UK) Ltd.
[8] *The HART Book,* Titchfield, Hants: GGH Marketing Communications.
[9] Bowden, R., *Hart field communications protocol. A technical description,* Fisher Rosemount.
[10] *Water Quality Measurement; the Impact on the Bottom Line,* SWIG.
[11] Jacobi, Th., Fussa, A. D., *Low Level Process Waste Water Monitoring Using Total Oxygen Demand,* paper presented at the '93 ISA Chicago Conference.
[12] *Biotecto TOC Analyser.* Field trial report available from BTG (UK) Ltd.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

15 Medical and Healthcare Sensors

V. M. OWEN, SciTec Management Consultants Ltd., Milton Keynes, UK

Contents

15.1	Introduction	432
15.2	Electrochemical Sensors	432
15.2.1	Ion Selective Electrodes	433
15.2.1.1	Applications	433
15.2.1.2	Market	434
15.2.2	ISFETs	435
15.2.2.1	Applications	435
15.2.2.2	Market	435
15.2.3	Gas Electrodes	436
15.2.3.1	Applications	436
15.2.3.2	Market	438
15.3	Optical Sensors	438
15.3.1	Waveguides	438
15.3.1.1	Applications	438
15.3.1.2	Market	439
15.3.2	Surface Plasmon Resonance	439
15.3.2.1	Applications	439
15.3.2.2	Market	440
15.3.3	Red and Near-Infrared Sensors	440
15.3.3.1	Applications	440
15.3.3.2	Market	441
15.4	Biosensors	441
15.4.1	Enzyme Electrodes	442
15.4.1.1	Applications	443
15.4.1.2	Market	443
15.5	Neural Activity Sensors	444
15.5.1	Applications	445
15.5.2	Market	445
15.6	Future Challenges	445
15.6.1	In vivo Sensing	446
15.6.2	Non-Invasive Sensing	448
15.7	Conclusion	449
15.8	References	449

15.1 Introduction

The healthcare professional needs information about the physical and chemical/biochemical status of the patient. This information is used to diagnose disease or to monitor the progress of treatment for the disease. Traditionally physical measurements such as temperature and blood pressure have been performed by the physician at the patient's side, while chemical and biochemical measurements of analytes such as serum sodium and cholesterol have been carried out in a centralized laboratory. Both of these approaches have the disadvantage that they give readings taken at one point in time. If changes in the patient's condition are to be monitored the procedures must be repeated with the attendant inconvenience to the patient and physician. Another disadvantage of using the centralized laboratory, with its current technology, is the time taken for the physician to receive the results.

Sensor technology opens up the possibility of continuous real-time or near real-time monitoring of a patient so that changes can be detected quickly in life threatening conditions or the effect of therapy can be seen immediately.

Sensor technology is opening up the possibility of providing analytical equipment which is easy to use and will provide quality analysis away from the central laboratory. Some of the new developments are exploring the possibilities of non-invasive procedures for the monitoring of biochemical analytes – an escape from the need to take and process blood samples.

The market for *in vitro* diagnostics, most of which are used in the laboratory, is maturing and by 1992 had reached about $ 13bn per annum. Of this nearly $ 15 bn was immunodiagnostics with nearly $4.5 bn for clinical chemistry. It is considered that the immunodiagnostics market may continue to grow at about 10% per year until the end of the decade. Beyond the year 2000 it is expected that the most rapid areas of innovation and growth will be in new methods of diagnosis and patient monitoring which will be closer to the patient and less-invasive.

This review of sensors for use in medicine and healthcare will examine the application and market for the different groups of sensor technologies. The technologies considered include electrochemical sensors, optical sensors, biosensors, and neural activity sensors. Then there is an examination of the future challenges including the requirements for *in vivo* sensing and the technologies for non-invasive sensing.

15.2 Electrochemical Sensors

The majority of the electrochemical sensors used in healthcare are ion selective potentiometric electrodes used for the measurement of specific ions such as sodium, potassium or hydrogen, or for the detection and measurement of specific gases such as carbon dioxide or ammonia. There has been considerable research interest in the development of ion selective field effect transistors (ISFETs) for these measurements; however these research devices have not yet reached the expected level of commerical success.

15.2.1 Ion Selective Electrodes

A potentiometric electrode is used in conjunction with an ion selective membrane to produce an ion selective electrode (ISE). ISEs for the measurement of sodium, potassium and lithium were developed in the 1920s and a glass electrode for the selective measurement of hydrogen ions, and therefore pH, has been commercially available since the mid 1930s. The technology was developed to include calcium selective electrodes, which were commercialized in about 1970.

15.2.1.1 Applications

The ISE which first entered the healthcare market was the pH electrode. It has been in common use in the clinical laboratory for over 35 years and has been used on whole blood samples. It is generally employed, in conjunction with gas electrodes for the measurement of pCO_2 and pO_2 (see below 15.2.3), to monitor patients during periods of critical care.

In contrast to the pH electrode, uptake of other ISEs in the clinical laboratory was slow. This was probably due to the availability of flame photometric technology for sodium and potassium analysis [1] and to the fact that ISEs measure ion activity and not the total concentration of the ion. The activity of an ion is an indication of the amount of free unbound ion present in the sample. The proportion of an ion which is in the free state is affected by the dilution of the sample and by the other constituents of the sample such as proteins and lipids. These constituents tend to bind ions. These factors mean that the results obtained by measuring ion activity (using an ISE) do not directly correlate to the analytical results obtained by flame photometry. This led to difficulties in standardizing and interpreting the results.

If the sample is diluted most of the matrix effects are removed and the result obtained using an ISE is comparable to flame photometry. For this reason clinical samples are frequently diluted before analysis using ISEs. Alternately a "factor" is applied to the analytical reading to convert it to a "total" ion concentration.

It appears that the extra information obtained by assaying for an ion activity as opposed to an ion concentration has not been found to give clinically useful information. The exception to this is calcium. Calcium ion activity is a more meaningful assay than total calcium, therefore calcium assays are carried out on undiluted samples.

ISEs are frequently used in critical care where there is a requirement for rapid test results. One study has shown that the turnaround time for receiving analytical results can be decreased from 25 minutes to less than three minutes by having an analytical device in the operating suite rather than using the services of a laboratory even when this was adjacent to the operating room [2].

The need for fast access to analytical results is particularly marked in open heart surgery patients. These patients are frequently older than the usual hospital patient, critically ill and may have multi-organ failure such as liver and kidney dysfunction and shock. They may also be receiving many drugs. The complexity and urgency of their disease means there is a need for timely analytical results, for the analytes listed in Table 15-1, for the management of the patient before, during and after surgery.

Table 15-1. Analysis necessary for the open heart surgey patient.

pH
pO_2
pCO_2
K^+
iCa^{++}
Hemoglobin or hematocrit
Coagulation studies (in managing heparin use)

15.2.1.2 Market

An ISE module is available for most large automated clinical chemistry analyzers used in the central laboratory. These modules may contain just the sodium and potassium ISEs or may have four options by including calcium and chloride. There are approximately 12000 large automated clinical chemistry analyzers in use in Western Europe and the annual market for these instruments is about 1300 units per year.

However, the majority of ISEs for clinical use are incorporated in small instruments which can be used in the laboratory, on the intensive care ward, or in the operating room, close to the patient. Table 15-2 lists manufacturers of small user-friendly instruments containing ISEs. These Instruments have been designed for operators with little or no knowledge of the technology. The number of calibration and maintenance steps performed by the operator is reduced because the instruments are designed to perform self-calibration and routine washing steps.

In the case of the Mallinckrodt Sensors GEM^{TM} the sensors, calibrants and flush fluids are packaged in a single disposable cartridge. This sensor system can be used on-line, directly on samples drawn from the patient via extra-corporeal tubing.

Table 15-2. Small user-friendly instruments incorporating ion selective electrodes.

Manufacturer/supplier	Instrument	Analyte
Beckman Instruments	Lablyte Electrolyte Systems	Na^+, K^+, Cl^-, Ca^{2+}, Li^+
Ciba Corning	600 series	Na^+, K^+, Li^+, Cl^-
Instrumentation Laboratory	System 501	Na^+, K^+
Kone	Microlyte Ion Selective Analyzer	Na^+, K^+, Li^+, Cl^-, Ca^{2+}, pH
Nova Biomedical	Nova Stat Profile	Na^+, K^+, Cl^-, Ca^{2+}, pH
Amdev	Lytening series	Na^+, K^+, Li^+, Ca^{2+}
Mallinckrodt	GEM	Na^+, K^+, Ca^{2+}, pH, blood gases
Radiometer	ABL	Na^+, K^+, Ca^{2+}, blood gases
AVL Scientific	AVL 980 series	Na^+, K^+, Ca^{2+}, blood gases

The market for these instruments is potentially very large as, in theory, one is required for each patient receiving intensive care. In the UK there are about 2500 intensive care beds spread over 500 intensive care units.

15.2.2 ISFETs

It was expected that FETs would have an advantage over conventional potentiometric electrodes given their planar construction, ease of large scale production, and the relative simplicity with which a multisensor device could be constructed. The small dimensions, fast response time, and direct application after storage were also seen as advantages over established technology. Although at least one small self-contained instrument based upon ISFET technology was developed and evaluated [3] FETs have enjoyed little commercial success in medical applications to date. However, these devices are beginning to compete in other markets for selective ion detection such as fermentation and bioreactors.

15.2.2.1 Applications

There is a need for miniature pH and pCO_2 sensors which could be introduced into the various chambers of the heart during cardiac diagnosis. These sensors could help in investigations of the blood flow through the various chambers of the heart.

At the same time there is a need to monitor the pH and temperature of the heart muscle (myocardium) during open heart surgery and heart transplant. The temperature of the myocardium may drop as low as 7 °C during this process and most existing pH probes are not appropriate for monitoring the pH corrected to 37 °C under these conditions.

Another potential application of the pH ISFET is in long term ambulatory pH monitoring of the gut. Presently monitoring of gut pH is carried out using pH electrodes, however these have problems. The electrodes are too large to be passed transnasally; glass electrodes are too expensive to be disposable; antimony electrodes are non-linear over the range required and have a slow response time. A pH ISFET has been developed which is smaller and less expensive to manufacture than an electrode. This ISFET has been shown to perform adequately in the gut environment [4].

15.2.2.2 Market

It is estimated that over half a million open heart operations are carried out worldwide each year and this number is increasing with the aging populations of the developed world. There are over 60000 operations in a year in England on the heart and surrounding blood vessels; of these about half involve open heart surgery.

Every year over 1 person in a thousand in the UK will require a study involving monitoring the pH of the gut. Over one third on the population may have some symptons of dyspepsia and many of these will have some symptoms of gastrointestinal reflux.

15.2.3 Gas Electrodes

The most important gas measurements which are made in healthcare are the partial pressures of oxygen and carbon dioxide. Blood gas analyzers based upon gas electrodes have been in use since the early 1960s. Although the ease-of-use and maintenance of the instruments has improved, the underlying analytical principles remain the same.

Oxygen tension is measured by a Clark electrode, a polarographic electrode. Carbon dioxide is measured by a pH electrode used in conjunction with an electrolyte solution and a membrane through which the CO_2 diffuses. As the CO_2 diffuses into the electrolyte solution it reacts with the water to form carbonic acid which dissociates to release hydrogen ions. These ions are detected by the pH electrode.

15.2.3.1 Applications

Oxygen and carbon dioxide measurements are two of the tests regarded as being essential in emergency and critical care. Most of the analyses are carried out on samples of arterial blood and the results indicate the status of the whole patient at the time the sample was taken. However there is a need for continuous monitoring of patients during critical care.

Blood gas instruments which include pH electrodes are found in the central laboratory but they are being increasingly used close to the patient. Blood gas analyzers have been developed for operators who have limited analytical training and no knowledge of the analytical technology. The calibration, quality control and maintenance procedures are designed for minimum operator involvement.

There are moves to bring testing even closer to the patient and analyses may be performed directly on blood withdrawn from the body via extra-corporeal tubing or by inserting sensors into the patient's body. Table 15-3 gives some working definitions of terms used when sensors are employed in contact with patients.

Table 15-3. Terminology for sensors used in contact with patients.

On-line	Blood is withdrawn from the patient, via extra-corporeal tubing, on request or at pre-programmed intervals is analyzed and then goes to waste
In-line	Blood is withdrawn from the patient, via extra-corporeal tubing is analyzed and then returned to the patient
In vivo	The sensor is inserted into a blood vessel or tissue in the patient and readings taken from that blood/tissue. The sensor may be used short-term (a few hours to one week) or long-term (over one week).
Transcutaneous	The sensor is on the patient's skin and readings are taken through the skin.

The need for continuous monitoring is particularly marked in the neonate receiving assisted ventilation where there are risks of hyperoxia (excess oxygen) as well as hypoxia (inadequate oxygen). There are also risks associated with elevated or depressed carbon dioxide levels as the former can cause internal bleeding and the latter can lead to vasoconstriction and reduced blood flow to the brain [5]. Efforts have been made to develop miniature intravascular oxygen

electrodes and these have had some limited success since the 1970s [6]. The electrodes are introduced into the umbilical cord of the neonate. Howevcr, the introduction of any catheter into an artery carries with it a risk of clot formation and these clots can move to other vulnerable parts of the body such as the brain.

There are problems with the fragility of glass pH electrodes when they are used invasively. There have been attempts to develop polymer-based ion selective membranes to overcome the risk of breakage. It appears that good results have been obtained using this technology [7].

The complications due to invasive monitoring procedures led the drive to find non-invasive methods of monitoring blood gases. In addition there is a need to monitor the blood profusion and oxygenation status of specific areas of the body such as a leg during healing of a wound. Since the early 1970s transcutaneous electrodes have been developed for these applications [8]. These sensors contain robust miniature oxygen and carbon dioxide electrodes, and heaters. They are applied to the skin of the patient which is warmed to 43 °C to 45 °C. Warming the skin has the effect of increasing the blood flow to the tissue immediately below the sensor and liquifying the lipids in the skin so that oxygen and carbon dioxide will diffuse across the skin.

Transcutaneous electrodes are particularly applicable and accepted for monitoring the neonate [9] as their skin is thinner and has a higher density of capillaries than that of an adult. In the neonate it is found that the transcutaneous values obtained for oxygen correlates more closely with the arterial level than is the case in adults.

This use of transcutaneous electrodes for monitoring babies has even been extended to use during birth. Transcutaneous electrodes have been inserted through the vagina and attached, by means of a suction ring, to the scalp of the baby [10].

In the adult, transcutaneous blood gas sensors are being used in intensive care. In addition they are finding other applications to assess the blood flow to a particular tissue or part of the body. Their applications include assessment of skin flap viability, drug evaluation, prediction of wound healing and selection of amputation level in peripheral vascular disease [11].

There is interest in extending the use of transcutaneous oxygen electrodes to monitor biochemical changes during exercise. This would overcome the problems associated with trying to obtain arterial blood samples for conventional oxygen partial pressure measurements during an exercise period. However, there is a delay of over two minutes before the transcutaneous electrode responds to a subject's change in oxygen uptake. This delay means that the transcutaneous device response may be too slow to detect when a subject cannot sustain exercise due to lack of oxygen.

This slow response time has also limited the use of the transcutaneous electrodes during anaesthesia where it may be necessary to get a fast warning of changes in patient status. Also surgery, anaesthesia and certain drugs used during surgery, can affect the flow of blood to the peripheral blood vessels. This will affect the amount of oxygen reaching the tissues beneath the skin and therefore the oxygen reaching the transcutaneous electrode. At the same time anaesthetic agents such as halothane and nitrous oxide can interact with the surface of the oxygen electrode and cause an upward drift in the reading obtained.

A series of papers have been published dealing with the technical aspects of transcutaneous electrodes and the application of the devices to the patient by The American Society for Testing and Materials [12], The International Electrotechnology Commission [13], American Academy of Pediatrics [14] and the International Federation of Clinical Chemistry [15].

15.2.3.2 Market

Transcutaneous electrodes are available from several commercial companies including Hellige GmbH, Germany; Novametrix Medical Systems Inc., USA; Radiometer A/S, Denmark; and Kontron AG, Switzerland.

15.3 Optical Sensors

A range of technologies have been used to produce optical sensors which can be used in medicine and healthcare. These include optical sensors based upon the conventional technology of shining a light of a particular wavelength from a source to a detector and noting changes due to light absorption by the intervening media. Interest in this approach has been extended recently into using the near-infrared range of wavelengths. Other optical sensor systems which received a lot of interest in the 1980s were based upon technologies using phenomena such as the evanescent component of light as it passes along a waveguide, and surface plasmon resonance.

15.3.1 Waveguides

A light passes along a waveguide or fibre optic by multiple internal reflections. At each reflection a component of the light, the evanescent wave, penetrates the surrounding medium and then returns to the waveguide. As the light only penetrates a fraction of a wavelength outside the waveguide it is only subject to changes on the surface of the waveguide. Changes in the bulk solution do not affect the evanescent component [16]. The changes in optical properties which can be monitored include light scattering, refractive index and light absorption. This system can be used to stimulate fluorescence and luminescence which can be monitored at right angles to the waveguide.

15.3.1.1 Applications

A disposable optical device based upon waveguide technology has been developed for immunoassay applications [17]. The device was designed for use by people with little or no technical training. It can be filled by capillary action to give a defined volume of sample. The baseplate of the device is a waveguide. An antibody is immobilized on this baseplate. Antigens in the sample and fluorophore-labelled antigen are allowd to compete for binding. Light of the wavelength required to excite the fluorophore is shone onto the baseplate causing the fluorophore to fluoresce. Some of the fluorescent light generated by the bound antigen/fluorophore enters the waveguide and is detected at the end of the waveguide. Fluorescence from the antigen/fluorophore complex still in the bulk solution will not enter the waveguide and will not be detected. This means there is no requirement for the washing steps necessary in conventional immunoassays.

This technology was developed by Unilever in the UK and then sold to Serono Diagnostics. Serono have continued to develop the devices but these had not entered the market by the early 1990s.

15.3.1.2 Market

No commercial devices based upon this technology had entered the market by 1993. If novel devices based on this technology are to succeed commercially they must find a niche application that requires quantitative immunoassays which are to be performed outside the central laboratory eg. close to the patient. The majority of immunoassays which are performed close to the patient or as a patient self-test are to do with conception. Self-testing pregnancy and fertility kits are available over-the-counter (OTC) (the European market is worth over £150 m and growing at about 25% per annum). These kits only offer a binary positive/negative test result and do not give a quantitative continuous reading over a concentration range.

The increase in assisted human conception is driving a requirement for increased assay of hormones of the reproductive cycle. The development of the follicle is monitored using a panel of hormone assays [18]. The results of these assays need to be known quickly so the patient, staff and operating room can be prepared for oocyte recovery.

The demand for *in vitro* fertilization (IVF) is growing. In the UK nearly 10000 patients underwent this treatment in 1990 and the number is increasing by about 10% per year.

15.3.2 Surface Plasmon Resonance

Surface plasmon resonance involves the movement of electrons in the surface of a metal conductor excited by light of a certain wavelength and at a particular angle. Energy is transferred from the incoming photons to electrons in the metal surface. On a diffraction grating, it has the effect of absorbing light energy at a particular wavelength, which depends upon the dielectric contstant of the medium in contact with the metal surface. This produces a gap in the spectrum of the light reflected from the grating. Absorption can be measured as a dip in reflectance intensity at a sharply defined angle. This is called the resonance angle. The dielectric constant is affected not only by changes in ion absorption at the surface but also by the nature of the biological molecules binding to the surface.

15.3.2.1 Applications

Surface plasmon resonance has the advantage over other techniques that it offers good sensitivity without the use of labels. Also it can used with complex mixtures with no need for prior purification. However the initial applications of this technology have been in research, as a tool to study interactions of biomolecules. The technology can be used to study the affinity, specificity, kinetics, multiple binding pattern and cooperativity of the interactions of molecules such as antibodies and their antigens.

15.3.2.2 Market

The first commercial instrument based on this technology was the Pharmacia BIAcore (Biospecific Interaction Analysis) which was introduced onto the market in late 1990. The system has a full supporting processing unit with system control and data evaluation, however the system was expensive at about $ 134000. It was expected that only about 12 systems per year would be sold.

In 1993 Fisons Instruments brought out an instrument based upon similar surface plasmon resonance technology.

15.3.3 Red and Near-Infrared Sensors

Devices using light in the red and near-infrared regions can be used to monitor the oxygen saturation of blood. Oximetry is proving to be one of the most successful advances in the technology for monitoring the blood oxygen content. Oximeters are constructed using small light emitting diodes (LEDs) which are capable of giving out enough light to penetrate even darkly pigmented skins. The light is detected using a photodetector which is not specific for any particular wavelength.

Oxyhemoglobin and deoxyhemoglobin absorb light at known functions of wavelengths. It is possible to determine the relative percentage of each constituent and thus the oxygen saturation. To obtain true arterial oxygen saturation values and to filter out absorption due to venous blood and other tissue constituents, signals from only the pulsatile arterial blood are used in the measurements.

15.3.3.1 Applications

Non-invasive oximetry is used to measure the oxygen saturation of blood by observing absorption of optical waves as they pass through the skin and interact with the red blood cells. The technique was first introduced in the 1930s. However applications were limited by the use of a single wavelength which limited the ability to compensate for a range of skin and tissue constituents which interfered with the measurement of oxyhemoglobin. In the late 1940s a device using two wavelengths was described, and later a device utilizing eight wavelengths was commercialized [9]. A more recent refinement is the pulse oximeter. This instrument uses two wavelengths of about 660 nm in the red region and anywhere between 800 and 1000 nm in the near-infrared region. A pulsatile light signal associated with the volume changes in cutaneous blood is used. This has the effect of eliminating the 'noise' due to the pulses of blood through the vessels. Pulse oximetry also uses reflections from the main arteries rather than just information from the surface capillaries. This reduces the need to arterialize or increase the blood flow to the capillaries.

Problems still remain with the non-ideal characteristics of LEDs and with light scattering, however modern micro-chip computing power can be used to compensate for these. The spectral output from LEDs is affected by temperature change but these problems can be overcome by including a temperature sensor in the device.

The use of two light sources limits the ability to compensate for carboxyhemoglobin (COHb), methemoglobin (MetHb) and bilirubin. In most patients the interference from these other compounds does not contribute significantly at the wavelengths used and therefore does not limit the clinical use of pulse oximetry. However there are certain categories of patients with elevated COHb and MetHb for which present oximetry technology is less than perfect. These categories include suicide attempts, patients with a genetic lack of the enzyme methemoglobin reductase, smokers, patients receiving AIDS therapy, and those exposed to the air pollution from heavy city traffic [19]. Another group of patients for which pulse oximetry is sub-optimum is the new born baby. Fetal carboxyhemoglobin has spectral absorption characteristics which can lead to miscalculation of oxygen saturation. Pulse oximetry looks at oxygen saturation using an adult algorithm which is inappropriate for use in neonate. In the adult the risk is hypoxemia (inadequate oxygen supply) while in the treated neonate there is a risk of hyperoxemia (excess oxygen). It is important that this is detected and treated, but pulse oximeters are not well designed for this application.

Near-infrared technology, using wavelengths between 700 and 1000 nm and adequate energy to pass through the head of a baby, is being developed for monitoring the state of tissue and cell oxygenation in the brain. Bedside instrumentation has been developed which is capable of measuring total hemoglobin, indicating the proportion which is oxidized, and indicating the blood flow to the brain. Patents that protect the rights to this sort of technology are appearing [20].

Non-invasive near-infrared technology is particularly attractive for patients who need repeated analyses over a long period of time. Diabetics are the largest group of patients in this category. There have been several reports of technology and patents published dealing with this application. Wavelengths in the range 700 to 1100 nm can be used to transmit through parts of the body such as the wrist or finger, while wavelengths such as 1600 to 2000 nm can be used for reflectance methods. The peaks due to glucose can be picked up using data analysis [21]. Diasense announced the start of clinical trials in June 1993 at the same time their patent was published [22].

15.3.3.2 Market

In the UK, of the 700000 babies born annually, about 7000 or 1 in 100 babies require intensive care from delivery. These are mainly premature babies and very low birth weight babies (0.7% of babies are classified as being of very low birth weight with a weight below 1.5 kg) [23].

For information on the market for monitoring diabetic patients see Section 15.4.1.2.

15.4 Biosensors

Biosensors, defined as analytical devices incorporating biological-derived sensing elements in intimate contact with or integrated with a physical transducer, have been promoted as devices which can fit a vast range of applications including some in medicine and healthcare. However, their commercial success has been limited.

Figure 15-1 indicates the potential combinations of transducers and biological components which could be used to construct a biosensor. Many of these combinations have been investigated for clinical applications and a recent review has assessed their advantages, disadvantages and the state of the art for use in a glucose biosensor [24].

However, of all the potential combinations of transducers and biological components only the enzyme electrode has achieved commercial success.

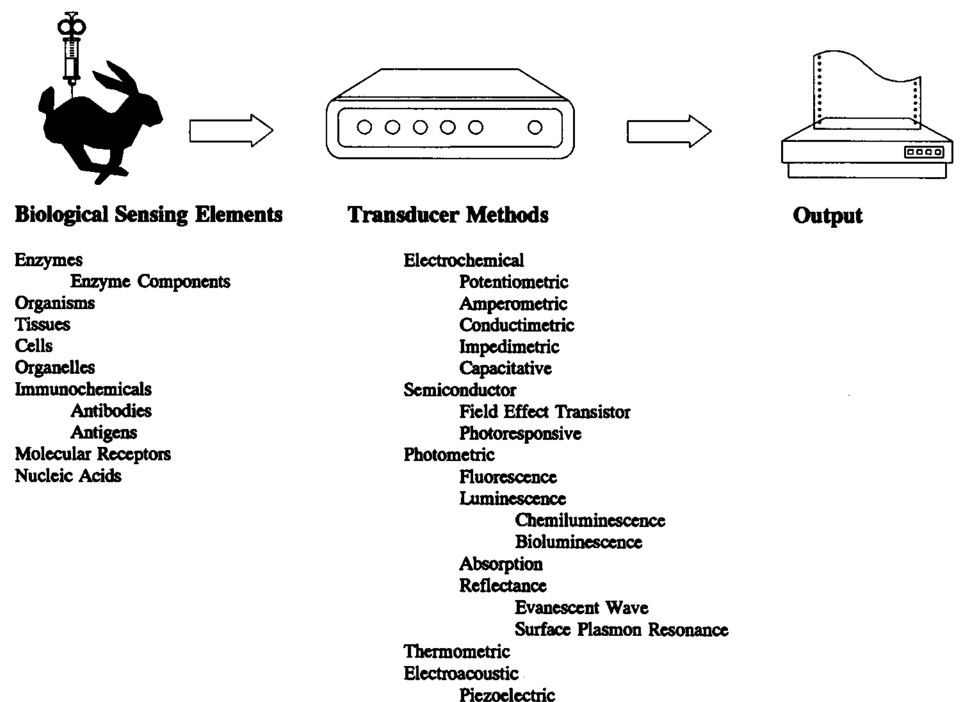


Figure 15-1. Potential combinations of transducers and biological components which can be used to construct a biosensor.

15.4.1 Enzyme Electrodes

Virtually all the enzyme electrodes used in medical applications are glucose electrodes based upon glucose oxidase used in combinations with an oxygen electrode, a hydrogen peroxide electrode, or with medicated electron transfer. Other enzyme electrodes which have been introduced onto the market include detectors of lactate based upon a hydrogen peroxide electrode used in combination with lactate oxidase immobilized on a membrane, and a urea electrode incorporating the enzyme urease and an ammonia-selective electrode.

15.4.1.1 Applications

Although the measurement of glucose forms one of the battery of tests used in general healthcare its most important and largest application is in monitoring the blood glucose levels in diabetes.

There are two main sites for this monitoring of diabetics. It may be performed in a clinic for diabetics where a dedicated biosensor instrument such as the YSI 2300 STAT glucose analyzer or the more automated Eppendorf ESAT 6660 is used.

However, the largest proportion of glucose testing is patient self-testing using dry chemistry reagent sticks. These sticks were introduced in the 1950s to test for glucose in urine and by the 1960s the technology had been extended to include semi-qualitative assays of glucose in serum or plasma. They are now capable of quantitative measurements in whole blood. In 1988, MediSense Inc, launched the ExacTech, the first biosensor for single-shot glucose determinations using a disposable enzyme-electrode.

These developments in analytical technology combined with the use of education programs, multiple injection regimens, insulin pumps and specially designed pocket computers have gone a long way to help control blood glucose levels in Type I (insulin dependent) diabetic patients [24]. However, patients find the testing laborious and inconvenient. Another problem is that as control is improved to normoglycemia there is a greater risk of hypoglycemia. Other patients never manage to adequately control their blood glucose levels. These problems lead to the need for a real-time glucose sensor which could initially act as a hypoglycemic alarm. This could be used overnight and should be capable of working for about 12 hours [25]. In the longer term there is a need for a glucose sensor which could be linked to an insulin pump in the closed loop system.

Technical progress in hemodialysis, for patients with damaged kidneys, has had the effect of halving the time required for dialysis. It is now possible for a patient to be successfully maintained on three four-hour dialyses per week. Attempts to reduce this time further has led to increased morbidity and mortality, however, it has been demonstrated that new high efficiency dialysis technology can reduce the treatment time drastically [26]. This shortening of the dialysis time increases the risk that if the dialysis is not set up correctly it will not run to the optimal point. At present there are no on-line monitors for dialyzers.

There is a need to monitor the removal of waste products, especially urea, and to monitor the fluid and electrolyte balance by measuring the electrolytes (especially pH and K^+) [27]. Sensors could be used in the dialysis fluid. Monitoring directly in the blood would present problems of biocompatability (see 15.6.1 *In vivo* sensing), but sensors are being developed for this application [28] [29]. Monitoring in the dialysate may mean the measurement of much lower concentrations of analyte. This sensitivity has been reported to be feasible [30].

15.4.1.2 Market

It is estimated that the European market for biosensors for use in healthcare reached nearly $ 40 m in 1992 with an annual growth rate of about 45% during the early 1990s. Most of this growth was due to the single-shot electrode system used to monitor blood glucose in diabetic patients.

About 30 million people worldwide have been diagnosed as suffering from diabetes including between 1 and 2% of the populations of Europe and North America. Worldwide, be-

tween 1000 m and 2000 m self-monitoring blood glucose tests are performed each year. Diabetes is becoming more prevalent and the detection rate is growing. In Britain some 60000 new cases are detected each year and 70% of diagnosed diabetes in the UK used a blood glucose monitoring system.

The market for monitoring blood glucose levels should be given a boost by the results of the Diabetes Complications and Control Trial conducted in the USA and Canada. This showed that tight regulation of blood glucose levels in Type I diabetics, achieved by frequent blood tests and insulin injections (three to four times per day) could very significantly reduce the incidence of complications. When the tightly regulated group was compared to a control group there was a 70% reduction in the onset of diabetic retinopathy, a 60% reduction of neuropathy, and a 50% reduction of kidney desease [31]. It is expected that these benefits seen in Type I diabetes would be repeated in Type II diabetes.

The worldwide population of patients receiving hemodialysis has increased dramatically over the past two decades and is set to rise even further with the aging populations of the developed world. In 1980 it was estimated that there were about 150000 dialysis patients worldwide, this had risen to about 400000 by 1990 and is expected to reach nearly 600000 by 1990. All these patients are potential benefactors of sensors developed for use with hemodialysis machines.

Biosensors have been suggested as the technology that could bring a range of tests into the doctor's office as they offer the possibility for user-friendly test devices. However the development of this market will depend, in part, upon its ability to compete with existing technology in terms of cost. Changes in reimbursement for testing in the American healthcare system are seen as a negative influence on the development of doctor's office testing. In the UK a recent study [32] suggests that the introduction of this form of near-patient testing would *increase* the overall testing rate by 16 percent with little or no impact on patient care. The study examined the use of tests for cholesterol, hemoglobin, γ-glutamyltransferase, electrolytes (sodium and potassium), midstream urine, and Chlamydia (vaginal). Of these only the midstream urine test was found to be cost-effective.

15.5 Neural Activity Sensors

The activity of the brain and nervous system can be detected using non-invasive electrodes which will detect the impulses transmitted by a nerve or the activity in part of the brain. The signals are detected using electrodes sited so they collect the signal of interest and avoid interfering signals from surrounding tissues. For example, in electroencephalography (EEG), electrodes are sited over the cerebral cortex must be positioned to avoid unwanted signals from eye or jaw movement. The interpretation of the data received is difficult and different approaches are being adopted to resolve this problem. Recent developments in this area are the use of density spectral array (DSA) and spectral edge frequency (SEF). These appear to be giving a style of data output which is relatively easy to interpret and the hardware required is not overly expensive [33].

15.5.1 Applications

The electrodes used for detecting neural activity are applied to the skin, particularly on the head, by means of a gel which gives adequate electrical contact. The electrodes are used during an examination of brain activity following an accident or in disease such as a stroke or multiple sclerosis.

There is interest in expanding the use of neural activity sensors in anaesthesia and neurosurgery. During anaesthesia it is difficult to monitor the depth of anaesthesia in the patient paralysed by drugs. Existing tests of patient's responses are difficult to standardize and it is difficult to relate them to the state of awareness in the patient [34].

15.5.2 Market

The development of a market for nerve activity sensing in anaesthesia and analgesia will depend upon a better understanding of which signals need to be monitored and the significance of the signals. However when the technical issues have been resolved the market of monitoring depth of anaesthesia may be driven by lawyers and court cases. A patient who has not remained fully unconscious during an operation or who considers that they have suffered excessive pain might be able to sue an anaesthetist. The anaesthetist will be under pressure to use the state-of-the-art neural activity monitors.

Neurosurgery tends to be a highly specialized discipline which is carried out in a limited number of centers of excellence in each country. For example there are 33 hospitals in the UK which have specialist neurosurgery intensive care units and there are about 40000 neurosurgical operations per year. Sensors for monitoring nerve activity during surgery will tend to be concentrated in these centers of excellence.

15.6 Future Challenges

The demand is for sensors which are capable of real-time patient monitoring. These should cause the minimum degree of disturbance to the patient. The two extremes of this approach are the use of a sensor which will work long-term in the human body with an infrequent requirement for service and calibration. This approach is analogous with the modern pacemaker which can function "unattended" for 10 years. The other extreme is the non-invasive or non-touch sensor which can monitor the patient with no discomfort or inconvenience to the patient. This is the approach seen in futuristic films such as Star Trek: The Next Generation!

A recent report [35] has identified best business opportunities for advanced sensors in healthcare. These are listed in Table 15-4.

Table 15-4. Best opportunities for advanced sensors in the healthcare sector.

In the early 1990s:
Extra-corporeal hemodialysis monitor
Fitness monitor
Cholesterol sensor
Microbiology screening sensor
Fertility hormone sensor
Neonatal blood oxygenation monitor
Cardiac muscle pH (ISFET) sensor
Cardiac chamber pH sensor
Nerve activity – neurosurgery monitor
Optic nerve blood flow sensor
Longer term:
Lipoprotein sensor
In vivo glucose monitor
Non-invasive glucose sensor
Bioelectronic/biomagnetic sensor

15.6.1 *In vivo* Sensing

In vivo sensing can be considered as the ultimate goal for monitoring a patient's long term condition. It opens up the door to closed-loop systems of drug delivery and to mimicking the body's own systems of feedback control.

The possibility of making direct chemical measurements in biological tissue has been recognized since the 1950s [36] when early microsensors were used, in research, for the measurement of pH and oxygen. The transfer of this technology into clinical use has been difficult. *In vivo* sensing is finding only limited applications such as monitoring umbilical cord pH at the time of delivery of a baby.

In vivo sensors may be used intro-vascularly (in the blood vessels) or through the skin into the tissues. The latter approach is considered desirable as changes may take place in peripheral tissues before the changes are reflected in the blood. However attempts to correlate the results obtained in the tissues and in the blood have not always been easy and this may be due to the tissue's reactions to damage or its reaction to a foreign body, the sensor.

The results of a multinational survey by the Biomedical Engineering Committee, a Concerted Action on "Chemical sensors for *in vivo* monitoring" are listed in Table 15-5 [37].

The same survey also listed some infrequently requested applications for *in vivo* sensors listed in Table 15-6.

Although it may not be possible to specify the ultimate characteristics of an implantable *in vivo* sensor, specifications for a glucose sensor for use in diabetics have been drawn up [38] (see Table 15-7).

Biocompatability presents one of the greatest challenges for the designer of the *in vivo* sensor. The sensor surface is likely to stimulate a sequence of events listed in Table 15-8. These events can lead to thrombus formation with potentially fatal consequences.

Even if the body's reactions do not lead to such an extreme outcome there will be some disturbance of the tissue surrounding the implanted sensor. This will affect the physiology and

biochemistry of the tissues around the sensor so that the measurements obtained in this tissue may be unrepresentative of the patient's body. Diffusion of molecules through the fibrin and platelet layer will reduce the concentration of components such as oxygen and glucose at the sensor surface, and may lead to a build up of components such as carbon dioxide.

At the same time the interaction of the body's defense mechanism with the surface of the sensor may affect the performance of the sensor. This will necessitate the recalibration of the sensor which is not easy when it is implanted in the human body.

Table 15-5. Main clinical problems where *in vivo* chemical sensors were considered to be helpful.

Clinical Problem	Analyte
Diabetes mellitus (25)	glucose (24), ketones (20), K^+ (2), insulin (2), Lactate (1), pH (1)
vital function monitoring in intensive care/ anaesthetics/prolonged surgery (16)	O_2 (15), CO_2 (10), pH (8), K^+ (2), electrolytes (unclassified) (2), gases (unclassified) (1), Na^+ (1), glucose (2), hemoglobin (3), osmolality (1), lactate (1)
Renal failure/monitoring dialysis (11)	urea (4), creatinine (2), K^+ (2), atrial natriuretic peptide (1), pH (1)

The numbers in parentheses are the number of responders.

