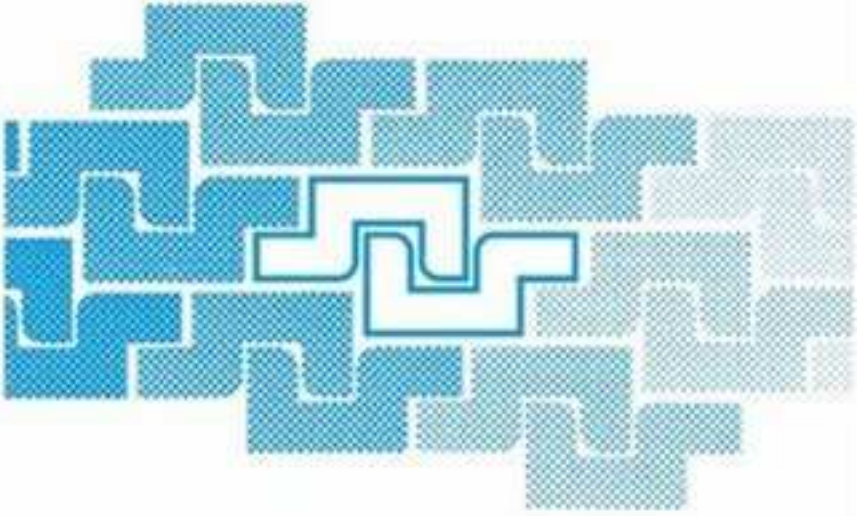


Smart Sensors and MEMS

Edited by

Sergey Y. Yurish and
Maria Teresa S.R. Gomes

NATO Science Series

II. Mathematics, Physics and Chemistry – Vol. 181

Smart Sensors and MEMS

NATO Science Series

A Series presenting the results of scientific meetings supported under the NATO Science Programme.

The Series is published by IOS Press, Amsterdam, and Kluwer Academic Publishers in conjunction with the NATO Scientific Affairs Division

Sub-Series

I. Life and Behavioural Sciences	IOS Press
II. Mathematics, Physics and Chemistry	Kluwer Academic Publishers
III. Computer and Systems Science	IOS Press
IV. Earth and Environmental Sciences	Kluwer Academic Publishers
V. Science and Technology Policy	IOS Press

The NATO Science Series continues the series of books published formerly as the NATO ASI Series.

The NATO Science Programme offers support for collaboration in civil science between scientists of countries of the Euro-Atlantic Partnership Council. The types of scientific meeting generally supported are "Advanced Study Institutes" and "Advanced Research Workshops", although other types of meeting are supported from time to time. The NATO Science Series collects together the results of these meetings. The meetings are co-organized bij scientists from NATO countries and scientists from NATO's Partner countries – countries of the CIS and Central and Eastern Europe.

Advanced Study Institutes are high-level tutorial courses offering in-depth study of latest advances in a field.
Advanced Research Workshops are expert meetings aimed at critical assessment of a field, and identification of directions for future action.

As a consequence of the restructuring of the NATO Science Programme in 1999, the NATO Science Series has been re-organised and there are currently Five Sub-series as noted above. Please consult the following web sites for information on previous volumes published in the Series, as well as details of earlier Sub-series.

http://www.nato.int/science
http://www.wkap.nl
http://www.iospress.nl
http://www.wtv-books.de/nato-pco.htm

Series II: Mathematics, Physics and Chemistry – Vol. 181

Smart Sensors and MEMS

edited by

Sergey Y. Yurish
National University Lviv Polytechnic,
Ukraine

and

Maria Teresa S.R. Gomes
University of Aveiro,
Portugal

Kluwer Academic Publishers

Dordrecht / Boston / London

Published in cooperation with NATO Scientific Affairs Division

Proceedings of the NATO Advanced Study Institute on
Smart Sensors and MEMS
Povoa de Varzim, Portugal
8-19 September 2003

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 1-4020-2928-4 (PB)
ISBN 1-4020-2927-6 (HB)
ISBN 1-4020-2929-2 (e-book)

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands.

NATO ASI '*Smart Sensors and MEMS*' participants and lecturers (8-19 September 2003, Povoa de Varzim, Portugal).

Contents

List of Contributors

Sergey Y. Yurish – NATO ASI Co-director, Dr., Associate Professor, IFSA Vice President, National University Lviv Polytechnic, Ukraine, 79013 Lviv, Bandera str., 12,
Telephone: +380 322 970857
Fax: +380 322970857
E-mail: info@sensorsportal.com
Http://www.sensorsportal.com

Maria Teresa Seabra dos Reis Gomes, NATO ASI Co-director, Professor, Dr., Department of Chemistry University of Aveiro, 3810-193 Aveiro, Portugal
Telephone: +351 234370722
Fax +351 234370084
E-mail: mtgomes@dq.ua.pt

Fabien Josse, Profesor, Dr., Electrical, Computer and Biomedical Engineering, Marquette University, P.O.Box. 1881, Milwaukee, WI 53201-1881, USA
E-mail: fabien.josse@marquette.edu

Vittorio Ferrari, Professor, Dr.**,** Dip. Elettronica per l'Automazione, Facoltá di Ingegneria - Universitá di Brescia, Via Branze, 38 - I25123 Brescia – Italy
E-mail: vittorio.ferrari@unibs.it

Elke Mackensen, Professor, Dr.,
Albert-Ludwigs-University Freiburg,
Institute of Microsystem Technology,
IMTEK, Laboratory for Electrical
Measurement and Testing,
Georges-Köhler-Allee; Geb. 103,
D-79110 Freiburg, Germany
E-mail: mackense@imtek.uni-
freiburg.de

Pavel Ripka, Professor, Dr., Czech
Technical University, Faculty of
Electrical Engineering, Dept. of
Measurement, Technicka 2,
Prague 6, 166 27, Czech Republic
E-mail**:** ripka@feld.cvut.cz

Radivoje Popovic, Professor, Dr.,
Swiss Federal Institute of
Technology Lausanne,
EPFL – DMT – IMS, 1015
Lausanne, Switzerland
E-mail: radivoje.popovic@epfl.ch

Ulrich Mescheder, Professor, Dr.**,**
FH Furtwangen - University of
Applied Sciences, Furtwangen,
Department Computer and Electrical
Engineering, Robert-Gerwig-Platz 1,
D-78120 Furtwangen, Germany
E-mail: mes@fh-furtwangen.de

Panos Datskos, Professor, Dr., Oak Ridge National Laboratory, Engineering Science and Technology Division, Bldg. 4500S, Rm H-160, MS 6141, Bethel Valley Rd., MS6141, P.O. Box 2008, Oak Ridge, TN 37831-6141, USA
E-mail: pgd@ornl.gov

Eugenii Katz, Professor, Dr., Institute of Chemistry, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel
E-mail: ekatz@vms.huji.ac.il

Dmytro V. Fedasyuk, Professor, Dr., CAD Departament, Institute of Computer Sciences and Information Technologies, National University Lviv Polythechnic, Lviv, Ukraine
Department of Microelectronics and Information System, Technical University of Lodz, Poland
E-mail: fedasyuk@polynet.lviv.ua

Preface

'Smart Sensors and MEMS' was the title of a NATO Advanced Study Institute (ASI) held in Povoa de Varzim, Portugal, from 8 to 19 September 2003. The purpose of the meeting was to discuss and disseminate the latest knowledge in field of smart sensors, transducers and MEMS technologies with a view of advanced signal processing and novel conversion methods. With that in mind, a broad range of physical, chemical and biosensors design principles, technologies and applications were included in the programme.

It was a first attempt to discuss in the same event different physical, chemical, biological sensors and MEMS technologies in point of view of smart sensors creation. Fourteen experts from Czech Republic, Germany, Italy, Israel, Netherlands, Portugal, Russia, Switzerland, Ukraine and USA have been invited to give lectures on latest achievements in sensors area and technologies.

The main task of measuring instruments, sensors and transducers designing has always been to reach high metrology performances. At different stages of measurement technology development, this task was solved using technological methods, consisting in technology perfection, as well as structural and structural-algorithmic methods. Historically, technological methods have received prevalence in the USA, Japan and Western European countries. The structural and structural-algorithmic methods have received a broad development in the former USSR and continue developing in NIS countries. The improvement of metrology performances and extension of functional capabilities are being achieved through the implementation of particular structures designed in most cases in heuristic way, and using advanced calculations, algorithms and signal processing (for example, Lyapunov's characteristic functions, weight functions, inferential calculations for soft sensors, advanced self-adaptive methods for frequency-to-digital conversion, etc.). Digital and quasi-digital smart sensors and transducers are not the exception [1].

The last crisis in the area of high technologies has evidently shown, that it is not enough to use only the technological methods. Despite of crisis, we need now smart sensors with increased accuracy, reliability and speed. Nowadays, intelligent sensors are extremely necessary for such applications, as electronic noses and tongues, smart vision systems, personnel (human body) detection, authentication systems, building monitoring system, etc. Most effectively for achievement of this purpose is a combination of technological methods and structural-algorithmic methods. It allows to achieve the same performances (or even better) at reduced material and human costs, with a much faster response.

The NATO ASI has brought together experts from the various and highly diverse areas of this broad field and to diffuse their work in the field of smart sensors and MEMS. The diffusion and share of concepts is very important and timely for the future developments in this emerging scientific area. In fact, the main objective of the NATO ASI was to disseminate up to date knowledge concerning the new and emerging applications of modern smart sensors (including their usage in the defense against terrorism) and to stimulate active international cooperation in the area of novel physical and (bio) chemical sensing principles, modern microsystem technologies and new methods of signal processing and conversion.

The NATO ASI 'Smart Sensors and MEMS' has presented the state-of-the-art and given an excellent opportunity to provide a systematic, in-depth treatment of the new and rapidly developing field of smart sensors and MEMS. The ultimate goal of the NATO ASI - to encourage the participants from many countries to work together on different smart sensors and to develop joint projects in the future has been achieved in full.

In addition to the round table and two panel discussions, the addition discussion on 'Smart Sensor Systems' has been initiated by lecturers and participants and held during the NATO ASI. It has been considered that there are a lot of new technologies, suitable for smart sensors creation [2], for example, micro- or nano-cantilevers (lecturer *P. Datscos*, Oak Ridge National Laboratory, USA) porous silicon (lecturer *U. Mescheder*, University of Applied Science, Furtwangen, Germany), thick films (lecturer *V. Ferrari*, Universitá di Brescia, Italy), resonant piezoelectric (lecturer *F. Josse*, Marquette University, USA), molecular architecture and nanotechnologies (lecturer *E. Katz*, The Hebrew University of Jerusalem, Israel) and other modern technologies let to produce different sensors classes: physical, chemical and biological. But still there is a problem how to joint them and use in a frame of smart sensor systems. One of points of view was to use novel frequency-to-digital conversion methods and converters (lecturer *S. Yurish*, International Frequency Sensor Association) in order to move from a traditional analog (voltage and current) signal domain to frequency-time signal domain. The last one lets to eliminate a lot of technical problems due to properties of frequency as informative parameters. No output standardization is necessary as in the case of analog signal domain. Nevertheless, as rule, the sensitive response of many sensors is in *mV*, it is expediently to use the voltage-to-frequency intermediate conversion. In additional such approach will give an opportunity to create new self-adaptive smart sensors (*S. Yurish*).

New conversion methods and advanced signal processing will play role a bridge between many different technologies at smart sensor systems design. The best modern approach for smart sensor systems creation is to

use both modern technologies and advanced methods for signal processing and measurement, especially in the frequency-time signal domain. Many type of sensing elements and read-out circuitry can be merged by this way on a single chip or in SoC. However, the combination of monolithic and hybrid integration with advanced processing and conversional methods in many cases allows also to achieve good results.

Such kind of research has been initiated by International Frequency Sensor Association IFSA [3] and now is a hot topic for a big international joint research project.

The aims of this volume are to disseminate wider and in-depth theoretical and practical knowledge about smart sensors and its applications, to create a clear consciousness about the effectiveness of MEMS technologies, advanced signal processing and conversion methods, to stimulate the theoretical and applied research in these very important areas, and promote the practical using of these techniques in the industry.

The book is an excellent guide for the practicing engineer, the researcher and the student interested in this crucial aspect of actual smart sensor design.

REFERENCES

[1] Yurish S.Y. Modern MEMS Technologies and Advanced Signal Processing: How to Obtain More Benefits in Smart Sensors Systems ? Sensors & Transducers 2003; 36:I:III.

[2] Smart Sensors and MEMS: Tutorials and Posters Abstracts, NATO ASI 'Smart Sensors and MEMS'. Preprints, ed. by Maria Teresa Gomes and Sergey Y. Yurish, Povoa De Varzim, 8-19 September 2003 (ISBN 0-9733840-0-X).

[3] Yurish S.Y., Novel Measurement Techniques up for License, IFSA Press Release, 21 November 2002, Toronto, ON, Canada, http://www.sensorsportal.com

Lviv, January 2004 *The editors*

Acknowledgements

The rapidly growing interest in smart sensors design as observable through the activities of International Frequency Sensor Association (IFSA) has been the motivation to start the preparation of a NATO Advanced Study Institute on the topic 'Smart Sensors and MEMS'.

The NATO ASI and this book would not have been possible without the support and enthusiasm of many people and organizations. But still the book cannot reflect the stimulating mood of the many discussions during the course, which were due to the interest and active contribution of the ASI lecturers and students.

For the excellent and very responsive hosting in Povoa de Varzim, the congress manager Vanessa Pinto Campos (LusoCongressos, Portugal) deserve our acknowledgement.

All lecturers did a great job in the preparation of the contents of their contributions both during ASI and for this book. Since only a few books are available on this hot topic.

Finally, however, neither the NATO ASI nor this book would have been possible without the considerable financial support of the NATO Scientific Council who deserves our very special acknowledgement.

The NATO ASI 'Smart Sensors and MEMS' was sponsored by

NATO – North-Atlantic Treaty Organization
(Grant No. PST.ASI.979712)

and organized with the technical cooperation of

University of Aveiro, Department of Chemistry
LusoCongressos (Portugal)
International Frequency Sensor Association (IFSA)
Sensors Web Portal, Inc. (Canada) – media support

The directors of the NATO ASI on Smart Sensors and MEMS,
Povoa de Varzim, 2003,

Sergey Y. Yurish,
National University Lviv Polytechnic, UA

Maria Teresa S.R. Gomes
University of Aveiro, Department of Chemistry, Portugal

Chapter 1

SMART SENSORS FOR ELECTRICAL AND NON-ELECLRICAL, PHYSICAL AND CHEMICAL VARIABLES: STATE-OF-THE-ART

Sergey Y. Yurish
National University Lviv Polytechnic, UA

Abstract: The chapter gives an overview of industrial types of smart sensors (temperature, pressure, rotation speed, optical, humidity, gas, chemical, acceleration, biosensors etc.) and contains quasi-digital, smart sensors and MEMS definitions. Digital and quasi-digital (frequency and duty-cycle output) sensors and transducers are considered. The obvious tendency of sensors accuracy increasing up to 0.01 % or better and wide conversion frequency range from several hundredth parts of Hz up to several MHz are observed. Main advantages of frequency-time domain signal as informative parameters for modern smart sensors are described.

Keywords: smart sensor; MEMS; frequency-time domain; digital sensor; quasi-digital sensor; transducer; integrated sensor; intelligent sensor; parametric sensor; self-generating sensor; modulating sensor, biosensor

1. INTRODUCTION

Smart sensors and MEMS are of great interest in many fields of industry, control systems, biomedical applications, etc. Very fast advances in IC technologies have brought new challenges in the physical design of integrated sensors and Micro-Electrical-Mechanical Systems (MEMS).

Modern microsystem technology (MST) offers new way of combining sensing, signal processing and actuation on a microscopic scale and allows both traditional and new sensors to be realised for a wide range of applications and operational environments. The term "MEMS" is used in different ways: for some, it is equivalent to word "MST", for others, it comprises only surface-micromechanical products. MEMS in the latter sense are seen as an extension to IC technology: "*an IC chip that provides sensing and/or actuation functions in addition to the electronic ones"* [1]. The definition of a smart sensor is based on [2] and can be formulated as: "*smart sensor (or intelligent sensor) is one chip, without external components,*

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 1-50.

including the sensing, interfacing, signal processing and intelligence (self-testing, self-identification, self-validation or self-adaptation) functions".

During measurements different kind of measurands are converted into limited number of output parameters. The mechanical displacement was the first historical type of such (unified) parameter. The mercury thermometer, metal pressure gauge, pointer voltmeter, etc. are based on such principle [3]. Amplitude of electric current or voltage is another type of unified parameter. Now almost all properties of substance and energy can be converted into current or voltage with the help of different sensors. All these sensors are based on the usage of an amplitude modulation of electromagnetic processes. They are so called analog sensors.

Digital sensors appeared, when a necessity to input results of measurement into computer arisen. According to researches of International Frequency Sensor Association (IFSA) the following division of output sensor signals is observed today (Figure 1).

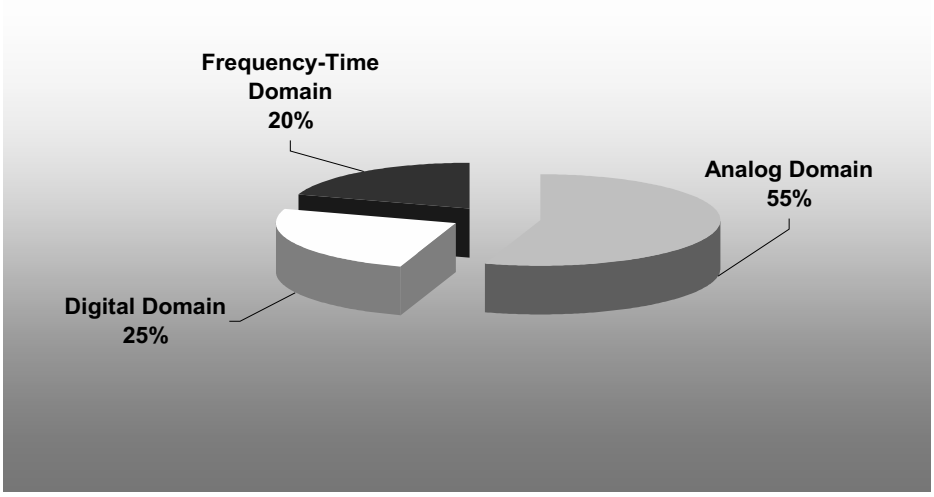


Figure 1. Classification of in Terms of Output Signals (studied by IFSA, 2003).

Firstly, the design task of digital sensors was solved by transforming of an analog quantity into a digital code by an analog-to-digital converter (ADC). The creation of quasi-digital sensors, in particular, frequency sensors, was another very promising direction [3]. *Quasi-digital sensors are discrete frequency-time domain sensors with frequency, period, duty-cycle, time interval, pulse number or phase shift output.* Today, group of frequency output sensors is the most numerous among all quasi-digital sensors (Figure 2). Such sensors combine a simplicity and universatility that is

inherent to analog devices and accuracy and noise immunity, proper to sensors with digital output. The further transformation of a frequency-modulated signal was reduced to counting of periods of a signal during reference time interval (gate). This operation exceeds in a simplicity and accuracy all other methods of analog-to-digital conversion [4].

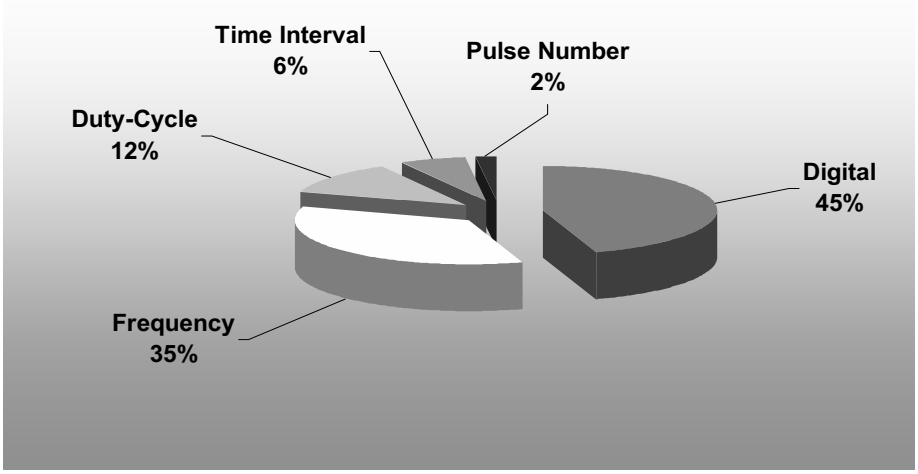


Figure 2. Classification of Sensors from Discrete Group in Terms of Output Signals (studied by IFSA, 2003).

Separate types of frequency transducers have been known for many years ago. However, the output frequency of such sensors (before digital frequency counters appeared) was measured by analogue methods and consequently substantial benefit from the usage of frequency output sensors practically has not been achieved.

The situation has dramatically changed since digital frequency counters and frequency output sensors attracted the increasing attention. As far back as 1961 professor P.V. Novitskiy wrote: "... In the future we can expect, that a class of frequency sensors will get such development, that the number of now known frequency sensors will exceed the number of now known amplitude sensors..." [3]. Although there are frequency output sensors practically for any variables, this prognosis has not been fully justified in the full because of some reasons.

With appearance sensor microsystems and heady development of microsystem technologies all over the world, the technological and cost factors have been modified for the benefit of digital and quasi-digital sensors. Modern technologies allow to solve rather complicated tasks, concerned with creation of different sensors. Up to now, however, have still

been some major obstacles preventing industries from largely exploiting such sensors in their systems. Most likely, there are only some subjective reasons:

- The lacking awareness of the innovation potential of modern methods for frequency-time conversion in many companies, as this processing techniques have mainly been developed in Former Soviet Union;
- Tendency of the companies to return, first of all, major expenditures, invested in development of conventional ADC;
- Lack of emphasis being placed on the business and market benefits which such measuring technologies can bring to companies and etc.

Today situation has changed dramatically. According to *Intechno Consulting*, the non-military world market for sensors will grow at an annual rate of 5.3% to reach US $ 42.2 billion. Under very conservative assumptions it is expected to reach US $ 50-51 billion by 2008; assuming more favorable but still realistic economic conditions, the global sensor market volume could even reach US $ 54 billion by 2008. Sensors on semiconductor basis will increase their market share from 38.9% in 1998 to 43% in 2008. Strong growth expected for sensors based on MEMS-technologies, smart sensors and sensors with bus capabilities [5]. According to The Freedonia Group the USA market for sensor products (sensors, transducers and associated housing) is projected to increase 6.7 % per year through 2006 to US $ 13.4 billion. The fastest growths will occur in sensors based on advanced, sophisticated technology – especially MEMS and optoelectronics – and/or used in dynamic applications such as automotive telematics and information technology. Also holding good prospects are imaging sensors [6]. It is reasonable to expect that silicon sensors will go on to conquer other markets, such as the appliances, telecommunications and PC market [7].

2. SMART SENSOR ARCHITECTURES AND DATA ACQUISITION

The processing and interpretation of information arriving form the outside are the main tasks of data acquisition systems and measuring instruments based on computers. Data acquisition and control systems need to get real-world signals into the computer. These signals come from a diverse range of transducers and sensors. According to [8] *Data Acquisition (DAQ) is collecting and measuring electrical signals from sensors and transducers and inputting them to a computer for processing.* The further processing can include the sensors' characteristic transformation, joint processing for many parameters as well as statistical calculation of results and represent them in a user-friendly manner.

According to the output signal, sensors and transducers can be divided into potential (amplitude), current, frequency, pulse-time and digital. As a result, the task of adequate sensor interfacing with PC arises before developers and users of any data acquisition systems. For solution of this problem, the special attention must be paid to problems of output conversion into a digital format as well as to high accuracy and speed conversion methods.

In general, a sensor is a device, which is designed to acquire information from an object and to transform it into an electrical signal. A classical integrated sensor can be divided into four parts as it is shown in Figure 3. The first block is a sensing element (for example, resistor, capacitor, transistor, piezo-electric material, photodiode, resistive bridge, etc.). The signal produced from the sensing element itself is often influenced by noise or interference. Therefore, signal conditioning and signal processing techniques as amplification, linearization, compensation and filtering are necessary (second block) to reduce sensor non-idealities.

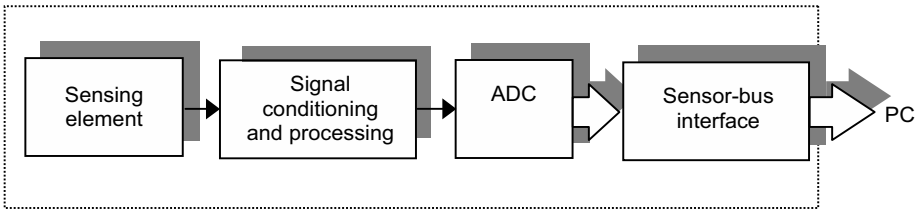


Figure 3. Integrated Sensors.

Sometimes if some sensing elements are used on the same chip, the multiplexer is necessary. In cases of data acquisition, the signal from the sensor must be in a serial or parallel digital format. This function can be realised by the analog-to-digital or frequency-to-digital converter. The last block is a sensor-bus interface. A data acquisition system can have a star configuration in which each sensor is connected to a digital multiplexer. When using a large number of sensors, the total cable length and the number of connection at the multiplexer can become very high. For this reason it is much more acceptable to have a bus-organized system which connects all data sources and receivers. This bus system handles all data transports and is connected to a suitable interface that sends accumulated data to the computer [9].

Dependent on the smart sensor architecture, various data acquisition schemes are possible. In the smart sensor architecture shown in Figure 4, the analog output of the sensor element S is at first amplified and corrected for offset, non-linearity, etc. Then the voltage-to-frequency conversion takes place. The frequency-time domain signal (frequency, period, time interval,

duty-cycle, etc.) is converted into digital. The format of the frequency-to-digital converter is such that the signal is transferred to the bus system at the command of the bus controller.

In the second example (Figure 5) some sensors elements form a sensors array. A single multiplexing circuit feeds signal-conditioned signals from sensing elements one after the other into a single frequency-to-digital converter and from here, the signals are transferred to the bus.

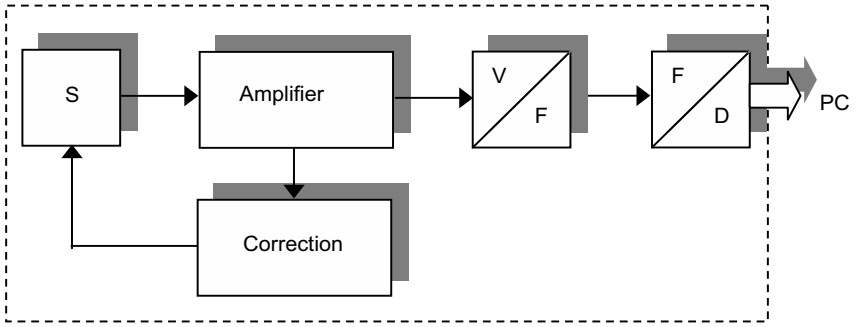


Figure 4. Smart Sensor Architecture with Preliminary Correction in Analog Domain and Further Conversion into Frequency-Time Signal Domain.

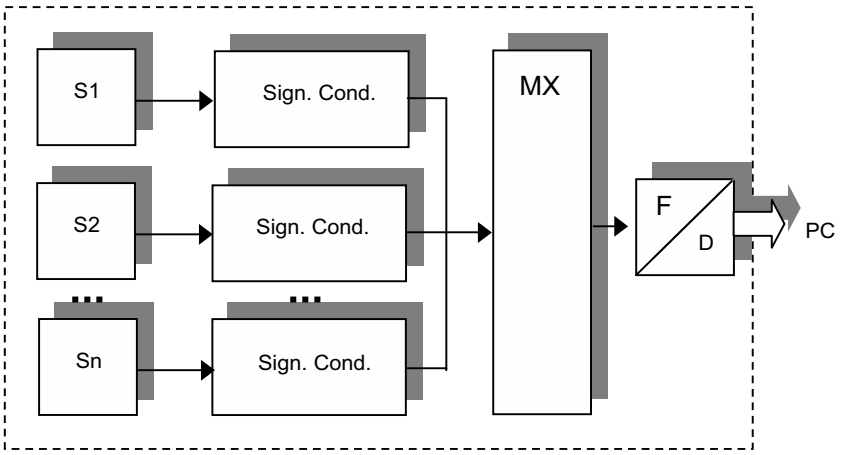


Figure 5. Architecture of Smart Sensor Array.

Such a sensor array or a multiparameter sensor can, for example, measure different variables like the temperature, the pressure, the humidity, etc. at a certain location.

In the third example of the smart sensor architecture, a sensor element is connected via a frequency-to-digital converter to a microcontroller

(Figure 6). The microcontroller can store the sensor's characteristic data in its internal ROM and, based on this information and the sensor signal, the microcontroller transfers the corrected signal to the bus. A very useful feature of such a smart sensor architecture is that the microcontroller also permits the central computer to send data back to the sensor, which can be used to change the measuring range, to exert a recalibration or to adjust the offset.

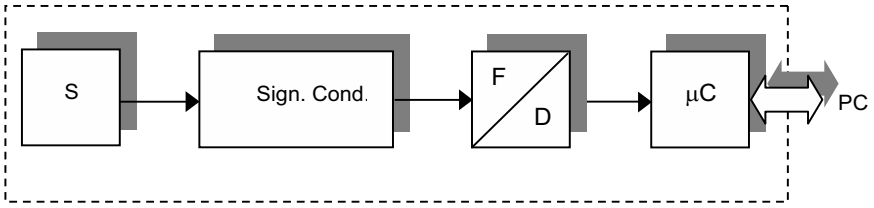


Figure 6. Smart Sensor Architecture with Microcontroller.

The microcontroller can itself realize the frequency-to-digital conversion with the help of co-called program-oriented frequency-to-digital conversion methods. In this case, the smart sensor architecture becomes simpler (Figure 7).

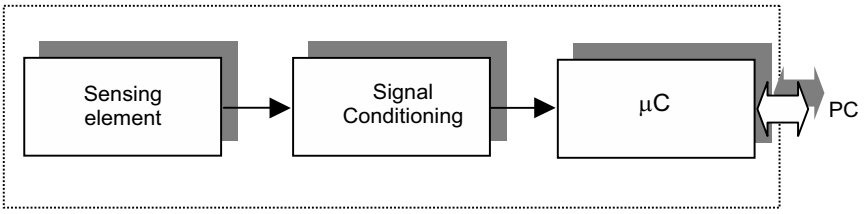


Figure 7. Smart Sensor with Program-Oriented Frequency-to-Digital Conversion.

A microcontroller is typically used for digital signal processing (for example, digital filtering), analog-to-digital or frequency-to-digital conversions, calculations and interfacing functions. Microcontrollers can be very well combined or equipped with standard interface circuits. Many microcontrollers include the two-wire I^2C bus interface, which is suited for communication over short distances (several meters) [10] or the serial interface RS-232/485 for communication over relatively long distance.

However, the essential difference of the smart sensor from the usual integrated sensor with embedded data processing circuitry is intelligence capabilities (self-diagnostics, self-identification or self-adaptation (decision-making)) functions. As a rule, these functions are implemented due to a

built-in microcontroller (microcontroller core ("microcontroller like" ASIC) or Application Specific Instruction Processor (ASIP)) or DSP. New functions and a possibility to modify sensor's performances are main advantages of such a smart sensor. Due to a smart sensor adaptability the measuring process can be optimized for maximum accuracy, speed and power consumption. Sometimes "smart sensors" are called "intelligent transducers".

The smart sensor architecture (Figure 7) is used with the aim to realize such a property of smart sensors as self-adaptation when dependent on measuring conditions or measurands the parameters of the method for frequency-to-digital conversion will be varied, for example, conversion time or accuracy of measurements.

In such a subsystem with the bus architecture of each sensor or a group of sensors can also contain a circuit, which can recognize addresses, this means the circuit can detect, when the communication between the sensor and the central computer is desired.

For many years, most of the described components of a bus system were separated and had their own housing. However, more recent developments have enabled the components indicated by the dashed lines in Figure 4-7 to be integrated into a single chip by implementation of standard microelectronics library cells.

These typical considered smart sensor architectures are used for creation of digital output smart sensors. Data transmission in the digital form excludes the inphase interference and the voltaic coupling of sensor's output with the computer can be provided if it will be necessary. Instead of buses, the binary encoded information can be transmitted into a parallel or a serial port. When parallel data transmission is feasible, the pulses representing the information arrive simultaneously at the central computer input making fast data transfer possible. However, in most cases the distance between the sensor and the computer is too far to permit the parallel data transfer and here the serial data transmission is required. The main advantage of such an approach is, that for its realization it is not necessary to use any additional computer boards and specialized software drivers. All connections are external and drivers are standard. External connection provides an additional coupling in comparison with the usual data acquisition system with the bus connection. As the cost of microcontroller continues to decrease, having microcontroller at each measurement location will become affordable. In the future, a lot of different signal processing circuits can be integrated into the sensor ship. This approach is the next step to much wider distribution of intelligence. From Figures 4-7 the most important elements of a bus-oriented data-acquisition system can be deducted. It can be expected that starting with large data acquisition systems, the analogue data transfer will gradually be replaced by digital systems.

3. SMART SENSORS: TECHNOLOGIES, INFORMATIVE PARAMETERS AND PROPERTIES

At present, a lot of different types of sensors are available. Rapid advancement of the standard process for VLSI design, silicon micromachining and fabrication provide the technological basis for the realization of such a type of sensors, opens an avenue that can lead to custom integrated sensors to meet the new demands in performance, size and cost. This suggests a smooth merging of the sensor and electronics and the fabrication of complete data acquisition systems on a single silicon chip. The essential issue is rather the fabrication compatibility of the sensor, sensor-related analogue microelectronic circuits and digital interface circuits [11]. In fact, for any type of a silicon sensing element and read-out circuitry, a process can be developed to merge them on a single chip. However, process development is very expensive and therefore only a huge production volume will pay off the development cost. Successful integrated-sensor processes must have an acceptable complexity and/or applicability for a wide range of sensors [12]. MEMS technologies allow to miniaturize sensors and, at the same time, to integrate their sensor elements with microelectronic functions in minimal space. Only MEMS technologies make it possible to mass-produce sensors with increasing cost-effectiveness while improving their functionality and miniaturizing them.

Of course, the implementation of the microcontroller in one-chip together with sensing the element and signal conditioning circuitry is an elegant and rather preferable engineering solution by creation of modern integrated smart silicon sensors. However, the combination of the monolithic and hybrid integration with advanced processing and conversional methods in many cases allows to achieve magnificent technical and metrological performances for the shorter time-to-market period without additional expenditures for expensive CAD tools and the long-time smart sensor design process. By implementation of smart sensors with hybrid-integrated processing electronics the hardware minimization is a necessary condition to reach the reasonable price and high reliability. In this case, we have the so-called "hybrid smart sensor" in which a sensing element and an electronic circuit are placed in the same housing.

Frequency-time domain sensors are rather interesting from a technological and fabrication compatibility point of view, the simplifications of the signal conditioning circuitry and measurand-to-digital converter, as well as metrology performances and the hardware for realization. The last one essentially influences the chip area. Such sensors are based on resonant phenomena and variable oscillators, whose information is embedded not in the amplitude but in the frequency or the time parameter of the output signal. First of all these are sensors with the frequency *(f_x),* the period *($T_x = 1/f_x$)*, the pulse width *(t_p),* the spacing interval *(t_s),* the duty-cycle *(t_p/T_x),* the on-line

time ratio or the off-duty factor (T_x/t_p), the pulse number (N), the phase shift (φ_x) or the single time interval (τ) output. These informative parameters are shown in Figure 8. Because of such parameters used as informative ones with properties of analogue and digital signals simultaneously, these sensors have been called "quasi-digital". Frequency output sensors group is the most numerous among all quasi-digital sensors (Figure 2). Let us consider main advantages of the frequency as the informative sensor's output signals.

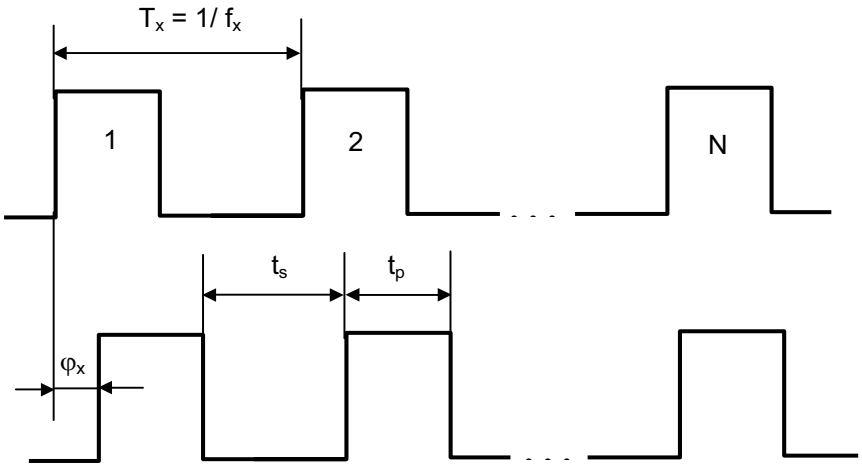


Figure 8. Frequency-Time Domain Sensor Informative Parameters.

- **High noise immunity.** In frequency sensors, the key opportunity to reach a high accuracy in comparison with the analog sensors with analog-to-digital conversion is greeted. This objective property of the high noise immunity proper to a frequency modulation is apparently, the principal premise perspective of frequency sensors in comparison with analog ones. The frequency signal can be transmitted by communication lines for the too much greater distance than analog and digital signals. The signal transmitted by the frequency way practically represents a serial digital signal. Thus, all advantages of digital systems are shown on it. Alongside with it, only the two-wire line is necessary for transmission of such a signal. In comparison with the usual serial digital data transmitting it has the advantage of not requiring of any synchronization. A frequency signal is ideal for high noise industrial environments.
- **High output signal power.** The sensor's signal can be grouped into six energy domains: electrical, thermal, mechanical, chemical, radiant and magnetic. Electrical signals are currently the most preferred signal form. Therefore, the sensor design is focused on developing of transducers that convert the signal from one or the other energy domain into a quantity in

the electrical domain. From the power point of view, the section from a sensor output up to an amplifier input is the heaviest section in a measuring channel for signal transmitting. Here the signal is transmitted by a very small level of energy. The losses, originating in this section can not be filled any more by any signal processing. Output powers of frequency sensors are, as a rule, considerably higher. In this case, the power affecting the generated frequency stability is the oscillation (reactive) power of the oscillating loop circuit and by a higher quality-factor of the oscillating loop its power is higher.

- **Wide dynamic range.** Since the signal is in the form of the frequency, the dynamic range is not limited by the supply voltage and noise. The dynamic range of over 100 dB may be easily obtained.
- **High accuracy of frequency standards.** The frequency reference, for example, crystal oscillators, can be made more stable, than the voltage reference. It is possible to explain it with the help of the same objective differences of information properties of amplitude modulated and frequency modulated signals.
- **Simplicity of commutation and interfacing.** Parasitic emf, transient resistances and cross-feed of channels in analog multiplexer by the usage of analog sensors are reduced to the occurrence of complementary errors. The frequency-modulated signal is not sensitive to all listed above factors. Multiplexers for frequency sensors and transducers are simple enough and do not introduce any errors into observed results.
- **Simplicity of integration and coding.** The precise integration in time of frequency sensor's output signal can be realized simply enough. The adding pulse counter is an ideal integrator with unlimited time of measurement. The frequency signal can be processed by microcontrollers without any additional interface circuitry.

All this makes the design and usage of different frequency-time domain smart sensors very efficient.

The most important properties of smart sensors have been well described in [9]. Here we will briefly describe only basic focus points for an intelligent frequency-time domain smart sensors design.

Adaptability. A smart sensor should be adaptive in order to optimize the measuring process. For example, depending on measuring conditions, it is preferable to have an opportunity to exchange measurement accuracy for speed and conversely, and also to moderate power consumption, when the high speed and accuracy are not required. It is desirable also to have an opportunity to adjust a clock crystal oscillator frequency depending on the environment temperature. The last opportunity also essentially influences the following focus point – *the accuracy*.

Accuracy. The measuring error should be programmable. The self-calibration will allow to reduce the systematic error, caused, for example, by the inaccuracy of the system parameters. The usage of statistical algorithms

and composited algorithms of the weight average would allow to reduce random errors caused, for instance, by interference, noise and instability.

Reliability. It is one of the most important requirements especially in industrial applications. Self-diagnostics is used to check the performance of the system and the connection of the sensor wires.

For the subject analyses of modern state in the quasi-digital smart sensor area, it is expedient to use the following classification. Depending on conversion of the primary information into frequency, all sensors are divided into three groups: sensors with measurand-to-frequency conversion, with measurand-to-voltage-to-frequency conversion and with measurand-to-parameter-to-frequency conversion.

I. Sensors with x (t) ➡ F (t) conversion. These are sensors that, due to the underlying principle, themselves generate a frequency output. Electronic circuitry might be needed for the amplification of the impedance matching, but it is not needed for the frequency conversion step itself. The obtaining of measuring information like the frequency or the frequency-pulse form is most simply reached in inductive, photoimpulse, string, acoustic and scintillation sensors, since its principle of operation allows to realise the direct conversion $x(t) \rightarrow F(t)$. One group of such sensors is based on resonant structures (piezoelectric quartz resonators, SAW (surface acoustic wave) dual-line oscillators, etc.) whereas another group is based on the periodic geometrical structure of the sensors, for example, angle encoders.

II. Sensors with x (t) ➡ V (t) ➡ F (t) conversion. This group is rather numerous in number of different electric circuits. These are Hall sensors, thermocouple sensors and photosensors based on valve photoelectric cells. When a frequency output is required, a simple voltage-to-frequency or current-to-frequency conversion circuit can be applied to obtain the desired result.

III. Sensors with x (t) ➡ P (t) ➡ F (t) conversion. The sensors of this group are rather manifold and numerous. These are the so-called electronic-oscillator based sensors. Such sensors are based on the usage of electronic oscillators in which the sensor element itself is the frequency-determining element. These are, first of all, inductive, capacity and ohmic parametric (modulating) sensors.

Parametric (modulating) sensors are devices producing the primary information by the way of respective alterations of any electrical parameter of some electrical circuit (inductance, capacity, resistance, etc.), for measuring of which it is necessary to have an external auxiliary power supply. Examples of such types of sensors are pressure sensors based on the piezoresistive effect and photodetectors based on the photoelectric effect.

In turn, *self-generating sensors* are devices permitting to receive a signal immediately by the way of a current $i(t)$ or voltage $V(t)$ and not requiring any source of power other than the signal being measured. Examples of such types of sensors are Seebeck effect based thermocouples and photo effect

based solar cells. Self-generating sensors are also called in literature "active" sensors, while modulating sensors are called "passive" sensors.

The signal power of modulating sensors is the largest and, therefore, from the noise-reduction point of view their usage is recommended.

The distinctiveness of these three sensor groups is the absence of conventional ADC. With the aim to design digital output smart sensors in this case, it is expedient to use a microcontroller for the frequency- to-digital conversion. The production of such smart sensors does not require extra technological steps. Moreover, modern CAD tools contain microcontroller cores and peripheral devices as well as voltage-to-frequency converters (VFC) in the library of standard cells. So, for example, Mentor Graphic CAD tool includes different kinds of VFC like AD537/650/652, CAD tool from *Protel* includes a lot of library cells of different *Burr-Brown's* VFC.

In comparison with the data capturing method using traditional Analog-to-Digital Converters (ADC), the data capturing method using VFC has the following advantages [13]:

- Simple, low-cost alternative to the A/D conversion;
- Integrating input properties, excellent accuracy and low non-linearity provide performance attributes unattainable with other converter types, make VFC ideal for high noise industrial environments;
- Like a dual-slope A/D converter, the VFC possesses a true integrating input and features the best, much better than a dual-slope converter, noise immunity. It is especially important in the industrial measurement and data acquisition systems. While a successive approximation A/D converter takes a "snapshot" in time, making it susceptible to noise peaks, the VFC's input is constantly integrating, smoothing the effects of noise or varying input signals;
- It has the universality. First, it is the user-selected voltage input range (± Supply). Second – it is the high accuracy of the frequency-to-digital conversion (up to 0.001 %). The error of such conversion can be neglected in a measuring channel. This is not true for traditional analog-to-digital conversion. The ADC error is commensurable with sensor's error, especially if we use the modern high precision sensors with relative error up to (0.01 %).
- When the data capturing method with the voltage-to-frequency converter is used, a frequency measurement technique must also be chosen which meets the conversion speed requirements. While it is clearly not a "fast" converter in a common case, the conversion speed of a VFC system can be optimized by using efficient techniques. Such optimization can be performed due to advanced methods of frequency-to-digital conversion, quasi pipeline data processing in a microcontroller and the usage of novel architectures of VFC.

Pointing on well-known advantages of frequency sensors, it is necessary to mark, that the number of physical phenomena, on the basis of which the

sensors with frequency and digital outputs can be designed, is essentially limited. Therefore, now analog sensors with current and voltage outputs have received broad dissemination. On the one hand, it happens because of high technological working off analog sensor units, and also because of heady development of the analog-to-digital conversion in the last years. On the other hand, voltage and current are used rather widely as unified standard signals in many measuring and control systems.

By the choice of those or other sensors the important role is played by the technological and cost factors. Therefore, the common statements, what sensors are the best - frequency or analog - disregarding concrete conditions of the usage, are not correct enough. With the appearance for the last years of sensor microsystems and the heady development of microsystem technologies all over the world, the technological and cost factors were modified for the benefit of frequency sensors.

Sensor types with the highest demand volumes are temperature sensors, pressure sensors, flow sensors, binary position sensors (proximity switches, light barriers, reflectortype photosensors), position sensors, chemical sensors for measurement in liquids and gases, filling sensors, speed and rpm-sensors, flue gas sensors and fire detectors worldwide. The fastest growing types of sensors include rain sensors, thickness sensors, sensors that measure the quality of liquids, navigation sensors, tilt sensors, photodetectors, glass breakage sensors, biosensors, magnetic field sensors, and motion detectors [5].

The part of the frequency-time domain sensor group is constantly increasing. At first, it is connected with the fast development of modern microelectronic technologies, secondly with the further development of methods of measurement for frequency domain parameters of signals and methods for frequency-to-digital conversion, and thirdly, with advantages of frequency as the informative parameter of sensors and transducers. Today it is difficult to find physical or chemical variables, for which the frequency output or digital sensors do not exist. It is naturally, this book does not claim for the scale completeness of the description of all existing sensors and their principles of operation. To the readers wishing to acquaint with smart sensors development history we would like to recommend the article [7].

4. SMART SENSORS STATE-OF-THE-ART

This review is made with the aim to illustrate frequency-time domain IC sensors state-of-the-art by the original solutions with high metrology performances, and also to formulate the basic requirements for such an important smart sensor's part, as the frequency-to-digital converter.

Frequency-time domain sensors can be grouped in several different ways. We will group them according to the measurand domains of the desired

information. There are six signal domains with the most important physical parameters, shown on Figure 9. Electrical parameters usually represent a signal from one of the non-electrical signal domains.

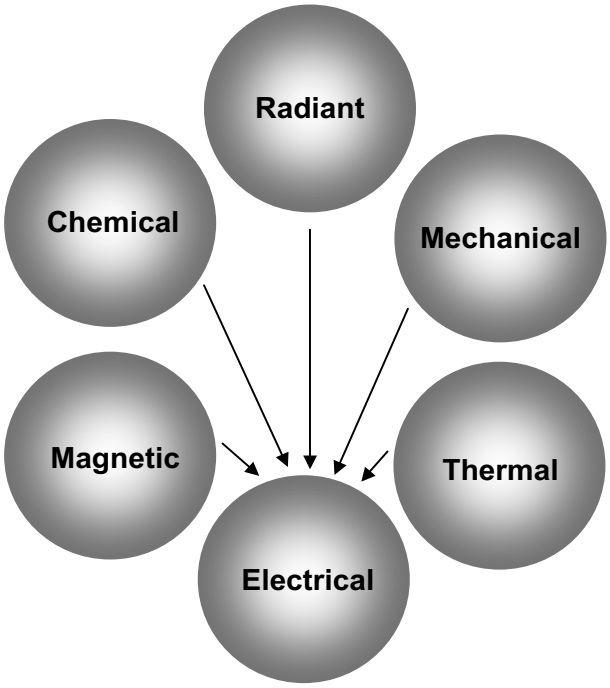


Figure 9. Sensor Classifications.

4.1 Temperature Sensors

Temperature is one of the most widely measured variables in industrial, consumer, and computer applications. It is an analog quantity, but digital systems often use temperature to implement measurement, control, and protection functions. In most applications in which temperature is measured, the measured value must be converted from analog to digital form. There is now a broad selection of digital and quasi-digital sensor types, which should match system's needs.

Four temperature sensor types (analog, digital, “analog plus” and system monitor are described in [16]. An ideal analog sensor provides an output voltage that is a perfectly linear function of temperature. In the digital I/O class of sensor, temperature data in the form of multiple 1s and 0s are passed to the microcontroller, often via a serial bus. Along the same bus, data are sent to the temperature sensor from the microcontroller, usually to set the

temperature limit at which the alert pin's digital output will trip. The alert interrupts the microcontroller when the temperature limit has been exceeded.

"Analog Plus" sensors are available with various types of digital outputs. The V_{OUT} vs. temperature curve is for an IC whose digital output switches when a specific temperature has been exceeded. In this case, the "plus" added to the analog temperature sensor is nothing more than a comparator and a voltage reference. Other types of "plus" parts ship temperature data in the form of the delay time after the part has been strobed, or in the form of the frequency or the period of a square wave.

Digital sensors include a digital interface that permits communication with a microcontroller. The interface is usually an I^2C or SMBus serial bus, but other serial interfaces such as SPI are common. In addition to reporting temperature readings to the microcontroller, the interface also receives instructions from the microcontroller. Those instructions are often temperature limits, which, if exceeded, activate a digital signal on the temperature sensor IC that interrupts the microcontroller. The microcontroller is then able to adjust fan speed or back off the speed of a microprocessor, for example, to keep temperature under control [16].

This type of device is available with a wide variety of features, among them, remote temperature sensing. To enable remote sensing, most high-performance CPUs include an on-chip transistor that provides a voltage analog of the temperature. (Only one of the transistor's two p-n junctions is used.) Other applications use a discrete transistor to perform the same function. Another important “smart” feature found on some of these types of sensors is the ability to interrupt a microcontroller when the measured temperature falls outside a range bounded by high and low limits. On other sensors, an interrupt is generated when the measured temperature exceeds either a high or a low temperature threshold. These limits are transmitted to the temperature sensor via the SMBus interface. If the temperature moves above or below the circumscribed range, the alert signal interrupts the processor.

This digital I/O class of devices finds widespread use in servers, battery packs, and hard disk drives. Temperature is monitored in numerous locations to increase a server's reliability: at the motherboard (which is essentially the ambient temperature inside the chassis), inside the CPU die, and at other heat-generating components such as graphics accelerators and hard-disk drives. Battery packs incorporate temperature sensors for safety reasons and to optimize charging profiles, which maximizes battery life.

The system monitor is the most complex IC of the four. In addition to the functions provided by the digital I/O type, this type of device commonly monitors the system supply voltages, providing an alarm when voltages rise above or sink below limits set via the I/O bus. Fan monitoring and/or control are sometimes included in this type of IC. In some cases, this class of device

is used to determine whether a fan is working. More complex versions control the fan as a function of one or more measured temperatures [16].

Reading temperature with a microcontroller is simple in concept. The microcontroller reads the output code of an analog-to-digital converter (ADC) driven by a thermistor-resistor voltage divider, analog-output temperature sensor, or other analog temperature sensor. The ADC built into some controllers can simplify this design. ADCs require a reference voltage, which can be generated by an external device. For example, the reference voltage for a thermistor sensor is usually the same as that applied to the top of the resistor-thermistor voltage divider. However, the following complications can arise in these systems: the sensor's output-voltage range is significantly smaller than the ADC's input-voltage range; limited number of ADC and/or microcontroller's pins; etc.

The design problems are simplified by usage a temperature sensor with a digital interface. Similarly, temperature sensors with time- or frequency-based outputs can alleviate the measurement problem when ADC's inputs and microcontroller's I/O pins are in short supply.

Temperature sensors play an important role in many measurements and other integrated microsystems, for example, for biomedical applications or self-checking systems and the design for the thermal testability (DfTT). IC temperature sensors take advantage of the variable resistance properties of semiconductor materials. They provide the good linear frequency, the duty-cycle or the pulse width output proportional with the temperature typically in range from –55 ^{0}C to +150 ^{0}C at the low cost. These devices can provide the direct temperature reading in a digital form, thus eliminating the need for an A/D converter. Because IC sensors can have a memory, they can be very accurately calibrated, and may operate in multisensor environments in applications such as communications networks. Many IC sensors also offer communication protocols for the use with bus-type data acquisition systems; some also have addressability and the data storage and retrieval capabilities.

Smart temperature sensors need to be provided with some kind of the output digital signal adapted to microprocessors and digital processing systems. This signal can be a time-signal type, where the measurement is represented by the duty-cycle or the frequency ratio, or the fully digital code that is sent to the processor in a serial way through the digital bus [15]. Some important restraints, caused by the integration of sensing and digital processing function on the same chip are [17] (a) the limited chip area, (b) the tolerances of the device parameters and (c) the digital interference.

Since CMOS is still the most extensively used technology the integration of temperature sensors in high-performance low-cost digital CMOS technologies is preferred in order to allow signal conditioning and digital processing on the same chip [14].

In the framework of the COPERNICUS EC project CP0922, 1995-1998, THERMINIC (THermal INvestigations of ICs and Microstructures), the

research group from Technical University of Budapest has dealt for several years with the design problem of small-size temperature sensors that must be built into the chip for thermal monitoring [18-21]. One of such sensors is based on a current-to-frequency converter [18, 19]. The analog signal of the current output CMOS sensor is converted into a quasi-digital one using a current-to-frequency converter. The block diagram of the frequency output sensor is shown on Figure 10. The I_{out} output current and its "copy" generated by a current mirror charge and discharge the capacitor C_x. The signal of the capacitor is led to a differential comparator the reference voltage of which is switched between the levels V_C and V_D [19]. The resulting frequency is

$$f = \frac{I_{out}}{2 \cdot C_x (V_C - V_D)}. \quad (1)$$

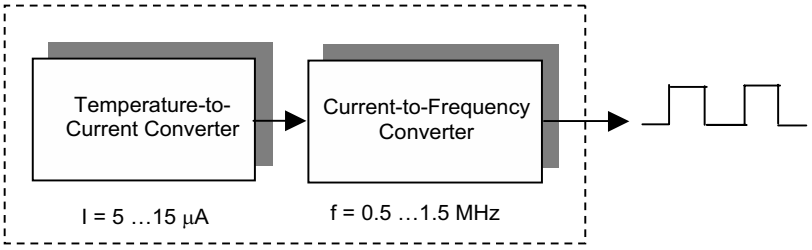


Figure 10. Temperature Frequency Output Sensor.

The sensitivity is –0.808% / ^{0}C. The output frequency 0.5 ÷ 1.3 MHz is in the convenient range. The complete circuit requires only an area of 0.018 mm^2 using the ECPD 1 μm CMOS process. The low sensitivity on the supply voltage is a remarkable feature: ± 0.25 V charge in V_{DD} results only in ± 0.28 % charge in the frequency. The latter corresponds to the ± 0.35 ^{0}C error. The total power consumption of this sensor is about 200 μW [20].

The characteristic of this sensor is quite linear and the output frequency of these sensors can be approximately written as

$$f_{out} = f_{20Cels} \exp(\gamma(T_{Cels} - 20^0 C)), \quad (2)$$

in the –50 … + 120 ^{0}C temperature range, where γ is the sensitivity, f_{20Cels} is the nominal frequency related to T=20 ^{0}C. Using the AMS 0.8 μm process, the area consumption is 0.005 mm^2 [21]. The THSENS-F [22] sensor

characteristic is shown in Figure 11. This sensor can be inserted into CMOS designs, which can be transferred and re-used as cell (layout level) entities or as circuit netlists with transistor sizes. The sensor's sensitivity is ≈ -0.8 % / ^{0}C; the temperature range is -50 … +150 ^{0}C; the accuracy is ≈ ± 2^{0}C for (0…120 ^{0}C). Two last parameters depend on the process.

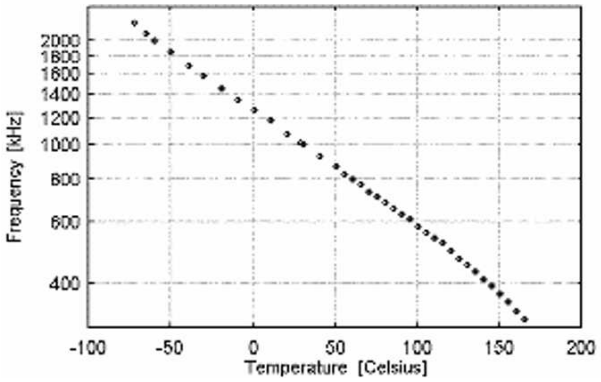


Figure 11. Sensor Characteristic (Output Frequency vs. Temperature).

If temperature sensors described above are inserted into a chip design, additional circuitry must be implemented in order to provide the access to such sensors [21]. Built-in temperature sensors can be combined with other built-in test circuitry. The boundary scan architecture [23] is suitable for monitoring temperature sensors. This architecture has led to the standard IEEE 1149.1 and is well suitable for incorporating frequency output temperature sensors. The end user has to encapsulate the prototype thermal measurement chips (TMC-s) into the package that has to be characterized by thermal measurements. The TMC9 and TMC81 chips are based on the same basic cell that is mainly covered by dissipating resistors and also contains a CMOS frequency output temperature sensor. These basic cells are organized into arrays of different size. Control of measurements is provided via a digital interface or standard boundary scan interface providing a digital read-out of measured temperature values. The TMC81 layout is shown in Figure 12.