Reprinted from Pickup, J. C., Alcock, S., *Biosensors and Bioelectronics* **6** (1991) 639–646; courtesy Elsevier Science Publishers Ltd.

Table 15-6. Infrequently mentioned applications for *in vivo* sensors.

Clinical problem	Analyte
Septicemia/severe infection (3)	gentamycin or other aminoglycoside (2)
Status asthmaticus (3)	theophylline or aminophylline (3)
Heart failure/arrhytmia (3)	lignocaine, procainamide or related anti-arrhytmic agent (3)
Hypoparathyroidism following thyroidectomy or parathyroidectomy (2)	calcium (2)
Exercise/work capacity or Sports medicine (2)	lactate (2), cathecholamines (2), pH (1), blood gases (1), electrolytes (1)
Infertility (2)	luteinizing hormone /follicle stimulating hormone
Cerebral injury/metabolism (2)	glucose (2), lactate (1)
Sensitivity of tumors to cytotoxic agents (2)	tumor pO_2 (1), tumor drug levels (1)
Therapeutic drug monotoring (2)	methotrexate (2)
Chronic bronchitis/chest medicine (2)	O_2 (2), CO_2 (1), pH

The numbers in parentheses are the number of responders.

Reprinted from Pickup, J. C., Alcock, S., *Biosensors and Bioelectronics* **6** (1991) 639–646; courtesy Elsevier Science Publishers Ltd.

Table 15-7. Specifications for implantable glucose sensor.

Specification	Ideal specification	Minimal requirement
Equilibrium time	seconds	10 minutes
Lifespan	10 years	1 month
Stability	within 1%	within 10%
Reproducibility	within 2% of specified range	within 10% of specified range
Drift	correctable	correctable
Specificity	glucose specific within physiological range	glucose specific within physiological range (with minimum interference from body chemicals)
Size	needle size	200 gm, 10 cm diameter
Safety	no hazards	non-lethal with redundant safety feature
Shelf life	indefinite	one month
Fragility	none	resists trauma

Reprinted from Wilkins, E., Wilkins, M., Schade, D., *J. Environ. Sci. Health, Part A* **26** (1991) 635–654; courtesy Marcel Dekker Inc.

Table 15-8. Sequence of interactions of tissues to foreign bodies.

protein adsorption
platelet adhesion
platelet aggregation
fibrin formation
clot formation

If the sensor incorporates a biological component (a biosensor) the biological component, usually an enzyme, will deteriorate over time at room temperature. This deterioration in enzyme activity is more rapid at human body temperature, thus decreasing the potential useful life of the biosensor [39].

15.6.2 Non-Invasive Sensing

There is a move away from the requirement to take blood samples for diagnostic testing. This trend would eliminate the invasive procedure of venipuncture which is uncomfortable for the patient. Also venipuncture carries a small risk to the patient as it breaks the skin barrier and can allow entry of disease agents. It carries a small risk to the phlebotomists that they will be infected by a disease from the patient, eg AIDS or hepatitis.

Currently available technology used for non-invasive sensing is limted to transcutaneous electrodes and pulse oximetry, however there is interest in extending the use of current technology and using other technologies for assessing the biochemical and physiological status of the human body.

The human body produces magnetic fields as the result of the creation of ionic fields, the incorporation of ferromagnetic components, or by a paramagnetic/diamagnetic response to

an applied field. These magnetic fields can be detected using a SQUID (superconducting quantum interference device) magnetometer. It has been shown that it is possible to detect a build up of iron in the liver during certain diseases of the blood and to detect a build up of dust containing ferromagnetic particles in the lungs of workers in certain hazardous occupations. Other biomagnetic measurements have been used to show electrophysiological disturbances in heart muscle and in the brain [40].

Transthoracic electrical impedance has been investigated as an indicator of cardiac output. However the resistivity of blood has been shown to vary with the hematocrit and with the concentration of some of the plasma components such as sodium chloride and proteins [41]. If this non-invasive approach to monitoring patients in intensive care is to succeed it will be necessary to develop models of the changes which take place in the critically ill patient and use these to correct the readings obtained for transtoracic electrical impedence.

New systems for non-invasive sensing are centred on the use of laser doppler methods to analyze blood flow through the human skin. The most promising field of application is in dermatology where it is being used to study psoriasis and to assess burn wounds. New systems sold by Moor (UK) and Lisca (Sweden) enable sections of the body to be scanned which significantly enhances the diagnostic capability.

15.7 Conclusion

Sensor technology has been used in medicine and healthcare for some 35 years since the initial realisation of the value of pH measurements; however the applications have been limited due to two factors. The use of many sensors involves an invasive step such as taking blood or a probe being inserted into the body. The other limiting factor has been the difficulty of use of the device. There has been the need for frequent maintenance and the device may be fragile.

A major step forward has been the development of the pulse oximeter, a relatively inexpensive device which does not require consumables, is inexpensive to use, robust and non-invasive. This sensor technology is leading the way to a range of devices which can monitor the patient without discomfort to the patient. It is also capable of generating real-time results which are available to the doctor promptly.

There is an increasing need for data refinement so the healthcare professional can quickly, easily interpret the results, and their attention is drawn to clinically significant findings.

15.8 References

[1] Czaban, J. D., *Anal. Chem.* **57** (1985) 345A-356A.
[2] Davis, Z., Pappas, P., Foody, W., *Medicial Laboratory Observer* (special issue) **23** (1991) 12-15.
[3] Thompson, J. M., Emmett, C., Smith, S. C. H., Cramb, R., Hutton, P., *Ann. Clin. Biochem.* **26** (1989) 274-280.
[4] Duroux, Ph., Emde, C., Bauerfeind, P., Francis, C., Grisel, A., Thybaud, L., Armstrong, D., Depeursinge, C., Blum, A. L., *Gut* **32** (1991) 240-245.

[5] Rennie, J. M., *Archives of Disease in Childhood* **65** (1990) 345-346.
[6] Vadgama, P., Alberti, K. G. M. M., *Recent Advances in Clinical Biochemistry* **3** (1985) 255-272.
[7] Rolfe, P., *Current Anaesthesia and Critical Care* **2** (1991) 122-129.
[8] Huch, A., *Pflügers Archiv* **337** (1972) 185-198.
[9] Clark, J. S., Votteri, B., Ariagno, R. L., Cheung, P., Eichhorn, J. H., Fallat, R. J., Lee, S. E., Newth, C. J. L., Rotman, H., Sue, D. L., *Am. Rev. Respir. Dis.* **145** (1992) 220-232.
[10] Okane, M., Shigemitsu, S., Inaba, J., Koresawa, M., Kubo, T., Iwasaki, H. J., *Perinat. Med.* **17** (1989) 399-407.
[11] Rithalia, S. V. A., *Journal of Medical Engineering & Technology* **15** (1991) 143-153.
[12] American Society for Testing and Materials Committee. Cutaneous gas monitoring devices for oxygen and carbon dioxide. Standard specification, F 984-986 (1986). Available from American National Standard Institute, 1430 Broadway, New York, NY 10017, USA.
[13] International Electrotechnology Commission. Particular requirements for the performance of transcutaneous oxygen and carbon dioxide partial pressure monitoring equipment. IEC Technical Committee 62D, Draft Publication 601-603 (1987).
[14] American Academy of Pediatrics; Task force on transcutaneous oxygen monitors; *Pediatrics* **83** 122-126.
[15] Wimberley, P. D., Burnett, R. W., Covington, A. K., Maas, A. H. J., Mueller-Plathe, O., Siggaard-Andersen, O., Weisberg, H. F., Zijlstra, W. G., *Clinica Chemica Acta* **190** (1990) S41-S50.
[16] Sutherland, R. M., Dahne, C., Place, J. F., Ringrose, A. S., *Clinical Chemistry* **30** (1984) 1533-1538.
[17] Bradley, R. A., Drake, R. A. L., Shanks, A. M., Stephenson, P. R., presented at the Royal Society, May 1986.
[18] Howles, C. M., Macnamee, M. C., *British Medical Bulletin* **46** (1990) 616-627.
[19] Moran, R. F., *Journal of the International Federation of Clinical Chemistry* **5** 170-182.
[20] Hamamatsu Photonics K. K., EP 0 549 835 A1, 7 July 1993.
[21] Noble, D., *Clin. Chem. News* **18/6** 1, 10.
[22] Diasense Inc., EP 0 548 418 A1, 30 June 1993.
[23] Green, A., Morgan, I., *Neonatology and Clinical Biochemistry;* London, ACB Venture Publications (1993).
[24] Fischer, U., *Diabetic Medicine* **8** (1991) 309-321.
[25] Shaw, G. W., Claremont, D. J., Pickup, J. C., *Biosensors and Bioelectronics* **6** (1991) 401-406.
[26] von Albertini, B., *Cont. Nephrol.* **61** (1988) 37-45.
[27] Own, V. M., in *Immobilized Biological Compounds for Detection, Medical, Food and Environmental Analysis;* Guilbault, G. G., Mascini, M. (eds.); Dordrecht: Kluwer Academic Press (in press).
[28] Campanella, L., Sammartino, M. P., Tomassetti, M., *Analyst* **115** (1990) 827-830.
[29] Thavarungkul, P., Håkanson, H., Holst, O., Mattiasson, B., *Biosensors & Bioelectronics* **6** (1991) 101-107.
[30] Sansen, W., Jacobs, Claes, A., et al., abstract from *2nd Workshop of Biomedical Engineering Action of the European Community in Chemical Sensors for in vivo Monitoring.* 12-15 November 1989 Firenze, Italy.
[31] Kahn, J., *Clinical Chemistry News* **19** (1993) 1, 3.
[32] Rink, E., Hilton, S. Szczepura, A., Fletcher, J., Sibbald, B., Davies, C., Freeling, P., Stilwell, J., *British Medical Journal* **307** (1993) 775-778.
[33] Shearer, E. S., O'Sullivan, E. P., Hunter, J. M., *Anaesthesia* **46** (1991) 750-755.
[34] Jessop, J., Jones, J. G., *British Journal of Anaesthesia* **66** (1991) 635-637.
[35] Owen, V. M., *Market Requirements and Opportunities for Advanced Sensors in the Healthcare Sector;* Advanced Sensors Technology Transfer Programme, Department of Trade and Industry, UK (1992).
[36] Rolfe, P., *Med. & Biol. Eng. & Compt.* **28** (1990) B34-B47.
[37] Pickup, J. C., Alcock, S., *Biosensors and Bioelectronics* **6** (1991) 639-646.
[38] Wilkins, E., Wilkins, M., Schade, D., *J. Environ. Sci. Health, Part A,* **26** (1991) 635-654.
[39] Schoonen, A. J. M., Schmidt, F. J. Hasper, H., Verbrugge, D. A., Thiessen, R. G., Lerk, C. F., *Biosensors & Bioelectronics* **5** (1990) 37-46.
[40] Swithenby, S. J., *Phys. Technol.* **18** (1987) 17-24.
[41] Fuller, H. D., *Annals of Biomedical Engineering* **19** (1991) 123-129.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

16 Environmental Sensors

KEITH W. JONES, CRL, HAYES, UK

Contents

16.1 Introduction 453

16.2 The Need for Environmental Sensors 454
16.2.1 Classification 454
16.2.2 Major Applications and Measurands 455
16.2.2.1 Drinking Water Quality 455
16.2.2.2 Waste Water Treatment 455
16.2.2.3 Effluent Discharge 457
16.2.2.4 Water Intake 463
16.2.2.5 Ambient Water Pollution 463
16.2.2.6 Gaseous Emissions 464
16.2.2.7 Ambient Air Pollution 467
16.2.2.8 Land Pollution 468
16.2.2.9 Future Applications 469

16.3 The Driving Factors for the Development of Environmental Sensors . . . 470
16.3.1 Introduction 470
16.3.2 Legislation 470
16.3.2.1 Europe 470
16.3.2.2 UK 472
16.3.2.3 USA 473
16.3.2.4 Trends in Legislation 474
16.3.3 Economic Pressure 475
16.3.4 Public Pressure 476

16.4 Technologies for Environmental Sensors 476
16.4.1 Introduction 476
16.4.2 Laboratory Techniques 476
16.4.3 Fixed In-Situ Monitors 478
16.4.4 Portable Sensors 479
16.4.5 Advanced and Emerging Sensors 479
16.4.5.1 Fiber Optic Chemical Sensors 479
16.4.5.2 Biosensors 480
16.4.5.3 Solid-State Sensors 481

16.5 Market Opportunity 483
16.5.1 Introduction 483
16.5.2 The Market for Environmental Sensors 483
16.5.3 Industry Attitudes to the Environment 486
16.5.4 The Environmental Monitoring Supply Industry 487
16.5.5 Market Trends 488

16.6 Bibliography 488

16.1 Introduction

This century has seen ever-increasing pressure on the environment through the explosive growth of industries and populations worldwide. If anything, this trend will increase in pace as more and more countries embrace industrialization. Already the potential for breakdown in the natural self-correction mechanisms in the environment can be seen in heavily over-crowded and polluted cities such as Bangkok and Mexico City and in the legacy of uncontrolled industrialization in Eastern Europe.

Fortunately, industrial companies, national governments and citizens worldwide have begun to see the protection and maintenance of the environment as key to their future. The pressure of public opinion, through government, is coming to bear ever more powerfully on those who can take action to influence the environment and this has resulted, amongst other effects, in a strong incentive to understand the condition of the environment in greater detail through closer monitoring and measurement. In particular, the ability to handle large quantities of data rapidly using computer-based monitoring systems has created a need for measurement technology which can provide data rapidly from close to the test site, ie, sensor technology, as distinct from off-line analysis in a remote, central laboratory.

At the same time, the past twenty-five years has seen explosive growth in the development of advanced sensor technologies for all applications, as detailed elsewhere in this Book Series. These trends combine to create a powerful market opportunity for sensor products for environmental monitoring.

A broad definition of environmental monitoring would include all aspects of air and water quality, ground contamination, radiation, noise, even heat and light pollution. However, the major opportunities for environmental sensors lie in the monitoring of pollution in air, water and ground and we have restricted the scope of this chapter to these three sectors. Similarly, there is no clear distinction between sensors for environmental monitoring and those for health and safety. The latter include, for example, the detection of flammable or toxic gases in the home or workplace, which may be regarded as pollutants in the personal environment. We have elected, somewhat arbitrarily, to exclude sensors intended to provide warning of imminent hazard to person or industrial plant from flammable, explosive or toxic pollutants and those that relate to the comfort of the working or home environment, such as temperature or humidity. In essence, our working definition of environmental sensors includes those which monitor the presence or extent of chemical pollutants which may be detrimental to the air, water or ground environment.

In the next section we summarize the major applications and requirements for environmental sensors in air, water and ground monitoring and, in particular, those for which suitable sensor products do not yet exist. We then describe the most significant factors which drive the market opportunity, legislation, economics and public opinion. Section 16.3 briefly discusses the most significant sensor technologies for the environment, both those in current usage and advanced technologies likely to be the basis of future products. Finally we summarize the potential market size for environmental sensors in various applications and geographical territories and project the emergence and evolution of the industrial sector that will supply the demand.

16.2 The Need for Environmental Sensors

16.2.1 Classification

The major applications may be classified as in Table 16-1. Sensors may be used in fixed installations, typically as part of a long term infrastructural monitoring program, or may be mobile, either for shorter timescale applications or for those concerned with the personal environment.

Table 16-1. Classification of environmental monitoring applications.

	Fixed monitors		Mobile monitors	
	Stationary source	Ambient	Portable	Personal
Water	Drinking water, effluent	Water pollution, intake monitoring	Water pollution, pollution tracing	Drinking water?
Air	Industrial emissions, leaks, car exhausts, biochemicals	Air quality	Air quality, surveys	Gas alarms
Land	Waste disposal		Remediation, leaks	

Fixed installations sub-divide into stationary source monitors and ambient monitors. Stationary source monitors are mostly associated with monitoring potential sources of pollution, such as gas or waste discharges from factories or water treatment plants. Their purpose is to monitor the long-term trends in the contamination load *input* into the environment to enable control of contamination at source through improved discharge treatments or production processes. Stationary source monitors also have a vital role in detecting sudden changes in the contamination load input due to process or treatment failures.

Ambient monitors are concerned with monitoring the contamination load *in* the environment, that is after dispersal from the various sources. Their purpose is to identify contamination problems in order to enable remedial action and, where possible, to indicate the sources of the contamination. These also act as alarms to indicate sudden changes in the condition of the environment due to unmonitored or unregulated sources of pullution.

Mobile monitoring applications subdivide into portable instruments and personal monitors. Portable instruments are mainly used to obtain a "snap shot" of the condition of the environment, for short-term source or ambient monitoring or for assessment and audit of contaminated sites prior to remediation. They are also vital in tracing the source of contamination detected by ambient monitors, following contamination trails back to source where no other instrumentation may be present.

Personal monitors are at present restricted mainly to personal gas alarms, used to detect hazardous or harmful concentrations of gases in the local working or home environment of the individual. As such they are excluded from consideration here. However there is a real

possibility that personal environmental monitors may represent a rapid growth area in the near future.

16.2.2 Major Applications and Measurands

16.2.2.1 Drinking Water Quality

Drinking water, all of which derives from natural water sources, has the widest possible interaction with people. The vital importance of its quality is obvious. All points in the processing of water from nature to domestic use can cause contamination of the final product. In this section we consider those aspects which are, or could be monitored in the final product itself, that is, the monitoring of the contamination of the "final product". Monitoring of earlier steps in the processing chain are dealt with in later sections.

The principal monitored contaminants are residuals from water treatment processes such as chlorine and aluminium, metals such as lead and copper which derive from the distribution system itself and bacteria in insufficiently disinfected water. Drinking water is also routinely tested for a range of substances prescribed as potentially hazardous through regulation or legislation in various countries (see Section 16.3). However, these are principally monitored through sampling for laboratory analytical tests and there is at present no pressure to develop on-site sensor-based tests.

A full list of prescribed parameters, concentrations or values, taken from the EC Drinking Water Directive, is given in Table 16.2.

Chlorine is added to water as it enters the distribution chain to remove active bacterial contamination. If too much is added, residual chlorine above the prescribed limits remains in the water coming from the tap and a characteristic taste is apparent. Moreover, the water supply companies would wish to minimise the chlorine added for reasons of cost control. A similar issue is becoming apparent with the increasing use of ozone treatment for disinfection, although the legislation lags the technology at present.

However, insufficient chlorine leads to residual bacterial activity, particularly from coliforms and streptococci which persist from treated waste water, both of which are potential health hazards. Less common, but also hazardous bacteria are cryptosporidium and, particularly in the USA, giardia.

Aluminum and iron compounds are added as flocculants in the treatment of waste water and can, in certain circumstances, persist in drinking water through the distribution chain. The health hazards of aluminum are well known and it is routinely and continuously monitored at all points in the treatment chain.

With the increasing age of the distribution network for drinking water in most developed countries, metals such as lead and copper can enter the water directly from the pipework. Although massive investment is being made to modernise the network and eliminate the sources of this contamination, particularly in the USA and UK, this will not be completed for many years and monitoring will still be necessary. In the USA, for example, legislation now demands that all domestic water supplies be regularly tested for lead levels.

The major needs for sensor-based monitors will be for stationary ambient and portable instruments for the detection of disinfection and treatment residuals, rapid portable tests for

Table 16-2. Prescribed substances unter the EC drinking water directive.

Parameter	Units	Maximum concentration or value
Colour	mg/l Pt/Co scale	20
Turbidity	Formaxin turbidity units	4
Odour (inc. hydrogen sulphide)	Dilution number	3 at 25 °C
Taste	Dilution number	3 at 25 °C
Temperature	°C	25
Hydrogen ion	pH value	9.5 5.5 (minimum)
Sulphate	mg SO_4/l	250
Magnesium	mg Mg/l	50
Sodium	mg Na/l	150
Potassium	mg K/l	12
Dry residues	mg/l	1500
Nitrate	mg NO_3/l	50
Nitrite	mg NO_2/l	0.1
Ammonium (ammonium and ammonia ions	mg NH_4/l	0.5
Kjeldahl nitrogen	mg N/l	1
Oxidizability (permanganate value)	mg O_2/l	5
Total organic carbon	mg C/l	No significant increase over that normally observed
Dissolved or emulsified hydrocarbons	µg/l	10
Phenols	µg C_6H_5OH/l	0.5
Surfactants	µg/l as lauryl sulphate	200
Aluminium	µg Al/l	200
Iron	µg Fe/l	200
Manganese	µg Mn/l	50
Copper	µg Cu/l	3000
Zinc	µg Zn/l	5000
Phosphorus	µg P/l	2200
Fluoride	µg F/l	1500
Silver	µg Ag/l	10

bacteria, possibly stationary ambient monitors for bacteria and portable instruments for heavy metals.

There are also opportunities for monitors of physical parameters such as color and turbidity, but we exclude these as physical instruments rather than opportunities for sensor-based products.

16.2.2.2 *Waste Water Treatment*

Water used in processes, either industrial or domestic, is eventually returned to natural water courses, rivers, lakes and so on. It is common for industrial effluent to return directly after internal processing has brought it within prescribed limits. It is then monitored at its point of entry to the natural environment. (This is dealt with in more detail in Section 16.2.2.3).

Urban waste water, on the other hand, is treated by the water supply companies to reduce the contamination load before release into the natural environment.

Monitoring of the quality is typically more integrated with the treatment process. The regulatory limits for contamination load are the same in both cases.

The key measurands relate to the composite loading of organic compounds in the water (the sources of which are familiar to us all and need not be detailed) and provide general indicators of pollution level, rather than specific chemical contaminant levels. These are biological oxygen demand (BOD), chemical oxygen demand (COD), total organic carbon (TOC) and dissolved oxygen (DO). Water returned to natural courses with low oxygen levels or with contaminants which consume oxygen significantly will have a damaging effect on the plant and animal life in the water.

BOD is a measure of the biodegradable substances in a sample, COD the oxygen required to oxidise the organic *and* inorganic substances present. TOC indicates the total carbon compounds present but does not differentiate between degradable and nondegradable compounds. DO in a sample is lowered if BOD or COD are high. Typically, BOD is used to indicate final pollutant load in effluent. Either COD, TOC or DO act as control parameters for the blending and treatment processes. The parameters are related and not usually all measured on the same site.

Specific chemical contaminants are also monitored during treatment and prior to effluent release, including inorganic ions (NH_4^+, NO_3^-, NO_2^-, CN^-, F^-, Cl^-, SO_3^-, SO_4^{2-}), and metal ions (Na^+, Mg^{2+}, Ca^{2+}, Fe^{2+}, Cu^{2+}, Zn^{2+}, Pb^{2+}, Cd^{2+}, Cr^{2+}). As noted in section 16.2.2.1, aluminum may persist in treated water from flocculants added during processing and this is commonly monitored continuously prior to discharge. pH, redox potential and conductivity are also widely used, both as process control parameters and as general indicators of water quality on discharge to the environment.

Bacterial activity, particularly faecal coliforms and streptococci, are also routinely tested for, although there are as yet no practical sensor-based monitors for these.

Future needs will be mainly for stationary ambient monitors. There is a need for more rapid, sensor-based monitors for oxygen demand, ideally for BOD which is currently only measured in the laboratory. Trends in legislation are promoting TOC as a compromise test for process control and pollutant loading and this may represent a good opportunity for new instruments. As with drinking water, rapid bacterial activity monitors would find wide acceptance in the waste water treatment industry. There has been a consistent call from the water treatment industry (amongst many others) for improved pH sensors.

16.2.2.3 Effluent Discharge

The discharge of certain substances in industrial effluent into natural water courses is regulated in most developed countries (see Section 16.3). This places an onus on the company making the discharge to monitor its own effluent. The substances monitored in each case will depend on the particular industrial process, but most regulators work from a list of substances deemed to be harmful to the water environment. As an example, prescribed substances for release to water in the UK form the "Red List" (Table 16-3). In addition, other substances must be notified to the regulator (the National Rivers Authority) as being present in effluent discharges (Table 16-4) and may also be subject to limits and therefore monitoring. In the

Table 16-3. The Red List.

Mercury and its compounds
Calcium and its compounds
Gamma-Hexachlorocyclohexane
DDT
Pentachlorophenol
Hexachlorobenzene
Hexachlorobutadiene
Aldrin
Dieldrin
Endrin
Polychlorinated biphenyls
Dichlorvos
1,2-Dichloroethane
Trichlorobenzene
Atrazine
Simazine
Tributyltin compounds
Triphenyltin compounds
Trifluralin
Fenitrothion
Azinphos-methyl
Malathion
Endosulfan

Table 16-4. Notifiable substances if in effluent.

Mineral oils
Alkalis
Metals and their compounds
Iron, aluminium, antimony, arsenic, beryllium, chromium, copper, lead, nickel, selenium, silver, tin, vanadium, zinc, cadmium, mercury
Cyanides or compounds containing cyanide
Salts including: nitriles, chlorates, fluorides, sulphates, hypochlorites, nitrates, nitrites, perchlorates, sulphides, carbides
Phenols, cresols and simple derivatives
Tar and tar oils
Mineral oils
Oil emulsions
Grease
Ammonia or ammoniacal compounds
Paint wastes (as sludges)
Pharmaceuticals including steroids and hormones
Surface acting agents
Organohalogen compounds, including pesticides and degreasing agents
Organosulphur compounds containing nitrogen
Organophosphorus or organosilicon compounds
Acrylonitrile

Table 16-4. (continued)

Formaldehyde
Carbohydrates
Yeast
Petroleum or any other substances capable of producting flammable or toxic vapours including calcium carbide and carbon disulphide
Solvents including: alcohols and nitration products, ketones, esters, ethers; hydrocarbons including benzene, toluene and xylene and their industrial equivalents
Cooling water
Boiler blowdown
Scrubbing water
Any other substance known to be toxic or hazardous

USA, prescribed substances form an even more extensive list (the Priority Pollutant and Target Compound Lists, Table 16-5).

The extensive and specific nature of these prescribed lists makes off-line laboratory analysis the only viable option for monitoring at present. However, the nature of many industrial processes and the increasing severity of the penalties for exceeding consent limits is forcing a trend towards more on-line monitoring (continuous or periodic). These would provide a far more rapid warning of increasing contaminant levels due to process failures, minimising the discharge of out-of-limits effluent. This trend is creating a demand for new sensor-based stationary source instruments for specific contaminants.

As with waste water treatment, less specific measurands are also used increasingly, particularly BOD and COD. There is also a trend towards the use of total toxicity as a measurand, as new sensor techniques based on biotechnology emerge, and this may well become more significant as a basis for regulatory discharge consents in future.

In the long-term, regulatory authorities are encouraging industries to control and abate the production of prescribed substances in effluent by improvement of the process, at source. Whilst there will still be a need for continuous on-line monitoring to provide early warning against process failures, the balance of the opportunity may shift towards portable instruments used for spot checks, either by the company or the regulators.

Table 16-5. Priority Pollutant (PPL) and Target Compound (TCL) Lists.

Parameter	Alternate nomenclature	PPL	TCL
Metals			
Aluminium			×
Antimony		×	×
Arsenic		×	×
Barium			×
Beryllium		×	×
Cadmium		×	×
Calcium			×
Chromium		×	×
Cobalt			×

Table 16-5. continued.

Parameter	Alternate nomenclature	PPL	TCL
Copper		×	×
Iron			×
Lead		×	×
Magnesium			×
Manganese			×
Mercury		×	×
Nickel		×	×
Potassium			×
Selenium		×	×
Silver		×	×
Sodium			×
Thallium		×	×
Vanadium			×
Zinc		×	×
Cyanide		×	×
Other			
Asbestos		×	
2,3,7,8-TCDO	2,3,7,8-Tetrachlorodibenzo-p-dioxin	×	×
Volatile organics			
Chloromethane	Methyl chloride	×	×
Bromomethane	Methyl bromide	×	×
Vinyl chloride	Chloroethylene	×	×
Chloroethane		×	×
Methylene chloride	Dichloromethane	×	×
Acetone	2-propanone		×
Carbon disulphide			×
1,1-dichloroethene		×	×
1,1-dichloroethane		×	×
CIS 1,2-dichloroethane			×
TRANS 1,2-dichloroethane		×	×
Chloroform		×	×
1,2-dichloroethane		×	×
2-butanone	Methyl ethyl etone		×
1,1,1-trichloroethane		×	×
Carbon tetrachloride		×	×
Vinyl acetate			×
Bromodichloromethane	Dichlorobromomethane	×	×
1,1,2,2-tetrachloroethane		×	×
1,2-dichloropropane		×	×
CIS 1,3-dichloropropene		×	×
TRANS 1,3-dichloropropene		×	×
Trichloroethene		×	×
Benzene		×	×
Bromoform	Tribromomethane	×	×
2-hexanone	Methyl butyl ketone		×

Table 16-5. continued.

Parameter	Alternate nomenclature	PPL	TCL
4-methyl-2-pentanone	Methyl isobutyl ketone		×
Dibromochloromethane	Chlorodibromomethane	×	×
1,1,2-trichloroethane		×	×
Toluene	Methyl benzene	×	×
Chlorobenzene		×	×
Eithyl benzene		×	×
Styrene			×
Xylenes (o, m and p)			×
2-chloroethyl vinyl ether		×	
Acrolein	2-propenal	×	
Acrylonitrile	Propenenitrile	×	
Semi-volatiles			
Phenol		×	×
Bis (2-chloroethyl) ether		×	×
2-chlorophenol		×	×
1,3-dichlorobenzene		×	×
1,4-dichlorobenzene		×	×
Benzyl alcohol			×
1,2-dichlorobenzene		×	×
2-methyl phenol			×
Bis (2-chloro isopropyl) ether		×	×
4-methylphenol			×
N-nitroso-dipropylamine		×	×
Hexachloroethane		×	×
Nitrobenzene		×	×
Isophorene		×	×
2-nitrophenol		×	×
2,4-dimethylphenol		×	×
Benzoic acid			×
Bis (2-chloroethoxy) methane		×	×
2,4-dichlorophenol		×	×
1,2,4-trichlorobenzene		×	×
Naphthalene		×	×
4-chloroaniline			×
Hexachlorobutadiene		×	×
4-chloro-3-methylphenol	para-chloro-meta-cresol	×	×
2-methylnaphthalene			×
Hexachlorocyclopentadiene		×	×
2,4,6-trichlorophenol		×	×
2,4,5-trichlorophenol			×
2-chloronaphthalene		×	×
2-nitroaniline			×
Dimethyl phthalate		×	×
Acenaphthylene		×	×
2,6-dinitrotoluene		×	×
3-nitroaniline			×

Table 16-5. (continued)

Parameter	Alternate nomenclature	PPL	TCL
Acenaphthene	Naphthyleneethylene	×	×
2,4-dinitrophenol		×	×
4-nitrophenol		×	×
Dibenzofuran			×
2,4-dinitrotoluene		×	×
Diethyl phthalate		×	×
4-chlorophenyl phenylether		×	×
Fluorene		×	×
4-nitroaniline			×
4,6-dinitro-2-methylphenol	4,6-dinitro-o-cresol	×	×
N-nitroso-diphenylamine		×	×
4-bromophenyl phenyl ether		×	×
Hexachlorobenzene		×	×
Pentachlorophenol		×	×
Phenanthrene		×	×
Anthracene		×	×
Di-N-butylphthalate		×	×
Fluoranthene		×	×
Pyrene		×	×
Butyl benzyl phthalate		×	×
3,-dichlorobenzidine		×	×
Benzo (a) anthracene	1,2-benzanthracene	×	×
Chrysene		×	×
Bis (2-ethylhexyl) phthalate		×	×
Di-N-octyl phthalate		×	×
Benzo (b) fluoranthene	3,4-benzofluoranthene	×	×
Benzo (k) fluoranthene	11,12-benzofluoranthene	×	×
Benzo (a) pyrene	3,4-benzopyrene	×	×
Indeno (1,2,3-cd) pyrene	2,3-o-phenylenepyrene	×	×
Dibenz (a,h) anthracene	1,2,5,6-dibenzanthracene	×	×
Benzo (g,h,i) perylene	1,12-benzoperylene	×	×
Benzidine		×	
1,2-diphenylhydrazine		×	
N-nitrosodimethylamine		×	
Pesticide/PCB			
Alpha-BHC		×	×
Beta-BHC		×	×
Delta-BHC		×	×
Gamma-BHC	Lindane	×	×
Heptachlor		×	×
Aldrin		×	×
Heptachlor epoxide		×	×
Endosulfan I		×	×
Dieldrin		×	×
4,4′-DOE		×	×

Table 16-5. (continued)

Parameter	Alternate nomenclature	PPL	TCL
Endrin		×	×
Endosulfan II		×	×
4,4′-DDD		×	×
Endosulfan sulfate		×	×
4,4′-DDT		×	×
Endrin ketone			×
Methoxychlor			×
Alpha-chlordane		×	×
Gamma-chlordane		×	×
Toxaphene			×
Aroclor 1016		×	×
Aroclor 1221		×	×
Aroclor 1232		×	×
Aroclor 1242		×	×
Aroclor 1248		×	×
Aroclor 1254		×	×
Aroclor 1260		×	×
Endrin aldehyde		×	

16.2.2.4 Water Intake

Water treatment, both of waste water and of drinking water, reduces the pollutant load in the water taken into the process by a broadly constant proportion. In order to maintain the output within consent limits, water taken into the process may also be monitored for specific pollutant levels significantly above the normal range. Similarly, water abstracted from natural sources for industrial processes commonly needs to be held within quality limits and intake monitoring may also be used here to warn of upstream pollution.

Measurands which have been monitored in intake protection systems include pH, DO, ammonia, formaldehyde, phenols, cresols, pesticides, hydrocarbons, heavy metals, cyanide and total toxicity. In practice, the monitoring suite would reflect the known potential pollution sources upstream of the intake and the particular requirements of the treatment or industrial process.

Intake monitoring in water treatment plants is increasingly being used in France and Germany, although research in the UK has been stalled by the privatisation of the water supply industry. The trend, however, is likely to be greater use of intake monitoring through fixed ambient sensors for specific contaminants, particularly pesticides, metals and hydrocarbons, and for general measures such as toxicity.

16.2.2.5 Ambient Water Pollution

Ambient water pollution is routinely monitored in two forms, surface water such as rivers, lakes and coastal waters, and ground water such as underground aquifers and borehole sources. In both cases the purpose of monitoring is to detect excessive pollution likely to harm

lifeforms in the water and to protect sources used for abstraction, for drinking water or for industrial use.

For surface water, the range of contaminants monitored is broadly the same as that covered by effluent discharge consents, ie oxygen demand, the appropriate prescribed substances list and toxicity. This provides a means of detecting pollution through excessive discharges and provides the starting point for tracing back to the source of the pollution. Other sources of pollution of surface waters arise from run-off from agricultural land, carrying nitrates and phosphates from fertilizers and pesticides into the water, and metals dissolved from soil and bedrock, particularly lead. Nitrates and, particularly, phosphates can cause excessive algal growth, leading to eutrophication. Sampling for laboratory analysis is most commonly used but, increasingly, portable instruments that measure a range of quality parameters are being introduced.

Groundwater monitoring is mainly for the purpose of detecting leakage from storage tanks and industrial premises and leachate from landfill sites, into the ground and, thence, into the water system. Particular measurands are therefore linked to these potential sources, hydrocarbons from fuel storage tanks, organochlorides from landfill, for example. Groundwater monitoring may be by portable monitors or, more usually, by fixed ambient monitors in suitably placed test boreholes.

A particular need not satisfied by available instruments is the monitoring of pharmaceutical intermediates, which may enter the system in effluent but which are not broken down by conventional water treatment methods. Similarly, excessive treatment of highly polluted water can give rise to reaction products, particularly with chlorine-based treatment. These reaction products should also be monitored.

The future trend is likely to be towards greater use of continuous monitors in boreholes for groundwater protection and of portable instruments for specific measurands in surface water for tracing and identification of pollution sources. In the longer term, however, regulators are likely to place the onus for effluent control and monitoring increasingly on the generator of the effluent and to carry out less field testing themselves. There will be a need for the foreseeable future, however, for borehole sensors and for noneffluent pollution sources.

16.2.2.6 Gaseous Emissions

Emissions of pollutants to the atmosphere occur as a result of a range of processes. The most significant are industrial processes, particularly combustion and solvent-based processes and fugitive emissions (leaks) from industrial plants, vehicle emissions and emissions from landfill sites. These are also the processes where fixed source monitoring is either a current or future issue. Other sources of gaseous pollutants, such as farm animals and domestic processes, whilst potentially significant in terms of pollutant loading, are not amenable to source monitoring. Also included in this section are particulate pollutants since they are generally dealt with by the same regulations as gaseous emissions.

As with effluent discharge to water, gaseous emissions from prescribed processes are regulated in most of the developed world by the granting of consents based on emission limits and agreed programmes of monitoring and pollution abatement. The measurands monitored depend on the specific process. The main prescribed pollutants under EC and USA regulations, for example, are shown in Tables 16-6 and 16-7, respectively.

Table 16-6. Releases to air: prescribed substances.

Oxides of sulphur and other sulphur compounds
Oxides of nitrogen and other nitrogen compounds
Oxides of carbon
Organic compounds and partial oxidation compounds
Metals, metalloids and their compounds
Asbestos (suspended particulate matter and fibres), glass fibers and mineral fibers
Halogens and their compounds
Phosphorus and its compounds
Particulate matter

Table 16-7. Hazardous air pollutants subject to the provisions of Section 112 of the 1990 Clean Air Amendments.

Acetaldehyde	Chloromethyl methyl ether	Ethyl carbamate
Acetamide	Chloroprene	Ethyl chloride
Acetonitrile	Chromium compounds	Ethylene dibrimide
Acetophenone	Cobalt compounds	Ethylene dichloride
Acetylaminofluorene (2)	Coke oven emissions	Ethylene glycol
Acrolein	Cresol (m)	Ethylene imine
Acrylamide	Cresol (o)	Ethylene oxide
Acrylic acid	Cresol (p)	Ethylene thiourea
Acrylonitrile	Cresols/Cresylic acid	Ethylidene dichloride
Allyl chloride	Cumene	Formaldehyde
Aminobiphenyl (4)	Cyanide compounds	Glycol ethers
Anilinc	D(2,4), salts and esters	Heptachlor
Anisidine (o)	DDE	Hexachlorobenzene
Antimony compounds	Diazomethane	Hexachlorobutadiene
Arsenic compounds	Dibenzofurana	Hexachlorocyclopentadiene
Asbestos	Dibromo-3-chloropropane	Hexachloroethane
Benzene	Dibutylphthalate	Hexamethylene-l
Benzidine	Dichlorobenzene (1,4)	o-diisocyanate
Benzotrichloride	Dichlorobenzidene (3,3)	Hexamethylphosphoramide
Benzyl chloride	Dichloroethyl ether	Hexane
Beryllium compounds	Dichloropropene (1,3)	Hydrazine
Biphenyl	Dichlorvos	Hydrochloric acid
Bis(2-ethylhexyl)-	Diethanolamine	Hydrogen fluoride
phthalate (DEHP)	Diethyl aniline (N,N)	Hydroquinone
Bis(chloromethyl) ether	Diethyl sulfate	Isophorone
Bromoform	Dimethoxybenzidine (3,3)	Lead compounds
Buradlene (1,3)	Dimethyl	Lindane
Cadmium compounds	aminoazobenzene	Maleic anhydride
Calcium cyanamide	Dimethyl benzidine (3,3′)	Manganese compounds
Caprolactam	Dimethyl carbamyl	Mercury compounds
Captan	chloride	Methanol
Carbaryl	Dimethyl formamide	Methoxychlor
Carbon disulfide	Dimethyl hydrazine (1,1)	Methyl bromide
Carbon tetrachloride	Dimethyl phthalate	Methyl chloride

Table 16-7. (continued)

Carbonyl sulfide	Dimethyl sulfate	Methyl chloroform
Catechol	Dinitro-o-cresol (4,6)	Methyl ethyl ketone
Chloramben	Dinitrophenol (2,4)	Methyl hyrazine
Chlordane	Dinitrotoluene (2,4)	Methyl iodide
Chlorine	Dioxane (1,4)	Methyl isobutyl ketone
Chloroacetic acid	Diphenylhydrazine (1,2)	Methyl isocyanate
Chloroacetphenone (2)	Epichlorohydrin	Methyl methacrylate
Chlorobenzene	Epoxybutane (1,2)	Methyl tert butyl ether
Chlorobenzilate	Ethyl acrylate	Methylene bis (2-chloroaniline)
Chloroform	Ethyl benzene	Toluene
Methylene chloride	Phosphorus	Toluene diamine (2,4)
Methylene diphenyl diisocyanate	Phthalic anhydride	Toluene diisocyanate (2,4)
Methylenedianiline (4,4)	Polychlorinated biphenyls	Toluidine
Mineral fibres	Polycyclic organic matter	Toxaphene
Naphthalene	Propane sultone (1,3)	Trichlorobenzene (1,2,4)
Nickel compounds	Propiolactone (beta)	Trichloroethane (1,1,2)
Nitrobenzene	Propionaldehyde	Trichloroethylene
Nitrobiphenyl (4)	Propoxur	Trichlorophenol (2,4,5)
Nitrophenol (4)	Propylene dichloride	Trichlorophenol (2,4,6)
Nitropropane (2)	Propylene oxide	Triethylamine
Nitroso-N-methylurea (N)	Propyleneimine (1,3)	Trifluralin
Nitrosodimethylamine	Quinoline	Trimethylpentane (2,2,4)
Nitrosomorpholine (N)	Quinone	Vinyl acetate
Parathion	Radionuclides	Vinyl bromide
Pentachloronitrobenzene	Selenium compounds	Vinyl chloride
Pentachlorophenol	Styrene	Vinylidene chloride
Phenol	Styrene oxide	Xylene (m)
Phenylenediamine (p)	Tetrachlorodibenzo-p-dioxin	Xylene (o)
Phosgene	Tetrachloroethane	Xylene (p)
Phosphine	Tetrachloroethylene	Xylenes (mixed)
	Titanium tetrachloride	

Most combustion processes monitor SO_2, NO_x, CO, CO_2 and hydrocarbons in the flue gases, and possibly also O_2 and particulates, with fixed in-stack sensors. SO_2 and some oxides of nitrogen are the major causes of acid rain, whilst CO_2 is the major "greenhouse gas". Other NO_x compounds are photochemical ozone precursors, causing smogs. New and larger processes are commonly required to carry out continuous monitoring, whereas periodic checks by the operator or regulator under defined process conditions are specified for older or smaller processes.

Processes which involve the incineration of organic compounds, such as rubber and plastics, or the use of organic solvents, would require monitoring for the release of more complex specific contaminants such as chlorinated hydrocarbons and fluorocarbons (CHCs and CFCs) both of which deplete the stratospheric ozone layer, volatile organic compounds (VOCs) which are photochemical ozone precursors, polychlorobiphenols (PCBs) and polyaromatic hydrocarbons (PAHs) both of which are highly toxic. Particulates, dust and so on, are generated by

most industrial processes and, if the process itself is prescribed, must be monitored. In addition to their nuisance value, particulates are increasingly implicated in respiratory diseases such as asthma, allergies and, in the case of asbestos, cancer.

As with effluent discharging processes, the emphasis of future regulation is likely to be on pollution abatement, ie reduction at source, and contaminant extraction and recovery, particularly for solvents and VOCs. This is likely, in the longer term, to slow the expansion in the market for continuous fixed monitors, whilst that for portable instruments, particularly for specific toxic gases, is likely to grow. Fixed monitors for flue gases are always likely to represent a strong market opportunity, as much for process control as for emission monitoring.

Decomposition of waste in landfill sites causes the emission of the greenhouse gases methane and CO_2. As well as direct dispersal into the atmosphere, these gases can migrate and accumulate in nearby buildings. Monitors, particularly for methane, are sited either on the surface, in the affected buildings or within boreholes driven into the landfill. Although landfill will become less common as a means of solid waste disposal in future, existing landfill sites will provide a strong market opportunity for robust reliable sensors for use in boreholes and for integrative free-space methane monitors for surface use.

Emissions from vehicles have been greatly reduced in the USA and Europe by the introduction of catalytic converters and oxygen (lambda) sensors for combustion control. However regulations are becoming even more restrictive, particularly in the USA, and more extensive monitoring of exhaust gases is likely to be necessary. Not only will emissions of CO, O_2, NO_x and SO_2 be monitored as part of an active emission abatement system but the integrity of that system will also need to be monitored. Thus there will be massive opportunities for very low cost sensors of these contaminants compatible with automotive use. The technology developments which would need to be accomplished in order to serve this market would also have a significant effect on the industrial gaseous emission monitoring market. There is also likely to be a need for remote monitoring of vehicle exhausts for police and regulatory users, although whether sensor-based instruments can satisfy this is arguable.

16.2.2.7 Ambient Air Pollution

Monitoring of the ambient atmosphere is most commonly carried out by regulatory authorities or by local or national government organisations. The means is a network of sensors from which an overall picture of the types and levels of airborne pollution in a defined area from day to day. This is used partly as a means of tracing the sources of excessive emissions from industrial and other sources but, probably more significantly, to provide an air quality report for the residents of the area. In the UK, for example, O_3, NO_2 and SO_2 are monitored by the Department of the Environment to provide daily ratings of air quality as "very poor", "poor", "good" or "very good" which are broadcast with local weather information. Other measurands monitored include carbon monoxide, particulates and airborne lead (these six measurands also correspond to the criteria polluants monitored in the USA). A national Enhanced Urban Monitoring initiative is also being established in the UK to monitor hydrocarbons such as benzene and 1,3-butadiene at 12 urban sites. The USA also monitors hazardous air pollutants, including vinyl chlorine, asbestos, beryllium, mercury and arsenic, and intends to control, through monitoring, substances on a list prescribed under the

Clean Air Act (see Table 16-8 and Section 16.3). Many of these substances, although subject to regulation, cannot yet be monitored satisfactorily. Future requirements will be for reliable long-term monitoring instruments for all these measurands which require little maintenance or other attention.

Table 16-8. Releases to land: prescribed substances.

Organic solvents
Azides
Halogens and their covalent compounds
Metal carbonyls
Organo-metallic compounds
Oxidising agents
Polychlorinated dibenzofuran and any congener thereof
Polychlorinated dibenzo-p-dioxin and any congener thereof
Polyhalogenatedbiphenyls, terphenyls and napthalenes
Phosphorus
Pesticides, that is to say, any chemical substance or preparation prepared or used for destroying any pest, including those used for protecting plants or wood or other plant products from harmful organisms; regulating the growth of plants; giving protection against harmful creatures; rendering such creatures harmless; controlling organisms with harmful or unwanted effects on water systems, buildings or other structures, or on manufactured products; or protecting animals against ectoparasites
Alkali metals and their oxides and alkaline earth metals and their oxides

16.2.2.8 Land Pollution

Pollutants are monitored in soil for two main purposes, as an early warning of the presence of contaminants which may migrate into water courses and for the assessment of contaminated land for remediation. In both cases the major contaminants are those prescribed for water and air discharge, in particular the EC prescribed lists I and II and the US Priority Pollutant and Target Compound Lists (see Sections 16.2.2.3 and 16.2.2.6). The UK regulations also prescribe a separate list of substances for discharge to land (see Table 16-9). However the particular characteristics of the medium in both cases present a challenge to sensor-based monitoring.

Land contamination can occur through the use of agricultural chemicals, particularly pesticides which persist for many years in the soil, through the deposition of solid or sludge waste from treatment plants or as a result of industrial processes. Pesticides are a particular problem for specific monitoring because of the very wide range of substances used (a 1988 survey in the USA found 46 different pesticides in the groundwater of 26 states). Periodic monitoring by land owners and regulators requires low-cost easy to use instruments which can detect both broad spectrum and specific pesticide contamination. Assessments for land remediation also need to cover a wide range of contaminants. Table 16-10 lists the most common contaminants found at USA sites for remediation through the Superfund mechanism (see Section 16.3). Whilst laboratory analysis and wet chemistry field test kits provide the necessary monitoring function at present, there is a need for portable field test instruments

Table 16-9. 25 most frequently identified substances at 546 superfund sites.

Substance	Percent of sites
Trichloroethylene	33
Lead and compounds	30
Toluene	28
Benzene	26
Polychlorinated biphenyls (PCBs)	22
Chloroform	20
Tetrachloroethylene	16
Phenol	15
Arsenic and compounds	15
Cadmium and compounds	15
Chromium and compounds	15
1,1,1-Trichloroethane	14
Zinc and compounds	14
Ethylbenzene	13
Xylene	13
Methylene choride	12
Trans-1,2-dichloroethylene	11
Mercury	10
Copper and compounds	9
Cyanides (soluble salts)	8
Vinyl chloride	8
1,2-Dichloroethane	8
Chlorobenzene	8
1,1-Dichloroethane	8
Carbon tetrachloride	8

Table 16-10. Markets for environmental monitoring instruments ($M).

	1991			1995			2000		
	Water	Air	Soil	Water	Air	Soil	Water	Air	Soil
EC	174	217	87	255	318	127	410	513	205
USA	214	268	107	314	392	157	506	632	253
RoW	182	228	91	267	333	133	429	537	215
Total	570	713	285	835	1044	418	1345	1681	673

for the most common contaminants, possibly using disporsable elements, for speed and consistent accuracy which present methods cannot offer.

16.2.2.9 Future Applications

There is a number of potential applications for sensors of environmental parameters that may emerge in the near future, in some cases very near. Two illustrative examples follow.

Awareness of the hazards which may be present in the environment is now high at the level of individual citizens, at least in the developed world. At present this is channelled into public,

political and consumer pressure on national governments to take appropriate action (see Section 16.3). However there is a trend for individuals to take responsibility for their own environment in terms of health and safety, through the use of personal alarms for toxic and flammable gases and domestic alarms for smoke and, more recently, CO. It is likely that sensor-based personal monitors of air quality or water quality (drinking or bathing water for example) with similar product characteristics would find a ready and substantial market. The monitors would need to be small enough to carry unobtrusively, very low cost and reliable over a period of one year or more. Such products are in development, particularly for gas monitoring, but no serious product is as yet commercially available.

The growth of the biotechnology industry worldwide has created a new class of hazardous substance, the emission of which into the environment needs to be regulated. Enzymes, antibiotics, hormones, microorganisms and bacterial toxins are all potential environmental hazards, whilst the potential dangers of releases of genetically engineered material are clear. As yet no suitable monitoring instruments for any of these are available. There will be opportunities in the near future for fixed source and ambient monitors for airborne releases and for portable and personal monitors.

16.3 The Driving Factors for the Development of Environmental Sensors

16.3.1 Introduction

Three major factors provide a pressure on industrial and government bodies to invest in environmental monitoring (and abatement) technology. These all arise from a growing recognition by the general public, over a number of years, of the need for significant and urgent action to reverse the effects of industrialization and urbanization. Whilst different countries vary in the level of this public recognition and the degree of focus it provides for pressure to change, the phenomenon is virtually world-wide.

The most immediate driver for industry, and increasingly on the public too, comes through *legislation* by national and local government organizations. Of growing significance is an *economic* pressure, partly through penalties for non-compliance with legislation but more directly through the attitudes of investors, shareholders, insurers and customers. Finally the *public* has a direct, focused voice through a range of lobbying and campaigning organizations which have largely wrought the change in governmental attitudes to the environment in recent years.

16.3.2 Legislation

16.3.2.1 Europe

The principal legislative driving force for the nations of the European Union, at least on the environment, is the series of directives issued by the European Commission (EC). These

are intended to provide a basis for specific legislation enacted by each member state. The overall philosophy is to bring all member states up to the environmental standards of the best.

The key EC directives (with later amendments and updates) affecting pollution control in air, water and ground are as follows:

Urban Wastewater Directive (91/271): Reduction of pollution in freshwater, estuarial and coastal waters due to domestic sewage, industrial effluent and agricultural run-off.

Drinking Water Directive (80/778): Quality standards for water for drinking and food processing.

Groundwater Directive (80/68): Regulates direct and indirect discharges of waste into groundwaters.

Dangerous Substances in Water Framework Directive (76/464): Establishes a system for prescribing substances as needing control or limits. Substances for which an emission reduction programme must be developed are entered on the "grey list", more toxic substances requiring strict limitation go on the "black list".

Bathing Water Directive (76/160): Maximum and guide levels for pollutants in bathing waters.

Surface Water Directive (75/440): Quality of surface waters intended for abstraction for drinking water.

Regulation on Substances that Deplete the Ozone Layer (594/91): Elimination of CFCs and CCl_4 by 1997, halons by 2000.

Incinerator Directives (89/369 and 89/429): Emission limits for new municipal waste incineration plants.

Large Combustion Plant Directive (88/609): Limits for SO_2 and NO_x for plants with thermal output greater than 50 MW.

Framework Directive on Air Pollution (84/360): Acid rain reduction, imposes BATNEEC principle on defined processes. (BATNEEC – Best Available Technology Not Entailing Excessive Cost).

Air Quality Directive (80/779): Limits smoke and SO_2 levels anywhere in the Union.

Diesel Emissions Directive (72/306): Limits on emissions from diesel vehicles.

Vehicles Emission Directive (70/220): Limits on CO and unburnt hydrocarbons.

Hazardous Waste Directive (91/689): Licensing of waste operators and compliance monitoring (repealed 78/319).

Sewage Sludge in Agriculture Directive (86/278): Correct use of sewage sludge on land.

Framework Directive on Waste (75/442): Ensures waste treatment does not harm human health.

These directives are implemented as a variety of legislative instruments in member countries of the EU, and are reflected in the legislation of many nonmember countries. Together with its equivalent in the USA, the framework provided by the directives is regarded as a model for environmental legislation in the rest of the world.

16.3.2.2 UK

In order to illustrate the effect of the directives on industry in particular countries we use the example of their implementation in the UK. Although neither the most nor least advanced in Europe in terms of implementation, UK legislation provides some indication of the trends for other countries in future.

At present the UK has one principal general legislative instrument, the Environmental Protection Act (1990) (EPA) supported by a raft of more specific regulations. Part I of the EPA introduces the concept of Integrated Pollution Control (IPC) by which industrial processes designated as most polluting (Schedule A processes) must identify and implement the Best Practical Environmental Option (BPEO) to minimize the total environmental impact of the process. This means that the balance of all emissions to water, air and of solid waste must be optimized. However, the EPA insists only that the Best Available Technique Not Entailing Excessive Cost (BATNEEC) be applied (including abatement, prevention and monitoring techniques). Schedule A processes require authorization from the regulatory body created to implement the EPA, Her Majesty's Inspectorate of Pollution (HMIP). Less polluting processes (Schedule B processes) are required only to control emissions to air and authorization and regulation of these is devolved to Local Government Authorities. Implementation of IPC through the EPA has been phased in over a five year program nearing its conclusion at the time of writing.

The EPA also governs other aspects of environmental management. Part II of the Act controls waste disposal and establishes Waste Regulation Authorities, Waste Disposal Authorities and Waste Collection Authorities to implement this control. In the main these are Local Government Authorities. The act also places a duty of care on any person involved in the handling of waste to prevent escape and to transfer it only to authorized persons. Part III of the act concerns statutory nuisances and clean air for processes not scheduled under Part I.

The second major instrument of legislation is the Water Act (1989), with its supporting regulations, which resulted in the privatisation of the old Water Authorities as public limited companies (plcs) and the formation of the National Rivers Authority (NRA) as the regulatory body for the purposes of the act. The NRA licenses the abstraction of water, by the water companies and by industrial users, and the discharge of water back into natural water courses. There is some overlap in responsibility for water discharge control between the NRA and HMIP, who are responsible for the control of discharge of Red List substances in water for example (see Table 16-3). Drinking water quality on the other hand is overseen by the Drinking Water Inspectorate (DWI), an agency of the Department of the Environment, and marine

pollution regulated by the Ministry of Agriculture, Fisheries and Food. This has resulted in some loss of focus in implementing and enforcing the legislation and a loss in the momentum gained during the late 1980s. However in 1996 responsibility for implementing and regulating the complete spectrum of environmental legislation in the UK will be unified in a new body, the Environment Agency (EA) and this will replace HMIP and the NRA and take on relevant activities of the other bodies. It is hoped that this will restore the rate of progress in management of the environment in the UK.

Other notable UK legislation is as follows:

Health and Safety at Work Act (1974): imposes a responsibility on employers for the safety of their employees handling waste and other hazardous substances.

Food and Environment Protection Act (1985): protects habitats from poor agricultural practice.

Control of Substances Hazardous to Health (COSHH) Regulations (1988): further legislation on employee and consumer safety.

Town and Country Planning Act (1990): implements the EC Directive on Environmental Assessment.

Clean Air Act (1968): now applies only to those activities not scheduled under EPA and covers the emission of smoke and grit.

16.3.2.3 USA

As in Europe, environmental regulation in the USA is a mixture of Federal, State and local government legislation. In several cases, State legislation exceeds the baseline set by the Federal act in severity, notably in California which has put itself at the leading edge of environmental regulation in the world. Federal regulations are administered by the Environmental Protection Agency (EPA) who also carry out and sponsor considerable quantities of research in pollution protection, prevention and monitoring technologies. The EPA had a $6Bn budget in 1991.

The major Federal legislative instruments are as follows:

Clean Water Act: provides for the US Army Corp of Engineers to issue permits for disposal of dredged or fill material on wetlands.

Ocean Dumping Ban Act: controls the indiscriminate dumping of waste in the oceans.

The Safe Drinking Water Act: places a duty on the suppliers of water for public consumption to ensure that it meets, at least, Federal standards. States are given primary responsiblity and can set more stringent standards.

The Lead and Copper Rule: an amendment to the above act mandating the procedure for testing of all drinking water for these metals.

Clean Air Act: provides broad regulation of ambient air quality through a series of standards subject to review at a minimum of five year intervals in the light of current research.

Toxic Substances Control Act: intended to prevent new chemicals from becoming environmental hazards.

Pollution Prevention Act: signifies a shift from pollution management to pollution prevention in the EPA's policies.

Resource Conservation and Recovery Act (RCRA): provides for the protection of human health and the environment by control of the management, storage and disposal of municipal, medical and hazardous substances.

Comprehensive Environmental Response, Compensation and Liability Act (CERCLA): enables Federal authorities to initiate the clean-up of contaminated land sites. Where the source of the contamination is identified, those responsible pay the costs. Where a site has been abandoned or those responsible cannot be identified, a Federal fund (Superfund) is available. This is itself funded by a levy on the major polluting industry sectors in the USA.

16.3.2.4 Trends in Legislation

A number of trends can be observed in the strategy behind environmental legislation worldwide. These include a move from "command and control" policies to the application of market forces to encourage change behavior and the search for competitive advantage by local legislation moving ahead of central legislation.

The UK has developed a typical command and control policy towards the environment. Potential polluters, that is those determined by agencies of the UK Government, must seek licence to operate from the regulatory body appointed by that Government. The operation in terms of output to the environment and, in some cases, the process itself, is subject to strict and often detailed control by the legislation. The performance of the operation is then monitored by the regulatory body and penalties are imposed on transgressors. In other European countries, greater use is made of market forces to discourage, rather than limit, the growth of these processes. This policy shifts the incentive from the regulator to the polluter and from control to reduction of pollution. An example is a proposed "carbon tax" to be applied to companies whose processes result in CO_2 or methane releases into the atmosphere. This has been mooted as a Europe-wide measure and is on the current agenda of the EC. However, the governments of the Netherlands and Belgium have elected to define an "optional carbon tax" for countries wishing to introduce such as scheme ahead of EU-wide legislation. This would impose additional fiscal burdens on the major contributors of carbon to the atmosphere, including the gas, coal and electricity industries, with rebates for those implementing abatement measures. It is not likely that the UK or USA will implement market force policies in the very near future, although both embrace the principle of "polluter pays", as discussed in the next section, and will almost certainly follow the trend if and when it is introduced through EC directives.

In the EU in particular, the central EC policy is to set a framework within which member nations must set local legislation. This is based on best practice within the EU, enhanced by central policy makers. This gives a competitive advantage to those nations which lead on legislative techniques, in effect preempting the EC framework. This role is most commonly taken by a number of EU nations, including Germany, Denmark and the Benelux countries. The UK on the other hand has led the EC in the definition of a fully integrated pollution control policy and the establishment of a single regulatory body, the UK Environment Agency, to oversee it. The EC is considering Europe-wide equivalents to these UK policies, which should provide a degree of competitive edge for the UK.

16.3.3 Economic Pressure

Ultimately, the only meaningful pressure on an industrial concern is economic. This can operate directly, through the imposition of financial penalties for polluters and tax rebates for those using abatement techniques for example, or indirectly through the attitudes of key stakeholders such as shareholders, investors, insurers and employees.

Investors take increasingly significant account of the environmental practices of companies in which they invest. So-called ethical investment trusts have grown in recent years ahead of the general rate of increase in investment. In the UK in 1990 for example, ethical trusts increased on average by approximately 16% in a generally flat climate for investment. The reasons are not entirely altruistic of course. There is a general belief that, in the medium to long term, companies which do not act in a demonstrably responsible way towards the environment will suffer financially. Companies with positive environmental policies therefore appear a better long term investment. For similar reasons there is pressure on public companies to follow good environmental practices from their shareholders.

Recent major environmental incidents such as Seveso, Chernobyl and the Exxon Valdez and a host of smaller pollution incidents have put significant financial pressure on major insurers such as Lloyds of London. For such as the oil, chemical and nuclear industries this has led to considerably increased costs for public liability insurance and the effect is felt by all companies with the potential to pollute. In many cases the premiums can be discounted where the company demonstrates a positive environmental strategy and, particularly for small and medium sized companies, this is a significant further pressure.

The pressure on companies from their employees is more subtle but cannot be ignored. It is a manifestation of the general public pressure discussed below, but with a very focused target, the specific employer, and a position of direct influence. This can make itself felt through the line management structure of the company and trade unions or, increasingly commonly, through workers councils. In the majority of companies in Europe, the USA and the Pacific Rim, the principle of total quality places the needs and aspirations of the employee alongside those of the customer, supplier and shareholder as all represent a financial interest in the company. The attitudes of the workforce to the environment therefore represent a very real pressure on the company which, although financial in nature, is largely derived from altruistic motives.

16.3.4 Public Pressure

The general public has a voice in two main ways, through organized pressure groups such as Greenpeace, Friends of the Earth and the Green political movement, and through local organization, usually in response to a threat to the local environment, the "not in my backyard" (NIMBY) syndrome. In both cases the motives for organization are largely altruistic in nature, although threats to the value of local property can sometimes be a powerful driver. The pressure is effected through media coverage, which could reflect negatively on the target company, through the statutory review and public consultation mechanisms associated with major projects and through political lobbying from within or outside the political establishment.

The growth in the weight and coherence of public opinion is illustrated by the increase in membership of UK environmental pressure groups between 1985 and 1990, for example Greenpeace (30,000 to 385,000), Friends of the Earth (60,000 to 200,000), The Royal Society for Nature Conservation (166,000 to 213,000) and the Green Party (4,500 to 14,900). The increase in effectiveness of the pressure is evidenced by burgeoning environmental legislation as a direct result of changing public opinion.

16.4 Technologies for Environmental Sensors

16.4.1 Introduction

The techniques and technologies used in environmental sensors have developed significantly to meet the growing need for monitoring which is both technically effective and cost-effective. However, the legislative, public and economic pressures outlined in Section 16.3 have accelerated beyond the capabilities of conventional sensor technology to meet these needs in recent years. A number of trends in the way environmental monitors are used in practice serve to illustrate the growing limitations of conventional methods and the opportunities for advanced sensors based on microelectronic, optical and biophysical technologies which these limitations represent. Amongst the more significant are the following:

- the use of *in-situ* monitors rather than remote laboratory-based analysis;
- the move towards continuous monitoring rather than periodic sampling;
- the use of monitors in closed-loop process and abatement control;
- the use of direct, specific measurands rather than indirect, inferred measurements.

These are illustrated in the following sections.

16.4.2 Laboratory Techniques

Laboratory analysis has traditionally been the basis for much environmental monitoring, particularly for water and ground pollution. In many of these applications, the need is for

accurate determination of the levels of specific pollutants rather than a rapid response, as for example in the surveying of polluted ground sites. However, lab techniques are also used in dynamic applications, mainly in water quality and pollution monitoring, where there are no suitable alternatives. Examples include BOD5 measurements (biological oxygen demand over a five day period) and bacterial activity monitoring by culture growth and colony counting. There are increasing pressures to move towards *in-situ* monitoring for these and most other measurands. However, the sheer number of specific measurands required to be monitored, as illustrated by the prescribed substances lists in Tables 16-2 to 16-5, will ensure that laboratory techniques are an essential element in the monitoring armoury for the foreseeable future.

The full range of laboratory analytical techniques are used in environmental applications. Commonly used analytical instrument techniques include flow injection analysis (FIA), thermal analysis (TA), high performance liquid chromatography (HPLC), gas chromatography (GC), mass spectrometry (MS) and optical spectrometry in the infrared, visible and ultraviolet spectral regions. In addition, chemical analytical techniques such as colorimetry, oxidation analysis and chemiluminescence are widely employed. Almost all monitoring of bacterial activity is based on colony counting techniques, requiring several hours or days of culture growth in a laboratory.

In many cases, the appropriate analytical technique for a particular analyte in a particular medium is also prescribed by the regulatory bodies and based on laboratory instrumentation. For example, analysis of water in the UK is guided by the socalled "Blue Book" ("Methods for the Examination of Water and Associated Materials", published by HMSO), whilst in the USA the Environmental Protection Agency lists approved analytical methods for water, gas and oil analysis. This is a significant barrier to the introduction of monitoring instruments based on other methods, as most in-situ monitors are.

The main benefits of laboratory analysis are the range of measurands which can be monitored and the quality of the results obtained. This is maintained by extensive training of operators and by rigid calibration and quality assurance programmes in most central laboratories. However, there are also drawbacks to the use of a central laboratory, principally the physical and temporal separation of the measurement from the monitoring site. These techniques can therefore only provide a historic view of environmental conditon. This is acceptable where conditions are static and tolerable where conditions change only slowly, as in some water treatment monitoring. Increasingly however, monitoring is being used to provide control of treatment and pollution abatement techniques, which should offer a more cost-effective route to meeting consent levels for most industrial processes, given the availability of suitable *in-situ* instruments.

A second situation for which laboratory analysis provides insufficient response is in monitoring for acute pollution incidents, either water borne or air borne, particularly where intake water for industrial processes or public health and safety are at risk. Fugitive emissions of toxic gases from industrial processes and illegal discharges of pollutants into rivers would be examples.

Even where the response of laboratory analysis is appropriate, there are additional issues to consider, particularly the problems of taking samples and preserving their integrity so that they accurately represent the condition of the test site when they are finally tested, and the cost of maintaining central laboratory facilities. Samples taken from natural water courses typically contain a complex mixture of dissolved and suspended substances and living organisms. The constitution of the sample may well change over time and not necessarily in

a way that can be predicted or corrected for. The advantage in making measurements as close to the test site as possible are clear. The cost of a laboratory facility includes not only the capital costs of buildings and equipment, which can be amortised over many thousands of individual tests, but more significantly the ongoing costs of staff training and quality assurance. This is a major factor in the growth of *in-situ* monitoring.

16.4.3 Fixed *In-Situ* Monitors

In-situ monitors are used in dynamic environmental conditions and where sample integrity could not be maintained through a laboratory test programme. In particular the needs of the water industry for sensors to control treatment processes, for other industries to monitor pollution levels in intake and discharge waters, and of combustion-based industries for process and gaseous emission monitoring, have led to a wide range of instruments designed to operate in or near the process. Applications include the control of chlorine disinfectant addition to drinking water, where a fine balance is sought between economy and the safety and acceptability of the water to the consumer, the monitoring of nitrate levels in treated waste water for blending control and the control of combustion processes through *in-situ* monitoring of flue gases.

In-situ monitors currently used are mainly based on electrochemical, colorimetric and optical techniques. The most common and widely used in the water environment is the pH sensor, based on a glass ion-selective electrode structure, both for control of treatment processes and as a measure of general pollution levels. Several other ion-selective electrode based instruments are also used in the water industry, for example to monitor nitrate, phosphate or chlorine ion levels. Many substances in water are monitored using self contained versions of laboratory analytical techniques which can operate automatically or semiautomatically at the testing site. These are mainly colorimetric or photometric analyzers for metals and other ions. Optical techniques are the mainstay of in-process flue gas monitoring, particularly infrared absorption spectrophotometers for NO_2, SO_x, CO and CO_2.

Two further application areas would also benefit from the use of fixed *in-situ* monitors. These are industrial discharge monitoring, where the response of laboratory and portable instrument tests may not be sufficient to provide early warning of acute pollution incidents, and ground monitoring around storage sites to provide early warning of leaks. In these cases, however, the cost of purchase and ownership of presently available instruments has been a significant barrier. The increasing shift of responsibility for the protection of the environment from the centralized regulator to the local polluter will increase the pressure for major industries to install fixed discharge monitors.

The major problems with fixed monitoring instruments centre on their reliability. To ensure the calibration and proper function of the instruments requires a labor intensive maintenance program. Moreover, many of the *in-situ* automatic analysers require regular (and frequent) resupply with reagents. A second aspect is the nature of the medium in which the monitors must function, particularly in waste water and natural water courses. These can include heavy concentrations of suspended solids and fibrous material which collect on any sensor element protruding into the flow and mask the sensor, a phenomenon known as "ragging". Algal growth on sensitive elements exposed to natural waters is a similar problem. *In-situ* instruments are therefore frequently designed with self-cleaning mechanisms. An alternative ap-

proach is to position the sensitive element out of the flow and to sample the medium. This transfers the problem to one of obtaining a representative sample for measurement, which is not straightforward in these difficult media.

The trend towards more in-situ monitoring presents a significant opportunity for advanced and emerging sensor technologies, as discussed in Section 16.4.5.

16.4.4 Portable Sensors

Portable sensors are mainly used in water and ground pollution monitoring applications where a rapid, general view of the condition of the environment is required. Assessment surveys of contaminated land are a typical example, where samples may subsequently be analysed in detail in the laboratory. Spot checks carried out on water courses and the tracking of pollution incidents are further examples.

The main techniques used in portable instruments are electrochemical, ionisation detection and, in a few cases optical. Examples include PID and FID instruments (photo- and flame-ionization detectors respectively), to measure hydrocarbon and VOC levels in ground surveys and leak detection, and water quality monitors which include electrochemical sensors for pH, ammonia and dissolved oxygen and sensors for conductivity, turbitiy, temperature and other parameters in one dippable probe. A further example is the personal gas monitor, of which there is an increasing number, which monitors CO, explosive or toxic gases in the local environment using electrochemical cells. Advanced sensor techniques promise to make a significant impact in all of these areas.

16.4.5 Advanced and Emerging Sensors

Environmental monitoring applications are widely seen as one of the major opportunities for advanced sensor technologies. In particular fiber optic chemical sensors, biosensors and solid state sensor techniques are expected to make an impact in this market. In many cases this is a result of particular benefits which the advanced technique brings over more conventional sensors, but in other cases the benefits are less obvious. The main driving force for the introduction of advanced techniques is reduced cost of ownership.

16.4.5.1 Fiber Optic Chemical Sensors

Fiber optic chemical sensors for environmental applications fall into two braod categories, remote fiber spectrometers and optrodes. Remote fiber spectrometers use conventional spectrophotometric techniques but with the optical signals transmitted to and from the sample via optical fibers. In this way the active parts of the instrument can be remote from the sample area, by many meters in some cases. The optical properties of the sample itself are monitored in this technique, the absorption, fluorescence or raman spectra for example. Optrodes monitor the optical properties of a species-specific reaction immobilized at the end of an optical fiber, to which the sample is introduced through a membrane or sampling system.

A great deal of research and development work has gone into such sensors for environmental monitoring, particularly in the USA, because of a research program sponsored and undertaken by the Environmental Protection Agency since 1982. Examples of sensors under development under this program and others include remote spectrometers and optrodes for organochlorides such as trichloroethylene, benzene and toluene, and infrared absorption-based sensors for methane. A recent development supported by the UK grovernment scheme ETIS is of an optrode for pesticides in groundwater.

The principal perceived benefit of fiber optic technology stems from the massive investment made in optical communications. The costs of the fibers themselves and components such as sources, detectors and connectors have reached very low levels. Thus fiber optic sensors are seen as potentially very low cost. This view unfortunately ignores the fact that the major costs of instrumentation are associated with ensuring the reliable operation of the sensor, for example the housing, sampling and self-cleaning aspects. Additional benefits arise from the ability to separate electronically and optically active elements of the sensor from the measurement site. A passive all-optical sensor element should have very high reliability. This is of value where the measurement environment is hostile, at high temperature or in an explosive atmosphere for example, or where maintenance of the sensor element may be difficult, with leak monitoring sensors buried beneath storage sites for example.

There is, however, a number of barriers to the success of fiber optic chemical sensors in this or any other market. Technologically, it has proven difficult to devise sensors which will operate successfully and reliably in "real world" environments such as waste water or natural water courses. There is also a long and typically costly development path from demonstrating sensitivity in the laboratory to a practically viable sensor. Although the total market potential for environmental sensors is very attractive (see Section 16.5) it is fragmented, that is the market for an individual sensor type within the territory avialable to a particular manifacturer is in most cases limited. The commercial justification for investing in this development is therefore not often very strong. From the end-user's viewpoint, fiber optic sensors are unproven. There is not the track record of reliable operation that system designers would look for, particularly in established applications like water treatment monitoring. Moreover established manufacturers are reluctant to risk hard-earned reputations for reliability by launching instruments based on unproven technology. Finally, there are as yet no standards for data protocols or specifications of fiber optic sensors. Many of these comments apply equally well to other advanced sensor technologies.

It is likely that fiber optic sensors will find applications in specific niche applications, such as groundwater sensors or sensors for hostile operating environments. In these cases the benefits may be argued to justify the high cost of development. There may also be opportunities for hybrid sensors using fiber optic technology in conjunction with other advanced techniques such as microengineered systems or biosensors. There is a long road, however, to the widespread penetration of fiber optic sensors into the environmental market.

16.4.5.2 Biosensors

Biosensors, particularly those based on whole cell responses, have greater potential for the monitoring of oxygen demand and general toxicity in water. Whole cell oxygen demand sensors are currently available, operating on a batch process which reduces the measurement time

from five days to one hour and fully continuous versions with a response time of a few minutes are in development. These will provide far more meaningful data for the control of water treatment. Toxicity biosensors are also in development which monitor the fall in activity levels of bacteria in the presence of a contaminated sample. However, as both of these examples use the metabolic activity of the cell as a primary measurand there is a potential problem with cross-sensitivity between oxygen demand and toxicity. Biosensors based on enzyme mediated reactions are also in development, particularly for measuring heavy metals, phenols and other organics in drinking water. The inhibition of enzyme reactions by the presence of toxic gases has also resulted in research on personal toxic gas monitors based on enzyme biosensors – so called "chemical canaries". The same phenomenon is also exploited by sensors for pesticides and herbicides, intended for use in ground water and contaminated land monitoring.

The major reason for the extensive interest in biosensors, particularly for the monitoring of pollutants harmful to humans, is the perception that their response to contaminants will be closer to that of the body than would a sensor based on chemical or physical phenomena, giving more meaningful and rapid early warning. There is also a major worldwide research effort into biosensors for healthcare applications, for largely the same reason, and it is hoped that environmental sensors can be adapted from some of these developments. There is also the strong prospect that biosensors can be made physically very simple and low cost. At least a portion of the enthusiasm, however, is a simple reflection of a very healthy climate for investment in biotechnology-based industries, particularly in the USA.