Another interesting fully-CMOS temperature sensor designed by this research team is based on the temperature dependence of the internal thermal diffusion constant of the silicon. In order to measure this diffusion constant an oscillating circuit is used in which the frequency-determining element is realized by a thermal delay line. The temperature difference sensors used in this delay line are the Si-Al thermopiles. This circuit is the Thermal-Feedback Oscillator (TFO). The frequency of this oscillator is directly related to the thermal diffusion constant and thus to the temperature. This constant can be defined as

$$D_{th} = \lambda / c, \tag{3}$$

where λ is the thermal conductivity and c is the unit-volume heat capacitance. This diffusion constant shows a reasonably large (-0.57 % / ^{0}C) temperature dependence in the silicon. In order to measure this diffusion constant the oscillating circuits were used in which the frequency-determining feedback element is realized by a thermal time-delay line. If the feedback element is a thermal two-port (thermal delay line) then the frequency of the oscillator is directly related to the thermal diffusion constant and thus shows similar temperature dependence as the thermal diffusion constant [18].

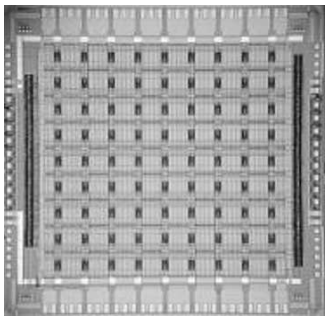


Figure 12. TMC81 Chip Layout (Reproduced from http://www.micred.com).

The thermal delay line requires, however, a significant power input. Because of this disadvantage, the circuit is not really suitable for on-line monitoring purposes [19]. However, these sensors and the sensor principal can probably be used for other applications.

Among a lot of temperature sensors it is necessary to mark the low power consumption smart temperature sensor SMT160-30 from *Smartec* (The Netherlands) [24]. It is a three terminal full silicon integrated temperature sensor, with a duty-cycle output. Two terminals are used for the power supply of 5 V and the third terminal carries the output signal. A duty cycle modulated output is used because this output is interpretable by a microprocessor without AD converter, while the analogue information is still available. The duty cycle of the output signal is linearly related to the temperature according to the equation:

$$D.C. = \frac{t_p}{T_x} = t_p \cdot f_z = 0.320 + 0.00470 \cdot t, \tag{4}$$

where t_p is the pulse duration; T_x is the period; f_x is the frequency; t is the temperature in ^{0}C. This sensor is calibrated during the test and burn-in of the chip. The sensor characteristic is shown on Figure 13.

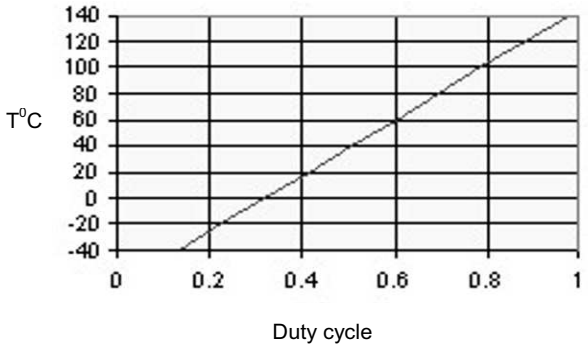


Figure 13. SMT 160-30 Sensor Characteristic (Temperature vs. Duty-cycle).

If the output signal is measured with an analogue device, the temperature can be obtained as well. The mean voltage (and the RMS value as well) is directly proportional to the duty cycle and the supply voltage. So V_{out}/V_{dd} represents t_p/T_x just as well. In this way the temperature can be measured in an analogue way as well as digitally equally simple and accurate.

The SMT160-30 (TO18 model) has an overall accuracy of ± 0.7 °C in the range from -30 °C to +100 °C and an accuracy of 1.2 °C from -45 to +130°C. This makes the sensor especially useful in all applications where "human" (climate control, food processing etc.) conditions are to be controlled. The relative error is 0.47 %, the frequency range is 1 ÷ 4 kHz.

The one wire CMOS output of the sensor can be connected to all kinds of microcontrollers and handle cable length up to 20 meters. This makes the SMT160-30 very useful in remote sensing and control applications.

This smart temperature sensor represents a significant totally new development in the transducer technology. Its novel on-chip interface meets the progressively stringent demands of both the consumer and industrial electronics sectors for a temperature sensor directly connectable to the microprocessor input and thus capable of direct and reliable communication with microprocessors.

In application where more sensors are used, easy multiplexing can be obtained by using more microprocessor inputs or by using simple and cheap digital multiplexers.

The next specialized temperature sensor is interesting due to the high metrology performances and, first of all, its high accuracy. It is the SBE 3F

temperature sensor with the initial accuracy of 0.001 ^{0}C and typical stable to 0.002 ^{0}C per year [25]. It is used for custom oceanographic profiling systems or for high-accuracy industrial and environmental temperature-monitoring applications. Depth ratings to 6 800 and 10 500 meters (22 300 and 34 400 ft) are offered to suit different application requirements.

The sensing element is a glass-coated thermistor bead, pressure-protected in a thin-walled 0.8 mm diameter stainless steel tube. Exponentially related to the temperature, the thermistor resistance is the controlling element in the optimized Wien Bridge oscillator circuit. The resulting sensor frequency is inversely proportional to the square root of the thermistor resistance and ranges from approximately 2 to 6 kHz, corresponding to temperatures from -5 to +35 ^{0}C.

Speaking about digital and quasi-digital output IC and smart temperature sensors it is necessary to mention interesting developments of such companies as *Analog Devices*, *Dallas Semiconductor*, *National Semiconductor*, *Maxim, Texas Instruments* and some others.

The TMP03/TMP04 are monolithic temperature sensors from *Analog Devices* [26] that generate a modulated serial digital output that varies in the direct proportion to the temperature of the device. The TMP03 is a complete temperature data-acquisition system on a monolithic silicon chip. Including a silicon-based sensor, internal voltage reference, and sigma-delta A/D converter, it fits in a 3-pin (power, common, and output) TO-92 transistor package. Its digital output is a low-frequency variable-duty-cycle serial data stream, available at an open collector with 5-mA sink-current available. A companion product, the TMP04, is identical but has a CMOS/TTL-compatible output. The quiescent power requirement is a modest 1.3 mA at +5 V (4.5 to 7V range).

The TMP03's digital output allows multi-channel systems to be constructed easily (additional sensors share a single-channel low-cost digitally-multiplexed decoder).

Typical applications of the TMP03 include isolated sensors, environmental control systems, computer thermal monitoring, thermal protection, industrial process control, and power-system monitors.

Analog Devices bandgap references generate both a constant voltage and a PTAT (proportional to absolute temperature) voltage. In the TMP03, these are applied as inputs to a first-order sigma-delta modulator. The device output is a 35-Hz (nominal) accurately mark-space-modulated digital signal that is insensitive to frequency. The TTL/CMOS compatible output allows the TMP03 to interface directly to standard logic.

Thus the TMP03 and TMP04 are well suited to interface directly to a microcontroller timer/counter input port and programmable logic arrays. The TMP04 provides a high-output-current logic output capable of driving a load capacitance of 1000 pF with minimal loss of switching-edge definition.

Since it is completely self-contained, the TMP03/TMP04 has specifications that are close to the final system specifications. A single temperature accuracy specification for the TMP03/04 combines errors due to the sensor's transfer function, signal conditioning and conversion. Typical accuracy (-25 to +100°C) is to within 1.5% (4% max), with non-linearity of 0.5°C and power supply sensitivity of 0.7°C/V (1.2°C/V max). The device's operating temperature range is -55°C to +150°C.

The digital output format of the TMP03/04 design allows this temperature transducer to be located away from the host computer system without degrading system accuracy; and the 35-Hz low-frequency output further insures data integrity over long distances. Cable capacitance between the TMP03 and its host computer will of course round the rising and falling edges of the square-wave output, but delays of the order of microseconds add negligible error relative to a 29-ms clock period. In most applications, temperature is a slowly changing variable, and a 35-Hz carrier has little effect on the measurement dynamic accuracy.

If a 100° higher output frequency, say 3 kHz, had been chosen, the output circuitry would need to drive high currents into the load capacitance to keep the logic transitions acceptably short; a 1-ms asymmetry between the rise and fall times would have added about 1°C error. High output current requirements also increase the required supply current. An alternative solution for high frequency, low-level transducer output: i.e., adding a local RS-232 (or RS-485) interface to drive a long cable again increases the remotely supplied current required.

The TMP03 sensor uses mark-space ratio modulation, embodying the relationships

$$T_{\exp C} = 235 - \frac{T_H}{T_L} \cdot 400 \tag{5}$$

$$T_{\exp F} = 455 - \frac{T_H}{T_L} \cdot 720, \tag{6}$$

Where T(H) and T(L) are the high and low periods of the square-wave output.

With the basic TMP03 sensor error specs, errors introduced in measuring T(H) and T(L) are the only other parameters needed to determine system accuracy. For example, if T(H) and T(L) are measured using a 125 kHz clock frequency and 12-bit counters, gated by the edges of the square-wave output, the quantization error is less than 0.5°F.T(H) and T(L) may be conveniently measured using discrete counters, programmable logic arrays, or from a microprocessor with an on-board timer/counter port. If absolute temperature is required, it can be calculated using a microprocessor or PC.

Digital output temperature sensors from *Analog Devices* are adduced in Table 1 The AD7818 (single-channel) and AD7817 (4-channel) [26] are on-chip temperature sensors with 10-bit, single and four channel A/D converters. These devices contain an 8 ms successive-approximation converter based around a capacitor DAC, an on-chip temperature sensor with the accuracy of ± 1^0C, an on-chip clock oscillator, inherent track-and-hold functionality and an on-chip reference (2.5 V ± 0.1%).

Table 1. Digital Temperature Sensors from Analog Devices.

Model	Description	Interface	# ADC Channel	Accuracy	Temperature Range
AD7416	10-bit TDC + prog. OTI	I^2C	None	±2°C	-55°C to 125°C
AD7417	AD7416+4 ADC input channels	I^2C	4	±2°C	-55°C to 125°C
AD7418	AD7416+1 ADC input channel	I^2C	1	±2°C	-55°C to 125°C
AD7816	SPI version of AD7416	SPI	None	±2°C	-55°C to 125°C
AD7817	SPI version of AD7417	SPI	4	±1°C	-55°C to 125°C
AD7818	SPI version of AD7418	SPI	1	±2°C	-55°C to 125°C
AD7814	10-Bit TDC in SOT23	SPI	None	±2°C	-55°C to 125°C
AD7414	10-Bit TDC in SOT23 prog OTI[1]	I^2C/Smbus (SMBus Alert)	None	±2°C	-40°C to +85°C
AD7415	10-Bit TDC in SOT23	I^2C/SMbus	None	±2°C	-40°C to +85°C

Dallas Semiconductor offers a broad line of factory-calibrated 1-, 2-, 3-Wire® or SPI buses temperature sensors/thermometers that can provide straightforward thermal management for a vast array of applications. This unparalleled product line includes a variety of "direct-to-digital" temperature sensors that have the accuracy and features to easily improve system performance and reliability [27]. These devices reduce the component count and the board complexity by conveniently providing digital data without the need for dedicated A-to-D converters. These sensors are available with accuracies ranging from ± 0.5 ^{0}C to ± 2.5 ^{0}C (guaranteed over wide temperature and power-supply ranges), and they can operate over the temperature range of –50 ^{0}C to +125 ^{0}C.

The conversion time range for the temperature into a digital is 750 ms ÷ 1.2 s. The 1-Wire and 2-Wire devices have the multi-drop capability, which allows multiple sensors to be addressed on the same bus.

In addition, some devices (DS1624, DS1629 and DS1780) combine the temperature sensing with other valuable features including EEPROM arrays, real-time clocks, and the CPU monitoring. One more interesting "smart" feature of Dallas Semiconductor's temperature sensors is expandable from 9 to 13 bits or user configurable to 9, 10, 11, or 12 bits resolution.

Dallas Semiconductor's DS1616 Temperature Data Recorder with the 3-Input Analog to Digital Converter adds the potential for three powerful external sensors to the base design of the DS1615 Temperature Data Recorder. It permits to logging of not only the temperature, but also the humidity, the pressure, the system voltage, external temperature sensors, or any other sensor with the analog voltage output. The DS1616 provides all of the elements of a multi-channel data acquisition system on one chip. It measures the selected channels at user-programmable intervals, then store the data and a time/date stamp in the non-volatile memory for the later downloading through one of the serial interfaces.

The DS1722 Digital Thermometer and Thermostat (-55°C to +120°C temperature range and ±2.0°C) with SPI/3-Wire Interface provides temperature readings which indicate the temperature of the device. No additional components are required; the device is truly a temperature-to-digital converter. Temperature readings are communicated from the DS1722 over a Motorola SPI interface or a standard 3-wire serial interface. The choice of interface standard is selectable by the user. For applications that require greater temperature resolution, the user can adjust the readout resolution from 8 to 12 bits. This is particularly useful in applications where thermal runaway conditions must be detected quickly. Applications for the DS1722 include personal computers/ servers/workstations, cellular telephones, office equipment, or any thermally-sensitive system [28]. Some other temperature sensors from Dallas Semiconductor are adduced in table 2.

National Semiconductor also proposes some digital temperature sensors [29] with different temperature ranges from –55 ^{0}C up to +150 ^{0}C: SPI/MICROWIRE plus the sign digital temperature sensor LM70 (10-bit) and LM74 (12-bit); the digital temperature sensor and the thermal watchdog with the two-wire (I^2C™ Serial Bus) interface LM75 (±3 ^{0}C); digital temperature sensors and the thermal window comparator with the two-wire interface LM76 (±1 ^{0}C), LM77 (±1.5 ^{0}C), LM88 (±3 ^{0}C, factory programmable dual remote-diode temperature sensor with 3 digital comparators) and LM92 (±0.33 ^{0}C). The sensors LM70, LM74 and LM75 include the delta-sigma analog-to-digital converter.

The window-comparator architecture of the sensors eases the design of the temperature control systems conforming to the ACPI (Advanced Configuration and Power Interface) specification for personal computers.

Table 2. Digital Temperature Sensors from Dallas Semiconductor.

Model	Description	Accuracy
1-Wire Digital Output Temperature Sensors		
DS18B20	1-Wire™ temperature sensor with alarm function. 64-bit ROM address allows multiple DS18B20 to reside on the same 1-Wire bus	±0.5°C (Flip Chip is ±2.0°C)
DS1821	1-Wire temperature sensor and standalone thermostat	±1.0°C
DS1822	1-Wire temperature sensor with alarm function. 64-bit ROM address allows multiple DS18B22 to reside on the same 1-Wire bus	±2.0°C
2-Wire SMBus Temperature Sensors		
DS1631	9-12 bit temperature sensor and thermostat	±0.5°C
DS1721	9-12 bit temperature sensor and thermostat	±1.0°C
DS1731	9-12 bit temperature sensor and thermostat	±1.0°C
DS1775	9-12 bit temperature sensor and thermostat	±2.0°C
DS75	9-12 bit temperature sensor functionally compatible with LM75	±2.0°C
SPI/3-Wire Temperature Sensors		
DS1626	3-wire, 9-12 bit temperature sensor with standalone thermostat function	±0.5°C
DS1722	SPI or 3-wire, 8-12 bit temperature sensor compatible with 1.8V logic	±2.0°C

Maxim Integrated Products provides the largest selection of thermal measurement integrated circuits (ICs) in the industry [30]. *Maxim's* single-wire output digital temperature sensors provide high-performance temperature sensing with a choice of period, frequency, or pulse-width modulation (PWM) output. *Maxim's* 2-wire SMBus™ temperature sensors provide high reliability protection for CPUs, FPGAs, and ASICs. They include features such as SMBus time-out, fault queue, and fail-safe hard wire connection. Models are available in single and multi-channel versions with a variety of performance and alarm configurations. Maxim's pulse-width modulation (PWM) digital temperature sensors provide high-performance with a single-wire output. They are available in variety of performance, supply voltage, and output configurations. *Maxim's* SPI™ digital temperature sensors provide high-performance temperature measurement and are available in a variety of product configurations. The digital temperature sensors from *Maxim* are adduced in table 3.

Table 3. Digital Temperature Sensors from MAXIM.

Model	Description	Accuracy
1-Wire Digital Output Temperature Sensors		
MAX6666/ MAX6667	Single wire, 35Hz PWM, open-drain or push-pull outputs (90μA current typ).	±1.0°C
MAX6672/ MAX6673	Single wire, 1.4KHz PWM, open-drain or push-pull outputs (150μA current typ.)	±0.5°C
MAX6675L/H	Allows μP to interface with up to eight sensors on a single control wire (150μA current typ).	±3.0°C
MAX6676/ MAX6677	Single wire, 1.8KHz PWM, open-drain output (80μA current typ.)	±1.0°C
2-Wire SMBus Temperature Sensors		
MAX6625/ MAX6626	9-bit/12-bit temperature sensor with digital OVERT-bar temperature alarm	±1.0°C
MAX6633/ MAX6635	9-bit/12-bit temperature sensor with 0, 1 or 2 comparator outputs.	±1.0°C
MAX6648/ MAX6692	Remote/local temperature sensor with ALERT-bar and OVERT-bar outputs.	±0.8°C
MAX6649	145°C remote/local temperature sensor with ALERT-bar ar OVERT-bar outputs.	±1.0°C
MAX6652	System monitor includes temperature sensors and four voltage measurements with ALERT-bar output.	±1.0°C
MAX6654	Remote/local temperature sensor with ALERT-bar output.	±1.0°C
MAX6655/ MAX6656	Dual remote temperature sensor, one internal sensor, and four voltage monitors.	±1.0°C
MAX6657/ MAX6658	Remote/local temperature sensor with ALERT-bar and OVERT-bar outputs	±1.0°C
MAX6659	Remote temperature with dual OVERT-bar outputs.	±1.0°C
MAX6680/ MAX6681	Fail-safe pin-programmable ALERT-bar threshold, factory-set OVERT-bar threshold.	±1.0°C
MAX6690	Remote/local temperature sensor with ALERT-bar output.	±2.0°C
SPI/3-Wire Temperature Sensors		
MAX6627/ MAX6628	12-bit remote temperature sensor with SPI interface.	±1.0°C
MAX6629/ MAX6632	12-bit local temperature sensor with SPI interface and either a 0.5 second conversion rate (200μA typ supply) or an 8 second (35μA typ supply).	±1.0°C
MAX6661	12-bit temperature sensor with linear fan control, fan excitation, and SPI interface.	±1.0°C
MAX6662	12-bit local temperature sensor with two alarms and SPI interface.	±1.0°C
MAX6674	Thermocouple-to-digital converter with 10-bit, 0.125°C resolution from 0°C to 128°C, SPI interface.	±2.0°C
MAX6675	Thermocouple-to-digital converter with 12-bit, 0.25°C resolution from 0°C to 1024°C, SPI interface.	±3.0°C
MAX6682	Thermistor-to-digital converter, 10-bit, 1 LSB accuracy, SPI interface.	±3.0°C

The MAX6666/MAX6667 are high-accuracy, low-cost, low-power temperature sensors with a single-wire output. The MAX6666/MAX6667 convert the ambient temperature into a ratiometric PWM output with temperature information contained in the duty cycle of the output square wave. The MAX6666 has a push-pull output and the MAX6667 has an open-drain output.

The MAX6666/MAX6667 operate at supply voltages from +3V to +5.5V. The typical unloaded supply current at 5.0V is 200μA. Both devices feature a single-wire output that minimizes the number of pins necessary to interface with a microprocessor (μP). The output is a square wave with a nominal frequency of 35Hz (± 20%) at $+25^0$C. The output format is decoded as follows:

$$T^0C = 235 - \frac{400 \cdot t_1}{t_2}, \tag{7}$$

where t_1 is the fixed with a typical value of 10 ms and t_2 is the modulated by the temperature . The MAX6666/MAX6667 operate from -40^0C to $+125^0$C.

Texas Instruments manufactures the programmable digital temperature sensors TMP122 and TMP124 with SPI interface and 1.5÷2 ^{0}C accuracy. It has 9-12 bit temperature resolution and programmable temperature thresholds. The temperature range is -40^0 C to $+125^0$ C. Programmable resolution, programmable set points and shut down function provide versatile for any application. The TMP122 and TMP124 are ideal for extended thermal measurement in a variety of communication, computer, consumer, environmental, industrial, and instrumentation applications [31].

Applied Microsystem Ltd. has proposed the smart temperature sensor with 0.0050 C accuracy and RS-232 ASCII communications. It is a low cost, real-time alternative ideal for integration into others' systems [32].

Thermometrics is now manufacturing direct-to-digital temperature sensors. These units provide a 9-bit digital output over the temperature range -55° to +125°C (-67° to +257°F) in increments of 0.5°C (0.9°F). 1 wire, 2 wire, 3 wire and SPI buses are available. Each unit contains a unique serial number integral to the silicon chip so that multiple units can be wired onto a single 1-wire bus, and each is recognized by its own serial number. This feature suggests applications in HVAC and other areas where multiple temperatures need to be economically monitored by a central station.

Interface (both hardware and software) are available to provide readout and alarm functions right through the PC's serial port. Accuracy is ± 0.5°C from -10° to +85°C with 9-bit resolution [33].

Another example of the digital output sensor is +GF+ SIGNET 2350 with the temperature range from -10 to +100 ^{0}C and the accuracy ± 0.5 ^{0}C.

The temperature sensor's digital output signal allows for wiring distances between sensor and temperature transmitter of up to 61 m. An integral adapter allows for the integration of the sensor and the transmitter into the compact assembly.

4.2 Pressure Sensors

Similarly to temperature sensors, pressure sensors are also very widely spread. Pressure transducers are found in numerous OEM applications, and they are used widely in process control. The introduction of the microcontroller has increased the functionality and expanded the use of pressure transducers over the past 15 years.

In Europe, the first truly integrated pressure sensor was designed in 1968 by Gieles at Philips Research Laboratories [34], and the first monolithic integrated pressure sensor with the digital (i.e, frequency) output was designed and tested in 1971 at Case Western Reserve University [35] as a part of a program addressing biomedical applications. Miniature silicon diaphragms, with the resistance bridge at the center of the diaphragm and sealed to the base wafer with the gold-tin alloy, were developed for implant and indwelling applications.

A wide variety of pressure sensing technologies is available today. Pressure sensors convert the external pressure into the electrical output signal. To accomplish this, semiconductor micromachined pressure sensors use the monolithic silicon diffused piezoresistors. The resistive element, which constitutes the sensing element, resides in the thin silicon diaphragm. The applying pressure to the silicon diaphragm causes its deflection and changes the crystal lattice strain. This affects the free carrier mobility, resulting in a change of the transducer's resistance, or piezoresistivity. The diaphragm thickness as well as the geometrical shape of resistors, is determined by the tolerance range of the pressure. Advantages of these widely spread transducers of such a type are:

- high sensitivity;
- good linearity;
- minor hysteresis phenomenon;
- small response time.

The output parameters of the diffused piezoresistors are temperature dependent and require from the device to be compensated if it's used over a wide temperature range. However, with occurrence smart sensors and MEMS, the temperature error can be compensated due to the usage of built-in temperature sensors.

Most of today's MEMS pressure transducers produced for the automotive market consist of the four-resistor Wheatstone bridge, fabricated on a single monolithic die using the bulk etch micromachining technology. The

piezoresistive elements integrated into the sensor die are located along the periphery of the pressure sensing diaphragm, at the points appropriate for the strain measurement [36].

Now the designers can choose between two architectures for the sensor compensation: the conventional analog sensor signal processing or the digital sensor signal processing. The last one is characterized by the fully digital compensation and the error-correction scheme. A very fine geometry, mixed-signal CMOS IC technologies have enabled the incorporation of the sophisticated digital signal processor (DSP) into the sensor compensator IC. The DSP was designed specifically to calculate the sensor compensation, enabling the sensor output to realize all the precision inherent in the transducer.

As it was considered in [37] 'as the CMOS process and the microcontroller/DSP technology have become more advanced and highly integrated, this approach may become increasingly popular. The debate continues as to whether the chip area and the circuit overhead of standard microprocessor designs used for this purpose will be competitive with less flexible (but smaller and less costly) dedicated DSP designs that can be customized to perform the specific sensor calibration function'. For example, the integrated pressure sensor uses a custom digital signal processor and the non-volatile memory to calibrate and temperature-compensate a family of pressure sensor elements for a wide range of automotive applications.

This programmable signal conditioning engine operates in the digital domain using a calibration algorithm that accounts for higher order effects beyond the realm of most analog signal conditioning approaches. The monolithic sensor provides enhanced features that typically were implemented off the chip (or not at all) with traditional analog signal conditioning solutions that use either laser or electronic trimming. A specially developed digital communication interface permits the calibration of the individual sensor module via connector pins after the module has been fully assembled and encapsulated. The post-trim processing is eliminated, and the calibration and the module customization can be performed as an integral part of the end-of-line testing by the completion of the manufacturing flow. The IC contains a pressure sensor element that is coprocessed in a submicron, mixed-signal CMOS wafer fabrication step and that can be scaled to a variety of automotive pressure sensing applications [37].

Now, let us give state-of-the-art and some industrial examples of modern pressure sensors and transducers. There are many different manufacturers on the modern sensor market who are producing pressure sensors and transducers with accuracy range from 1 % FS up to 0.01 % FS, frequency, digital output (RS-232/485, IEEE-488) and bus capabilities (CANbus, Modbus, HART, Fieldbus, etc.).

The major attention was given to the creation of pressure sensors with the frequency output in the USSR [38, 39]. The first of them was based on the usage of voltage-to-frequency converters and had the accuracy up to 1 %, the effective range of measuring frequencies 0 ÷ 2 kHz in the pressure range 0 ÷ 40 MPa. The second was founded on the usage of the piezoresonator. The connection of this device into the self-oscillator circuit has allowed to receive a frequency signal, proportional to the force. The relation between the measured pressure *p* and the output frequency signal *f* is expressed by the following equation:

$$p = (f - f_0)/K_p; \quad K_p = K_F \cdot S_{eff}, \tag{8}$$

where f_0 is the frequency at $p = 0$; f is the measurand frequency; K_p is the conversion factor of pressure-to-frequency; K_F is the force sensitivity factor; S_{eff} is the membrane's effective area.

The silicon pressure sensor with based on the bulk micromachining technology and VFC based on CMOS technology was described in [40]. It has 0 ÷ 40 kPa measuring pressure range, 280 ÷ 380 kHz frequency output range and the main error ± 0.7 %.

Kulite company produces the frequency output pressure transducer ETF-1-1000. The sensor provides an output, which can be interfaced directly to a digital output. The transducer uses a solid-state piezoresistive sensing element, with the excellent reliability, repeatability and accuracy. The pressure range is 1.7 ÷ 350 bar, the output frequency is 5 ÷ 20 kHz, the total error band is ± 2%. Other examples are pressure transducers VT 1201/1202 from *Chezara* (Ukraine) with 15 ÷ 22 kHz frequency output range and the standard error ± 0.25 % and ± 0.15 % accordingly.

The shining example of the sensor with $x(t) \rightarrow V(t) \rightarrow F(t)$ conversion is the pressure sensor from *ADZ Sensortechnik GmbH* (Germany). The IC LM 331 was used as VFC. The output frequency of the converter can be calculated according to the following formula:

$$f_{out} = \frac{(U_{in} - U_{offset}) \cdot R6}{2.09V \cdot R4 \cdot R8 \cdot C6}, \tag{9}$$

where

$$U_{in} = P_{abs} \cdot 0.533 \frac{V}{bar} + 0.5V \tag{10}$$

The measuring range is 0 ÷ 8.8 bar, the frequency range is 1 ÷ 23 kHz.

Honeywell manufactures pressure sensors and transducers for a variety of applications designed to provide 0.1 ÷ 0.03 % FS accuracy with digital and analog functionality [41].

Omega Engineering Inc. has announced the frequency output type (1÷6 kHz), 1 % FS accuracy microprocessor compatible pressure transducers PX106 series. Ten pressure ranges from 0-6 psi up to 0-2 000 are available [42].

The high-accuracy (0.01 %) fibre-optic pressure transducers have been developed in *ALTHEN GmbH* by applying the optical technology to resonator-based sensors [43].

Further development of microelectronic technologies and smart sensors has declined in the rise of the high-precision (up to 0.01 % FS) digital and quasi-digital output pressure sensors and transducers. Some of them are described below.

The *Paroscientific Inc*. Digiquartz® Intelligent Transmitter [44] consists of a unique vibrating quartz crystal pressure transducer and a digital interface board in the integral package. Commands and data requests are sent via the RS-232 channel and the transmitter returns data via the same two-way bus. Digital outputs are provided directly in engineering units with the typical accuracy of 0.01 % over a wide temperature range. The use of a frequency output quartz temperature sensor for the temperature compensation yields the achievable accuracy of 0.01 % full scale over the entire operating temperature range. The output pressure is fully thermally compensated using the internally mounted quartz crystal specifically designed to provide a temperature signal. All transmitters are programmed with calibration coefficients for full plug-in interchangeability. The Intelligent Transmitter can be operated either as a stand-alone standard output pressure sensor with the display, or as a fully integrated addressable computer controlled system component. Transducers use crystalline quartz as the key sensing elements for both the pressure and the temperature because of its inherent stability and precision characteristics. The pressure sensing element is a quartz beam, which changes frequency under the axial load. The transferred force acts on the quartz beam to give a controlled, repeatable, and stable change in the resonator's natural frequency, which is measured as the transducer output. The load dependent frequency characteristic of the quartz crystal beam can be characterized by a simple mathematical model to yield highly precise measurements of the pressure and pressure related parameters. The output is a square wave frequency [44].

Over thirty full scale pressure ranges are available - from a fraction of an atmosphere to thousands of atmospheres (3 psid to 40 000 psia). Absolute, gauge, and differential transducers have been packaged in a variety of configurations including intelligent transmitters, depth sensors, portable standards, water level systems and meteorological measurement systems. Intelligent electronics have two-way digital interfaces (bi-directional RS-232 and RS-484) that allow users to adjust resolution (up to 0.0001%), update rates, engineering units, sampling commands and other operational parameters.

Other examples are intelligent pressure standards (series 960 and 970) from *Pressure Systems* [45]. In 960 series, the pressure is measured via the change in the resonant frequency of the oscillating quartz beam by the pressure-induced stress. Quartzonix™ pressure standards produce the output frequency between 30 and 45 kHz and can achieve the accuracy of ± 0.01% FS. The precise thermal compensation is provided via the integrated quartz temperature sensor used to measure the operating temperature of the transducer. The Series 970 uses a multi drop, 9 600 baud ASCII character RS-485 type interface, allowing a network of up to 31 transducers on the same bus. The output pressure measurement is user programmable for both the pressure units and update rate.

The PDCR 450 and DPS 4000 series CANbus digital output pressure sensors are produced by *Druck* [46]. Fully temperature corrected pressure readings are output as a digital word in any one of 24 engineering units, requiring no user system set-up or calibration. The integral digital electronics enhance performance to levels unmatched by traditional analogue transducers. "Smart" functions include a device self-checking and diagnostics.

The Atmodule AT2640-16A precision digital-output pressure sensor from *Atmos* uses smart sensor techniques to correct pressure readings for temperature and linearity, achieving a total error band of under 0.1% FS (0.016 PSI) over the -58 F to +185 F temperature range. The module contains an aerospace-grade silicon pressure sensor, sensor interface electronics (four-wire serial interface, three-wire user interface), a 21-bit A/D converter, EEPROM calibration memory, and a sensor signal processor [47].

The differential pressure sensors SDP1000 and SDP2000 with CMOSens® technology are available from *Sensirion.* It combines an economical yet highly precise sensor element with electronics for signal processing on a single chip (ASIC). SDP1000/SDP2000 are suited for numerous applications, such as HVAC medical applications, air and environmental protection, and portable instruments [48].

Instrumentation Northwest's AquiStar® PT2X submersible smart sensor represents the latest in state-of-the-art level measurement technology. This digital (RS485/RS232 interface Modbus protocol) output device offers improved noise immunity, thermal performance and transient protection. In addition, this device stores over 130,000 records of pressure, temperature and time data, operates with low power. The PT2X comes with INW's easy-to-use Aqua4Plus software. Aqua4Palm is available for Palm®-based handhelds.

This sensor also provides updated accuracy and resolution and simple software field calibration, which increases data integrity but reduces overall system costs [49].

4.3 MEMS Accelerometers

Another very popular silicon sensor from the mechanical signal domain is the accelerometer. The measurement of acceleration or one of its derivative properties such as vibration, shock, or tilt has become very commonplace in a wide range of products.

The first attempt at developing a piezoresistive accelerometer for the shock and vibration measurement community occurred in 1962 [50] and incorporated a patented butterfly bulk semiconductor gauge. During the 1970s, MEMS accelerometers began to become commonplace in the market. The past decade has increasingly seen MEMS accelerometers married to a separate chip containing electronic circuitry and packaged in a hermetically sealed housing. This has enabled a full-scale output signal level on the order of volts. In 1968 [50], a patent was filed to incorporate a 2-wire IC (FET) into a piezoelectric accelerometer. In time, all manufacturers of piezoelectric accelerometers endorsed this principle and more complex circuits were developed. Over the past 10 years, the demand for piezoelectric accelerometers with integral electronics has continued to increase [51].

The focus of the 1990s, guided by IEEE (Institute of Electrical and Electronics Engineers) standards, is smart transducers. IEEE 1451.2 provides a definition of transducer electronic data sheets (TEDS) that are capable of interfacing with one or more transducers in individual smart transducer interface modules (STIMs). This standard will allow the transducer industry to design stand-alone STIMs for marketing to military or aerospace system integrators or end test users. The STIM could integrate the transducer(s), signal conditioning, digitizers, and address logic into a single package. What is the thermal sensitivity change, thermal zero correction, last calibration, and so forth of a given transducer? The TEDS could contain this information. The signal from a transducer could be corrected in near real-time. IEEE P1451.3 and P1451.4, when complete, are intended to move this technology further toward the marketplace. To envision the implications of smart transducers in the aerospace market, it is necessary only to imagine the structural dynamics qualification of a new airframe. Not only would the opportunity for misidentification of channels be lessened, but also much of the thousands of feet of cabling necessary for signal transmission could be eliminated [51].

The types of a sensor used to measure the acceleration, shock, or tilt include the piezo film, the electromechanical servo, the piezoelectric, the liquid tilt, the bulk micromachined piezoresistive, the capacitive, and the surface micromachined capacitive. Each has distinct characteristics in the output signal, the development cost, and the type of the operating environment in which it functions best of all [52]. The type, which has been used for many years, is the piezoelectric and which is relatively new is the surface micromachined capacitive. To provide useful data, the first type of

accelerometers requires the proper signal conditioning circuitry. For the last years, the working range of these devices has been broadened to include frequencies from 0.1 Hz to above 30 kHz.

Capacitive spring mass accelerometers with the integrated electronics that do not require external amplifiers are proposed by *Rieker Inc*. These accelerometers of *Sieka* series are available with the analog DC output, digital pulse-width modulated, or frequency-modulated outputs [53].

The surface micromachined products provide the sensor and the signal conditioning circuitry on the chip, and require only a few external components. Some manufacturers have taken this approach one step further by converting the analog output of the analog signal conditioning into a digital format such as a duty cycle. This method does not only lift the burden of designing of the fairly complex analog circuitry for the sensor, but also reduces the cost and the board area [52].

A very simple circuit can be used to measure the acceleration on the basis of ADXL202/210 accelerometers from *Analog Devices*. Both have the direct interface to popular microprocessors and the duty cycle output with 1 ms acquisition time [54]. For interfacing of the accelerometer's analog output (for example, ADXL05) with microcontrollers, *Analog Devices* proposes acceleration-to-frequency circuits based on AD654 VFC to provide a circuit with a variable frequency output. A microcontroller can then be programmed to measure the frequency and compute the applied acceleration [54].

The ADXL202E is a 2-axis acceleration sensor on a single IC chip, with a measurement range of ±2 *g*. The ADXL202 can measure both dynamic acceleration (e.g., vibration) and static acceleration (e.g., gravity).

The outputs are duty-cycle modulated (DCM) signals whose duty cycles (ratio of pulse width to period) are proportional to the acceleration in each of the 2 sensitive axes. These outputs may be measured directly with a microprocessor counter, requiring no A/D converter or glue logic. The DCM period is adjustable from 0.5 ms to 10 ms via a single resistor (R_{SET}). If an analog output is desired, an analog output proportional to acceleration is available from the X_{FILT} and Y_{FILT} pins, or may be reconstructed by filtering the duty cycle outputs.

The dual (X, Y) axis accelerometer inclinometer KXG20 from *Kionix* also has a duty-cycle output [55]. This silicon micromachined linear accelerometer consists of a sensor element and an ASIC. The sensor element functions on the principle of differential capacitance. Acceleration causes displacement of a silicon structure resulting in a change in capacitance. An ASIC, using a standard CMOS manufacturing process, detects and transforms changes in capacitance into a duty-cycle signal, which is proportional to acceleration. The sense element design utilizes common mode cancellation to decrease errors from process variation and environmental stress.

ST Microelectronics' LIS3L02 linear accelerometer is the first one-chip three-axis MEMS-based accelerometer sensor with an interface chip [56].

The Dual Axis MEMS - based accelerometer MXA2500U has been introduced by *MEMSIC*. The device includes the integrated circuits and sensor on one chip based on standard CMOS process. The layout is shown in Figure 14.

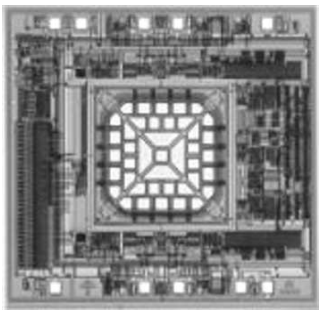


Figure 14. MXA2500U Accelerometer Layout
(Reproduced from *Sensors & Transducers e-Digest*, No.5, May 2002).

Accelerometer's features include the continuous self-test and custom programmable specifications [57].

Self-testing is a desirable requirement in certain safety applications such as crash detection for airbag actuation. The self-test procedure for a micromachined silicon accelerometer realised using a commercially available microprocessor is described in [58]. The self-test procedure must be performed at resonance and the microprocessor is used to identify the individual resonant frequency of each device and confirm the operation of the PZT elements. The microprocessor could also be used in the future to fully test and calibrate the device thereby ensuring correct and accurate operation.

Summit Instruments Platinum Series Smart Accelerometer Systems are complete, easy-to-use, user-configurable acceleration measurement systems containing one, two or three accelerometers, a temperature sensor, signal processor, IRIG-106 PCM encoder, RS-485 interface and three analog outputs in a small, easy-to-install package [59].

All channels are sampled simultaneously to avoid data skewing. The digital signal processor takes 16-bit samples, filters, ranges, and calibration compensates at up to 42,500 samples/sec/channel. Digital data can be streamed out at up to 3 Mbit/sec.

The output range, filter frequency and calibration of each channel, as well as telemetry configuration, can be set by the user via the RS-485 command processor. The CRC-16 error checking is used to ensure command and data integrity.

The built-in temperature sensor can be used by critical applications to correct for any residual temperature effects.

For a frequency output sensor, *Honeywell* produces the Accelerex RBA-500 accelerometer [60]. It is primarily used to supplement GPS navigation systems. It's a good choice where frequency output, high-g, small size, low power, and light weight are necessary. The acceleration is measured as a function of the frequency difference between two vibrating quartz beams. The output of the Accelerex accelerometer is also thermally compensated through the use of an internal temperature sensor and Honeywell-supplied coefficients; when integrated over time, delta velocity is directly provided to the user's system.

4.4 Rotation Speed Sensors

There are many known rotation speed sensing principles and many commercially available sensors. The overwhelming majority of such sensors are based on the Hall-effect and magnetoresistors sensing principles or inductive sensing principle. According to the nature, rotation speed sensors are from the frequency-time domain. Pulses are generated on its output. Its frequency is proportional to the measured parameter. The measuring frequency of the rotation speed is given by

$$n_x = f_x \cdot \frac{60}{Z}, \tag{11}$$

where Z is the number of modulation rotor's (encoder's) gradations.

For modern applications the rotation speed sensor should provide the digital or the quasi-digital output compatible with standard technologies. This means that the sensor and the signal processing circuitry (the microcontroller core) can be realized in the same chip. An excellent solution in many aspects is when this signal is a square-wave output of an oscillator the frequency of which is linearly dependent on the rotating speed and carries the information about it.

The semiconductor active position sensor of relaxation type designed by authors together with *Autoelectronic* company (Kaluga, Russia) can serve as an example [61, 62]. It was developed on the basis of the crankshaft position sensor. Its principle of action is based on the effect of the continuous suppression of oscillations of the high frequency generator by passing of each metal plate of the modulating rotor in front of the active sensor element and its subsequent resumption. Due to that, the rectangular pulses with the constant amplitude ($+V_{cc}$) are continuously formed on the sensor output. The frequency of these pulses is proportional to the rotating speed. If the metal

tooth of the rotor-modulator comes nearer to the active element (generator coil), the logic level "1" ($+V_{cc}$) is formed on the sensor output. When the active sensor part appears between the teeth, the logic level "0" is formed on the output. Thus, the active sensor forms the pulsing sequence, the frequency of which is proportional or equal to the rotating speed. This sensor does not require any additional buffer devices for the tie-in measuring system and has a very easy interface to be integrated in the microsystem. Besides, the sensor meets to requirements of the technological compatibility with other components of the microsystem.

By the sensor design, the “chip & wire” technology was used, which combines the advantages of both monolithic and hybrid integrated technologies. All electronics was realized in a single chip, only the inductance, two resistors and the stabilitron were implemented in accordance with the hybrid technology.

The comparative features of the modern non-contact sensors of different principles of function are adduced in Table 4. Here sensors A5S07/08/09 are made by *BR Braun* (Germany), DZXXXX – by *Electro Corporation* (USA); VT1855, OO020 - by *NIIFI* (Penza, Russia); 4XXXX- by *Trumeter* (UK); LMPC- by *Red Lion Controls* (USA).

Table 4. Comparative Features of Non-Contact Sensors of Rotation Speed.

Sensors	**Frequency Range, *kHz***	**Supply Voltage, *V***	**Current Consumption, *mA***	**Type**
ASRS	0 ÷ 50	4.5 ÷ 24	7 ÷ 15	active
A5S07	0.5 ÷ 25	8 ÷ 28	15 + load current	hall-effect
A5S08/09	0.5 ÷ 25	8 ÷ 25	15	hall-effect
DZ375	0 ÷ 5	4.5 ÷ 16	20 ÷ 50	magnetic
DZH450	0 ÷ 5	4.5 ÷ 30	20	hall-effect
DZP450	1 ÷ 10	4.5 ÷ 16	50	hall-effect
VT1855	0.24 ÷ 160	27	3	inductive
OO 020	0.24 ÷ 720	27	100	photo
4TUC	0.3 ÷ 2	10 ÷ 30	200	mag./inductive
4TUN	0.3 ÷ 2	6.2 ÷ 12	3	mag./inductive
45515	0.002 ÷ 30	25	20	hall-effect
LMPC	up to 10	9 ÷ 17	25	mag./inductive

Active, magnetic and Hall-effect sensors are more suitable for the determination of the object status “Stop” (a shaft is stopping). The advantage of active semiconductor sensors is the possibility of the operation with non-magnetic modulating rotor’s teeth (steel, copper, brass, aluminium, nickel, iron). Therefore, the modulating rotor can be made of plastic and its teeth – of the metallized coating. It essentially raises the manufacturability and decreases the cost value. With the exception of the non-contact rotation speed sensing, such sensors can be used like an angular position sensor, a position sensor, a metallic targets counter and an end-switch. In addition, a

smart sensor on this basis allows to realise the measurement of the rotation acceleration.

Active semiconductor sensors are not influenced by run-out and external magnetic fields in comparison with the usage of Hall-effect sensors. By usage of the Hall-effect sensors, it is also necessary to take into account the availability of the initial level of the output signal between electrodes of the Hall's element by absence of the magnetic field and its drift. It is especially characteristic for a broad temperature range. A Hall-effect rotation speed sensor needs encoders with magnetic pole teeth.

Another good example of a smart sensor for the rotation speed is the inductive position, speed and direction active micro-sensor MS1200 from *CSEM* (Switzerland) [63]. The device has the output that can switch up to 1 mA and is compatible with CMOS digital circuits, in particular with microprocessors. The frequency range is 0 ÷ 40 kHz, the air-gap is 0 ÷ 3 mm. The core is a sensor chip with one generator coil and two sets of detection coils (Figure 15).

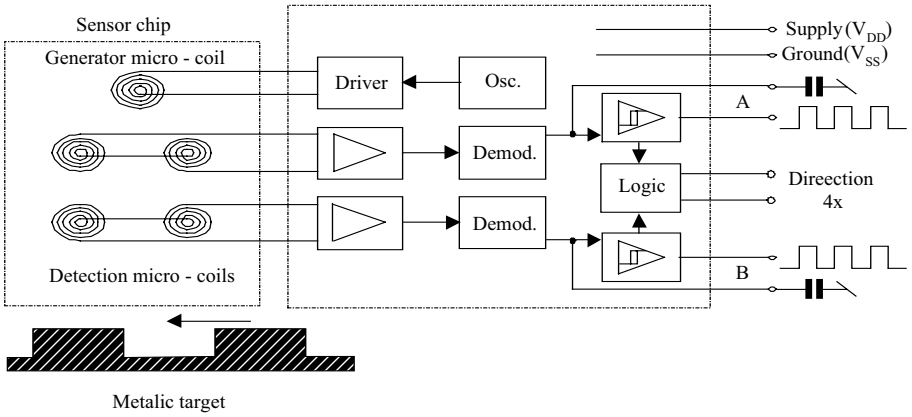


Figure 15. MS 1200 Functional Block Diagram.

The detection coils are connected in a differential arrangement, to reject the common mode signal. The sensor also includes an electronic interface, which is composed of a high frequency excitation for the generator coil and two read-out channels for the two sets of detection coils (channels A and B). The read-out electronics extract the amplitude variation of the high frequency signal due to the presence of a metallic target. The output stage is a first order low-pass filter and a comparator. For a nominal target period of 2 mm, the outputs are two channels in quadrature (A quad B) as well as a direction signal and a speed signal (4X interpolation). It is composed of two silicon chips, one for the integrated micro-coil and the other for the

integrated interface circuit. The sensor produces a two-channel digital output, as well as a direction signal.

ATS660LSB true zero-speed Hall-effect adaptive gear-tooth sensor is manufactures by *Allegro* [64]. It is a gear-tooth sensor solution for uniform teeth targets as found in today's demanding transmission applications. The ATS660LSB incorporates patented self-calibration circuitry (U.S. Pat. 5,917,320) that nulls out the effects of installation air gap, ambient temperature, and magnet offsets to provide superior timing accuracy with symmetrical targets over large operating air gaps - typical of targets used in speed-sensing applications (pitches varying from below 0.5 to over 1.2 teeth per diametric millimeter). The self-calibration at power up keeps the performance optimized over the life of the sensor. The ATS660LSB has an open-collector output for direct digital interfacing with no further signal processing required.

The integrated circuit incorporates a dual-element Hall-effect sensor and signal processing that switches in response to differential magnetic signals created by the ferrous gear teeth. The circuitry contains a sophisticated digital circuit to eliminate magnet and system offsets and to achieve true zero-speed operation. D-to-A converters are used to adjust the device gain at power on and to allow air-gap independent switching, which greatly reduces vibration sensitivity of the device.

The sensor PicoTurn-SM5.x from *Transducers Direct, LLC* is made out of a simple coil mounted on a ferrite core. In combination with a resistor this forms an R-L network. The time constant of this R-L network is measured precisely by use of a TDC (Time-to-Digital Converter). If a vane of the compressor wheel is brought in front of the sensor, the time constant changes by attenuation through eddy currents. A DSP, succeeding the TDC, takes this information to calculate the rotational speed [65].

The time measurement is done with high precision, the resolution is about 125ps. The time constant of the R-L network can be chosen to be very short. The rotational speed detectable can be very high in consequence.

Attenuation by eddy currents is best for dia- and paramagnetic materials. Thus the sensor is especially suited to detect aluminium wheels. The maximum spacing between sensor and compressor wheel depends on the shape of the vanes. It typically is in the range of 0.7 to 1.0 mm. The cable length may reach up to 3m. Guaranteed by the special measuring method, the system is insensitive to variations of the dielectric constant due to dirt. Measurement range is 200 ÷ 350 000 rpm.

4.5 Intelligent Opto Sensors

Next, we shall examine a technique of delivering the output from optical (light) sensors into the frequency-time (quasi-digital) domain. Light is a real-

world signal that is often measured either directly or used as an indicator of some other quantity. Most light-sensing elements convert light into an analog signal in the form of the current or the voltage, then a photo diode current can be converted into the frequency output. Light intensity can vary over many orders of the magnitude, thus complicating the problem of the maintaining resolution and signal-to-noise ratio over a wide input range. Converting the light intensity to a frequency overcomes limitations imposed on the dynamic range by the supply voltage, the noise and the ADC resolution.

The device to realise this is a low-cost programmable silicon opto sensors TSL230/235/245 from *Texas Instrument* with the monolithic light-to-frequency converter [66]. The output of these devices is a square wave with the frequency (0 ÷ 1 MHz) that is linearly proportional to the light intensity of the visible and short infrared radiation. Additionally the devices provide the programming capability for the adjustment of the input sensitivity and the output scaling. These capabilities are effected by a simple electronic technique, switching in different numbers of the 100 elements of the photo diode matrix. For the cost reason, the low cost microcontroller with the limited frequency range may wish to be used for the frequency-to-digital converter due to the output scaling capability. Options are an undivided pulse train with the fixed pulse width or the square wave (50% duty-cycle) divided by 2-, 10- or 100-outputs. The light levels of 0.001 to 100 000 $\mu W/am^2$ can be accommodated directly without filters [66].

Since the conversion is performed on-chip, effects of the external interference such as noise and leakage currents are minimized and the resulting noise immune frequency output is easily transmitted even from remove locations to other parts of the system. The isolation is easily accomplished with optical couples or transformers.

Texas Instrument manufactures also the TSL2550 two-wire serial SMBus ambient light sensors (ALS) [67]. Conventional silicon detectors respond strongly to infrared (IR) light that peak in the range of 700-1100 nm which the human eye does not see. Accordingly, if conventional detectors are used as ambient light sensors, inaccurate results will occur due to differences in the IR spectrum of various light sources. This problem is overcome in the TSL2550 through the use of two photodiodes without the use of an expensive optical filter. One of the photodiodes (Channel 0) is sensitive to both visible and infrared light, while the second photodiode (Channel 1) is sensitive primarily to infrared light. An integrating ADC converts the photodiode currents to a digital output which in turn is input to a microprocessor where illuminance (i.e. ambient light level) in lux is derived from the two channels to approximate the human eye response.

The linear array (LA) TSL3301 from the same manufacturer is a linear 300 dpi optical sensor array with integrated 8-bit analog-to-digital converter that operates with a supply voltage in the range of 3 to 5 volts and has a

programmable sleep mode. It is intended for high-performance, power and cost sensitive applications including scanning, rotary and linear encoders, bar code readers and OCR readers to name a few. The array consists of 102 pixels spaced on a 300 dpi centers. Associated with each pixel is a charge integrator/amplifier and sample-hold circuit. All pixels have concurrent integration periods and sampling times. Data communication is accomplished through a three-wire serial interface. The array is split into three 34 pixel zones, with each zone having programmable gain/offset levels [68].

The Smart Vision Sensor is a combination of sensor and general-purpose image RISC processor on the same semiconductor chip (Figure 16).

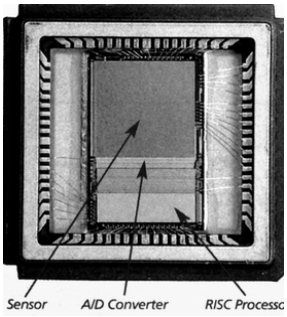


Figure 16. Smart Vision Sensor MAPP2500

MAPP2500 (Matrix Array Picture Processor) is a two dimensional Smart Vision Sensor containing a full matrix of 512x512 sensors and 512 processor units. The built in processor array handles a full row of pixels in parallel. The MAPP2500 is based on an SIMD (Single Instruction stream, Multiple Data stream) architecture containing 512 parallel processing elements. To simplify programming, all registers also have a uniform width of 512 bits [69]. The device handles all of the critical tasks found in an image processing system; image sensing, digitization and data reduction. Its performance is in many cases superior to that provided by other programmable solutions.

Another interesting examples are the integrated smart optical sensors developed in Delft University of Technology [70, 71]. Integrated on-chip colour sensors have been designed and fabricated to provide a digital output in the IS2 bus format. The readout of photodiodes in the silicon takes place in such a way that pulse series are generated with the pulse frequency proportional to the optical intensity (luminance) and the duty cycle to the colour (chrominance). The colour information is obtained using the wavelength dependence of the absorption coefficient in the silicon in the

optical part of the spectrum, so no filters are required. The counters and the bus interface have been realized in a bipolar and CMOS version with the enhanced resolution, which is being investigated.

4.6 Humidity and Moisture Sensors

Frequency-time domain humidity and moisture sensors can be created based on humidity – capacitance – frequency (or duty-cycle) converters. Taking into account that accuracy of such sensors is not so high (1÷ 5%), the analog low cost IC timer 555 or 556 can be use as a voltage-to-frequency converter in order to produce a quasi-digital sensor output.

Although an RH sensor might seem a simple device, its proper operation is highly dependent on careful humidity calibration [72].

Relative humidity sensors HS1100/HS1101 from *Humirel* (USA) are based on a unique capacitive cell. Together with two types of frequency (duty-cycle close to 50%) output circuits, these sensors are with $x(t) \rightarrow C(t) \rightarrow F(t)$ conversion. The circuits are based on the IC TLC555. Though these timers are not precise, the conversion error does not exceed 1%. The typical frequency range is 5 978 ÷ 7 285 Hz. Based on the rugged HS1101 humidity sensor, HF3223 / HTF3223 is a dedicated humidity transducer designed for OEM applications where the reliable and accurate measurement is needed. The direct interface with a microcontroller is made possible with the module's linear frequency output (8 030 ÷ 9 560 Hz) for 10 ÷ 95 RH (%) measurements [73].

E+E Elektronik (Germany) manufactures the humidity/ temperature transmitter with the frequency output (EE05, EE 25 series) [74]. It provides a pulsed signal for both the humidity and the temperature. Every microprocessor system is able to read these data by simply counting the pulses without the expensive A/D converting. The measuring range is 0 ÷100% RH, the frequency range is 62.3 ÷ 47.1 kHz and 12.5 ÷ 9.4 kHz dependent on the type and the accuracy is ± 2% RH (for EE 25 series).

Another humidity sensor was fabricated by *Galltec* (Germany). It is also the sensor with $x(t) \rightarrow C(t) \rightarrow F(t)$ conversion. One application circuit with the frequency output for FE09/1, FE09/2 and FE09/4 humidity sensing elements is designed on the discrete components (5 ÷ 95% RH corresponding to 54 ÷ 47 kHz). The second application circuit with the frequency output for the same sensing elements is based on the IC 555 (5 ÷ 95% RH corresponding to 33 ÷ 27 kHz and 3 ÷ 2 kHz for FE09/1000 sensing element).

The Soil Moisture Smart Sensor from *ONSET* integrates the field-proven ECH2O™ Sensor and a 12-bit A/D, providing ± 3% accuracy in typical soil conditions, and ± 1% accuracy with soil-specific calibration. Readings are

provided directly in volumetric water content. The sensor has relatively low sensitivity to salinity and temperature effects [75].

Following to the modern trends, Swiss-based *Sensirion AG* has introduced a new generation of integrated, digital, and calibrated humidity sensor using CMOS micromachined chip technology [76]. The new product, SH11 (Sensmitter), is a single chip relative and temperature multi sensor module with a calibrated digital output, which allows for simple and quick system integration. The device includes two calibrated microsensors (Figure 7) for relative humidity and temperature, which are coupled to an amplification, A/D conversion and serial interface circuit on the same chip.

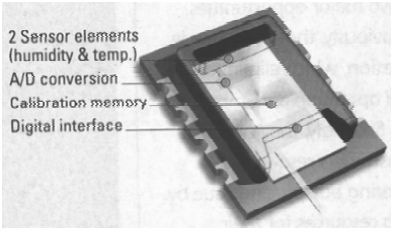


Figure 17. SH11 (Sensmitter)
(Reproduced from Control Engineering Europe, April/May 2002).

The A/D conversion makes the signal extremely insensitive to noise. A checksum generated by the chip itself is used for additional reliability. Other advantage includes very short response times (4 sec at 1/e). The sensor chip can be connected directly to any microprocessor system by means of the 2-wire interface [76]. The SH15, SHT75 series of the chip has increased accuracy up to ± 2 % RH [77].

4.7 Chemical and Biological Intelligent Sensors

An academic/industrial UK LINK project is currently being out to develop the hand held electronic nose (H^2EN) - an array of sensors simulating the human olfactory response. The approach described in [78] places the resistive sensor in the feedback loop of a 'digital' RC oscillator the output of which is a square wave with a frequency inversely proportional to the sensor resistance. The frequency measurement technique used (period counting) counts the number of cycles of the internal FPGA oscillator module n_o over a period equal to a fixed number of cycles of the sensor oscillator n_s. The current prototype H^2EN uses an embedded ELAN SC400 486SX microcontroller. In order to reduce the interference due to the noise

the acquired data are averaged by increasing n_s until the noise performance becomes acceptable. This approach achieves two objectives: firstly, random errors due to noise tend to be minimised and secondly the resolution of the final measurement is improved. Both are achieved at the expense of the measurement speed.

The Portable Electronic Nose PEN2 - intelligent chemical sensor for gases and gas mixtures – is manufactured by *AirSense Analytics* [79]. The detection of the gases is performed with an array of gas sensors. Single compounds or mixtures of gases can be identified after a training step by using the pattern generated by the sensors.

The sensors constituting the array are selected for their chemical affinities and are typically based on chemisorbing polymer films. Many of the following sensors can be used, and a serial polling of each sensor reading creates outputs. In the ideal array response each output corresponds to only one analyse or chemical compound. One of such solid-state SAW sensors is described in [80]. This sensor is an uncoated, high-Q piezoelectric quartz crystal with the natural resonating frequency of 500 MHz. Its surface temperature is controlled by a small thermoelectric element that cools the surface to promote the vapour condensation and then heats it for cleaning between analyses.

The added mass of the analyte condensing on crystal's surface lowers the vibration frequency in direct proportion to the amount of condensate. This frequency is mixed with a reference frequency, and the intermediate frequency (typically 100 kHz) is counted by a microprocessor.

The acoustic gas sensor is described in [81]. In this sensor, the sound velocity is continuously measured with the high resolution in a gas-filled cell, by controlling the frequency of an oscillator via the transit time of the sound between an ultrasound transmitter and a receiver element. The circuitry for control, signal processing and communication is based on the microcontroller PIC-17C44. This microcontroller can handle frequency input signals directly, and also provides analog output signals by means of the pulse width modulation (PWM).

Chemical signal domain sensors can be used not only for measuring different chemical quantities and compositions of mixed gases, but also for measuring such a "non-chemical" quantity like the rotation speed [82], for which measuring the electrical and magnetic signal domain sensors are usually used. Electrochemical oscillations have often been observed in iron immersed in the solution containing phosphoric acid and hydrogen peroxide. This oscillation has been interpreted by the cross linkage between electrochemical reactions and mass transport processes in the vicinity of an electrode. Therefore, the oscillation frequency is expected to reflect the flow rate around the electrode. A rod of carbon steel S50C was mounted in the rotating disc apparatus. The intersection of the rod rotated concentrically. In the mixture of 0.5M phosphoric acid and 1.5M hydrogen peroxide the highly

stable electrochemical oscillation appeared. The amplitude of the potential change was about 600 mVp-p. The oscillating frequency was a sensitive function of the rotation rate in the range of 100 rpm to 2 000.

Last five years the big attention was given to so-called intelligent biosensors (intelligent material-based sensors).It is high technology devices incorporating a biological sensing element either closely connected to, or integrated within, a transducers system [83]. The usual aim using a biosensor rather than any other sensor is to produce an electrical signal being proportional in magnitude or frequency to the concentration of a chemical or biochemical to which the biological element reacts. Biosensors combine the specificity and sensitivity of biological systems with the computing power of the microprocessors or microcomputers [83].

Biosensors represent a rapidly expanding field, at the present time, with an estimated 60% annual growth rate; the major impetus coming from the health-care industry (e.g. 6% of the western world are diabetic and would benefit from the availability of a rapid, accurate and simple biosensor for glucose) but with some pressure from other areas, such as food quality appraisal and environmental monitoring. The estimated world analytical market is about £12,000,000,000 year-1 of which 30% is in the health care area. There is clearly a vast market expansion potential as less than 0.1% of this market is currently using biosensors. Research and development in this field is wide and multidisciplinary, spanning biochemistry, bioreactor science, physical chemistry, electrochemistry, electronics and software engineering [84].

4.8 Multiparameters Sensors

Modern micro-electro-mechanical and semiconductor technologies enable the integration of microelectronics circuits and multifunctional sensor arrays fabricated on silicon substrates for detecting of different kinds of chemical and physical parameters. Such sensors are called *multiparameters sensors* and represent the single construction, allowing the perceiving of the set of physical or chemical quantities, which action is concentrated in a small, local area. It is not a simple unit of one-functional sensors, but rather its structurally advanced combination is created with the aim to reduce the chip area and share the usage of the digital or quasi-digital output, etc. Advanced sensors for detecting simultaneously of various parameters such as temperature, pressure, gas and vapour concentration, odour, acceleration, inertia, electric and magnetic fields, etc. can generate frequency-time domain output signals carrying the information provided by sensing elements. It must provide not only high signal-to-noise ratio in a wide dynamic range but also good cross-sensitivity. In general, the sensors consist of two elements: a detector and a platform that communicates with the detector through an

active interface with variable electrical, mechanical, optical, or chemical impedance. The platform must be able to generate output electrical signals carrying the information provided by the detector [85].

Modern silicon technologies offer many advantages in the design of multifunctional (multiparameters) smart sensors. It is expedient to have multi-sensor arrays in electronic noses and tongues as well as in medical implemented sensors; to join temperature, pressure and humidity detection in one multiparameters sensor for different environmental tests and a very broad range of applications (e.g. biotechnological, food industry, etc.). The perspective is also the creation of multifunctional sensing elements.

The first multiparameters (multifunctions) sensors were a natural combination of temperature and pressure or temperature and humidity sensors in the single device. The temperature sensors have been used also for temperature error compensation. Some examples of such kind of sensors are here: SBE49 FastCAT CTD integrated temperature, conductivity and pressure sensor [86]; 9401-ADM smart sensor for temperature, RH and dew/frost points from *Graftel Inc*. [87]; AH31 (Sensmitter) - fully calibrated digital relative humidity sensor, temperature and dewpoint sensor from *Sensirion* [48]; Combined pressure and temperature sensor (WEPS) with analog, frequency or digital output from Read Group [88]; WX200/WM918 temperature/humidity sensor [89], etc.

Another example of multiparameters sensor is the MPS-D sensor from *SEBA Hydrometrie GmbH* for the simultaneous measurement of six parameters: pH-value, redox-potential, c, temperature, water level and dissolved oxygen [153]. The sensor has a digital output for connection to the data logger MDS (RS-232), directly to the PC, or to data transmission units (modem, radio-modem, etc).

5. CONCLUSIONS

The huge number of frequency-time domain sensors certainly is not restricted by the review adduced above. However, from this survey it is possible to draw the following conclusions. The rapid development of microsystems and microelectronics in a whole promotes the further development of different digital and quasi-digital smart sensors and transducers. Today there are frequency-time domain sensors practically for any physical and chemical, electrical and non-electrical quantities. The obvious tendency of the accuracy increasing up to 0.01 % and above is observed. These devices are working in broad frequency ranges: from several hundredth parts of Hz up to several MHz. The extension of their "intelligent" capabilities including the intelligent signal processing is traced.

The process of miniaturization boosts the creation of multichannel, multifunction (multiparameter) one-chip smart sensors and sensors arrays.

More detailed information can be obtained from the monthly updated *Sensors Web Portal* (http://www.sensorsportal.com) in appropriate sensors sections.

REFERENCES

[1] MST Benchmarking Mission to the USA, Report of a European Visit to Microsystems Research and Production Sites in the USA; 1997, November 13-25, NEXUS, 1998.
[2] Huijising J.H., Riedijk F.R. and Gert van der Horn, Developments in Integrated Smart Sensors. Sensors and Actuators (A: Physical) 1994; 43:276-288
[3] Novitskiy P.V., Frequency Sensors Design Problem for all Electrical and Non-Electrical Quantities. Measurement Technology 1961; 4:16-21 (In Russian).
[4] Novitskiy P.V., Knorring V.G., Gutnikov V.S., Digital Measuring Instruments with Frequency Sensors. Leningrad: Energia,, 1970 (In Russian).
[5] Sensor Markets 2008: Worldwide Analyses and Forecasts for the Sensor Markets until 2008, Press Release, INTECHNO CONSULTING, Basle (Switzerland), 1999.
[6] Sensors Forecasts to 2006 & 2011. The Freedonia Group, 2000.
[7] Middelhoek S., Celebration of the Tenth Transducers Conference: The Past, Present and Future of Transducer Research and Development. Sensors and Actuators (A: Physical), 2000; 82:2–23
[8] Measurement and Automation Catalog, National Instruments, USA, 2000.
[9] Rutka M.J., *Integrated Sensor Bus*, Ph.D. Thesis, Delft University of Technology, 1994.
[10]Gerard C.M. Meijer, Concepts and Focus for Intelligent Sensor Systems. Sensors and Actuators (A: Physical) 1994; 41-42:183-191
[11]Wolffenbuttel R. F., Fabrication Compatibility of Integrated Silicon Smart Physacal Sensors. Sensors and Actuators (A: Physical) 1994; 41-42:11-28
[12]Bartek M. *Selective Epitaxial Growth for Smart Silicon Sensor Application*, Ph.D. thesis, Delft University of Technology, 1995.
[13]Kirianaki N.V., Yurish S.Y., Shpak N.O. Smart Sensors with Frequency Output: State-of-the-art and Future Development. Proceedings of IFAC Workshop on Programmable Devices and Systems; 2000, February 8-9; Ostrava, Czech Republic
[14]Bianchi R.A., F.Vinci Dos Santos, J.M. Karam, B. Courtois, F. Pressecq, S.Sifflet, CMOS-compatible Smart Temperature Sensors, Microelectronics Journal 1998; 29:627-636
[15]Riedijk F.R., J.H. Huijsing, An Integrated Absolute Temperature Sensor with Sigma-Delta A/D Conversion. Sensors and Actuators (A: Physical) 1992; A24:249-256
[16]Jay Scolio, Temperature Sensor ICs Simplify Designs. Sensors 2000; 17, No.1
[17]Krummenacher P., Smart Temperature Sensor in CMOS Technology. Sensors and Actuators 1990; A21-A23:636-638
[18]Székeley V., Márta Cs., Kohári Zs., Rencz M. New Temperature Sensors for DfTT applications; 1996 September 25-27, Budapest, Hungary.
[19]Székeley V., Rencz M., Thermal Monitoring of Self-Checking Systems. Journal of Electronic Testing: Theory and Application 1998; 12:81-92
[20]Székeley V., Rencz M., Török S., Márta Cs., Lipták-Fegó L., CMOS Temperature Sensors and Built-in Test Circuitry for Thermal Testing of ICs. Sensors and Actuators 1998; A71:10-18
[21]Székeley V., Thermal Testing and Control by Means of Built-in Temperature Sensors. Electronics Cooling 1998; 4:36-39
[22]http://www.micred.com
[23]Bleeker H., P. van den Eijnden, F.de Jong, *Boundary-Scan Test*. Dordrecht: Kluwer Academic Publishers, 1993.

[24]SMT 160-30 Temperature Sensor, *Specification Sheet*, Smartec, 1998.
[25]http://www.seabird.com/products/spec_sheets/3Fdata.htm
[26]Digital Temperature Sensor Selection Guide, Analog Devices.
[27]http://www.dalsemi.com
[28]http://dbserv.maxim-ic.com/quick_view2.cfm?qv_pk=2766
[29]http://www.national.com
[30]http://www.maxim-ic.com
[31]http://www.ti.com
[32]http://www.AppliedMicrosystems.com
[33]http://www.thermometricscorp.com/digitalsensors.html
[34]Gieles A.C.M., Integrated miniature pressure sensor, *Dutch patent application* No. 6817089, 1968.
[35]Blaser E.M., Ko W.H. and Yon E.T. A miniature digital pressure transducer. Proceedings of 24 Annual Conference on Engineering. in Medicine. & Biology; 1971 November; Las Vegas, Nevada, USA
[36]Makdessian A., Parsons M., DSSP-based Pressure Sensors. Sensors 2001; 18, No.1.
[37]Czarnocki W.S., Schuster J.P. The Evolution of Automotive Pressure Sensors. Sensors 1999; 16, No.5.
[38]Gadzhiev N.D., Garibov M.A., Kasimov M.C., Leschinskiy Y.B. Pressure Transducer with Frequency Output. Measuring Instruments and Control Systems 1979; 2:27-28 (In Russian).
[39]Maleiko L.V., Malov V.V., Rudenko A.P., Yakovlev I.V., et al. Piezoresonant Pressure Transducers. Measuring Instruments and Control Systems 1984; 9:19-21 (In Russian).
[40]Beshliu V.S., Kantser V.G., Beldiman L.N., Beshliu V.V., Coban R.A. Integral Gauge Pressure Sensor with Frequency Output Signal. Proceedings of International Semiconductor Conference (CAS'99); 1999 October 5-9; Sinaia, Romania.
[41]http://www.pressuresensing.com/
[42]http://www.omega.com/ppt/pptsc.asp?ref=PX106&Nav=preb06
[43]Paros J. M. Fiber-Optic Resonator Pressure Transducers. Measurements & Control 1992; pp.144-148.
[44]Busse D.W., Wearn R.B. Intelligent Digital Pressure Transmitters for Aerospace Applications (http://www.paroscientific.com/intdigaeroapp.htm)
[45]http://www.pressure-systems.com
[46]http://www.druck.com
[47]http://www.atmos.com/product_details/at2640.htm
[48]http://www.sensirion.com
[49]http://inwusa.com/pt2x.htm
[50]Walter P. History of the Development of the Accelerometer, 50 Years of Shock and Vibration History, The Shock and Vibration Information Analysis Center (SAVIAC) 1996, Arlington, VA:376-385
[51]Walter P. Trends in Accelerometer Design for Military and Aerospace Applications. Sensors 1999; 16, No.3
[52]Bowling S., Richey R. Two approaches to Measuring Acceleration. Sensors 2000; 17, No.2
[53]http://www.riekerinc.com/Accelerometers.htm
[54]http://products.analog.com/products/info.asp?product=ADXL202
[55]http://www.kionix.com/sheets/KXG20.pdf
[56]http://www.st.com/stonline/prodpres/dedicate/mems/products/products.htm
[57]Ultra Low Noise and Low Cost Dual Axis Accelerometer.Sensors & Transducers e-Digest 2002; No5, May (http://www.sensorsportal.com)
[58]Beedy S.P.,Grabham N.F., White N.M. Micromachined Accelerometer with Microprocessor Controlled Self-Test Procedure. Sensor Review 2001; 21:33-37
[59]http://www.summitinstruments.com/smart/X5203A.htm
[60]http://www.inertialsensor.com/rba500.shtml

[61]Deynega V. P., Kirianaki N. V., Yurish S. Y.. Microcontrollers Compatible Smart Sensor of Rotation Parameters with Frequency Output. Proceeding of 21st European Solid State Circuits Conference (ESSCIRC'95); 1995 September 19-21 Lille, France.
[62]Deynega V.P., Kirianaki N.V., Yurish S.Y. Intelligent Sensor Microsystem with Microcontroller Core for Rotating Speed Measurements. Proceedings of European Microelectronics Application Conference, Academic Session (EMAC'97); 1997 May 28 - 30 Barcelona, Spain.
[63]Digital Inductive Position Sensor 1200, Product Description, CEMS, Switzerland, 2000.
[64]http://www.allegromicro.com/sf/0660/index.asp
[65]http://www.transducersdirect.com/converters/picoturn.asp
[66]Intelligent Opto Sensor. Data Book SOYDE02B. Texas Instruments, 1996.
[67]http://www.taosinc.com/category.asp?cateid=4
[68]http://www.taosinc.com/product_detail.asp?cateid=3&proid=34
[69]http://www.ivpvision.com
[70]G.de Graaf, Wolffenbuttel R.F. Light-to-Frequency Converter Using Integrated Mode Photodiodes. Proceedings of IMTC'96; 1996 June 4-6; Brussels, Belgium.
[71]G.de Graaf, Riedijk F., Wolffenbuttel R.F. Colour Sensor System with a Frequency Output and an ISS or 12C Bus Interface. Proceedings of Eurosensors X Conference; 1996 8-10 September, Leuven, Belgium.
[72]Brown R.M. Ensuring RH Sensor Repeatability with Capacitance Testing. Sensors 1997, July.
[73]http://www.humirel.com
[74]Sensor Technology, E+E Elektronik Calalogue, Edition' 99
[75]http://www.onsetcomp.com
[76]New CMOS Humidity Sensors Provide Digital Accuracy. Control Engineering Europe 2002; April/May:18-19
[77]http://www.sensirion.com/en/sensors/humidity/default.htm
[78]Hatfield J.V., Daniels A.R., Snowden D., Persaud K.C., Payne P.A. Development of a Hand Held Electronic Nose (H^2EN). Proceedings of EUROSENSORS XIII; 1999 September 12-15; The Hague, The Netherlands
[79]http://www.airsense.com/english/produkte_pen.html
[80]Staples E. J. A New Electronic Nose. Sensors 1999*;* Vol. 16, No.5
[81]Hök B., Blückert A., Löfving J., Acoustic Gas Sensor with ppm Resolution. Proceedings of EUROSENSORS XIII; 1999 September 12-15; The Hague, The Netherlands
[82]Akimitsu Ishihara, Shukuji Asakura, A. Possible Motion Sensor with Frequency Output. Proceedings of the 18th Chemical Sensor Symposium; 1994 April 3-5; Tohoku University, Japan
[83]Powner E.T., Yalcinkaya F. Intelligent Biosensors. Sensor Review 1997; 17:107-116
[84]http://www.lsbu.ac.uk/biology/enztech/biosensors.html
[85]Ivanov D.V. Advanced Sensors for Multifunctional Applications. JOM-s 2000; 52(10)
[86]http://www.seabird.com/products/spec_sheets/49data.htm
[87]http://www.graftel.com/graftelcom/model9401/9401ADM.html
[88]http://www.readgroup.com/matre/subsea.asp
[89]http://www.qsl.net/zl1vfo/wx200/temp_sensor.htm
[90]http://www.seba.de/html/mps-d_e.htm

Chapter 2
NOVEL CONVERSION METHODS FOR SELF-ADAPTIVE SMART SENSORS

Sergey Y. Yurish, Nikolay V. Kirianaki
International Frequency Sensor Association (IFSA)

Abstract The chapter discusses classical (standard counting technique, indirect, interpolation and combined counting methods) and advanced (suitable for self-adaptive smart sensors) methods for frequency (period)-to-code conversion: reciprocal, ratiometric, constant elapsed time (CET), M/T, single (SB) and double buffered (DB), direct memory access (DMA), self-adapting methods of dependent count (MCD) and method with non-redundant reference frequency and methods for duty-cycle – to – digital and phase-shift – to – digital conversion. The comparative analysis is given.

Keywords: frequency-to digital conversion methods; duty-cycle – to – digital conversion methods; phase-shift – to – digital conversion methods; method of dependent count (MDC); method with non-redundant reference frequency.

1. INTRODUCTION

One of the main parts of frequency-time domain smart sensors is the frequency (period, duty-cycle or time interval)-to-digital converter. This unit directly influences such sensor metrological characteristics, as the accuracy and the conversion time as well as the power consumption. In spite of the fact that the frequency can be converted into digit most precisely in comparison with other informative parameters of the signal, in practice, it is not a trivial task of simple time-window counting. Besides, very often a consumer or a sensor manufacturer is not an expert in the area of frequency-time measurements. Let us consider most popular frequency-to-digital conversion discrete methods and give an analysis of metrological performances, conversion frequency ranges and requirements for realization with the aim to admonish against mistakes by the choice of this or that conversion method.

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 51-90.

2. CLASSICAL FREQUNCY-TO-DIGITAL CONVERSION METHODS

2.1 Standard Direct Counting Technique

The frequency counting scheme according to the standard counting method for conversion of the average frequency during a fixed reference gate time is shown in Figure 1 [1-3]. It is one of the most commonly used techniques for converting of the output of a sensor to a numerical quantity. Due to advantages and simplicity of implementation the method is still used today [4]. The rest circuitry, which must be used to reset the counter before the next gate period occurs, is not shown in this simplified diagram.

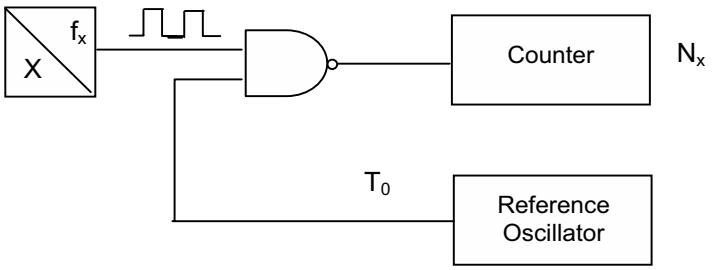


Figure 1. Simplified Diagram of Standard Counting Method.

The conversion method consists of counting a number of periods T_x of the unknown frequency f_x during the gate time (the reference time interval) T_0. The gate time is formed by dividing a reference frequency f_0 down to a suitable period. The output pulses of the sensor are simply accumulated during the time the gate signal is high. The result of conversion can be calculated by the following way:

$$N_x = T_0 / T_x = T_0 f_x \,, \tag{1}$$

If T_0 is equal to one second, for instance, the output N_x is equal to the sensor frequency f_x. In common case, the converted frequency is determined according to the following equation:

$$f_x = N_x \cdot f_0 = \frac{N_x}{T_0} \tag{2}$$

The time diagram of the described circuit (Figure 2) shows that the absence of synchronization of the beginning and end of the gate time T_0 with pulses f_x results in an error of the measurement; its absolute value is determined by values Δt_1 and Δt_2. It is easy to see, that actual time of measurement is

$$T_0^{'} = N_x \cdot T_x = N_x / f_x = T_0 + \Delta t_1 - \Delta t_2 \tag{3}$$

therefore

$$N_x = T_0 \cdot f_x + (\Delta t_1 - \Delta t_2)/T_x \tag{4}$$

$$T_0 = N_x T_x + \Delta t_1 - \Delta t_2 = N_x T_x + \Delta t = N_x T_x \pm \Delta_q \tag{5}$$

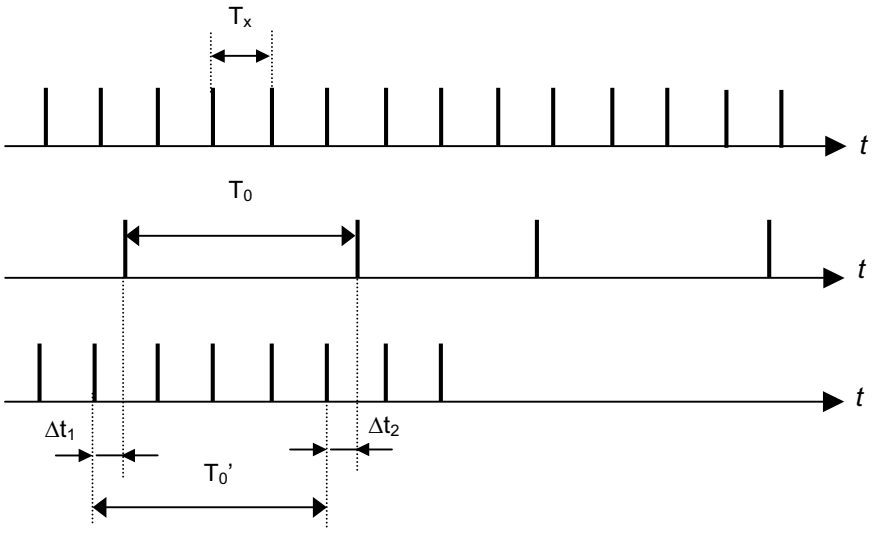


Figure 2. Time Diagrams of Direct Counting Method.

Time intervals Δt_1 and Δt_2 can be changed independently from each other, accepting values from 0 up to T_x with the equal probability. Then the maximum relative quantization (discretization) error caused by the absence of synchronization, will be

$$\delta_q = \pm \frac{1}{N_x} = \pm \frac{1}{T_0 \cdot f_x} \tag{6}$$

Without special measures the absolute error Δ_q with maximal value does not exceed ± 1 count pulse, will be distributed according to the triangular law (Simpson's distribution law).

The moment of the beginning of measurement can be synchronized with the converted frequency but the moment of the ending of measurement cannot be synchronized. The quantization error for this case may be determined from the equation (4), if $\Delta t_1 = 0$. Then the quantization error will be determined according to the following equation:

$$\delta_q = -\frac{1}{T_0 \cdot f_x} \tag{7}$$

The value δ_q can be reduced, if f_x pulses are shifted compulsorily in a half-period in reference to the beginning of the gate time T_0. In this case $\Delta t_1 = T_x/2$, and the quantization error will be determined

$$\delta_q = \pm\frac{1}{2T_0 \cdot f_x} \tag{8}$$

In this case, the distribution of the methodical quantization error is the uniform symmetric (unbiased) law.

It is obvious, that by the calculation of the gate time T_0 according to the equation (5) its maximum value will be determined by the lower frequency f_{xmin} of the conversion frequency range. The high range of frequency conversion is limited by the maximum pulse counter speed.

By the direct frequency measurement, there are two most essential error components. These are the frequency reference error δ_{ref} and the above considered, quantization error δ_q.

The frequency reference error δ_{ref} is the systematic error, caused by inaccuracy of the initial tuning and the long-term instability of the quartz generator frequency. The casual component of the resulting error is determined by the short-term instability of the frequency f_0.

As known, the frequency deviation of the non-temperature-compensated crystal oscillator from the nominal due to the temperature change is $(1 \div 50) \cdot 10^{-6}$ in the temperature interval $-55 \div +125\ ^0$C. This error's component may be essential in the measurement of high frequencies. For reduction of the systematic frequency reference error an oven-controlled crystal oscillator is used at which the maximal value of frequency reference error $\delta_{ref} = 10^{-6} \div 10^{-8}$ remains in the given limits for a long time.

The maximum permissible absolute quantization error is

$$\Delta_q = \pm f_0 = \pm \frac{1}{T_0} \tag{9}$$

In turn, the limits of the absolute error of frequency-to-digital converter based on the standard direct counting method can be calculated as:

$$\Delta_{max} = \pm \left(\delta_{ref} f_x + \frac{1}{T_0} \right) \tag{10}$$

Limits of the relative error of such a frequency-to-digital converter in percent is

$$\delta_{max} = \pm \left(\delta_{ref} + \frac{1}{f_x \cdot T_0} \right) \cdot 100 \tag{11}$$

The quantization error δ_q depends on the converted frequencies. It is negligible for high frequencies above 10 MHz, grows at the frequency reduction and may reach an inadmissible value in the low and infralow frequency range. For example, by the measurement of the frequency f_x = 10 Hz at gate time T_0 = 1 s the quantization error will be 10 %. On the other hand in order to reach the reasonable quantization error of at least 0.01%, a frequency of 10 Hz will result in an increased conversion time up to T_0 = 1 000 s. Effective methods of reducing quantization error in the standard direct counting method are:

a) multiplication of converted frequencies f_x in k times and subsequent measurement of the frequency $f_x \cdot k$, that is reduced to increase the pulse number inside the gate time;

b) usage of weight functions.

Both ways result in some increase of hardware or increase in the chip area for realization of the additional frequency multiplier, increase of the conversion time by the realization of the weight averaging with the help of the microcontroller core.

One of the demerits of this classical method is the redundant conversion time in all frequency ranges, except the nominal frequency.

2.2 Indirect Counting Method

The indirect counting method [5-6] is another classical method for the frequency-to-digital conversion. This method is rather effective for the conversion of low and infralow frequencies. According to this method the

number of pulses of the high reference frequency f_0 is counted during one T_x or several n periods T_x (Figure 3).

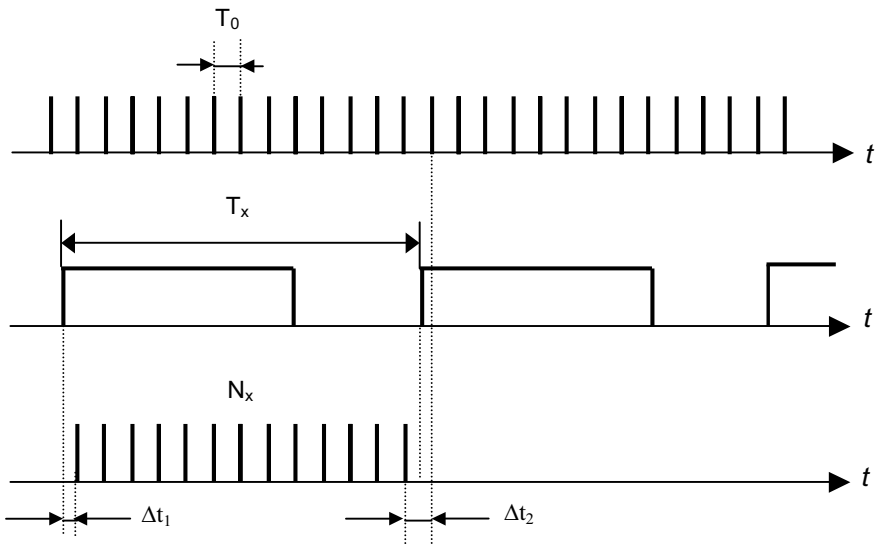


Figure 3. Time Diagrams of Indirect Counting Method.

Thus, the following number will be accumulated in the counter:

$$N_x = n \cdot T_x / T_0 = n \cdot f_0 / f_x , \tag{12}$$

where n is the number of periods T_x. The number N_x is equal to the converted period. By the same way, it is possible to convert the pulse width t_p, or the time interval τ between start- and stop- pulses into the code.

The number of pulses N_x counted by the counter is determined by the number of periods $T_0=1/f_0$ during the time interval T_x. Therefore the average value of T_x is equal to:

$$T_x = N_x \cdot T_0 \tag{13}$$

In view of the absolute quantization error because of non-synchronization

$$T_x = (N_x - 1)T_0 + \Delta t_1 + (T_0 - \Delta t_2) = N_x T_0 + \Delta t_1 - \Delta t_2 = N_x T_0 \pm \Delta_q \tag{14}$$

By synchronization of the reference frequency f_0 with the beginning of the converted interval T_x, it is possible to provide $\Delta t_1 = 0$. However, it is impossible to synchronize the ending of this interval in a similar way.

In order to have the result equals to the converted frequency it is necessary to calculate the ratio

$$N_{fx} = \frac{1}{N_x} \tag{15}$$

By using the microcontroller core, this operation is carried out without any problems and in parallel to the conversion process. It is also possible to use various functional converters or digital integrators working in the mode of hyperbolic function modelling so that the conversion results in units of frequency.

Commonly, the error of frequency-to-digital converters of the periodic signal of any form based on the conventional indirect counting method is determined by the instability of the reference frequency, the quantization error and the trigger error due to internal and input signal noises. Trigger errors occur when a time interval of the measurement starts or stops too early or too late because of the noise on the input signal. There are two sources of this noise: the noise on the signal being measured and the noise added to this signal by the counter's input circuitry.

The relative quantization error can be calculated according to the following equation:

$$\delta_q = \frac{f_x}{n \cdot f_0} \cdot 100 \tag{16}$$

The quantization error can be reduced by increasing the reference frequency f_0 or n - numbers of converted periods T_x. It increases with increasing number of converted frequencies. Limits of the relative error of the frequency-to-digital converter based on the indirect counting method are

$$\delta_{max} = \pm\left(\delta_{ref} + \frac{1}{f_0 T_x n} + \frac{\delta_{TriggerError}}{n} \right) \tag{17}$$

In turn, the trigger error can be calculated as

$$\delta_{TriggerError} = \frac{1.4\sqrt{\left(V_{noise-input}\right)^2 + \left(V_{noise-signal}\right)^2}}{S \cdot T_x}, \tag{18}$$

where S is the signal slew rate (V/s) at trigger point. At rectangular pulses with the wavefront duration no more than $0.5T_0$ the trigger error is equal to

zero. Such output signals are used in the majority of modern frequency output sensors.

The quantization error of the frequency-to-digital converter using the indirect counting method also depends on the measurand frequency f_x (16).

From the given equations it follows, that the conversion of short time intervals causes the large quantization error. It can be reduced by three possible ways. Two of them are obvious direct solutions - increasing of the reference frequency f_0 and converting a greater number of intervals n accordingly. The first way require a high-frequency generator and counter.

There is another conversion method, which reduces the quantizaton error. It is the co-called interpolation method, in which instead of an integer number of reference periods filling out the converted time interval, fractional parts of this period between the reference and the first counting pulse as well as the last counting and reference pulse are also taken into account [7]. This conversion method is illustrated in Figure 4.

The first counting pulse during T_x, is delayed relative to the period's wavefront for time Δt_1, and the tail of T_x and the next counting pulse appearing after the tail - for time Δt_2. If it is possible to take precisely into account intervals Δt_1 and Δt_2, the quantization error would be excluded. The task of measuring the intervals Δt_1 and Δt_2 can be solved in the following way.

During the time interval Δt_1 a capacitor is linearly charged, and then discharged in 1 000 times slowly. This interval is filled out by the same counting pulses and the number N_1 is accumulated. The time interval Δt_2 is measured in the same way. As a result, the required time interval T_x is measured with absolute quantization error $T_0' = T_0/10^3$, which is equivalent to the filling by the counting pulses with a frequency in 10^3 times more than f_0.

The speed of the indirect counting method is determined by the time interval T_x and the latency of the new measuring cycle. The latency can be reduced by two times in the case of the pulse signal or up to zero by using two counters working alternatively. The indirect counting method is a method with a non-redundant conversion time. However, its main fault is a high quantization error in the medium and high frequency range. So, for example, for conversion of a frequency f_x = 10 kHz and reference frequency f_0 = 1 MHz the quantization error will be 1 %.

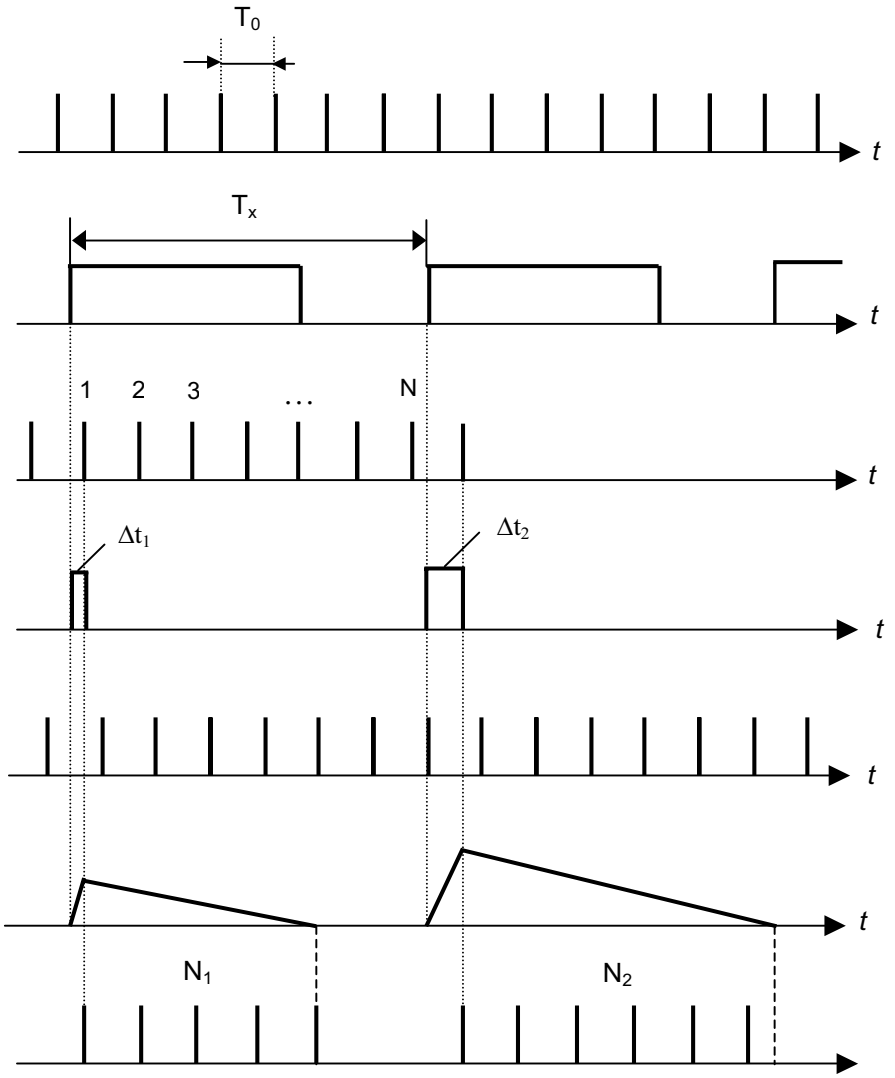


Figure 4. Time Diagrams of Interpolation Method.

2.3 Combined Counting Method

A modification of classical frequency-to-digital conversion methods is possible using the combined conversion method with adaptation possibilities according to the frequency range. In the middle and high frequency range the conversion is carried out according to the standard counting method and in

the low and infralow frequency range - according to the indirect counting method [2, 8].

The boundary frequency of adaptation at which methods are switched, is determined by the condition of maximum quantization errors (6) and (16) equality at n=1 and determined as

$$f_{xbound} = \sqrt{\frac{f_0}{T_0}} \tag{19}$$

Such a combination results in a number of properties, which expands the application of the combined method. The quantization error is essentially reduced at the ends of the conversion range of frequencies. With the increase of T_0 the quantization error δ_q is decreased, and f_{xbound} is shifted into the low frequencies. In turn, with the increase of f_0 the boundary switching frequency of conversion methods is shifted into the high frequencies.

The combined frequency-to-digital conversion method may be easily realized based on a microcontroller. According to the initial data, the microcontroller determines the boundary frequency f_{xbound} at which conversion modes will be switched, analyses the actual value of the converted parameter and forms the command for the mode change. The particular choice of the microcontroller will depend upon the required performance and the operating range as well as other system requirements. Performance factors such as the accuracy and measurement range depend upon available on-board peripherals and the operating speed of the microcontroller.

3. ADVANCED AND SELF-ADAPTING FREQUNCY-TO-DIGITAL CONVERSION METHODS

3.1 Ratiometric Counting Method

In spite of the fact that today frequency can be measured by the most precise methods in comparison with others physical quantities, precise frequency-to-digital conversion with the constant quantization error in a wide specified measuring frequency range (from 0.01 Hz up to some MHz) and with non-redundant conversion time can only be realized based on novel methods of measurement for frequency-time parameters of the electric signal. This requires additional hardware costs and arithmetic operations: multiplication and division for calculation of the final result of conversion. Therefore, additional measuring (conversion) devices should be included in the microsystem. These include two or more multidigital binary counters, the

multiplier and the code divider, logic elements, etc. There are some different design approaches in this case. In the authors' opinion, a successful solution is the use of a microcontroller core in such microsystems. In this case, with the aim to minimize the built-in hardware, it is expedient to take advantage of the program-oriented methods of measurement for frequency-time parameters of signals.

We shall first consider the idea of the original method of the discrete count [4, 7], called the *ratiometric counting method*, which allows frequency-to-digital conversion with a small constant error in a wide frequency range, we shall than consider how this method can be realized.

Let's assume, that the converted periodic signal is in the form of sine wave. By means of input forming device it will be transformed into a periodic sequence of pulses, the period T_x of which is equal to the period of the converted signal. There are various devices and principles for transformation of periodic continuous signals into a sequence of rectangular pulses.

Regardless of this sequence the first reference time interval (gate time) T_{01} is formed (Figure 5). It is filled out by N_1 pulses of the periodic sequence. The number N_1 is accumulated in the first counter. The converted frequency f_x' is determined according to the following

$$f_x^{'} = \frac{N_1}{T_{01}} \tag{20}$$

The frequency deviation from the value f_x is determined by the quantization error, the reduction of which is the aim of this method.

Simultaneously, the second gate time T_{02}, is formed. Its wavefront coincides with the pulse, appearing right after the start of the first gate time T_{01}, and the wavetail, with the pulse appearing right after the end of the first gate time T_{01}. Thus, the duration of the second gate time T_{02} is precisely equal to the integer number of periods of the converted signal, i.e.

$$T_{02} = N_1 \cdot T_x \tag{21}$$

The wavefront and the wavetail of the formed gate time are synchronized with the pulses of the periodic input sequence generated from the input signal, therefore the rounding error is excluded. The second gate time is filled out by pulses of reference frequency $f_{0,}$ whose number is accumulated in the second counter.

The formula for calculation of the converted frequency can be obtained in the following way. The number of pulses, which have got into the second gate time, as can be seen from Figure 5, is determined by the ratio

$$N_2 = N_1 \cdot T_x / T_0 = N_1 \cdot f_0 / f_x \tag{22}$$

hence

$$f_x = \frac{N_1}{N_2} \cdot f_0 \tag{23}$$

where f_0 is the reference frequency. This can be done in a PC or a microcontroller core along with the offsetting and scaling that must often be performed.

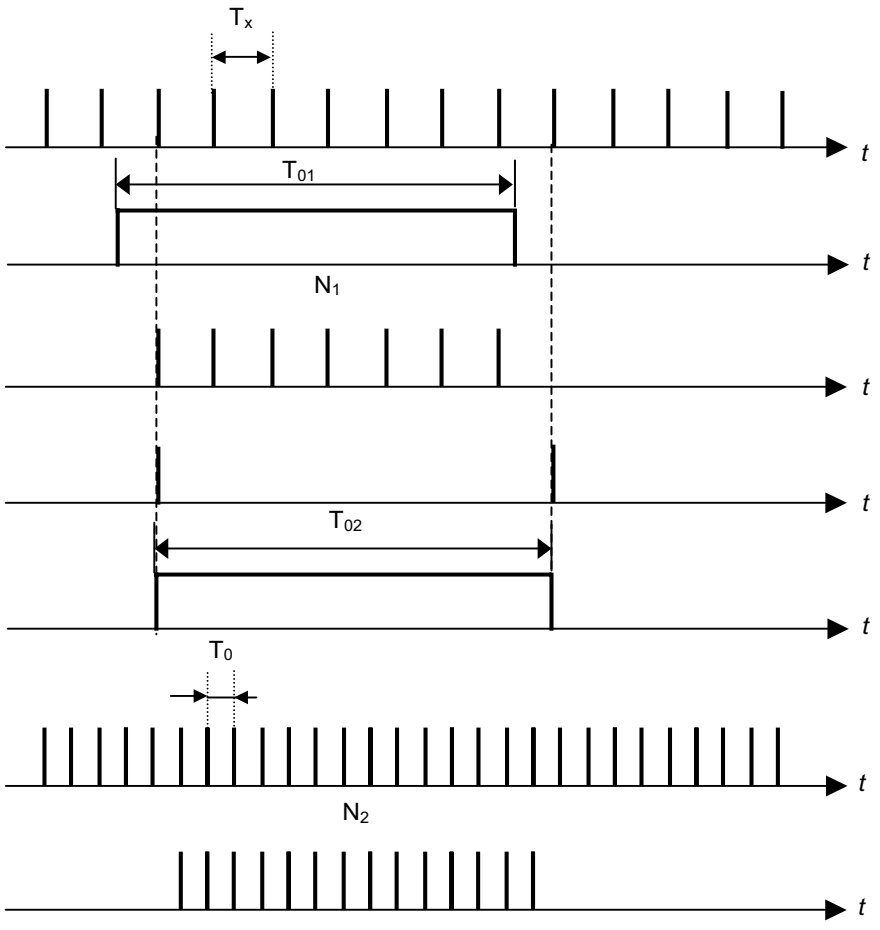


Figure 5. Time Diagrams of Ratiometric Counting Method.

The accuracy of the frequency measurement is determined by the quantization error of the time interval N_1T_x.

Let's withdraw the equation for the relative quantization error δ_q for the frequency-to-digital conversion. First of all we shall determine the maximum value of the relative quantization error for the time interval $T_{02}=N_1T_x$. As this interval is filled out by counting pulses with the period T_0, the maximum absolute error is $\Delta_2 = \pm T_0$, and the maximum relative error is:

$$\delta_2 = \pm\frac{T_0}{T_{02}} = \pm\frac{T_0}{N_1 \cdot T_x} \tag{24}$$

The equality $N_1T_x = T_{02}$ can be presented as $f_x = N_1/T_{02}$. Then according to rules of the error calculation for indirect measurements the measurement error of the function f_x is connected with the measurement error of the argument T_{02} by the ratio (with the accuracy of the second order of smallness): $\delta_q = \delta_2$. After substitution the δ_2 from (24) we shall receive

$$\delta_q = \pm\frac{T_0}{N_1 \cdot T_x} = \pm\frac{f_x}{N_1} \cdot T_0 \tag{25}$$

According to the standard direct counting method it is possible to write the equality $T_{01} = N_1/f_x'$. Substituted the relation $f_x'/N_1 = 1/T_{01,}$ into (25) instead f_x/N_1 we obtain

$$\delta_q = \pm\frac{T_0}{T_{01}} = \pm\frac{1}{f_0 \cdot T_{01}} \tag{26}$$

This formula let us draw the conclusion that the maximum value of the relative quantization error for the frequency-to-digital conversion for this method does not depend on converted frequencies and, hence, is constant in all conversion ranges of frequencies.

For the reference frequency f_0 = 1 MHz and the first gate time T_{01} =1s the maximum value of the relative quantization error will be $\delta_q = \pm 10^{-4}$ %.

If, by the measurement of the time interval $T_{02}=N_1T_x$ using the interpolation method, considered in 2.2 with the same frequency and gate time we obtain $\delta_q = \pm 10^{-7}$ %.

Finally, we consider the block diagram of the converter (Figure 6), which realizes the frequency-to-digital conversion according to the considered ratiometric counting method.

It contains two counters and a D-trigger clocked by the sensor output. All counter functions can be provided by the counter/timer peripheral component, which interfaces to many popular microcontrollers.

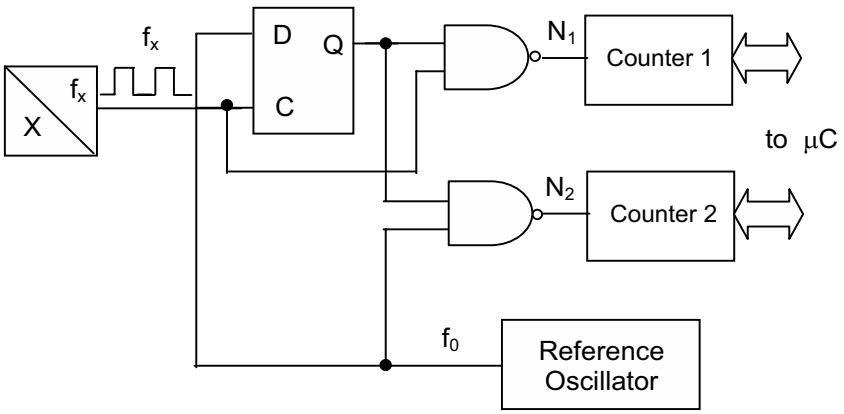


Figure 6. Ratiometric Simplified Counting Scheme.

This method can also be used for the period *(T_x)*- to-code conversion. Thus, the period is calculated according to the following formula

$$T_x = \frac{N_2}{N_1} \cdot T_0 \tag{27}$$

The main demerit of the considered conversion method is the redundant conversion time.

3.2 Reciprocal Counting Methods

A variations of the method described above is the reciprocal counting method (Figure 7). The cyclicity of the conversion cycle T_{cycle} is determined by the reset pulse. The beginning of the counting interval T_{count} coincides with the next pulse of sequence f_x, appearing after the ending of the 'Start' pulse, and the end coincides with the next pulse f_x, appearing after the 'End' pulse [9]. The frequency is calculated during the interval T_{calc}.

The frequency or the period are calculated similarly as described above according to equations (5.4) and (5.11) accordingly. The quantization error is calculated according to the following:

$$\delta_q = \pm \frac{1}{f_0 \cdot T_{count}} \tag{28}$$

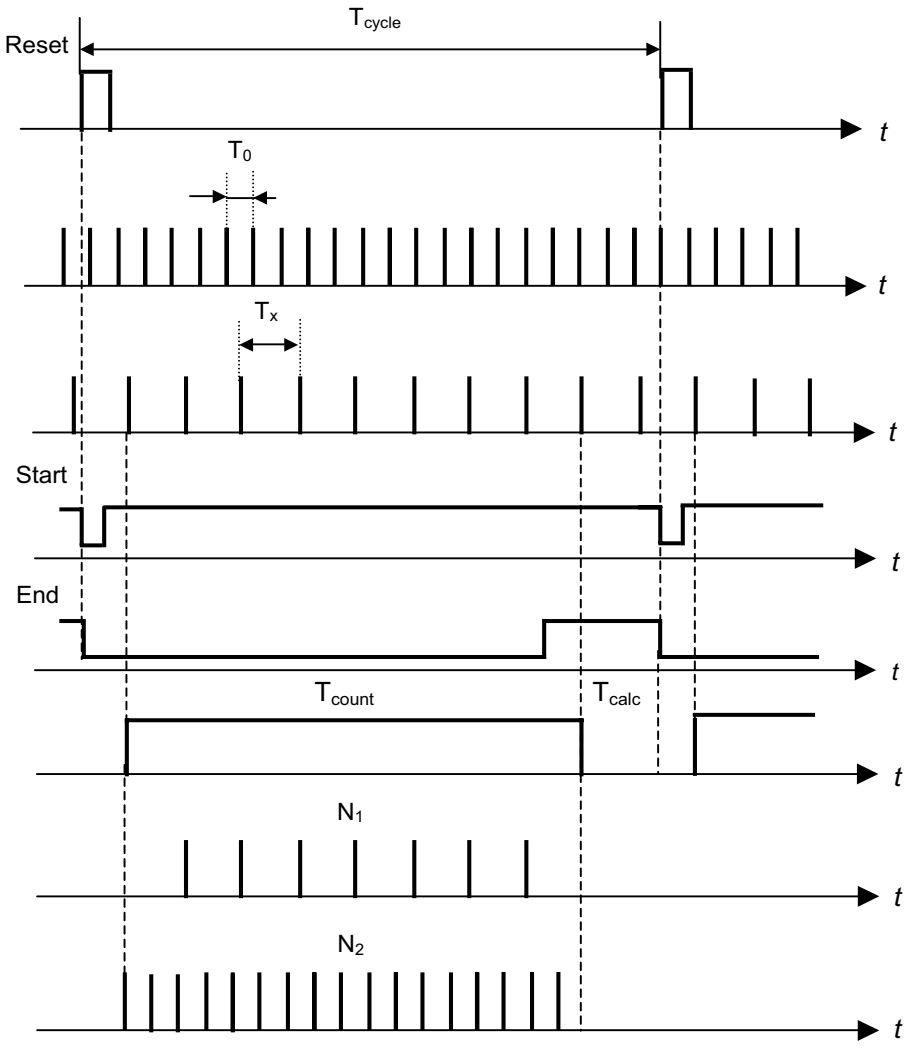


Figure 7. Time Diagrams of Reciprocal Counting Method.

However, in this method, the value of the first gate time T_{count} can be determined only approximately. Having set the conversion cycle equal to 1 s, it is possible to assume only roundly, that $T_{count} \cong 1$s. It is a little bit inconvenient for engineering calculation of the quantization error δ_q.

This method has the same demerit like the ratiometric counting method - the redundant conversion time.

3.3 M/T Counting Methods

The so-called *M/T counting method*, which also overcomes demerits of conventional methods (standard direct counting method and indirect counting method) and achieves high resolution and accuracy for a short detecting time was described in [10].

Time diagrams of the M/T counting method is shown on Figure 8. The detecting time T_{02} is determined by synchronizing the generated output pulse first after a prescribed period of time $T_{01.}$

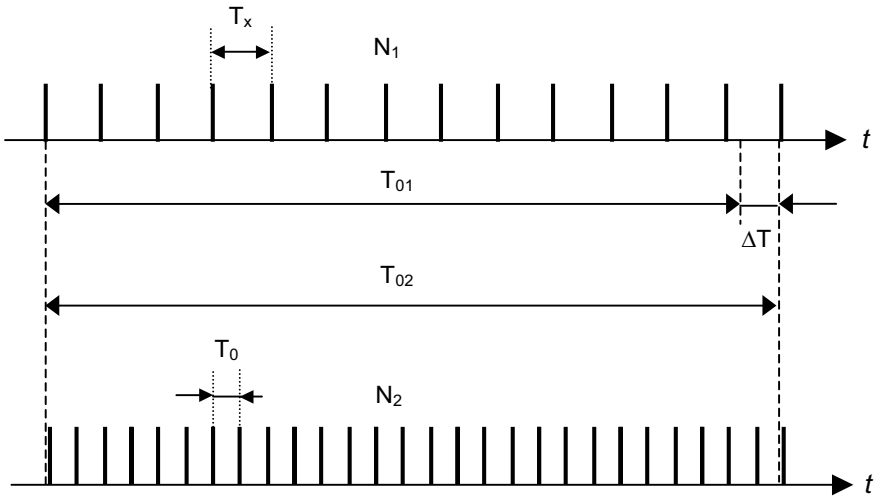


Figure 8. Time Diagrams of M/T Counting Method.

It is easy to notice, that the M/T method differs from the above described ratiometric counting method by the synchronization of the first reference time interval T_{01} with the pulse of the converted frequency f_x. The converted frequency or the period are determined similarly according to equations (23) and (27). The quantization error does not depend on the converted frequency and is constant in all frequency ranges.

The detecting time is calculated by the following formula:

$$T_{02} = T_{01} + \Delta T \tag{29}$$

The method uses three hardware timer/counters. One of them works in the timer mode in order to form the first time interval $T_{01,}$ the rest are used in the counter mode.

The M/T method has the same demerit as the previous two methods - the redundant conversion time.

3.4 Constant Elapsed Time (CET) Method

The *Constant Elapsed Time (CET)* method is another advanced method with the constant quantization error in a wide specified conversion frequency rage [11]. Like the M/T counting method, the CET method is based on both counting and period measurements. Both measurements starting at a rising edge of the input pulse. The counters are stopped by the first rising edge of the input pulse occurring after the constant elapsed time T_{01}. For its realisation the method uses two software timers with two hardware timer/counters inside the microcontroller.

In order to eliminate repetitive starting and stopping counters that require a reinitialization, the time counter run continuously and the pulse counter is read at the end of each time interval. This method has the same demerit: the redundant conversion time.

3.5 Single and Double Buffered Methods

The single buffered (SB) method is based on both pulse counting and measurement of the fractional pulse period before the interrupt sample time [12]. Instead of repetitive enabling/disabling of the capture register function, this function is always enabled. Each rising edge of the synchronized input pulse f_x stores the content of the timer in the timer capture register. The same pulse is the clock for the pulse counter. The interrupt requests are generated using the interval timer, without any link to the pulse rising edges.

The time difference N_2 in this method is determined using the difference between two readings of the capture register during the current and previous interrupt service routine accordingly.

The pulse difference N_1 is determined using the difference between two readings of the pulse counter during the current and previous interrupt service routine accordingly.

The frequency f_x is determined using the pulse difference, the time difference and the clock frequency f_0 of the timer as (23).

Each frequency-to-digital converter consists of a free running timer with a timer capture register, a pulse counter and an interrupt generator. The hardware complexity for such a system, using a software timer instead of an interval timer, is similar to that of the CET method. Hardware for the SB system is available in some microcontrollers.

The SB method has the following disadvantages [12]:

1) The reading of the timer capture register can be erroneous, if the reading is performed during the store operation of the content of the free running timer. The other problem occurs if the reading of the pulse counter is done during counting. Both problems require the synchronization hardware.
2) The rising edge of an external pulse at the time interval between the reading of the timer capture register and the reading of the pulse counter will cause the inadequacy of the content of the pulse counter compared to the necessary content. This is a source of a numerical error, especially during the low frequency measurement.

The converter based on the *double-buffered (DB) method* consists of an interval timer with an associated modulus register, a timer capture register, an additional capture register and a pulse counter with the counter capture register.

The problems of the SB method are inherently solved using synchronization with the rising and falling edges of the system clock. The maximum measured frequency for the DB method is limited by the synchronization logic. The equivalent hardware complexity for both methods is similar, because of the free running the timer consists of the register and associated logic. The maximum measured frequency for the DB method is, however, not limited by the software loop but only by the hardware.

The measurement error for the SB and DB methods is caused by the synchronization of the pulse signal. The worst-case error is caused by missing one count of the system clock f_0. The lower error limit of the SB and DB methods is the same as the error limit of the ratiometric, reciprocal, M/T and CET methods, while the higher error limit of the SB and DB methods is twice as much as the error limit of these methods. Like all mentioned methods SB and DB methods also have the redundant conversion time in all specified conversion frequency ranges except the nominal frequency.

3.6 DMA Transfer Method

The Direct Memory Access (DMA) method [13], provides an average frequency measurement of the input pulses, based on both pulse counting and time measuring for constant sampling time. The time is measured by counting pulses of the reference clock with the frequency f_0 in a free-running timer. Each rising edge of the input pulses activates a DMA request. The DMA controller transfers the content of the free-running timer into the memory (analogous to the timer capture register in the SB method) and decrement the DMA transfer counter (analogous to the pulse counter in the SB method). After each constant sampling time, the interval timer generates an interrupt request to the microcontroller, which reads the contents of both

the DMA transfer counter and the memory in an interrupt routine. The measured frequency is calculated as (23).

Due to some specific errors for the DMA method the maximum error is greater than the error of all the above described advanced conversion methods (sometime more than 10 times greater) in the case of the coincidence between an input pulse and the execution of the longest microcontroller instruction, just prior to reading in an interrupt routine.

3.7 Method of Dependent Count

The Method of the Dependent Count (MDC) was proposed in 1980 [14]. It combines the advantages of the classical methods as well as the advanced methods ensuring a constant relative quantization error in a broad frequency range and at high speed. It is suitable for a frequency conversion in a wide frequency range: from parts of Hz up to several MHz with the constant, beforehand given, quantization error and non-redundant conversion time. This method is the most perspective method for application in self-adaptive smart sensors.

One of the essential advantages of the MDC is the possibility to convert the frequency $f_x \geq f_0$. In this connection we shall use the following denotations for the deducting of main mathematical formulas: F is the greater of the two frequencies f_x and f_0; f is the lower of the two frequencies f_x and f_0.