Very real problems remain in the development of practical biosensors which will operate in environmental applications. Principal amongst these is the stability of the bioelement, be it whole cell or enzyme. In many sensor architectures the bioelement must be immobilised on a substrate so that it can be monitored in a controlled and consistent way over an extended period. However there is a trade off between the strength of this immobilisation and the sensitivity of the sensor. The compromise which is frequently required sacrifices long-term stability for acceptable short term performance. Continuous monitors and reusable sensor elements are therefore difficult to achieve.

The potential benefits of biosensors are eagerly avaited by many in the environmental area, particularly in the water sector. However it may require a significant breakthrough in the medical area before these benefits are seen to be practically achievable. It is also possible that biosensitive elements will be combined with the other advanced sensor technologies discussed here and applied in new and unconventional instruments for monitoring the environment.

16.4.5.3 Solid-State Sensors

Solid-state sensor techniques include microelectronic sensors, thick and thin film sensors and microengineered silicon devices. All are based on fabrication techniques familiar in the microelectronics industry and refined in that field for high quality, high volume and very low cost. They are also all characterised by small physical size and rugged, monolithic construction. These features are of significant potential benefit in environmental applications.

Chemically sensitive field effect transistors (CHEMFETs) can be configured to exploit ion-selective elements (ISFETs), gas-sensitive membranes (GASFETs), immobilized enzyme-mediated reagents (ENFETs) and other mechanisms. This flexibility has led to the laboratory

development of a very wide range of sensors with potential applications in air, water and ground monitoring, including sensors for pH, nitrates, ammonia and metal ions. However, the only example to reach the market in a significant way is the pH sensitive ISFET. Despite very heavy investment, it has proven difficult to achieve long term stability in more complex CHEMFETs, as would be essential for more widespread penetration of the environmental monitoring market. Cross-sensitivity has also been a considerable problem. However, the market prospects for ISFETs in particular, as replacements for ion-selective glass electrodes in water-based applications, are very high and there is still sigificant pressure to develop practical devices.

Thick film fabrication techniques, principally screen printing of reagents and electrodes onto plastic, glass or ceramic substrates, are increasingly being used for electrochemical sensor elements. Commercially available examples include disposable sensors for lead and copper in drinking water based on anode-stripping voltammetry and vibrating-electrode sensors for residual chlorine and ammonia. A UK collaborative project is under way to develop a multisensor array fabricated using thick film techniques for water quality monitoring, including sensors for temperature, conductivity, pH and dissolved oxygen, with reference electrodes, all on a single substrate. Chlorine and heavy metals sensors will be added in a later phase of the project. Thick and thin film sensors are also being researched for exhaust emission monitoring and engine management in automobile engines. In addition to lower cost alternatives to the conventional lambda oxygen sensor, NO_x and possibly other gases may also be sensed using low cost film-based sensors.

The principal attraction of thick film technology in particular is the very low investment required to set up a production facility, compared for example to a silicon-based sensor fabrication line. Whilst the flexibility of silicon sensors is not available from thick film sensors, this economic advantage will ensure a significant future market opportunity.

Microengineered silicon sensors have mainly appeared as physical sensors, temperature, strain or acceleration for example. However, there is a growing number of developments, mostly at very early stages, aiming to reproduce the functionality of laboratory analytical instruments on a microscale. Examples include a gas chromatograph on a chip, which has been commercially available for some years, flow injection analysis systems and various forms of mass spectrometer integrated in silicon. The techniques are well established to form flow channels, reaction chambers, valves, pumps, heaters and other elements of conventional analytical instruments on a millimetre scale and with features on a micron scale. Moreover, the silicon substrate can, in principle, also support the integration of sensor elements such as CHEMFETs or optical sensors and of signal processing electronics. The long term prospect is therefore for devices with the specific analytical capabilities of laboratory instruments but small enough and cheap enough to operate in the field. The investment required to set up manufacturing facilities for microengineering will be high but, as with microelectronics fabrication, many different sensor designs for diverse markets could be made in a single facility. The investment cost could therefore be amortised across the broadest possible market. Much development work is required before these devices become practically usable but they would revolutionize environmental monitoring as well as many other application areas.

16.5 Market Opportunity

16.5.1 Introduction

The size of the market for sensors for environmental monitoring has proven particularly difficult to predict with any great accuracy. This is a consequence of the blurred distinction between sensors for environmental purposes and those for process control or health and safety in some industries. Also, in most market reports the total market for environmental control and management is given with no clear breakdown to show the proportion appropriate to sensors. However, a great deal of market research has been applied to the environmental area and we can synthesize a view of the potential for sensor products from this. All sources of market information used are given in the Bibliography.

16.5.2 The Market for Environmental Sensors

CEST (1991) forecast the total expenditure required in the ten years to 2000, to address the major environmental issues discussed in Section 16.2, in the EC and USA to be $1,380Bn and $1,700Bn respectively. We make an assumption that the expenditure in the rest of the world will be broadly similar to that for the EC (as in generally the case with instrumentation products), giving an estimated total world market for environmental technologies of $4,512Bn, over the period 1991 to 2000. This is a generous estimate as it includes all costs associated with environmental protection and makes the assumption that expenditure will be forthcoming in all areas of environmental interest. In the light of environmental investment in the first three years of the period, which has not been as extensive as was assumed, and taking into account the more limited range of issues relevant to the applications considered in this analysis, we take an estimated figure of $2,500Bn as the total world market size over the ten year period.

From a study of the US market for environmental monitoring instruments by the DTI (1993) we can determine the rough proportion of the total expenditure appropriate to sensor products as 1%. Assuming that the proportion is the same across all territories, we deduce a total estimated world market for environmental sensor products of $25Bn over the period.

From the same DTI report we determine that the US markets for sensors in the three main media categories, water, air and land, are roughly in the proportion 4:5:2. According to many market reports, the environmental market has grown on average at 10% over the past few years and is expected to continue at this rate. In practice, individual media and territorial sectors will grow at somewhat different rates between 6% and 13%, but the average is valid.

The estimated markets for water, air and land-based environmental sensors, in the EC, USA and the rest of the world (ROW), for 1991, 1995 and 2000 are therefore as shown in Table 16-10. Within the EC figures, Germany, France and the UK account for approximately 32%, 16% and 10% of the total respectively.

The proportions of the UK market associated with ambient and source monitors, fixed and mobile monitors is shown in Figure 16-1, derived from a report by NEDC and GAMBICA (1992). Again, whilst there will be local differences, the proportions will be broadly valid in all sectors.

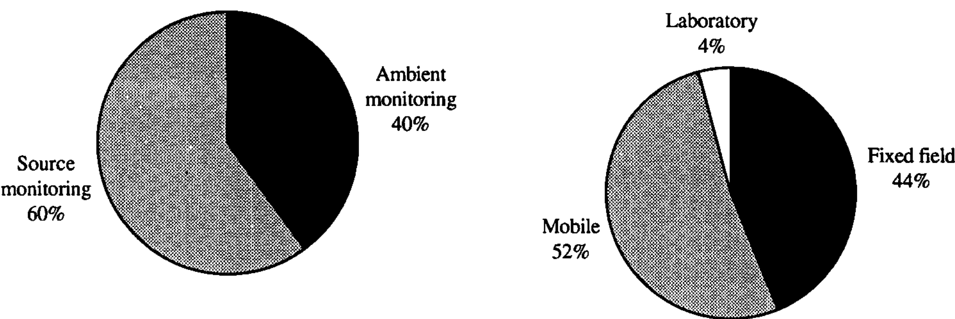


Figure 16-1. UK expenditure on environmental monitoring 1990/91.

Four examples of instruments for specific applications serve to illustrate the growth trend in the overall market.

Figure 16-2 shows the forecast worldwide market for portable analysers for environmental monitoring, derived from Frost & Sullivan, 1992. These are assumed here to comprise 11% of the total portable analyser market. In practice, this proportion is likely to increase as monitoring functions move from the laboratory into the field, particularly for contaminated land assessment.

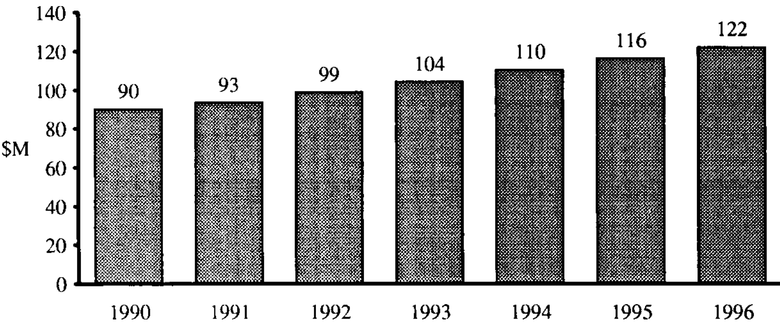


Figure 16-2. World market for portable analyzers for environmental monitoring.

Figure 16-3 shows a similar forecast of the worldwide market for sensors for wastewater monitoring. The average annual growth of 11% is more typical of environmental applications and represents an attractive potential market. The figures are derived from a report by Frost & Sullivan, 1992.

The world market for gas sensors, including but not exclusive to environmental applications, is illustrated in Figure 16-4 (derived from MIRC, 1992). Steady market growth from the early 1990s on reflects the increasing importance of air quality and industrial emission monitoring, as well as the emergence of domestic gas sensors as a significant environmental monitoring market.

Finally, Figure 16-5 shows the US market for chemical sensors in vehicles, most of which are concerned with exhaust gas monitoring or engine management to minimise emissions

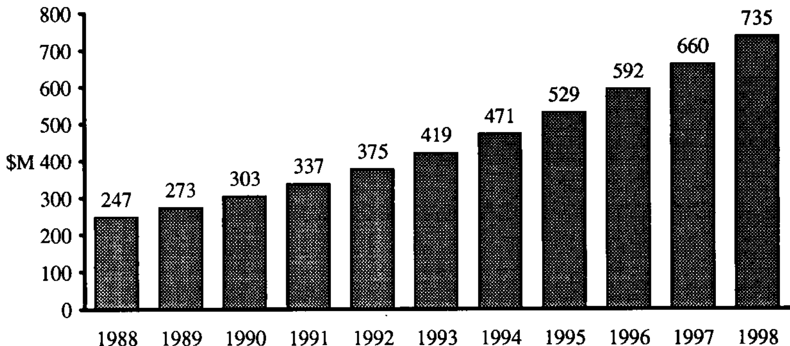


Figure 16-3. World market for sensors for wastewater monitoring.

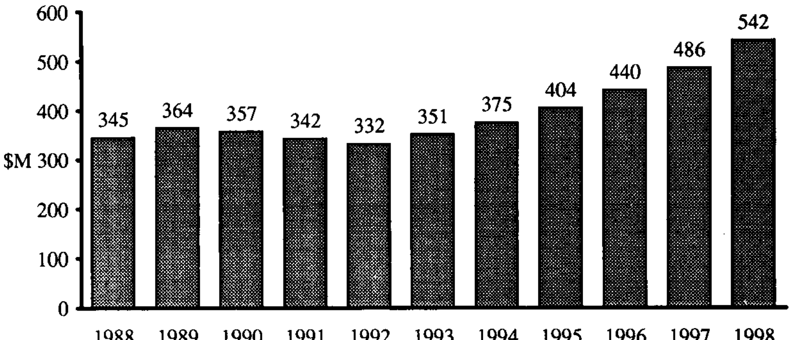


Figure 16-4. World market for gas sensors.

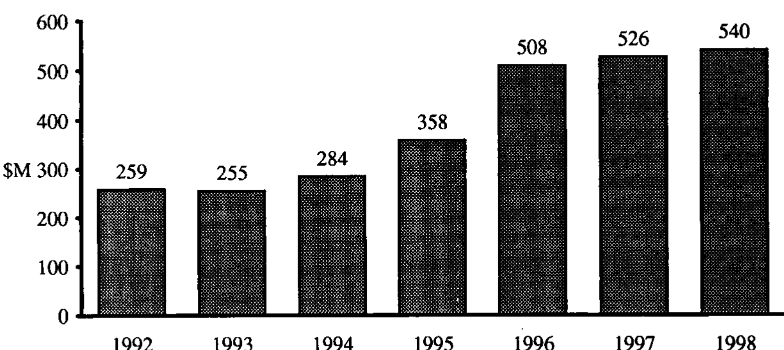


Figure 16-5. US market for automotive chemical sensors.

(Frost & Sullivan, 1993). The surge in demand around 1995 reflects the tightening of vehicle emission regulations, particularly in the USA and Europe. Growth settles back to the level of automobile sales after the initial introduction program.

16.5.3 Industry Attitudes to the Environment

It is interesting to note the response so far of industry to the pressures outlined in Section 16.3. A number of surveys have been carried out in the UK in recent years on the attitudes of companies affected by environmental regulations.

Seventy nine metals-based businesses, subject to Integrated Pollution Control Part 1 or Local Authority Air Pollution Control under the Environmental Protection Act, responded to a survey organized by EEF in 1993. One in three thought that their business would be "seriously affected" by the implications of the Act or, worse, that they were likely to go out of business. It is estimated that the cost of compliance for a company operating one prescribed process is roughly the same under either control regime at £375,000 to £400,000 and that a company operating two such processes must spend almost £1,000,000.

In 1994 the UK Chemical Industry Association reported that UK chemicals companies have consistently fallen short of planned capital spending on pollution prevention. Table 16-11 illustrates this by the percentage of total capital investment dedicated to pollution prevention planned each year since 1991, and the actual investment made.

Table 16-11. Chemical industry spending on the environment (% of total capital investment).

Year of report	1991	1992	1993	1994	1995	1996
1991	23	23				
1992	10	14	21	24		
1993	10	14	15	16	23	
1994	10	14	14	15	16	20

Conversely, a survey of companies supplying goods and services to the UK environmental market, carried out by Ecodata in 1993, revealed a far more optimistic view of the market trends. Figure 16-6 shows the percentage of respondents who experienced a change in their level of sales in 1992/3 and those anticipating a change in 1993/4. The 1992/3 figures indicate a bouyant market, with over half the respondents increasing their sales, but the expectation

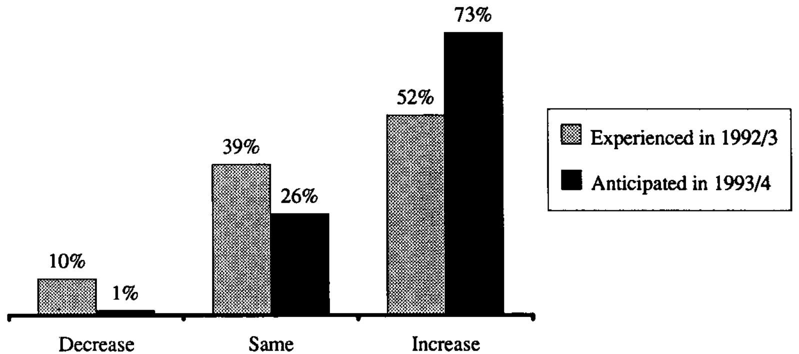


Figure 16-6. Change in sales of respondents.

for 1993/4 is strikingly positive. The potential value of the market to the respondents is further illustrated by their expectation of increased competition, both from the UK and elsewhere, as shown in Figure 16-7.

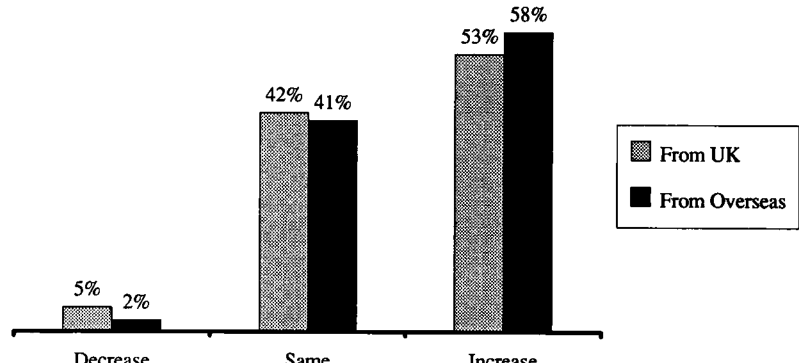


Figure 16-7. Anticipated change in competition to respondent.

16.5.4 The Environmental Monitoring Supply Industry

As with many instrumentation industries, the supply of environmental monitors is dominated by a small number of companies, with diverse interests across a number of application areas. In the USA for example, just three firms hold 60% of the domestic enivironmental instrumentation market (DTI 1992). The industry is, however, relatively small and very fragmented. The vast majority of participants are small companies, serving specific niches in the market.

Since new advanced sensor products also tend to address niche applications, the competitive situation is very favourable to the introduction of new products or the entry of new suppliers. Unfortunately, the corollary of this is that the potential market for many new products is simply too small to justify the development costs. Where the product is aligned with the needs of the regulatory bodies, for example if it allows a specific pollutant to be monitored effectively for the first time, the major regulators have available funds for research and development and most sponsor extramural research and development. The Environmental Protection Agency in the USA, National Rivers Authority in the UK, NOVEM in the Netherlands and other national bodies are both generators and sponsors of new sensor development. Product specification and positioning for new developments must therefore bear the regulators in mind as much as the eventual end user.

A major role in the supply of environmental monitoring products is played by the environmental engineering and consulting companies. These have the responsibility for specifying all of the subsystems required to deliver a system solution to the enduser, a waste water treatment plant or flue gas desulphurisation system for example. The criteria on which they make this choice include technical performance but, because of the typical structure of their contract with the end user, have purchase price as a very high priority. The vast majority of consulting companies positively avoid involvement with new product development in order to

maintain their independence. They are, however, of critical importance to the sensor manufacturer as their position in the value chain is very powerful.

It is likely that, with the evident growth of the opportunity for environmental products and services, major industrial concerns will seek to become involved. The most probable strategy would be one aquisition and merger with existing players. The industry in eight to ten years time is likely to be dominated by a small number of giant, vertically integrated suppliers acting as product developer and consultant, supplying turnkey solutions to the major end users, utilities and regulatory authorities.

16.5.5 Market Trends

The need for investment by industry and governments worldwide becomes ever more urgent as we understand more of the effect we have on the environment. The environment is high on the agenda of most national governments and regulatory pressure is being applied to polluting industries. However, the upbeat market forecasts for all environmental products and services are based on the premise that this pressure will be rigorously applied and backed up by significant penalties for offenders. Recent evidence suggests that, in some territories at least, this is not the case. Industrial companies will only in rare circumstances invest serious resources to environmental issues if the cost of compliance with regulations clearly exceeds the cost of noncompliance.

It is also true, however, that companies are taking an increasingly responsible and broad view of what constitutes the cost of noncompliance. The "corporate conscience" is becoming another, growing driving force for the environmental market.

The rare combination of legislative and public pressure, rapidly emerging new technologies, including sensor technologies, and the responsible attitude of many major industrial concerns will ensure a healthy rate of growth in the market for environmental sensors for the foreseeable future.

16.6 Bibliography

Periodicals:

Control & Instrumentation (Morgan-Grampian plc).
Environmental Business (Information for Industry Ltd).
Environmental Business Magazine (Information for Industry Ltd).
Environmental Sensors (IOP Publishing Ltd).
Industrial Waste Management (Faversham House Group).
Integrated Environmental Management (Blackwell Scientific Publications).
International Environmental Technology (Labmate Ltd).
Physics World (IOP Publishing Ltd).
Pollution Prevention (MacDonald Communications Inc).
Process Engineering (Morgan-Grampian plc).

Scottish Envirotec (Don-Mor Productions Ltd).
The ENDS Report (Environmental Data Services Ltd).

Reports, books etc.

Contaminated Land: Market and Technology Issues, CEST, 1992.
Environment Industry Yearbook, The Environmental Press, 1994.
Environmental Opportunities: Building Advantage out of Uncertainty, CEST, 1992.
European Sensor Markets, MIRC, 1992.
Field/On-site Waste Water Analytical Instrumentation Markets in the US, Frost & Sullivan, 1993.
Financial and Structural Analysis of the World Sensor Industry, Frost & Sullivan, 1993.
Hazardous Waste Management, Financial Times Business Information, 1990.
Industry and the Environment: A Strategic Overview, CEST, 1991.
Market Opportunities for Advanced Sensors in Health and Safety, R W Bogue, 1992.
Portable Biosensors: Technology and Markets, Futurescope, Arthur D Little, 1990.
Requirements and Market Opportunities for Advanced Sensors in Chemical, Pharmaceutical & Biotechnology Industries, Adams Business Associates, 1991.
Sector Report: USA Environmental Instrumentation, DTI, 1992.
Sector Report: USA Water Pollution Issues, Technologies and Opportunities, DTI, 1993.
The Environmental Monitoring Business, NEDC, 1992.
The European Market for Industrial Gas Sensors, Frost & Sullivan, 1991.
The European Market for Portable and Transportable Analytical Instruments, Frost & Sullivan, 1992.
The Role of Advanced Sensors in Environmental Monitoring and Pollution Control Practices, R W Bogue, 1990.
The Role of Advanced Sensors in the Power and Energy Industries, R W Bogue, 1991.
Waste Management, CEST, 1990.
Water, Resource and Opportunity, CEST, 1992.
World Emerging Sensor Technologies, Frost & Sullivan Market Intelligence, 1993.
World Fiber Optic Sensor Markets, Frost & Sullivan Market Intelligence, 1993.
World Niche and Specialised Sensor Markets, MIRC, 1992.
World Sensor Markets, Frost & Sullivan, 1993.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

17 Automotive Sensors

PETER COCKSHOTT, Broughton in Furness, UK

Contents

17.1 Introduction 493
17.1.1 Sensors and Systems in Vehicles 493
17.1.2 The Market for Automotive Sensors 493
17.1.3 Reasons for Sensor Requirements 494
17.1.3.1 Emission Control and Fuel Economy 494
17.1.3.2 Safety, Handling and Vehicle Performance 494
17.1.4 Sensor Performance Requirements 494
17.1.4.1 Sensor Performance and Environment 494
17.1.4.2 Packaging 495
17.1.4.3 Cost 495

17.2 Sensors for Engine Systems 495
17.2.1 Gasoline (Spark-Ignition) Engines 495
17.2.1.1 System Requirements 496
17.2.1.2 Gasoline Engine Sensors 498
17.2.1.3 Gasoline Engine Sensors: Conclusions and Market Trends 506
17.2.2 Diesel (Compression-Ignition) Engines 507
17.2.2.1 System Requirements 507
17.2.2.2 Diesel Engine Sensors 508
17.2.2.3 Discussion and Conclusions: Diesel Engine Sensors 510

17.3 Sensors for Body Systems and Vehicle Dynamics 510
17.3.1 Braking System and Traction Control 510
17.3.1.1 System Requirements 510
17.3.2 Steering 511
17.3.2.1 System Requirements 511
17.3.3 Suspension 511
17.3.3.1 System Requirements 511
17.3.4 Safety and Comfort 512
17.3.5 Sensors for Vehicle Dynamics 512
17.3.5.1 Wheelspeed Sensors 512
17.3.5.2 Accelerometers 512
17.3.5.3 Steering Wheel Angle Sensors 513
17.3.5.4 Steering Wheel Torque Sensors 513
17.3.5.5 True Groundspeed Sensors 513

17.3.5.6 Fiber Optic Gyros 514
17.3.5.7 Solid State Gyros 514
17.3.5.8 Suspension Height Sensors 514
17.3.5.9 Crash Sensors for Airbag Deployment 514
17.3.6 Vehicle Dynamics Sensors: Conclusions and Market Trends 515
17.3.6.1 Market Trends 515

17.4 Sensors for the External Environment 515
17.4.1 Intelligent Cruise Control 515
17.4.2 Lane Following Control 516
17.4.3 Collision Avoidance 516
17.4.4 Vision Enhancement 517
17.4.5 Navigation 517
17.4.6 Driver Status Monitoring 517

17.5 Silicon Sensors in Vehicles 517
17.5.1 Silicon Properties used in Sensors 518
17.5.2 Silicon Micromachining 518
17.5.3 Pressure Sensors 518
17.5.4 Accelerometers 519
17.5.5 Other Silicon Sensors 519
17.5.6 Automotive Silicon Sensors: Conclusions 519

17.6 Smart Sensors and System Integration 519
17.6.1 Smart Sensors 520
17.6.2 System Integration 520

17.7 Conclusions 521

17.8 References 522

17.1 Introduction

The market for sensors for use in electronic control systems in road vehicles is in many ways a unique one. Many of the control systems are complex, their number and complexity increasing year by year. The sensors involved need to be of high performance: high accuracy, tolerance of a hostile environment, and ability to function with minimal maintenance over the life of the vehicle must all be achieved at a cost perhaps an order of magnitude lower than that of an equivalent sensor for industrial use. However, the market is large, and seems set to continue to grow throughout the present decade.

17.1.1 Sensors and Systems in Vehicles

The concept of sensors and control systems in vehicles is less new than is commonly perceived. The car of the 1950's, with no electronics more sophisticated than a valve radio, contained a few sensors for driver information: fuel gauge, oil pressure, coolant temperature. In addition, however, there were mechanical control systems which were the direct forbears of present day electronic systems. The distributor, controlling the spark, sensed engine speed and manifold pressure; the carburetor sensed inlet air flow; the diesel governor sensed engine speed; the automatic transmission sensed engine speed and throttle position; the power assisted brakes and steering had simple means of sensing driver demand and providing assistance from hydraulic or pneumatic mechanisms.

So it appears that *mechanical* control systems have been available for many years, developing steadily from the earliest days of the road vehicle: what has taken a leap forward in recent years has been the incorporation of *electronics*, and electronic sensors which are integral parts of the electronic systems. The reasons why this has become necessary (and indeed possible) will be discussed in detail later in this section.

17.1.2 The Market for Automotive Sensors

In 1990, 10% of the value of a passenger vehicle was the electronics, including of course sensors. By 2000, it is predicted that this proportion will double to 20%.

Much of the growth in electronics usage in vehicles since the early 1970's has been driven by legislation, initially for reasons of pollution reduction and fuel economy, and also to a lesser extent for driver and passenger safety.

The market may be divided into two sectors, passenger cars and commercial vehicles. The commercial vehicle is a piece of capital equipment, and features (including electronic systems and sensors) not required by legislation will only be called for if they offer a perceived improvement in reliability or cost of operation. The passenger car, on the other hand, is a consumer item: comfort, driving pleasure and status provide additional drives for incorporating sophisticated features.

The market for vehicle components has always been very competitive, and pressures to minimise costs are always strong. For this reason, it is often difficult, in the absence of legislation, to sell an electronic system for a higher price than of the mechanical system it replaces.

17.1.3 Reasons for Sensor Requirements

17.1.3.1 Emission Control and Fuel Economy

These factors are primarily concerned with engine controls, and so are conveniently grouped together. From the early 1970's, air pollution from vehicle emissions was recognised as a problem, and regulations to limit emissions have since been introduced progressively throughout the developed world, led by California and followed by the rest of the USA, Japan, and Europe. Concurrently with this process, successive oil crises in the Middle East led to regulations to limit fuel consumption in the USA. This, combined with the steep rise in real terms of the price of petroleum-based fuels, provided strong incentives for improving the fuel efficiency of vehicles. The result is that today the average car travels 33% further on a gallon fuel than a pre-1970 car, but emits 16% of the pollution [1].

17.1.3.2 Safety, Handling and Vehicle Performance

Increasingly, vehicles are sold on safety features. Anti-skid braking (ABS) is now available on many models, and features such as Traction Control and Adaptive Suspension contribute to the safety as well as to the comfort and driveability of a vehicle. Airbags are rapidly gaining acceptance as additional safety features to supplement seatbelts for driver and front-seat passenger. In recent years, road deaths have decreased by 26% while the number of vehicles on the roads has increased by 32%, but nonetheless this level of accidents is still unacceptable, and improved safety features remain a high priority.

Improvement in comfort of vehicle occupants is also an important sales feature, particularly in high-line passenger vehicles. Air conditioning, and also vehicle suspension and steering systems are using electronic controls to an increasing degree.

17.1.4 Sensor Performance Requirements

17.1.4.1 Sensor Performance and Environment

Accuracy requirements for sensor information depend on the requirements of the System in which it is to be used, but in areas such as engine control high accuracies are required, particularly in the primary sensors, in order to meet the legislative requirements on emissions. For such sensors accuracies of 1% or better may be required, and this performance must be maintained over a temperature range of zero to 80 °C. Devices must continue to operate with relaxed accuracy but no permanent drift or damage between −40 °C and +150 °C. They must operate in an environment of mechanical vibration, thermal shock, oil leaks, fuel, hydraulic fluids, salt spray and mud.

Road vehicles, and particularly passenger cars, do not have service schedules supervised by regulatory authorities, and servicing is often carried out by non-specialised personnel. In spite of this, a life of 100,000 or 150,000 miles is normally called for.

17.1.4.2 Packaging

For these reasons, suitable protective packaging of the sensor becomes a key feature: not only must the basic sensor be able to deliver the required accuracy, but steps must be taken to protect it from the environment. High temperature electronics are also required for signal conditioning and control units.

17.1.4.3 Cost

As components for a consumer product, parts for passenger vehicles need to be highly competitively priced. If, say, an industrial pressure transducer of sufficient performance were selected for manifold pressure sensing in an engine control system, its cost would be a significant fraction of that of the engine itself. This is considered to be an unacceptable cost, and suppliers of manifold pressure sensors must be content with $9–$12 per device. How suppliers are able to meet this kind of sensor cost target and remain profitable will be explained later in the chapter for the various sensors described, but basically there has been a continuous process of development and cost reduction capplied to engine control sensors over the years since they first began to be manufactured.

A key factor is of course the large volumes required (currently 20 million automotive silicon pressure transducers are used annually throughout the world) [2] so development costs are quickly amortised. Devices are manufactured in wafer form in Silicon, and they are assembled, packaged and tested using techniques perfected in the Semiconductor industry.

Another factor is standardisation: the number of different types of device, in terms of pressure range, and other characteristics, is strictly limited. For example, one pressure range for naturally aspirated systems will be required, and a higher maximum pressure will by required for turbocharged systems, etc, but this simplicity is advantageous for low cost manufacture.

17.2 Sensors for Engine Systems

Engine controls were the first area in which electronics and sensors began to be used. Electronic ignition and fuelling systems for gasoline engines have been in use for many years, and they continue to increase in complexity. The reasons for this lie in legislation for emissions control, because purely mechanical systems are no longer able to provide sufficiently tight control to keep emissions within the limits required in some countries.

17.2.1 Gasoline (Spark-Ignition) Engines

As gasoline and Diesel engines operate in quite fundamentally different ways, we will consider first the gasoline engine.

17.2.1.1 System Requirements

In a gasoline engine, a mixture of gasoline and air is prepared in the inlet manifold, and this mixture is then drawn into the combustion chamber where it is ignited by the spark. The energy released by this combustion process provides the motive power for the vehicle. For satisfactory operation in terms of power output from a given quantity of fuel, of exhaust emissions, and of "driveability" (ie, that the vehicle responds smoothly and predictably to changes in driver demand), the air/fuel ratio and the timing of the spark must be controlled for all engine conditions. Thus it is the control of ignition timing and air/fuel ratio that is the basic function of the gasoline engine control system.

Now the running conditions of an engine can be expressed in terms of *engine speed* and *engine load*, and it is common practice to "map" the various requirements (fuel per stroke, ignition timing, etc) as a contour map in the speed – load plane (Figure 17-1). Measurement of engine speed is of course straightforward at least in principle, but engine load, which in this context represents the force exerted at the piston by the combustion process, has to be inferred from one of a number of alternative measurements. The most common ways of inferring load are Manifold Absolute Pressure (MAP) or Air Mass Flowrate (AMF) into the manifold.

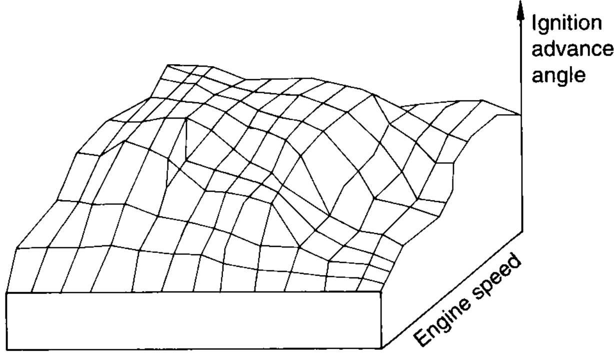


Figure 17-1.
Map of ignition advance versus manifold absolute pressure (engine load) and engine speed.

It is interesting to compare this with the traditional mechanical system using a carburetor and distributor: the distributor, in addition to its basic function of directing the HT supply for the spark to each cylinder in the correct firing order, actually controls the ignition timing as a function of engine speed (providing advance with speed by means of centrifugal bobweights) and engine load, providing a variation in timing with manifold pressure from the vacuum capsule. The carburetor measures airflow, which is related to power (the product of speed and load), by passing the air flow through a venturi. The depression produced by this venturi draws fuel at the required rate through a calibrated jet.

However, these mechanical systems are not sufficiently precise to provide ignition timing and fuelling with the accuracy necessary to meet present day emissions regulations in many parts of the world, so modern systems measure speed and load electronically and the optimum conditions are determined by an electronic engine management system, such as that shown schematically in Figure 17-2.

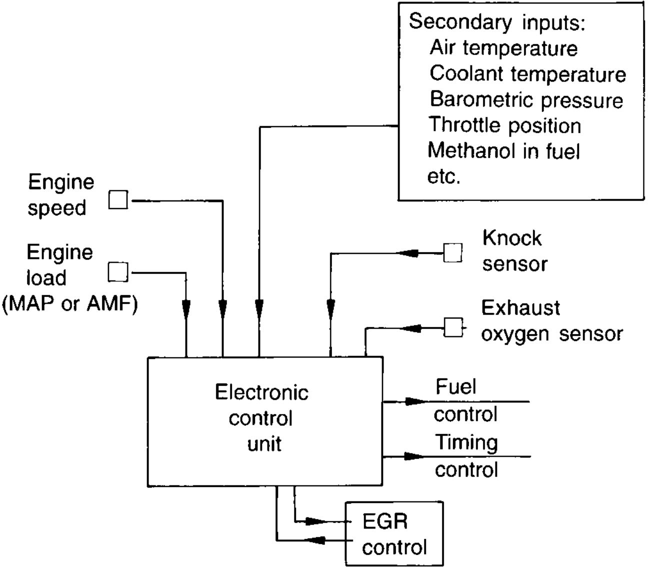


Figure 17-2. Gasoline engine control system.

In addition to these *primary* sensors for speed and load mentioned above, a crankshaft position reference is required from which to reference ignition and fuel injection timing, and also various *feedback* sensors are specified.

Exhaust gas oxygen sensors provide feedback on the air-fuel ratio. Many systems run at their optimum at a degree of ignition advance very close to the conditions which produce "knock", an undesirable effect in which the combustion wavefront, instead of progressing smoothly across the combustion chamber, produces a shock wave. This is harmful to the engine and emits the familiar "pinking" sound. The prevent this, manufacturers fit a "knock" sensor, frequently an accelerometer, to the engine and the control system retards the ignition, (or in the case of some turbocharged systems, reduces the boost pressure) until knocking ceases.

Exhaust gas recirculation (EGR) is used in systems where it is necessary to reduce emissions of nitrogen oxides (NO_x) by reducing the combustion temperature. The introduction of non-combustible exhaust gas (which by its nature contains little or no oxygen) at the inlet is controlled by a valve connecting the exhaust with the inlet. The position of this valve is determined, like other requirements, by mapping in the speed/load plane, and controlled by an actuator and position feedback sensor.

In addition to these "real time" measurements, a number of other "*secondary*" measurements will be required, such as temperatures (of inlet air, coolant, etc), barometric pressure, oil level, coolant level, etc.

In countries where methanol is added to fuel in varying proportions, a sensor is required to measure the proportion of methanol. In the USA, for instance, so-called "flexible fuel" vehicles can use fuels ranging from ordinary gasoline to 15% methanol in 85% gasoline. The addition of methanol has a beneficial effect on exhaust emissions, but methanol has significantly different properties to gasoline, and a signal is required for the electronic control unit to make adjustments to ignition timing, fuel supply, and exhaust gas recirculation.

"On-board-diagnostics" (OBD) are now specified for vehicles for California (1993). This requires that certain key systems relating to emissions are continuously monitored and that a warning is given to the driver in the event of malfunction.

17.2.1.2 Gasoline Engine Sensors

In this section we discuss the range of sensors for Gasoline engine control systems. We will consider first the *primary* sensors, which give the basic speed and load signals, then the *feedback* sensors, providing correction signals such as exhaust gas oxygen and knock, and lastly the *secondary* sensors which give correction for ambient conditions and other factors.

Primary Sensors: Engine Speed and Crank Position

Engine speed and crank position are measured by detecting marks or teeth either on the flywheel or on a ring attached to the crankshaft. The position reference may be given either by an additional mark placed on a different diameter and detected by a separate sensor, or by a "missing" mark. The various types of sensor for detecting these marks will now be discussed.

"Variable Reluctance" sensing (Figure 17-3). This is in essence an alternator. It detects "bumps" on the wheel to be sensed by means of a magnetised bar round which a coil is wound. The change in airgap caused by the bump gives rise to a change in flux in the magnetic circuit and thus an emf is generated in the coil. The output waveform will be the differential of the flux (as shown in the figure) so the most accurate way of obtaining a position reading is to detect the zero-crossing of the waveform.

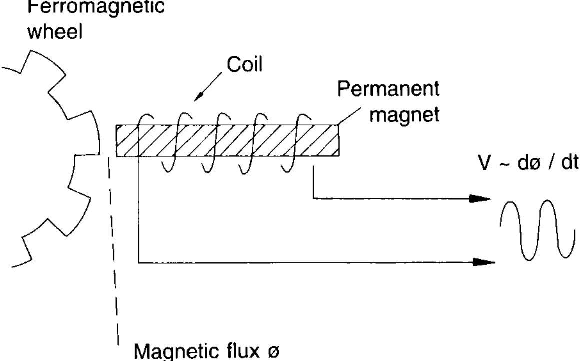


Figure 17-3. Variable reluctance geartooth sensor.