The time diagram of method is shown in Figure 9. The method consists in the following. With arrival of the impulse of signals with the lower frequency f (this corresponds to the moment t_1 in Figure 9), the counters start to calculate the impulses of both signals. The number of impulses N_i, stored in the up-down counter, which calculates the impulse with the frequency F, is compared with the number N_δ. This number is set up previously in the counter by the microcontroller. At some moment of time (this corresponds to the moment t_2 in Figure 9), when the number of impulses, calculated by this counter, will be N_δ, e. g. $N_i = N_\delta$ with the arrival of next impulse (after the moment t_2) with the lower frequency f (moment t_3), the impulse count will be stopped. The number of impulses (of signal of lower frequency f) counted by one of the counters is n, and the number of impulses (of signal of higher frequency F), counted by the second counter, is $N = N_\delta + \Delta N$. The conversion time t_x always equals to an integer number of periods of a signal with the lower frequency f:

$$t_x = \tau \cdot n = \frac{n}{f} \tag{30}$$

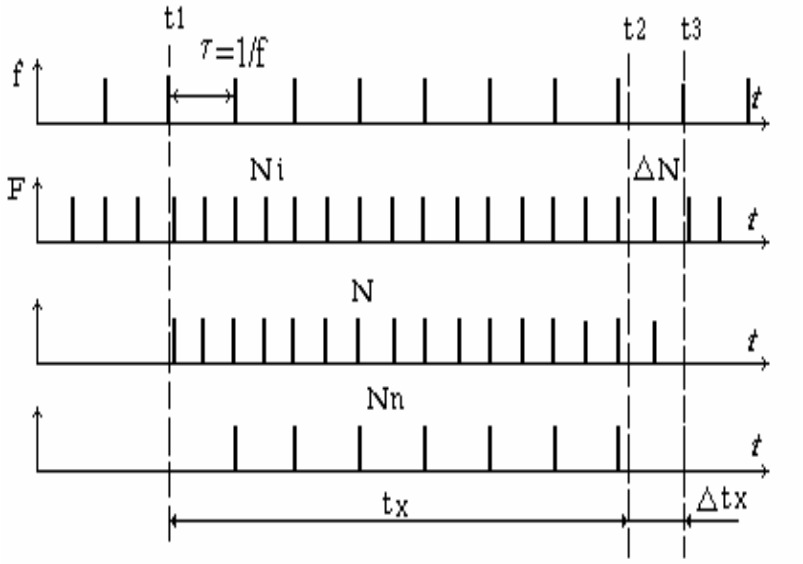


Figure 9. Time Diagrams of the Method of the Dependent Count.

This interval can be given also as:

$$t_x = T \cdot N = \frac{N}{F} = \frac{N_\delta + \Delta N}{F} == \left(\frac{1}{\delta} + \Delta N\right) \cdot T \qquad (31)$$

From the equations (30) and (31) follows, that $f = F\ (n/N)$ or $F = f\ (N/n)$. If the measured frequency f_x is lower frequency, e.g. $f_x = f$, and reference frequency f_0 is high ($f_0 = F$), that

$$f_x = f_0 \cdot \frac{n}{N}$$

or (32)

$$f_x = f_0 \cdot \frac{N}{n},$$

when $f_x = F$ and $f_0 = f$.

The microcontroller calculates the unknown frequency from the equations (32). The program for calculation is determined by the command, which to be prepared on the basis of earlier entered into the microcontroller information which from the frequencies is lower.

For the period τ or T, the conversion is carried out similarly. The

microcontroller calculates it values from the following equations:

$$\tau = \frac{N}{f_0 \cdot n}$$

or (33)

$$T = \frac{n}{f_0 \cdot N}$$

The quantization error for considered conversions caused by that, the interval t_x (conversion time) is not equal to the interval, which is determined by the integer number N of periods of a signal with high frequency, e.g.

$$t_x \neq N \cdot T \neq N/F \tag{34}$$

At change of the lower frequency in known limits, the interval t_x will be changed (or, on the contrary, will be changed the N/F at change of higher frequency). It will result in change of the number of impulses N from $\Delta N = 0$ up to $\Delta N = \Delta N_{max}$ calculated by the counter. Here ΔN_{max} is the number of impulses in the interval $\Delta t_{xmax} = \tau$. Taking into account the fact, that the period of these pulses is equals to T, the following equation will be true:

$$\Delta N_{max} = \frac{\tau}{T} = \frac{F}{f} \tag{35}$$

The maximum quantization error arises in that case, when the number of impulses N counted by the counter is minimum and equal to the N_δ (one of frequencies is changed). Then

$$\delta_{max} = \frac{1}{N_{min}} = \frac{1}{N_\delta} \tag{36}$$

Hence, the maximum error is determined by the value N_δ only and practically does not depend on the measured frequency. The minimum value of the error will be at $N = N_{max}$. But as $N_{max} = N_\delta + \Delta N_{max}$, then

$$\delta_{min} = \frac{1}{N_{max}} = \frac{1}{N + \Delta N_{max}} \tag{37}$$

With the aim to compare the method of the dependent count with the standard counting methods as well as with other advanced methods for the

frequency-to-digital conversion let us determine the coefficient of variation for the quantization error $\alpha = \delta_{q\,max}/\delta_{q\,min}$ for these methods.

So, for the method of the dependent count if f_x is lower of two frequencies $(f_x = f)$, and f_0 is greater of the frequencies $(f_0 = F)$, i.e. $f_x < f_0$, from the equations (36), (37) and taking into account the equation (35) we will have

$$\alpha = \frac{N_\delta + \Delta N_{max}}{N_\delta} = 1 + \frac{1}{N_\delta} \cdot \frac{F}{f} \tag{38}$$

or

$$\alpha_1 = 1 + \frac{1}{N_\delta} \cdot \frac{f_0}{f_x} \tag{39}$$

At $f_x > f_0$

$$\alpha_2 = 1 + \frac{1}{N_\delta} \cdot \frac{f_x}{f_0} \tag{40}$$

Let's determine, in how many times the quantization error will be varied by the measuring frequency f_x = 2 Hz, if $f_0 = 10^6$ Hz and $N_\delta = 10^6$ ($\delta = 10^{-6} \cdot 100\ \% = 0.0001\ \%$). From the equation (39) we shall receive

$$\alpha_1 = 1 + \frac{1}{10^6} \cdot \frac{10^6}{2} = 1.5\,,$$

i.e. the greatest error $\delta_{max} = 1/N_\delta = 10^{-6}$, and the lowest $\delta_{min} = 0.67 \cdot 10^{-6}$. As the greatest quantization error for the method of the dependent count is constant for any measurand, it is possible to characterise the possible range of variation of this error in the specified measuring range of frequencies by the coefficient of variation α. From this example follows, that the error variation is not more than 1.5 times in the frequency range $2 \div 10^6$ Hz (by this, the time of the measurement is constant for the given quantization error). By the usage of the standard direct counting method or the indirect method measuring period, the variation of the quantization error will be 500000 (at the same time of measurement).

The method of dependent count has the highest speed at measurement for all frequencies from the frequency range and consequently, allows reducing essentially the dynamic error.

According to (31) the maximum time of conversion is

$$t_x = \frac{N_\delta + \frac{f_0}{f_x}}{f_0} \tag{41}$$

in the case, when $f_x \le f_0$ and

$$t_x = \frac{N_\delta + \frac{f_x}{f_0}}{f_x}. \tag{42}$$

Thus, with the aim to cover a lot of existing smart sensors, working in various parts of frequency range, the following size of variables changing were used: $f_x \in$ [0.1÷10 000 000] Hz; $f_0 \in$ [100 000÷1 000 000] Hz; $N_\delta \in [1\ 000 \div 10^7]$ (it is corresponds to the size of given quantization error changing from 0.1 up to 10^{-7} %; $\Delta N \in [0 \div \Delta N_{max}]$. The plot of modelling results of function $t_x = \varphi(N_\delta, f_x, f_0)$ in the wide frequency range from 1 Hz up to 10 MHz at relatively high accuracy of conversion $N_\delta = 10^6$ (the quantization error does not exceed 10^{-4} %) is shown in Figure 10 [15].

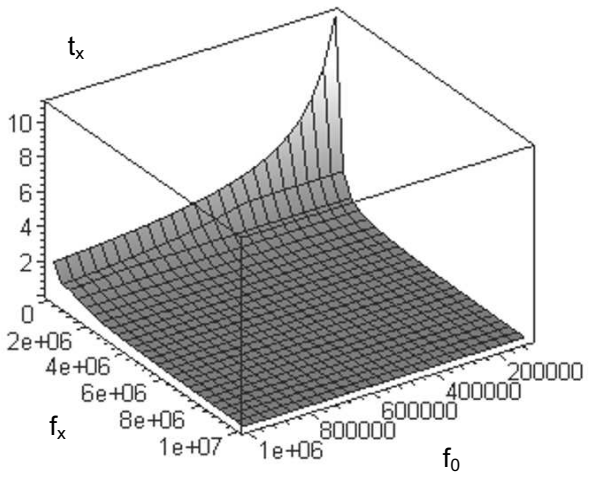


Figure 10. Modelling Result of $t_x = \varphi(N_\delta, f_x, f_0)$ Function at $N_\delta = 10^6$; $f_0 = 100\ 000 \div 10^6$ Hz; $f_x = 1 \div 10\ 000\ 000$ Hz.

This example illustrates one more essential advantage of the method of dependent count - an opportunity to convert frequencies exceeding the reference frequency $f_x \ge f_0$. Besides, the given example testifies about the

necessity to use the high reference frequency only in low and infralow ranges. This opens a prospect for the adaptive control of reference frequency during the conversion that results in reduction of power consumption in smart sensors.

3.7.1 Examples

Let it is necessary to convert the frequency $f_x = 2 \cdot 10^4$ Hz at $f_0 = 10^6$ Hz and $N_\delta = 10^6$ ($\delta = 10^{-4}$ %). According to the method of dependent count the maximum conversion time is

$$t_x = \frac{10^6 + \dfrac{10^6}{2 \cdot 10^4}}{10^6} \approx 1\,\text{s}$$

In turn, according to the standard counting method, the time of measurements necessary for the same accuracy is calculated according to the following formula:

$$t_x = \frac{1}{\delta \cdot f_x} = \frac{N_\delta}{f_x} = \frac{10^6}{2 \cdot 10^4} = 50\,\text{s}$$

Let's consider the conversion of same frequency, but with the help of indirect counting method. In spite of fact that it is a conversion method with non-redundant conversion time on the nature, in order to have the required quantization error for the given frequency range it will be necessary to convert much more than one periods of f_x:

$$N_T = \frac{f_x}{f_0 \cdot \delta} = \frac{f_x \cdot N_\delta}{f_0} = \frac{2 \cdot 10^4 \cdot 10^6}{10^6} = 2 \cdot 10^4.$$

In this case, the conversion time will be calculated according to the following formula:

$$t_x = N_T \cdot T_x = \frac{2 \cdot 10^4}{10^4} = 2\,\text{s}$$

The usage of other advanced conversion methods with the fixed conversion time, for example, ratiometric, reciprocal, CET, M/T, DMA, etc.

methods also demands to increase the conversion time like to the standard counting method - up to 50 s.

In other words, the time of measurement for the method of dependent count is non-redundant in all specified measuring range of frequencies. In the standard counting method and modern advanced, the time of measurement is redundant, except the nominal frequency. Moreover, for the method of dependent count the time of measurement can be varied during measurements depending on the assigned error.

3.8 Method with Non-Redundant Reference Frequency

Modern integrated one-chip smart sensors are very sensitive to power consumption. Taking into account a very high degree of integration of such embedded systems and different wireless applications, the power consumption becomes a critical parameter.

Smart sensors are highly programmable and based on embedded microcontrollers or its cores. The actual power-supply current depends on many factors. One of them is an operation frequency. The power consumption of the device is directly proportional to the system clock. The power dissipation in CMOS integrated circuits is given by the following equation [16]:

$$P_{avg} = V_{dd}^2 \cdot C \cdot f_{clc}, \tag{43}$$

where V_{dd} is the supply voltage, C is the average switched capacitance per clock cycle, and f_{clc} is the clock frequency. If the clock speed doubles, the current doubles. It is evident from (43) that a large improvement in power dissipation can be achieved by scaling down of $V_{dd.}$ However, as rule, the V_{dd} as well as C is fixed for the concrete circuit design.

Obviously, power can be saved by operation the device at the lowest clock speed possible that still meets the specification for the device and the requirements of the application. But this condition is contradicted with the requirements of high accuracy and speed. There is a strong tension between the need for both high metrological performances and minimal power consumption in low-power smart sensor embedded systems. So, the quantization error for frequency and time measurements (conversions) is inversely proportional to the reference frequency, formed by the system clock. Hence, scaling down of the clock frequency is associated with reduced performance.

However, smart sensors are devices with intelligent capabilities. It is also includes a self-adaptation. So, smart sensors should support therefore the flexible clocking. So, dependent on conditions of measurement when high

conversion accuracy is not required, the clock frequency should be reduced with the aim to reduce the power consumption. On the other hand, if conditions of measurement demand a high precision accuracy the clock frequency should be increased during a certain time interval, with the aim to reduce the dominant error of conversion for frequency-time parameters - a quantization error. For realization of such possibilities, it is necessary to use novel so called frequency-to-digital conversion method with non-redundant reference frequency [17]. Hence, a technique to improve performance has to be considered along with the clock frequency scaling.

In the quest for low power and high metrological performance, a flexible clocking might be the best solution. The essence of the method with non-redundant reference frequency consists in the following. The embedded microcontroller or arithmetic unit calculates the value of necessary reference frequency according to the given quantization error δ_{gi}:

$$\mathrm{f_{0i}} = \frac{\mathrm{k}}{\delta_{\mathrm{gi}}}, \tag{44}$$

where $k = 1/T_o$ = const (T_o is the first reference gate time). Both the reference frequency f_{0i}, which is received by division/multiplication of the clock frequency f_{clc} and the measurand frequency f_x are calculated by the counters CT_1 and CT_2 accordingly. The conversion time is equal to the integer number of periods of frequency f_x. The time diagrams of method are the same as for the ratiometric counting method [4]. The frequency f_x is calculated similarly to any of advanced frequency-to-digital conversion method:

$$\mathrm{f_x} = \frac{\mathrm{N_1}}{\mathrm{N_2}} \cdot \mathrm{f_{0i}}. \tag{45}$$

The quantization error does not exceed the beforehand given:

$$\delta_{\mathrm{gi}} = \frac{1}{\mathrm{T_0} \cdot \mathrm{f_{0i}}}, \tag{46}$$

Similarly to any advanced frequency-to-digital conversion method, this method ensures the constant quantization error in the whole specified measuring range of frequencies - from infralow up to high frequencies. Besides that, the reference frequency f_{oi} is non-redundant and determined by the given error of measurement $\delta_{gi.}$

The modern achievement in microelectronics and the usage of program-oriented conversion methods [8], the realization with minimum hardware is

possible. As on-chip capacitances are much lower than off-chip capacitances, it is good to implement as much of the sensor as possible on a single chip [18]. The MSP430 16-bit RISC ultra-low power consumption 20-pin version of the new microcontroller family for metering application from *Texas Instruments* [19] is most appropriate for realization of such conversion method for programmable smart sensors. These microcontrollers have performances, which render it suitable for low-power embedded applications like smart sensors. First of all, it is the flexible clocking. The microcontroller can also operate with a low voltage from 1.8 up to 3.6 V.

The system clock frequency f_{system} in this microcontroller depends on two variables:

$$f_{system} = N \cdot f_{crystal}, \tag{47}$$

where N *(3÷127)* is the multiplication factor; $f_{crystal}$ is the frequency of crystal (normally 32768 Hz). The normal way to change the system clock frequency is to change the multiplication factor *N*. The System Clock Frequency Control register SCFQCTL is loaded with *(N-1)* to get the new frequency [19].

Figure 11 shows the possible smart sensors architecture based on the MSP430 microcontroller.

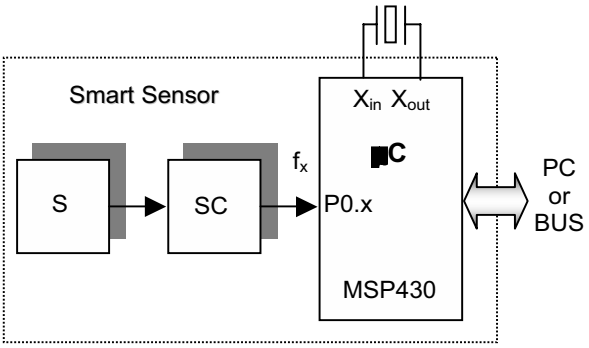


Figure 11. Smart Sensor Architecture based on MSP430 Microcontroller (*S* -sensing element*; SC* - signal conditioning).

The frequency signal can be connected to any of the eight inputs of Port0 and counted via the external interrupt. If the frequencies to be measured are above 30 kHz then the Universal Timer/Port or the 8-bit Interval Timer/Counter may be also used for counting. The first gate time is formed by the Basic Timer. Although the timer, as rule, runs independently of the CPU, the speed of the timer is still based on CLCOUT. Consequently, the current requirement of the timer changes with processor clock speed just as

other functions do. The current use of the timer also depends on its activity. The more frequently the timer reloads, the higher the operating current is.

The function of $P_{avg} = \phi(\delta)$ for adaptive method *1* with non-redundant reference frequency and any of advanced methods *2* at the gate time $T_0 = 0.1\ s$ is shown in Figure 12.

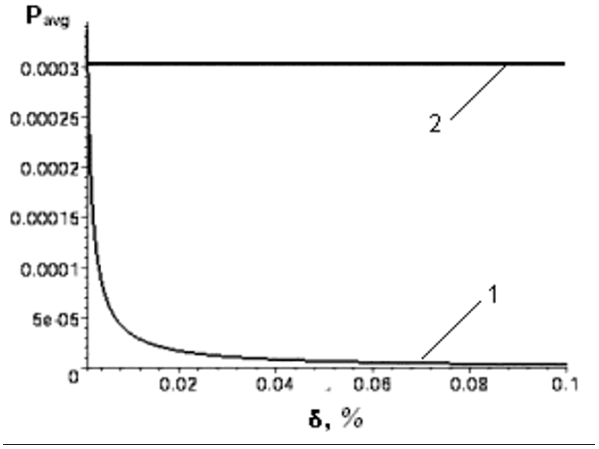


Figure 12. The Dependence of $P_{avg} = \phi(\delta)$ for Adaptive Method with Non-redundant Reference Frequency (1) and any of Advanced Method (2) at $T_0 = 0.1$ s.

Due to the redundant and constant reference frequency in the last case, the power consumption is essentially higher.

3.9 Comparison of Methods

For choice of methods frequency-to-digital conversion, it is expedient to prefer one that has high metrological characteristics and simple realization by its universality. The main performances of all the methods considered above are adduced in Table 1. Here number 1 is the indirect counting method (period measurement); 2 is the standard direct counting method (frequency measurement); 3 is the ratiometric counting method; 4 is the reciprocal counting method; 5 is the M/T counting method; 6 is the constant elapsed time (CET) method; 7 is the single buffered method and the double buffered method; 8 is the DMA transfer method; 9 is the method with the non-redundant reference frequency; 10 is the method of the dependent count.

As it can be seen from the table, the majority modern advanced counting methods overcame the demerit inherent to classical conversion methods.

Namely, it means the inconstancy of the quantization error in all specified ranges of converted frequencies. However, most of them (methods with the constant conversion time as well as with the slightly varied conversion time) have a redundant conversion time for all frequencies, except the nominal one. The exceptions are the method of the dependent count and the classical indirect counting method (period measurement). Only one of methods, namely the method of the dependent count measures the frequency $f_x \geq f_0$, and only one method, namely the method 9 has the non-redundant reference frequency.

Table 1. Main Performances of Methods for Frequency-to-Digital Conversion.

Method	Quantization Error, δ_q	Conversion Time	Conversion Range, D_f	Calculation of Result	Error Variation, γ
1.	$\frac{f_x}{f_0}$	T_x	$f_x << f_0$	$f_x = \frac{f_0}{N_2}$	$f_{x\,max}/f_{x\,min}$
2.	$\frac{1}{T_0 \cdot f_x}$	T_0	$f_x >> \frac{1}{T_0}$	$f_x = \frac{N_1}{T_0}$	$f_{x\,max}/f_{x\,min}$
3.	$\frac{1}{f_0 \cdot T_{01}}$	$T_{01} + \Delta T$	$f_x \leq f_0$	$f_x = \frac{N_1}{N_2} \cdot f_0$	1 ÷ 1.5
4.	$\frac{1}{f_0 \cdot T_{count}}$	$T_{count} + \Delta T$	$f_x \leq f_0$	$f_x = \frac{N_1}{N_2} \cdot f_0$	1 ÷ 1.5
5.	$\frac{1}{f_0 \cdot T_{01}}$	$T_{01} + \Delta T$	$f_x \leq f_0$	$f_x = \frac{N_1}{N_2} \cdot f_0$	1 ÷ 1.5
6.	$\frac{1}{f_0 \cdot T_{01}}$	$T_{01} + \Delta T$	$f_x \leq f_0$	$f_x = \frac{N_1}{N_2} \cdot f_0$	1 ÷ 1.5
7.	$\frac{1}{f_0 \cdot T_0}$	T_0	$f_x < f_0$	$f_x = \frac{N_1}{N_2} \cdot f_0$	1 ÷ 2
8.	$\frac{1}{f_0 \cdot T_0}$	T_0	$f_x < f_0$	$f_x = \frac{N_1}{N_2} \cdot f_0$	1 ÷ 10
9.	$\frac{1}{T_{01} \cdot f_{0i}}$	$T_{01} + \Delta T$	$f_x \leq f_{0i}$	$f_x = \frac{N_1}{N_2} \cdot f_{0i}$	1 ÷ 1.5
10.	$\frac{1}{N_\delta}$	$\frac{N_\delta + \Delta N}{f_0} = \frac{N_2}{f_0}$	$f_{xmin} < f_0 \leq f_{xmax}$	$f_x = \frac{N_1}{N_2} \cdot f_0$	1 ÷ 1.5

The relation between the maximum relative error and the frequency of the input pulses for main conversion methods is shown in Figure 13.

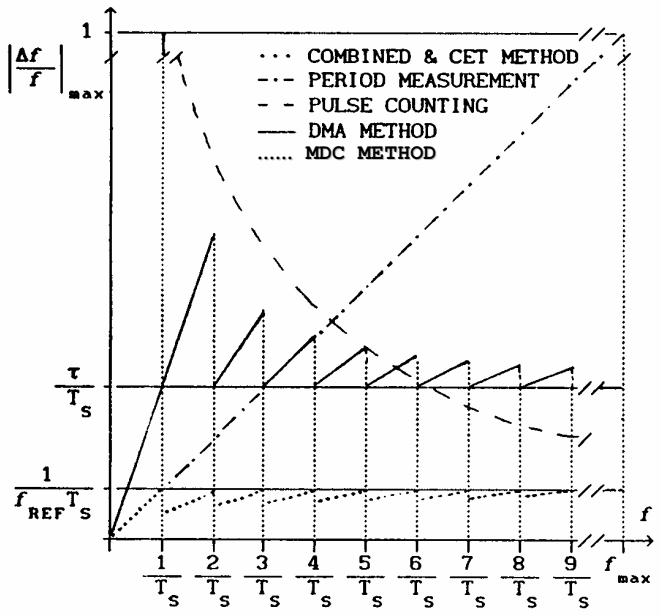


Figure 13. Relation Between the Maximum Relative Error and Frequency of the Input Pulses.

4. DUTY-CYCLE -TO -DIGITAL CONVERSION METHODS

A duty-cycle output signal is widely used as informative output signal for different quasi-digital sensors (Figure 1.2). For example, there are the temperature sensors SMT160-30, TMP04, MAX6666/67, accelerometer ADXL202E and KXG20-L20 accelerometer inclinometer, optical sensors humidity sensors, etc. [20]. All these sensors produce an output that is a duty-cycle modulated quasi-digital signal. Such kind of signal can be easily interfaced with modern microcontrollers. In comparison with a frequency output signal the duty-cycle is rather immune to interfering signals, such as spikes [21], and the ratio does not depend on the absolute value of any component [22].

Various methods exist to measure duty-cycle of an impulse signal. For example, some simple PWM A/D converter can use the classical approach: to measure the pulse width and period of signal, then calculate the ratio:

$$\text{D.C.} = \frac{\tau_x}{T_x}, \tag{48}$$

where τ_x is the pulse width, T_x is the period of pulse signal. Main error's components are quantization errors for pulse width and period. Both components can be big enough. If a high accuracy is needed, a very high clock frequency should be used. The result also depends on the frequency $f_x = 1/T_x$.

Another approach to measure a duty-cycle is to take random samples of a digital signal (random-sampling method) [23]. The method can be realized very easy by program-oriented way. But this method is suitable only for low-resolution conversions for which the necessary resolution is a maximum of 9 bits.

A new method of reading the time-domain sensor signals is described in [24]. It can eliminate the part of quantization error without increase of clock frequency. The method uses the internal clock frequency as 2^N times of the signal frequency. So, it means that T_x doesn't change with the sensor output signal. However, very often, the frequency (period) of signal is changing. In this case this method cannot be used.

The novel proposed method [25] is based on the determination of average pulse width and average period during the conversion time T_q. The last one is determined by the beforehand given quantization error δ and equals to the integer number of periods. Due to this, the component error by reason of non-multiplicity of conversion time T_q and period T_x is eliminated. The time diagrams of the method are shown in Figure 14.

At the beginning of conversion the given relative error $\delta_{Tx} = 1/N_\delta$ is set up. The beginning of T_q is coincided always with the wavefront of first pulse with τ_{x1} duration, and the wavetail – with the τ_{Nx+1} pulse. Hence, the requirement of the method of the dependent count [14] - the multiplicity of number of periods T_x to the conversion time T_q is fulfilled. The numbers $N_{\tau x}=N_\delta\ \tau_x/T_0$ and $N_{Tx}=N_\delta\ T_x/T_0$ are counted by two counters during the conversion time T_q. The duty-cycle is calculated according to the following equation:

$$N_{D.C.} = \frac{N_{\bar{\tau}x}}{N_{\bar{T}x}} = \frac{\bar{\tau}_x}{\bar{T}_x} = \overline{D.C.} \tag{49}$$

With the purpose to determine the relative quantization error, having calculated the full differential of the equation (48), proceeded to the final increments and relative units, we shall receive the following equation:

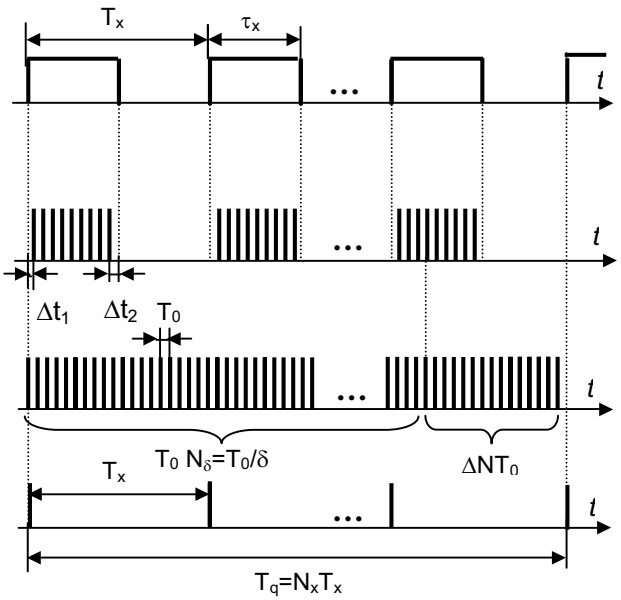


Figure 14. Time Diagrams of Novel Method for Duty-cycle – to – Digital Conversion.

$$\delta_{D.C.} = \frac{\Delta\tau_x}{\tau_x} = \frac{\Delta N_{\tau x}}{N_\delta \cdot \tau_x} \cdot T_0 + \frac{\Delta N_{Tx}}{N_\delta \cdot T_x} \cdot T_0 = \frac{1}{N_\delta \cdot f_0}\left(\frac{\Delta N_{\tau x}}{\tau_x} + \frac{\Delta N_{Tx}}{T_x}\right) \qquad (50)$$

Without the usage of any advanced conversion methods, the $\Delta N_{\tau x}$ and ΔN_{Tx} are equal to ± 1. However, using the principle incorporated in the method of the dependent count [14], the conversion interval can be chosen by multiple to the period of input signals. Hence, in this case the $\Delta N_{Tx} = 0$. Then the quantization error will be calculated according to the following equation:

$$\delta_{D.C.} = \frac{1}{N_\delta \cdot f_0 \cdot \tau_x} \cdot 100, \qquad (51)$$

As it is visible from this equation, the quantization error does not depend on the converted frequency of signal and is determined basically by the pulse width τ_x. The dependence of $\delta_{D.C} = f(\tau_x)$ for this advanced duty-cycle – to – digital conversion method is shown in Figure 15.

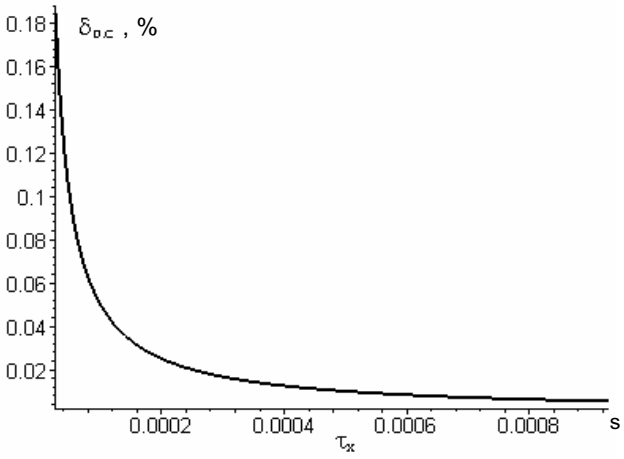


Figure 15. Dependence of $\delta_{D.C} = f(\tau_x)$ for Advanced Conversion Method.

5. CLASSICAL PHASE-SHIFT – TO – DIGITAL CONVERSION METHODS

The phase shift (φ_x) - to- digital conversion can be reduced to the conversion of the time interval t_x on which two periodic sequences of pulses with the period T_x [7] are shifted.

The essence of the classical widespread method of the phase-shift – to – digital conversion consists in the following. Sinusoidal voltage V_1 and V_2, the phase shift between which it is necessary to measure are converted into short unipolar pulses (Figure 16). The strobing pulse t_x is formed from the first pair of pulses 1 and 2. It is filled out by the reference frequency pulses f_0. The number of pulses coming into the counter during the interval t_x:

$$n = f_0 \cdot t_x \tag{52}$$

A strobing pulse, equal to the period of the converted sinusoidal voltage, is formed in parallel. This pulse is also filled out by the reference frequency pulses f_0. The number of counted pulses during the period T_x is:

$$N = f_0 \cdot T_x \tag{53}$$

The phase-shift is calculated according to the following equation

$$\varphi = \frac{360 \cdot n}{N} \tag{54}$$

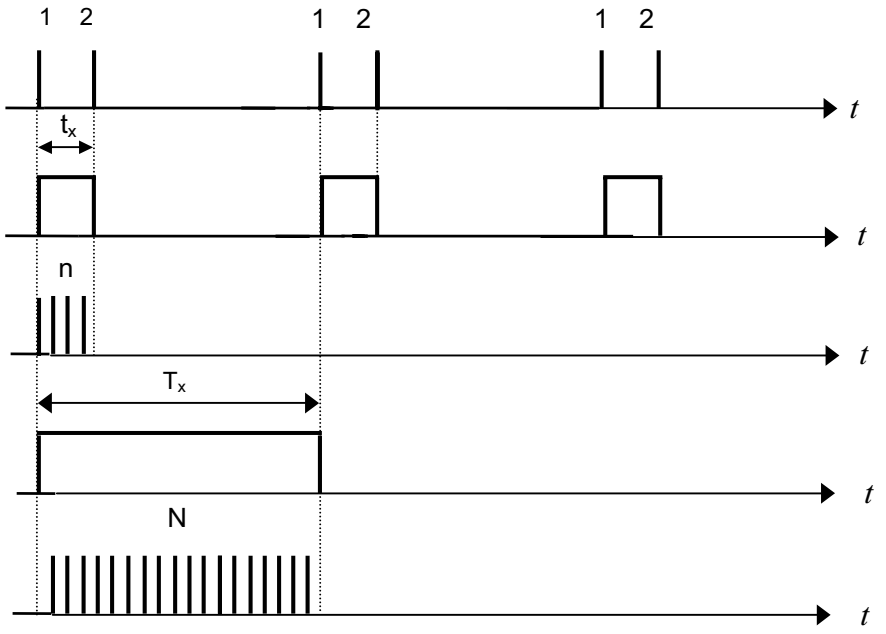


Figure 16. Time Diagrams of Method for Phase Shift-to-Digital Conversion.

The microcontroller core embedded in a smart sensor allows the choice of period necessary to determine the phase shift as well as to observing phase shifts fluctuations. There is also the possibility to convert average phase shifts during q periods.

However, by using this method for the phase shift-to-code conversion, high accuracy can be achieved only in low and infralow frequency ranges. In order to increase the conversion resolution and accuracy, the interpolation method, a method based on multiplication by f_x of the time interval proportional to the converted shift, or multiplication of the reference frequency f_0 by f_x can be used.

Example: At $\Delta\varphi_q = 0.1^0$ and $T_0 = 10^{-6}$ ($f_0 = 1$ MHz) the greatest possible frequency of the converted signal should not exceed $f_{xmax} = 277.77$ Hz. The low frequency is not limited.

Another classical method of the phase shift-to-digital conversion consists in the phase shift conversion for some periods during the constant conversion time T_{cycle}. Thus, the high converted frequency considerably extended; however, the conversion time will be increased.

6. ADVANCED PHASE-SHIFT – TO – DIGITAL CONVERSION METHODS

The phase shift φ_x between two periodic sequences of pulses with the period T_x can be converted by the *method of coincidence* [26]. Time diagrams of the method are shown in Figure 17. In this case the number N_1 pulses with the period T_0 and the number N_1' pulses of the first sequence T_x between coincident pulses of these sequences is counted. Then

$$N_1 \cdot T_0 = N_1^{'} \cdot T_x \tag{55}$$

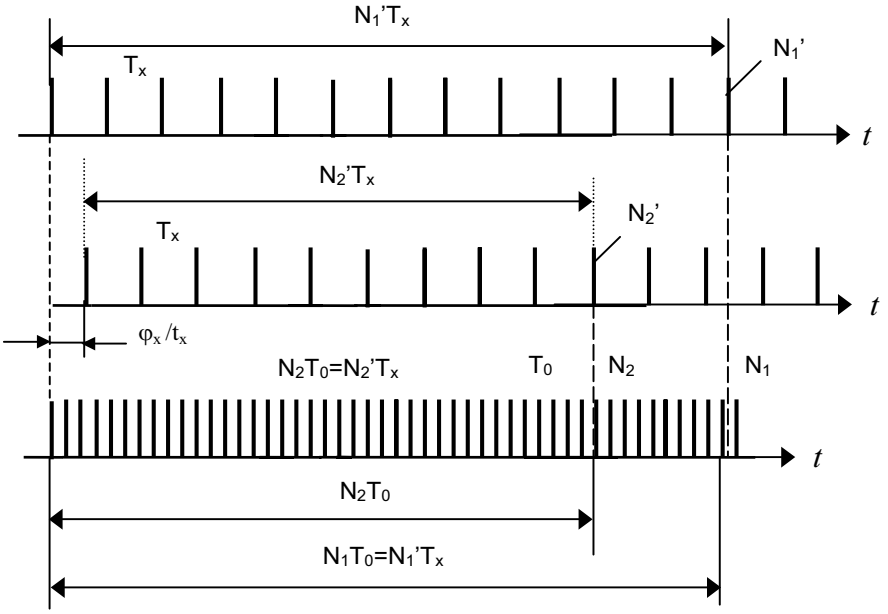


Figure 17. Time Diagrams of the Method of Coincidence for Phase Shift-to-Digital Conversion.

Similarly, the number N_2 of pulses T_0 and the number N_2' of the second sequence with the period T_x, shifted on the t_x and taking place between the first moment of coincidence of the first pair of pulses and the nearest moment of coincidence of the second pair of the pulse are counted. Then

$$N_2 \cdot T_0 = N_2^{'} \cdot T_x + t_x \tag{56}$$

From these two equations, we receive the formula for the phase-shift calculation:

$$\varphi_x = \left(\frac{N_1^{'} \cdot N_2 - N_2^{'} \cdot N_1}{N_1} \right) \cdot 360 \qquad (57)$$

and for the converted time interval:

$$t_x = \left(\frac{N_1^{'} \cdot N_2 - N_2^{'} \cdot N_1}{N_1} \right) \cdot T_0 \qquad (58)$$

The analysis of equations (57) and (58) shows, that φ_x and t_x do not depend on the period T_x. Conversion errors will be determined mainly by pulses duration only. For reduction of these errors, the method of forming of pulse packets of coincidences can be used. Thus, the absolute error of measurement for t_x can be reduced up to $0.5 \cdot 10^{-12}$s and the absolute conversion error for the phase shift φ_x - up to 0.05^0 at 1 MHz.

The considered above method of coincidence has the following disadvantages: the difficulty of exact indication of coincidences, and also the complexity of subsequent processing for result (three operations of multiplication, one division and one subtraction). In turn, it results in the increased conversion time.

In order to eliminate the mentioned disadvantages, the advanced method for phase-shift – to – digital conversion suitable for the usage in smart sensors and systems has been proposed [27]. The method is based on determination of average time interval and average period during the conversion time, multiple to the period of signal T_x. Due to this, the error by reason of non multiplicity of conversion time T and period T_x is eliminated. Besides, frequency conversion range is extended up to infralow frequencies. In comparison to the advanced method, described in [26], the conversion time is determined by the error of period conversion $\delta_{Tx} = 1/N_{\delta}$, e.g. $T=N_{\delta}\ T_0 + (0...T_x)=nT_x$. The time diagrams of method are shown in Figure 18.

The calculates for the phase-shift should be done according to the following formula:

$$N_{\varphi x} = 360 \frac{N_{\bar{t}x}}{N_{\bar{T}x}} = 360 \frac{\bar{t}_x}{\bar{T}_x} = \bar{\varphi}_x^{\ 0} \qquad (59)$$

or

$$N_{\varphi x} = 2\pi \frac{\bar{t}_x}{\bar{T}_x} = \bar{\varphi}_x (\text{rad}) \tag{60}$$

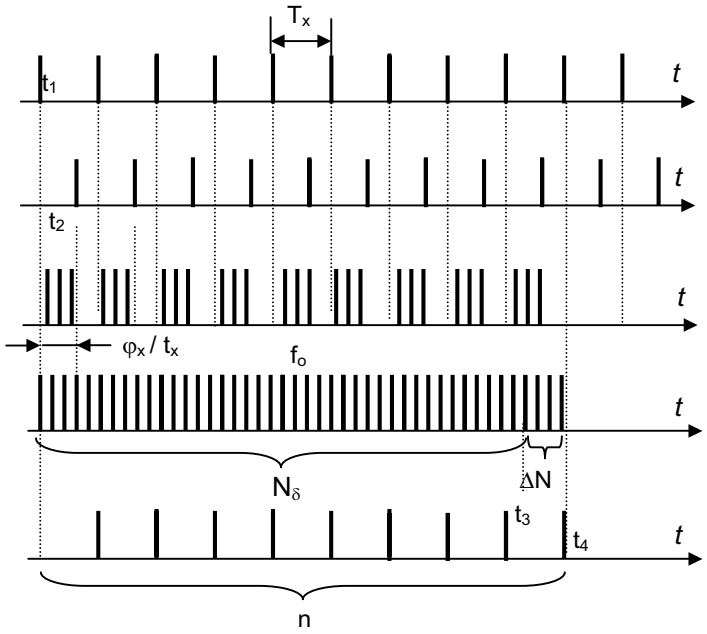


Figure 18. Time Diagrams of Advanced Method for Phase Shift – to – Digital Conversion.

The microcontroller can also calculates the derived parameters as *cos* φ_x, *sin* φ_x, etc.

The relative error for frequency- or period-to code conversion is $\delta_{Tx} = \delta_{fx} = 1/N_\delta$. It does not depend on the frequency in all range of converted frequencies.

With the purpose to determine the relative quantization error, having calculated the full differential of the equation (1), proceeded to the final increments and relative units, we shall receive the following equation:

$$\delta_{\varphi x} = \frac{\Delta\varphi_x}{\varphi_x} = \frac{\Delta N_{tx}}{N_\delta \cdot t_x} \cdot T_0 + \frac{\Delta N_{Tx}}{N_\delta \cdot T_x} \cdot T_0 = \frac{1}{N_\delta \cdot f_0} \left(\frac{\Delta N_{tx}}{t_x} + \frac{\Delta N_{Tx}}{T_x} \right) \tag{61}$$

Without the usage of any advanced conversion methods, the ΔN_{tx} and ΔN_{Tx} are equal to ± 1. However, using the principle incorporated in the method of the dependent account [4], the conversion interval can be chosen

by multiple to the period of input signals. Hence, in this case $\Delta N_{Tx} = 0$. Then the quantization error will be calculated according to the following equation:

$$\delta_{\varphi x} = \frac{1}{N_{\delta} \cdot f_0 \cdot t_x} \tag{62}$$

As it is visible from this equation, the quantization error does not depend on converted frequency of signal and is determined basically by the duration of pulses t_x. The range of measurement for time intervals at the usage of modern direct conversion methods is $2 \cdot 10^{-9} \div 1$s [28]. However, for realization of the proposed conversion method for the phase shift, the following condition is necessary: $T_0 << t_x$.

The dependence of $\delta_{\varphi x} = f(t_x, f_0)$ at $N_{\delta} = 10\ 000$ is shown in Figure 19. The proposed conversion method allows to measure the phase shift φ_x in a wide range of frequencies of input signals.

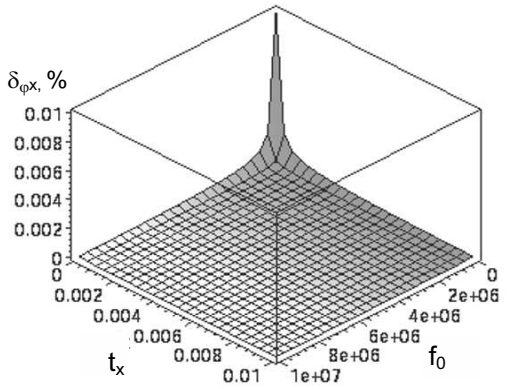


Figure 19. Dependence of $\delta_{\varphi x} = f(t_x, f_0)$ at $N_{\delta} = 10\ 000$.

7. CONCLUSIONS

Further development of microelectronic technologies, microsystems and smart sensors stimulate perfecting the known and development of new advanced methods for frequency-time domain parameters-to-digital conversion with aim to:

- to increase accuracy, speed and metrological reliability of the measurement of absolute and relative values, and their ratio as well;
- to expand functionality and the conversion range;

- to automatize completely all procedures of measurement, control, digital processing, parametric adaptation and self-diagnostics;
- to simplify the circuitry or minimize a chip area;
- to reduce the cost, weight, dimensions, power consumption, etc.

Due to numerous disadvantages, the classical frequency, duty-cycle and phase-shift conversion methods cannot be used in modern smart sensors.

The method of the dependent count and method with non-redundant reference frequency are optimal for the microcontroller based frequency-to-digital converters of a new generation for the measurement of absolute and relative frequencies in self-adaptive smart sensors. The accuracy of conversion is one of most important quality factors for smart sensors. First of them allows to choose automatically the required conversion time with the aim to provide the set value of the conversion error. The second method allows to choose automatically the reference frequency on the set value of the conversion error, saving thus the power consumption in those conditions of the measurement when a precise conversion is not required. In comparison with other advanced conversion methods having the redundant reference frequency, the power consumption reduction is possible in more than two orders.

All these open a prospect to use such programmable sensors in different real time applications, for example, ABS, as well as for measuring and data acquisition systems for various high-speed behaviour processes.

The novel advanced duty-cycle and phase-shift conversion methods can be used in different smart sensors and transducers, as well as in data acquisition (DAQ) boards for frequency-time parameters of electric signals.

REFERENCES

[1] Kirianaki N.V., Dudykevych V.B., *Methods and Devices of Digital Measurement of Low and Infralow Frequencies.* Lviv: Vyshcha Shkola, 1975 (In Russian).
[2] Kirianaki N.V., Gayduchok R.M., *Digital Measurements of Frequency-Time Parameters of Signals.* Lviv: Vyshcha Shkola, 1978 (In Russian).
[3] Haward, A.K., Counters and Timers. Electronic Engineering 1979; **51** (630): 61-70.
[4] *BurrBrown Applications Handbook*, USA, 1994.
[5] Kasatkin A.S. *Automatic Processing of Signals of Frequency Sensors*. Moscow: Energiya, 1966 (in Russian).
[6] Yermolov R.S. *Digital Frequency Counters*. Moscow: Energija, 1973 (in Russian).
[7] Mirskiy G. Y. *Electronic Measurements.* Moscow: Radio i Svyaz', 1986 (in Russian).
[8] Yurish S.Y. Program-oriented Methods and Measuring Instruments for Frequency-Time Parameters of Electric Signals. Ph.D. Thesis, State University Lviv Polytechnic, 1997 (In Ukrainian).
[9] Shvetskiy B.I. *Electronic Digital Measuring Instruments*. Kiev: Tekhnika, 1991 (in Russian).
[10]Ohmae T., Matsuda T., Kamiyama K., Tachikawa M. A Microprocessor-Controlled High-Accuracy Wide-Range Speed Regulator for Motor Drives. IEEE Transactions on Industrial Electronics 1982; IE-29:207-211

[11]Bonert R., Design of a High Performance Digital Tachometer with Microcontroller. IEEE Transactions on Instrumentation and Measurement1989; 38:1104 –1108
[12]Prokin M., Double Buffered Wide-Range Frequency Measurement Method for Digital Tachometers. IEEE Transactions on Instrumentation and Measurement 1991; 40:606-610.
[13]Prokin M. DMA Transfer Method for Wide-Range Speed and Frequency Measurement. IEEE Transactions on Instrumentation and Measurement 1993; 42:842 - 846.
[14]Pat. 788018 (USSR), Method of Measurement of Frequency and Period of Harmonic Signal and Device for its Realisation, Kirianaki N.V., Berezyuk B.M., 1980 (in Russian).
[15]Kirianaki N.V., Yurish S.Y., Shpak N.O.. Dynamic Performances of Smart Sensors Based on Self-Adaptive Frequency-to-Code Conversion Method. Proceedings of IFAC Workshop on Programmable Devices and Systems (PDS'2001); 2001 November 22-23; Gliwice, Poland.
[16]Kaushik R. "Power Analysis and Design at System Level". In *Low Power Design in Deep Submicron Electronics,* W. Nebel and J. Mermet, ed. London: Kluwer Academic Publishers, NATO ASI Series, Vol. 337, 1997.
[17]Yurish S. Y., Kirianaki N. V., Shpak N. O. Accurate Method of Frequency-Time Measurement with Non-Redundant Reference Frequency. Proceedings of the Conference on Actual Problems of Measuring Technique (Measurement-98); 1998 September 7-10; Kyiv. Ukraine.
[18]MSP430 Newsflash, No. 1, April 2002.
[19]MSP430 Family Metering Application Report, Texas Instrument Data Book, Version 3.0, April 1997.
[20]Sensors Web Portal http://www.sensorsportal.com
[21]Gerard C. M. Meijer. Concepts and Focus Point for Intelligent Sensor Systems. Sensors and Actuators 1994*;* A 41-42:183-191
[22]Middelhoek S., French P.J., Huijsing J.H. and Lian W.J, Sensors with Digital or Frequency Output. Sensors and Actuators 1988*;* 15:119-133
[23]Jarkko Vuori, Simple Method Measures Duty Cycle. EDN Magazine 1997; 3
[24]Guan Chao, Gerard C.M. Meijer. A Novel Method of Reading the Time-Domain Sensor Signals. Proceedings of ProRISC; 2001 November *29 – 30;* Veldhoven, The Netherlands.
[25]Yurish S.Y., Kirianaki N.V. Novel Duty-Cycle – to – Digital Conversion Method. Proceedings of 8th International Workshop on ADC Modelling and Testing (IWADC'03); 2003 September 8-10; Perugia, Italy.
[26]Ornatsky, P. P., *Automatic Measurements and Measuring Instruments (Analogue and Digital)*. Kiev: Vyshcha shkola, 1986 (in Russian).
[27]Kirianaki N.V., Yurish S.Y., Shpak N.O. Programmable Measurement Converter Based on Aadvanced Methods for Phase Shift- and Frequency- to - Code Conversion. Proceedings of IFAC Workshop on Programmable Devices and Systems (PDS' 03)*;* 2003 February 11-13; Ostrava, Czech Republic.
[28]Denbnovetskiy S.V., Kokoshkin S.M., Shkuro A.N. *Nanosecond Chronometry*. Kiev: Tehnika, 1991 (in Russian).

Chapter 3
RESONANT PIEZOELECTRIC DEVICES AS PHYSICAL AND BIOCHEMICAL SENSORS

Fabien Josse
Microsensor Research Laboratory and Department of Electrical and Computer Engineering, Marquette University

Richard W. Cernosek
Micro-Analytical Systems Department, Sandia National Laboratories

Abstract: Acoustic wave devices based on piezoelectric crystals and used for materials characterization and biochemical sensor applications are covered. The various acoustic wave devices used for physical and biochemical sensing applications are briefly described. Two types of sensors under development are discussed in details. They are the thickness shear mode (TSM) resonators and the guided shear horizontal surface acoustic (guided SH-SAW) devices, also commonly known as Love wave devices. It is noted that the two types of devices can be used for sensing in gas and/or liquid phase. The effectiveness of the TSM resonator for polymer material characterization is presented and discussed. The impedance-admittance characteristics of the equivalent circuit models of both the unperturbed and coated resonators are analyzed to extract the polymer storage modulus and loss modulus (G' and G''). The design and performance of guided shear horizontal surface acoustic wave (guided SH-SAW) devices being investigated and under development for high sensitivity chemical and biochemical sensors in liquids are discussed. It is noted that despite their structural similarity to Rayleigh SAW, SH-SAWs often propagate slightly deeper within the substrate, hence preventing the implementation of high sensitivity detectors. The device sensitivity to mass and viscoelastic loading can be increased using a thin dielectric guiding layer on the device surface. Suitable design principles for these sensor platforms are discussed with regard to wave guidance, electrical passivation of the interdigital transducers (IDT) from the liquid environments, acoustic loss, and sensor signal distortion. Results of chemical sensing and biosensing experiments are presented and discussed.

Keywords: resonant piezoelectric device, biochemical sensor, TSM resonator, guided SH-SAW device, polymer material characterization, acoustic wave device.

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 91-123.

1. INTRODUCTION

Acoustic wave technology has spawned a new class of sensors that use piezoelectric substrates to convert electrical energy into mechanical energy or vice versa. An electrical signal applied to a transducer is converted to an acoustic wave via the piezoelectric effect. In delay line devices, the wave is received at an output transducer where it is converted back to an electrical signal. Any interaction between a measurand and the acoustic wave causes a change in the propagating (or resonating) characteristics of the wave such as its frequency and energy loss.

Acoustic wave devices have been widely investigated for the detection of hazardous compounds in gas environments [1-15]. In those applications, the device is coated with a chemically sensitive material and the results are obtained via changes in the physical properties (mass, viscoelasticity, permittivity and conductivity) of the coating. With a few exceptions, changes in the device response are most often attributed to change in mass of the coating at the surface of the crystal. It is noted that other factors such as changes in the viscoelasticity of the coating also affect the device response. These changes manifest themselves in shifts in operating frequency and attenuation, a measure of detected compounds. Although research in the area of gas sensing with coated-acoustic wave devices has been widely reported [1-15], few studies have investigated the direct detection of organic contaminants in aqueous solutions. Most investigations for liquid phase detection involved biosensor applications or uncoated devices for the characterization of liquid properties, such as density, viscosity, dielectric constant and conductivity [1-3]. Over the past decade, there has been a great deal of interest in the development of acoustic wave-based liquid sensors [1, 16-25]. Direct detection in liquid environments has proven to be difficult, due to various reasons ranging from the type of acoustic wave used to the sensing coating. A number of acoustic wave devices have been investigated for liquid-phase chemical sensors and biosensors. They include thickness shear mode (TSM) resonator, shear horizontal acoustic plate mode (SH-APM) device, surface acoustic waves (SAW) devices with exclusively shear horizontal (SH) particle displacement, lamb (or flexural plate) wave, and Love wave devices. They facilitate the direct and reagent-free detection of molecules and are suitable for real-time, on-line monitoring.

In addition to their use as detectors of biochemical analytes in gas or liquid phase, acoustic wave devices are also ideally suited to thin film characterization due to their extreme sensitivity to surface perturbations. Acoustic waves provide nondestructive probing, since they typically operate

at low powers and do not change the properties of the entities being probed. They can monitor a wide range of physical properties in an in-situ manner. Of all the acoustic waves, the TSM resonator is the most commonly used acoustic wave device for material physical property characterization because of its simplicity and ability to support shear horizontal waves, which are ideal for the characterization of the complex shear modulus of thin viscoelastic films.

In the present chapter, two types of acoustic wave devices with predominantly shear horizontal particle displacement used as physical and biochemical sensors in liquid environments will be presented. They are the thickness shear mode (TSM) resonator and the guided shear horizontal surface acoustic wave (SH-SAW) device also commonly known as Love wave device. In the first part, the modeling of TSM resonator for sensing and material characterization will be reviewed. Expressions are derived and used for the rapid fitting of an equivalent circuit model to the measured admittance data, which allow the rapid extraction of the physical properties of the surface perturbations. Results are presented for a polymer material, an organic resin known as SU-8. In the second part, the design and performance of guided SH-SAW (Love wave) devices being investigated for high-sensitivity biochemical sensors in liquids are presented. Results are presented for direct chemical sensing and biosensing in liquid environments using sensor platforms on 36^{o} rotated Y-cut $LiTaO_3$.

2. THICKNESS-SHEAR MODE (TSM) RESONATOR

The TSM resonator is very sensitive to surface perturbations. These surface perturbations can include mechanical loadings such as an added ideal mass layer, a Newtonian or non-Newtonian liquid, a viscoelastic layer, or combinations of the individual types. Several methods have been used to interpret TSM resonator response depending on the properties of the perturbation; among them are the Sauerbrey equation and the crystal impedance method [1, 3]. The Sauerbrey equation relates mass accumulated on the crystal to the change in the resonant frequency [26]. However, this method is only valid for a rigid and acoustically thin film. Since the mass effect is not the only effect that influences the acoustic behavior of the TSM device, the Sauerbrey method is unable to distinguish between changes in mass load on the surface and the viscosity changes of a liquid or changes in the viscoelastic properties of a coating. To overcome these limitations, the

crystal impedance method has been developed [27-29]. This method allows the extraction of physical parameters (viscosity, density, shear modulus) from complex mechanical loading on the TSM resonator.

The TSM resonator consists of a thin disk of AT-cut quartz with metal electrodes deposited on both sides. Due to the piezoelectric properties and crystalline orientation of the quartz, application of alternating potential to the electrodes of the resonator results in an internal mechanical stress and consequently a shear deformation of the crystal. Resonance will occur when the crystal thickness is an odd multiple of half the acoustic wavelength. If a medium is in contact with one or both resonator surface, the excitation of the crystal subjects the medium on the crystal to an oscillatory driving force. Due to the electromechanical coupling, the mechanical properties of the contacting medium are reflected in the electrical response of the resonator.

It is convenient to use an equivalent circuit model to describe the electrical behavior of the TSM resonator. With only a few lumped elements, the modified Butterworth-VanDyke (BVD) model simulates the electrical characteristics of the TSM resonator over a range of frequencies near resonance. This model can explicitly relate the lumped elements in the circuit to physical properties of the TSM, and the surface load (liquid load and/or viscoelastic film).

2.1 Theory

The quartz theory used in this paper is essentially identical to that used by most authors working in the acoustic sensor area and, with only a few simplifying assumptions, is the same as that using rigorous acoustic wave propagation theory with appropriate boundary conditions [30-31]. All of these other authors use electroacoustic theories that treat the quartz resonator as a second order system, and that, near mechanical resonance, the equivalent circuit (the so-called BVD model) description can be used. This theory of the quartz resonator is well-known and extremely precise for most sensor and material characterization studies.

The BVD model can be used to describe both the unperturbed and perturbed quartz crystal resonators as shown in Figure 1. The TSM resonator equivalent circuit consists of two branches: a static branch and a motional branch. A static capacitance C_0 arises between the electrodes located on opposite sides of the insulating quartz. A parasitic capacitance C_p also arises due to the test fixture. The net capacitance of the static branch, C_0*, is the sum of C_0 and $C_{p.}$ Since the quartz is also piezoelectric, electromechanical

coupling gives rise to additional motional contributions (L_1, C_1 and R_1) in the unperturbed resonator. The static branch dominates the electrical behaviors away from resonance, while the motional branch dominates near resonance.

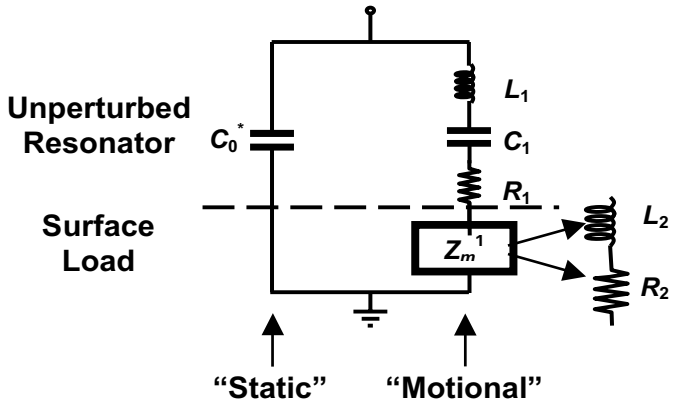


Figure 1. The BVD Equivalent Circuit Model of the Unperturbed TSM Resonator and the Modified BVD Model of the Perturbed TSM Resonator.

The modified BVD model also shown in Figure 1 can be used to describe the electrical response of the perturbed device. When the resonator has a surface perturbation, the motional impedance, Z_m, changes, as indicated by the introduction of the complex electrical load impedance Z^l_m in Figure 1 [32]

$$Z_m = R_1 + j\omega L_1 + \frac{1}{j\omega C_1} + Z_m^1 \tag{1}$$

with

$$Z_m^1 = R_2 + j\omega L_2 = \frac{N\pi}{4K^2\omega_s C_o}\left(\frac{Z_s}{Z_q}\right) \tag{2}$$

In equation (2), N is the vibration mode index, K^2 is the quartz electromechanical coupling coefficient, ω_s is the angular series resonant frequency for the unperturbed TSM resonator, $Z_q=(\rho_q\mu_q)^{1/2}$ is the quartz shear wave characteristic impedance, where ρ_q and μ_q are the mass density and shear stiffness of the quartz. The parameters for AT-cut quartz are readily available

from the literature. Z_s is the shear mechanical impedance due to the load at the device surface [1]:

$$Z_s = \left. \frac{T_{xy}}{v_x} \right|_{y=0} \tag{3}$$

where T_{xy} is the peak sinusoidal steady state shear stress imposed on the contacting medium by the resonator and v_x is the resulting x-directed surface shear particle velocity displacement. It is noted that Z_s is a complex quantity.

The electrical impedance element R_2 and L_2 can be related to the real and imaginary components of the surface mechanical impedance as:

$$R_2 = \frac{N\pi}{4K^2\omega_s C_0}\left(\frac{\mathrm{Re}(Z_s)}{Z_q}\right) \tag{4}$$

$$L_2 = \frac{N\pi}{4K^2\omega_s C_0}\left(\frac{\mathrm{Im}(Z_s)}{Z_q}\right) \tag{5}$$

A general model has been developed [29], which can incorporate a physically diverse set of single-component loads, including rigid mass layers, Newtonian or Maxwellian liquids and viscoelastic media. Models and descriptions of the TSM resonator loaded with each type of layer has been treated and can be found in the published literature. In general, a different layer presents a different surface mechanical impedance, which is characterized by a given acoustic phase shift, ϕ, and attenuation of the wave across the layer thickness. In general, multiple surface loadings on the TSM resonator cannot be treated in a linearly additive fashion. An exception to this rule exists for several ideal mass layers or a mass layer next to the resonator surface with a viscous medium contacting it. In the case of a TSM resonator with simultaneous viscoelastic and Newtonian liquid loading - representing a typical liquid chemical sensing system - the total impedance is not equal to the sum of the characteristic impedance of the individual layers. This is due to the phase shift caused by the viscoelastic layer before it interacts with the liquid (the transmission line effect). Instead, the total surface mechanical impedance at the resonator surface due to the viscoelastic coating/liquid combination can be defined as [1, 29]:

$$Z_s = Z_0\left[\frac{Z_l \cosh(\beta h_f) + Z_0 \sinh(\beta h_f)}{Z_0 \cosh(\beta h_f) + Z_l \sinh(\beta h_f)}\right] \tag{6}$$

where Z_0 , Z_l are the characteristic mechanical impedance of the viscoelastic film and the Newtonian fluid, respectively. $\beta = j\omega(\rho_f / G)^{1/2}$ is the complex wave propagation constant. ρ_f and h_f are the density and thickness of the film, respectively. Without loss of generality, equation (6) can be rearranged into a more useful form:

$$Z_s = \frac{Z_l + Z_0 \tanh(\beta h_f)}{1 + \left(Z_l / Z_0\right)\tanh(\beta h_f)} \tag{7}$$

The combined surface effects due to the viscoelastic coating and the liquid overlayer are heavily dependent on the impedance ratio, Z_l / Z_0, at the coating/liquid interface. If Z_l is comparable in magnitude to Z_0, the shear acoustic wave propagating in the viscoelastic film loses a portion of its energy to the liquid. However, if $|Z_l| << |Z_0|$, the acoustic wave sees a low impedance boundary at the coating/liquid interface and essentially all of the energy is reflected back into the viscoelastic layer. This latter case is representative of most liquid-phase sensor systems, in which high modulus polymers are used as coatings and low viscosity Newtonian liquids, primarily water, contact the coating. Under these conditions, the total surface impedance is a simple linear combination of the motional impedances due to the liquid and a finite thickness viscoelastic film: $Z_s = Z_l + Z_0 \tanh(\beta h_f)$ It is noted that when the TSM resonator is only loaded with a viscoelastic coating, $Z_l = 0$, and equation 7 reduces to the input impedance at the quartz surface/coating interface [1]:

$$Z_s = jZ_0 \tan\left(\omega\left(\frac{\rho_f}{G}\right)^{1/2} h_f\right) \tag{8}$$

In obtaining equation (8), it is noted that the identity tanh(jx) = jtan(x) has been used. Z_0 is the characteristic impedance of the viscoelastic coating, and is given by $(\rho_f G)^{1/2}$.

Since polymer films are commonly applied as sorbing layers in gas and liquid-sensing applications, it is desired to use this modeling structure to assist in characterizing them – the total system resembles the layered sensor geometry. More recently, polymer materials are being investigated as effective waveguiding layers in the implementation of guided shear-horizontal surface acoustic wave (SH-SAW) sensor platforms [25]. The TSM resonator material characterizations (using shear mode interactions) are directly applicable to the guided SH-SAW sensor problem. Polymers being viscoelastic materials, their elastic and viscous properties must be described by a complex modulus, $G = G' + iG''$. G' is the shear storage modulus or the elastic contribution and G'' represents the shear loss modulus or the viscous contribution. Polymers are usually classified by the values of G' and G''. A glassy film exists when $G' > 10^8$ dyne cm^{-2} and $G'' << G'$; a rubbery polymer when $G'' \leq G'$ and a Newtonian liquid as G' tends to zero and $G'' = \omega\eta$ (η is the liquid viscosity). It is important to understand how the TSM resonator responds to a polymer coating as described by the surface load in equation (8). A description of the dynamic behavior of viscoelastic films on the TSM is given in references [1, 33-34], and is summarized here for convenience. If the viscoelastic layer is sufficiently thin and/or rigid so that $\phi << \pi/2$, then the entire layer tends to move synchronously with the TSM surface. Under these conditions, the layer often can be treated as a simple mass layer. For $\phi \leq \pi/2$, displacement at the upper surface of the layer lags that at the device surface. This condition is typical for a lossy polymer, adding some motional (mechanical) resistance to the dynamic action. For $\phi \approx \pi/2$, film resonance occurs and the coated resonator exhibits the characteristics of coupled resonant systems. For $\phi > \pi/2$, the upper surface of the layer is 180^o out of phase with the resonator surface and the resonant frequency of the coated TSM resonator is higher than that of the uncoated resonator.

Equation (8), which relates the surface mechanical impedance to the physical parameters of the viscoelastic load, can be used to extract G' and G". This may be done numerically or analytically, typically after further mathematical simplification. Because the shear modulus G is a complex quantity, the mechanical impedance described by equation (8) does not readily decompose into simple analytical expressions needed to describe R_2 and L_2. Thus, approximate forms of Z_s are needed. The tangent function can be expanded in a Taylor series and the first two terms retained, $\tan(x) \approx x + {}^1/_3\, x^3$. This leads to [30]:

$$\tan\left(\omega h_f\sqrt{\frac{\rho_f}{G}}\right)\approx\omega h_f\sqrt{\frac{\rho_f}{G}}+\frac{1}{3}\omega^3 h_f^3\left(\frac{\rho_f}{G}\right)^{3/2} \tag{9}$$

The surface mechanical impedance of the viscoelastic layer can then be written as:

$$Z_S \approx \frac{\omega^3\rho_f^2 h_f^3}{3}\frac{G''}{|G|^2}+j\left(\omega\rho_f h_f+\frac{\omega^3\rho_f^2 h_f^3}{3}\frac{G'}{|G|^2}\right) \tag{10}$$

The first term (real component of Z_s), proportional to G" is the attenuation due to the coating. The second term (imaginary component of Z_s) represents the energy storage in the film; it consists of a mass component, $\omega\rho_f h_f$, and a viscoelastic component proportional to G'.

The resistance R_2 and inductance L_2 due to the viscoelastic layer can be deduced from equations (4, 5 and 10):

$$R_2 = \frac{N\pi}{4K_0^2\omega_s Z_q C_0}\left(\frac{\omega_s^3\rho_f^2 h_f^3 G''}{12}\right) \tag{11a}$$

$$L_2 = \frac{N\pi}{4K_0^2\omega_s^2 Z_q C_0}\left(\omega_s\rho_f h_f+\frac{\omega_s^3\rho_f^2 h_f^3 G'}{3|G|^2}\right) \tag{11b}$$

By setting

$$A=\frac{G'}{|G|^2}=\frac{12R_2K_0^2C_0Z_q}{N\pi\omega_s^2h^3\rho_f^2} \tag{12}$$

$$B=\frac{G''}{|G|^2}=\frac{12K_0^2\omega_s C_0 Z_q L_2-3\rho_f h_f N\pi}{N\pi\omega_s^2h^3\rho_f^2} \tag{13}$$

both the storage modulus *G'* and loss modulus *G''* can be extracted as

$$G' = \frac{B}{A^2 + B^2} \tag{14}$$

$$G'' = \frac{A}{A^2 + B^2} \tag{15}$$

A more precise approximation can be obtained by retaining the first three terms of the Taylor series expansion for the tangent function [35]. In this case,

$$G' = \frac{B}{A^2 + B^2}$$

$$G'' = \frac{A}{A^2 + B^2},$$

where

$$A = \left(\frac{\left(\frac{19\rho_f h}{24} - \frac{L_2}{A_0} \right) + \left[\left(\frac{L_2}{A_0} - \frac{19\rho_f h}{24} \right)^2 + \frac{R_2^2}{\omega_s^2 A_0^2} \right]^{\frac{1}{2}}}{\frac{4\omega_s^4 \rho_f^3 h^5}{15}} \right)^{\frac{1}{2}} \tag{16}$$

$$B = \frac{15R_2}{4\omega_s^5 \rho_f^3 h^5 A_0} \frac{1}{A} - \frac{5}{4\omega_s^2 h^2 \rho_f} \tag{17}$$

with

$$A_0 = \frac{N\pi}{4K_0^2 \omega_s C_0} \frac{1}{\sqrt{\mu_q \rho_q}} \tag{18}$$

The analysis for the extraction of the shear moduli can be performed using equation (8) for an exact description of the viscoelasticity or equations (11-17) when *a priori* information about the film is known and simplifying assumptions can be accommodated. Modeling of viscoelastic films involves

finding the solutions to four parameters *(G', G'', h_f,* and *ρ_f)*. However, admittance measurements contribute only two parameters, the real and imaginary part of the mechanical impedance *Re (Z_s)* and *Im (Z_s)*. Both h_f and ρ_f are used as constraints in the modeling programs. Usually a tight constraint on h_f and ρ_f can generate unique solutions of *G'* and *G''*. The crystal impedance method involves using an automatic network analyzer (ANA) to measure the reflection coefficient, $S_{11}(\omega)$, of the polymer-loaded TSM quartz resonator. The admittance (or impedance) can be obtained in the form of complex quantities from the reflection coefficient. Fitting the circuit model of Figure 1 to the measured data allows the extraction of the moduli *G'* and *G''* of the layer using steps described in Figure 2.

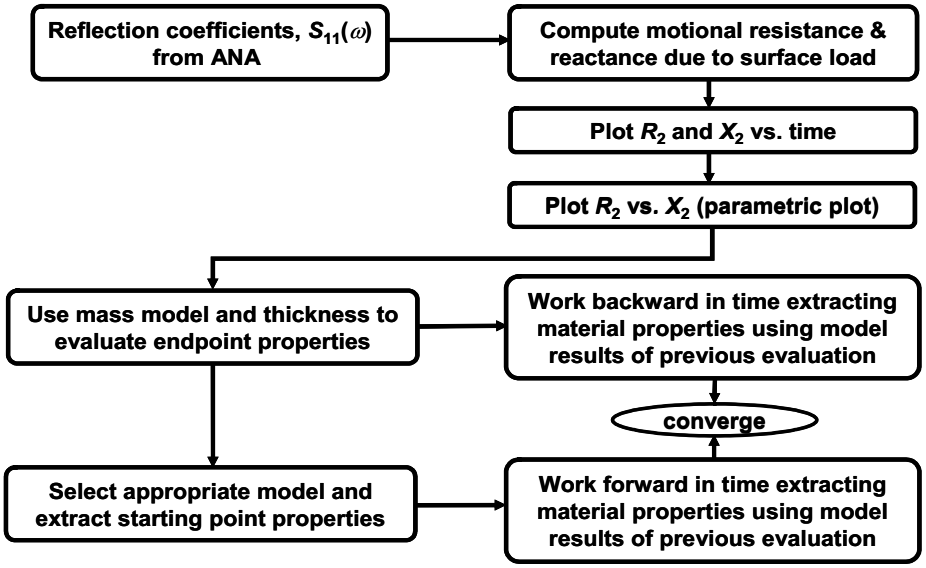


Figure 2. Flow Chart for the Response Analysis and Viscoelastic Material Parameter Extraction.

2.2 Polymer Material Characterization Using TSM Resonators

The crystal impedance method involves using a network analyzer to measure the reflection coefficient, $S_{11}(\omega)$, of the surface-loaded TSM quartz

resonator. Viscoelastic properties of several different polymers have been studied using this method; we discuss here only one glassy polymer, SU-8, as an example of the technique. SU-8 polymer is an epoxy resin being investigated as a guiding layer in guided SH-SAW sensors. A TSM resonator of 9-MHz fundamental frequency is used in the study. From the measured reflection coefficient the complex admittance or impedance can be obtained. Experimentally obtained impedance data and film thicknesses are then used in the analysis. The density of the SU-8 film ρ_f is taken to be 1.2 g/cm^3, as provided by the manufacturer. The analysis extracts the shear moduli as indicated by equations (9-17). Modeling of viscoelastic films involves finding the values for five parameters *(G', G'', h_f, ρ_f,* and *C_p)*. The parasitic capacitance C_p is included to account for practical cases in which the supporting structure and dielectric interactions may change substantially after the coating. Impedance is measured at the fundamental resonance frequency as well as the 3rd harmonic frequencies. R_1, L_1 and C_1 and C_p are obtained by fitting the experimental electrical impedance with the theoretical model of the unperturbed TSM device. These parameters for the unperturbed device are stored for later use. Once thin films are spin-deposited on the resonator surface, further measurements are gathered for the reflection coefficients and the impedance can be derived.

While there are five parameters being extracted from the analysis, admittance measurements contribute only two parameters, the real and imaginary part of mechanical impedance *Re (Z_s)* and *Im (Z_s)*. The film density, ρ_f , is taken to be that of the bulk value; the thickness, h_f , is independently measured. Both h_f and ρ_f are used as constraints in the modeling programs. C_p is estimated to be close to the unperturbed value but is allowed to vary in the fitting routine. Usually a tight constraint on h_f and ρ_f can generate unique solutions of *G'* and *G''*. Admittance magnitudes for the bare and SU-8 coated 9-MHz TSM quartz resonator are shown in Figure 3. As expected, the series resonant frequencies of the coated devices are shifted to lower values and the admittance magnitudes exhibit damping due to the film losses as the coating thickness increases. Accordingly, the TSM resonator experiences a drop in the quality factor (Q-factor), which can be observed by loss in magnitude and the widening of the series resonant peak. When the thickness of SU-8 film reaches 5 μm, the resonant frequency shift apparently begins to deviate from the linear region (frequency versus mass) and enters a nongravimetric regime, as indicated in Figure 4. This implies that the viscoelastic properties of the polymer influence the device response and significant phase lag occurs on the upper film surface.

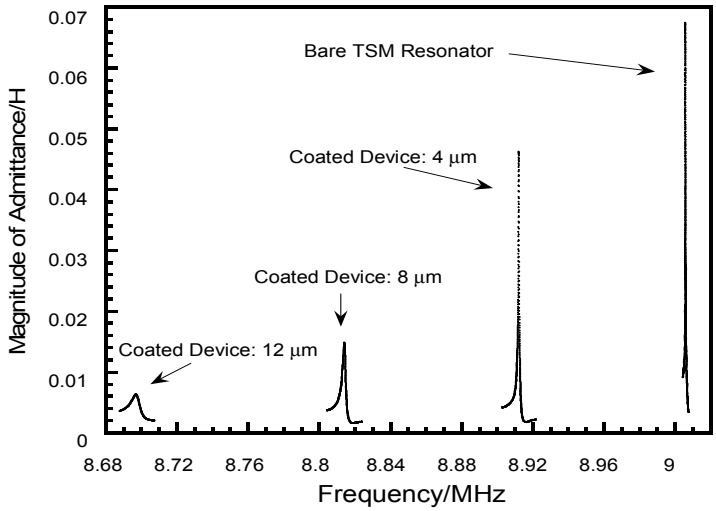


Figure 3. The Crystal Admittance Spectra of the Bare and SU-8 Coated TSM Resonators. SU-8 is Cured Following Procedures Provided by the Manufacturer.

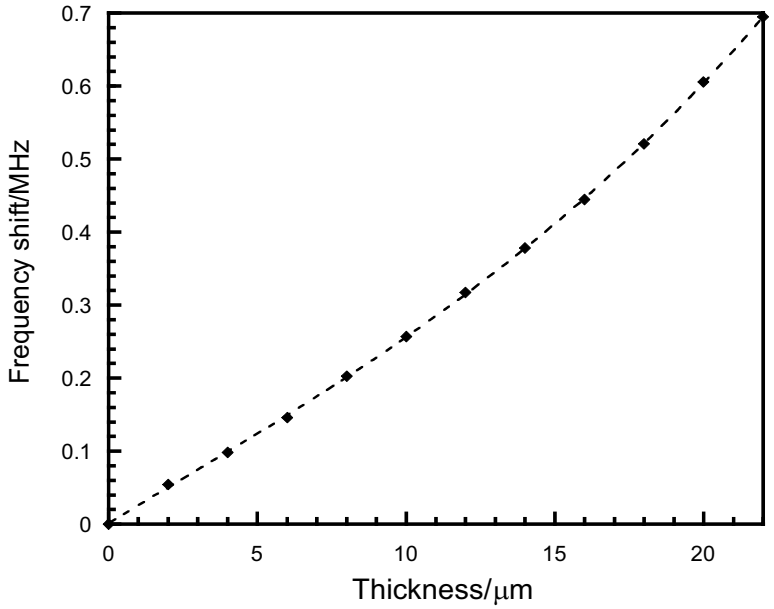


Figure 4. Measured Series-resonant Frequency Shifts as a Function of the Thickness of SU-8 Film in the Range of 0-22 μm at the Fundamental Frequency. The Device is a 9-MHz TSM Quartz Resonator.

The dynamical behavior of the viscoelastic film's displacement across the thickness h_f is determined by the acoustic phase shift, ϕ, across the film [1, 33-34] as explained in the previous section.

Using the extracted value for the shear modulus $G=1.6\times10^{10} dyne/cm^2$, the shear wave phase velocity across the polymer film can be estimated as 1,160 *m/s*. The phase shift ϕ across the layer of 5 μm polymer film is calculated to be

$$\phi = 2\pi f \times h / v \approx 0.155 \times \frac{\pi}{2} < \frac{\pi}{2} \tag{19}$$

Thus, a linear relationship between frequency shift and film thickness is observed in the range of about 0-5 μm. Beyond that, frequency shift increases more rapidly with phase shift ϕ (in other words, the thickness of the film) until $\phi = \pi/2$. In this region, one can determine the shear modulus of the films and study the cross-linking of polymer films initiated by heat or light exposure.

The above discussion is confirmed by the measured device resistance R ($R=R_1+R_2$) shown in Figure 5 for a TSM resonator coated with SU-8 films at varying thicknesses. The resistance R represents the mechanical energy dissipation of the system. The change in R remains small (less than 20 ohm) when $0 < h_f < 5$ μm (corresponding to a phase shift $\phi < 0.155 \times (\pi/2)$). The damping of the admittance is not large in this region as indicated in Figure 3. For film thickness such that 5 μm $< h_f <$30 μm, R increases sharply then gradually approaches a maximum value. Film resonance occurs when $\phi = \pi/2$ and the resistance becomes maximum [34]. According to the experimental value, the SU-8 experiences film resonance near $h_f \cong 30$ μm. Film resonance represents the strongest coupling between the resonator and the coated film, resulting in a greater extraction of acoustic energy from the resonator. The general resistance change as a function of the film thickness in Figure 5 concurs with the results reported by Martin and Frye [28] for another polymer film, (polyisobutylene, PIB). The resistance of SU-8 is approaching but has not reached the maximum value at the thickness of 22 μm. This result is quite consistent with the prediction that the film resonance occurs at around $h_f \cong 30\ \mu m$. Furthermore, it verifies the reliability of the fitting results of the shear moduli (G', G'') for the SU-8 film.

Measurements were also performed with the TSM resonator operating at the third harmonic (~27 MHz). The phase shift across the polymer coating at this frequency will be approximately three times as well.

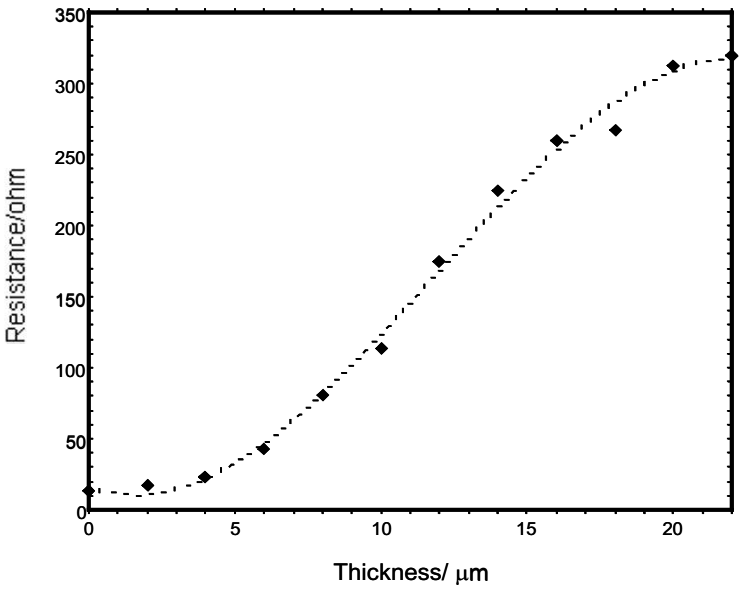


Figure 5. The Resistance R of the Device ($R=R_1+R_2$) as a Function of Polymer Thickness Measured at the Fundamental Frequency. R_2 is the Resistance Associated with the Viscoelastic Polymer. The Device is a 9-MHz TSM Quartz Resonator.

This results in a larger "effective" thickness when performing the analysis at the third harmonics. The measured device resistance approaches saturation at a thickness of approximately 10 μm, agreeing with the earlier predictions that for film resonance. Figure 6 shows the results for shear moduli extracted from measurements made at both the fundamental frequency and the third harmonic for the SU-8 coated TSM device. The moduli increase with film thickness and the shear storage modulus exceeds the shear loss modulus by more than an order of magnitude. As the film thickness increases above 10 μm, both the storage modulus (G') and loss modulus (G'') approach saturation indicating that the extracted viscoelastic properties are now independent of thickness. A thicker film layer is desirable for this analysis because of the larger viscoelastic contribution, leading to smaller errors in the analysis. For a relatively small film layer thickness, the extracted viscoelastic properties do not reflect that of a bulk film material. This essentially means that the phase shift across the thin film layer is practically negligible. Under these conditions, the extraction of the storage and loss modulus could be quite

inaccurate. These figures confirm earlier discussion that meaningful values of viscoelastic properties can only be obtained at sufficiently large polymer thicknesses. The storage modulus (*G'*) of the SU-8 film (assuming $\rho_f = 1.2\ g/cm^3$) approaches its saturation value of 1.66×10^{10} dyne/cm^2 for relatively thick films (10-20 μm), while the loss modulus (*G''*) reaches its maximum at 6.0×10^8 dyne/cm^2. Although the accuracy of the extracted shear modulus parameters occurs at high thickness where viscoelastic contribution is substantial, it is noted that, at these thicknesses, the measured loss is quite large and approaches saturation as shown in Figure 5. The modeling used here will result in relatively large errors for acoustically thin films. For rigid layers, the loss contribution is negligible (coating mostly represents a mass load), and extraction of *G'* and *G''* is not possible. In this layer thickness region (gravimetric region), the electrical response of the devices is only dependent upon the surface mass density ρ_s. It must be pointed out that the polymer thickness and density (taken to be fixed values in the analysis) are the major sources of error in the extraction of *G'* and *G''*.

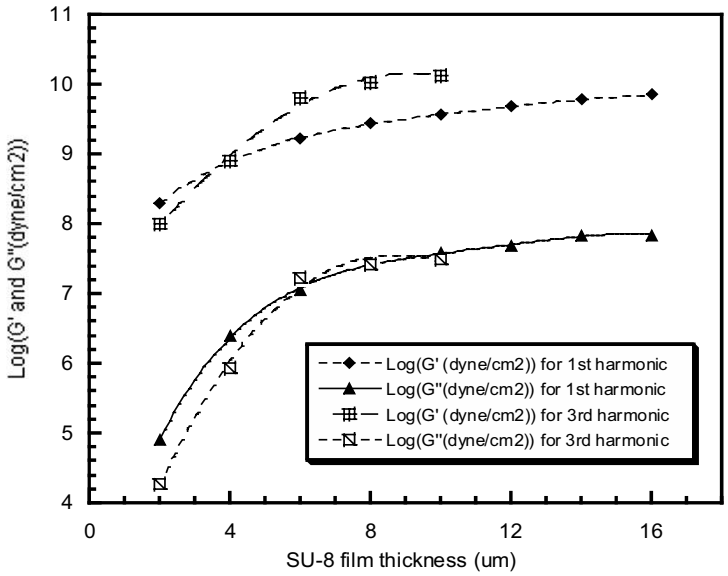


Figure 6. The Extracted Shear Moduli as a Function of Polymer Thickness for SU-8 Films Using a 9-MHz TSM Resonator. Measurements are Conducted at the Fundamental Frequency and Third Harmonic. Shown are Both Storage and Loss Moduli.

Apparently, the SU-8 polymer film is glassy due to the high value of *G'* and high ratio *G'/G''*. The storage modulus of SU-8 is comparable to that of PMMA, which was reported to be 1.43×10^{10} dyne/cm^2 at a thickness of 17.6 μm [35].

When the film density and thickness are not available, a unique solution usually cannot be obtained if the crystal impedance measurements are made only at the fundamental frequency. By measuring the electrical response of the TSM resonator at various harmonics, more information is available and extraction of unique moduli is possible [30]. However, the measurement and analysis using the multi-harmonic impedance method overlooks the frequency dependence of shear moduli in the viscoelastic regime.

3. GUIDED SH-SAW SENSORS FOR CHEMICAL AND BIO-CHEMICAL DETECTION IN LIQUIDS

Acoustic wave devices for liquid-phase sensing applications [1, 16-25], where the device is in direct contact with the solution are presented. Here, new challenges are encountered, including additional loss contributions and signal distortions due to the liquid being in contact with the acoustic wave and/or the interdigital transducers (IDTs). Other challenges include the effects of the liquid seals on signal quality and reproducibility, and the transport of the target analyte or antigen to the sensor surface.

Of all acoustic wave devices, SH-SAW devices appear most promising for (bio-)chemical detection in liquid environments: (1) shear horizontal surface waves are more sensitive than bulk waves to perturbations produced from the environment without the excessive loss associated with Rayleigh SAW in liquids; (2) the selected piezoelectric materials and transducer designs lead to very high *Q* (quality factor) structures; (3) device frequencies can be scaled to > 100 MHz, so device sensitivity can be high provided that noise is decreased; and (4) devices are small, robust, and easy to incorporate into on-line low cost systems. However, despite their structural similarity to Rayleigh SAWs, SH-SAWs often propagate slightly deeper within the substrate (in some cases, referred to as surface skimming bulk waves) [36-39], hence preventing the implementation of high sensitivity detectors. The device sensitivity to mass and viscoelastic loading s can be increased using a thin guiding layer on the device surface. The effect of the overlayer is to trap the acoustic energy near the sensing surface [36-39], thus increasing the sensitivity to surface perturbations. The resulting acoustic wave is analogous to a Love wave on an

isotropic substrate with overlayer. The two most commonly used crystal orientations are the rotated Y-cuts quartz 90°- X, and the 36° rotated Y-cut $LiTaO_3$.

Various dielectric materials such as silicon dioxide (SiO_2), silicon nitride (Si_3N_4), and most polymers can be used as the waveguide material. Polymers have an advantage over other waveguide materials due to their relatively low shear wave velocity [18] and the ease of surface layer preparation. However, because of their viscoelasticity, cross-linking or curing is necessary to avoid excessive acoustic loss [18]. Cross-linking allows the acoustically lossy polymer to exhibit an equilibrium elastic stress, thus representing a stable waveguide layer.

However, it is not desirable to simply minimize the acoustic loss of the overlayer. An overlayer with moderate acoustic loss will help to suppress the triple transit echo (TTE) (for delay lines), which is one of the major sources of signal distortion for low loss devices. In addition, the overlayer will decrease the acoustic velocity of the SH-SAW, thus reducing signal distortion due to overlap and interference with adjacent bulk waves.

The dielectric overlayer can also act to passivate the IDTs from the contacting liquid, which often is conductive or has a high dielectric constant (like water). Polymers as the passivating coating have a relatively low dielectric constant and will provide an insulating barrier, assuming the polymer is properly cross-linked or cured to eliminate fluid uptake.

Finally, the overlayer has to provide a suitable basis for attachment of selective layers, like antibodies, enzymes, or chemically selective polymers that will be used to complete the total sensor.

All the above issues must be taken into account when choosing the appropriate waveguide material, determining its thickness, and pretreatment (deposition rate, cross-linking process, cure schedule, etc.), and investigating the general design parameters for acoustic wave sensors like device configuration (delay line or resonator), IDT geometry and substrate material.

In the present section, guided SH-SAW devices as sensor platforms are presented for the implementation of high sensitivity bio-chemical detectors in liquid environments.

3.1 Fundamentals of Guided SH-SAW Sensors

Figure 7 shows the basic configuration of a guided SH-SAW device as a sensor platform. It consists of a SH-SAW device (in this case a delay line) with an overlayer of thickness, h, in the x_3 direction, having a lower shear

wave velocity. A thin metal layer is used between the two IDTs, representing an electrical short and eliminating acoustoelectric interactions with the load. As a result, only sensing caused by mechanical loading is considered here. The overlayer can be, in reality, a composite layer, which consists of the waveguiding layer (h_1) and the biochemically sensitive layer (h_2) with or without attached receptors. A liquid layer ($X_3 > h$) provides the transport medium for the target species. Perturbation theory, which yields approximate sensitivity expressions, is often used provided that the perturbing film(s) is sufficiently thin in comparison to the acoustic wave wavelength. The problem may be solved as a 3-layer or a 4-layer problem. In a 3-layer geometry, the layer serves as both the guiding layer and bio-chemically sensitive layer. The guided acoustic wave is assumed to propagate along the X_1 direction and have no variation along X_2.

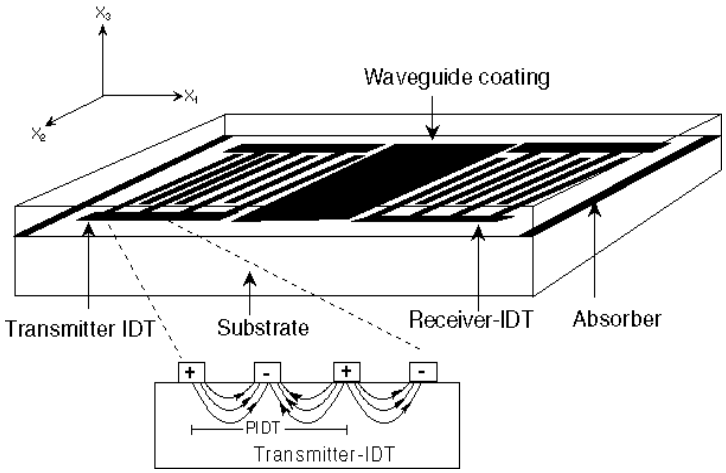


Figure 7. Geometry of the Guided SH-SAW Sensor Platform. SAWs with Shear Horizontal (SH) Displacements are Excited by the IDTs. Crystal Orientations Include the 36° Rotated Y-cut $LiTaO_3$, and the Rotated Y-cuts Quartz 90°-X.

Dependence of the amplitude of the wave upon time t and path of propagation x_1 was assumed as: $exp[j(\omega t-\beta x_1)] = exp[j\omega t-j(k-j\ \alpha)x_1] = exp[j\omega t]\ exp[-jkx_1-\alpha x_1]$ and α is real number. A complex propagating factor β is defined by the wave number *k* and attenuation α as

$$\beta = k - j\alpha = \frac{\omega}{V} - j\alpha \tag{20}$$

The variation of the complex propagating factor can be derived as

$$\frac{\Delta\beta}{k} = -\frac{\omega\Delta V}{V^2}\cdot\frac{V}{\omega} - j\frac{\Delta\alpha}{k} = -\frac{\Delta V}{V} - j\frac{\Delta\alpha}{k} \tag{21}$$

The next step in solving the problem consists of evaluating the mechanical loading perturbation due to the overlayers. Approximate expressions for the variations in wave velocity and attenuation, which define the sensing, can be obtained using mechanical perturbation theory together with surface impedance techniques. The details of this theory can be found in the literature. The fundamental equation for the mechanical loading is expressed by

$$\frac{\Delta\beta}{k} = -\frac{jV(v^*\cdot Z_A^{'}\cdot v + v\cdot Z_A^*\cdot v^*)}{4\omega P} \tag{22}$$

where * is the complex conjugate, " ′ " implies a perturbed parameter, V is the phase velocity, ω the angular frequency and P the power flow per unit width; Z_A and v are the surface acoustic impedance and particle velocity matrices, respectively.

Velocity change and attenuation change are then given by the real part and imaginary part, respectively, of the equations

$$\frac{\Delta V}{V} = \mathrm{Re}[\frac{jV(v^*\cdot Z_A^{'}\cdot v + v\cdot Z_A^*\cdot v^*)}{4\omega P}] \tag{23}$$

$$\frac{\Delta\alpha}{k} = \mathrm{Im}[\frac{jV(v^*\cdot Z_A^{'}\cdot v + v\cdot Z_A^*\cdot v^*)}{4\omega P}] \tag{24}$$

In the case of a guided SH-SAW, waveguiding by a suitable layer (oxide coating, polymer and polyimide) of appropriate thickness occurs when the shear wave velocity in the layer (V_M) is less than that in the substrate (V_{SH}). Assuming that a predominantly shear horizontal wave is coupled to the IDT, the fractional change in the wave velocity - a measure of the sensitivity to

surface mechanical perturbations - can be obtained from perturbation theory [11] as

$$\frac{\Delta V}{V} = -\frac{V_{SH}}{4}(\rho h)\left[1-(\frac{V_M}{V_{SH}})^2\right]\left|U_2\right|^2, \tag{25}$$

where V_{SH} is the unperturbed velocity of the SH-SAW, ρ is the layer mass density, h is the layer thickness and U_2, the normalized particle velocity displacement amplitude at the surface. It is noted that when the resulting wave suffers no dispersion, and the propagation path length, *L*, is equal to the sensing path length, l, equation (25) also describes the relative frequency shift, $\Delta f/f$, of the guided SH-SAW sensor. As a result, the device sensitivity to mass and viscoelastic loading s due to the thin guiding layer can be calculated [14-17]. Clearly, the device sensitivity increases as U_2 at the sensing surface increases, i.e., as the acoustic energy is trapped to the sensing surface. Note the effect of the layer viscoelasticity is implicitly represented by the term $[1-(V_M/V_{SH})^2]$ in equation (25). A similar, but somewhat more complicated, expression can be obtained for the change in the attenuation coefficient of the guided SH-SAW in terms of the complex shear modulus of the layer, which includes the shear viscosity coefficient of the layer. For the 36°YX-$LiTaO_3$, such an expression for the attenuation coefficient must also account for the effect of the shear vertical (SV) wave component, which is not negligible in evaluating the loss.

Figure 8 shows the frequency shift (sensitivity) of the $LiTaO_3$ guided SH-SAW as a function of poly (methyl methacrylate) (PMMA) and cyanoethylcellulose (CEC) layer thickness. It can be seen that the slope of the curve in Figure 8 reaches a maximum at a waveguide thickness of between 1600 and 2000 nm for the PMMA, indicating the near optimum thickness for achieving high mass sensitivity. For CEC, an optimum thickness of approximately 2400 nm was determined. An effective determination of the optimum thickness, however, requires looking at the device insertion loss in liquid environments as well as the frequency shift.

In another example, using the above derivation, the sensitivity to viscous liquid has been evaluated. Assuming Newtonian liquid of viscosity, η, and density, ρ, and assuming that a predominantly shear horizontal wave is coupled to the IDT, the fractional changes in wave velocity and attenuation with respect to pure water are found to be

$$\frac{\Delta V}{V} = -\frac{V_{SH}}{4}\left|U_2\right|^2\left[\left(\frac{\omega\eta\rho}{2}\right)^{1/2} - \left(\frac{\omega\eta_W\rho_W}{2}\right)^{1/2}\right] \tag{26}$$

$$\frac{\Delta \alpha}{\beta} = \frac{V_{SH}}{4}\left|U_2\right|^2\left[\left(\frac{\omega\eta\rho}{2}\right)^{1/2} + \left(\frac{\omega\eta_W\rho_W}{2}\right)^{1/2}\right] \tag{27}$$

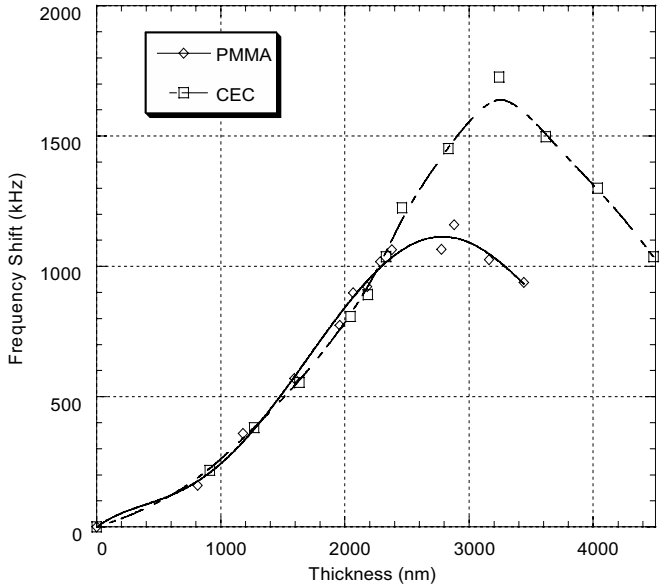


Figure 8. Guided SH-SAW Frequency Shifts (or Sensitivity) for Poly(methyl methacrylate) (PMMA) and Cyanoethylcellulose (CEC) Guiding Layers as a Function of Layer Thickness. (Ref. 42)

In the above equations, which indicate the sensitivity of SH-SAW to viscous liquid loading, η_w and ρ_w represent the viscosity and density of pure water, respectively. U_2 is the normalized particle velocity displacement amplitude at the surface. Clearly, as the particle velocity displacement amplitude at the surface increases due to guidance by a thin isotropic layer, the sensitivity to fluid viscosity increases. The above can easily be extended to a Maxwellian liquid case where the fluid exhibits a first order relaxation process [29, 47].

$$\eta(\omega) = \frac{\eta_0}{(1 + j\omega\tau)}, \tag{28}$$

where η_o is the low-frequency shear viscosity and τ is the relaxation time ($\tau = \eta_o / G_{hf}$, with G_{hf} the high-frequency rigidity modulus).

For SH-SAW on 36^0 rotated Y-cut $LiTaO_3$ which is not a pure mode axis, equations (23, 24) must be modified to include the shear vertical, SV, component of the excited wave. Such component will contribute, especially, to additional losses. However, for a metallized surface as it is the case here, the slightly leaky SH-SAW is converted to a pure SH-SAW.

3.2 Sensor Design Considerations

The first design principles and considerations are identical to those used for Rayleigh SAWs [1]. The design of the IDTs for the generation and detection of SH-SAW uses the delay line configuration often employed for SAW filters. In the present case, a dual delay line design is used with one line as a sensing line and the other as a reference line. Common environmental interactions elicit responses from both lines and are removed by subtraction. The chemical reaction is designed to take place only on the sensing line and a unique differential signal results. In addition, two possible platform geometries shown in Figure 9 with emphasis on the encapsulation for confinement of the liquid can be used for the lines. While the design of the sensor platform is usually made with few engineering considerations, analysis and/or prediction of the sensor response requires that the sensor effect be accounted for in the device response. The sensor effect can be easily incorporated into the device unperturbed transfer function, $T_{12}(f)$ to allow for the variations of delay time and attenuation. The sensor response can then be modeled by the following transfer function

$$T'_{12}(f) = T_{12}(f) e^{-i\frac{2\pi L}{\lambda}} e^{i\frac{2\pi\delta l}{\lambda}} e^{-\alpha l}, \tag{29}$$

where $\delta \equiv \Delta V/V$ is the fractional velocity change of the SH-SAW due to the guiding/sensing effect, $\lambda \equiv$ wavelength, $\alpha \equiv$ the attenuation coefficient introduced by the waveguide layer and chemically sensitive layer composite,

and L and l are the IDT center-to-center separation and sensing path length, respectively.

It is noted that when the resulting wave suffers no dispersion, and the sensing path length, l, is different from the propagation path length, L, equation (25) is appropriately modified to describe the relative frequency shift, $\Delta f/f$, of the guided SH-SAW sensor in terms of the velocity change as

$$\frac{\Delta f}{f} = \frac{1}{L} \cdot \frac{\Delta V}{V} \tag{30}$$

As a result, the device sensitivity to mass and viscoelastic loading s due to the thin guiding layer can be calculated [41, 44-46]. Writing $\Delta m = \rho h$, the device sensitivity to mass loading can be defined in terms of $\Delta V/V$ or $\Delta f/f$ as

$$S_m = \lim_{\Delta m \to 0} \left(\frac{\Delta f}{f} \right) \Big/ \Delta m \tag{31}$$

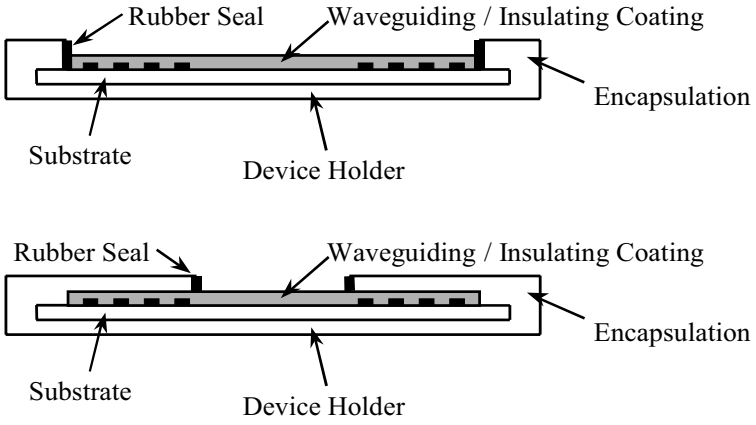


Figure 9. Two Examples of Guided SH-SAW Sensor Platform Geometries with Emphasis on the Encapsulation for Confinement of the Liquid.

The above assumes a purely elastic, lossless film. It is also noted that for a lossless film the term $e^{-\alpha l}$ in equation (29) is equal to one. However, for

viscoelastic film and/or chemically sensitive film loading, the device sensitivity must be recalculated to account for changes in viscoelasticity. In that case, the waveguiding layer and/or the chemically sensitive layer are characterized by a complex shear modulus, G' + jG", with G" = ωη. G' and G" are the shear storage modulus and loss modulus, respectively, η is the shear viscosity coefficient and $\omega = 2\pi f$ is the angular frequency. For chemical sensors applications, while mass loading is often assumed the dominant factor contributing to the frequency change, viscoelastic properties also contribute to Δf changes, and lead to changes in device attenuation. In the latter case, changes in the device loss can be used as a second measurand to characterize the sensor. For low-loss surface materials, and assuming $h << \lambda$, approximate analytical expressions can be obtained for both the velocity change and the attenuation. For the 36°YX-$LiTaO_3$, such an expression for the attenuation coefficient must also account for the effect of the shear vertical (SV) wave component, which is not negligible in evaluating the loss. However, for arbitrary viscoelastic film loading, effective computation of both parameters can only be achieved numerically. This decision is made depending on the value of the mechanical quality factor, Q, of the load. It is noted the inverse of Q is defined as the loss tangent, $\tan\theta \equiv 1/Q = G'' / G'$, and if $\tan\theta > 0.01$, then the load material may be treated as a high-loss medium [48]. As a result, for chemical sensors in which a chemically sensitive and lossy polymer is deposited on the guiding layer, the value of the mechanical quality factor, Q, will dictate the approach to use in evaluating the device attenuation in terms of the layer thickness and viscosity. For biosensing applications in which a second layer (the layer of attached receptors) are applied on the waveguide layer, equation (31) is sufficient to approximate the mass sensitivity of the device [15-18]. This is because the bio-layer consisting of the receptors and the bound antigens is of the order of a few molecules, and has negligible viscoelastic contribution to the sensor response.

3.3 Examples of Bio-chemical Detection Using Guided SH-SAW Sensors

In this section, the results from devices using the PMMA guiding layers are shown and discussed in biosensing applications. The guided SH-SAW devices used are on 36° YX-$LiTaO_3$ substrates with a polymer coating as the waveguide layer. The device is designed and fabricated with 100/800 nm thick Cr/Au interdigital transducers (IDTs) having a periodicity of 40μm. This

corresponds to an operating frequency of approximately 103 MHz for the uncoated devices [42]. In order to evaluate the mass sensitivity of the guided SH-SAW device, biosensing experiments are conducted with goat immunoglobulin G (IgG) adsorbed on one delay line of the dual delay line device while the other delay line is blocked with bovine serum albumin (BSA). The device is then exposed to rabbit anti-goat IgG in the liquid environment. In Figure 10, the responses (difference signal) of devices with different PMMA coating thicknesses to injection of 24 μg/ml rabbit anti-goat IgG are shown. It is seen that mass sensitivity increases with waveguide thickness, while differences in baseline noise are small. This is because, as the film thickness increases, the acoustic energy is trapped more to the sensing surface, drastically increasing the device sensitivity. The frequency shift produced in a sensor with 1.95 μm of PMMA is more than four times greater than that for a sensor with no PMMA waveguide. The responses are not corrected for the influence of temperature, and no effort is made to control laboratory or device temperature since differential signal response is measured for the dual line.

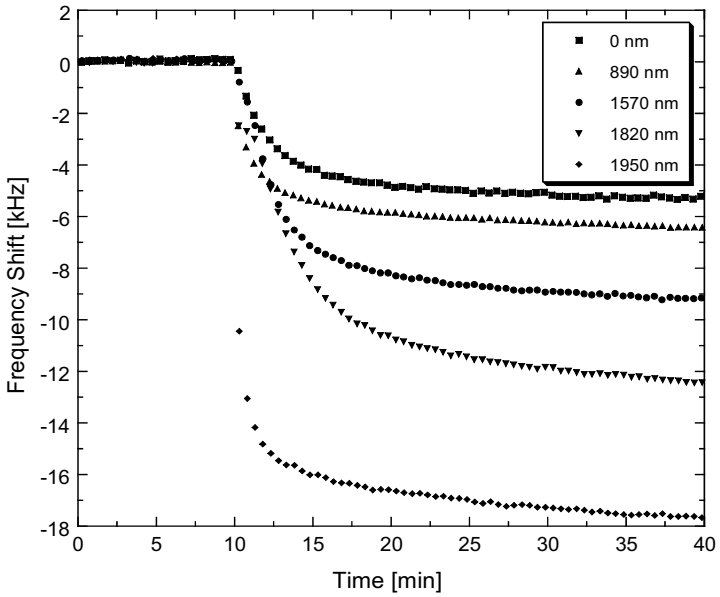


Figure 10. Guided SH-SAW Frequency Shifts (or Sensitivity) for Poly(methylmethacrylate) (PMMA) and Cyanoethylcellulose (CEC) Guiding Layers as a Function of Layer Thickness on 36° YX-$LiTaO_3$ Substrates. (Ref. 42).

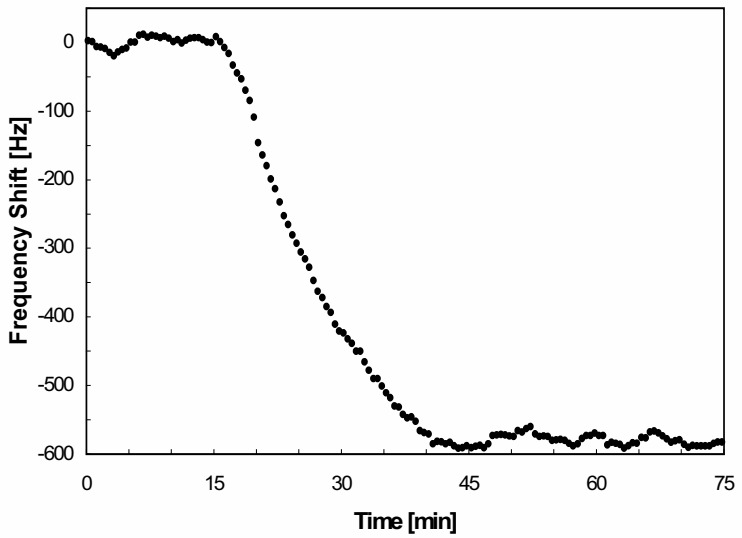


Figure 11. Response of a Guided SH-SAW Device with 1.57 μm Thick PMMA Coating to 100 ng/ml Rabbit Anti-goat IgG, Injected After 15 min. The Difference Signal (Sensing Minus Reference Lines) is Shown (Ref. 42).

To further test the mass sensitivity of the devices, the detection of 100 ng/ml rabbit anti-goat IgG by a 1.57 μm thick waveguide device (note this PMMA layer thickness is non-optimal) is shown in Figure 11. Measured noise levels of 32 Hz (peak-to-peak) and 8 Hz (rms) are obtained. These results indicate that the most sensitive device can detect 68 pg/mm^2 or 17 pg/mm^2, respectively, the latter corresponding to 0.1% of a closely-packed monolayer of antibodies.

3.4 Examples of Chemical Detection in Liquids Using SH-SAW Sensors

Guided SH-SAW devices with an operating frequency of approximately 103 MHz for the uncoated devices and similar to those described in the previous section are used. A dual delay line configuration is used and consists of a reference line coated with polymethyl methacrylate (PMMA) which acts as a guiding layer only, and a sensing line coated with poly(isobutylene)

(PIB), polyepichlorhydrin (PECH) or poly(ethyl acrylate) (PEA) which acts as both a guiding layer and a chemically sensitive layer. Typical sensor response of a device to varying concentrations of xylene (10-60 ppm) in DI water is shown in Figure 12. On that device, the sensing line is coated with 0.8-μm-thick PIB cured at 40°C for 15 min while the same thickness of PMMA, cured at 180°C for 2h is used for the reference line. The result shows that for direct detection of xylene in water, the sensor exhibits excellent reversibility when the chemical analyte is removed by subjecting the device to DI water. In Figure 12, response time of 10 min is observed. However, the observed time doesn't represent the sensor response time due to the use of the flow mixing system. A typical response time of the sensor is observed to be less than 20 seconds as chemical analyte is directly placed in contact with the device.

The results of Figure 12 are not sufficient to determine the major contribution to the sensor response. These contributions can be due to mass loading and changes in viscoelastic properties of the polymer. Further insight into the various contributions to the sensor response can be obtained by simultaneously measuring the device loss as a function of analyte concentration. Figure 13 shows insertion loss in the xylene detection using the guided SH-SAW device with 0.8μm thick PIB on the sensing line and 0.8μm PMMA on the reference line. Clearly, the observed loss can be explained by changes in the viscoelastic properties of the load, i.e. chemically sensitive layer. As a result, both Figures 12 and 13 clearly indicate that the two contributions to the sensor response are from the added analyte mass and subsequent changes in viscoelastic properties.

Figure 14 shows the frequency shift (sensitivity) of the $LiTaO_3$ guided SH-SAW sensor device in the xylene detection with different coating thicknesses The reference line is coated with 0.5-, 0.8- or 1.0- μm -thick PMMA cured at 180°C for 2h while the chemical sensing line is coated with the same thickness of PIB, cured at 40°C for 15 min. It is seen that increasing the polymer thickness increases the sensitivity. In part, this is due to increased mass-loading in the thicker films. However, it is also expected that contributions from change in viscoelasticity will increase, especially the loss contribution. This is because the increase analyte concentration in the coating results in increase expansion in the polymer volume.

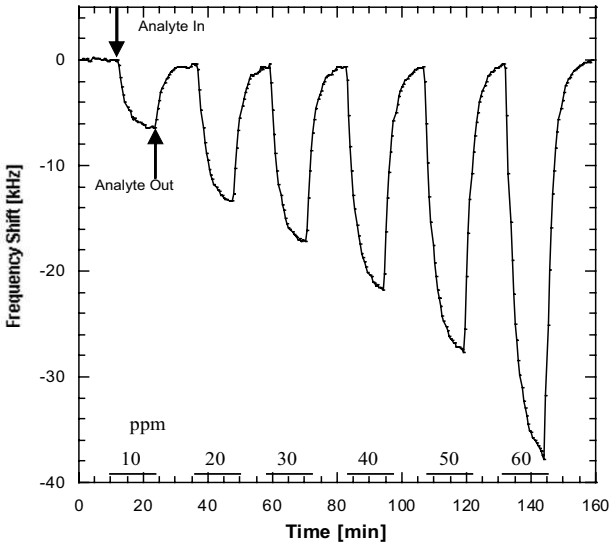


Figure 12. Detection of 10 to 60 ppm of Xylene Using a Guided SH-SAW Device with 0.8µm Thick PIB on the Sensing Line and 0.8µm Thick PMMA on the Reference Line. The PIB Layer is Uncured.

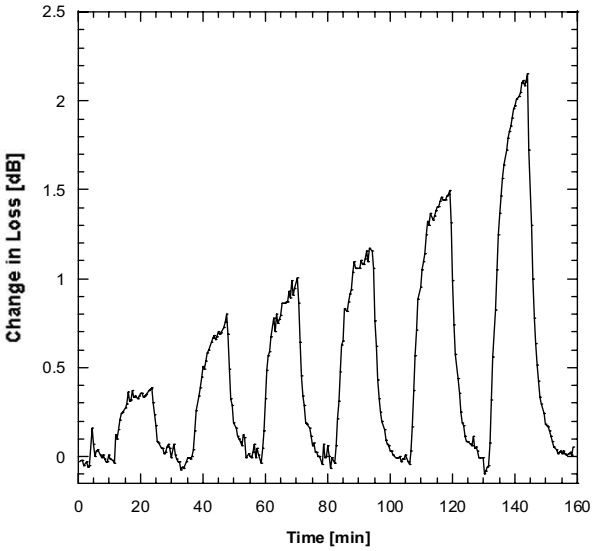


Figure 13. Change in Loss in the Detection of 10 to 60 ppm Xylene Using the Guided SH-SAW Device with 0.8µm Thick PIB on the Sensing Line and 0.8µm PMMA on the Reference Line.

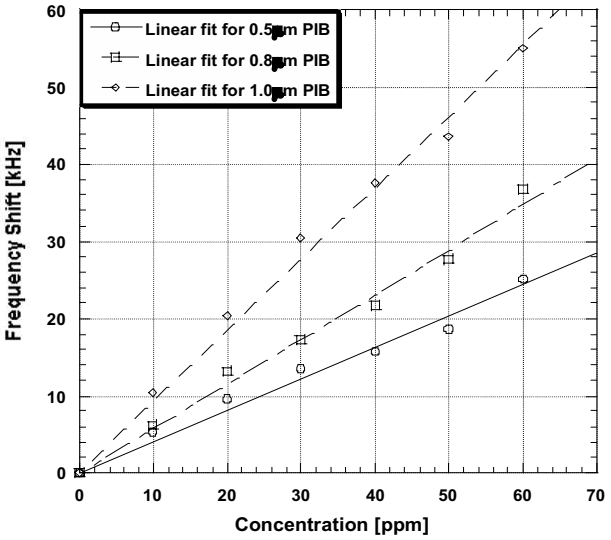


Figure 14. Comparison of Sensitivity of PIB (0.5μm, 0.8μm and 1.0μm) –Coated Guided SH-SAW Sensor Platform in the Detection of Xylene in Water (PIB is Cured at 40°C for 15 min).

4. SUMMARY

Thickness-shear mode (TSM) resonators and guided shear-horizontal surface acoustic wave (guided SH-SAW) devices are described as two resonant piezoelectric devices that can be used effectively as physical and bio-chemical sensors in both gas and liquid environments. The measured sensor parameters for these devices include resonant frequency shifts and/or attenuation (device loss) changes, which occur during the probing action or detection. Other characteristics include sensitivity, resolution, limit of detection and reproducibility, and are well documented in the published literature. It is noted here that increasing frequency of operation of these devices results in increased resolution, but will result in increased sensitivity (not proportionally) only if the sensor signal noise can also be decreased. The latter could be achieved depending on the driving circuitry and its performance. It is also noted that the sensor signal noise level can be decreased through the use of "intelligent" signal processing. The limit of detection (LOD) of these devices (as of any other sensor) depends on both the sensitivity and the signal stability, i.e. signal noise and drift.

Specifically, the TSM resonator is an effective probe for rapid characterization of the complex shear modulus ($G = G' + G''$) of thin polymer films, provided that both the film density and thickness can be independently and accurately determined. The theory used is simple and, near crystal resonance, is as good as other rigorous acoustic wave propagation theories. Near mechanical resonance, a modified equivalent circuit (the so-called BVD model) description is used. This simplified theory of the quartz resonator is well-known and extremely precise for all sensor and material characterization studies. For viscoelastic materials, at relatively high thickness (2-4 μm depending on the polymer), accurate shear modulus parameters (approaching that of the bulk material) can be extracted where viscoelastic contribution is substantial. As the operating frequency of the device increases, the frequency dependence of the moduli must be accounted for.

Of all acoustic wave devices, guided SH-SAW sensors (also known as Love wave sensor) appear most promising for bio-chemical detection in liquid environments, with the possibility of tailoring the device sensitivity. Optimized sensor implementation requires investigating various bio-chemically sensitive layers, guiding coating types and thicknesses, curing methodologies for high sensitivity and stability. Chemical sensor platform has been demonstrated using PMMA polymer guided SH-SAW on 36°-YX $LiTaO_3$, with limit of detection in the low ppb range for some chemical analytes in aqueous solutions. Similarly, biosensing experiments using identical devices indicate that the device has good sensitivity and can detect 17 pg/mm^2, corresponding to 0.1% of a closely-packed monolayer of antibodies. It is noted that these devices have yet to be optimized.

REFERENCES

[1] Ballantine, D. S., White, R. M., Martin, S. J., Ricco, A.J., Zellers, E. T., Frye, G. C., and Wohltjen, H., *Acoustic Wave Sensors, Theory, Design, and Physico-Chemical Applications*. San Diego: Academic Press, 1997.