The variable reluctance method is the cheapest and simplest form of sensor: its weakness lies in the fact that its output is related to the *rate of change* of the flux, and therefore its amplitude is proportional to speed. This means that the output may be unacceptably small at low speeds and is of course zero at rest. The last mentioned fact is a disadvantage when setting the timing of the engine during manufacture or servicing, and for these reasons many engine manufacturers specify a device giving output at zero speed.

Hall-effect sensing. When a current is passed through a semiconductor in a direction perpendicular to a magnetic field, an emf is generated in the mutually perpendicular direction. (Figure 17-4). This effect is used in magnetic field sensing in a number of applications and is frequently used in conjunction with a magnet in crank sensing.

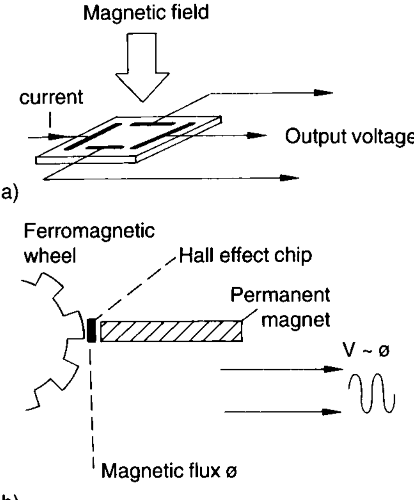


Figure 17-4.
Hall effect wheelspeed sensor, a) Principle of Hall effect; b) Hall effect geartooth sensor.

Magnetoresistive sensing. Magnetoresistance is another semiconductor effect. Here, if a current is passed through a semiconductor in a magnetic field, a resistance change occurs which is related to the magnetic field. This can be used with a magnet in a similar way to a Hall sensor to detect "bumps" in the wheel. The change in resistance in typical values of magnetic field is quite small (perhaps a few percent) so the device is often used in a bridge arrangement which can also be used to provide compensation for changes of resistance with temperature (Figure 17-5).

Optical sensing. Optical sensors have been used in speed and position sensing both for crankshafts and in distributors. An LED light source and photodetector are used to detect slots or teeth in the wheel. Cost and problems of contamination mean that optical methods are seldom to be preferred in original equipment, although multi-bit position encoders are frequently used in test-bed systems where the highest levels of accuracy are required.

Discussion of speed and position sensing. Although speed and position sensing appears to be a relatively simple matter, the problems of required accuracy coupled with the operating environment are typical of those encountered wherever engine controls are considered. The crank positional accuracy needed will usually be of the order of one degree of rotation, and since position detection is in most cases performed by detecting a certain level of output from the sensor the problem of drift over the operating temperature range is not trivial. Furthermore, at an engine speed of 6000 rpm, the crankshaft takes 27 μs to move one degree, so there are also speed considerations to be taken into account in the sensors and the circuitry.

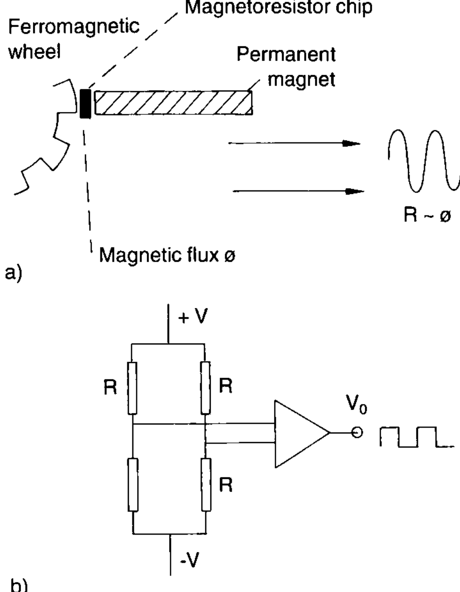


Figure 17-5.
Magnetoresistor. a) Geartooth sensor device; b) bridge arrangement.

As regards choice of technology, variable reluctance devices are cheapest, simplest and most robust (they can operate up to 250 °C). Their main disadvantage is their low output at low speeds (falling to zero at zero speed). This difficulty is avoided by the more expensive and less robust Hall device.

Speed and position sensors are discussed by Wells [3].

Primary Sensors: Manifold Absolute Pressure

Silicon piezorestive sensors. The measurement of manifold pressure for ignition and fuelling systems was one of the first major challenges to be met in the development of the range of automotive sensors. To meet the requirement for around 1% accuracy over a significant temperature range seemed a formidable task in the early systems. The problem was solved using silicon piezoresistive diaphragms with integral compensation, fabricated in large numbers on wafers using techniques developed for the production of integrated circuits in the semiconductor industry. A large number of manufacturers are now able to supply sensors to automotive specification based on this technology, covering pressure ranges for naturally aspirated or turbo-charged installations.

Manifold pressure sensors are usually designed to be mounted in the electronic control unit, and connected to the manifold by a pipe. This provides a more benign environment and reduces the need for temperature compensation. Some units, however, are designed for direct mounting on the manifold, thus being exposed to the engine environment but giving faster response.

The sensing part of these devices is made using silicon micromachining techniques to define the dimensions of a thin diaphragm, and the wafer containing the diaphragm is then attached to a silicon substrate so that a vacuum cavity is formed under the diaphragms. Resistors are

applied at appropriate points on the diaphragm by IC techniques (diffusion or ion implanation), and pressure changes deform the diaphragm and cause changes in value of these resistors, which are usually connected electrically to form a bridge. The vacuum cavity of course provides a reference for absolute pressure measurements (Figure 17-6).

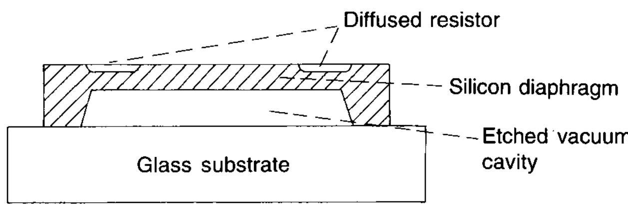


Figure 17-6. Silicon piezoresistive pressure sensor.

Silicon capacitive. As an alternative to piezoresistance sensing, pressure sensors can be formed in silicon by micromachining techniques as described above, but instead of the resistors capacitive electrodes are applied to the diaphragm and the substrate, resulting in a capacitance which varies with pressure.

Conclusions: manifold pressure sensors. The technology of automotive silicon pressure sensors is now fairly mature, the main trends in development being towards greater integration of signal processing and compensation within the device.

Primary Sensors: Air Mass Flow

It has been realized since the earliest days of electronic fuel systems that the measurement of air *mass* flow into the manifold would give the most direct measure of engine load, and considerable effort was devoted to the development of a suitable sensor. Very few methods are available for the direct measurement of mass flow of gases: the majority of conventional flowmeters measure an approximation to volume flow. With a gas, changes in density with temperature and pressure mean that a measure of volume flow is not well correlated with mass flow.

The first successful air flowmeter was marketed by Bosch, and is still to be found on some systems. It comprises a spring loaded flap which is deflected by the air flow, the degree of deflection being detected with a potentiometer.

Heated Element flowmeters. The majority of flowmeters in use today are of the "hot wire" or "hot film" type. In this method, a wire element with an appreciable temperature coefficient of resistance is mounted in the air flow. A current is passed through it which heats it up so that its resistance reaches a set value, (ie its temperature is at a certain level). If an airflow is applied, this tends to cool the wire, so that its resistance will tend to fall, but the electronics are arranged so that the current is increased to keep the resistance at a constant level. The current required to maintain constant resistance is the output of the device. The wire is usually arranged as one arm of a bridge, with a balancing arm which is not heated and whose resistance value provides temperature compensation (Figure 17-7).

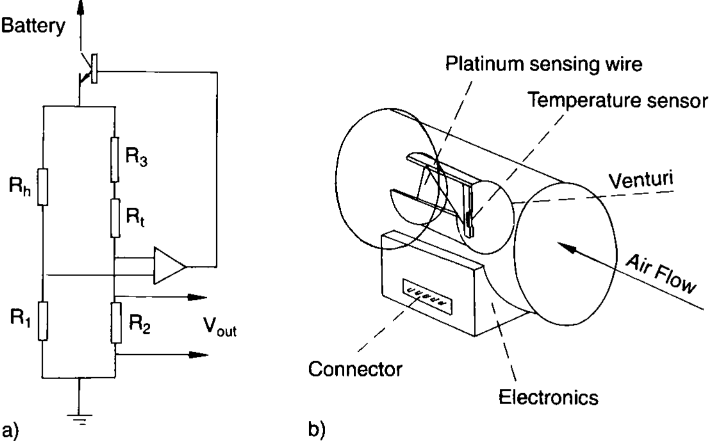


Figure 17-7. Hot-wire airflow sensor. a) Sensor circuit (R_h: hot wire sensor, R_t: temperature sensor, R_1, R_2: balancing resistors); b) sensor system.

In these devices, the heated element is inevitably exposed to intake air, and is therefore prone to contamination, which alters its heat transfer characteristics and thus the calibration of the device. This is minimised in some installations by placing the element in a bypass duct [4], while in other types the heated element can be raised to a high temperature to "burn-off" contamination at some point in the operating cycle, such as engine switch-on or switch-off.

Some airflow sensors based on these principles but made using silicon micromachining methods have also been described [5, 6].

Conclusion: air mass flow sensors. As with pressure sensing, little development work in the field of air flow measurement has been reported in recent years but silicon micromachined devices may make an impact in the future.

Feedback Sensors: Exhaust Gas Oxygen

Exhaust gas oxygen sensing gives an important feedback signal indicating the air/fuel ratio supplied to the combustion process. For emissions control, two alternative strategies are employed by engine manufacturers. The engine may either be run at the *stoichiometric* air/fuel ratio, ie the amount of fuel supplied is exactly enough to consume the oxygen, in which case the amount of oxygen in the exhaust gas is expected to be zero. Alternatively, the manifold and combustion chamber may be carefully designed so that the engine is able to run on extremely weak mixtures, (ie mixtures in which there is a significant excess of air). In these systems there will of course be a proportion of oxygen in the exhaust. The choice of strategy is strongly influenced by the prevailing emissions regulations, and they each pose different problems with regard to exhaust gas oxygen sensing.

Stoichiometric exhaust gas sensors. The majority of exhaust gas sensors in use at present are of this type. Of these, nearly all are a solid-electrolyte cell of zirconia (ZrO_2), which generates

an emf when there is a difference in oxygen concentration between its two sides. Such a device together with its output characteristic is shown in Figure 17-8.

The sensor takes the form of a closed ended tube of zirconia mounted in a casing which allows it to be screwed into the exhaust manifold. The inside of the tube is connected to atmosphere to provide the reference oxygen concentration, while the outside of the tube is exposed to the exhaust gas. The outside will be coated with a precious metal catalyst layer so that the gas in contact with it is brought to chemical equilibrium, and at the exhaust gas temperature the structure acts as a solid electrolytic cell.

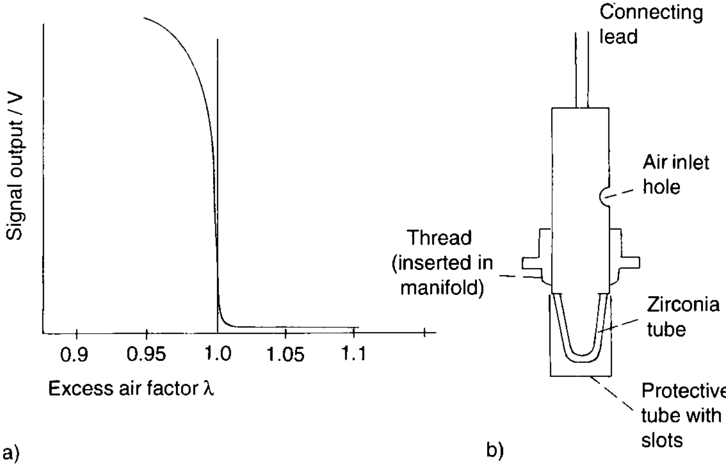


Figure 17-8. Stoichiometric exhaust-gas sensor. a) Characteristic of lambda sensor; b) sensor device.

The output is close to zero when there is an appreciable oxygen concentration on the outside (the exhaust gas side), but rises rapidly to a value of the order of 1 volt as the oxygen concentration approaches zero. Thus the device gives a precise indication as the stoichiometric ratio lambda passes through zero, but gives zero output for all air/fuel ratios significantly weaker than stoichiometric. The device as described only operates at raised temperatures, and needs to be installed at a point where the exhaust gas temperature exceeds 350 °C. This of course still leads to incorrect operation during warm-up, and *heated* sensors are also available which are less dependent on exhaust temperature for proper functioning. They incorporate a electrical heating element mounted inside the zirconia tube.

Sensors for weak (or *lean*) ratios will be described in the next section.

Wide-range oxygen sensors. As we have seen, the straightforward zirconia cell only distinguishes between the presence or absence of oxygen. In engines which are designed for operation on weak (or *lean*) fuel-air mixtures, oxygen is always present in the exhaust, the quantity of oxygen being directly related to the air-fuel ratio, and for these cases a *wide range* sensor (Figure 17-9) is required.

The commonest form of wide range sensor uses a zirconia electrolyte as in a stoichiometric device but it is coated with a *diffusion layer* of porous ceramic. The amount of oxygen diffus-

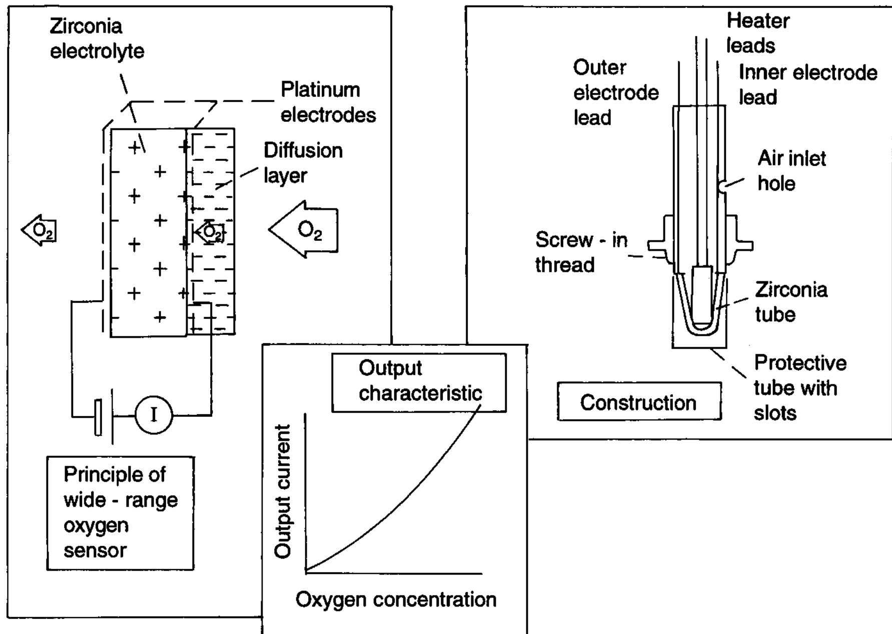


Figure 17-9. Wide-range exhaust-gas oxygen sensor.

ing through this layer to the surface of the electrolyte is related to the oxygen concentration in the exhaust, and when the electrodes are maintained at a fixed potential difference the current passing is nearly linearly related to the oxygen concentration. This type of sensor requires a heater to keep its temperature constant. A good account of the operation of the wide range sensor is given in [7], and the application of such a sensor is described in [8].

Other oxygen sensing methods. Some metal oxide materials behave like semiconductors whose resistivities vary with the concentration of oxygen to which they are exposed. These semiconductor sensors operate by partial oxidation or reduction of the material in the presence or absence of oxygen.

Sensors formed in thick films of titania (TiO_2) are available, and work has been described on other materials such as gallia (Ga_2O_3) and strontium titanate ($SrTiO_3$) [9]. They require a heater and temperature control as their resistance values are temperature sensitive, but are said to give much faster response to changes in oxygen concentration than zirconia sensors enabling air-fuel ratio to be measured for individual engine strokes.

Conclusions: Oxygen sensing. The market for stoichiometric sensors is mature and showing steady growth. Development work on semiconductor type sensors continues, and a desirable goal would be a "linear" sensor which gave an output proportional to oxygen concentration without the complication of pumping. Development work is also active in pumped sensors.

Feedback Sensors: Knock

The characteristic vibrations caused by knock can be observed most clearly by monitoring the combustion chamber pressure, and knock produces a "ringing" signal superimposed on the normally smooth pressure waveform. Pressure sensors are routinely mounted in engines on test bed installations for the study of combustion waveforms and the observation of knocking phenomena, but these are too costly for use as original equipment in vehicles.

Recently, low cost cylinder pressure sensors have been proposed [10] which might make this method of knock sensing feasible, but difficulties of drilling an additional opening through the coolant jacket and into the combustion chamber make it unlikely that this will be the method of choice.

The great majority of current knock sensors are in the form of specially designed piezoelectric or magnetostrictive accelerometers mounted at selected points on the engine block [11]. The signal given by knock has a characteristic frequency for a given engine, usually in the range of 5 kHz to 7 kHz, so it is possible by suitable filtering to detect knock from the clutter of other engine vibrations.

Feedback Sensors: EGR (Exhaust Gas Recirculation) Valve Position

Exhaust gas recirculation is controlled from the main control unit by a valve with a solenoid actuator, and a feedback sensor for its position is required. A plastic film potentiometer sensor may be used, although the environment and life requirements are severe, with the high temperature of the exhaust system. Some systems specify a non-contacting electromagnetic sensor which adds complexity and expense. Non-contacting technologies include differential variable inductors and other forms of variable inductor.

Secondary Sensors: Throttle Position

Throttle opening gives an alternative measure of engine load, since it directly influences manifold pressure or air flow. Systems have been built using throttle position as the primary engine load input, but many manufacturers use it as an auxiliary input. Sometimes a switch may be added to give a signal to indicate when the engine is idling.

The throttle position is measured if required with a rotary plastic film potentiometer. These have been developed to give extremely long life (some have been reported having life of 10^8 operations or 10^9 "dithers" on one spot). Specifications for passenger vehicles would be between one and two orders of magnitude lower than this so the technology of plastic potentiometers for this type of application appears to be sound.

A typical throttle position sensor takes the form of a rotary potentiometer fixed to the throttle spindle.

Secondary Sensors: Temperatures

Temperatures need to be measured at a number of points: inlet air, coolant, lubricant, automatic transmission, etc. Thermistors in specially designed packages for mounting in the required location are supplied, and some manufacturers supply thermistor pellets or chips for

customers to mount in their own packages. Thermistors may be of the non-linear negative temperature coefficient (NTC) type, or the linear positive coefficient (PTC) type, depending on requirements.

Secondary Sensors: Barometric Pressure

Sensors for barometric pressure are absolute pressure devices similar to the manifold pressure sensor. They will be mounted in the control unit and open to the atmosphere.

Secondary Sensors: Methanol-in-Gasoline Sensors

Refractive Index and Dielectric Constant Methods. Methanol has different refractive index and dielectric constant to gasoline, and both quantities vary in a simple way with methanol concentration in a methanol and gasoline mixture. Methanol may be determined by a refractive index measurement, but problems arise with variations in refractive index of different grades of gasoline and with contamination of the optical surfaces. A number of methods based on dielectric constant have been proposed [12, 13, 14].

Inaccuracies can be introduced by the absorption of water, which also has a high dielectric constant, but these are comparatively small, and temperature compensation is also required.

Infra-Red Methods. Some work has also reported on sensors using infra-red absorption [15], but these are in a comparatively early stage of development.

17.2.1.3 Gasoline Engine Sensors: Conclusions and Market Trends

Broadly stated, the technologies of gasoline engine control sensors are mature: advances in engine control such as adaptive control, in which the system automatically searches for the optimum operating conditions, will involve developments in the electronics and in the control philosophies rather than in the individual sensor requirements.

The main areas where sensor developments are taking place are in lean region (ie wide range) oxygen sensors, improved displacement sensors for applications such as EGR, and in alcohol-in-fuel sensors for multi-fuel installations.

Market Trends

At present, gasoline engine controls form by far the largest market sector for automotive sensors (86% of the market in 1990). As regards important specific sensors the worldwide market for speed and position sensors was $1200M in 1992 and will rise to $2000M in 1997; for air mass flow the corresponding figures are $300M and $450M. For exhaust gas oxygen, the market now (1993) is $150M, predominantly zirconia stoichiometric type; this will rise to $240M in 1997.

17.2.2 Diesel (Compression-Ignition) Engines

A Diesel engine differs from a gasoline engine in that the fuel is injected directly into the combustion chamber where the air charge has already been raised to a high pressure by the compression stroke, and is therefore at such a temperature that ignition takes place. The other important difference is that there is no throttle. Air passes direct into the manifold and thence into the combustion chamber; the fuel is injected from a high pressure pump and the power output is controlled by the quantity of fuel per stroke. The pump is the component which performs the principal control functions of the diesel engine, and, as occurred earlier with gasoline engines, it has become necessary to improve the precision of fuel metering and injection timing in order to meet emissions legislation.

17.2.2.1 System Requirements

In order to understand the requirements for an electronic diesel system it is again helpful to consider a traditional mechanical pump system. There are various designs of diesel pump, rotary and in-line, but they all perform the same basic function. Since a diesel engine is inherently unstable (the pump efficiency increases with increasing speed as leakages become less significant), it is necessary to provide some form of speed control in order to prevent the engine speed rising to damaging levels. This is done by incorporating a centrifugal governor, which sets the idle speed and also the maximum speed by controlling the fuel supply. Between these limits, the driver selects the desired speed from the position of the "accelerator" pedal. The fuel control map for a diesel engine is shown in Figure 17-10.

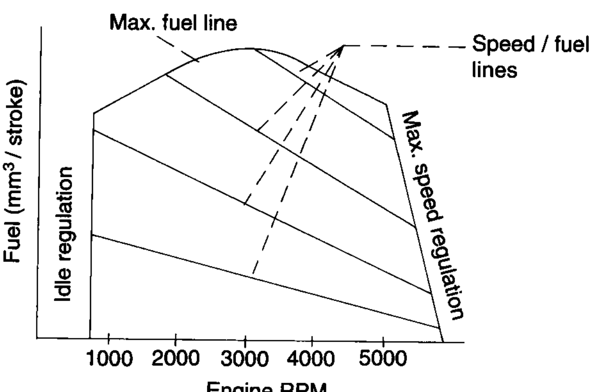


Figure 17-10.
Fuel delivery map for Diesel engine.

To control emissions it is necessary to control the quantity of fuel per stroke and also the timing of injection. In order to provide sufficient accuracy of control for emissions legislation, electronic diesel control systems have recently been introduced. Injection timing affects hydrocarbon emissions, and if the maximum fuel for the operating conditions is exceeded, then smoke is emitted. An electronic system will have an engine speed sensor and crank position reference just as does a gasoline engine, and the fuel/stroke will be controlled either by varying the stroke of the injection pump or alternatively by the use of a "spill valve".

The method of varying the pump stroke will depend on the design of the pump, but there is often a moving part (a control rod in some cases or a sliding conical rotor in others) whose position is measured with a displacement sensor. Injection timing will also be effected by rotating a cam and this position is also measured with a displacement sensor. An example of a diesel control system is described in [16].

As in gasoline engines, NO_x emission is controlled by exhaust gas recirculation, which is controlled in a similar way to gasoline systems. Also, many installations are turbocharged, and the degree of EGR and turbocharging will affect the maximum fuel value so that the control system will require inputs of boost pressure and EGR value.

The *primary sensor* requirements in a diesel are thus engine speed (together with crank position), and driver demand, as determined by pedal position. The ECU determines the fuel supply and timing requirements from these inputs and the pump responds through mechanical acturators. *Feedback sensors* are required for the precise control of these functions. Figure 17-11 shows schematically the operation of a pumped system.

Other systems, at present mainly fitted to heavy trucks, use a "unit injector" principle [17]: here, there is no pump as such, but each injector is supplied by a piston driven directly from the camshaft. Start and finish of injection (and thus quantity of fuel injected and injection timing) is controlled by "spill valves" which divert the fuel from the injector and back into the supply line.

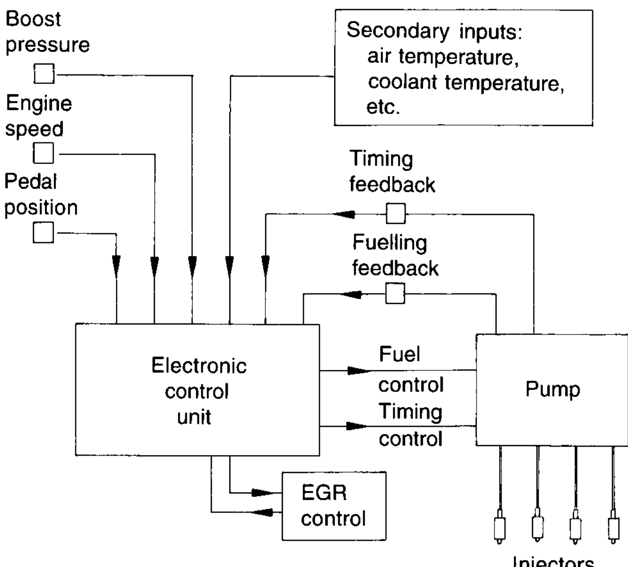


Figure 17-11. Diesel control system.

17.2.2.2 Diesel Engine Sensors

In spite of the differences in the basic system, many of the sensor requirements for diesels are similar to gasoline engines, the main difference being the need to provide controls for the pump which is the heart of the engine control system.

Primary Sensors: Engine Speed and Crank Position Reference

The requirements for engine speed and crank position are similar to those for gasoline engines, and similar solutions are available, the main choice again being between variable reluctance, Hall effect, and magnetoresistor.

Primary Sensors: Pedal Position

Again, requirements are similar to those for gasoline engines – a plastic film potentiometer, rotary or linear, designed to suit the configuration of the pedal.

Primary Sensors: Boost Pressure

A naturally-aspirated diesel engine has no throttle, so that its manifold pressure is approximately atmospheric for all operating conditions. In turbo-charged applications, however, the manifold boost pressure is an important input parameter. A pressure sensor basically similar to a gasoline device will be used, except that the pressure range (absolute pressure) will be from 1 bar (atmospheric) to 2 or maybe 3 bar.

Feedback Sensors: Diesel Pump Sensors

Fuel delivery and injection timing controls. The main feedback sensors in a diesel are devices to measure control functions in the pump. Now a diesel pump is a complex piece of machinery with some extremely fine mechanical tolerances, and these sensors are usually closely integrated with the structure of the pump.

The appropriate fuelling rate and injection timing are determined by the ECU and the pump controls for fuel and timing are operated by actuators. In a rotary pump, such as the Lucas EPIC system, there are two measurements to be made. The first is the axial displacement of a conical rotating member which controls the fuel supply by limiting the stroke of the pump pistons, and the second is the rotational position of a cam which determines the timing.

Both of these measurements are made using linear displacement sensors in the form of variable inductors: some installations use a moving iron core in a coil to vary the inductance, others [18] use a "Short-circuit ring" – a copper or aluminium ring which slides over the core of a coil which varies the inductance of the coil by "shading" the magnetic field.

Other Feedback and Secondary Sensors

No exhaust gas oxygen sensor is required since diesel engines always run with excess oxygen, and similarly knock sensors are not required, since knock is not a problem (at any rate so far as the control system is concerned) in diesels.

EGR sensor requirements are again as for gasoline.

Secondary sensors. The sensor requirements for temperatures, fluid levels, barometric pressure, etc are again similar to gasoline.

17.2.2.3 Discussion and Conclusions: Diesel Engine Sensors

The technology and market for Diesel electronics is much less developed than for gasoline installations, and electronic systems are only at present installed in comparatively few vehicles. Electronic controls are now being fitted and it is expected that electronic systems will be on larger diesel cars by 1996 to meet European emissions legislation [19].

17.3 Sensors for Body Systems and Vehicle Dynamics

Vehicle body and dynamics systems, (braking, steering, suspension, etc) are to some extent interrelated: for example a suspension system could be designed to keep the vehicle level during cornering or braking. Many of the quantities sensed, such as acceleration and displacement, are common to several of the systems, so the sensors will be considered after the main systems and their requirements have been introduced.

17.3.1 Braking System and Traction Control

Electronics are used in braking systems to provide anti-skid braking (ABS), and also to provide Traction Control in which the brakes and modulation of the engine speed are used to prevent spinning of a wheel under conditions of low friction. In more advanced systems full "brake-by-wire" is being evaluated, and in the future, braking systems may be integrated with systems such as "*intelligent cruise control*" and "*collision avoidance*" which will be discussed later in this chapter (Secs. 17.4.1, 17.4.3).

17.3.1.1 System Requirements

The basic ABS anti-skid system uses a speed sensor on each wheel. If during braking one wheel decelerates faster than the others this indicates that that wheel is sliding. Various algorithms are used to control this situation, but the effect is to reduce braking on a wheel on which sliding is detected. Problems can arise if all wheels are sliding simultaneously, in which case there is no reference speed and the system will not function correctly. More advanced systems require an absolute ground speed reference which may be provided by a *True Groundspeed Sensor* (Optical or Microwave) or more simply by integrating the signal from one or more accelerometers. Other inputs might include yaw-rate or lateral acceleration for proportioning of brakes. Fully electronic "brake-by-wire" systems are being considered but will have to overcome considerable market resistance.

Such "brake-by-wire" systems require some means of sensing driver demand: a strain gauge in the brake pedal push rod or a pressure sensor in the master cylinder would provide this. The pressure range requirement for master cylinder sensing would be in the region of 200 bar.

A Traction Control, or "Anti-Spin" system has very similar requirements to ABS: the basic system using speed sensors and more advanced ones having a ground speed or accelerometer input also.

17.3.2 Steering

Electronic power-assisted steering systems are also under development to replace the hydraulic systems in common use in heavier vehicles to-day. The hydraulic systems are dependent on the engine for the belt-driven booster pump, and so lose power in the event of engine stall. Furthermore, the pump is extracting power from the engine continuously, whereas an electronic system will only take electrical power when demanded by the driver, which is an advantage for fuel economy.

17.3.12.1 System Requirements

The electronic steering system will require an input of driver demand: systems have been proposed using steering wheel angular position [20], but systems which only provide assistance when high steering effort is called for (when manoeuvering at low speed, or during violent high-speed manoeuvers) require a steering wheel torque sensor.

There may be applications for accelerometers and yaw-rate sensors in more advanced electronic steering systems: accelerometers to measure lateral acceleration would give feedback of steering forces, and a yaw-rate sensor (ie a gyro) could compare driver demand with actual yaw-rate to correct for deviations caused by changes in road camber or side winds. A good review of these requirements is given by Frere [21].

Some vehicles have four wheel steering in which the rear wheels are steered automatically, the steering angle being a programmed function of front-wheel steering angle and vehicle speed. This can be designed to give improvements in stability at high speeds and manoeuverability at low speeds.

17.3.3 Suspension

Suspension systems come with a variety of degrees of sophistication. *Adaptive* systems vary damping rate in response to various inputs: vertical acceleration, vertical velocity, suspension height, yaw rate, steering wheel angle, etc. In *Active* systems the springs and dampers themselves are replaced by hydaulic actuators, enabling the vehicle to absorb road irregularities and respond in a totally programmed way. Inputs to such systems will be similar to those for a more sophisticated adaptive system.

17.3.3.1 System Requirements

Basic adaptive suspension systems require inputs of vertical acceleration or vertical height as measured by damper strut extension. To give damping control the acceleration signal may be integrated to give velocity, or conversely the height signal may be differentiated. To counteract the tendency of the vehicle to pitch during acceleration or deceleration, and to roll when cornering, more advanced systems have inputs of acceleration and of steering wheel angle [22]. Measurement of pitch and roll may alternatively be made using rotation sensors or gyros [23]. These inputs are used to control damping rate by means of valves incorporated in the damper struts.

In a fully active system [24] a similar range of inputs are used to control hydraulic actuators which take the place of springs and dampers.

17.3.4 Safety and Comfort

At present the most important safety system is the airbag. Airbags are triggered by accelerometers located at several points on the vehicle. Since in a crash the vehicle deforms and does not behave as a rigid body, the location of the sensors and the algorithm used to determine triggering of the airbag need to be carefully chosen.

Passenger comfort systems include air conditioning, with requirements for temperature sensors, and a system is available which detects the ingress of toxic gases (carbon monoxide, nitrogen oxides) into the passenger compartment and switches the air conditioning or heating system to recirculating air.

17.3.5 Sensors for Vehicle Dynamics

17.3.5.1 Wheelspeed Sensors

Present Anti-skid (ABS) braking systems simply incorporate wheelspeed sensors in each wheel hub. Any of the technologies mentioned in the section on engine speed is suitable, but variable reluctance probes are often suitable as the system needs no calibration at zero speed and is not required to operate at very low speeds. Proximity to the brakes means that high temperatures can be encountered, which is another point in favour of the VR probe. Hall effect probes are also frequently used.

17.3.5.2 Accelerometers

Accelerometers are likely to be required for a number of future systems including braking, suspension, steering and airbag collision sensing, and it is considered probable that their use will increase dramatically over the next few years. Their use in braking systems is probably preferable to true groundspeed sensing because of their lower cost, and in addition they can be used in more advanced "brake-by wire" concepts to provide feedback directly on braking retardation, thus enabling the retardation for a given level of braking effort to be independent of vehicle mass (ie loaded or empty for goods vehicles).

An accelerometer consists in essence of a mass, the *seismic mass*, whose position is constrained by a spring. Acceleration produces an apparent force on the mass which is detected by a deformation of the spring. A number of technologies exist for implementing this at low cost, of which silicon micromachining is probably the most promising. In a micromachined silicon device the mass and spring will be formed in a single piece of silicon (Figure 17-12), and the deformation of the spring may be detected either by piezoresistors mounted on the spring member, or by capacitance. Other technologies include thick-film strain gauge, and piezo-electric. A good review of automotive accelerometers and applications is given by MacDonald [25]. Design and manufacture of silicon accelerometers are discussed by Bryzek et al [26].

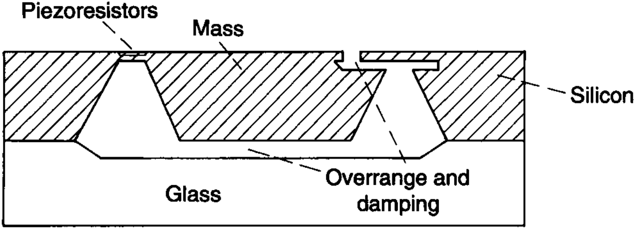


Figure 17-12. Silicon piezoresistive accelerometer.

The acceleration range requirement for braking and traction control would be $+/-1\ g$, and for body vertical motion in suspension systems, $+/-2\ g$. Air bag crash sensors need to detect $50\ g$.

17.3.5.3 Steering Wheel Angle Sensors

The simplest form of electronic steering system would use a steering wheel angle sensor, and optical rotary encoders have been proposed for this [27]. The environment on the steering column is comparatively benign, so that optical solutions are acceptable.

17.3.5.4 Steering Wheel Torque Sensors

Systems with position sensors do not give driver "feel" as well as those which sense driver torque demand, and steering column torque sensors have been described which use magnetostriction [28], and optical techniques [29].

17.3.5.5 True Groundspeed Sensors

Systems under consideration will counter the problem of all-wheel locking by use either of true groundspeed sensors or of accelerometers to provide an absolute reference for vehicle speed.

Work on true groundspeed sensing has included microwave Doppler systems and optical systems. Microwave Doppler systems are already supplied for agricultural equipment. The beam is directed forwards from an antenna mounted on the front of the vehicle and tilted down towards the ground. Energy is scattered back to the antenna from the ground and a Doppler-shifted frequency is observed dependent on the speed.

Optical methods investigated have included correlation and spatial filtering, of which spatial filtering appears the simplest. It uses a spatial filter in the form of a parallel slit grating to convert random signals reflected from an illuminated area of road beneath the vehicle into a frequency signal which is proportional to the vehicle speed.

Examples of both the above are discussed in [30].

17.3.5.6 Fiber Optic Gyros

A fiber optic gyro is based on the Sagnac interferometer principle [31], and a great deal of work has been applied to the development of FO Gyros for guidance systems in aircraft and missiles. A discussion of the principles and practicalities of these devices is beyond the scope of this chapter and it would appear to be an inherently complex and costly device.

Nonetheless, however, a number of proposals have been made for Automotive-grade devices, for example by Washimi et al [32]. This reference describes a device, applied to production vehicles in Japan, of sufficient accuracy for Navigation systems, which would be more than sufficient for steering applications. Application to steering systems is therefore a question of whether or not a system using a gyro could be manufactured at an acceptable cost.

17.3.5.7 Solid State Gyros

A number of workers have recently described solid state gyros which, while they may not have the potential for navigational systems, promise to give a low cost device with sufficient performance for steering and other applications in vehicle dynamics.

Varnham [33] describes a device in the form of a radially vibrating piezoelectric cyclinder. If the cyclinder rotates about its axis, Coriolis forces couple energy from one mode of vibration into a second mode. The energy in this second mode is detected and gives a linear measure of rotation rate.

17.3.5.8 Suspension Height Sensors

Sensors are designed either as linear displacement devices, to be mounted in or on the damper cylinder, or as rotary angle sensors, mounted on a swinging arm. Many use plastic film potentiometers, although some are available which use LVIT or LVDT sensing.

Ultrasonic proximity sensors have also been used for sensing suspension height with respect to the road. These are at present being proposed for headlamp levelling.

17.3.5.9 Crash Sensors for Airbag Deployment

Present-day airbag systems use "ball and tube" acceleration switches [34] in which a ball, restrained by a permanent magnet, breaks free and activates a mechanical switch, which initiates the decision algorithm for firing the airbag.