[2]. Janata, J. "Mass Sensors", In *Principles of Chemical Sensors*. New York: Plenum Press, Kluwer Academic, 1990.

[3] Thompson, M., and Stone, D. C., *Surface-Launched Acoustic Wave Sensors, Chemical Sensing and Thin-Film Characterization*. New York: Wiley & Sons, 1997.

[4] Mandelis, A., and Chrisofides, C., *Physics, Chemistry, and Technology of Solid State Gas Sensor Devices*. New York: Wiley, 1993.

[5] Bastiaans, G. J. "Piezoelectric Biosensors." In *Chemical Sensor Technology,* S. Yamauchi, ed., Amsterdam: Elsevier, 1992.

[6] Hierlemann, A., and Baltes H. "Mass-Sensitive Sensors." In *Semiconductor-Based Chemical Sensors, MEMS Handbook*, J. Korvink, O. Paul, William Andrew Publishing, ed. New Jersey: Noyes Publications, 2003.
[7] Zelenka, J., *Piezoelectric Resonators and Their Applications*, Amsterdam: Elsevier, 1986.
[8] Auld, J., *Acoustic Fields and Waves in Solids, Vols. I and II*. Florida: Krieger Publishing, 1990.
[9] Crooks R. M., Ricco A. J. New Organic Materials Suitable for Use in Chemical Sensor Arrays. Accounts of Chemical Research, Special Issue on Chemical Sensors 1998; 31 (5): 219-227.
[10]Ricco A. J., Crooks R. M., Osbourn G. C. Surface Acoustic Wav Chemical Sensor Arrays: New Sensitive Interfaces Combined with Novel Cluster Analysis to Detect Volatile Organic Compounds and Mixtures. Accounts of Chemical Research, Special Issue on Chemical Sensors 1998; 31 (5):289-296.
[11]Grate J. W., Martin S. J., White R. M. Acoustic Wave Microsensors. Anal. Chem.1993; 65(22), [Part I, 65(21):940A-948A; Part II, 65(22):987A-996A].
[12]Drafts B. Acoustic Wave Technology Sensors. IEEE Trans. Microwave Theory and Techniques 2001; 49(4):795-802.
[13]Grate J. W., Abraham M. H. Solubility Interactions and the Design of Chemically Selective Sorbent Coatings for Chemical sensors and Arrays. Sensors and Actuators B 1991; 3:85-111.
[14]Hierlemann A., Zellers E. T., Ricco A. J. Use of Linear Solvation Energy Relationships for Modeling Responses from Polymer-Coated Acoustic-Wave Vapor Sensors. Anal. Chem. 2001; 73(14):3458-3466.
[15]Zellers E. T., Batterman S. A., Han M., Patrash S. J. Optimal Coating Selection for the Analysis of Organic Vapor Mixtures with Polymer-Coated Surface Acoustic Wave Sensor Arrays. Anal. Chem.1995; 67(6):1092-1106.
[16]Andle J. C., Vetelino J. F., Lade M. W., McAllister D. J., An Acoustic Plate Mode Biosensor. Sensors and Actuators B 1992; 8:191-198.
[17]Andle J. C., Weaver J. T., McAllister D. J., Josse F., Vetelino J. F. An Improved Acoustic Plate Mode Biosensor. Sensors and Actuators B 1993; 13:437-442.
[18] Gizeli E., Goddard N. J., Stevenson A. C., Lowe C. R.. A Love Plate Biosensor Utilising a Polymer Layer. Sensors and Actuators B 1992; 6:131–137. Also IEEE Trans. Ultrasonics, Ferroelectr. Freq. Control 1992; 39:657-659.
[19]Gizeli E., Stevenson A. C., Goddard N. J., Lowe C. R. Acoustic Love Plate Sensors: Comparison with Other Acoustic Devices Utilizing Surface SH Waves. Sensors and Actuators B 1993; 13-14:638-639.
[20]Baer R. L., Flory C. A., Tom-Moy M., Solomon D. S., STW Chemical Sensors, Proc. IEEE Ultrason. Symp.; 1992.
[21]Dahint R., Grunze M., Josse F., and Renken J. Acoustic Plate Mode Sensor for Immunochemical Reactions. *Anal. Chem.* 1994; 66:2888–2892.
[22]Kondoh J., Shiokawa S. SH-SAW Devices as Effective Identification System for Liquids. Proc. IEEE Ultrason. Symp. 1994.
[23]Rapp M., Wessa T., Ache H. J. Modification of Commercially Available Low-loss SAW Devices Towards an Immunosensor for In-situ Measurements in Water. Proc. IEEE Ultrason. Symp.; 1995.
[24]Welsch W., Klein C., von Schickfus, M., Hunklinger S. Development of a Surface Acoustic Wave Immunosensor. Anal. Chem. 1996; 68:2000–2004.
[25]Josse F., Zhou R., Patel R., Zinszer K., Cernosek R. W. Real-time Detection of Chemical Compounds in Liquid Environments Using Polymer Coated Thickness Shear Mode Quartz

Resonators. Anal. Chem. 2000; 72:4888-4898.
[26]Sauerbrey G. The use of oscillators for weighing thin layers and for micro-weighing, Z. Physik 1959; 155: 206–212
[27]Benes E. Improved quartz crystal microbalance technique. J. Appl. Phys. 1984; 56: 608
[28]Martin SJ, Frye GC. Polymer film characterization using quartz resonators. Proc. IEEE Ultrason. Symp.;1991.
[29]Bandey H. L., Martin, S. J., Cernosek, R. W. Modeling the responses of thickness-shear mode resonators under various loading conditions. Anal. Chem. 1999; 71:2205-2214
[30]Wolff O., Seydel E., and Johannsmann, D., Viscoelastic properties of thin films studied with quartz crystal resonators. Faraday Discuss Chem. Soc.1997; 107:91-104
[31]Domack A., Prucker O., Ruhe J., Johannsmann D. Phys. Rev. E 1997; 56:680-689
[32]Martin S. J., Frye G. C., Ricco A. J., Senturia S. D. Effect of surface roughness on the response of thickness-shear mode resonators in liquid. Anal. Chem. 1993; 65:2910-2922
[33]Martin S. J., Frye G. C., Senturia S. D. Dynamics and Response of Polymer-Coated Surface Acoustic Wave Devices: Effects of Viscoelastic Properties and Film Resonance. Anal. Chem. 1994; 66(14): 2201-2219
[34]Behling C., Lucklum R., Hauptmann P. The non-gravimetric quartz crystal resonator response and its application for determination of polymer shear modulus. Meas. Sci. Technol. 1998; 9:1886-1893
[35]Morray B, Li S, Hossenlopp J, Cernosek R W, Josse F. PMMA polymer characterization using thickness-shear mode (TSM) quartz resonator. Proc. IEEE Freq. Control Symp. and PDA Exhibition, 2002.
[36]Lee D. L., Analysis of energy trapping effects for SH-type waves on rotated Y-cut quartz. IEEE Trans. Son. Ultrason. 1981; SU-28:330–341
[37]Josse F., Lee D.L. Analysis of excitation, interaction and detection of surface and bulk acoustic waves on piezoelectric substrates. IEEE Trans. Son. and Ultrason.1982; SU-29:262-273.
[38]Harding G. L., Du J., Dencher P. R., Barnett D., Howe E., Love wave acoustic immunosensor operating in liquid. Sensors and Actuators 1997; A61:279-286.
[39]Bender F., Cernosek R.W., Josse F. Love-wave biosensors using cross-linked polymer wave guides on $LiTaO_3$ substrates. Electronics Letters 2000; 36 (19):1672-1673
[40]Leidl A., Filser H., Labatzki A., Passivation of the interdigital transducers of acoustic sensors based on surface localized HPSWs for measurements in liquids. Sensors and Actuators A 1995; 47:353–356
[41]Du J., Harding G. L. A multilayer structure for Love-mode acoustic sensors. Sensors and Actuators A 1998; 65:152–159
[42]Josse F., Bender F., Cernosek R.W. Guided shear horizontal surface acoustic wave sensors for chemical and biochemical detection in liquid. Anal. Chem. 2001; 73:5937-5944
[43]Auld, B. A., *Acoustic Fields and Waves in Solids*. Vol. II, Florida: Krieger, 1990.
[44]Andle JC, Vetelino JF, Josse F. A theoretical study of acoustic plate modes as biosensing elements, Proc. IEEE Ultrason. Symp.;1991.
[45]Kovacs G., Venema A. Theoretical comparison of sensitivities of acoustic shear wave modes for (bio)chemical sensing in liquids. Appl. Phys. Lett.1992; 61:639–641
[46]Wang Z., Cheeke J. D. N., Jen C. K. Sensitivity analysis for Love mode acoustic gravimetric sensors. Appl. Phys. Lett. 1994; 64:2940–2942
[47]White, F. M., *Viscous Fluid Flow*. New York: McGraw-Hill, 1991.
[48]Kielczynski P., Attenuation of Love waves in low loss media. J. Appl. Phys.1997; 82:5932-5937.

Chapter 4

ACOUSTIC-WAVE PIEZOELECTRIC AND PYROELECTRIC SENSORS BASED ON PZT THICK FILMS

Vittorio Ferrari
Dip. Elettronica per l'Automazione and INFM
University of Brescia, Italy

Abstract: The operating principles, design and manufacturing in thick-film technology (TFT) of sensors based on the piezoelectric and pyroelectric properties of lead zirconate titanate (PZT) thick films are described. In particular, the attention is focused on acoustic-wave piezoelectric sensors for physical and chemical quantities, and on pyroelectric sensors and arrays for light spot position measurement.

Keywords: piezoelectric sensor, pyroelectric sensor, acoustic-wave sensor, sensor array, thick-film technology, lead zirconate titanate (PZT).

1. INTRODUCTION

Solid-state sensors and sensor systems are constantly expanding into new application areas, currently representing one of the most topical and rapidly evolving fields of science and technology. It is a field where a host of different requirements unavoidably coexist, such as demands in terms of performances, operating conditions, employed materials, integration with the electronics, size, power consumption and cost.

Nowadays, there is not one single technology capable of addressing all the above requirements, and the situation is not likely to change in the near future. Though silicon micromachining and MEMS (Micro-Electro-Mechanical System) technology are offering a real breakthrough for sensor fabrication, there are technical and economical issues that make their diffusion in real-world applications slower than expected or desired. Therefore, in many cases other solutions are required, and a number of different sensor technologies are currently in use.

The thick-film technology (TFT) is one of the most reliable, flexible and widespread technologies for sensor and sensor system fabrication that can count on the possibilities offered by commercially-available screen-

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 125-154.

printable pastes, such as piezoresistive, thermoresistive, conductive and dielectric pastes. In addition, however, the TFT is also very attractive when combined with less conventional materials and innovative sensor designs. In this outline, this chapter illustrates some examples of sensors based on the piezoelectric and pyroelectric properties of lead zirconate titanate (PZT) thick films.

2. THICK-FILM TECHNOLOGY

Historically, the thick-film technology (TFT) started some decades ago as a technology for electronics with the manufacturing of hybrid circuits. Among the key factors for ongoing success of TFT in such fields as the automotive, military, and telecommunications there are a number of distinctive features, like capability of operation in harsh environments, reliability, integration and miniaturization, functional trimming, and low design cost and capital investment.

The TFT consists of a series of processes whereby, starting from properly formulated pastes (or inks), the deposition on insulated substrates of layers having different electric and functional properties is obtained. In the standard TFT, the process steps involved are basically three, namely the screen printing of the paste, the drying of the wet film, and its subsequent firing at high temperature [1].

All the processes are carried out in air or sometimes in controlled atmosphere, but generally not in a vacuum. This contributes to the cheapness of TFT and represents the most important difference with other film deposition techniques, such as the thin-film technology, which almost invariantly require vacuum. The typical film thickness obtainable with TFT is in the range from microns to several hundreds microns, which is higher than that achievable by most other deposition methods.

Thick-film pastes are colloidal systems that typically include a functional part in the form of powder, a binder, and a vehicle. The functional part defines the final electrical properties of the film. The composition depends on use, i.e. conductor, resistor, insulator, and the average particle size in the micron range. The binder promotes vitrification during firing, and it is generally composed by glass frits or oxides. The vehicle gives the best rehology for printing and it is usually an organic solvent with possible additives. Also available and widely used are polymeric thick-film (PTF) pastes that do not require firing but stabilize their properties with low-temperature curing.

The substrate materials can be ceramics, insulated metals or plastics. Among the ceramics, the most typically used are alumina (Al_2O_3), aluminum nitride (AlN), and beryllia (BeO) in order of rising thermal conductivity, and also zirconia (Y_2O_3). Insulated metals are most

commonly steel, aluminum, or copper. Plastics (e.g. polyester) are used as flexible substrates and/or lightweight low-cost substrates coupled to polymer thick films. In the last years there is an increasing interest in using silicon as a substrate for TFT, toward the development of MEMS based on screen-printing on silicon.

The screen-printing process, as shown in Figure 1a, consists of the transfer of the paste onto the substrate by extrusion through a patterned screen mesh, by means of the pressure exerted by a squeegee. In place of the screen mesh a stencil can be used. The patterning is obtained by an emulsion mask applied onto the screen, as illustrated in Figure 1b, into which the desired shape has been photo-reproduced by light exposure and development starting from a computer-generated artwork.

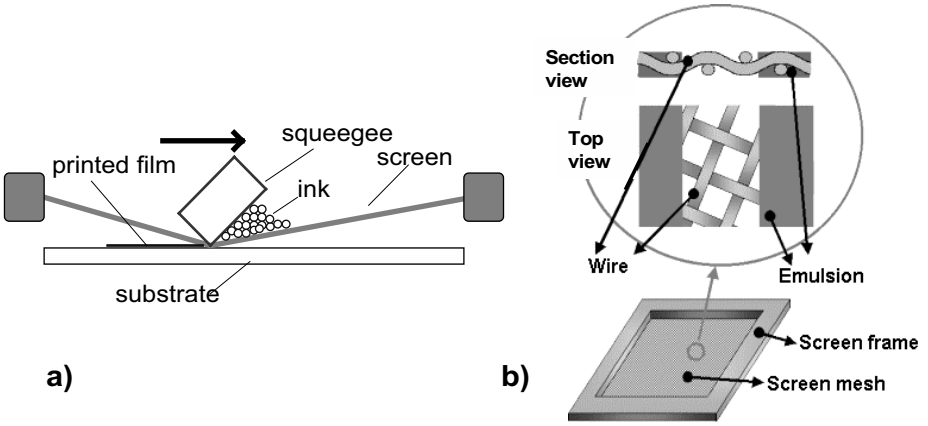


Figure 1. Screen-printing Process (a); Screen Mesh (b) (adapted form www.noritake-elec.com).

The film thickness and print resolution depend, among other factors, on the characteristics of the screen mesh. In particular, the mesh count, which defines the number of openings per linear inch (typical values are 250, 325, or 400), and the wire diameter jointly determine the open fraction of the screen area.

The standard achievable resolution is between 0.05 and 0.1 mm for line width and pitch. Better figures are obtainable with photoimaging and etching techniques. A screen-printing machine is shown in the picture of Figure 2a.

The wet film obtained after printing is allowed to level for a settling period of 10-30 min. Then the substrate is heated in an oven at 100-150°C for 15-30 min for the drying process, during which volatile compounds evaporate and the film becomes more mechanically stable.

During the following firing process, the ceramic or metallic substrate is put into a belt furnace, of the type shown in Figure 2b, for a thermal cycle that typically reaches temperatures in the order of 800-1000°C. At this step, the organic components still present in the film are burned out, the film ancors to the substrate, and the functional properties develop and stabilize to their final level.

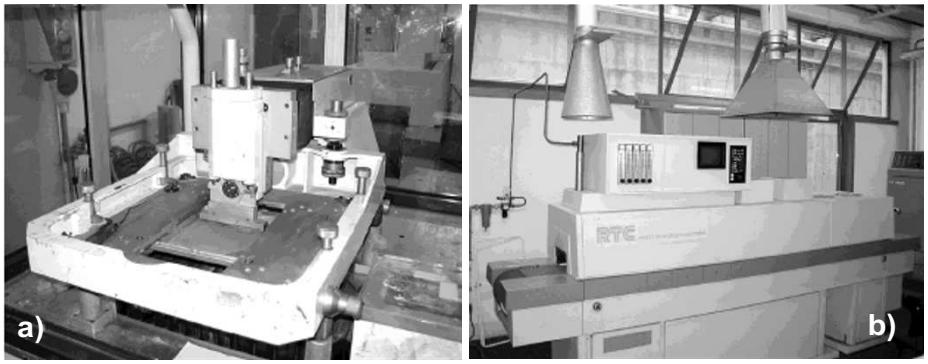


Figure 2. Screen-printing Machine (a) – Infra-red Belt Furnace (b).

An important factor in the firing process is the temperature profile, which should be optimized for the given film composition. The firing atmosphere is open air most of the times, or, occasionally, an inhert atmosphere.

With plastic substrates, that cannot withstand the high temperature involved in the standard firing process, appropriate pastes are used that can be cured at lower temperatures typically by infrared heating.

Besides its primary role as a technology for electronics, TFT is also an enabling and attractive technology for sensors [2, 3].

The main advantages offered by TFT for sensor manufacturing are a high degree of flexibility given by the choice of paste compositions and substrate materials and shapes, reliable operation in harsh environments, low cost at small production volumes ($<10^4$/yr), capability of integration with electronics in hybrid systems, miniaturization at the mesoscale level.

TFT also represents a flexible platform and a supporting technology for sensor systems. Hybrid units of significant functional complexity can be cost-effectively developed which, for instance, combine TFT substrate, passives and interconnections, silicon MEMS, and microelectronic circuits in the form of single chip or multi-chip modules (MCM).

In addition to the standard TFT, a number of technology "variants" have been developed that are attractive for mesosystems and, such as the Green Tape and the Low-Temperature Co-fired Ceramic (LTCC) technologies [4].

3. PIEZOELECTRICITY AND PYROELECTRICITY

The piezoelectric effect, discovered in 1880-81 by the Curie brothers, is the property of certain materials in which an imposed mechanical stress induces an electric charge (direct effect), and an applied electric field develops a mechanical strain (converse effect) [5, 6].

The piezoelectric effect is linear, in the sense that it conserves the sign between the forcing and the generated quantities as schematically shown in Figure 3. Therefore, by inverting the former an opposite effect is produced in the latter. This is different from electrostriction, a property shown by all materials to various degree, where the relationship between produced strain and applied electric field is quadratic.

Piezoelectricity is exploited in many applications that can be coarsely grouped in two categories: sonic and ultrasonic wave generation and detection, and electromechanical or mechanoelectrical conversion.

The former comprises sensors and transducers, filters and resonators. The latter includes actuators, transformers, gas igniters, motors, machining tools, nebulizers and humidifiers, and more.

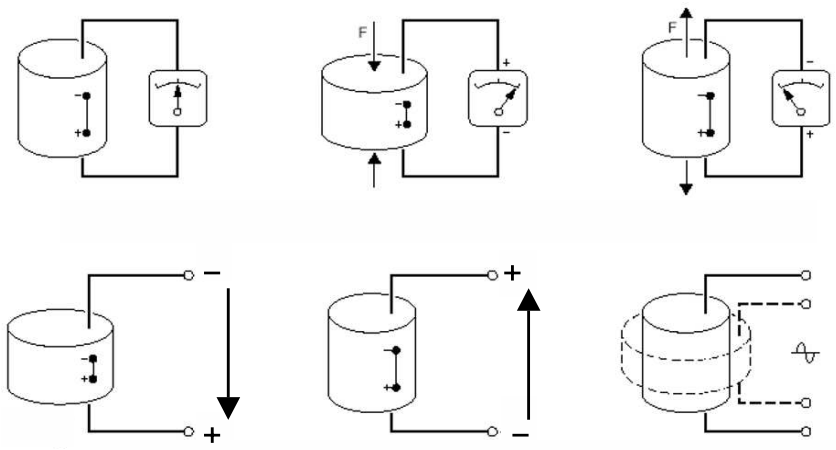


Figure 3. Piezoelectric Effect (adapted from Philips-Morgan Electroceramics).

The pyroelectric effect, which was named in 1824 by Brewster but known as a phenomenon since before, is the property of certain materials that have a spontaneous electrical polarization which is a function of temperature [7], as shown in Figure 4. The polarization is along a unique

axis that is called the polar axis. Static polarization induces charges that are neutralized by losses due to finite material resistivity, therefore only temperature variations can be detected.

Pyroelectricity is exploited in thermal radiation detectors [8], i.e. those in which the radiation produces a temperature change that is converted into an electric signal, as opposed to photon detectors, like photodiodes, where the radiant to electrical conversion is direct. Pyroelectric detectors generally have wide spectral response, high speed, and require no cooling to improve the signal-to-noise ratio. Typical applications of pyroelectric sensors are in occupancy detection systems, burglar alarms, contactless temperature measurement (pyrometers), infra-red cameras and vidicons, gas analyzers, laser power meters, and more.

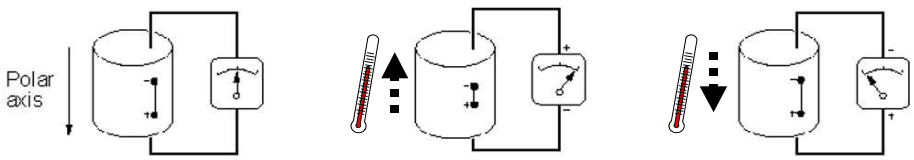


Figure 4. Pyroelectric Effect.

The degree of structural symmetry of a given material directly affects its piezoelectric and pyroelectric properties, as shown in Figure 5. Out of the 32 crystal classes (point groups) into which any one crystal can be classified, 11 are centrosymmetric, therefore they posses no polar properties and are not piezoelectric nor pyroelectric. The remaining 21 do not have a center of symmetry, and 20 of them are piezoelectric. Out of such 20 classes, 10 have a unique polar axis of spontaneous polarization, therefore they are pyroelectric. Ferroelectrics are a subclass of pyroelectrics whose polarization is reversible under an applied electric field.

The materials belonging to the class of ferroelectrics exhibit the strongest piezoelectric and pyroelectric effects. Ferroelectrics can be either monocrystals, e.g. Rochelle salt, or polycrystals, e.g. barium titanate or lead zirconate titanate (PZT) ceramics [9, 10].

In ferroelectrics, the electrical dipoles are arranged in domains. Above a certain temperature dependent on the particular material, called the Curie temperature T_C, the domains are randomly oriented so they average out and produce a zero net polarization.

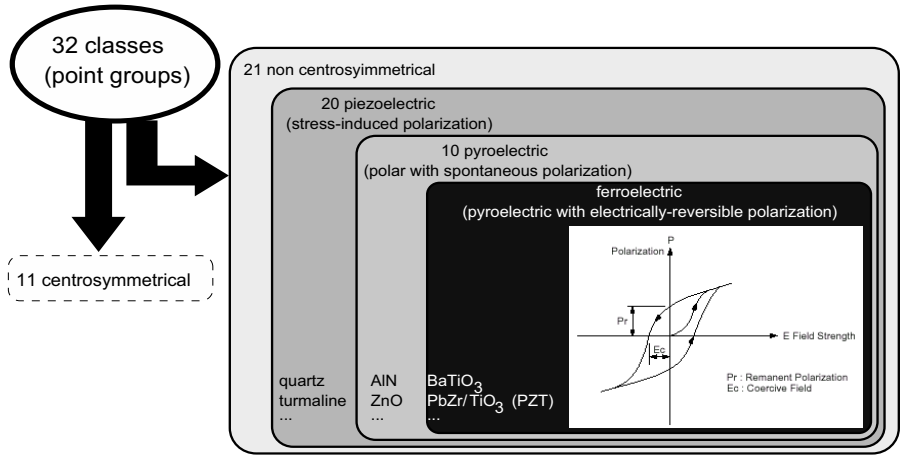


Figure 5. Point Groups and Material Properties.

A treatment called poling is applied to set up a permanent polarization in ferroelectrics, as illustrated in Figure 6. An electric field is applied usually at high temperature to orient the domains, and maintained while cooling back to room temperature. At this point, the material maintains a remnant polarization along the poling axis and exhibits both the piezoelectric and pyroelectric effects. Poled ferroelectrics must operate at a temperature well below their T_C, otherwise thermal depoling occurs.

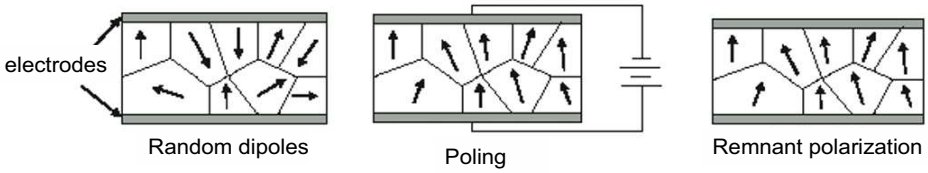


Figure 6. Poling of Ferroelectrics.

Piezoelectricity involves a mutual interaction between the electrical and mechanical domains that can be described by linear piezoelectric constitutive equations [5, 6].

There are several options for constitutive equations, according to which of the electrical and mechanical quantities are chosen as the independent and dependent variables. In the monodimensional simplified hypothesis, the following four formulations are possible:

$$\begin{cases} S = s^E T + dE \\ D = dT + \varepsilon^T E \end{cases} \quad (1.a) \qquad \begin{cases} T = c^E S - eE \\ D = eS + \varepsilon^S E \end{cases} \quad (1.c)$$

$$\begin{cases} S = s^D T + gD \\ E = -gT + D/\varepsilon^T \end{cases} \quad (1.b) \qquad \begin{cases} T = c^D S - hD \\ E = -hS + D/\varepsilon^S \end{cases} \quad (1.d)$$

where:
S = strain [m/m];
T = stress [N/ m^2];
D = dielectric displacement [C/m^2];
E = electric field [V/m]
s^X = elastic compliance [m^2/N] at constant X
$c^X = 1/s^X$ = elastic stiffness [N/ m^2] at constant X
ε^X = dielectric permittivity [F/m] at constant X
d, e, g, h = piezoelectric constants

The piezoelectric constants, the elastic stiffness and compliance, and the dielectric permittivity are linked by the following relationships:

$$d = \varepsilon^T g = s^E e = [\text{C/N}] = [\text{m/V}]$$
$$e = \varepsilon^S h = c^E d = [\text{C/m2}] = [\text{N/mV}]$$
$$g = 1/\varepsilon^T \; d = s^D h = [\text{m2/C}] = [\text{Vm/N}]$$
$$h = 1/\varepsilon^S \; e = c^D g = [\text{N/C}] = [\text{V/m}]$$
$$\varepsilon^T = \varepsilon^S + de \;\; ; \;\; c^D = c^E + he$$

The energy transfer between the electrical (E) and mechanical (M) domains is expressed by the electromechanical coupling coefficient k:

$$k^2 = \frac{(\text{stored energy})_M}{(\text{supplied energy})_E} = \frac{(\text{stored energy})_E}{(\text{supplied energy})_M} = \frac{d^2}{s^E \varepsilon^T} \quad (2)$$

In the more general case of a tridimensional nonisotropic medium the tensorial notation becomes necessary for generalizing the Eqs.(1) [11]. For example, under the usual assumption of summation over repeated indexes, Eq.(1.a) becomes:

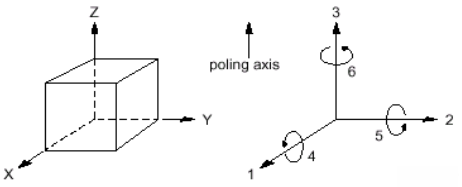
$$\begin{cases} S_{ij} = s^{E}_{ijkl} T_{kl} + d_{kij} E_k \\ D_i = d_{ikl} T_{kl} + \varepsilon^{T}_{ik} E_k \end{cases} \qquad (3)$$

S_{ij} = 2nd-rank strain tensor
T_{kl} = 2nd-rank stress tensor
s^{E}_{ijkl} = 4th-rank compliance tensor
ε^{T}_{ik} = 2nd-rank permittivity tensor
d_{ikl} = 3rd-rank piezoelectric tensor

Equations analogous to Eq.(3) can be written for the other three formulations as a generalization of Eqs.(1.b), (1.c), and (1.d).

Thanks to the symmetry of strain and stress tensors the reduced index notation illustrated in Figure 7 can be used, and Eq.(3) becomes:

$$\begin{cases} S_I = s^{E}_{IJ} T_J + d_{kI} E_k \\ D_i = d_{iJ} T_J + \varepsilon^{T}_{ik} E_k \end{cases} \qquad (4)$$



Index ij	*Reduced Index I*	*Matrix Representation*
11	1	
22	2	
33	3	$\begin{pmatrix} T_1 & T_6 & T_5 \\ T_6 & T_2 & T_4 \\ T_5 & T_4 & T_3 \end{pmatrix}$
23 or 32	4	
13 or 31	5	
12 or 21	6	

Figure 7. Reduced Index Notation.

Pyroelectricity implies that the polarization P of a material is a function of temperature T. In ferroelectrics, P includes both a spontaneous polarization P_S and a field-induced polarization P_{ind}. Limiting to scalar notations, the dielectric displacement D is given by:

$$D = \varepsilon_o E + P_S + P_{ind} = P_S + \varepsilon E \qquad (5)$$

The variation of D with respect to T, assuming constant E, is called the generalized pyroelectric coefficient p_g of the material. That is:

$$p_g = \frac{\partial D}{\partial T} = \frac{\partial P_S}{\partial T} + E\frac{\partial \varepsilon}{\partial T} = p + E\frac{\partial \varepsilon}{\partial T} \qquad (6)$$

Therefore, p_g has two contributions given by the two terms at the second

member of Eq.(6). The first term $p=\partial P_S/\partial T$ is the so-called true pyroelectric coefficient. The second term represents the induced pyroelectric effect. Sensors exploiting the induced pyroelectric effect are said to operate in the dielectric bolometer mode. They do not require poling and are not limited to work below the Curie temperature of the constituent material, but need an applied electric field E [8].

When the material shape and volume are kept constant, i.e. at constant strain, the definition of Eq.(6) describes to the so-called primary pyroelectric effect. When the material is free to undergo thermal expansion, i.e. at constant stress, an additional contribution results from the piezoelectrically induced charge. This second phenomenon is called the secondary pyroelectric effect. In the case of films on a substrate, the partially-clamped boundary conditions and the given materials determine the relative importance of the primary and secondary effects [12].

Limiting to the true pyroelectric effect, if electrodes are placed perpendicular to P in a material where the temperature T is assumed to be uniform and they are wired in a circuit, a time-variation of T induces a current in the circuit, as shown in Figure 8a.

The electrical output can be related to the radiant energy input by means of the thermoelectrical model of Figure 8b, where R_T, R_P and C_T, C_P represent the thermal and electrical resistances and capacitances respectively. Assuming an incoming radiant power W harmonic at the angular frequency ω, the voltage and current responsivities are respectively defined as [7]:

$$R_V(\omega) = \frac{V_p(\omega)}{W(\omega)} \tag{7.a}$$

$$R_I(\omega) = \frac{I_p(\omega)}{W(\omega)} \tag{7.b}$$

The schematic Bode plots of the relevant quantities shown in Figure 8c allow to analyze the dynamic properties of a pyroelectric sensor. In particular, both a thermal and an electrical corner frequencies are present, respectively given by $\omega_T=1/R_TC_T$ and $\omega_E=1/R_PC_P$. The relative position of ω_T and ω_E depends on the material and geometrical parameters. For bulk sensors generally $\omega_E > \omega_T$ while the opposite is true for thin film sensors.

The use of a transimpedance amplifier allows to make R_P and C_P virtually uninfluential, and the electrical time constant can be tailored by means of the external components R_F and C_F.

The simplified model of Figure 8 becomes inadequate when the sensor has a multilayered structure and the thermal profile along the thickness and across the different materials becomes important [13].

4. PZT SCREEN-PRINTED THICK FILMS

Lead zirconate titanate (PZT) is a ferroelectric polycrystalline ceramic made by a solid solution of two perovskite type oxides, i.e. oxides with a formulation of the type ABO_3, namely lead zirconate ($PbZrO_3$) and lead titanate ($PbTiO_3$) [5].

Above the Curie temperature of around 330-350°C, or lower depending on composition, the structure is cubic, therefore the associated symmetry determines a nonpolar behavior. Below the Curie temperature the structure is rhombohedral or tetragonal, therefore the associated asymmetry determines a polar behavior.

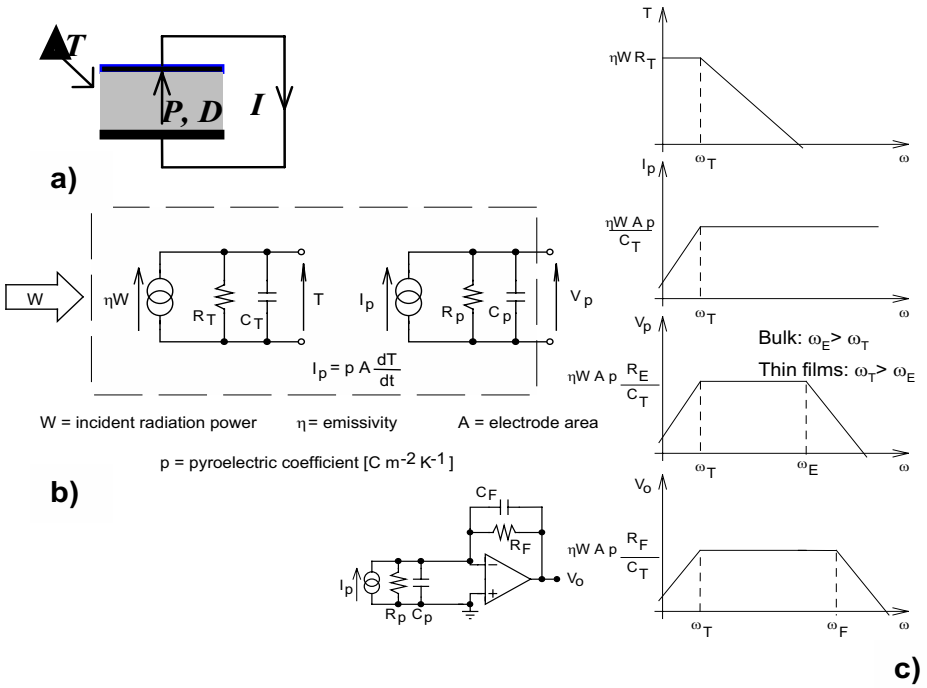


Figure 8. Pyroelectric Sensor Representation (a); Electrothermal model (b); Electrical responses (c).

PZT in general exhibits large piezoelectric and pyroelectric effects, and has a high electromechanical coupling coefficient that makes it ideal in a lot of fields both in sensing and actuating applications. PZT is prepared and commercialized in various compositions with different grades and denominations, such as PZT-4, PZT-5, PZT-5H, and others [5]. Another commercial name of PZT is PXE, standing from piezoxide [9].

The use of PZT films is widespread in sensors and microsystems and there are a number of methods to obtain them, one of which is screen

printing [14]. Compared to other deposition methods for PZT films [8,15], such as physical vapor deposition (PVD), chemical vapor deposition (CVD), chemical solution deposition (e.g. sol-gel by spin coating), screen printing offers several opportunities. First of all, higher film thickness can be obtained, namely in the range 10-100 μm, compared to less than 5 μm/layer for a typical sol-gel process. A high film thickness is often desirable for actuating and sensing purposes. In addition, the manufacturing can be comparatively simple with less critical process parameters, thereby often leading to lower cost.

PZT screen-printable pastes are not generally available as commercial products as readily as the standard TFT inks for electronics and sensor use, such as, for example, the piezoresistive inks. Typically, most of the researchers involved in sensor and microsystem applications of PZT thick films prepare their paste in house, or obtain it from other specialized laboratories. The typical formulation of a PZT screen-printable paste and the flow in the preparation process are described in Figure 9.

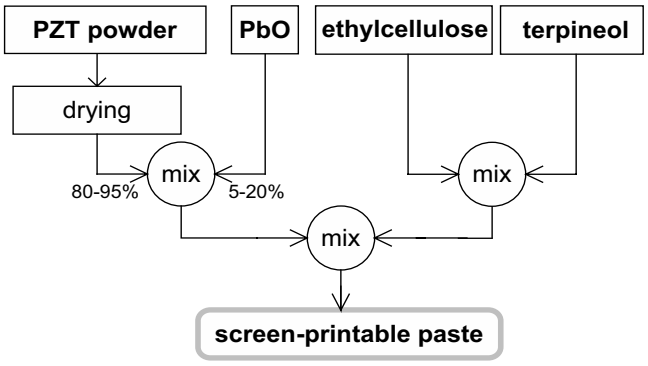


Figure 9. Preparation of PZT Screen Printable Paste.

The functional part is the PZT powder that can have different characteristics depending on specific composition, with average granulometry in the micron range. As the binder, PbO is often used in a mixture of 5 to 20%wt. Alternatively, other materials such as a glass frit or lead borosilicate can be used. As the vehicle, a solution of terpineol in ethylcellulose provides good rehology characteristics, but commercial thinners are used as well. Optionally, grinding of the powder mixture and wet ball milling of the paste are performed to improve homogeneity.

The obtained PZT paste is then screen printed on the substrate as described in Section 2. The most commonly used substrate when using the standard TFT process is alumina (Al_2O_3), but also insulated metals are possible. The electrodes are most often obtained by screen printing

commercial conductor pastes based, for instance, on PdAg compositions. Evaporation of metals is also sometimes used for the topmost electrodes.

After a suitable drying and leveling process, the paste is fired at high temperature in conditions that may be between 900 and 1000°C for 20-30 min. Poling is than performed in typical conditions of applied field between 2 to 5 MV/m at 150°C or more for around 30 min.

The typical characteristics obtainable for PZT thick films by the above described processes are compared in Table 1 to those of the bulk ceramic made by the same powder composition used for the paste.

*Table 1.*Typical Characteristics of a PZT Bulk Ceramic Versus Thick Film of the Same Powder Composition.

parameter	**BULK CERAMIC**	**THICK FILM**
rel. dielectric const. $\varepsilon^T/\varepsilon_o$	1300	570±25
piezoelectric coeff. d [C/N]	$320 \cdot 10^{-12}$	$(130 \pm 20) \cdot 10^{-12}$
density ρ [g/cm^3]	7.7	5±0.1
elastic modulus $1/s^D$ [N/m^2]	$9 \cdot 10^{10}$	$(3.1 \pm 0.3) \cdot 10^{10}$

It can be observed that the thick film globally exhibits lower values for all of the considered properties compared to the bulk ceramic, for instance the mass densities are in a ratio of about 2/3. This is consistently related to differences in the microstructure, whereby the PZT thick films typically exhibit a much high degree of porosity than the bulk counterpart.

Isostatic pressing before firing the films has been proposed to reduce porosity and improve their overall characteristics [16].

Microstructural features of PZT thick films strongly depend on the starting powder, paste composition, and processing parameters [17]. An example is shown in the pictures of Figure10.

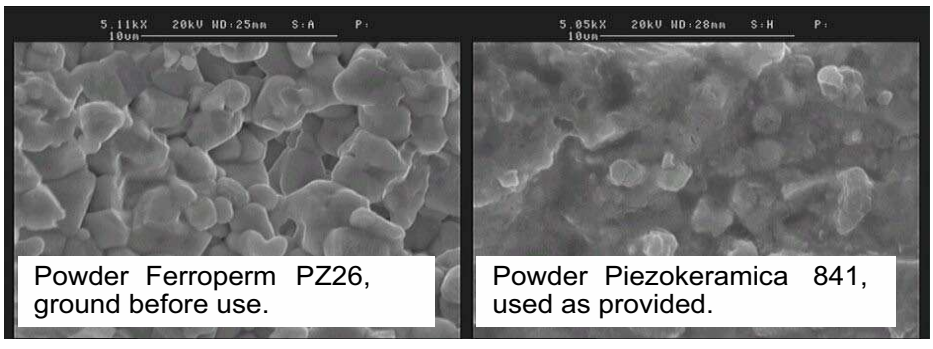


Figure 10. Microstructural Comparison of PZT Thick Films Prepared at the University of Brescia.

In addition to the standard processes and materials of TFT, thick films of PZT for sensing applications can also be obtained on other substrates and with somewhat different deposition methods.

An interesting and topical research field is the screen printing of PZT thick films on silicon for microsensing and MEMS applications [18,19]. Screen-printed-on-silicon PZT is attractive in all those applications where the higher thickness obtainable compared to different deposition methods is at premium. Apart from the aspects related to the dimensional resolution and the feasibility of printing on fragile suspended elements such as beams and membranes, an important issue to be considered in the overall process is the lead diffusion into silicon driven by the high firing temperature [20].

Other interesting variants for deposition methods of PZT thick films are, for instance, the use epoxy resins as binders that allow for low-temperature curing [21], the preparation of piezoelectric "paints" that can lead to films on plastics and flexible substrates [22], or sol-gel spraying on metals and curved surfaces [23] by which a large thickness is obtainable.

5. PIEZOELECTRIC ACOUSTIC-WAVE THICK-FILM SENSORS

Different types of piezoelectric transducers can be obtained depending on the configurations in which piezoelectric elements are arranged [24, 25], as summarized in the four cases of Figure 11 described in the following.

Figure 11a represents the case where the input signal is the electrical (E) stimulus to the piezoelectric element, and the output signal is the mechanical (M) response. Therefore, the configuration is based on EM conversion and the piezoelectric transducer works as an actuator. Examples of this type are micropositioners, motors, sonic or ultrasonic emitters.

Figure 11b represents the opposite case, where the input signal is the mechanical stimulus to the piezoelectric element, and the output signal is the electrical response. The configuration is now based on ME conversion and the piezoelectric transducer works as a sensor. Examples of this type are vibration pick-ups, sonic or ultrasonic receivers.

Figure 11c represents a sensor configuration based on an EM-ME open-loop conversion, where an electrical stimulus to an input piezoelectric element generates a mechanical effect in a medium that is converted back into an electrical response by a second piezoelectric element. The input quantity, i.e. the one to be measured, modulates some property of the mechanical propagation path between the two piezoelectric elements, and it is transduced into the output signal in the form of a perturbation of the electrical transfer function between stimulus and

response.

The mechanical propagation path can be external to the transducer, such as for instance in ultrasound flowmeters, or be an integral part of the transducer itself. In this case, the transducer sustains the propagation of mechanical elastic vibrations, i.e. acoustic waves, and the sensor can be called an acoustic-wave (AW) sensor [26, 27]. The presence of two separate piezoelectric elements is not strictly required, in fact a single element as well can be used. In this case, the mechanoacoustic propagation path becomes the piezoelectric element itself.

Figure 11d extends the concept of acoustic-wave sensors by inserting the configuration of Figure 11c as the feedback element of an amplifier, therefore creating an EM-ME closed-loop conversion. If proper gain conditions are satisfied, the system can be made resonating at one of the natural frequencies of the mechanoacoustic propagation path that, in turn, depends on the input measurand quantity. Therefore, the resonant frequency can be conveniently taken as the output signal. The resultant device can be called an acoustic-wave resonant sensor [28-33].

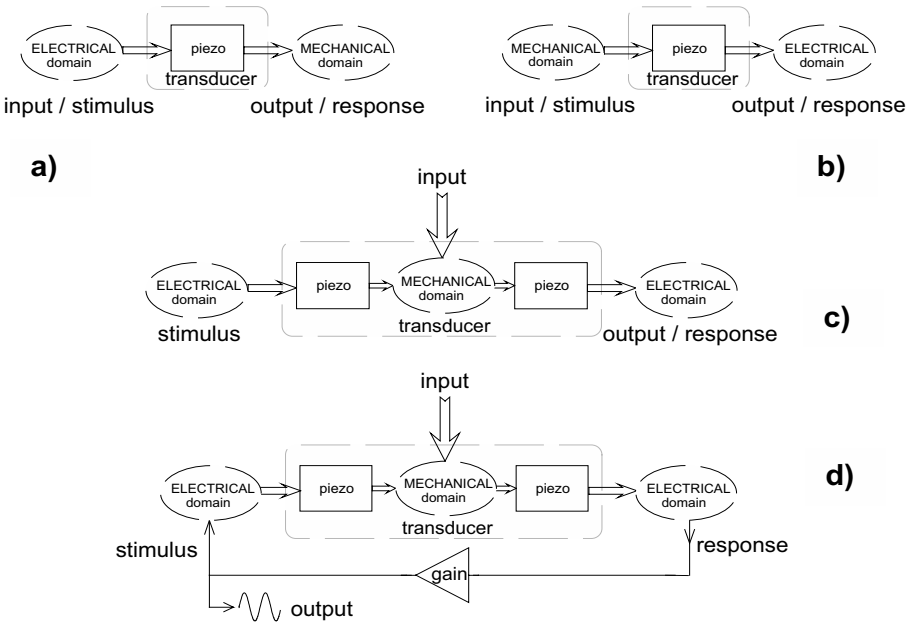


Figure 11. Types of Piezoelectric Transducers: EM Conversion (a); ME Conversion (b); EM-ME conversion (c); Closed-loop EM-ME Conversion (d).

AW resonant sensors offer a number of advantages related to their operating principle and frequency output, such as high potential sensitivity which is solely determined by mechanical properties and is virtually

unaffected by piezoelectric coefficients and amplifier gain, good stability, resolution and noise immunity, direct interfaceability with digital systems.

Piezoelectric acoustic-wave sensors base their operation on piezoelectrically-excited matter vibrations propagating in elastic media. Several wave propagation modes are possible, including longitudinal or transverse (shear) waves, for which particle motion is respectively parallel or normal to the wave propagation direction, or combinations of the two. Such a diversity in wave modes gives rise to a number of different sensor devices, such as the bulk acoustic wave (BAW), surface acoustic wave (SAW), and flexural plate wave (FPW) devices, to name a few [26-27,34].

Apart from their specific configuration and structure, piezoelectric AW sensors can be grouped into two general categories, namely the one-port, and the two-port sensors.

As shown in Figure 12, one-port sensors have a single pair of connections for both stimulus and response, while two-port sensors have two separate connection pairs. Indicating with H the electrical transfer function being modulated by the measurand input, for one-port sensors H is the impedance (or the admittance) relating current and voltage at the sensor port, while for two-port sensors H is a current-current, voltage-voltage or mixed ratio that mutually couples the two ports.

In both cases, mechanical resonances in the sensor reflect into electrical resonances of the transfer function H, and therefore the former ones can be detected and referred to the measurand by measuring the latter ones.

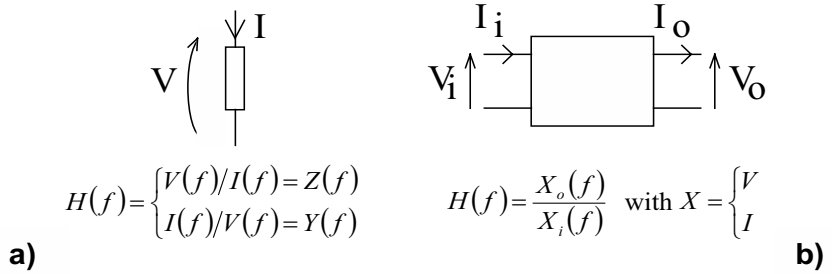


Figure 12. One-port (a) and Two-port (b) Piezoelectric Acoustic-wave Sensors.

AW sensors in the closed-loop configuration of Figure 11d are of two types, namely resonators, and delay lines. They are shown in Figure 13.

Resonators around each resonant frequency to first approximation can be thought as 2-nd order lossless mechanical systems with equivalent mass M and spring constant K, respectively related to the material density ρ and stiffness c and both depending on a frequency-determining dimension l. The resonant frequency in question can be expressed as:

$$f = \frac{1}{2l}\sqrt{\frac{c}{\rho}} = \frac{1}{2\pi}\sqrt{\frac{K}{M}} \tag{8}$$

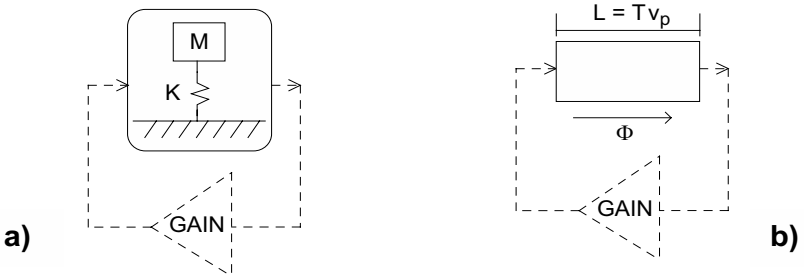


Figure 13. Resonator (a) and Delay Line (b) Configurations for Acoustic-wave Resonant Sensors.

Delay lines behave as acoustic propagation paths of length L and phase velocity v_p. The resonant condition is that for which the phase Φ around the loop is equal to $2n\pi=\omega L/v_p$, where n is an integer. Therefore, the

$$f = n\frac{v_p}{L} \tag{9}$$

resonant frequencies are harmonics given by:

The sensitivity s_x and fractional sensitivity S_x towards a generic measurand x can be defined as [35]:

$$S_x = \frac{1}{f}\frac{df}{dx} = \frac{1}{f}s_x \tag{10}$$

For a resonator sensor, from Eqs.(8) and (10) S_x can be derived as:

$$S_x = \left(\frac{dc}{2c} - \frac{d\rho}{2\rho} - \frac{dl}{l}\right)\frac{1}{dx} = \frac{dK}{2Kdx} - \frac{dM}{2Mdx} \tag{11}$$

Eq.(11) shows that if the measurand x produces an increase (decrease) in the equivalent spring constant K by increasing (decreasing) the stiffness c, then the resonant frequency increases (decreases). Conversely, if the x produces an increase (decrease) in either the equivalent mass M by increasing (decreasing) the density ρ, or the dimension l, or both, then the

resonant frequency decreases (increases).

For a delay-line sensor, from Eqs.(9) and (10) S_x can derived as:

$$S_x = \frac{1}{v_p}\frac{dv_p}{dx} - \frac{1}{L}\frac{dL}{dx} \tag{12}$$

Eq.(12) shows that if the measurand x produces an increase (decrease) in the phase velocity v_p, then the resonant frequency increases (decreases). Conversely, if the x produces an increase (decrease) in the propagation length L, then the resonant frequency decreases (increases).

The above considerations, though derived starting from simplified models, are useful to qualitatively predict the effect of a given measurand on a resonant sensor output and to provide guidelines for design.

Some examples of AW piezoelectric resonant sensors made in TFT will be briefly discussed in the following.

One of the first TFT piezoelectric resonant sensors ever reported is a pressure sensor [36, 37]. The sensor has a two-port resonator configuration based on an alumina circular diaphragm excited in flexural resonance by properly driven PZT films. The differential pressure applied across the diaphragm deforms it and generates a radial tension. This in turn increases the equivalent stiffness, therefore, according to Eq.(12), the resonant frequency rises. The reported sensitivity was around 150 Hz/mm_{Hg} at a frequency of 57.8 kHz.

In another device, a surface-wave sensor structure was realized by screen printing on a rectangular alumina substrate two PZT elements with interdigitated electrodes respectively working as emitter and receiver of acoustic waves along the substrate [38]. In this case, the sensor is configured as a two-port delay line that, equipped with an electronic amplifier, behaves as a delay-line oscillator. A force applied in the center of the substrate due to a mass load generates a strain or, equivalently, an increase of the acoustic path length L. This in turn, according to Eq.(12), produces a decrease in the frequency. A frequency decrement of around 40 Hz for a load of 100 g was reported at a resonant frequency of 8 MHz.

Figure 14 shows an hydrostatic pressure resonant sensor based on a rectangular alumina beam on which two PZT thick-film elements are positioned [39]. They work as drive and sense elements in a two-port configuration and, by closing the loop with a proper amplifier circuit, they allow to excite the beam in flexural resonance. As shown in Figure 15a, the beam was mounted in a DIL package supported by the bonding wires in an approximately free-free condition that provided a marked resonance at the fifth flexural mode located at 198 kHz. The hydrostatic pressure of a surrounding gas has the effect of increasing the beam inertia, therefore the resonant frequency decreases. This is shown in Figure 15b that reports the

sign-inverted frequency shift at the oscillator output as a function of the pressure increment over the ambient value. A linear behavior up to 4 bar can be observed, with a sensitivity of around –35 Hz/bar.

This sensor configuration has an extended measuring range and a high overload capability since no deformable diaphragms are present that may fracture. The operation limit is determined by the amount of tolerable damping before the resonance becomes so loosely detectable by the electronic circuit that oscillations stop. On the other hand, the sensitivity depends on the particular gas.

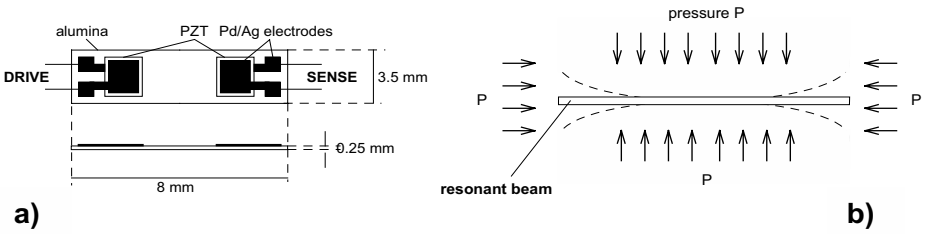


Figure 14. Layout (a) and Operating Principle (b) of a Hydrostatic Pressure Resonant Sensor.

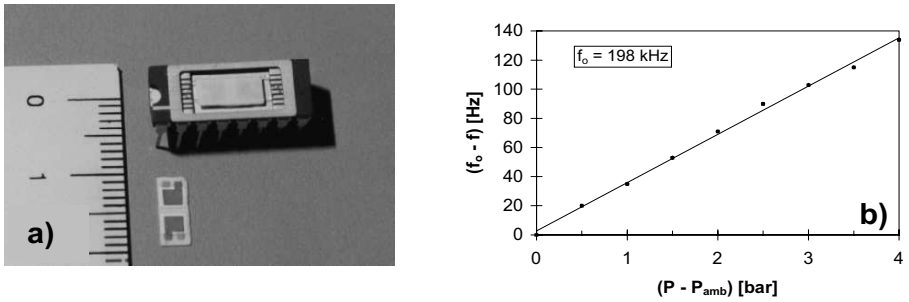


Figure 15. Picture (a) and Resonant Frequency Versus Pressure (b) of the sensor of Figure 14.

Figure 16a illustrates a PZT-on-steel resonant tuning fork force sensor [40]. On a steel beam a dielectric insulation layer was screen printed followed by three PZT thick films with associate electrodes. The two elements at the beam ends connected in parallel work as a driver and the central element as a sensor in a two-port configuration which, arranged in a closed loop with an amplifier, allows to excite the beam in flexural

resonance.

When the beam is clamped at both ends, an axial traction force increases the equivalent stiffness, therefore the resonant frequency rises. The relationship between frequency and applied force is linear for small forces. The beam was mounted on a clamped-free aluminum frame to transform an applied vertical force into an elongation of the beam as shown in Figure 16b. A sensitivity of 1.1 Hz/N at the resonant frequency of 35 kHz was measured, with a linear range up to 800 N.

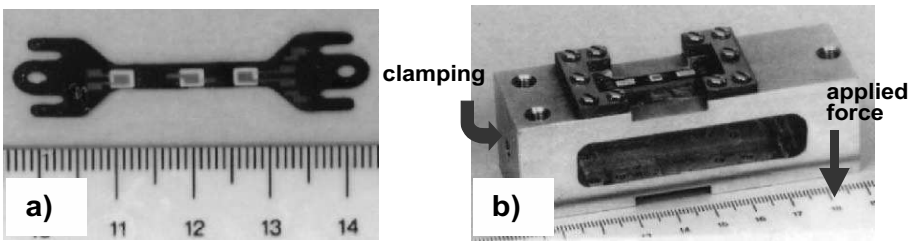


Figure 16. PZT-on-steel Resonant Tuning Fork Force Sensor (a); Tuning Fork Sensor Mounted on its Supporting frame (b).

An important topic of study and field of application of PZT thick films is variable-mass AW sensors [32, 41] or, more generally, AW sensors responsive to an applied acoustic load. Variable-mass, also called gravimetric, AW sensors work as resonant microbalances. By functionalizing the surface with a sorption coating they can be used for chemical sensing, as schematically shown in Figure 17.

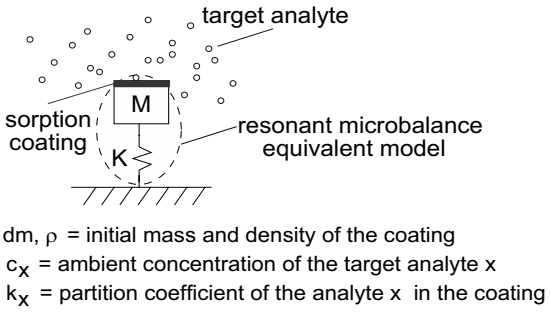


Figure 17. Schematic Representation of a Resonant-microbalance Chemical Sensor.

With the meaning of symbols given in Fig.17 and indicating by S_m the mass fractional sensitivity of the sensor as a microbalance, the relative frequency shift in the presence of a given ambient concentration of the target analyte can be expressed as [26]:

$$\frac{df}{f} = S_m dm\left(1+\frac{k_x}{\rho}c_x\right) \tag{13}$$

PZT thick-films were firstly used in gravimetric sensing as part of two-port resonator sensors with polymeric coatings for the detection of humidity in air [42,43]. A significant increase in mass sensitivity was later obtained by using the PZT thick film itself as the resonator forming the microbalance, leading to a device named resonant piezo-layer (RPL) sensor [44, 45].

As shown in Figure 18, the structure of a thick-film RPL sensor is made by the superposition of a non-piezoelectric substrate, specifically alumina, a bottom electrode layer, a PZT layer poled along its thickness, and a top electrode layer. The PZT film is acoustically coupled to the substrate, leading to a thickness-expansion composite resonator, which can be used as a bulk acoustic-wave sensor responsive to an acoustic surface load.

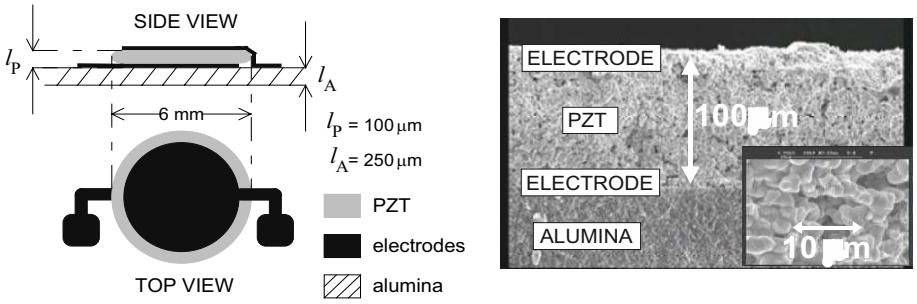


Figure 18. Structure of a Resonant Piezo-layer (RPL) Sensor.

Typical dimensions are 5 mm for the electrode diameter, 250, 10 and 90 μm for the thickness of the substrate, electrodes and PZT respectively. This results in a resonant frequency of around 7 MHz, while higher values can be achieved for thinner devices.

A surface mass load on the top electrode determines a frequency decrement with a sensitivity that exceeds 500 kHz/mg at 7 MHz [46]. This is comparable to what obtainable with a quartz crystal microbalance (QCM) sensor at the same frequency. As an advantage, the RPLs offer the opportunity to fabricate low-cost multisensor arrays on the same substrate,

as shown in Figure 19a. The array also integrates a resistive heater driven in the constant-temperature mode, as illustrated in Figure 19b, to minimize the thermal drift of the sensor resonant frequencies that typically can be as high as 800-1000 kHz/°C.

Figure 20a reports the measured mass sensitivity of RPL sensors having different thickness of the PZT films while maintaining the same 250-µm thick alumina substrate. By the use of proper sorption coatings, the achieved mass sensitivity is suitable for gravimetric chemical sensing [47, 48]. As an example, Figure 20b shows the comparison of the cyclic responses to toluene of two 7-MHz RPL and two 10-MHz QCM sensors sensitized with cavitand receptor films [49, 50].

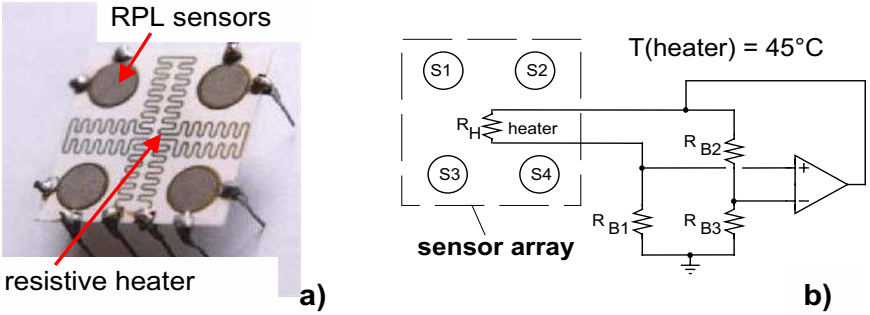


Figure 19. Array of Four RPL Sensors Plus Integrated Resistive Heater (a); Constant-temperature Driving Mode of the Heater for Thermal Stabilization (b).

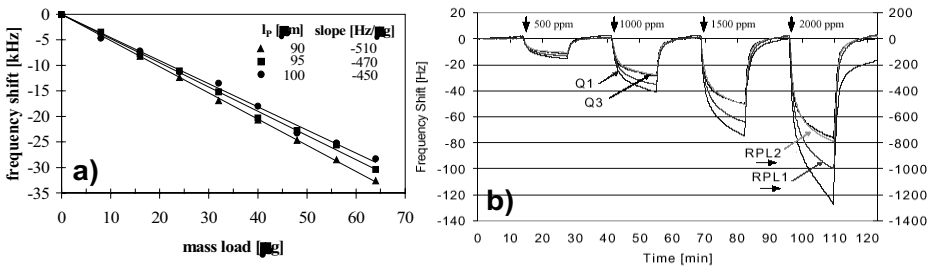


Figure 20. Measured Mass Sensitivity of RPL Sensors of Different PZT Thickness (a); Response to Toluene of RPL and QCM Sensors Sensitized with Cavitand Receptors (b).

While RPL sensors have been so far investigated for gravimetric chemical sensing in air, other AW sensor configurations based on PZT thick films could be desirably envisaged for in-liquid measurements. Interesting results in this respect are, for instance, the use of silicon microcantilevers actuated by PZT thick films for bio-chemical analysis in solution [51], and the exploitation of PZT elements to generate flexural plate waves in micromembranes starting from bulk modes [52].

6. PYROELECTRIC THICK-FILM SENSORS

The same as with the piezoelectric properties, also in the case of the pyroelectric characteristics, PZT thick films exhibit differences compared to the bulk ceramic counterpart [53, 54].

In particular, the reported pyroelectric coefficient p is typically between 1.2 and 1.5×10^{-4} $Cm^{-2}K^{-1}$ for thick films, depending on paste composition and poling parameters, while is around 3.8×10^{-4} $Cm^{-2}K^{-1}$ for bulk PZT. Densification processes based on the application of isostatic pressure before firing has been proposed to increase p [55]. On the other hand, the much lower dielectric permittivity of thick films favorably determines a voltage responsivity that can be similar or higher than for bulk ceramics.

Pyroelectric sensors employing PZT or modified-PZT screen-printed thick films have been reported as single-element detectors [54, 56-57], as well as in the form of multisensor arrays [53, 58].

Often, the attention has been focused more on material properties rather than on the design of sensor systems for particular applications. Typically, measurements have been carried out mainly with relatively high-power light sources, such as lamps and lasers.

For example, thick films of lead lanthanum zirconate titanate (PLZT) were used to fabricate a linear array that was tested under excitation by means of a flash lamp [58]. A pyroelectric detector made by a PZT thick film with buried electrodes was proposed and tested by means of an IR chopped source [54].

Multi-element thick-film pyroelectric sensors were also proposed for wide spectral range radiation detection (from IR to UV), and laser beam alignment and profiling. Figure 21a and Figure 21b respectively show an array of 3x3 pixels [59] and a four-quadrant detector [60] made by PZT thick films on alumina intended for laser spot localization.

It has been demonstrated that PZT thick films can produce significant pyroelectric signals when excited by low-power light sources, such as a light-emitting diode (LED) in the visible.

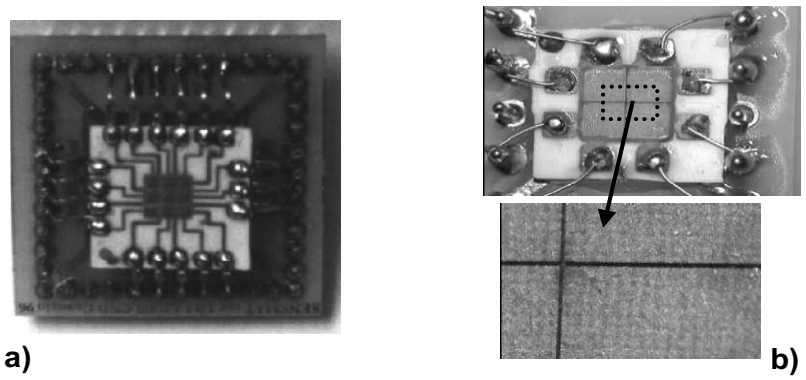


Figure 21. Pyroelectric Multielement Sensors: Array of 3x3 Pixels (a) and Four-quadrant Detector (b).

If, as shown in Figure 22, the LED is driven in amplitude modulation and the corresponding pyroelectric current is converted into a voltage, synchronously rectified, and low-pass filtered, then a DC voltage V_o is obtained which depends on the light intensity collected from the LED. Interferences from thermal and light background are rejected. The maximum responsivity is typically obtained for a modulation frequency in the order of 100 Hz. The typical output signal of the transresistance amplifier shown in Figure 22, taking into account of the gain, sensor dimensions and pyroelectric coefficient, corresponds to peak temperature variations in the PZT film of about 70 μK.

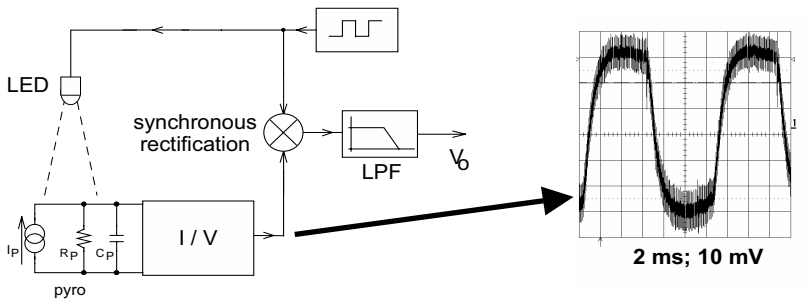


Figure 22. LED Excitation and Synchronous Detection of a Pyroelectric Signal.

Based on this principle, pyroelectric contactless displacement sensors were proposed. For example, the sensor of Figure 23a and a coupled LED

respectively work as the reference element and the movable light-spot cursor to contactless measure linear displacement [61].

The sensor consists of a planar multi-layer structure on alumina substrate where a rectangular film of PZT is enclosed between two PdAg electrode layers at the top and bottom. Each layer is actually composed by a pair of triangular split electrodes (left, L, and right, R). The resulting four electrodes are combined in reverse-parallel with respect to the common poling direction, so that the pyroelectric signals from the R and L subsections subtract in a push-pull configuration.

The LED spot position along the sensor length is converted into a suitably linear DC output by means of the readout method of Figure 22. The readout method, combined with the differential configuration of the sensor, provides rejection of vibrations and microphonics caused by the piezoelectric response of PZT, and of external thermal and light sources.

Figigure 23b shows the measured sensor response over the of range ±10 mm for different driving currents of the LED. The sensitivity ranges from 66 to 120 mV/mm with a nonlinearity of about 4% of the span.

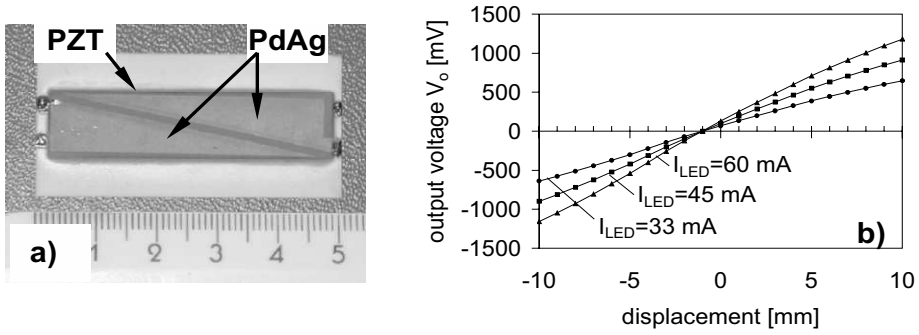


Figure 23. Contactless Linear Displacement Sensor to be Coupled with a LED Cursor (a); Sensor Output Versus LED Position for Different Driving Currents (b).

An extension of the same principle was proposed for biaxial position measurements by means of the pyroelectric sensor array shown in Fig.24.

The sensor array is made by 8x8 pixels of PZT screen printed on an alumina substrate [62]. The electrodes are configured as 8 columns and 8 rows evenly spaced, whose respective crossings form the 64 pixels of the array.

When an amplitude-modulated LED is placed in front of the array and moved in the XY plane parallel to the array surface, as illustrated in Fig.25a, the array and the LED respectively work as the reference element and the contactless cursor of a biaxial position sensor. By means of a purposely designed method of readout and processing of the pyroelectric signals, a continuos and linear dependence of the output voltages on XY-

position of the LED were obtained. The sensitivity was about 1 V/mm over the range ±3.5 mm, with a nonlinearity within ±1% of the span. The resolution threshold was 4 μm, therefore much better than the pixel-to-pixel pitch.

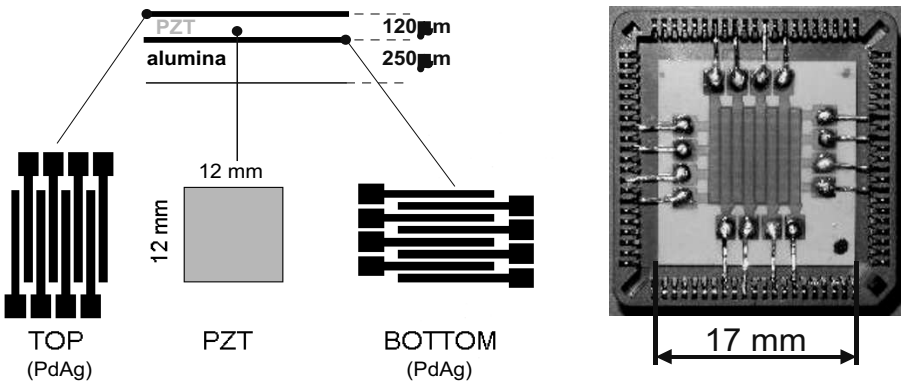


Figure 24. Pyroelectric Array with 8x8 Pixels for Biaxial Measurement of LED Spot Position: Array Layout (a) and Picture (b).

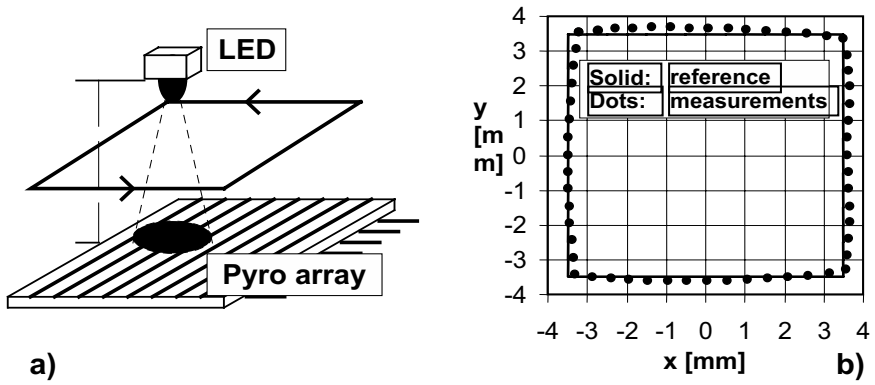


Figure 25. Test Configuration for the Pyroelectric Array (a) and Measurement Results Along a Square Reference Path (b).

Fig.25b shows the results of biaxial measurements along the worst-case squared test path with side of 7 mm, i.e. on the borders of the linearity region. A good reconstruction of the LED trajectory can be observed.

Compared to semiconductor position-sensitive detectors, thick-film pyroelectric sensors and driven LEDs for contactless displacement measurement offer the advantage of lower cost at larger sensing areas. On

the other hand, due to their poor high-frequency response with standard film and substrate thickness, they have been restricted so far to quasi-static applications.

7. CONCLUSIONS

Both driven by an increasing number of new potential applications and stimulated by technological advances, the field of sensors is evolving along a trend from single-element sensors to sensor systems with complex functionality. As a consequence, a successful sensor technology is ever increasingly required to comply with and support such a trend.

In this respect, mono-technology solutions seem to be not that efficient, with a reason for this certainly represented by the peculiarities of the sensor field compared to other fields, like, for instance, the microelectronics. A good example is the role and potential of silicon micromachining and MEMS technology for sensors [63]. The main characteristics and strengths of silicon micromachining are miniaturization to the microscopic scale, standardization of materials and processes, leading to batch production and, in turn, to low cost provided that large-volume markets are available.

On the other hand, the typical requirements for sensors are miniaturization often to the mesoscopic scale, wide diversification of materials related to the sensing principles and target operating conditions, consequent need for nonstandard processes and packaging, low to medium cost in markets that, apart from few exceptions, are low to medium in volume.

In fact, it appears that the sensor and microsystem field intrinsically demands for a range of different microfabrication technologies [64], possibly combined together in hybrid optimized solutions.

In this framework, TFT traditionally has been playing a role that can be expanded when combined with non-conventional pastes and new sensor designs, such as with the piezoelectric and pyroelectric devices illustrated in this chapter.

ACKNOWLEDGEMENTS

Several among the sensor prototypes here presented were developed in the Sensor and Electronic Instrumentation Laboratory at the University of Brescia, as a cooperative effort of a number of colleagues. Credit and appreciation go in particular to Andrea Taroni, Daniele Marioli, Emilio Sardini, Damiano Crescini, Alessio Ghisla, and Marco Ferrari.

REFERENCES

[1] Holmes P.J., Loasby R.G., *Handbook of Thick Film Technology*. Ayr (Scotland): Electromechanical Publications Limited, 1976.

[2] Prudenziati M. Ed., *Thick Film Sensors*. Amsterdam: Elsevier, 1994.

[3] White N.M., Turner J.D., Thick-film sensors: past, present and future. *Meas. Sci. Technol*., 1997; 8, 1: 1-20.

[4] Gongora-Rubio M.R., Espinoza-Vallejos P., Sola-Laguna L., Santiago-Avlés J.J., Overview of low temperature co-fired ceramics tape technology for meso-system technology (MsST). *Sensors and Actuators A,* 2001; 89: 222-241.

[5] Ikeda T., *Fundamentals of Piezoelectricity*. Oxford: Oxford University Press, 1989.

[6] *IEEE Standard on Piezoelectricity, ANSI/IEEE Std 176-1987*, New York: Institute of Electrical and Electronic Engineers, 1987.

[7] Whatmore R.W., Pyroelectric devices and materials. *Rep.Prog.Phys.*, 1986; 49: 1335-1386.

[8] Muralt P., Micromachined infrared detectors based on pyroelectric thin films. *Rep.Prog.Phys.*, 2001; 64: 1339-1388.

[9] Waanders J.W., *Piezoelectric ceramics*. Ehindhoven: Philips Components, 1991.

[10] Brissaud M., Characterization of Piezoceramics. *IEEE Trans. on UFFC,* 1991; 38, 6: 603-617.

[11] Auld B.A., *Acoustic Fields and Waves in Solids*. Vol. 1-2, New York: John Wiley & Sons, 1973.

[12] Liu S.T., Long D., Pyroelectric detectors and materials. *Proc. of IEEE*, 1978; 66: 14-26.

[13] Samoilov V.B., Yoon Y.S., Frequency Response of Multilayer Pyroelectric sensors. *IEEE Trans. on UFFC*, 1998; 45, 5: 1246-1254.

[14] De Cicco G., Morten B., Prudenziati M., "Piezoelectric Thick-film Sensors." In *Thick Film Sensors*, Prudenziati M., Ed. Amsterdam: Elsevier, 1994, 209.

[15] Whatmore R.W., Zhang Q., Huang Z., Dorey R.A., Ferroelectric thin and thick films for microsystems. *Materials Science in Semiconductor Processing*, 2003; 5: 65–76.

[16] Yao K., Zhu W., Improved preparation procedure and properties for a multilayer piezoelectric thick-film actuator. *Sensors and Actuators A,* 1998; 71: 139-143.

[17] Torah R.N., Beeby S.P., White N.M., Improving the piezoelectric properties of thick-film PZT: the influence of paste composition, powder milling process and electrode material. *Sensors and Actuators A*, 2004; 110: 378-384.

[18] Beeby S.P., White N.M., Silicon micromechanical resonator with thick-film printed vibration excitation and detection mechanisms. *Sensors and Actuators A*, 2001; 88: 189-197.

[19] http://www.ikts.fhg.de/business/funktionskeramik/piezotechnik/keramik/si.pdf

[20] Glynne-Jones P., Beeby S.P., Dargie P., Papakostas T., White N.M., An investigation into the effect of modified firing profiles on the piezoelectric properties of thick-film PZT layers on silicon. *Meas. Sci. Technol*., 2000; 11: 526-531.

[21] Clegg W.W., Jenkins D.F.L., Cunningham M.J., The preparation of piezoceramic.polymer thick-films and their application as micromechanical actuators. *Sensors and Actuators A,* 1997; 58: 173-177.

[22] Egusa S., Iwasawa N., Preparation of Piezoelectric Paints and Application as Vibration Modal Sensors. *Journ. Intelligent Material Systems and Structures,* 1994; 5: 140-144.

[23] Kobayashi M., Olding T.R., Sayer M., Jen C.K., Piezoelectric thick film ultrasonic transducers fabricated by a sol-gel spray technique. *Ultrasonics*, 2002: 39: 675-680.

[24] Mason W.P., *Electromechanical Transducers and Wave Filters*. New York: Van Nostrand Company, 1948.

[25] Kino G.S., *Acoustic Waves: Devices, Imaging and Analog Signal Processing*.

Englewood Cliffs, NJ: Prentice-Hall, 1987.
[26] Ballantine D.S., White R.M., Martin S.J., Ricco A.J., Zellers E.T., Fryre G.C., Wohltjen H., *Acoustic wave sensors*. San Diego: Academic press, 1997.
[27] Vellekoop M.J., Acoustic wave sensors and their technology. *Ultrasonics*, 1998; 36: 7-14.
[28] Buser R. A., "Resonant Sensors" In *Sensors vol. 7: Mechanical Sensors*, H.H. Bau, N. F. de Rooij, B. Kloeck, Eds.. Weinheim: VCH, 1994; 208-279.
[29] Langdon R. M., Resonator sensors-a review. *J. Phys. E: Sci. Instrum.*, 1985; 18: 103-115.
[30] Hauptmann P., Resonant Sensors and Applications. *Sensors and Actuators A*, 1991; 25-27: 371-377.
[31] Eernisse E. P., Ward R. W., Wiggins R. B., Survey of Quartz Bulk Resonator Sensor Technologies. *IEEE Trans. Ultrason., Ferroelec., Freq. Cont.*, 1988; 35: 323-330.
[32] Benes E., Gröschl M., Burger W., Schmid M., Sensors based on piezoelectric resonators. *Sensors and Actuators A*, 1995; 48: 1-21.
[33] Zelenka J., *Piezoelectric Resonators and their Applications.* Amsterdam: Elsevier, 1986.
[34] Benes E., Gröschl M., Seifert F., Pohl A., Comparison between BAW and SAW sensor principles. *IEEE Trans. Ultrason., Ferroelec., Freq. Cont.*, 1998; 45, 5: 1314-1330.
[35] Vig J. R., On Acoustic Sensor Sensitivity. *IEEE Trans. on UFFC,* 1991; 38, 3: 311.
[36] Morten B., De Cicco G., Gandolfi A., Tonelli C., PZT-based Thick Films and the Development of a Piezoelctric Pressure Sensor. *Hybrid Circuits*, 1992; 28: 25-28.
[37] Morten B., De Cicco G., Prudenziati M., Resonant pressure sensor based on piezoelectric properties of ferroelectric thick films. *Sensors and Actuators A*, 1992; 31: 153-158.
[38] White N. M., Ko V.T.K., Thick-film acoustic wave sensor structure. *Electronics Letters*, 1993; 29: 1807-1808.
[39] Ferrari V., Marioli D., Taroni A., Trasduttori piezoelettrici in tecnologia a film spesso. *Alta Frequenza*, 1997; 9, 3: 52-56.
[40] Crescini D., Ferrari V., Marioli D., Taroni A., Thick-film resonant sensor for force measurement on stainless steel substrate. *Sensors and Microsystems: Proceedings of the First National Conference on Sensors and Microsystems AISEM,* Roma, 19-20 February 1996, Singapore: World Scientific Publishing, 1996; 256-260.
[41] Nieuwenhuizen M. S., Venema A., "Mass-Sensitive Devices." In *Sensors vol. 2: Chemical and Biochemical Sensors, Part I*, Göpel W., Hesse J., Zemel J. N., Eds., Weinheim: VCH, 1991; 647-680.
[42] Morten B., De Cicco G., Prudenziati M., A thick-film resonant sensor for humidity measurements. *Sensors and Actuators A*, 1993; 37-38: 337.
[43] Ferrari V., Marioli D., Taroni A., Ranucci E., Ferruti P., Development and Application of Mass Sensors based on Flexural Resonances in Alumina Beams. *IEEE Trans. on UFFC,* 1996; 43, 4: 601-608.
[44] Ferrari V., Marioli D., Taroni A., Self-resonant piezoelectric thick films for sensing applications. *Electron. Lett*., 1996; 32, 9: 855-856.
[45] Ferrari V., Marioli D., Taroni A., Thick film resonant piezo-layers as new gravimetric sensors. *Meas. Sci. Technol*., 1997; 8, 1: 42-48.
[46] Ferrari V., Marioli D., Taroni A., Theory, Modelling and Characterization of PZT-on-Alumina Resonant Piezo-Layers as Acoustic-Wave Mass Sensors. *Sensors and Actuators A,* 2001; 92, 1-3: 182-190.
[47] Ferrari V., Marioli D., Taroni A., Ranucci E., Ferruti P., Gravimetric chemical sensors in thick film technology with hybrid electronics. *Conference Proceedings of SAA'96 - National Meeting on Sensors for Advanced Applications*, Brescia, 16-17 May 1996; Bologna: SIF, 1997; 54: 3-10.
[48] Ferrari V., Marioli D., Taroni A., Ranucci E., Multisensor Array of Mass

Microbalances for Chemical Detection Based on Resonant Piezo-Layers of Screen-Printed PZT. *Sensors and Actuators B,* 68, 2000; 1-3: 81-87.

[49] Ferrari M., Ferrari V., Marioli D., Taroni A., Suman M., Dalcanale E., Comparative Experiments on Cavitand-coated PZT Resonant Piezo-Layer and QCM Sensors Exposed to Organic Vapors. Book of Abstracts *8th National Conference on Sensors and Microsystems*, Trento, 12-14 February, 2003; 156.

[50] Ferrari M., Ferrari V., Marioli D., Taroni A., Suman M., Dalcanale E., Investigation on Cavitand-Coated PZT Resonant Piezo-Layer and QCM Sensors at Different Temperatures Under Exposure to Organic Vapors. *Proc. of the Eurosensors XVII Conference,* Guimarães, Portugal, 21-24 September, 2003; 402-405, in press on *Sensors and Actuators B*.

[51] Kim H.J., Kim Y.B., Park J., Kim T.S., Biological Element Detection Sensor Application Of Micromachined PZT Thick Film Cantilever. *Proc. of IEEE Sensors 2003*, Toronto, CANADA, October 22-24, 2003; 1054-1058.

[52] Guo H., Lal A., Flexural Plate wave excitation using bulk modes. *IEEE Ultrasonics Symp*., 2001; 799-802.

[53] Baudry H., Delmas G., Thick film pyroelectric sensor array. *Proc. 8th European Hybrid Microelectronics Conference,* Rotterdam, 28-31 May 1991; 227-235.

[54] De Cicco G., Morten B., Dalmonego D., Prudenziati M., Pyroelectricity of PZT-based thick films. *Sensors and Actuators A*, 1999; 76: 409-415.

[55] Lucat C., Menil F., Von Der Mühll R., Thick-film densification for pyroelectric sensors. *Meas. Sci. Technol*., 1997; 8, 1: 38-41.

[56] Leppävuori S., Lozinski A.H., Ussimäki A., A thick-film pyroelectric PLZT ceramic sensor. *Sensors and Actuators A*, 1995; 46-47: 391-394.

[57] Lozinski A., Wang F., Ussimäki A., Leppävuori S., PLZT thick-films for pyroelectric sensors. *Meas. Sci.Technol.*, 1997; 8: 33-37.

[58] Lozinski A., Wang F., Ussimäki A., Leppävuori S., Thick-film pyroelectric linear array. *Sensors and Actuators A*, 1998; 68: 290-293.

[59] Capineri L., Ferrari V., Naldoni F., Masotti L., Marioli D., Taroni A., 3 x 3 matrix of thick-film pyroelectric transducers. *Electron. Lett.,* 1998; 34, 15: 1486-1487.

[60] Capineri L., Ferrari V., Marioli D., Masotti L., Taroni A., Characterization and modeling of PZT-based pyroelectric sensors. *Sensors and Microsystems: Proceedings of the 4th Italian Conference on Sensors and Microsystems AISEM,* Roma, 3-5 February 1999, Singapore: World Scientific Publishing, 2000; 316-321.

[61] Ferrari V., Marioli D., Taroni A., Displacement Sensor Based on Pyroelectric Thick Films and Contactless Light-Spot Cursor. *IEEE Trans. on Instrum. Meas.*, 2002; 51, 4: 819-823.

[62] Ferrari V., Ghisla A., Marioli D., Taroni A., Array of PZT Pyroelectric Thick-Film Sensors for Contactless Measurement of XY Position. *IEEE Sensors Journal,* 2003; 3, 2: 212-217.

[63] Kovacs G.T.A., *Micromachined Transducers Sourcebook*. New York: WCB/McGraw-Hill, 1998.

[64] Madou M., *Foundamentals of Microfabrication*. Boca Raton: CRC press, 2nd ed., March, 2002.

Chapter 5

WIRELESS PASSIVE SAW IDENTIFICATION MARKS AND SENSORS

- Surface Acoustic Wave Devices -

Elke Mackensen and Leonhard Reindl

Laboratory for Electrical Instrumentation, Institute of Microsystem Technology, Albert-Ludwigs-University Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany

Abstract: Surface acoustic wave (SAW) devices can be turned into identification and sensor elements, so called SAW transponders. Identification elements transmit an individual ID number and sensor elements measure physical quantities such as temperature, pressure, torque, acceleration, humidity, etc. The SAW devices do not require any power supply and may be accessed wirelessly.

The complete sensor system consists of such a SAW transponder and a local radar transceiver (reader unit). An RF burst transmitted by the radar transceiver is received by antenna of the passive SAW transponder. The passive transponder responses with an RF signal, like a radar echo, which can be received by the front-end of the local transceiver. Amplitude, frequency, phase and time of arrival of this RF response signal transmit information about the SAW reflection and propagation mechanisms which in many cases can be directly attributed to the sensor effect for a certain measurement value. Due to the high delay time of the SAW transponder in the order of a few microseconds, usually no intersymbol interferences due to environmental reflections occur when the system operates in harsh indoor/outdoor environments.

This chapter first surveys about classical wired SAW sensors. Then the functional principles of such wireless SAW transponder systems and the reader systems will be pointed out. Furthermore the design requirements for SAW transponders, such as identification tags and radio sensors, several application examples and their state-of-the-art performances are presented.

Keywords: Passive SAW transponder, surface acoustic wave, local radar transceiver, wireless measurement, sensor, identification, temperature, pressure, torque, acceleration, tire-road friction, magnetic field, water content of soil

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 155-202.

1. INTRODUCTION

Today there are a lot of applications in which sensors must measure logistical, physical, chemical, or biological quantities such as identification, temperature, pressure, torque, acceleration, humidity, etc. After the measurement the obtained data must be further processed. In many applications, a wired connection between the sensor and the data processing system is not applicable. The sensor may be in a remote room or a process chamber, pressure or radiation can inhibit a direct link. In this case a radio sensor system offers a solution. Apart from active sensor units [63], such sensor systems can be realized by combining radar techniques and wireless surface acoustic wave (SAW) devices. Compared with conventional telemetry systems the main features are a large readout distance of several meters, and that SAW devices do not require a battery or any other power supply. Wireless passive SAW sensors and identification marks can be advantageously placed on moving or rotating parts and in hazardous environments, as contaminated areas or high-voltage plants. They can also be used for contact-less measurement in high-vacuum process chambers, under concrete, in extreme heat, or strong radioactive environment, where the use of conventional sensors is impossible, dangerous or expensive. SAW sensors can also be employed to increase comfort or safety at home, e.g. with personal switches or SAW based security systems.

This chapter gives an overview on surface acoustic waves and the operating principle of classical wired SAWs. In section 2 the functional principle of wireless SAW transponder systems and the radio request or reader systems will be discussed. Section 3 and 4 describes the design requirements for SAW transponders, such as identification tags or radio sensors. Finally section 5 presents several application examples of such sensor systems and their state-of-the-art performances. A short conclusion with a synoptical table completes this chapter.

1.1 Surface Acoustic Waves (SAW)

It was Lord Rayleigh (John William Strutt) at the end of the 19th century who treated in theory the problem of acoustic wave propagation along the surface of a semi-infinite isotropic medium [64]. The physics are mathematically treated by a boundary value problem. One solution, now called the Rayleigh wave (Figure 1), describes a wave whose amplitude of the

particle displacement decreases exponentially with depth, resulting in hardly any motion in the substrate at depths greater than about one acoustic wavelength λ. This means, the wave is bound to the surface. The particles near the surface move along elliptic trajectories in the sagittal plane, which is the plane containing the surface normal and the propagation direction. The phase velocity v_{SAW} of the wave is independent of frequency (no dispersion), and smaller than that of the slowest bulk mode. Therefore, on a plane surface no radiation loss into the substrate occurs, because no phase-match condition between the Rayleigh wave and a bulk mode is satisfied. The velocity v_{SAW} is in the order of 3000 m/s. At the same frequency, the acoustic wavelength is 10^{-5} times smaller than the corresponding electromagnetic wavelength. Therefore several wavelengths can be stored in a small piece of a chip, thus allowing very steep filter characteristics.

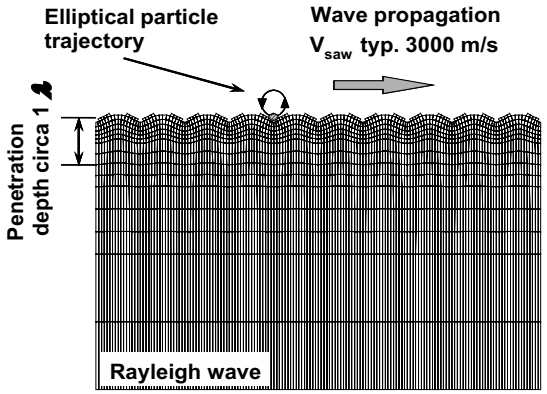


Figure 1. Particle Displacement of a Rayleigh Wave, Propagating Along the Surface of a Semi-infinite Isotropic Medium.

On piezoelectric substrates, there is a very convenient way for excitation and detection of an acoustic wave: the use of a so-called interdigital transducer, or IDT (Figure 2) [1].

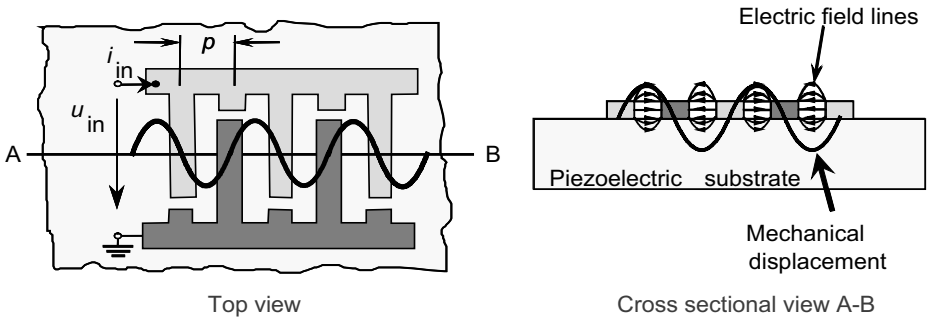


Figure 2. Top View and Cross Section of an Interdigital Transducer (IDT).

In its simplest form, an IDT consists of a series of parallel planar metal electrodes which are periodically spaced on the plain-polished surface of a piezoelectric substrate and alternately connected to each other via two bus bars. Typically, aluminium is used for the metallization. If a DC voltage is applied to the two bus-bars, then, due to the converse piezo effect, the electrical field stimulates an elastic deformation. If an AC voltage is applied to the bus bars, a harmonic deformation is generated. This, in turn, leads to the excitation of elastic surface waves, radiating out of the transducer in both directions normal to the electrodes. By reciprocity, the detection process works in the reverse: an acoustic wave entering an IDT generates an electrical charge distribution and therefore an electrical RF output signal between the two electrode combs.

For an IDT, the width and spacing of the electrodes and strips, together with the phase velocity v_{SAW} determine the operating frequency *f*. The sum of the electrode width and spacing gives the geometrical period *p* (pitch) of the IDT. In the simplest transducer, the pitch *p* is constant all over IDT. The condition for resonance is given by a SAW wavelength λ of:

$$\lambda = 2 \cdot p \tag{1}$$

The operating frequency *f* of the IDT thereby results to:

$$f = \frac{v_{SAW}}{\lambda} = \frac{v_{SAW}}{2 \cdot p} \tag{2}$$

The number N of pitches of the transducer determines its length in the time domain, which in turn results in the bandwidth B of operation of:

$$B \approx \frac{v_{SAW}}{2 \cdot N \cdot p} \tag{3}$$

Figure 3 shows a SEM photo of two SAW pulses, which were emitted by an IDT.

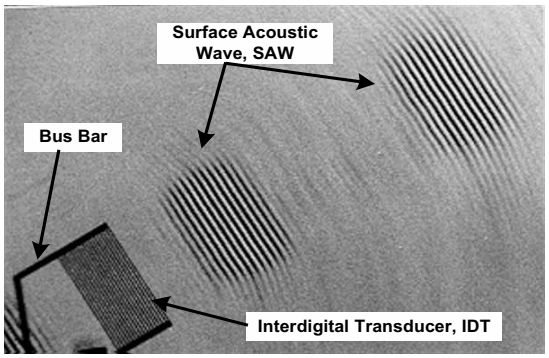


Figure 3. SEM-photo of an IDT and two SAW Pulses (Photo by Siemens AG).

Table 1 gives an overview of some important materials used in the SAW industry and their corresponding SAW parameters. The most used materials are Quartz (SiO_2), Lithium Niobate ($LiNbO_3$) and Lithium Tantalite ($LiTaO_3$). Quartz crystal cuts show excellent temperature stability, which is given by the first order temperature coefficient of delay (TCD). However the coupling coefficient k^2 of Quartz is low. On some cuts, the first order temperature coefficient of delay vanish, only a small quadratic dependence remain. The coupling coefficient k^2 is a measure of the efficiency to convert an applied microwave signal into mechanical energy associated with the acoustic wave. A high coupling coefficient is advantageous for good acoustic reflection. Thus the low coupling coefficient of Quartz limits the relative bandwidth, which can be electrically matched without losses to some percent. The coupling coefficient k^2 of $LiNbO_3$ is much larger than that of quartz, allowing medium loss devices for wideband applications with relative bandwidths up to some 10 %. $LiTaO_3$ shows extreme high coupling coefficients, which makes these

cuts highly suitable for low loss and wide band applications, as found in the RF stages of mobile phones. $LiNbO_3$ and $LiTaO_3$ show a temperature coefficient between 18 ppm/°C and 94 ppm/°C, thus these materials are very well suited for temperature sensors. The acoustic loss of all materials is nearly negligible at 434 MHz, but at 2.45 GHz, the loss of SAWs on Quartz or $LiTaO_3$ is extremely high.

Table 1. Properties of some frequently used SAW substrate material cuts. "Orientation Cut" denotes the crystal orientation of the substrate surface normal, "Orientation Prop" gives the crystal orientation of the wave-propagation direction.

Material	Orientation		V_{SAW}	k^2	TCD	Loss (dB/µs)	
	Cut	Prop			(ppm/°C)	433 MHz	2.45 GHz
Quartz	ST	X	3158	0.1	0	0.75	18.6
	37°rotY	90°rotX	5094	≈ 0.1	0	-	-
LiNbO3	Y	Z	3488	4.1	94	0.25	5.8
	41°rotY	X	4750	15.8	69	-	-
	128°rotY	X	3980	5.5	75	0.27	5.2
LiTaO3	36°rotY	X	4220	≈ 6.6	30	1.35	20.9
	X	112°rotY	3301	0.88	18	-	-

1.2 Design of SAW Devices

The design of a SAW device is based on signal theory, network theory, and field theory [58]. Impulse response modeling allows for a first order design since the impulse response *h(t)* of the IDT is directly related to the overlap and spacing of the electrodes as a function of position. The frequency response *H(f)*, which is given by the Fourier transform of the impulse response *h(t)*, can therefore be computed straightforward from the electrode pattern. This procedure can also be reversed and used for transferring an aimed transfer function in an appropriate electrode pattern for a filter synthesis, which usually incorporates some weighting techniques.

Signal-theory guidelines, however, are not capable of accounting for many microwave and microacoustic effects that have to be considered in accurate SAW design procedures. These so-called second-order effects include microwave reflections at strip and electrode edges, the effects due to longitudinal and transverse charge distribution in electrodes, mechanical loading effects (e.g. mass loading, stress loading, topographical loading at strip edges), the microacoustic attenuation and dispersion phenomena, the diffraction and refraction of SAW waves, the beam steering of SAW waves

due to the anisotropy of the substrate, the generation of spurious bulk-acoustic waves (BAWs) in the IDTs, the conversion of SAW waves into BAWs, ohmic losses of the metal strips, and the direct electromagnetic feed-through between the in- and out-coupling IDTs. The quantitative characterization of these second-order effects is done by applying field-theory approaches [59], device measurement, or a combination of both [60].

Usually, the results obtained using such experimental or theoretical methods are stored in databases to feed them into the final design procedure, which is based on network-theory techniques. The SAW structures are subdivided into substructures, which are themselves furthermore subdivided into basic two-port cells (with two acoustic ports) or three-port cells (with two acoustic ports and one electric port). A typical basic cell, for example, may be a single electrode or even a smaller portion of an electrode. The latter, extremely detailed subdivision into cells may be necessary if the charge distribution on the electrodes or edge effects have to be modeled very accurately. The electric and acoustic ports of the basic cells are then recursively combined using network-theory algorithms. Popular techniques used for cell modeling include various lumped-element methods, such as equivalent circuit modeling techniques, on the physical effects oriented methods like P-matrix formulations or coupling-of-modes techniques [65]. As a result of the sophisticated microacoustic design, the electric terminal behavior of a complete SAW device can be analyzed and optimized, also considering microwave parasitic and electrical matching structures.

1.3 SAW Fabrication

SAW manufacturing has been developed from the technology of integrated circuits [61]. Usually, SAW companies apply optical exposure, by which sub-micron patterning is already achievable. Line widths of less than 0.3 microns are possible [62]. Wafer discs, with diameters of up to 4 inches, are processed on a production scale. In contrast to semiconductor processes, only one single metallization layer has to be manufactured, but with a much higher demand for accurate line-width and thickness. In order to generating the metallization structure, etching or lift-off techniques are applied (Figure 4). The required pattern is repeated many times on the mask, so that many devices can be exposed simultaneously.

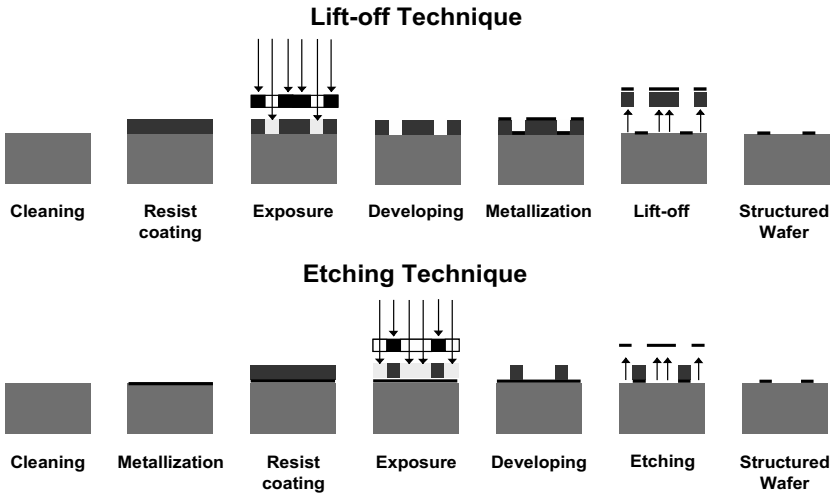


Figure 4. Process Flow for the Fabrication of SAW Devices.

A major issue is the minimum line width obtainable, because it determines the maximum operating frequency. Very fine lines can be obtained by X-ray or electron-beam exposure; however, up until now this has not been a useful commercial variant. As is seen from Figure 5, high resolution and good edge quality down to a pitch of 0.6 μm is routinely achieved using optical exposure.

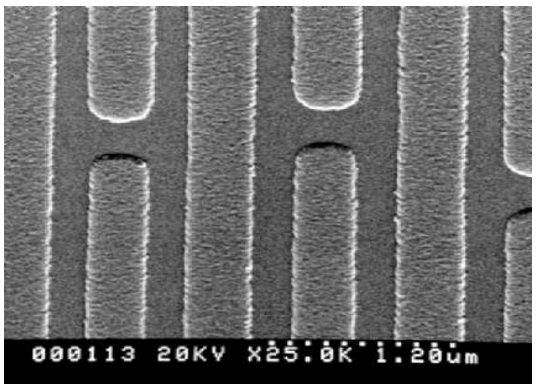


Figure 5. SEM Photograph of a Part of an IDT with Periodicity of 0.6 μm and Linewidth of 0.4 μm.

1.4 Classical SAW Based Band Pass Filter

Figure 6 depicts a typical classical SAW filter. It consists of one or several transducers (and sometimes some reflectors) on a piezoelectric substrate, like quartz, $LiNbO_3$, or $LiTaO_3$, which is die-bonded into a housing and wired to pins.

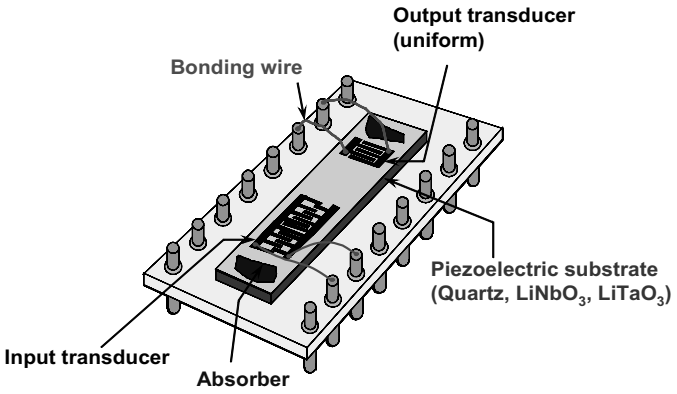


Figure 6. Schematic Drawing of a SAW Filter.

SAW technology has been applied in electronic analogue signal processing over the past 30 years. The main application for SAW filters is in the field of radio transmission, like portable or wireless phones, TV sets, satellite receivers, or wireless systems. Figure 7 depicts a block diagram of a typical transceiver in portable phones. In the lower part the transmitter section and in the upper part the receiver section is shown. All those gray blocks show parts of the circuitry, where SAW filters might be used. Thus in a typical portable phone 3 to 5 SAW filters are wired.

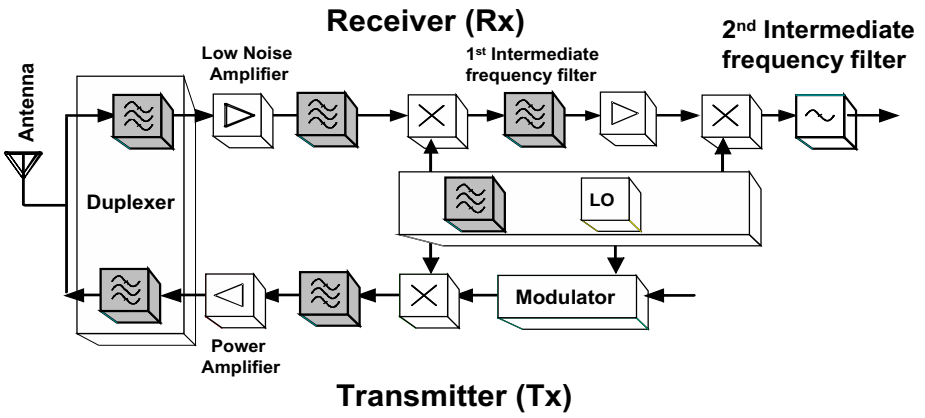


Figure 7. A Schematic of a Typical Cellular Front-end.

1.5 Sensor Effects in SAW Devices

In most sensing applications, the quantity to be monitored both strains the SAW chip and also changes the SAW velocity by influencing the elastic constants of the substrate crystal. Other quantities such as gas or vapor concentration, voltage or magnetic field can be measured indirectly by coating the chip with a layer whose acoustical properties depending on the observed quantity; the SAW then reacts to changes in these properties. In either case, a change Δy of the quantity y_0 results in a variation of the delay time τ, the corresponding phase φ and the center frequency f of the SAW device. The linear terms of a Taylor series expansion usually describe the situation well enough:

$$\Delta\tau(y_0) = \left.\frac{\partial\tau}{\partial y}\right|_{y=y_0} \Delta y =: \tau(y_0) S_y^{\tau} \Delta y \tag{4}$$

$$\Delta\varphi(y_0) = \left.\frac{\partial\varphi}{\partial y}\right|_{y=y_0} \Delta y =: \varphi(y_0) S_y^{\varphi} \Delta y \tag{5}$$

$$\Delta f(y_0) = \left.\frac{\partial f}{\partial y}\right|_{y=y_0} \Delta y =: -f(y_0) S_y^{f} \Delta y \tag{6}$$

In the equations S_y denotes above the sensitivity with respect to y. An example for this sensitivity is the well-known linear temperature coefficient of delay:

$$TCD = S_T = \frac{1}{\tau} \cdot \frac{\partial \tau}{\partial T} \tag{7}$$

2. OPERATING PRINCIPLES OF WIRELESS SAW SENSOR SYSTEMS AND OF THE REQUEST UNIT

In the following section the functional principles of wireless SAW transponder systems and several circuitries used for reader systems will be discussed.

2.1 Operating Principle of Wireless SAW Identification or Sensor Systems

Figure 8 shows the operating principle of a wireless SAW sensor system in detail. A radio frequency electromagnetic (RF) read-out signal is transmitted by a local radar transceiver (TRx). A small portion of this signal is picked up by the antenna of the passive SAW transponder where an IDT, connected to the antenna, converts the received signal into a SAW. The micro-acoustic wave propagates towards reflectors distributed in a characteristic barcode-like pattern and is partially reflected at each reflector. The micro-acoustic wave packets returning to the IDT are reconverted into electrical signals and retransmitted to the radar TRx unit by the transponder antenna. This response contains information about the number and geometries of the coding structures as well as the propagation and reflection properties of the SAW. Its evaluation in the radar reader unit will allow for the extraction of the desired information, which is, e.g. the sensor effect for a certain measurand or a specific ID number.

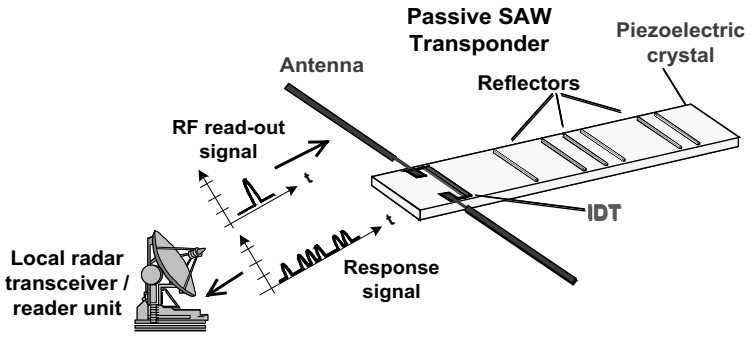


Figure 8. Schematic of a SAW-based radio-Link System, which uses a Reflective Delay Line.

Due to the low velocity of SAWs, long delay times in the range of some microseconds can be achieved using rather small SAW chips. Therefore, at VHF/UHF frequencies, environmental echoes caused by electromagnetic multipath propagation phenomena will have already safely faded away when the sensor response arrives at the radar TRx. Hence, the sensor response can easily be separated from environmental echoes in the time-domain. This fact incorporates a great advantage of wireless SAW-based sensor systems compared to other radio link systems. Figure 9 shows a typical response signal of a SAW ID tag together with the request impulse and some environmental echoes.

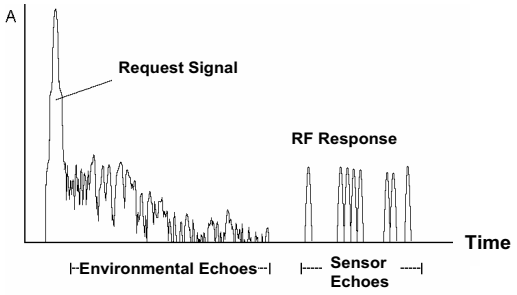


Figure 9. Request Pulse, Environmental Echoes and RF Response of a SAW ID Tag.

If a wirelessly SAW-based sensor technique is also used for measuring distances, linear and angular velocities of targets, then this time separation of most environmental echoes is a great advantage in comparison to conventional radar measurement techniques. The measurement accuracy of this radar values lies within 1 % of the electromagnetic free-space wavelength when attained by using a SAW transponder.

Passive SAW transponders do not require any power supply. In the VHF/UHF frequency range, the insertion attenuation A_{SAW} of SAW transponders is in the order of 20-60 dB ([10],[11],[12],[23],[24],[30]), depending on the operating frequency, the substrate material, and the number of reflectors. Furthermore, the achievable access rate is up to 10^5/s. The latter fact allows for communication with fast moving objects or vehicles.

2.2 The Request Unit

The reader units of wireless SAW sensor system applications resemble those used in traditional radar systems. In Europe, three frequency bands suitable for SAW devices are allocated to unlicensed low power devices (LPDs) such as industrial, scientific, or medical (ISM) appliances: 433.07…434.77 MHz, 868...870 MHz, and 2.4…2.483 GHz. Regardless of the type of passive radio sensor used, a wireless one-port response measurement has to be performed with a time division between the read-out signal and the sensor response. As is the case with radar systems, the receiver (Rx) usually is located nearby the transmitter (Tx) so that coherent detection is feasible. The actual measurement can be performed in either time or frequency domain.

2.2.1 Time Domain Sampling

Methods like pulse or pulse compression radar are applicable for the time domain sampling. Figure 10 depicts the functional principle of the time domain sampling. A short pulse or wave packet *s(t)* is transmitted by the reader unit. Each radio request signal generates a corresponding response signal *r(t)*. The resulting delay time of the response signal carries the information of the passive sensor. The read-out signal spans the total system bandwidth *B* at once. Therefore, to avoid ISI (intersymbol interference), a fast sampling has to be performed in the receiver unit with at least twice of the bandwidth *B*. Time-domain sampling is a single-scan measurement method,

whereby the read-out repetition can be up to 10^{-5} s. Hence, this method is especially well-suited for reading fast changing measurement values or for reading out of fast moving objects.

For non-spread spectrum signals, the duration of one read-out signal T_{TDS} is low compared to the read-out repetition. With a restricted peak power P_0, the energy content E_{TDS} *(=* P_0/B*)* of one read-out signal and therefore also the maximum read out distance will be small. Using pulse compression methods, E_{TDS} can be increased by the time-bandwidth product *TB* of the pulse compression system, thus enlarging the read-out distance. The maximum duration for T_{TDS}, however, is limited to the basic delay of the SAW sensor (some microseconds). Using pulse compression techniques a gain of up to 12 dB has been achieved, which doubles the read-out distance [20].

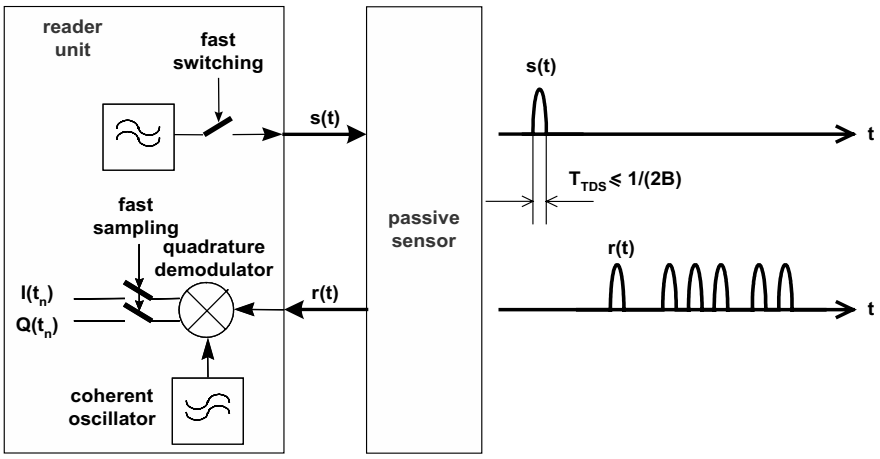


Figure 10. Time Domain Sampling Technique.

2.2.2 Frequency Domain Sampling

Frequency domain sampling can be performed using a network analyzer or a frequency-modulated continuous wave (FMCW) radar architecture ([30], [31]). Figure 11 depicts an example of a radar reader unit realized in the frequency domain. The total bandwidth is scanned in M steps in the frequency domain. The bandwidth of one step B_{res} can be rather low to achieve a high resolution. The Tx pulse can last for a relatively long time period T_{FDS}. If T_{FDS}

exceeds the basic delay of the SAW sensor, the duplexer switch at the Trx front-end has to be replaced by a circulator and a low noise amplifier with a high dynamic range. Frequency domain sampling techniques enhance the SNR (E_b/N_0) in the receiver at the expense of a lower time resolution of the measurement value. Magnitude and phase of a narrow-band signal have to be detected in the receiver, which lowers the complexity and cost of the sampling and signal processing unit. Similar to time-domain sampling systems, the RF signals can be derived from a single oscillator for a coherent detection.

Frequency domain sampling is a multi-scan measurement technique. To achieve the information of *M* points in time-domain, *M* frequencies have to be scanned. The information on the sensor signal in the time domain is received by a frequency to time transformation, e.g. FFT or other more sophisticated algorithms [27]. The total measurement time takes more than M times the minimum measurement cycle of a single time-domain sampling. Frequency domain sampling is therefore well suited for slow, high precision and long-distance measurements. In order to eliminate the transmitted signal and all environmental echoes, also in a frequency domain sampling technique the first 1-2 µs of the sensor response can be suppressed in the time-domain [29].