The technology of solid state accelerometers, discussed earlier in the section, is well suited to this application: the acceleration range required is $+/-50\ g$.

Because the airbag system is one which may well remain unused, possibly for the entire life of the vehicle, but which must be able to operate correctly at any time, accelerometers have been designed which incorporate a "self-test" facility which operates with a heated "bimetal" device which applies a momentary full-scale deflection to the sensor on switch-on [35].

17.3.6 Vehicle Dynamics Sensors: Conclusions and Market Trends

Vehicle dynamics systems appear to be the next important area of growth for electronics and sensors in vehicles. Use of accelerometers in suspension, braking and airbag systems will increase dramatically, and the technology of silicon accelerometers seems ready to fulfil the immediate requirements.

Technologies for other developing sensor requirements are less mature: displacement sensors for suspension height, torque and position sensors for steering, yaw-rate sensors for advanced steering systems, all need further development.

17.3.6.1 Market Trends

In 1990, the proportion of passenger cars having ABS was 10% worldwide; in 2000, the figure will be over 95%. The corresponding figures for electronic power steering are 6% rising to 50%, and for adaptive suspension, 2% rising to 15%.

The world market for accelerometers for vehicle dynamics, including braking, steering, suspension and airbags is expected to rise from $300M in 1992 to around $900M in 1997.

17.4 Sensors for the External Environment

This section will deal with some advanced concepts being developed with the objective of increasing traffic efficiency and safety, many of them under the European *Prometheus* program. It will include systems, all at present in the development stage, which enable a vehicle to "lock on" to a vehicle ahead and keep a constant safe distance from it. Other systems enable the vehicle to follow lane markings on the road surface; or provide warning to the driver of potential collision, and possibly take necessary avoiding action without intervention by the driver [36]. Systems are also being considered to adapt technology developed in night fighter aircraft to enhance the vision of the driver by infra red imaging, and systems for navigation and guidance will also be briefly discussed.

17.4.1 Intelligent Cruise Control

A conventional cruise control system enables the driver to select a vehicle speed, and the vehicle will maintain this speed by automatic adjustment of the throttle. This system is inconvenient on crowded roads, as the speed needs constantly to be adjusted to allow for the movements of other traffic.

The Autonomous Intelligent Cruise Control (AICC) system has rangefinding radar equipment mounted on the vehicle, and the control unit adjusts the vehicle speed by manipulation of the throttle and brakes to maintain constant headway between the vehicle and the one in front.

A major problem with headway measurement using radar has been recognized for many years (Figure 17-13 a): if the vehicle is travelling in a curve, objects which are not in the vehi-

cle's path (such as stationary roadside furniture, or vehicles travelling in adjacent lanes) can appear in the field of view directly ahead of the vehicle. In early work with comparatively low resolution radar and limited signal processing capability, this problem appeared insuperable, but recent work has overcome the problem by using higher frequency radar and up-to-date digital processing.

The system under development at present [37] uses 76 GHz or 94 GHz radar directed in three beams ahead of the vehicle (Figure 17-13 b). The 3 beam approach allows improved tracking on bends, and the signal processing is being developed to improve discrimination when multiple targets are present.

This system is planned for 1998 vehicles, and so is in a fairly advanced stage of development. Its acceptance will depend on extreme reliability, as false alarms will lead to the system not being used and failure to detect a target could of course have grave consequences. The driver has control of the vehicle at all times: operation of brakes or throttle immediately restore the vehicle to full manual control as in a conventional cruise control.

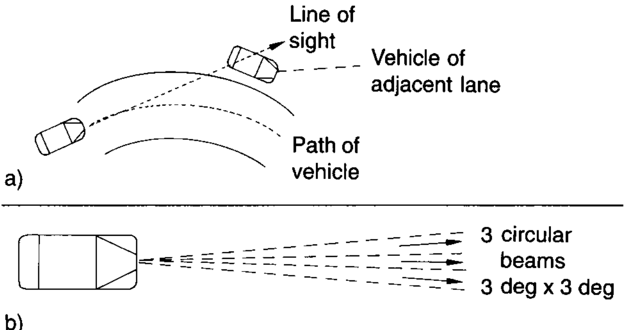


Figure 17-13. Radar system for intelligent cruise control and collision avoidance systems. a) The problem of obstacle detection on a bend; b) 3-beam radar system.

17.4.2 Lane Following Control

Like the previous system, this leaves the driver in full control of the vehicle, but gives a warning, either audibly or possibly a feedback signal through the steering "feel".

A video camera is used to view the road ahead and the image is processed to detect the lane markings, and if the vehicle deviates from the centre of the lane the driver is warned.

17.4.3 Collision Avoidance

This system uses a variety of sensors to warn the driver of imminent collision. Obstacles ahead of the vehicle are detected using a combination of the two techniques described in the previous two sections. Radar is used for range information and a video camera with advanced image processing is used to locate and identify objects. Problems in headway detection arise from the long range required for motorway-speed driving, and the problem of roadside objects on bends is more severe because of the increased range. The problems of reliably assessing

whether or not a situation presents a hazard are highly complex, and will involve advanced image processing and computing techniques, such as the use of neural networks.

For giving warning of side impacts, or when reversing or performing low speed manoeuvres such as parking, shorter range detection is sufficient, and wide-beam microwave or infra-red radar detectors monitor the sides and rear of the vehicle. It is interesting to note that at the present time small microwave devices are fitted to some school buses in the USA to warn the driver of children near the wheels.

17.4.4 Vision Enhancement

Systems are also under development which improve the driver's vision in conditions of poor visibility by superimposing the image from a far infra-red camera onto the view through the windscreen by means of a head-up display. The image is at infinity and arranged to coincide with the driver's view.

This IR camera takes the form of an array of pyroelectric detectors, which is comparatively inexpensive and avoids the need for cooling [38].

17.4.5 Navigation

A number of route finding and planning facilities have been proposed, many of them based on infrastructure such as roadside beacons which transmit information on traffic conditions, road repairs or diversions. Also in this class are satellite-based navigation systems, which again depend on radio or other receivers. These fall outside the scope of the present topic.

Intertial navigational systems, however, are our concern, and inertial-grade accelerometers and fiber optic gyros [39] for vehicles have been the subject of development work, particularly in Japan, where the layout of certain cities makes navigational systems attractive.

17.4.6 Driver Status Monitoring

There is evidence that a considerable proportion of accidents, especially on motorways, are caused by drivers falling asleep. Systems are proposed which will monitor the driver's state of alertness and give a signal if he becomes drowsy.

A system was demonstrated recently which used a video camera to monitor the driver's blink-rate, and combined this with data on steering wheel movements, pulse and breathing rates, etc, to detect the onset of drowsiness.

17.5 Silicon Sensors in Vehicles

Silicon is a material with remarkable electrical and mechanical properties, and its influence on electronic and sensor technologies has been profound. This section will summarise the impact of silicon sensors since the early days of automotive electronics and into the foreseeable

future when further developments, already mentioned in this chapter, will see even greater silicon sensor usage in vehicles.

17.5.1 Silicon Properties used in Sensors

In addition to having electrical properties which have made possible the modern electronics industry, silicon has excellent mechanical properties also. It is very hard, has the elastic modulus of steel and the density of aluminium, and it is almost perfectly elastic up to its fracture point. It is highly resistance to chemical attack, and it is routinely available in single crystal form at extreme levels of purity, in a market which has been built up to meet the demands of the semiconductor industry. It is also as a result of developments made in the semiconductor industry that technologies are available for the manufacture of silicon devices on the wafer scale in very scale quantities and at very low cost.

This availability of large scale manufacturing techniques is only of interest to the sensor industry because it is matched by a unique array of sensing properties. Stress can be measured by the piezoresistance effect, enabling the production of pressure, strain and acceleration sensors; magnetic fields can be sensed using Hall effect or magnetoresistance; light can be detected by the photoelectric effect, and by making use of the mechanical properties of silicon, resonant structures can be made whose frequency properties are influenced by quantities such as strain, temperature or chemical species. It is understandable, therefore, that silicon devices occupy an important place in the range of automotive sensors [40].

17.5.2 Silicon Micromachining

The combined efforts of the microelectronics and sensor industries have resulted in a range of chemical and electrochemical etching techniques which, combined with photolithography, enable silicon to be sculped into microstructures of very high precision [41]. In addition to the capabilities of modern photolithography, the degree of precision is due to the fact that silicon can be etched *anisotropically* by certain etchants: certain crystal planes etch very much more slowly than others, so that very flat surfaces at defined orientations can be formed in single crystal material. Also, *etch-stop layers* can be introduced by diffusion or ion implanation of certain impurities which modify the rate of etching. Another factor is that chemical etching introduces none of the stress that would be produced by mechanical cutting or abrading.

All these factors have combined to enable pressure diaphragms, beams, cantilevers, etc to be formed and produced in large quantities, tens or hundreds to a wafer, and at low cost.

17.5.3 Pressure Sensors

Manifold pressure was the first automotive sensor to be implemented in silicon. Early pressure sensors were based on rubber or metal diaphragms, and were bulky, expensive and inaccurate affairs. Silicon pressure sensors at that time were also expensive, and manufactured in small volume for specialist applications. The breakthrough came when the chemical

micromachining processes were combined with the diffusion, ion implanation, and photolithographic processes of the semiconductor integrated circuit industry, resulting in a device which was available at a cost of a few dollars and included temperature compensation to achieve the accuracy requirements of manifold pressure measurement.

17.5.4 Accelerometers

The success of silicon in providing pressure sensors appears likely to be repeated in the field of accelerometers. With the growth in use of airbags, and increasing interest in braking and suspension systems of greater sophistication, it is predicted that there will be a spectacular increase in the use of accelerometers over the next few years. Silicon (piezoresistive or capacitive) seems at present to be the only technology capable of supplying devices at the required low cost.

17.5.5 Other Silicon Sensors

Other silicon sensors include Hall probes for speed sensing: the European market for Hall probes (automotive) was around 12 million units in 1992, and may rise to 27 million by 1997.

Optical receivers (photodiodes, phototransistors etc, with or without integral amplifiers) are used with optical sensing techniques, such as optical speed and position sensors, the steering torque sensor described earlier, and fibre optic sensors such as the fibre optic gyro.

Other silicon sensors of interest include the airflow sensor mentioned earlier in this chapter.

17.5.6 Automotive Silicon Sensors: Conclusions

From the above it will be clear that the field of automotive electronics has been profoundly influenced by the development of low cost silicon sensors: firstly in pressure sensors, and it appears probable that this process will continue with accelerometers. According to a recent report by Frost and Sullivan [42, 43], the total European automotive market for silicon sensors is predicted to rise from the region of $100M to around $400M by 1997. The most rapid rise predicted is for accelerometers.

17.6 Smart Sensors and System Integration

The most basic form of sensor (Figure 17-14a) is sensitive not only to the intended variable but also to other perturbing effects, of which the most significant is usually temperature. Thus its output signal will be a function of several variables.

17.6.1 Smart Sensors

In an environment such as that of a road vehicle it is not practical to maintain these perturbing variables constant (as we have already seen, temperature can vary over a range of up to 200 °C), so some form of compensation must be provided. In cases where the functions relating the output to the perturbing variable are simple this can be done by analogue means. For instance, a temperature sensing element can be incorporated in the sensor package and used to apply a measure of correction. In the case where there are a number of perturbing variables, and if the transfer functions for these variables are nonlinear, digital techniques must be applied if a "true" signal is to be obtained from the device.

Figure 17-14b shows the concept for a fully "Smart" sensor. The signal to be measured is obtained by a sensor whose output, together with outputs from sensors for the various perturbing variables is passed via an analogue to digital converter to a correction computer. This is programmed with a set of parameters for the individual sensor, which describe the transfer functions for each variable, obtained during initial calibration. In this way a "true" digital signal is obtained at the output. Zabler et al [44] describe the application of these principles to a short-circuit-ring displacement sensor for a diesel pump. A smart sensor for methanol concentration in fuel which uses a microprocessor to correct for temperature variations in a capacitive sensor has also been described [45].

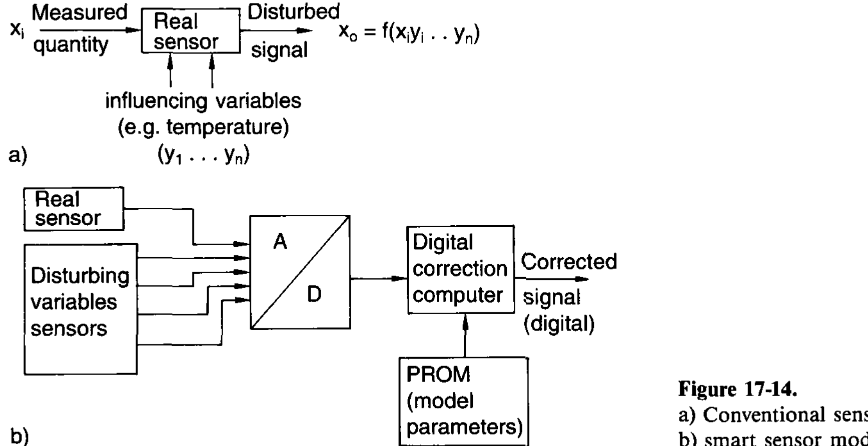


Figure 17-14.
a) Conventional sensor model;
b) smart sensor model.

17.6.2 System Integration

With the increasing complexity of vehicle systems and subsystems, it becomes apparent that the output of a sensor may be required by several subsystems. It thus becomes attractive to integrate the subsystems together in a "bus" architecture, and sensors having digital outputs as described in the previous section can communicate directly with the bus and avoid sensor duplication. In early electronic engine management systems even the ignition and fuelling systems were completely separate, and had separate sets of sensors, whereas it is now taken for granted that the manifold pressure sensor and engine speed and position sensors will be

used in the engine control unit to determine fuelling, ignition, and EGR requirements. Similarly, as mentioned earlier, it is logical to integrate the sensing requirements of braking, steering and suspension, because of the interactions of braking and cornering forces with suspension levelling. An example of integration between suspension and steering is given in [46].

Persuing this line of development, one can think of a total vehicle control system combining all the considerations of this Chapter in an overall system which inputs the drivers demands, combined with various automatic safety controls and guidance systems to control the behaviour of the vehicle in three axes. The article cited in [47] proposes such a system (Figure 17-15).

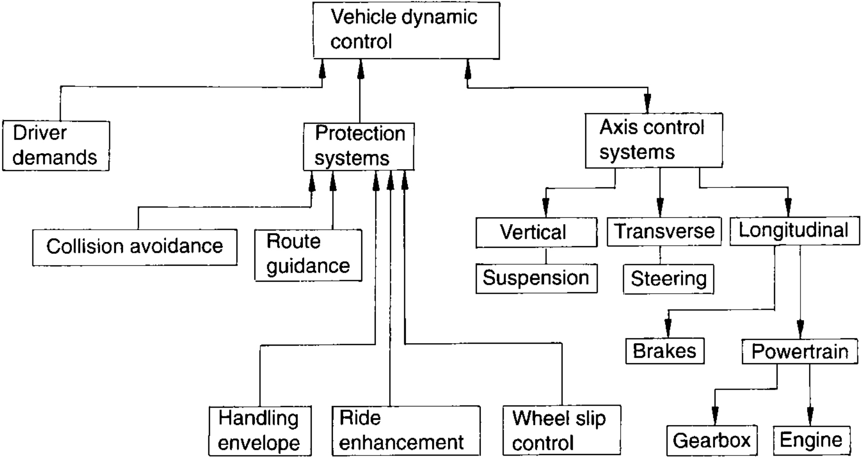


Figure 17-15. Hierarchical vehicle control system.

17.7 Conclusions

We have surveyed the field of Automotive Sensors, system by system, and have seen that while with gasoline engine controls the technology and market are mature (although still growing with technical progress and increasing market penetration), a number of other areas are poised for rapid growth, having only recently become important.

Thus diesel engine controls have only achieved low acceptance at present, but will grow as stricter emissions regulations are introduced; and vehicle dynamics (braking, steering and suspension) having important bearing on safety and comfort, are also at the beginning of a big growth curve. Airbag systems, already well established in the USA are also rapidly gaining acceptance in Europe.

The future of advanced systems using new techniques of computing and image processing has been briefly considered.

We have also examined the unique role of silicon in automotive sensor technology, and seen how it made possible the manifold pressure sensor, and how there is the potential for silicon accelerometers in a number of applications related to vehicle dynamics.

Lastly, smart sensors are likely to be used increasingly, both because of their ability to compensate for errors and changes in measurement conditions, and also because their digital output renders them capable of communicating directly with a bus, which opens up possibilities of system integration and the sharing of sensor data between subsystems.

17.8 References

[1] Quo-Tec Ltd., "Requirements and Market Opportunities for Advanced Sensors in Transport", *DTI Advanced Sensors Programme* (1990).

[2] Prosser, S. J., Moore, J. H., "Advances in Automotive Sensors", in: Grattan, K. T. V. (ed.), *Sensors: Technology, Systems and Applications*; Adam Hilger, 1991.

[3] Wells, R. J., "Non Contacting Sensors for Automotive Applications", *SAE Technical Paper* **880407** (1988).

[4] Takahashi, K., et al., "Hot Wire Air Flow Meter for Engine Control System", *SAE Technical Paper* **900258** (1990).

[5] Stephan, C. H., Zanini, M., "A Micromachined Silicon Mass-Air-Flow Sensor", *SAE Technical Paper* **920473** (1992).

[6] Defauw, P., "Microbridge, a New Technology for Air Flow and Differential Pressure Measurement", *USITT Colloquium*; Southampton: University of Southampton, April 1989.

[7] Kamo, T., et al., "Lean Mixture Sensor", *SAE Technical Paper* **850380** (1985).

[8] Hunt, F., et al., "A Study of a Wide Range Air-Fuel Ratio Sensor for Exhaust Emission Reduction", *SAE Technical Paper* **930233** (1993).

[9] Binder, J., "New Generation of Automotive Sensors to Fulfill the Requirements of Fuel Economy and Emission Control", *Sensors and Actuators* **A 31** (1992) 60-67.

[10] Mock, R., Meixner, H., "A Miniaturized High-Temperature Pressure Sensor for the Combustion Chamber of a Spark Ignition Engine", *Sensors and Actuators* **A 25-27** (1991) 103-106.

[11] Dues, S. M., et al., "Combustion Knock Sensing: Sensor Selection and Application Issues", *SAE Technical Paper* **900488** (1990).

[12] Pinnock, R. A., "A Petrol/Alcohol Ratio Sensor foe Flexible-Fuel Vehicles", *IEE Colloquium on Automotive Sensors,* Digest No. **1992/107** (1992).

[13] Meitzler, A. H., Saloka, G. S., "Two Alternative, Dielectric-Effect, Flexible-Fuel Sensors", *SAE Technical Paper* **920699** (1992).

[14] Binder, J., "New Generation of Automotive Sensors to Fulfill the Requirements of Fuel Economy and Emission Control", *Sensors and Actuators* **A 31** (1992).

[15] Carduner, et al., "Near Infrared Absorption Sensor for In-Vehicle Determination of Automotive Fuel Composition", *SAE Technical Paper* **92098** (1992).

[16] Martin, J., "Improvements in Transient Diesel Performances by Electronic Control of Injection", *I Mech E Paper* **C 382/140** (1989).

[17] Frankl, G., et al., "Emissions Developments of the EUI Systems Towards Improved Performances and Emissions Standards", *I Mech E Paper* **C 389/016** (1992).

[18] Zabler, Heintz, Dietz, Gerlach, "Mechatronic Sensors in Integrated Vehicle Architecture", *Sensors and Actuators* **A 31** (1992) 54-59.

[19] "Engine Control", *Automotive Engineer,* Feb/Mar 1993. 38-41.

[20] Wells, R. F., "Automotive Steering Sensors" *SAE Technical Paper* **900493** (1990).

[21] Frere, P. E. M., "Inertial Sensor Requirements for Automotive Control Systems", *I Mech E Meeting: Technology of Inertial Sensors and Systems,* 17th April 1991.

[22] Hennecke, D., et al., "EDC III – The New Variable Damper System for BMW's Top Models", *SAE Technical Paper* **900662** (1990).

[23] Colinot, J. P., et al., "Gyrometer Application for a Low-Frequency Active Suspension", *Sensors and Actuators* **A 37-38** (1993) 116-120.

[24] Yokoya, Y., et al., "Integrated Control System Between Active Control Suspension and Four Wheel Steering for the 1989 Celica", *SAE Technical Paper* **90178** (1990).
[25] MacDonald, G. A., "A Review of Low Cost Accelerometers for Vehicle Dynamics", *Sensors and Actuators* **A 21-23** (1990) 303-307.
[26] Bryzek, J., Petersen, K., Christel, L., Pourahmadi, F., "New Technologies for Silicon Accelerometers Enable Automotive Applications", *SAE Technical Paper* **920474** (1992).
[27] Smith, D. S., Jackman, P. R., "Optical Sensors for Automotive Applications", *IEE Colloquium on Automotive Sensors,* Digest No. **1992/107** (1992).
[28] Garshelis, I. J., et al., "Development of Non-Contact Torque Transducer for Electric Power Steering Systems", *SAE Technical Paper* **920707** (1992).
[29] Hazelden, R. J., "Optical Torque for Automotive Steering System", *Sensors and Actuators* **A 37-38** (1993) 193-197.
[30] Kidd, et al., "Speed over Ground Measurement", *SAE Technical Paper* **910272** (1991).
[31] See, eg, Culshaw, Dakin, *Optical Fiber Sensors: Systems and Applications*; Artech House, 1989.
[32] Washimi, K., et al., "Fiber Optic Gyroscope for an Automobile Navigation System", *SAE Technical Paper* **920704** (1992).
[33] Varnham, M. P., "Yaw Rate Sensor for Automotive Applications", *IEE Colloquium on Automotive Sensors,* Digest No. **1992/107** (1992).
[34] "Air Bag Impact Sensing", *Automotive Engineering,* April 1993, p. 11-13.
[35] Bryzek, J., Petersen, K., Christel, L., Pourahmadi, F., "New Technologies for Silicon Accelerometers Enable Automotive Applications", *SAE Technical Paper* **920474** (1992).
[36] Tribe, R., "Automotive Applications of Microwave Radar", *IEE Colloquium on Consumer Applications of Radar and Sonar,* Digest No. **1993/124** (1993).
[37] Martin, P., "Autonomous Intelligent Cruise Control Incorporating Automatic Braking", *SAE Technical Paper* **930510** (1993).
[38] Watton, et al., "IR Bolometer Arrays: The Route to Cooled, Affordable Thermal Imaging", *SPIE Vol. 2020 IR Technology* **19** (1993) 379.
[39] Washimi, K., et al., "Fiber Optic Gyroscope for an Automobile Navigation System", *SAE Technical Paper* **920704** (1992).
[40] Grace, R. H., "Semiconductor Sensors and Microstructures in Automotive Applications", *SAE Technical Paper* **910495** (1991).
[41] Zanini, M., Stevenson, P., "Silicon Microstructures: Merging Mechanics with Electronics", *SAE Technical Paper* **920472** (1992).
[42] Bogue, R. W., "Silicon Sensors in Western Europe", in: *Sensors VI: Technology, Systems and Applications*; IOP Publishing, 1993.
[43] McEntee, J., "Start Making Microsensors", *Physics World,* December 1993, p. 33-37.
[44] Zabler, Heintz, Dietz, Gerlach, "Mechatronic Sensors in Integrated Vehicle Architecture", *Sensors and Actuators* **A 31** (1992) 54-59.
[45] Kopera, et al., "Methanol Concentration Smart Sensor", *SAE Technical Paper* **930354** (1993).
[46] Yokoya, Y., et al., "Integrated Control System Between Active Control Suspension and Four Wheel Steering for the 1989 Celica", *SAE Technical Paper* **90178** (1990).
[47] "Electronic System Architecture", *Automotive Engineering,* April 1993, p. 15-17.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

18 Sensors in Manufacturing and Quality Assurance

N. K. Pratt, Technical Services, Northants, UK
Robert Jones, Cambridge Consultants Limited, Cambridge, UK

Contents

18.1 Introduction: The Manufacturing Processes and Quality Assurance 526
18.1.1 Process Research 526
18.1.2 Supplier Control 526
18.1.3 Process Control 526

18.2 Operational Constraints 528
18.2.1 Workpiece Access 528
18.2.2 Operator Skill 529
18.2.3 Speed of Operation 529
18.2.4 Calibration 529
18.2.5 Cost 529

18.3 Sensors, Gauging Systems and Applications 530
18.3.1 Mechanical and Electro-Mechanical Sensors 530
18.3.2 Optical and Opto-Electronic Sensors 532

18.4 Commercial and Technical Trends 537

18.1 Introduction: The Manufacturing Processes and Quality Assurance

Sensing is implicit in "Measurement", measurement contributes "Data" and data enables "Control". The application of sensors* as the key elements of gauging systems in manufacturing and quality assurance (QA) provides the most direct source of intelligence to manufacturing management and enables control of the process to be maintained. Data is collected at various stages of manufacturing and the measurement and acquisition requirements of each stage affect the specification of the gauging system. It is useful, therefore, to identify these stages and thereby to classify the general measurement procedures and diversity of requirement.

18.1.1 Process Research

Process research normally requires the output of the production machinery to be monitored and assessed using sequenced investigative checks and statistical process control (SPC). Typical sensors for this purpose are those for the measurement of dimension, surface roughness, roundness and geometry (ie, form).

18.1.2 Supplier Control

This is an aspect of QA that is of growing importance and requires data to be collected on site using portable and flexible measurement equipment.

18.1.3 Process Control

Process control is the main area of QA and is applied at various stages of manufacturing as summarised below:

In-Process Control. This requires closed loop feed back to the manufacturing process and demands an immediate and conditioned signal from the measurement system. The latter may be built into or onto the machine although at times (and especially where the process adjustments are trend based) the product may be manually or automatically transferred to a receiving fixture for measurement. Such equipment must be rugged, capable of withstanding a harsh and often corrosive environment, and be protected against random electrical interference. The data is usually stored and statistically evaluated for the purposes of production management.

---

* In this context the sensor should be regarded as the primary device by which the dimensions of the component are converted to an analogue signal. The gauging system incorporates either a single sensor or ensemble of sensors and the means by which they are presented mechanically to the work piece and the measurement data acquired and processed.

Immediate Post Process Control. This encompasses various gauging techniques, the form of which depends upon the type of production, ie, either series, or batch/small quantity. In series production dedicated measuring machines or fixtures are used, often measuring multi features and parameters simultaneously. This process is based traditionally on fixed accept/reject gauges capable of being used in relatively harsh environments by non-skilled operators. The wider use of CNC machines in series production has created a growing need for active, adaptive type gauging in which the use of statistical QA techniques is widespread and operation by at least semi-skilled personnel is assumed. Figure 18-1 shows a gauging system suitable for this application in which a fixed array of inductive transducers are used to perform multi-dimensional measurement. For batch and small quantity production, the measurement system is used mainly by skilled operators and needs to be adaptive to cover frequent changes in component size. Gauges ranging from active hand held devices to computer controlled, multi-axis measuring machines which are used to verify the overall conformance of the product. Linkage to SPC and hard copy output is a standard requirement.

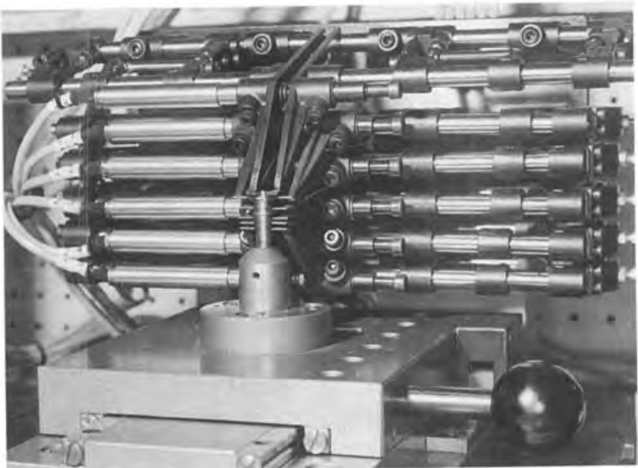


Figure 18-1. Inductive transducers incorporated in a multi-dimension measurement fixture (reproduced by permission of Thomas Mercer Limited).

The Final Inspection Process. The increasing integration of the measurement process with production is tending to eliminate the need for the final inspection process. When required, it uses the same type of equipment as that employed for immediate post-process gauging. Improved, environmentally controlled conditions allow, however, more sensitive measuring instruments to be sited adjacent to the shop floor. Data collection and analysis is also required.

The Calibration Laboratory. This is a specialist area and requires test and measuring equipment with an accuracy of order ten or more times better than the production equipment in use. The measurement machines are specially built for the purpose of gauge calibration and attention is paid to kinematics and sensor design. Close environmental control is essential and the outputs need to be directed to the production of calibration certificates. Computerized

gauge management systems are often connected directly to such systems and carry a record of all the gauges in use, their calibration status and history, plus relevant standards, conversion factors for data manipulation and certification purposes.

18.2 Operational Constraints

In order to assess the suitability of a gauging system for a particular application it is necessary to take into account a number of requirements. These include access to the workpiece, operator skill, speed of operation, and calibration as discussed in Sections 18.2.1 to 18.2.4. In practice speed of operation will be limited, not only by the frequency response of the sensor, but by the gauge adjustment, location and calibration procedures. It is particularly important that the latter constitutes an integral part of the system design and operation.

18.2.1 Workpiece Access

Access, or the ability to penetrate the workpiece in order to reach the point where the measurement is required, limits the type of sensor that can be used for a given application. In many instances, the sensor has to remain outside of the workpiece and a mechanical translation mechanisms used to relay the signal. Internal bore gauging (Figure 18-2) is an example of such an application in which an axial push rod terminating in a cone contacts three radial anvils which engage the bore. (By selecting a 22.5 degree cone angle, the diametral expansion of the anvils is precisely translated into axial movement of the push rod and the bore diameter thereby translated to an external micrometer reading).

Figure 18.2.
A trigger action, three point bore gauge with capacitant digital transducer (reproducted by permission of Bowers Internal Gauge Limited).

In air gauging the sensing medium is actually transmitted to the measuring position, via a system of ducts that terminate in a precisely controlled orifice. Under these conditions the fluid flow or back pressure which defines the object geometry is sensed at the external gauging unit. This simplicity of transmission makes air gauging extremely adaptable.

18.2.2 Operator Skill

Operator skill is another factor which restricts the use to particular types of gauge. In mass or high series production, the repetitive nature of the measurement and the ability to tool up for the semiskilled operatives dictates the much wider use of comparative, rather than adaptive, gauging. The automatic transmission of data into the management or machine control loop networks is necessary when data collection is required since data that is collected manually is recognized to be error prone.

18.2.3 Speed of Operation

The speed of response of a system may directly affect the choice of a sensor, but overall speed of operation is more likely to be affected by the rate at which it is possible to undertake the complete measurement process. This requires:

- adjustment and calibration of the gauge,
- gauge set and reset,
- gauge location during measurement,
- the maintenance of gauge stability during measurement.

Any gauge which is either difficult to calibrate or requires careful application and repeated adjustment will, in practice, fall into disuse.

18.2.4 Calibration

Good practice demands that gauging systems be easily set and calibrated against a master setting piece or sample. Failure to ensure this in the design of the system will lead inevitably to less frequent use and measurement error.

18.2.5 Cost

Having selected the sensor or measurement system which satisfies the required specification, the final arbiter is cost. Here it is important to add additional costs imposed by maintenance, calibration and repair to the initial capital expenditure. For example, fixed gauges are prone to high wear rates and each individual gauge requires recalibration at regular intervals (say four times per year). A more expensive, adaptive gauge, capable of measuring

many sizes and set to a master setting piece, is automatically adjusted for wear each time it is set. Under these conditions the single setting master represents the only calibration cost.

18.3 Sensors, Gauging Systems and Applications

In this section a summary of the main types of sensor used in gauging systems is presented and correlated with the main areas of application.

18.3.1 Mechanical and Electro-Mechanical Sensors

This class of sensor form the basis for the dominant range of metrology systems currently in use. Associated gauging systems and application are summarised in Table 18-1. The mechanical transduction mechanisms form the basis for the ubiquitous hand held and operated gauging systems that are to be found in most workshops. The size of the market is large as can be estimated from figures published by Mitutoyo. The latter are major manufacturers of such systems and in the early 1990s have quoted a total monthly production of callipers, dial indicators, height gauges and bore gauges approaching 250000. Electromechanical gauging systems are used throughout the manufacturing process and two important systems are discussed below.

(a) Contacting Surface Roughness and Geometric Form Measurement Instruments

Surface roughness-instruments are based on a diamond tipped stylus linked to an inductive transducer. The stylus is attached to a pick-up unit which traverses the object. (The pick-up may be fitted with a skid mounted in line with the stylus tip and traversing axis). As the traversing unit draws the skid along the surface to be measured, its large radiused contact face rides over the macro finish or waviness present, while the preceding stylus penetrates the microfinish structure of the surface. In this way the longer wave forms are filtered out mechanically and only the microstructure processed.

The horizontal magnification on the graph recorder is normally low and selectable from ×2, ×5, ×20, ×50, and ×100. The vertical magnification is selectable from ×100 to ×50000, with accuracy discrimination on the graph as fine as 5 nm.

Roundness, straightness and profile measurement instruments are also stylus based. They operate in a similar manner to the surface roughness instruments but at a lower range of vertical magnification. The output is also integrated over the fine surface structure to generate the macroscopic geometric form of the object. Precision references datums enable roundness and straightness accuracies of respectively ± 0.25 μm and 1 μm per 10^2 mm to be achieved. Profilers are operated at lower magnifications to accommodate the wider stylus accuracy to give height discrimination of typically 1 μm.

The design and manufacture of gauging systems with the above performance is demanding and is undertaken by limited numbers of quite small, specialised companies (for example,

Table 18-1. Summary of mechanical and electronic sensors and associated metrology systems.

Transduction mechanism	Sensor Type	Range (mm)[1]	Accuracy (μm)[1]	Gauging system and application	Comment
Mechanical	Micrometer screw	25	10	Internal, external and depth measurement. Bore gauging via mechanical linkage.	The basic sensors are now combined with linear scale position sensors and analogue LCD output displays which simplify measurement and reduce error.
	Vernier	300	20	Caliper, height and depth gauge.	
	Dial indicator	10 to 100 1	10 2	Comparative measurement (due to limited linearity).	
Pneumatic	Air gauge	0.1	2	Measurement of inaccessible points (eg, 1 mm bores).	Air/electronic transducer required at readout for data collection.
Electro-mechanical	Inductive LVDT	±0.5 ±6	0.2[2] 10[3]	Multi dimensional measurement using gauging fixtures designed for specific components (see, for example, Figure 18-1). Surface roughness and form measurement when combined with a stylus[4].	Fixture gauging exploits the high accuracy attainable over short ranges.
	Capacitant	20	0.4	Medium to high precision hand gauging	Optimum configuration of range and accuracy attainable.
	Kinematic electronic switch	see[5]	0.25[5])	Touch trigger probe surface position sensors combined with precision encoded translation stages.	Integral component of CMMs (coordinate measurement machines)

[1] Typical values.
[2] Short range (0.1 mm) accuracy.
[3] Full range accuracy.
[4] See Section 18.3.1.
[5] The range is defined by that of the stage to which the probe is attached and can be well in excess of 1 m for CMMs. This accuracy corresponds to that of the probe and does not take into account errors of the stage positional sensor.

Rank Taylor Hobson in the UK and Tencor in the USA). These and similar companies compete within a relatively small and specialized sector of the gauging system market.

(b) The Tough Trigger Probe and Coordinate Measurement Machines (CMMs)

The touch trigger probe is largely responsible for the rapid expansion in the design and application of the Coordinate Measuring Machine (CMM) that took place in the mid 1960s. CMMs now account for roughly 10% of the total dimensional metrology market which has

an estimated world value of between £250M and £300M. This corresponds to an average of 6000 machines at about £50000 per machine. The CMM measures the three-dimensional coordinates of an object and is based typically upon a large granite surface plate with highly accurate linear glass scale transducers on the three axes. The probe enables the full measurement precision of this arrangement to be achieved and comprises a ball stylus mounted into a probe body by a seating mechanism which incorporates a kinematically designed three dimensional switch. Switch contact is broken when the probe touches the surface and the position of the latter thereby detected with high repeatability.

The CMM has inherently a sub-micron linear measurement accuracy that is derived from its glass scale linear transducers*. The deflections and inertias involved in the measurement process reduce, however, its overall spacial measurement capability. Error plotting, against master calibration devices, can reduce such errors under known conditions and enable the trigger probe's unidirectional repeatability accuracy of typically 0.25 μm for a 50 mm probe to be exploited.

The manufacture and supply of touch trigger probes has been dominated by a relatively small number of companies (for example, Renishaw in the UK and Zeiss in Germany). These have maintained their market share by a combination of technical know-how and strong patent protection. It is interesting to note however that Renishaw primarily manufacture the probe whereas Zeiss manufacture both the probe and a wide range of CMMs.

18.3.2 Optical and Opto-Electronic Sensors

Optical sensors and gauging systems play a central role in metrology and QA and traditionally have achieved their accuracy by a combination of operator skill and inherent magnification (as, for example, in projection microscopy and autocollimation). Manual focus detection by different users does, however, compromise accuracy and extend measurement time. These problems are now overcome by the use of opto-electronic detection which extends measurement range, repeatability and speed. Diverse opto-electronic measurement systems now exist in which either single or arrays of detectors are combined with high precision optics, illumination and opto-mechanics (for example, linear optical scales), and low cost data acquisition. These are summarised in Tables 18-2 and 18-3. (Note that in these tables the x and y are co-ordinates in the object plane and the object profile is the variation in the normal z height as a function of x and y).

Products and markets for the system based on the traditional optical technique of profile projection, focusing and auto-collimation telescopes are relatively mature. Although the incorporation of opto-electronic detection has made them easier to use, their application is still centred primarily on relatively low speed alignment and off-line measurement. The remainder of the techniques have been exploited over the last ten to fifteen years and have led to the emergence of a second generation of optical gauging systems which are reviewed in Table 18-3. These systems have extended significantly the range of measurements that may be carried out

* Such encoders are capable of sub-micron resolution and accuracy over ranges of order a meter and are manufactured by, for example, Heidenheim and Mitutoyo. They may be regarded as an important subsection of the metrology market.

Table 18-2. Summary of noncontact optical and opto-electronic sensors and gauging systems.

Sensor Principle	Range	Accuracy	Gauging system and application
Profile projection	10 to 10^3 mm typically dependant on magnification (×5 to ×100)	±0.05%	Routine measurement of object profiles (nb. depth of object limited to those for which a sharp 2D image may be projected)[1]
Focusing telescope	±1/2 mm	±0.12 mm deviation at 30 m range	Remote measurement of misaligment and geometry of large structures
Auto collimation (ie, coaxial projection and detection of a collimated beam)	1°	1″ (typical)	High precision measurement of angular deviation
Image analysis	10^2 mm at ×10 demagnification[2]	Approximately 1/10 of the focal plane pixel dimension[3]	Dimensional measurement of surface features in x and y
Triangulation	10 mm to 10^3 mm[4]	Surface dependant (typically ±10 µm in 10 mm range)[5]	Medium to low precision z surface profile measurement. Remote measurement of large structures
Scanned confocal microscopy	1 mm	0.01 µm to 0.1 µm	High precision surface finish and profile measurement in z
Interference microscopy	15 µm	0.03 nm	High precision surface finish and profile measurement in z
Measurement of the spatial distribution of light scattered from surfaces	Peak to peak surface heights, 0.1 µm to 100 µm	see footnote 6)	Surface roughness, finish and fault measurement
Laser beam scanning and detection	15 µm to 1 mm 1 mm to 150 mm	±0.2 µm ±30 µm	Automatic edge detection

1) The accuracy noted is only attainable for objects that have clearly projected profiles which limits the range of component that can be measured using this technique.

2) Here it is assumed that a 10^2 mm object is observed at a ×10 demagnification to fill an image plane detector array of nominal dimension 10 × 10 mm.

3) The pixel dimension is typically 10 µm giving a 100 µm accuracy under conditions in 2). Note that processing to give sub-pixel resolution is used routinely to extend the resolution and dynamic range of measurement.

4) The principle of optical triangulation is used in a number of miniature opto-elctronic sensors to give typically 10 mm to 10^2 mm measurement ranges. System adapted from surveying systems measure large objects remotely (10^3 mm measurement range at a distance of 70 m, see also Table 18-3).

5) Image distortion due to the microscopic scattering characteristics of the surface limits fundamentally the accuracy of optical triangulation gauges.

6) The spatial distribution of the scattered light field provides a measure of the surface finish. The accuracy of this measurement is difficult to define since the indirect optical output has to be calibrated against values derived from established stylus contact techniques.

Table 18-3. Survey of commercially available optical gauge systems and a sample of manufacturers.

System	Application	Manufacturers[1)]
Dedicated 2D and 3D non-contact gauging systems based on image analysis *(x, y)* and *z* auto focus	Off-line, semi-automatic, high precision measurement of the profile and features of complex objects	Zeiss, Leitz, Vision Engineering, Nikon
CMM compatible 2D non-contact measurement probes based on image analysis	CMM based, high precision measurement of profile and features of complex objects	Zeiss, Leitz, Renishaw, Wegu
"Surveying" based automated triangulation	Measurement of geometry of large objects (for example planes, ships, civil engineering structures)	Leitz, Leica, IBP
Miniature triangulation sensors	Low to medium precision measurement of surface profile. Often integrated with external mechanical positioning devices and CMMs	Diffracto, TOM, Keyance, Limab, Remplir, Anritsu, Zeiss, Micro Epsilon
Scanned confocal microscope for point by point measurement	High precision and resolution off-line measurement of surface profile and microstructure	UBM
Automated interference microscope for field view measurement	High precision and resolution off-line measurement of surface profile and microstructure	Wyko
Self-contained unit for the illumination of surfaces and measurement of spatial distribution of scattered light	On-line and off-line measurement of quality of surface finish	General: Horstmann Coventry Gauge, Rodenstock Gloss Gauge: Sheen Instruments, Macbeth
Diameter measurement system based on scanned laser beam	On-line measurement of diameter of cylyndrical work pieces	Anritsu, Keyance

1) This is a sample of manufacturers and should not be regarded as an exhaustive list.

optically and in some cases have the speed, precision and robustness required for on-line measurement.

Image analysis based on two-dimensional CCD array detection and processing algorithms that enable edge coordinates to be detected to sub-pixel resolution is used to measure the x, y coordinate of object features. Precision stages and linear scale optical encoders that either position the object (ie, as in automated measurement microscopes) or the sensor (ie, as in miniature vision sensor attached to CMM probes) are key elements of such systems. Precise edge and line measurement is a critical element of the mask manufactured process in the semiconductor industry where the accuracy of the process has been driven to the fundamental limits imposed by diffraction effects. Lower performance systems are used for the routine metrology of a wide range of manufactured components. In a number of these automated and

encoded focus in the z direction is added which enables the system to be used for general 3D metrology.

In triangulation gauging systems the object is illuminated in a known direction by either a point or structured light field (for example, a projected linear or scanned laser beam or a moire pattern). The illuminated region is viewed in a different and known angular direction. The object profile can then be calculated from the viewing and observation vectors and the corresponding image plane coordinate of the illuminated region. Illumination and viewing heads mounted on surveying quality gimballed goniometers can be used at large base distances to measure distant structures as, for example, in the Leica/Leitz ECDS systems. Alternatively a miniature laser diode illumination element and position sensitive image detector may be combined in a single unit to form a miniature triangulation sensor. Lower quality versions of the latter are subject to large errors due to changes in the ambient background and variations in surface reflectivity that can cause signal saturation. These problems can be largely overcome by automatic gain control and the use of a CCD detector array (Figure 18-3 a) thereby enabling complex surface profiles to be measured (Figure 18-3 b). Under these conditions the accuracy is limited fundamentally by the surface structure which causes random image distortion which vary for different types of machined surface (see also note 5 of Table 18-2).

Scanned confocal and interference microscopy are characterized by their high measurement resolution and accuracy in the z direction of measurement. This enables them to be used to measure microscopic features and surface finish (see, for example, Figure 18-4). The confocal sensor measures the height of the surface by correlating the point of surface focus (as indicated by peak power detection at a confocal photodiode) with the encoded position of the microscopes z axis scan. The high precision of the interferometric system derives from the analysis of the 2D interference microscope fringe field using either phase stepping or coherence techniques based on spectrally broad band sources. Both systems are suitable for relatively slow, off-line measurement. It is to be expected that an increased demand for these types of system will result from the increase in the manufacture of microfabricated components.

The spatial distribution of the light reflected from a surface is influenced by the random microstructure of the illuminated region. This phenomenon has been exploited in a number of systems that are used to give a noncontact reading of surface finish. The latter is derived from parameters that are a function of the wavelength, spectral width, direction of illumination and measured spatial distribution of the scattered field. Such systems may be used to provide a rapid check on the quality of machined surfaces. (Note that the readings are not absolute and must usually be calibrated against a specimen of known, good quality, calibrated using a contact sensor such as the Talysurf). An important subset of this sensor is the gloss meter which uses the same principal to measure the quality of very shiny surfaces and play a key role in the QA of sprayed metals and plastic manufactured components.

The gauging system for diameter measurement incorporates a focused laser beam, which is constrained to an optimally collimated scan in its focal plane. The latter also defines the measurement gate through which cylindrical work pieces are passed. The position of beam intersection with the cylinder is determined from the scan encoder and the detected light field and the work piece diameter thereby determined. The precision and ruggedness of this system enables it to be used for high speed, on-line measurement in relatively harsh environments that include extruded cable and optical fibre manufacture.

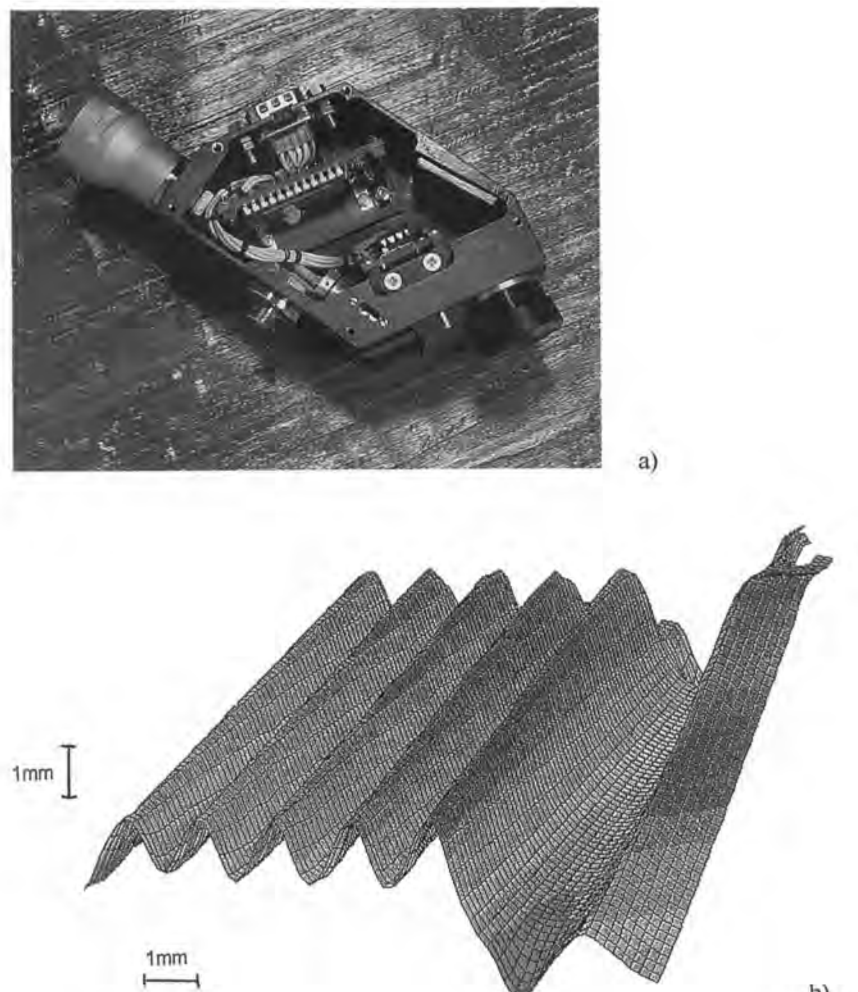


Figure 18-3. a) Miniature triangulation gauge incorporating a CCD array and automatic light level gain control (developed by Cambridge Consultants Limited). b) Profile of a turbine blade fixing measured using the probe in a) in combination with a CMM. (Figures a) and b) reproduced by permission of Rolls Royce.)

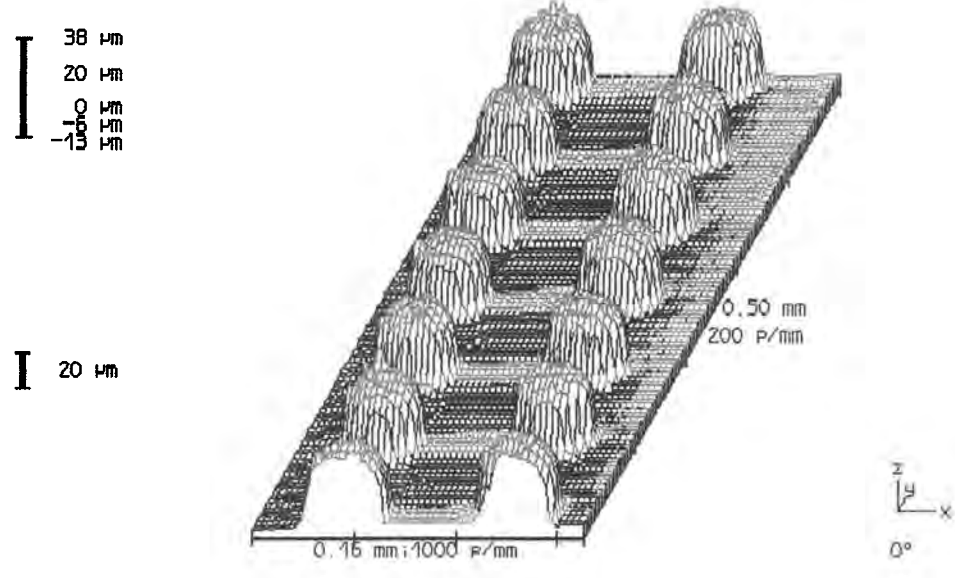


Figure 18-4. Profile of indium pads on silicon measured using a UBM confocal "micro focus" sensor.

18.4 Commercial and Technical Trends

Commercial and technical trends in metrology QA will be driven by the increasingly intense competition that exists in large scale manufacturing. Manufacturers must now ensure that not only does a component conform dimensionally but that it is fit for purpose, ie, the system in which it is incorporated will work to the customer specification. The latter imposes additional requirements on the metrology and QA process in which, for example, the rapid determination of surface microstructure as well as form is required. (Here it is noted that new 'functional characteristics' derived from microstructure measurements may enable faults in the machining process and component degradation to be detected prior to any indication of form error). The manufacturers must, at the same time, maintain and preferably increase production rates whilst minimising waste. Customers now also require information of a manufacturers current and emerging QA procedures and in return will discuss their future requirements and help formulate strategies. Under these conditions metrology and QA ceases to be a hidden procedure but becomes a key element in a manufacturers ability to sell business.

The efficiency (for example, measurement throughput) and data processing capability of existing metrology systems will need to be increased in response to the above requirements. CMMs are, in view of their flexibility, expected to play a central and expanding role with a trend toward improved computer correction of in built errors, cost reduction and the manufacture of small, ultra high precision units for the measurement of microcomponents.

It is also expected that there will be continued growth in the development of "second generation" optical/opto-electronic gauging systems for form, surface finish and feature measurement. This is because their non-contact mode of operation offers the scope for the high speed on-line measurement necessary to close the loop between computer aided manufacture (CAM) and design (CAD). They can thereby help to improve manufacturing quality and efficiency whilst providing additional component data. This introduces a number of technical challenges: the sensor must be compact, robust and capable of operating in harsh environments. Furthermore, noncontact systems will often not function well when the work piece surface is contaminated and methods which either compensate for, or eliminate this problem, must be conceived. These requirements will have a major influence on the design of emerging sensors and gauging systems and if satisfied, should generate new markets. Current developments in opto-electronic device technology such as the low-cost, intelligent digital cameras will help these developments.

Finally it is important to note that there is a strong movement towards the use of microfabrication techniques for the manufacture of very small, high precision components which is likely to stimulate an increased demand for new measurement systems.

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

Index

above-IC microsystems, examples 30–42
above-IC technology 24
absorption sensing, IO, free path 250 f
AC power sensors 58
acceleration sensor system
–, multidimensional 105 f
–, self-testing 107
acceleration sensors
–, capacitive *see* capacitive acceleration sensors
–, IO 232–236
–, LIGA technique 98–108
–, measurement system, design 37 f
–, SAW 148
–, Si technology 519
–, tunnelling detecting 340
–, vehicle 512 f
accelerometers *see* acceleration sensors
acceptor ceramic 189
acoustic plate modes 137
acoustic reflector, filter element 140
acoustic wave devices 135–180
acoustic wave types 136–139
acousto-optical tunable filters 241
active wavelength demodulation systems 241
actuators
–, magnetotronic 42–47
–, optical 271 f
–, smart, avionics 379
A/D-converters 60 ff
–, modulator architecture 71
adaptive control systems, spacecraft 387
advanced electronic design, market role 359 f
aerospace market
–, segments 366 ff
–, trends 404–408
aerospace payloads, special sensor instrumentation 366
aerospace sensors 365–411
–, classification 373
aerospace systems
–, air traffic safety requirements 371
–, constraints and requirements 368
–, design assurance 371
–, environmental loads 368
AFM 339 ff, 343 ff
–, cantilever oscillations 231
agglomeration, nanoparticles 278
air gauging 529
air mass flow, heated element flowmeter 501 f
air pollutants, hazardous, Clean Air Amendments (1990) 465 f
airbag deployment, car crash sensors 514
airbrush coating, QMB transducers 163
aircraft subsystems 374 f
ambient air pollution monitoring 467 f
ambient water pollution monitoring 463 f
analytical microsystem 116
analytics, non-destructive, nanotechnology 19
anisotropic etching 263, 518
annealed proton exchange 227
anti-aliasing filter 61
anti-resonant reflecting optical waveguide 235
apertures, nanometer-sized, for optical sensors 349
apodization, IDT 140
atomic force microscope *see* AFM
atomic layer epitaxy 20
attitude measurement sensors, spacecraft 387–390
automotive sensors 491–523
–, cost 495
–, market 493
–, packaging 495
–, performance requirements 494 f
–, purposes 494
aviation systems *see* aerospace systems
avionics 373
avionics system
–, development trends 378
–, elements, associated sensors 375–378

ball-and-tube acceleration switch 514
band scheme of YSZ 316
barometric pressure sensors, gasoline engine 506
beside-IC technology 24
bioaffinity sensors 322 f
biocompatibility, in vivo sensors, healthcare 446
biocomponents, sensor 13
bioelectronics 331 ff
biomagnetic measurements, healthcare 449
biomolecular recognition, catalytic reactions 323
biophotonics 362
biosensors 13 ff
–, cell function principle 325
–, environment 480 f
–, enzymatic 323
–, healthcare 441–444
–, healthcare applications 443
–, IO 244–248
–, market 443 f
–, membrane function principle 323 f
Bleustein-Gulyaev wave 137
boost pressure measurement, Diesel engine 509
bottom-up strategy, molecular engineering 297
brake-by-wire system, vehicle 510 ff
braking system control 510
bridge oscillator circuit, QMB transducers 170
building engineering 16
bulk materials, nanoparticulate 283–291
bulk micromachining, Si 26

cage compounds 319
–, dynamic 320 f
–, static 318 ff
caisson technique 225
calibration
–, gauging system 529
–, process control 527 f
cantilevers, optically waveguiding 234
capacitance measurements, force detection by cantilever displacement 345
capacitive acceleration sensors
–, beside-IC technology 36
–, housing 38 ff
–, LIGA technique 98–108
–, performance 41 f
–, temperature compensation 99–102
capacitive sensors, readout techniques 103 ff
car crash sensors, airbag deployment 514
CCD array detection 534
ceramic, gas sensor 187
chemical beam epitaxy 20
chemical canaries 481
chemical microanalyzer systems, resumé 131
chemical sensors 307–325
–, capacitive 334
–, IO 244–248
–, materials 334
–, QMB 157
–, SAW devices 152–156
–, techniques 13
chemical vapor deposition 53 f, 65
chemical vapor reaction 276, 288
chemosensors *see* chemical sensors
chirped transducers 140
clusters 21
–, creation 20
–, research 19
CMOS sensors
–, elements 56 ff
–, on-chip circuitry 59–62
CNC machines, gauging system 527
coatings
–, crack-free 277
–, of nanoparticles 278
Coble mechanism, sintering 278
collision avoidance, vehicle 516 f
colloidal techniques 287
colloids
–, stabilization 277
–, surface-stabilized 279
color measurement, process control 428
common land contaminants 468
compact optical disk 251
compression-ignition engine *see* Diesel engine
computer, molecular 299, 327
conduction mechanisms, high-temperature 185–190
conductivity, liquid, SAW 149
constructive and connective techniques, microsystems 7
contacting surface roughness measurement 530 f
control systems, world market size 416 f
conversion, mode 241, 249
converters
–, A/D 60 ff, 71
–, delta-sigma 37
–, thermo- 58, 69
–, TM/TE, acousto-optical 230
cook book, product design 29 ff
coordinate measurement machine 531 f

cost, gauging system 529
coupling, of microelectronics and (bio)chemical structures 297
crack detection, avionics 382
crank position determination, gasoline engine 498 ff
crank position reference, Diesel engine 509
cross-sensitivities
–, elimination 209
–, gas sensors, high-temperature 195–198
current sensor, fiber-optic 265 ff
current transformer, MO 266

data pick-ups, IO 251 f
decimator filter 76
defect models, $SrTiO_3$ 187–190
delay line, filter structure 141
delta-sigma converter, electromechanical 37
demodulation systems, active wavelength 241
de/multiplexer, multi-channel 242 f
deposition, post-processing 53
design assurance, aerospace systems 371
detection
–, crack, avionics 382
–, of electrochemical reactions 302
–, phase, active 227 f
–, –, passive 224–227
–, SAW velocity changes 145
detectors
–, CCD array 534
–, force 344 ff
–, pyroelectric 517
–, smoke 250
–, *see also* sensors; spectrometers
Diesel engine 507–510
Diesel engine sensors 508 ff
–, feedback 508
–, primary 508
Diesel pump sensors 509
Diesel system, electronic 507
difference interferometer
–, polarimetric 236
–, scheme of 246 f
diffractive optical elements 24
direct phase measurement, SAW 146
direct-writing methods 20
displacement sensors
–, IO 223–229
–, nano- and picometer 339–343
donor ceramic 189
drinking water quality measurement 455 f
Driscoll oscillator, QMB transducers 170
driver status monitoring, vehicle 517

Earth observation systems 399 ff
effluent discharge monitoring 457–463
electric field sensors, IO 239 f
electrical operation monitoring, avionics 382
electro-mechanical sensors 530 ff
electrochemical microanalytical system *see* ELMAS
electrochemical potential matching, for molecular recognition 315–318
electrochemical sensors, healthcare 432–435
electrodes
–, electron-conducting 316
–, ion-conducting Zr_2 317
–, micro/macro, information mediators 324
–, mixed conducting perovskites 317
electroforming 9
electromagnetic compatibility 239
electronic design, advanced 359 f
electronic nose 173, 297, 307, 326
electronics
–, bio- 331 ff
–, micro- 263
–, micro- and bio- 297
–, molecular 19, 326–330
–, opto- 18, 532–537
electroplating, LIGA technique 83
ELMAS 117–123
–, concept 118 ff
–, functional model 122 f
emission control, automotive 494
engine speed and load map 496
engine speed measurement
–, Diesel engine 509
–, gasoline engine 498 ff
engine system control, sensors 495–510
environment, industry attitude 486 f
environment protection
–, economic pressure 475
–, public pressure 476
environmental legislation
–, Europe 470 ff
–, trends 474 f
–, UK 472 f
–, USA 473 f
environmental loads, aerospace systems 368
environmental monitoring 17
–, atmosphere 384
–, supply industry 487 f

environmental sensors 451–489
–, biosensors 480 f
–, classification 454 f
–, fiber optic chemical sensors 479 f
–, in-situ monitors 478 f
–, laboratory techniques 476 ff
–, market 483 ff
–, market trends 488
–, portable 479
–, solid state 481 f
–, technologies 476–482
enzyme sensors 323
epitaxy 10
ESA product specification tree 371
ESA Scientific Program, Horizon 2000, payload sensor types 394 ff
etching
–, anisotropic 263
–, post-processing 53
etchstop layers 518
European space expenditures 406
evanescence field sensing *see* evanescent field sensors
evanescent field absorbance sensors *see* evanescent field sensors
evanescent field sensors 123, 245, 248 ff
–, IO 123 ff
–, –, performance 125 f
–, nanosensors 349
–, NIR, sensor system for chemical analysis 123–127
evanescent wave sensors *see* evanescent field sensors
exhaust gas recirculation, gasoline engine 497, 505
external environment sensors, vehicle 515 f

Fabry-Perot resonators 249
–, waveguide 243
Faraday effect 251, 265
feedback sensors
–, Diesel engines 508
–, gasoline engine 497 ff, 502 ff
fiber-optic chemical sensors, environment 479 f
fiber-optic current sensors 265 ff
fiber-optic gyros 237 f
–, avionics 380
–, vehicle 514
fiber-optic nanosensors 349 f
fiber optics 10
fiber technology 261
fibers
–, glass 9
–, polarization-maintaining 261 ff
film-coating techniques 10
films
–, gas-sensitive 185
–, magnetostrictive 150
–, nanoparticulate 283–291
–, thin and thick 183 ff
–, ultrathin 20 f
filter element, acoustic reflector 141
filter structure
–, delay line 141
–, resonator 141
filters
–, acousto-optical tunable 241
–, anti-aliasing 61
–, decimator 76
final inspection, process control 527
flow measurement, process control 417–420
flow-through cell, clip mounted 119
flowmeters
–, coriolis mass 419
–, gas and liquid flow-rate 149
–, heated element 501 f
–, magnetic 418
–, ultrasonic 419
fluorescence methods, IO 247–248
force detection methods, nanosensor 344 ff
force nanosensors 343–346
force sensors, IO 232–236
frequency-modulated microsensors 268
frequency-to-analog sensor 11
fuel control map, Diesel engine 507
fuel delivery control, Diesel engine 509
fuel economy, automotive 494
fugitive volatile organic compounds, measurement, process control 429

galvanic deposition 48
–, FeNi 26
gas and liquid flow-rate sensors, SAW 149
gas electrodes, healthcare 436 ff
–, –, applications 436 f
–, –, market 438
gas-sensitive films 185
gas sensors
–, high-temperature 182–201
–, –, arrays 200 f
–, –, measurement methods 192 f

–, –, selective 198 ff
–, IO 248–251
–, kinetics 190–194
–, QMB 158
gaseous diffusion, gas sensor kinetics 190 f
gaseous emissions monitoring 464–467
gasoline engine 495
gasoline engine sensors 498–506
–, feedback 497 ff
–, market trends 506
–, primary 497 ff
–, secondary 497 ff
gauging system
–, calibration 529
–, cost 529
–, operational constraints 528
–, operator skill 529
–, speed of operation 529
–, triangulation 535
–, workpiece access 528
gels, crystallization behavior 285
geometric form measurement 530 f
glass fibers 9
global navigation satellite systems 393
gloss meter 535
glucose sensor, implantable, specifications 448
grating coupler refractometer 245 f
grating microspectrograph
–, experimental results 93–97
–, layout 89 ff
gravitational wave sensors, tunnelling detecting 342
ground infrastructure market, spacecraft 408
growth, nanoparticles 278
guidance system, RLG-based, spacecraft 387
gyro compass, avionics 380
gyrometers
–, all-integrated optical 238 f
–, fiber-optical 237 f, 514
gyros
–, fiber-optic 237 f
–, ring laser 380, 387
–, solid state 514

Hall-effect sensing, gasoline engine 499
Hall probes, submicron 350 f
health monitoring 16
healthcare sensors 431–450
healthcare sensors, advanced 445
heated element flowmeter, air mass flow 501 f
heating structures, temperature sensors 203
heavy metal sensor, optochemical 114
heterodyne interferometer 229
high-temperature microsensors 181–219
–, *see also* temperature sensors
holographic structurization 20
Horizon 2000, ESA Scientific Program, payload sensor types 394 ff
host-guest coatings, QMB transducers 163
host-guest reactions 153
housing techniques, microsystems 7 f
humidity sensors
–, capacitive 56, 64–67
–, piezoresisistive 57
hybrid sensors, electro-optical 264 f
hydrostatic pressure sensors, SAW 148

IC sensors 52–56
–, additional processing 53–56
–, fabrication technology 52
–, plain IC technology 52
ID tag 141
image analysis 534
imaging film technique, waveguide 225
imaging spectrometer 384
in-IC microsystems, examples 42–47
in-IC technology 24
in-process control 526
in vivo sensors
–, clinical problems 447
–, healthcare 446 ff
–, unusual applications 447
induced fit, molecular cages 320 f
inertial measurement unit 237
inertial navigation units 387
information mediators, micro/macroscopic electrode 324
information storage, nanotechnology 18
information technology, molecular nanostructures 295–336
infrared space observatory 394
injection timing control, Diesel engine 509
integrated circuit *see* IC
integrated optics *see* IO
intelligent cruise control, vehicle 515 f
intelligent devices 6
intelligent sensor system, concept 102–108
interaction forces, intra- and intermolecular 306
interdigital transducers 139
–, chirped 140
–, modulated 140
interference microscopy 535

interferometers
–, difference 236, 246 f
–, heterodyne 229
–, Mach-Zehnder 233 f, 249
–, Michelson 223–228
–, Sagnac 514
–, waveguide 247
interferometric phase detection
–, active 227 f
–, passive 224–227
interferometry, laser 344
internal bore gauging 528
IO 261
IO sensors 221–258
ion selective electrodes *see* ISE
ion sensitive field effect transistors *see* ISFET
ion-sensitive membrane 118
IR sensor, tunnelling detecting 341
IR/NIR sensors
–, healthcare, applications 440 f
–, healthcare, market 441
ISE 118
–, healthcare 433 ff
–, –, applications 433 f
–, –, market 434 f
ISFET, healthcare 435
–, –, applications 435
–, –, market 435
– 118, 323

key-lock interactions *see* lock-and-key principle
knock sensors, gasoline engine 497, 505

Lamb wave 137
lambda probes 204, 210
–, gasoline engine 503
land pollution monitoring 468 f
lane following control, vehicle 516
Langmuir-Blodgett technique 20
–, nanotechnology 19
–, *see also* films
laser beam deflection, force detection by
cantilever displacement 344
laser-Doppler velocimeters 231 f
laser-induced photoacoustic spectroscopy 127
laser interferometry, force detection by cantilever
displacement 344
lateral magnetotransistors 57
lateral nanostructures, creation 20
leaky surface wave 137
level measurement, process control 421 f
life-support sensors, spacecraft 392
LIGA 9
–, capacitive acceleration sensor 98–108
–, microspectrograph, optochemical sensor 115
LIGA microstructrure 81
LIGA process 263, 267
LIGA technique 81–84
–, different level heights 85
light pointer, IO 246
linear focusing grating coupler 228
linear temperature coefficient of delay 141
liquid conductivity sensors, SAW 149
liquid permittivity sensors, SAW 149
lithography 9
–, X-ray 82 f
local intelligence 102
lock-and-key principle 15, 307 f
–, biosensors 322 f
–, molecular cages 318 ff
Love wave 138

Mach-Zehnder interferometer 249
–, pressure sensor 233 f
magnetic field sensors
–, SAW 150
–, tunnelling detecting 341
magnetic flowmeters 418
magnetic nanosensors 350–353
magnetic sensors 57 f, 67 ff
–, auxiliary conditioning circuit 67
magnetic spin sensors 353
magnetic stray field nanosensors 352
magneto-optic *see* MO
magnetoresistive sensing, gasoline engine 499
magnetostrictive films, SAW 150
magnetotransistors, lateral 57
–, *see also* magnetic sensors
magnetotronic actuator
–, in-IC technology 42–47
–, performance 47
malfunction susceptibility, microsystems 6
manifold absolute pressure *see* MAP
manufacturing process control 525–538
MAP
–, silicon capacitive sensors 501
–, silicon piezoresistive sensors 500 f
market
–, aerospace, segments 366 ff
–, –, trends 404–408
–, automotive sensors 493
–, biosensors, healthcare 443 f

–, control systems, size 416 f
–, environmental sensors 483 ff
–, –, trends 488
–, gas electrodes, healthcare 438
–, gasoline engine sensors, trends 506
–, IR/NIR sensors, healthcare 441
–, ISE, healthcare 434 f
–, ISFET, healthcare 435
–, measurement systems, size 416 f
–, microsystems, demands 27
–, neural activity sensors, healthcare 445
–, role, advanced electronic design 359 f
–, –, microfabrication 359 f
–, sensor, review 360 ff
–, surface plasmon resonance, healthcare 440
–, trends, sensor 357–363
–, vehicle dynamics sensors, trends 515
–, waveguides, healthcare 439
mask making, LIGA technique 82
mass measurement, process control 422
measurands, process control 415
measurement *see* sensors
measurement systems, world market size 416 f
mechanical engineering, nanotechnology 19
mechanical sensors 530 ff
–, vehicle control 493
medical engineering, nanotechnology 19
medical sensors 431–450
–, data acquisition 436
membrane combination, LIGA technique 85 f
metal nanoparticles, plasmon resonance 290
metallic microstructures, LIGA technique 87
metallization, QMB transducers 160
meteorological measurements 384
methanol-in-gasoline sensors 506
metric cameras, photogrammetry 384
metrology systems *see* gauging systems
micelle structures 280
Michelson interferometer scheme 223–228
–, double 226
micro-optics 9 f
micro-SQUID sensors 351
microaccelerometer, solid-state 402
microanalyzer
–, chemical 131
–, electrochemical 117–123
–, spectrochemical 126 f
microbalance, quartz 157–175
microbridges, optically waveguiding 234
microelectronics 263
microemulsions, stable 280
microengineering, nanoparticulate systems 290
microfabrication, market role 359 f
microgravity experiments, microaccelerometer 399
micromachining
–, bulk, mono-Si 26
–, Si 518
–, surface, poly-Si 26
micromechanics 8 f
micropumps
–, design 109
–, fabrication 109 ff
–, light-driven 272
–, measurements 111 f
–, thermoplastically molded 109–112
microscopy
–, atomic force *see* AFM
–, interference 535
–, scanning force *see* SFM
–, scanning near field optical 305, 349
–, scanning probe 300–307
–, scanning tunnelling *see* STM
microsensor technology, three-dimensional 79–133
microsensors
–, applications 15 ff
–, frequency-modulated 268
–, high-temperature 181–219
–, interface 60 f
–, optical 259–274
–, –, examples 264–271
–, potentiometric 120 ff
–, signal processing capabilities 60
–, speed/resolution requirements 59
–, thermocouple 348
microspectrograph 267 f
–, LIGA technique 89–97
–, –, optochemical sensor 115
microspectrometer *see* microspectrograph
microstructures
–, LIGA 81
–, metallic, LIGA 87
–, mobile, LIGA 88 f
–, optical, LIGA 86 f
–, production, LIGA technique 86–89
microstructuring technologies, optical 262
microsystems
–, analytical, optochemical 116
–, constructive and connective techniques 7
–, efficient design 29
–, engineering 3 ff

–, housing techniques 7
–, malfunction susceptibility 5
–, market demands 27
–, reliability 6
–, sensors 51–77
–, technologies 24 ff
–, testing 6
microtechnologies 3 ff, 8–15
minicell 333
minimal analog interface 71–76
mixed conductor dopant matching, for molecular recognition 312–315
MO current transformer 266
MO disk 251
mobile microstructures, LIGA technique 88 f
mode conversion
–, acousto-optical, gas sensor readout technique 249
–, SAW-induced 241
modulated transducers 140
modulators
–, A/D conversion, circuit analysis 73–76
–, –, design guidelines 75 f
–, –, input limits 74
–, –, quantization noise 73
–, –, system architecture 71 ff
–, basic structure 63
–, noise-shaping 62
–, phase 243
–, sigma-delta 62, 68 f, 72
moisture meter, process control 427
molding 9
molecular beam epitaxy 20
molecular computer 299, 327
molecular electronics 326–330
–, nanotechnology 19
–, self-repair mechanism 330
molecular engineering 297
molecular orbital matching, for molecular recognition 309–312
molecular recognition 307–325
–, biomimetic approach 322 f
–, dopant matching, mixed conductors 312–315
–, electrochemical potential matching, ion conductors 315–318
–, enzymatic 323
–, induced fit, molecular cages 320 f
–, key-lock interactions, molecular cages 318 ff
–, molecular orbital matching, electronic conductors 309–312
multi-sensor array, optical 268–271
multifunctional materials 8
multilayer, functionalized, molecular electronics 330

Nabarro-Herring mechanism, sintering 278
nanocomposites 278
–, optical properties 290
nanocrystallites, pore size distribution 285 ff
nanolithographic patterning 297
nanometer displacement sensors 339–343
nanoparticles
–, agglomeration 278
–, chemical synthesis 276 ff
–, coated 278
–, dispersions, optically transparent materials 289
–, fabrication 276
–, film-forming 283
–, growth 278
–, metal, plasmon resonance 290
–, microengineering 290
–, semiconductor, quantum confinement 290
–, sintering 284
nanoparticulate, bulk materials 283–291
nanosensors
–, application trends 17
–, evanescent field 349
–, fiber-optic 349 f
–, force 343–346
–, future 337–356
–, heat-flux 346 f
–, magnetic 350
–, magnetic spin 353
–, magnetic stray field 352
–, near-field optical 350
–, optical 349 f
–, thermal 346 ff
nanostructures
–, biomolecular 296
–, chemically synthesized 296
–, identification and manipulation 300–307
–, lateral 20
–, manipulation 306–307
–, molecular, information technology 295–336
–, smart 295–336
nanotechnology 17–21
–, cluster research 20 f
–, information storage 18
–, Langmuir-Blodgett technique 19
–, materials 275–291

–, mechanical engineering 19
–, medical engineering 19
–, molecular electronics 19
–, non-destructive analytics 19
–, optoelectronics 18
–, quantum electronics 18
–, solar energy engineering 19
–, top-down strategy 17
–, vacuum electronics 18
navigation sensors, spacecraft 390
navigation systems
–, avionics 382
–, vehicle 517
near field thermal lens system 127–131
–, miniaturized 130 f
neural activity sensors 444 f
–, applications 445
–, market 445
neural networks
–, process control 424 ff
–, sensor arrays 326
noise rejection filter 61
non-invasive sensing, healthcare 448 f
notifiable substances, effluent discharge 458 f
nucleation process 278

odor measurement, process control 429
odor recognition, biological 298
olfactometry 174
on-board-diagnostics, gasoline engine 498
on-chip signal processing 59
operational transconductance amplifier 68
operator skill, gauging system 529
optical actuators 271 f
optical elements, diffractive 24
optical lithography 20
optical microsensors 259–274
optical microstructures, LIGA technique 86 f
optical sensors 532–537
–, gasoline engine 499
–, healthcare 438–441
–, integrated 221–258
–, micro- 259–274
–, nanometer-sized apertures 349
–, SAW based 151
optical time domain reflectometry 270
optical waveguides 234 f
optically transparent materials, nanoparticulate dispersions 289
optochemical sensor system
–, construction 116
–, heavy metal sensor 114
–, LIGA microspectrometer 115
–, toxic gas sensor 115
optoelectronic sensors 532–537
optoelectronics, nanotechnology 18
optomechanical actuation 272
optothermal deflection, of fluid jets 272
oscillator circuits, QMB transducers 170 f
overlap weighting technique, IDT 140
oversampling interface, sensors 62–71
oxide semiconductor sensors, gasoline engine 504
oxygen sensors
–, motor exhaust 209–212, 502 ff
–, resistive, SrTiO3 187–190
–, resistive, TiO2, Ga2O3, CeO2 190
–, selective 205–209
–, wide-range, gasoline engine 503 f

packaging *see* housing
particle–particle interaction, control 277, 283, 289
passenger comfort system, vehicle 512
pattern recognition 297
payload sensor instrumentation 383
–, spacecraft 392–403
pedal position measurement, Diesel engine 509
pellistor, planar 212–217
permittivity, liquid, SAW 149
perovskite titanates, gas sensor 187
phase detection, interferometric 224–228
phase modulators, electro-optical 243
phenomenologicical sensing 307
photochromy 290
photodiode array, grating spectrograph 96
photogrammetry, metric cameras 384
physical sensors, SAW devices 147–152
Pierce oscillator, QMB transducers 170
plasma etching 233
plasmon resonance
–, metal nanoparticles 290
–, surface 245, 439 f
plastic molding, LIGA technique 83 f
Pockels effect 265
polarization-maintaining fibers 261 ff
portable sensors, environment 479
position sensors, remotely read, SAW 151
post-process control 527
post-processing deposition 53
post-processing etching 53
potentiometric microsensors 120 ff

power sensors
–, AC 58
–, thermoelectric 69 f
power supply rejection ratio 62
precipitation reaction 278
prescribed substances
–, air pollutants 465
–, EC drinking water directive 456
–, land pollutants 468
pressure measurement, process control 420
pressure sensors
–, barometric, gasoline engine 506
–, beside-IC technology, capacitive 31
–, –, piezoresistive 31
–, building blocks 34 f
–, design 31 f
–, housing 33 f
–, hydrostatic, SAW 148
–, IO 232–236
–, linearization scheme 32 f
–, Mach-Zehnder interferometer 233 f
–, MAP 500 f
–, Si technology 518 f
primary sensors
–, Diesel engine 509
–, gasoline engine 498 ff
printing head, magnetographic 42
Priority Pollutant and Target Compound List 459
process control 413–430
–, calibration 527 f
–, emerging trends 424–429
–, final inspection 527
–, immediate post-process 527
–, in-process 526
–, measurement requirements 415–416
process measurement, types 414
process sensing 413–430
–, emerging trends 424–429
product life cycle 29 ff
production process control 17
pseudo-surface wave 137
push-broom scanner 384
pyroelectric detectors, vehicle 517

QMB transducers 157–175
–, acoustic equivalent circuit 164–167
–, coating, active 160–164
–, –, from gas chromatography 162 f
–, –, host-guest 163
–, –, selective biological 163
–, electric equivalent circuit 167 ff
–, metallization 160
–, noise equivalent concentration 172 f
–, oscillator circuits 170 f
–, quartz substrate 159–160
–, signal processing 171 f
quality assurance 358
quality assurance sensors 525–538
–, commercial and technical trends 537 f
quantization noise, A/D conversion 73
quantum confinement, semiconductor nanoparticles 290
quantum dots 18
quantum electronics, nanotechnology 18
quartz microbalance *see* QMB

radio-readable sensors 12
ragging, of in-situ sensors 478
Rayleigh wave 136
–, thin film perturbation 145
recognition
–, molecular *see* molecular recognition
–, pattern 297
red and NIR sensors, healthcare 440 f
Red List, water pollutants 457
reflecting grating scheme, interferometer 228 f
reflectometry, time domain, optical 270
refractive index sensing 245 ff
refractive index shifting element 236
refractometer, grating coupler 245 f
reliability, microsystems 6
remote sensing instruments 383
rendezvous-docking sensor systems, spacecraft 390 f
repirometers, process control 428
resonance, surface plasmon 439 f
resonator, filter structure 141
ring laser gyros
–, avionics 380
–, spacecraft guidance systems 387
rotation sensors, IO 236–239

sacrificial layers, LIGA technique 85
safety system, vehicle 512
sagittal plane 136
Sagnac effect 236
Sagnac interferometer 514
–, *see also* gyros
satellite subsystems, associated sensors 385
SAW 10 ff
–, types 136–139

SAW devices 139–142
SAW ID tags 11
SAW sensing mechanisms 142–145
SAW sensors 4
SAW velocity changes, detection 145
scanned confocal microscopy 535
scanning force microscope *see* SFM
scanning near field optical microscope 305, 349
scanning probe microscope 300–307
scanning tunnelling microscope *see* STM
Scherrer´s equation 287
screen-printing method 183
secondary sensors, Diesel engine 509
security engineering 16
Seebeck effect 69
seismic mass, vehicle accelerometer 512
self-balancing bridge 33
self-contained units 5
self-organization, principles 21
semiconductor nanoparticles, quantum confinement 290
sensor arrays, neural networks 326
sensor flip chip 119
sensor markets, review 360 ff
–, trends 357–363
–, *see also* market
sensor requirements, review 360 ff
sensor systems development 5
sensor technologies, trends 357–363
sensors
–, absorption, IO 250 f
–, acceleration *see* acceleration sensors
–, aerospace 365–411
–, automotive 491–523
–, bio- *see* biosensors
–, bioaffinity 322 f
–, car crash 514
–, chemical 152–156, 307–325
–, –, IO 244–248
–, –, QMB 157
–, –, techniques 13
–, CMOS 56 ff
–, conductivity, liquid 149
–, current, fiber-optic 265 ff
–, definition of 300
–, Diesel engine 507–510
–, displacement, IO 223–229
–, –, nano- and picometer 339–343
–, electric field, IO 239 f
–, electro-optical, hybrid 264 f
–, electrochemical, healthcare 432–435
–, engine knock 505
–, environmental, 451–489
–, –, classes 454 f
–, enzyme 323
–, evanescent field 123, 248, 349
–, flow 418 f, 501 f
–, flow-rate, SAW 149
–, for phenomenological properties 307
–, force, IO 232–236
–, –, nanosensor 343–346
–, gas, high-temperature 182–201
–, –, IO 248–251
–, –, QMB 158
–, gasoline engine 495–506
–, gloss 535
–, glucose, healthcare 448
–, gravitational wave 342
–, gyration *see* gyrometers, gyros
–, healthcare 431–450
–, heavy metal, optochemical sensor 114
–, high-temperature 181–219
–, humidity 56 f
–, IC 52–56
–, in-situ , environment 478 f
–, in vivo, healthcare 446 ff
–, IO 221–258
–, IR 341
–, magnetic 57 f, 350
–, magnetic spin 353
–, magnetic stray field 352
–, magnetoresistive 499
–, manufacturing process control 525–538
–, mechanical and electro-mechanical 530 ff
–, medical 431–450
–, micro- and nanotechnologies 1–26
–, moisture, process control 427
–, neural activity, healthcare 444 f
–, non-invasive, healthcare 448 f
–, optical *see* optical sensors
–, optochemical 116
–, oxygen 187–190
–, –, motor exhaust 209–212, 502 ff
–, oxygen demand, process control 428
–, permittivity, liquid 149
–, physical, SAW 147–152
–, portable, environment 479
–, position, SAW 151
–, potentiometric 120 ff
–, power, AC 58
–, –, thermoelectric 69 f
–, pressure *see* pressure sensors

–, primary 498 ff, 509
–, red and NIR, healthcare 440 f
–, rotation, IO 236–239
–, secondary 509
–, smart 391, 520
–, –, process control 426 f
–, solid-state, environment 481 f
–, space systems 385–403
–, spacecraft and platform 387–392
–, SQUID 351
–, Taguchi 182
–, temperature *see* temperature sensors
–, thermocouple 348
–, toxic gas, optochemical 115
–, velocity, IO 229–232
–, –, laser-Doppler 231 f
–, vibration 268
–, –, IO 229–232
–, virtual, process control 424 ff
–, voltage, SAW 149
–, vortex, process control 419
–, *see also* gauging system
sensors with oversampling interfaces 62–71
Sezawa wave 138
SFM 20
–, for surface characterization 303
shear-horizontal polarization 136
Si-compatibility 333
Si planar pellistor 212–217
–, operating modes 215
signal processing 5 f
signal-to-noise ratio 25
silicon capacitive sensors, MAP 501
silicon micromaching 518
silicon piezoresistive sensors, MAP 500 f
silicon sensors
–, useful Si properties 518
–, vehicle 517 ff
sing-around arrangement, SAW 145 f
sintering, nanoparticulate systems 284
smart actuators, avionics 379
smart sensors 520
–, process control 426 f
smart structure sensors, spacecraft 391
smoke detector, IO 250
sol-gel process 278–282
solar energy engineering, nanotechnology 19
solid state gyros, vehicle 514
solid-state sensors, environment 481 f
sols, dip-coating 285
space observatory, infrared 394
space system sensors 385–403
spacecraft and platform sensors 387–392
spark-ignition engine *see* gasoline engine
spectrochemical microanalyzer 126 f
spectrographs, micro- *see* microspectrographs
spectrometers
–, imaging 384
–, on-chip 242 f
spectroscopy, laser-induced photoacoustic 127
spectrum analyzers, IO 241–244
speed and position sensing, gasoline engine 499
speed of operation, gauging system 529
spun hi bi fiber 262
sputtering process 184
SQUID sensors, microsized 351
stabilized platform sensors, spacecraft 391
star sensors, spacecraft 387
steering system, electronic power-assisted, vehicle 511
steering wheel angle sensors, vehicle 513
steering wheel torque sensors, vehicle 513
STM 19, 339 ff
–, for detection of electrochemical reactions 302
–, for surface characterization 300
–, lithography 20
Stoneley wave 138
structure monitoring, avionics 382
super-stoichiometric oxides, anionic and cationic 186
Superfund
–, common land contaminants 468
–, EPA 474
supramolecular structures *see* cage compounds
surface acoustic wave *see* SAW
surface and interface analysis, tools 308
surface micromachining, Si 26
surface models, gas sensors, high-temperature 195–198
surface-modifying molecules 279 ff
surface plasmon resonance 245
–, healthcare 439 f
–, –, applications 439
–, –, market 440
surface skimming bulk wave 137
surrogate measurements, process control 427 f
suspended solids measurement, process control 423 f
suspension height sensors, vehicle 514
suspension system, vehicle 511
system integration, vehicle 520 f

Taguchi sensors 182
telemetry, tracking and command, spacecraft 393
temperature measurement, process control 422 f
temperature sensors
–, gasoline engine 505 f
–, high- 181–219
–, metal oxide 202–205
–, remotely read, SAW 151
–, tunnelling 348
TE/TM mode conversion, SAW-induced 241
thermal lens effect 128 f
thermoconverter 58, 69
thermocouple sensor, microsized 348
thermometer *see* temperature sensors
thermoplastic molding, micropump 109–112
thick-film/thin-film technology 183 ff
three-dimensional microsensor technology 79–133
three-layer resist system 91 f
throttle position sensors, gasoline engine 505
TM/TE converter, acousto-optical 230
top-down strategy
–, microsystem design 49
–, nanolithography 297
–, nanotechnology 17
touch trigger probe 531 f
toxic gas sensor, optochemical 115
traction control 510
traffic engineering 16
transcutaneous electrodes 437
transducers
–, interdigital *see* interdigital transducers
–, quartz microbalance *see* QMB transducers
–, *see also* sensors
transfer reaction, gas sensor kinetics 191
transistors
–, ion-sensitive field effect *see* ISFET
–, magneto- 57
transthoracic electrical impedance 449
triangulation gauging systems 535
True Groundspeed Sensors, vehicle 510 510, 513
turbidity measurement, process control 423 f

ultra-thin films
–, analysis 21
–, production 20
ultrasonic evaporation, QMB transducers 163
unit injector principle, Diesel engine 508

vacuum electronics, nanotechnology 18
valve position sensors, gasoline engine 505
variable reluctance sensing, gasoline engine 498
vehicle body control 510
vehicle control systems
–, electronical 493
–, mechanical 493
vehicle dynamics control 510
vehicle dynamics sensors, market trends 515
vehicle handling, safety and performance 494
velocimeters
–, IO 229–232
–, laser-Doppler 231 f
Verdet constant 265
vibration sensors
–, bridge structure 268
–, IO 229–232
virtual sensors, process control 424 ff
vision enhancement, vehicle 517
voltage sensors, SAW 149
volume diffusion, gas sensor kinetics 191
vortex meter, process control 419

wafers, market demands 27
waste water monitoring 456 f
water intake monitoring 463
wave types, acoustic 136–139
waveguide grating 243
waveguide imaging film technique 225
waveguide interferometers 247
waveguides 10
–, anti-resonant reflecting optical 235
–, fabrication technologies 222
–, healthcare 438 f
–, –, applications 438 f
–, –, market 439
–, optical performance 92 f
–, planar 87, 228
wavelength filters, tunable 241 f
wavelength-selective devices, IO 241–244
weighting techniques, IDT 140
wet chemical analyzers, process control 429
wet chemistry, nanoparticulate systems, synthesis 276 ff
wheelspeed sensors, vehicle 512
wide-range O_2 sensors, gasoline engine 503 f
withdrawal weighting technique, IDT 140
workpiece access, gauging system 528

X-ray lithography, LIGA technique 82 f

Ytterbium-stabilized ZrO_2, band scheme 316

zeta potential, colloids 278

Sensors A Comprehensive Survey
Edited by W Gopel, J. Hesse, J. N. Zemel
©VCH Verlagsgesellschaft mbH,1995

List of Symbols and Abbreviations

Symbol	Designation	Chapter
a	constant (unspecified)	
	acceleration	
A	constant (unspecified)	
	area of capacitor	4
	area covered by electrodes	5
	(signal, vibration) amplitude	5,6,7
A_0	fundamental antisymmetric mode	5
$A^{\prime}$	singly ionized acceptor-type dopant	6
A_{eff}	effective catalyst surface area	6
b	constant	9
	width of stationary electrode	4
B	bandwidth	3
	magnetic induction density	3
c	phase velocity of acoustic wave (ie, sound velocity)	5
	spring constant	11
	vacuum light speed	8
C	capacitance	
	coordinate dependant constant	5
	gas concentration	6
C'	capacitance per unit length	5
C_{66}	elastic modulus	5
C_Q	dynamic capacity	5
d	distance	7
	grain diameter	9
	separation	7
	thickness	5
D	diffusion constant	9
	output code	3
$D^{\bullet}$	singly ionized donor-type dopant	6
E	coefficient of elasticity	5
	relative sensitivity (of IC)	4
	strain energy per unit volume	9
e, e′	electron	6
E_0	full-scale signal power	3
E_A	bulk acceptor energy	10
E_A	activation energy	6
E_{act}	activation energy of desorption	10

Symbol	Designation	Chapter
E_C	conductance band edge	10
ΔE	electrochemical potential difference	10
E_D	bulk donor energy	10
E_F	Fermi level	10
E_i	activation energy of species *i*	6
E_{kin}	kinetic energy	10
E_{QI}	quantization noise power (primary)	3
Err_{lin}	relative deviation from linearity	4
E_{ss}^{eff}	effective surface state	10
E_V	valence band edge	10
E_{vac}	vacuum energy level	10
f	decision function	6
	frequency	
f_0	resonance frequency	5
f_{mod}	upper frequency for temperature modulation	6
F	Faraday constant	10
	force	11
	frequency	
F_S	sampling frequency	3
F_u	bond breaking force	10
g	gain	3
g_i	parameter set	6
G	Gibbs energy	9
	transconductance	3
G_c	surface crack forming energy	9
h	hole (ie, missing electron)	6
h'	film height	5
ΔH	reaction enthalpy	6
I	current	
I	intensity of probe laser beam	4
dI/ds	STM signal	10
dI/dV	tunneling conductance	10
Im	imaginary part of a complex number (operator)	5
j	square root of -1	5
J_n	Bessel function	7
k	Boltzmann constant	2,6,9
	reaction rate	6
k^2	electroacoustic coupling factor (air)	5
k_L^2	electroacoustic coupling factor (liquid)	5
k_T^2	piezoelectric coupling coefficient	5
K	coefficient of compression	5
K_b	partition coefficient	10
K_b^0	normalized partition coefficient	10
K_i	mass action constant of species *i*	6
K_i	frequency independant constant ($i = x,y,z$)	5
$K_{x,y,z}$	coordinate dependant constants	5

Symbol	Designation	Chapter
l	particle diameter	10
	width of seismic mass	4
L	fiber line length	8
	inductance	
	length of acoustic waveguide	5
	transducer spacing	5
L'	inductance per unit length	5
L_D	Debye length	10
L_Q	dynamic inductance	5
m	inverse slope of curve	6
	mass	
	number of additional gas components	6
M	oversampling ratio	3
	seismic mass	4
n	charge carrier concentration	10
	concentration of electrons	6
	integer number	5
	mode integer	5
	number of gas components	6
	refractive index	4,7,8
	slope of curve	6
$n_1(f)$	power spectrum (primary quantization)	3
Δn	refractive index change	7
N	intensity	10
	number of electrode fingers	5
	number of turns	8
N_i	number of particles of species i	6
$O_{2,S}$	physisorbed oxygen molecule	6
O_h	octahedral interstitial site	10
O_O	oxygen atom at regular lattice site	6
O_S	chemisorbed oxygen atom	6
p	concentration of holes	6
	pressure	4,9
	spatial periodicity of electrode fingers	5
p_B	in-band noise	3
pCO_2	carbon dioxide partial pressure	15
p_{O_2}	oxygen partial pressure	10
pO_2	oxygen partial pressure	14
P	average power flow per unit width	5
	pressure	6
	output signal power	7
$P(O_2)$	oxygen partial pressure	6
P_1	in-band power (primary quantization)	3
P_{chem}	chemical power	6
P_{el}	electrical heating power	6
q	charge	10
Q	crystal quality factor	5
	reaction heat	10
Q_{st}	isosteric heat of adsorption	10

Symbol	Designation	Chapter
r	particle radius	9,10
R	gas constant	6
	resistance	6,17
	resistive loss	5
	sample number at averaging moment	3
Re	real part of a complex number (operator)	5
RF_S	oversampling rate	3
R_T	resistance at given temperature	6
s	distance probe tip / object (gap spacing)	10,11
S	magnetic sensitivity	3
	signal of thermal lens	4
	sintering rate	9
S_0	sticking coefficient	10
S_0	fundamental symmetric mode	5
SNR	signal-to-noise ratio	2,3
S_z^y	sensitivity of y to a change in z	5
t	layer thickness	9
	time	7,17
t_{90}	sensor response time	6
Δt	delay time (pulse length)	8
t_d	tetrahedral interstitial site	10
T	temperature (K or °C)	
	transducer time length	5
$T_Q(f)$	transfer function in frequency	3
U	voltage	4,10,11
$\dot{u}_i$	unperturbed particle velocity component	5
V	Verdet constant	8
	voltage	
V_O	oxygen lattice vacancy	6
$V_{O,S}$	oxygen vacancy at crystal surface	6
$V_O^{\bullet}$	singly ionized oxygen vacancy	6
$V_O^{\bullet\bullet}$	doubly ionized oxygen vacancy	6
V_O^{2+}	electron and oxygen vacancy	10
V_O^x	neutral oxygen vacancy	6
V_{Sr}	strontium lattice vacancy	6
V_{Sr}'	singly ionized strontium vacancy	6
V_{Sr}''	doubly ionized strontium vacancy	6
V_{Sr}^x	neutral strontium vacancy	6
V_V	defect volume (voids)	9
x	direction in space	6
	measurand	5
Δx	displacement	7
δx	position resolution	8
X^{gas}	molecule in gas-phase	10
X_{IN}	input measurand	3
X^{phys}	physisorbed molecule	10

Symbol	Designation	Chapter
y	direction in space	6
	gap width (capacitor)	4
z	direction in space	6
	distance between sample and pinhole in TL	4
Δz	deflection	11
Z	acoustic impedance	5
α	attenuation coefficient	5
	bandwidth ratio component:system	2
	coefficient of thermal expansion	4
	linear temperature coefficient of resistance	6
	rotation angle	8
	polarizability	10
β	coefficient of material transport	6
	quadratic temperature coefficient of resistance	6
	wavenumber	5
	wave propagation constant	7
γ	propagation constant	5
δ	cantilever deflection	7
	grain boundary thickness	9
	partial charge	10
Δ	comparator step size	3
ε	permittivity	4,10
ε_0	dielectric constant (ie, free-space permittivity)	4
$\varepsilon_Q\varepsilon_0$	electric permittivity of quartz	
ζ	degradation factor	3
η	viscosity	5
η'	dynamic viscosity	5
Θ	cut angle	5
κ	decay rate	11
κ'	volume elasticity of liquid	5
λ	exhaust gas excess air factor	6,17
	wavelength	5,7,11
$\Delta\lambda/\lambda$	spectral resolution	7
λ',μ'	Lamé constants	5
μ	dipole moment	10
μ'	shear stiffness	5
ν	frequency	7,8
	pulse repetition rate	8
ν_{gr}	group velocity	5
ν_{ph}	phase velocity	5
ξ	sensitivity coefficient	3
ρ	density	4
	mass density	5
ρ'	mass density	5
σ	electrical conductivity	5,6,10
	triaxial tensile stress	9
$\Delta\sigma$	pressure difference	9

Symbol	Designation	Chapter
σ_m	standard deviation	4
τ	delay time	5
τ'	shear relaxation time	5
ϕ	clock phase	3
	magnetic flux	17
	phase response	6
	phase angle	5
	work function	10
ϕ_0	elementary magnetic flux quantum	11
ϕ_{SB}	Schottky barrier height	10
Φ	cut angle	5
	phase response	5
χ	compressibility	5
	electron affinity	10
Ψ	cut angle	5
ω	angular frequency	5,11

Abbreviation	Explanation
A/D	analog/digital
ABB	Asea Brown Boveri
ABS	"Anti-Blockier-System" (anti-skid braking system)
ac	alternating current
AC	alternating current
AC	adenylate cyclase
ACS	American Chemical Society
ADC	analog to digital converter
ADF	automatic direction finder
ADI	attitude direction indicator
AES	Auger electron spectroscopy
AFM	atomic force microscopy
AICC	autonomous intelligent cruise control
AIDS	acquired immune deficiency syndrome
ALE	atomic layer epitaxy
AMF	air mass flowrate
ANSI	American National Standards Institute
AOCS	attitude and orbit control system
AOTF	acousto-optical tunable filter
APD	avalanche photodiode
APE	annealed proton exchange
APM	acoustic plate mode
ARROW	anti-resonant reflecting optical waveguide
ASI	Artificial Sensing Instruments
ASIC	application specific integrated circuits
ASME	American Society of Mechanical Engineers

Abbreviation	Explanation
AT	quartz cutting direction
ATC	air traffic control
ATP	adenosine triphosphate
AWDS	active wavelength demodulation system
B	transistor base
BATNEEC	Best Available Technique Not Entailing Excessive Cost
BGW	Bleustein-Gulyaev wave
BHC	1,2,3,4,5-hexachlorocyclohexane
BIA	biospecific interaction analysis
BiCMOS	bipolar complementary metal oxide semiconductor
BOD	biological oxygen demand
BOD5	biological oxygen demand over a five day period
BPEO	Best Practical Environmental Option
BSA	bovine serum albumin
C	cantilever
C	transistor collector
CAA	Civil Aviation Authority
CAD	computer aided design
CAM	computer aided manufacture
cAMP	cyclic adenosine monophosphate
CBE	chemical beam epitaxy
CCD	charge coupled device
CD	compact disk
CDS	coronal diagnostic spectrometer
CELIAS	charge, element, and isotope analysis system
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CFC	chlorofluorocarbon
CHC	chlorinated hydrocarbon
CHEMFET	chemically sensitive field effect transistor
CIS	cluster ion spectroscopy
CMM	coordinate measurement machine
CMOS	complementary metal oxide semiconductor
CMV	cytomegalovirus
CNC	cyclic nucleotide-gated ion channel
CNC	computer numerical control
CNR	carrier-to-noise ratio
COD	chemical oxygen demand
COHb	carboxyhemoglobin
COSHH	control of substances hazardous to health
COSTEP	comprehensive suprathermal and energetic particle analyzer
CRC	Chemical Rubber Company
CSEM	"Centre Suisse d´Electronique et de Microtechnique" (Swiss center of electronics and microtechnics)
CSVAL	chirasil-val (siloxane with valine groups)
CVD	chemical vapor deposition
CVR	chemical vapor reaction
D/A	digital/analog
dA	deoxyadenylic acid

Abbreviation	Explanation
DASA	"Deutsche Aerospace AG"
dc	direct current
DC	direct current
DDD	1,1-bis(4-chlorophenyl)-2,2-dichloroethane
DDT	1,1-bis(4-chlorophenyl)-2,2,2-trichloroethane
DEHP	bis(2-ethylhexyl)phthalate
DFB	defined distribution feedback
DGAC	Direction Générale de l'Aviation Civile (civil aviation authority)
DGG	"Deutsche Glastechnische Gesellschaft" (German society of glass technology)
DIAL	differential absorption lidar
DIMP	diisopropyl methylphosphonate
DL	delay line
DMMP	dimethyl methylphosphonate
DO	dissolved oxygen
DOE	Department of Energy
DOE	diffractive optical element
DSA	density spectral array
dT	deoxythymidylic acid
DTI	Department of Trade and Industry
DWI	Drinking Water Inspectorate
DWP	digital wave processor
E	transistor emitter
EA	Environment Agency
EAA	ethylacetoacetate
ECS	Electrochemical Society
ECU	European currency unit
EDI	electron drift instrument
EDP	ethylene-diamine-pyrocatechol
EDX	energy dispersive X-ray analysis
EEF	Engineering Employers Federation
EEG	electroencephalography
EELS	electron energy loss spectroscopy
EEPROM	erasable electrically programmable read-only-memory
EFAS	evanescent field absorbance sensor
EFW	electric fields and waves
EGR	exhaust gas recirculation
EIT	extreme-UV telescope
ELF	extremely low frequency
ELL	ellipsometry
ELMAS	electrochemical microanalytical system
EMC	electromagnetic compatibility
emf	electromotive force
EMI	electromagnetic interference
EMP	electromagnetic pulse
ENDS	Environmental Data Services Ltd.
ENFET	enzyme-mediated field effect transistor
ENVISAT	environment satellite
EPA	Environmental Protection Act

Abbreviation	Explanation
EPIC	European photon imaging camera
ERNE	energetic particle analyzer
ESA	European Space Agency
ESTEC	European Space Research and Technology Centre
ETH	"Eidgenössische Technische Hochschule" (Swiss federal institute of technology)
EUMILSATCOM	European military satellite communication
EUREKA	European Research Coordination Agency
EVA	extravehicular activity
FAA	Federal Aviation Authority
FCFD	fluorescent capillary fill device
FET	field effect transistor
FGM	fluxgate magnetometer
FIA	flame injection analysis
FID	flame-ionization detector
FOG	fiber-optical gyrometer
FOV	field of view
FPR	Fabry-Perot resonator
FS	fluorescence spectroscopy
FST	quartz crystal cutting direction
FT-IR	Fourier transform infrared spectroscopy
FWHM	full width at half maximum
G	G-protein
GASFET	gas-sensitive field effect transistor
GBF	"Gesellschaft für Biotechnologische Forschung" (biotechnological research society)
GC	gas chromatography
GC-MS	gas chromatography -mass spectrometry
GDP	guanosine diphosphate
GEO	geostationary earth orbit
GLONASS	global navigation satellite systems
GNSS	global navigation satellite systems
GOLF	global oscillations at low frequencies
GPS	global positioning system
GPWS	ground proximity warning system
GRIN	graded index
GTP	guanosine triphosphate
HART	highway addressable remote transducer
HBMC	hydroxybutylmethylcellulose
HMDV	hoof-and-mouth disease virus
HMPIP	Her Majesty´s Inspectorate of Pollution
HOMO	highest occupied molecular orbital
HPLC	high-performance liquid chromatography
HREELS	high resolution electron energy loss spectroscopy
HSI	horizontal situation indicator
IBP	"Fraunhofer-Institut für Bauphysik" (institute of civil engineering physics)
IC	integrated circuit
ICB	ion cluster beam
ID	identification
IDT	interdigital transducer

Abbreviation	Explanation
IEC	International Electrotechnical Commission
IEE	Institution of Electrical Engineers
IEEE	Institute of Electrical and Electronics Engineering
IgG	immunoglobulin-G
ILS	instrument landing system
IMT	"Institut für Mikrostrukturtechnik" (microstructuring institute)
IMU	inertial measurement unit
INU	inertial navigation unit
IO	integrated optics
IO-EFAS	integrated-optical evanescent field absorbance sensor
IOPO	integrated optical parametric oscillator
IOT	"Integrierte Optik Technologie" (IO technology)
IPA	Integrated Pollution Control
IR	infrared
IR-ATR	infrared spectroscopy of attenuated total reflection
IRCH	"Institut für Radiochemie" (radiochemical institute)
ISA	industry standard architecture
ISE	ion-selective electrode
ISFET	ion-sensitive field effect transistor
ISM	ion-sensitive membrane
ISO	International Standards Organization
ISO	infrared space observatory
ISS	ion scattering spectroscopy
IVA	intravehicular activity
IVF	in vitro fertilization
JAMIE	Joint Analogue Microsystems Initiative of Europe
KfK	"Kernforschungszentrum Karlruhe GmbH" (Karlruhe nuclear research center)
LASCO	white light and spectrometric coronograph
LBA	Luftfahrt Bundesamt (federal aviation authority)
LCD	liquid crystal display
LED	light emitting diode
LEED	low electron energy diffraction
LEO	low earth orbit
LFGC	linear focusing grating coupler
LIDAR	light detection and ranging
LIGA	"Lithographisch-galvanische Abformung" (lithography, electrodeposition, molding)
LIPAS	laser-induced photoacoustic spectroscopy
LLWAS	low level wind shear alert system
LMT	lateral magnetotransistor
LOD	limit of detection
LP	lactose permease
LPCVD	low pressure / liquid phase chemical vapor deposition
LUMO	lowest unoccupied molecular orbital
LVDT	linear variable differential transformer
LVIT	linear variable integral transformer
MAK	"Maximale Arbeitsplatzkonzentration" (maximum allowable concentration at workplaces)
MAMA	multi-anode microchannel array

Abbreviation	Explanation
MAP	manifold absolute pressure
MAU	million accounting units
MBE	molecular beam epitaxy
MDI/SOI	Michelson Doppler imager / solar oscillations monitor
MetHb	methemoglobin
MFH	multi-frequency heterodyne receiver
MFM	magnetic force microscopy
MLS	microwave landing system
MO	magneto-optical
MRS	Materials Research Society
MS	mass spectrometry
MT	magnetotransistor
MUA	mercaptoundecanoic acid
MUX	multiplexer
MZI	Mach-Zehnder interferometer
NAMAS	National Measurement Accreditation Service
NASA	National Aeronautics and Space Administration
NATO ARW	North Atlantic Treaty Organization Advanced Research Workshop
NEC	noise equivalent concentration
NEDC	National Economic Development Council
NEDO	New Energy and Industrial Technology Development Organization
NFTL	near-field thermal lensing
NIMBY	"not in my backyard"
NIR	near infrared
NOSE	Neotronics olfactory sensing equipment
NRA	National Rivers Authority
NTC	negative temperature coefficient
NTT	Nippon Telegraph and Telephone Corporation
OBD	on-board-diagnostics
OM	optical monitor
OSA	Optical Society of America
OTA	operational transconductance amplifier
OTC	"over-the-counter"
OTDR	optical time domain reflectometry
PA	product assurance
PAH	polyaromatic hydrocarbon
PAPMS	poly(aminopropyl)methylsiloxane
PAS	photoacoustic spectroscopy
PB	polybutadiene
Pc	phthalocyanine
PC	phosphatidylcholine
PC	personal computer
PCB	polychlorinated biphenyl
PCI	process control instrumentation
PCMS	poly(cyanopropyl)methylsiloxane
PCPMS	poly(cyanopropyl)methylsiloxane
PDMS	polydimethylsiloxane
PDS	photothermal deflection spectroscopy

Abbreviation	Explanation
PE	plasma etching
PE	phosphatidylethanolamine
PEACE	plasma electron and current analyzer
PECH	polyepichlorohydrin
PEI	polyethyleneimine
PI	polyimide
PIB	polyisobutylene
PID	photo-ionization detector
PIN	positive region/intrinsic semiconductor/negative region
PiPCMS	poly(isopropylcarboxylic acid)methylsiloxane
PLS	partial least squares
PMMA	poly(methylmethacrylate)
PPA	polyphenylacetylene
PPL	priority pollutant list
PPMS	polyphenylmethylsiloxane
PROM	programmable read only memory
PS	polystyrene
PSA	prostate-specific antigen
PSI	Paul-Scherrer Institute
PSRR	power supply rejection ratio
PTB	"Physikalisch-Technische Bundesanstalt Braunschweig" (federal institute of physical science and technology)
PTC	positive temperature coefficient
PTFE	polytetrafluoroethene
PTPS	photothermal phase shift spectroscopy
PTR	Pt resistance thermometers
PVDF	polyvinylidenedifluoride
PZT	$PbZrO_3$ - $PbTiO_3$
QA	quality assurance
QMB	quartz microbalance
R&D	research and development
radar	radio detection and ranging
RAM	random access memory
RAM	Raman spectroscopy
RAPID	research with adaptive particle imaging detectors
RC	resistor/capacitor
RCRA	Resource Conservation and Recovery Act
rf	radio frequency
RF	radio frequency
RGS	reflection grating spectrometer
RI	refractive index
RIE	reactive ion etching
RLG	ring laser gyrometer
RMI	radio magnetic indicator
rms	root mean square
RoW	"rest of the world"
RPM	revolutions per minute
RTD	resistance temperature devices

Abbreviation	Explanation
RVD	rendezvous-docking
RW	Rayleigh wave
SAE	Society of Automotive Engineering
SAM	scanning Auger microscopy
SAR	synthetic aperture radar
SAW	surface acoustic wave
sc	single crystal
SE	stationary electrode
SECM	scanning electrochemical microscopy
SEDX	scanning energy dispersive X-ray analysis
SEF	spectral edge frequency
SEL	"Standard Elektrik Lorenz A.G."
SELL	spectral ellipsometry
SEM	scanning electron microscopy
SFM	scanning force microscopy
SH	shear horizontal
SIMS	secondary ion mass spectrometry
SLAR	side-looking airborne radar
SM	seismic mass
SNOM	scanning near-field optical microscopy
SNOS	scanning near-field optical spectroscopy
SNR	signal-to-noise ratio
SP	sagittal polarization
SPA-LEED	spot-profile-analysis low electron energy diffraction
SPC	statistical process control
SPIE	Society of Photo-Optical Instrumentation Engineers
SPM	scanning probe microscopy
SPR	surface plasmon resonance
SQUID	superconducting quantum interference device
SSBW	surface skimming bulk wave
SSIMS	scanning secondary ion mass spectrometry
SSPR	scanning surface plasmon resonance
ST,X	quartz cutting direction
STAFF	spatio-temporal analysis of field fluctuations
STM	scanning tunnelling microscopy
SUMER	solar UV measurements of emitted radiation
SW	surface wave
SWAN	solar wind anisotropies
TA	thermal analysis
TAB	tape automated bonding
TCAS	traffic alert and collision avoidance system
TCD	temperature coefficient of delay
TCE	trichloroethene
TCL	target compound list
TCO	temperature coefficient of offset
TDS	thermal desorption spectroscopy
TE	transverse electric
TEM	transmission electron microscopy

Abbreviation	Explanation
TFGC	trifocal focusing grating coupler
TL	thermal lensing
TM	transverse magnetic
TMPyP	5,10,15,20-tetra(4-N-methylpyridyl)porphyrin
TNT	trinitrotoluene
TOC	total organic carbon
TOD	total oxygen demand
TT&C	telemetry, tracking and command
TU	"Technische Universität" (university of engineering sciences)
TV	television
UHF	ultra high frequency
UPS	UV photoemission spectroscopy
UV	ultraviolet
UVCS	UV coronograph spectrometer
VDE	"Verein Deutscher Elektrotechniker" (association of German electrical engineers)
VDI	"Verein Deutscher Ingenieure" (association of German engineers)
VHF	very high frequency
VIRGO	variability of solar irradiance
VIS	visible (light)
VLBI	very long baseline interferometry
VOC	volatile organic compounds
VSI	vertical speed indicator
WBD	wide band data
WEU	Western European Union
WHISPER	waves of high frequency and sounder for probing of electron density by relaxation
WIS	water industry specifications
X	quartz cutting direction
XPS	X-ray photoemission spectroscopy
YAG	Y-Al-garnet
YSZ	Y-stabilized ZrO_2

©VCH Verlagsgesellschaft mbH,1995

ELECTROANALYSIS

An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis

Published monthly.
ISSN 1040-0397. Language of publication: English.

Electroanalysis is a monthly international journal of high repute covering all branches of electroanalytical chemistry.
It offers you a complete source of information, hence it frees you from having to search through other journals to find the data you need.

Serving as a vital communication link between the research labs and the field, ***Electroanalysis*** helps you to quickly adapt the latest innovations into their practical clinical, environmental and industrial applications.

Yes, please send me a free sample copy of
ELECTROANALYSIS

Name

Address

X

Date, Signature

Please fill in the form and return it to:

VCH, Reader Service
P.O. Box 10 11 61
D-69451 Weinheim, Germany

Fax: +49 62 01/606 117
Tel.: +49 62 01/606 427