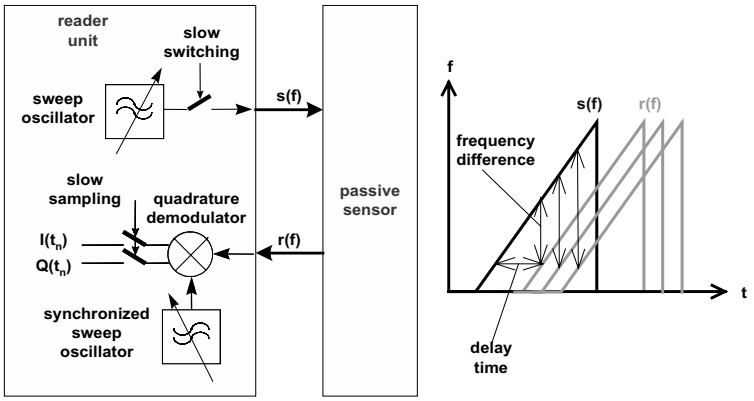


Figure 11. Frequency Domain Sampling Technique.

For the majority of applications, small and low cost local radar TRx (reader unit) and small antennas are necessary. Dipole, patch or loop type antennas are usually used for the transponders. Figure 12 shows the

implementation of a radar unit operating in the 434 MHz band in an enclosure of 7 cm x 5 cm x 2.6 cm.

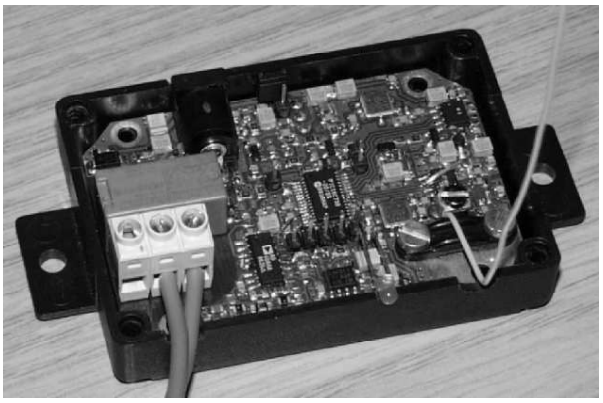


Figure 12. Example of a Radar Unit Operating in the 434 MHz Band (7 cm x 5 cm x 2.6 cm).

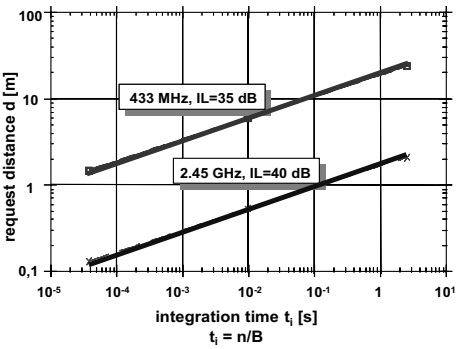
2.3 Request Distance

Very important is the achievable measurement distance between the reader unit and the SAW-device and how fast the measurement can be done. For free-space propagation conditions of electromagnetic waves the well-known radar equation predicts that the signal level received by the reader unit decreases with the fourth power of the distance *d* from the SAW transponder:

$$d = \frac{\lambda}{4\pi}\sqrt[4]{\frac{P_T \cdot G_T^2 \cdot G_R^2 \cdot n}{k \cdot T_0 \cdot B \cdot F \cdot \frac{S}{N} \cdot IL}} = \frac{\lambda}{4\pi}\sqrt[4]{\frac{P_T \cdot G_T^2 \cdot G_R^2 \cdot t_i}{k \cdot T_0 \cdot F \cdot \frac{S}{N} \cdot IL}} \tag{8}$$

Based on the data given in Figure 12, one obtains a maximum transceiver-transponder distance *d* of only 25 to 30 cm for a single read out cycle. In RF-shielded metallic process chambers the distances are larger because the Tx power level can be enhanced and the operating frequency can be chosen arbitrarily. If we allow a lowering of the time resolution t_i, the sampled signal can be averaged over several read out cycles. *n* predicts the number of averaging's. This technique however lowers the measurement system

bandwidth *B* but increases the readout distance *d*. An averaging factor n of 16 doubles the request distance. Using these techniques, request distances of 5 to 10 m in the lower and 1 to 2 m in the upper ISM band have been demonstrated (Figure 13).



Assumed values for estimation:

P_T:	radiated power	14 dBm
G_T:	transmitting antenna gain	0 dB
G_R:	receiving antenna gain	-3 dBi / 0 dBi
λ:	wavelength	70 cm / 12 cm
F:	receiver noise figure	12 dB
S/N:	signal to noise ratio for detection	20 dB
IL:	sensor insertion loss	35 dB / 40 dB
T_0:	antenna noise temperature	300 K
B:	Bandwith	
n:	numbers of averaging	

Figure 13. Estimation of the Achievable Measurement Distance Between the Reader Unit and the SAW-device.

2.4 Multiple Access of SAW Radio Sensors

Because SAW transponders are passive components without any active logic on chip, they cannot be addressed individually. To access more than one transponder a frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), space division multiple access (SDMA), or combinations of these must be implemented. For FDMA orthogonal frequency bands or sub-bands must be used for each individual transponder. This technique is feasible for several resonant transponders [21]. When implementing TDMA different time positions for each reflected signal has to be chosen in order to minimize ISI (intersymbol interference). Typically 10 TDMA chips, each with 3-4 reflectors, can be setup. Figure 14 shows an example for TDMA. Using CDMA and the corresponding signal processing, again approximately 10 code orthogonal transponders can be addressed ([41], [42], [43]). For SDMA techniques a space separation of the transponders has to be ensured. Due to the field attenuation of Rx power with r^4, the near-far problem limits the multiple access of passive radio transponders.

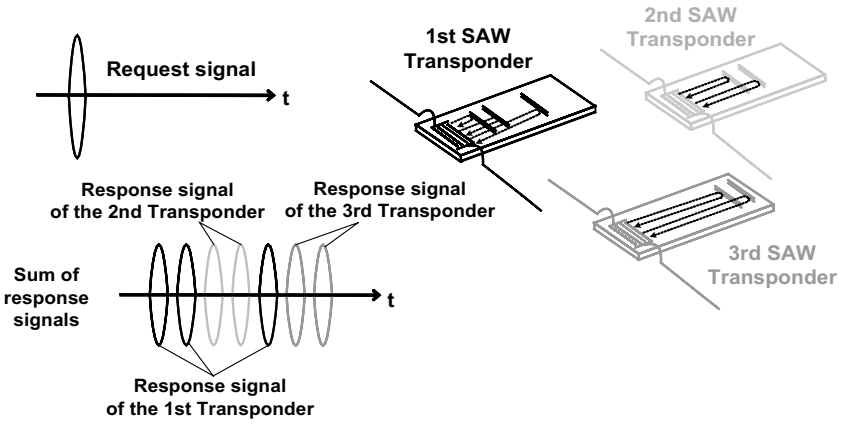


Figure 14. TDMA Example with Three Different Transponders.

3. SAW IDENTIFICATION TAGS

In this section the design requirements of SAW identification tags is presented. For ID tags, a large amount of individual symbols are required. A SAW ID tag usually incorporates a coupling IDT and a coding area. There are two ways of designing the coding area: either reflectors are used or, alternatively, several coding transducers are wired together with a common bus-bar.

3.1 Transducer Coded Delay Lines

Figure 15 depicts an example of a bus-bar connected transducer tag. At the left side you can see the coupling IDT. The coding area at the right side consists of several transducers connected via bus-bars.

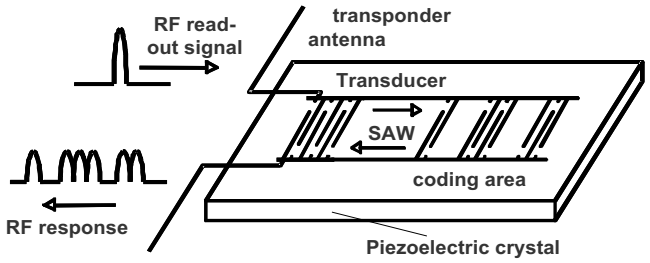


Figure 15. Schematic Layout of a SAW ID Tag with Several Transducers Wired Together to a Common Bus Bar.

The minimum length of this common bus-bar is given by the number of symbols N multiplied by the minimum distance ΔT between two symbols, which is determined by the bandwidth B of the system. The wavelength λ_{el} of an electric signal on $LiNbO_3$ is reduced by a factor of approximately 10 compared to free space, due to the high dielectric constant of this material. If the length of the bus-bar becomes larger than 10 % of λ_{el}, the quasi-static approximation for the electrical connection is no longer valid and electric the wave-guide effect on the bus-bar must be taken into account. If we demand, that the length of the bus-bar must remain smaller than 10 % of λ_{el}, and, if we take into account, that the minimum resolution is given by the system bandwidth B, we get an upper limit N_{max} for the number of symbols N, which can be lined up on $LiNbO_3$ in this technique, with

$$N_{\max} \le \frac{0.1\lambda_{el}}{v_{SAW}/B} = \frac{0.1 \cdot B \cdot 10^4 \cdot \lambda_{SAW}}{v_{SAW}} = \frac{10^3 \cdot B \cdot v_{SAW}}{v_{SAW} \cdot f_0} = 10^3 \frac{B}{f_0} \quad (9)$$

Using a relative bandwidth B/f_0 of 2%, less than 20 symbols are possible with this technique. For a identification system incorporating only a few symbols, or, if a substrate with a low dielectric constant like quartz is used, a tag of this kind may be a promising concept.

3.2 Reflective Delay Lines

Most SAW tags and passive sensors are designed using a reflective delay line, like the one sketched in Figure 16. The surface acoustic wave in a reflective delay line propagates towards reflectors distributed in a characteristic barcode-like pattern and is partially reflected at each reflector. The usage of reflectors halves the chip size compared to the non-reflective arrangement of Figure 15 ([17], [20], [22], [23]). Furthermore, the electrical impedance of the transducers remains unattached by the number of coding elements and the actual code.

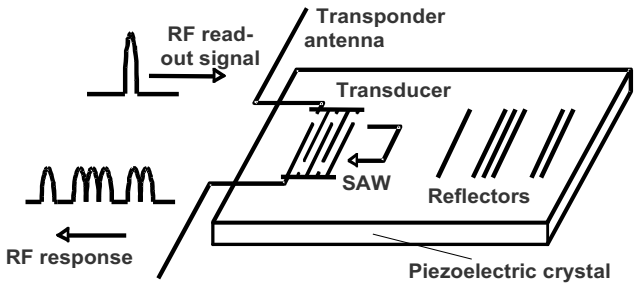


Figure 16. Schematic Layout of a Reflective Delay Line.

3.3 Coding Schemes

The coding of a SAW ID tag depends on the applied modulation technique and is obtained by a specific coding arrangement. In the simplest case, this coding is done by a binary amplitude shift keying ASK (on/off). Each predetermined possible symbol position is set either by a transducer or reflector (on-bit) or not (off-bit). By using an on/off-keying with 32 reflectors, 2^{32} (more than 4 billion) different tags can be coded. ID-tags with ASK modulation are already in use, e.g., in German subway systems ([23], [35]). The system operates in the ISM band at 2.45 GHz. Figure17 shows the layout and a measurement of an ID tag using a reflective delay line structure and a ASK modulation scheme.

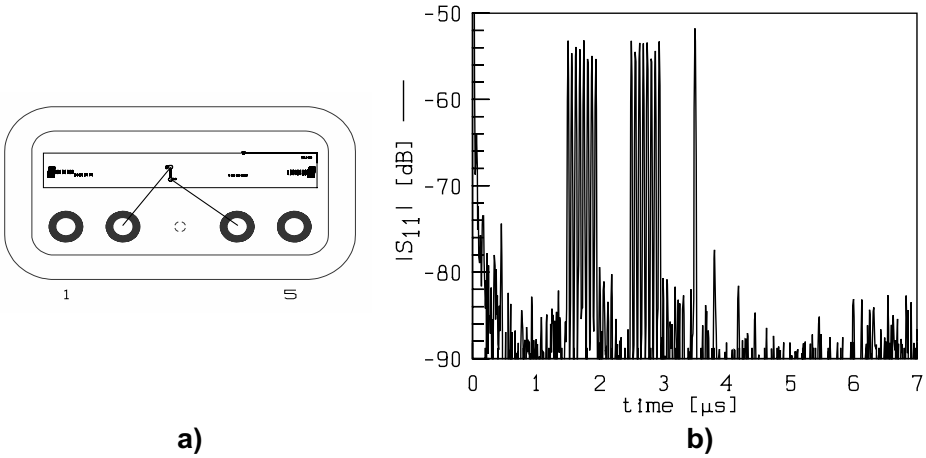


Figure 17. a) Layout of a SAW ID Tag Comprising 33 Reflectors in 4 Tracks
b) Measurement of the SAW ID Tag with 8 "ON", 8 "OFF", 8 "ON“, 8 "OFF", and 1 "ON".

The ASK on/off keying is outperformed by other modulation techniques such as phase shift keying PSK. Using a PSK modulation the phases of the time response are evaluated and not their amplitudes. A binary phase shift keying BPSK modulation obtains the same bit error rate with a 6 dB lower signal-to-noise ratio compared to an ASK modulation, thus enhancing the maximum readout distance. Higher-order PSK like quadrature PSK (QPSK) use fewer symbols (reflectors) and therefore need less chip size and obtain a lower insertion attenuation, but a higher signal-to-noise ratio is required for the detection. The requirements for a SAW transponder used in a phase modulation system, however, become more stringent, because all influences which change the delay time of a reflector, e.g., the SAW velocity or the accuracy of the fabrication process, must be controlled within a small fraction of one single SAW wavelength.

A good choice for the modulation technique used in SAW transponders might be a pulse position modulation (Figure 18), which achieves the code-density of a higher phase-modulation and avoids their stringent sensitivity on small inaccuracies during the fabrication process [36]. In a pulse position modulation scheme the symbol is set at one out of several time slots. The accuracy in the time resolution of one target is much higher than the time separation of two targets in most radar systems. Therefore, the time slots can be arranged much denser, compared to the symbol separation in a ASK

modulation. If each reflector is set in one out of 2^n time slots, each symbol codes for n bit.

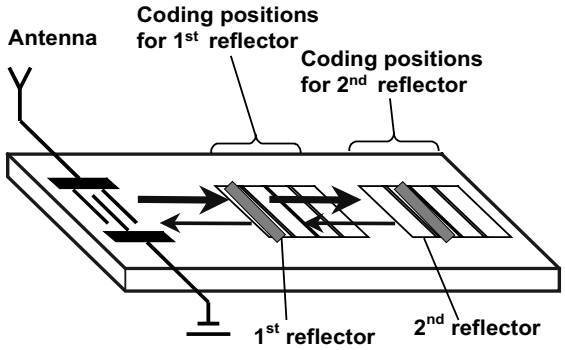


Figure 18. Pulse Position Coding Schema [36].

3.4 Design of Transducers

The uniform transducer, shown in Figure 19 *a*, consists of electrodes with a pitch of one quarter of a wavelength. Such a transducer offers the highest transduction for a given time length, and therefore for a given bandwith. The drawback of this structure is a high acoustic reflection which decreases the achievable signal to noise ratio. If all coding tracks are placed on the same side of the coupling IDT the best choice for the transducer layout would be a single phase unidirectional transducer (SPUDT), like sketched in Figure 19 *c*. This transducer stimulates SAWs mostly in one direction. A problem of all transducers occurs with the increasing size of the transducer. In this case it is advantageous to arrange different tracks on both sides, which can be done using e.g. a split finger transducer, sketched in Figure 19 *b*.

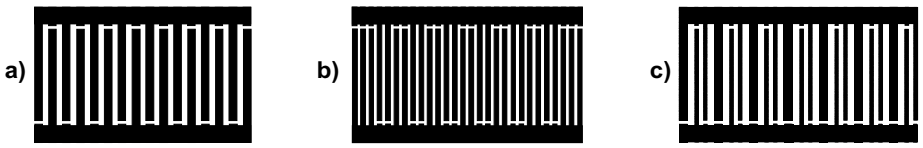


Figure 19. Layout of Transducers for SAW ID Tags and Radio Requestable Sensors:
a) Uniform Transducer b) Split Finger Transducer c) Single Phase Unidirectional Transducer.

The required relative bandwidth mostly restricts the geometrical length and thus the number of overlaps of the coupling transducers. Thus parasitic effects like ohmic losses and capacities of the pads become relevant, especially on transducers operated at higher harmonics. In most cases the metallization height is kept relatively low in order to minimize bulkwave generation in the coding tracks. Sometimes the bonding pads need a thicker metallization to ensure a faultless bonding process.

3.5 Design of Reflectors

Typical reflectors for SAWs are shown in Figure 20. Several types as the λ/4-stripes open or shorted, multistrip-couplers or chevron type reflectors are intensively discussed in the literature. All reflector types, however, may produce problems in the uniformity of the tag, which is the highest tolerable variation between all ON symbols. Each reflector reduces the transmitting wave, mainly due to bulk wave conversion. Thus, the amplitude of the incoming wave to a certain rear symbol position is a function of the number of proceeding ON symbols. This could generate strong code-dependent fluctuations in the uniformity of amplitude coded tags unless specific countermeasures are taken.

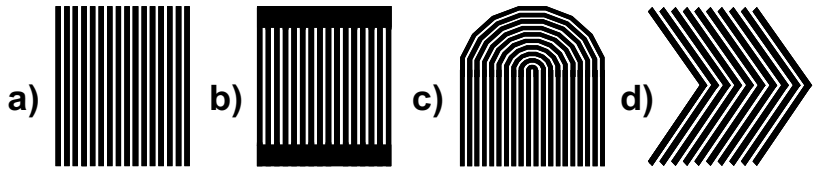


Figure 20. Layout of Typical Reflectors for SAW ID Tags and Sensors:
a) Uniform Open, b) Uniform Short, c) Multistrip and d) Chevron-Type

The code-dependency of the uniformity can be compensated if the magnitude of reflection of each reflector can be chosen individually, for example, by selecting the convenient reflector out of several feasible. A more feasible way, however, is to arrange a specific non-reflecting structure at each OFF position. The non-reflecting structure should show the same transmission loss like the corresponding ON structure of this position. Therefore no code dependence of the uniformity arises. For a non-reflecting structure, a shorted split finger transducer operating at the third harmonic can be used, for

example. In this structure no electrical reflection occurs, and the mechanical reflections of succeeding electrodes cancel out each other. The mechanical period, however, is large enough to allow for bulk wave conversion.

3.6 Multiple Reflections

If several symmetrical reflectors (as outlined in Figure 20 *a, b, c*), which results into reflections on both sides, are lined up in the same acoustic track, multiple reflections within the reflectors occur. The chevron type reflector is a non-symmetrical reflector, which prevent this multiple reflections. Triple reflections, like depicted in Figure 21, are more severe than 5th or even higher order reflections, because each individual reflection is small. In a system using an amplitude modulation technique the multiple reflections of the leading symbols can appear at the same time position where an OFF symbol is intended, thus debase the dynamic separation between ON and OFF symbols. By using other modulation techniques these multiple reflections can disturb the subsequent symbols and thus the bit error rate.

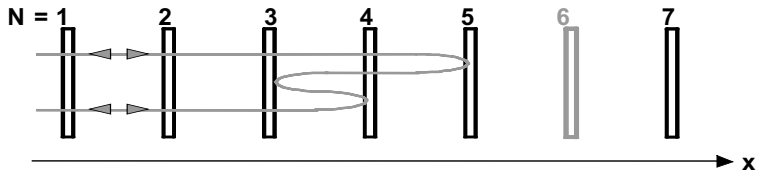


Figure 21. Example for a Triple Reflection if the Reflectors are Line up in the Same Acoustic Track.

As shown in Figure 21 multiple reflections occur if all reflectors are arranged in one track. So the idea is to arrange the reflectors in separate tracks, like this one sketched in Figure 22. But if each reflector is arranged in an individual track the problem of a high track-loss arises. Furthermore, either the aperture of the coupling transducer will become too large to stimulate a homogeneous SAW, or the aperture of each individual reflector will get very small. For systems with a low number of symbols, which is the case with SAW radio requestable sensors, a tag of this kind is a promising concept, since with this arrangement all multiple reflections are delayed beyond the time of interest and, furthermore, are very weak.

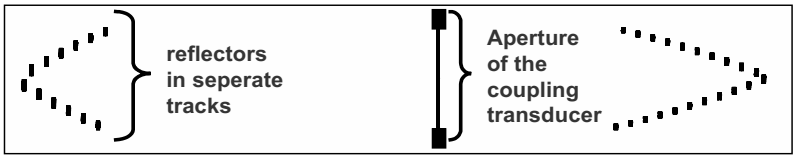


Figure 22. Example of an ID Tag Where Every Reflector is Arranged in a Separate Acoustic Track.

It was investigated how many reflectors in a track will lead to the best results. After a calculation of multiple reflections, one gets an equation on the maximum reflection for each reflector and thus an equation of the overall insertion loss of the tag. The results of these computations for a symmetrical reflector and a chevron type reflector are depicted in *Figure* 23. The design parameters are listed to the right of *Figure* 23.

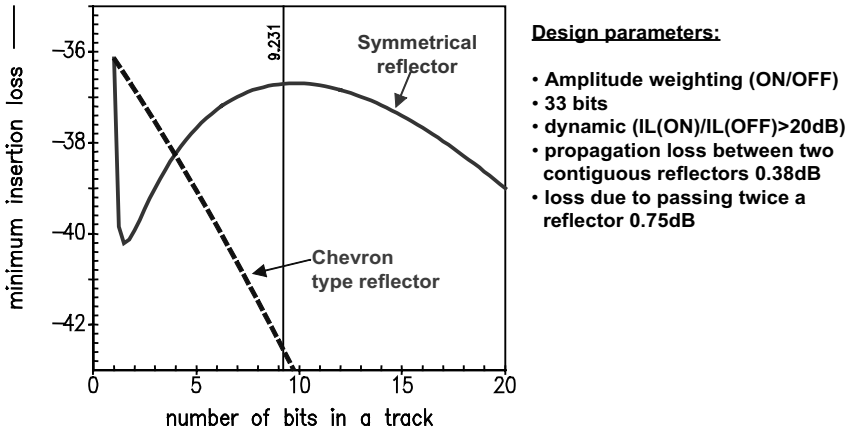


Figure 23. Evaluated Minimum Insertion Loss of an SAW ID Tag in Dependency of the Number of Reflectors in the Same Acoustical Track.

The solid line represents the computation for symmetric reflectors and the dashed line for non-symmetric ones (chevron type). For symmetric reflectors there are two optimum designs, one which consists of 4 tracks with 8 reflectors lined up in each track, and one design with each reflector in a

separate track. For a design with 2-4 reflectors lined up in each track an arrangement with chevron-type reflectors is best. More of those reflector types, however, cause higher insertion attenuation than other reflector types. A good choice for reflectors might be a full reflector together with 3-4 preceding chevron-type reflectors at the end of a track, and for the previous reflectors the use of symmetrical ones.

4. SAW RADIO REQUESTABLE SENSORS

In this section several approaches for the implementation of radio requestable sensors arel be presented.

4.1 Delay Lines

Most wireless SAW sensors are build up in a reflective delay line technique like the ID tags described in section 0. For tagging, a large quantity of reflectors is necessary, usually 32 or more. Using a delay line technique for sensing purposes, only three or four symbols are required for sensing purpose. *Figure* 24 shows a photo of a SAW radio sensor for temperature measurement, realized in a reflective delay line technique and sealed in a SMD package. On the left side the IDT can be seen and to the right there are 4 reflectors distributed on the chip.

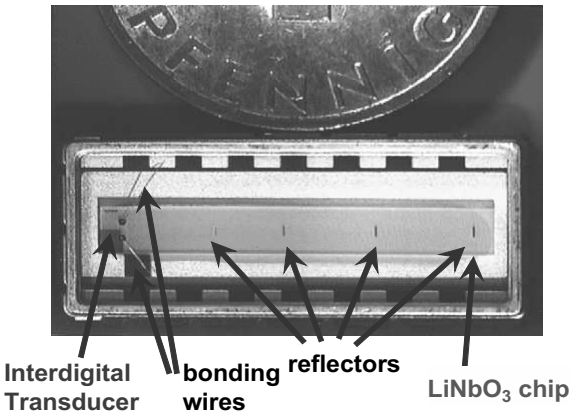


Figure 24. Photo of a SAW Radio Sensor for Temperature Realized in a Reflective Delay Line Technique.

The RF response of a delay line depends on the arrangement of the reflectors or coding transducers. Figure 25 shows the corresponding impulse response of the temperature sensor, both in a cartesian graph and in a polar graph. Each pulse corresponds to one of the reflectors. Due to multiple reflections at the chip edge some spurious signals follow the four useful pulses (Figure 25 *a*). The different phases between the pulses are pointed out in the polar graph of the impulse response (Figure 25 *b*).

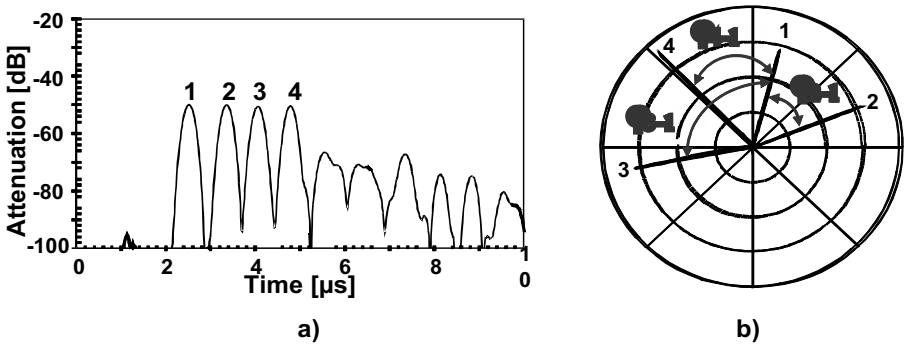


Figure 25. Impulse Response of a Delay Line Temperature Sensor a) Representation of the Amplitude in Cartesian Graph b) Representation of the Baseband in Polar Graph.

The sensor effect described by equation (4) causes the group delay time difference between two time response pulses #1 and #2 *($\tau_{2\text{-}1} = \tau_2 - \tau_1$)* to vary as:

$$\Delta\tau_{2-1} = \left[\tau_2(y_0)S^{\tau}_{y,2} - \tau_1(y_0)S^{\tau}_{y,1}\right]\Delta y \qquad (10)$$

In most sensing applications using a delay line, a differential arrangement is implemented and the change $\Delta\tau_{2\text{-}1}$ is evaluated. If, as usual, the frequency band used by the request signal is a subset of the operating frequency band of the sensor element, neither the center frequency nor the bandwidth of the sensor response is influenced by the sensor effect. Hence, all sensor responses can be demodulated by one fixed frequency, so that the time response pulses exhibit a constant phase. Consequently, the resolution is increased significantly because the phase differences $\Delta\varphi_{2\text{-}1}$ can be determined accurately in the time domain:

$$\Delta\varphi_{2-1} = 2\pi f_0 \Delta\tau_{2-1} =: S_y^{\varphi} \Delta y \tag{11}$$

Here, the term $2\pi f_0 \tau$ corresponds to the number of wave lengths stored in the relevant propagation path. With equation (10), the effective sensitivity of the reflective delay line becomes:

$$S_y^{\varphi} = 2\pi f_0 \left[\tau_2(y_0) S_{y,2}^{\tau} - \tau_1(y_0) S_{y,1}^{\tau} \right] \tag{12}$$

It should be noted that, according to equation (12) either the sensitivities or the delay times in the differential arrangement have to differ. This is achieved by a suitable selection of material and geometry.

4.2 Resonators

In a reflective delay line the acoustical path is used twice. But it is possible to achieve an acoustical path which is more frequently used. This aim can be achieved by placing the IDT between two reflective structures, forming a so called resonator ([21], [32]). Figure 26 shows the schematic layout of a SAW sensor using two 1-port resonators.

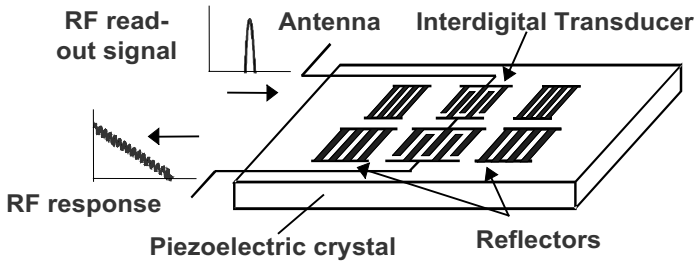


Figure 26. Schematic Layout of a SAW Radio Sensor Consisting of Two 1-port SAW Resonators.

A read out signal excites the resonators and the received energy is stored in the cavities of the resonators. After that, the resonator fades out and releases the stored energy in small portions on a time scale given by the decay time. The number of stored wave-lengths is given by the loaded quality factor Q, which also gives the decay time of the resonator (Figure 27 *a*). To eliminate the request signal and environmental echoes, the first 1 or 2 µs of the sensor response can be gated in the time domain. After a subsequent Fourier transform or another suitable algorithm, the frequency domain peaks are evaluated to retrieve the sensor information (Figure 27 *b*).

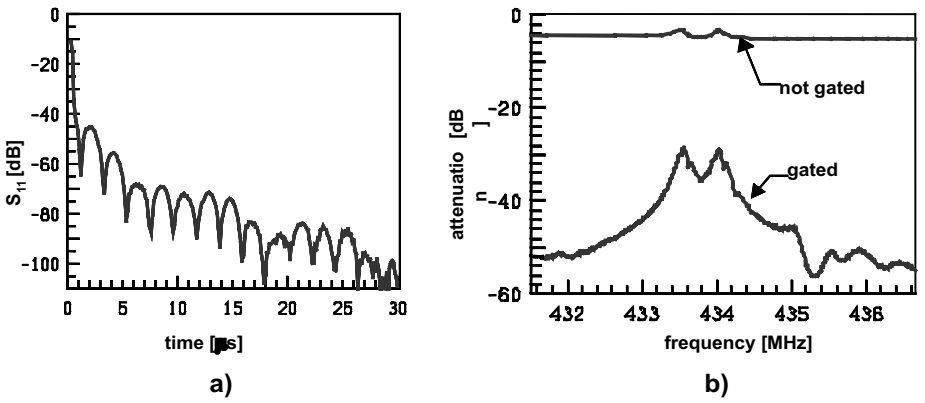


Figure 27. Measured Characteristics of a SAW Sensor Comprising Two 1-port Resonators: a) Time Response b) Frequency Response.

4.3 Dispersive Delay Lines

The sensitivity of the delay time of a SAW delay line transponder can be enhanced by a factor 10 to 20 by using dispersive, so called chirped reflectors ([22], [23], [37], [38]). Dispersive means frequency modulated. Figure 28 shows the schematic layout of a reflective delay line with two dispersive reflectors. One reflector has implemented an up-chirp FM law, and the other a down-chirp FM law.

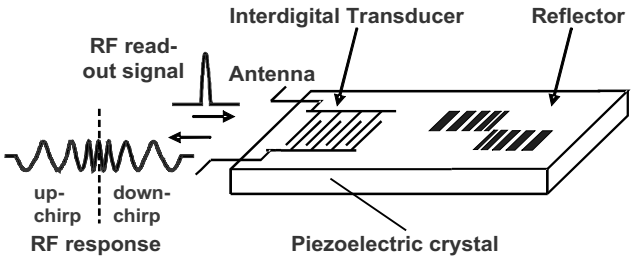


Figure 28. Schematic Layout of a Reflective Dispersive SAW Delay Line.

In a dispersive device, the group delay time τ of a transmitted signal depends on the frequency *f*. For a linear dispersion with bandwidth *B*, dispersion time *T* and centre frequency *f*, the group delay time τ can be calculated for all frequencies within the pass-band with:

$$\tau(f) = \tau_0 \pm \frac{T}{B} \cdot f \tag{13}$$

The "+" sign represents an up-chirp and the "-" represents a down-chirp law. A variation Δv_{SAW} of the SAW velocity v_{SAW} affects both the time τ_0 as well as the synchronous frequency *f*, resulting in a change of the group delay time τ. Therefore, we get an overall shift and a small tilt of the chirp function *f(t)*. This feature is similar to the impact of a Doppler shift to a FMCW radar system. Neglecting the small tilt, we obtain in first order:

$$\frac{\Delta\tau}{\tau_0} = \left[1 \mp \frac{T}{B}\frac{f_0}{\tau_0}\right] \cdot \frac{\Delta v}{v} \tag{14}$$

The expression in the brackets of equation (13) can be interpreted as an amplification factor for the sensitivity of the group delay, due to the usage of chirped devices. Figure 29 illustrates this amplification effect for the sensitivity of the delay time by using dispersive delay lines. The horizontal arrow gives the shift of the delay time to the area of the synchronous frequency and the vertical arrow the shift in the synchronous frequency by a

variation of the SAW velocity. The delay time to the new area with the correct synchronous frequency results due to both effects and the chirp rate.

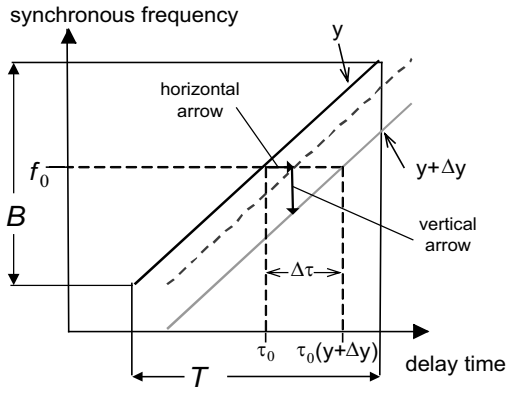


Figure 29. Illustration of the Amplification Effect for the Sensitivity of the Delay Time by Using Dispersive Delay Lines.

The sensitivity can be increased by a factor of 10 to 100. Even a change in the sign of the sensitivity is possible. Figure 30 shows a measurement of this effect. This amplification, however, is only effective to the group delay time. In the phase delay, we get the same dependence, as we would get by using a non-chirped device.

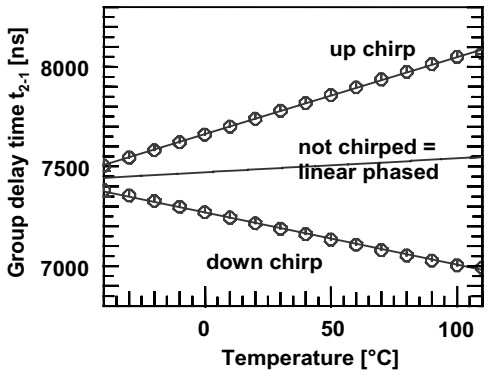


Figure 30. Measurement of the Amplification Effect for the Sensitivity of the Delay Time by Using Dispersive Delay Lines.

4.4 Combination with External Sensors

Also non-SAW sensors, or switches with varying impedance, can be addressed wirelessly when combined with a SAW transponder [11]. Hereby an external sensor is used as an electrical load of an IDT (Figure 31). The sensor quantity modulates the load impedance, what in turn changes the reflection behaviour of the IDT and thus the re-transmitted signal of the transponder ([11], [23], [39], [40]).

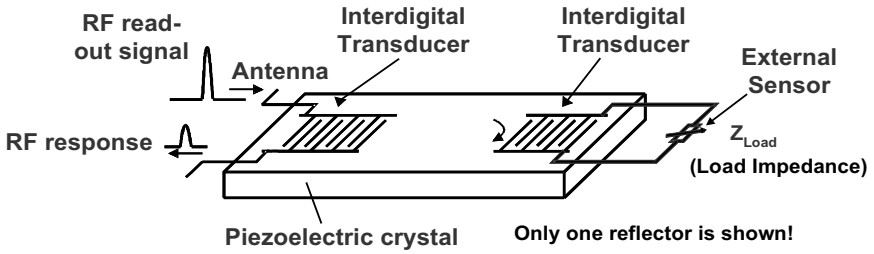


Figure 31. Schematic Layout of a Passive SAW Device Combined with an External Sensor.

The reflectivity of an interdigital transducer as a function of a complex termination impedance Z_{load} at its electrical port by given in the well known P-Matrix formalism [66] with P_{11}:

$$P_{11}(Z_{load}) = P_{11}^{sc} + \frac{P_{13}^2}{P_{33} + \frac{1}{Z_{load}}} \tag{15}$$

Figure 32 shows the acoustic reflectivity of a split-finger ($\lambda/8$) IDT as a function of its electrical load. To enhance the sensor effect a substrate with a high coupling coefficient should be used.

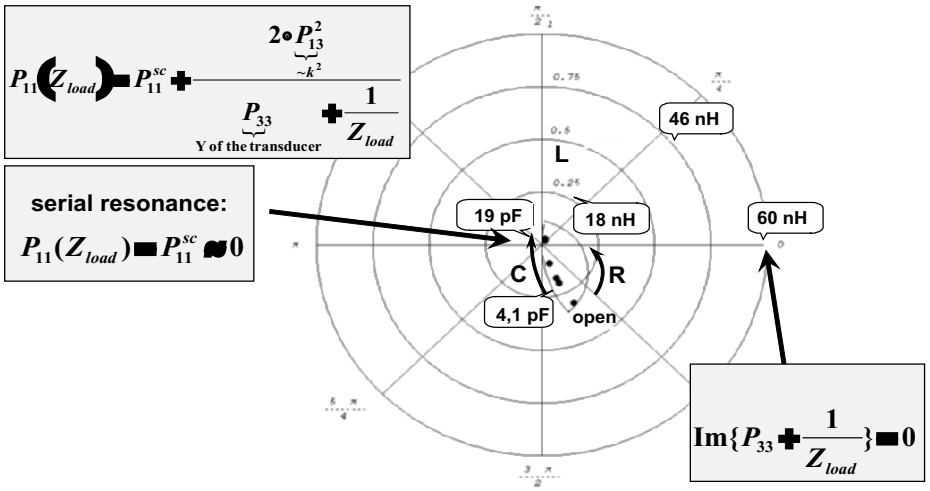


Figure 32. Measured Amplitude and Phase of the Acoustic Reflection Factor P11 (Polar Chart Representation) of a Split-Finger IDT with Capacitive and Inductive Load Impedance Z_{load}.

5. APPLICATION EXAMPLES

In this section some application examples of wireless passive SAW sensor systems and their state-of-the-art performance are presented. The wireless measurement of temperature, pressure, torque, acceleration, tire-road friction, magnetic field, and water content of soil will be discussed in further detail.

5.1 Identification Marks

Up to now, several SAW identification systems have been installed, e.g. in the car manufacturing industry [36].

Figure 33 shows an application of SAW ID tags in the Munich subway net operating in the 2.45 GHz band [35]. The SAW identification tags were mounted to the side of the railway vehicles. The interrogators were placed at selected points along the net, near the rails, and linked to a central computer. Identification under live railroad conditions proved to be a critical prerequisite for the system, given the broadband high energy interference level caused by motors and electronic alternators of the railway vehicles. The total

interference immunity, even under extremely adverse conditions, was convincingly demonstrated by the SAW identification system.

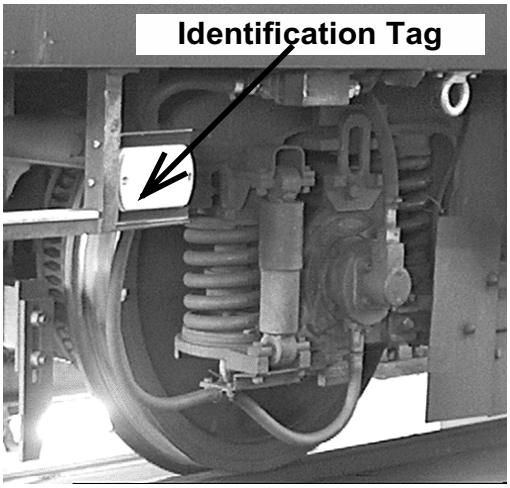


Figure 33. SAW ID Tag Mounted on the Side of a Railway Vehicle.

5.1 Temperature Sensors

Lithiumniobate, $LiNbO_3$, is an ideal material for temperature sensors, because of its large *TCD* of approx. –94 ppm/°C and its high electro-acoustic coupling factor ([17], [20], [23]).

SAW radio sensors for temperature have been used to monitor the temperature of break discs of a train, when slowing down entering a station [44], and to monitor the temperature of high-voltage surge arresters [30]. Figure 34 *a* shows the set up for monitoring the rotor temperature of an 11 kW asynchronous motor [30]. The temperature sensor was mounted on the short-circuiting ring of the rotor and read out with a loop antenna, which was mounted on the bearing plate. The motor was loaded with an eddy-current brake. Figure 34 *b* gives the measured rotor temperature. During start-up and operation under load conditions the motor heats up, as expected. When the motor stops working the temperature shows an exponential decay. Crosses mark discrete measured temperature values achieved with a reference thermometer. The slight difference between the measured values is explained due to the thermal contacts of the reference thermometer and the SAW device are not identical.

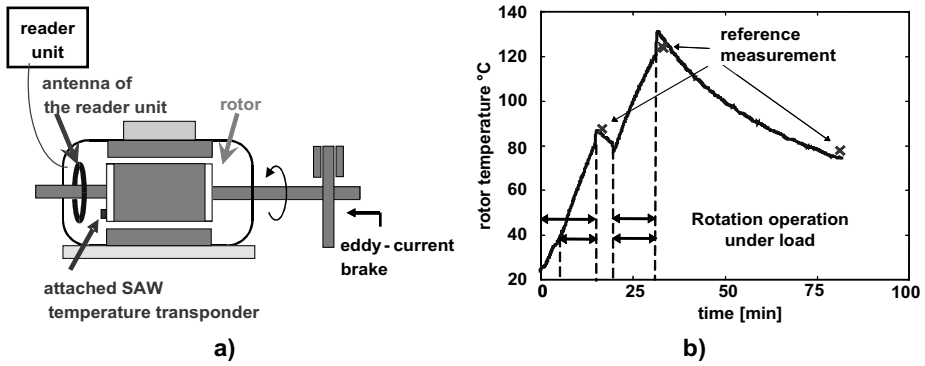


Figure 34. a) Experimental set up for Monitoring the Rotor Temperature in an 11 kW Asynchronous Motor. b) Measured Rotor Temperature; the Crosses Show Reference Measurements Derived with a Conventional PT100 Sensor.

Up to 200 °C, standard assembly, interconnect and packaging techniques can be applied, whereas higher temperatures ask for new solutions ([16], [45]). The same problem arises for the conventional SAW materials as aluminum electrodes or $LiNbO_3$ and quartz substrates. For temperatures greater than 400° C the aluminum electrodes and the IDT will be damaged.

The maximum temperature for SAW sensors, however, can be as high as 1000°C when using special substrate materials like Langasite or Gallium Orthophosphate, special electrode materials like platinum-titanium and platinum bonding wires. Figure 35 shows the photo of a mounted test device. It consists of two delay lines with different delay times, used as a test device for measuring the acoustic loss of the free surface of Langasite. The test chip operated up to a temperature of 1000°C and showed a surprisingly low increase in the insertion loss, as the temperature increased from 25°C to 1000°C (Figure 36). Between room temperature and 955°C only a very moderate increase is noticeable; however, at temperatures higher than 955°C, the insertion loss increases dramatically. A detailed analysis showed that the SMD housing used was destroyed at these high temperatures.

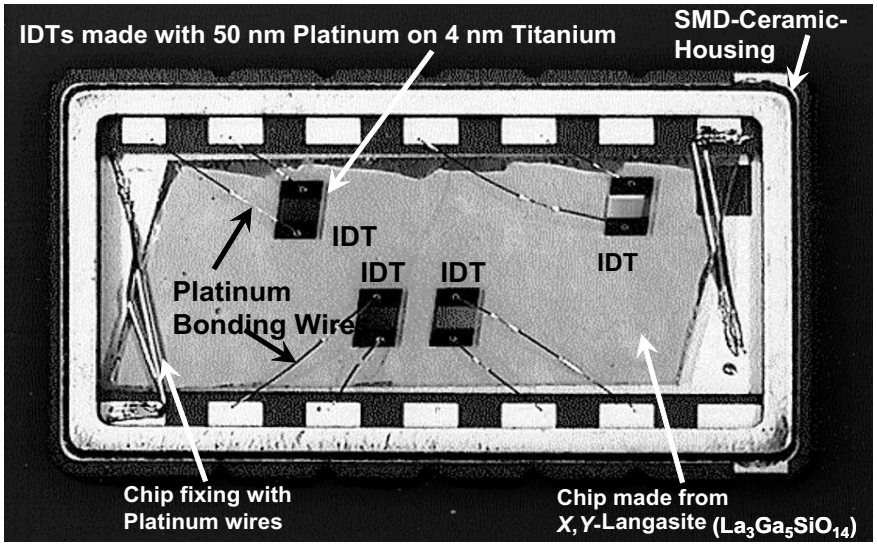


Figure 35. Delay Line for Testing the High Temperature Features of Langasit.

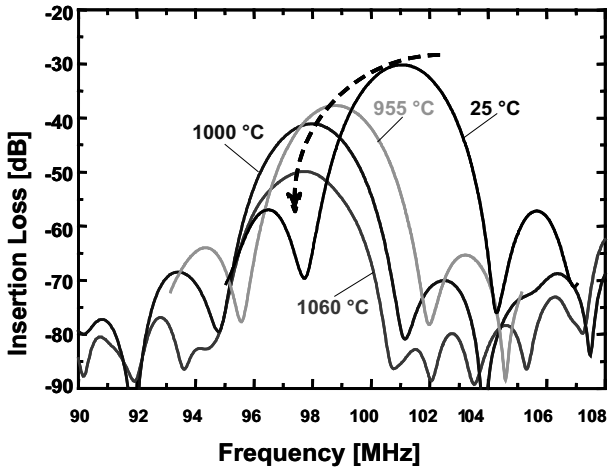


Figure 36. Insertion Loss of a Delay Line on X,Y-Langasite ($La_3Ga_5SiO_{14}$) with Platinum Electrodes in Dependency of the Temperature.

5.2 Sensors for Mechanical Quantities Like Torque, Pressure, Acceleration, Tire Friction

If SAW devices are exposed to mechanical forces like stretch, pressure or acceleration, the characteristics of the saw device are affected. Due to the bending of the substrate both the surface lengths as well as the stress condition will change. The first example for the measurement of mechanical quantities is the monitoring of torque with SAW transponders. Using such passive SAW torque transponders slip rings can be omitted ([16], [30], [52], [53], [54]). The SAW sensors measure, like resistive strain gauges, the torque indirectly by detecting the strain or stress distribution generated by a torque acting on the shaft. The fact, that the strain shows an opposite sign of ± 45° directions relative to the shaft axis, can be used for temperature compensation. Figure 37 shows a shaft with two mounted SAW torque sensors [54]. Quartz and $LiNbO_3$ as substrate materials have been investigated.

Figure 37. SAW Torque Sensors Mounted on a Shaft. (Photo: Sensor Technology, GB, [54]).

Wireless SAW pressure sensors are fabricated using a quartz diaphragm that bends under hydrostatic pressure. A reflective delay line is achieved by structuring one surface of the quartz diaphragm (Figure 38 *a*, [46], [48]). An "all-quartz package", consisting of the delay line chip itself and a lid of the same material, is expected to produce minimal thermal stresses and therefore reduces the cross-sensitivity to temperature [46]. One obtains a pressure resolution of about 1 % of the full range. Figure 38 *b* shows the pressure measurement of a sensor in a car tire ([44], [49]).

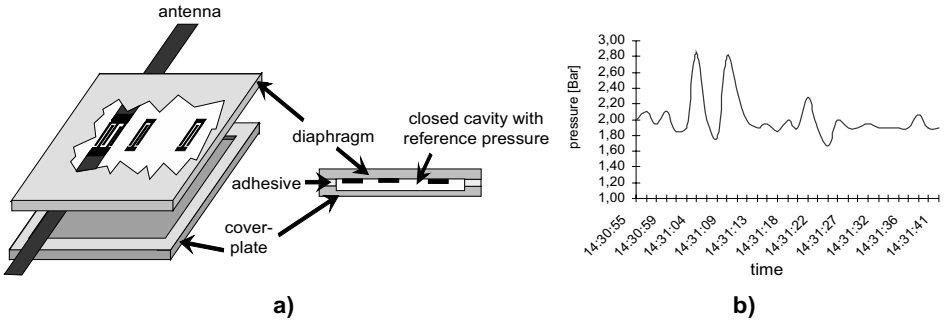


Figure 38. a) Schematic drawing of a SAW pressure sensor;
b) Pressure measurement of a sensor in a car tire ([44], [49]).

With a SAW transponder it is also possible to monitor the friction coefficient between a car tire and the road, a key parameter when stabilizing a vehicle in critical situations [24]. The friction coefficient can be measured by evaluating the mechanical strain in the tire surface in contact with the road, by utilizing the deformation of the tread elements. Figure 39 shows the expected deformation of a tread element during the road contact.

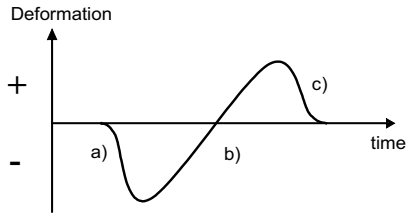


Figure 39. Deformation of a Tread Element During Road Contact
a) Entering; b) Contact Area; c) Running Out.

Siemens in cooperation with the Continental AG and the TU Darmstadt (FZD) have studied the possibility to apply SAW sensors for the measurement of the friction coefficient. A schematic drawing of an experimental encapsulation of a SAW bending beam that can be implemented into a tire tread element is depicted in Figure 40 *a*. Figure 40 *b* shows the variation in the phase difference of such a SAW friction sensor during the tire-road contact in time. The narrow line *A* gives the measurement on a dry road and the broad line *B* on a wet one, respectively. Both curves were recorded with a

velocity *v* of 40 km/h. The tread is flattened in the contact region due to the load of the vehicle. In the first part of the contact area the tread elements transfer a propulsion to the road and in the second part a breaking. During the contact a small slippage of the tread element occurs. If the tire/road contact is lowered, this slippage increases and the gradient of the strain curve is lowered.

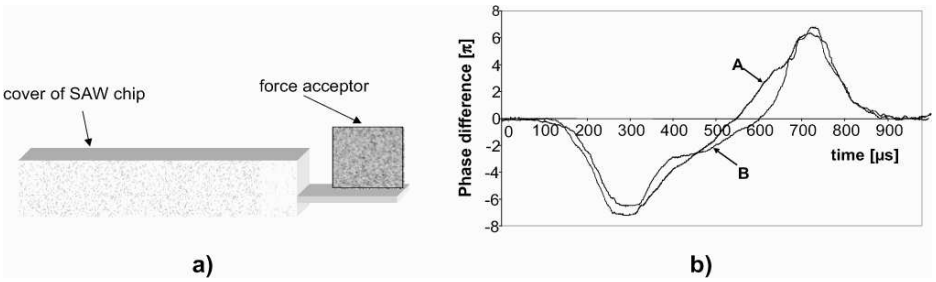


Figure 40. a) Mounting of a SAW Sensor for Tire Friction Control, which Can be Implemented into a Tire Tread Element. b) Variation in Phase Difference of a SAW Friction Sensor During the Tire-Road Contact Time. A (arrow line): Dry Road, B (broad line): Wet Road; Both Curves Were Recorded at v = 40 km/h.

With an additional seismic mass, a SAW radio readable accelerometer can be realized ([49], [50]). Figure 41 *a* shows an accelerometer configuration, using a flexured SAW cantilever beam. This sensor was wired to an antenna and fixed to a dart (Figure 41 *b*). With this wireless accelerometer system the deceleration of the dart invading the target is shown in *Figure* 42.

Figure 41. a) SAW Accelerometer Configuration using a Seismic Mass and a Flexured Cantilever Beam b) Deceleration of the Dart Invading a Target.

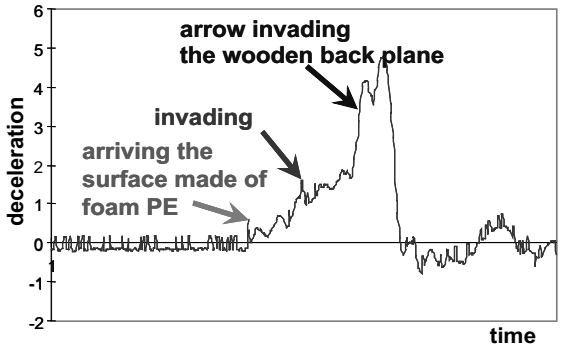


Figure 42. Measured Deceleration of a Dart.

5.4 Impedance Varying Sensors Like Current and Water Content Sensors

In these section examples of radio sensor configurations are presented, where the SAW device is used only as a transponder. The actual measurement is performed by an external (classical) sensor element, which modulates the reflection properties of the SAW transponder according to the actual measurement value. Currents produce magnetic fields, which can be detected by SAW devices, coated with a magneto-strictive layer [55]. Another possibility is an impedance-type sensor as shown in section 0 with magneto-resistors or giant magnetic impedance (GMI) elements acting as variable load Z_L ([23], [56]). Figure 43 shows the schematic drawing of such a current sensor: two magneto resistors exposed to two biasing magnetic fields of opposite polarity, each loading an IDT-reflector.

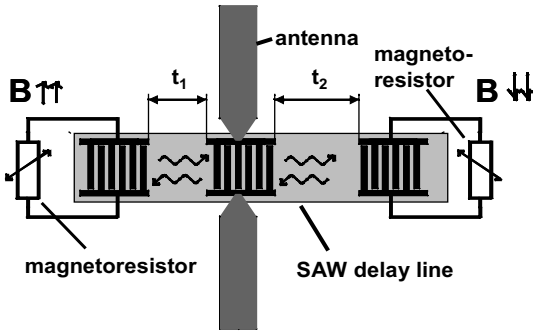


Figure 43. Schematic Drawing of a SAW Magnetic Field Sensor.

The sensor information is calculated from the ratio of the two reflected pulses (Figure 44a) The current resolution of this radio measurement is approximately 5 % of the full scale. *Figure* 44 *b* gives some measurement values as a function of the current.

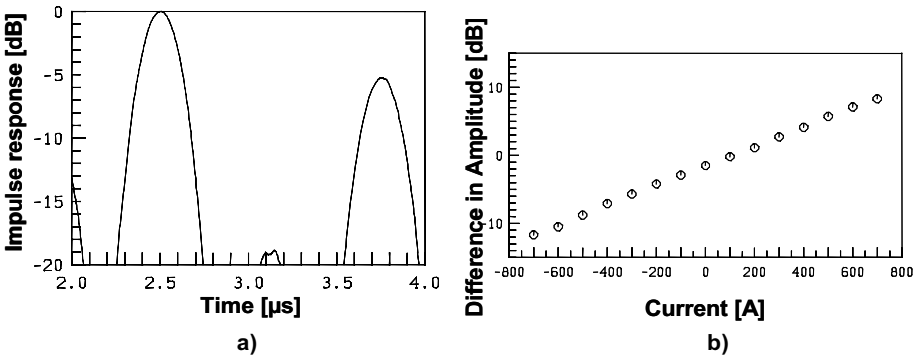


Figure 44. Sensor Characteristics of the Wireless SAW Current Sensor Outlined in Figure 43 a) Impulse Response b) Measured Difference Between the Amplitudes.

By connecting an impedance-type SAW transponder to a parallel-wire electrical transmission line with two rods, a passive radio readable water content sensor can be designed [57]. Fig ure 45 shows the complete water content sensor element. There are two rods at the bottom, matching elements, and the SAW device. On the top an antenna and an additional inductive element L2 are mounted, to prevent static discharge of the SAW device.

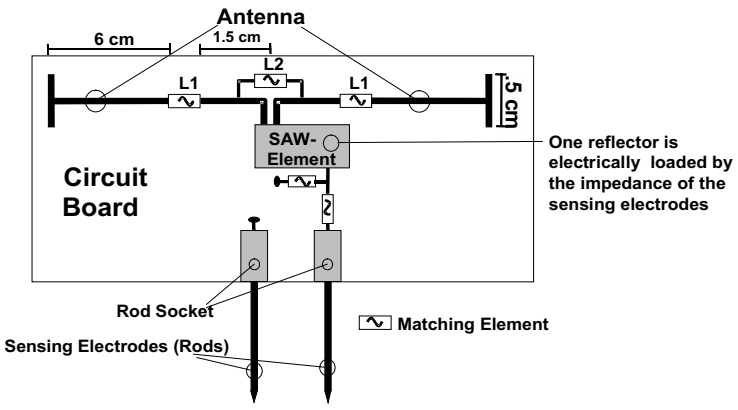


Figure 45. Schematic Drawing of a Water-content Sensor.

The rods are stuck into sandy soil. The permittivity ε of sandy soil is strongly dependent on the water content due to the high dielectric constant of free water (*Figure* 46).

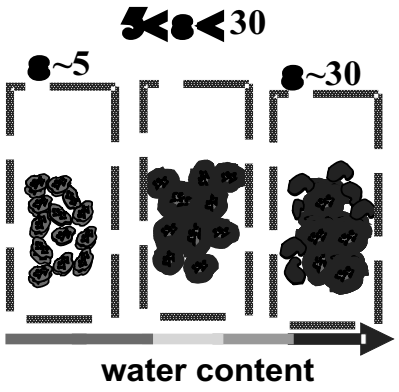


Figure 46. Change of the Permittivity ε of Sandy Soil with Increasing Water Content.

Figure 47 shows the variation of the electrical reflection coefficient at the center frequency of the sensing electrodes, including the matching circuitry, when are put into a sandy soil, with respect to a change of the water content of the soil. The dashed line shows the simulated behaviour and the triangles represent measured points. Essentially, the sensing electrodes can be simulated by an open ended transmission line. The effective electrical length of the line varies with the water content of the soil due to the change in permittivity. Additionally, the losses increase as the water content increases. The matching circuit was in such a way chosen, that the variation of the impedance of the sensor range from minimal acoustic reflection up to the point of maximum reflection of the split finger IDT. The theoretical dynamics of the reflection with this load impedance range from -5 dB to -30 dB, which results in a dynamic range of 25dB.

A read-out distance of about 2 m can be achieved with this water content sensor used in dry and in moist soil. Figure 48 shows the impulse response of the water content sensor for dry and moist soil. There are three designed signals at 3, 4, and 5 μs. At 2 μs, there is an additional echo. This is due to the fact that the rods build up an antenna, too. The signal is generated by an acoustic wave traveling only one way between the reflector #2 and the input transducer. The sensor signal is evaluated using the amplitude difference of

the signals at 3 and 4 µs. The signal at 5 µs is weakened significantly, if the reflector #2 is operating with maximum acoustic reflection.

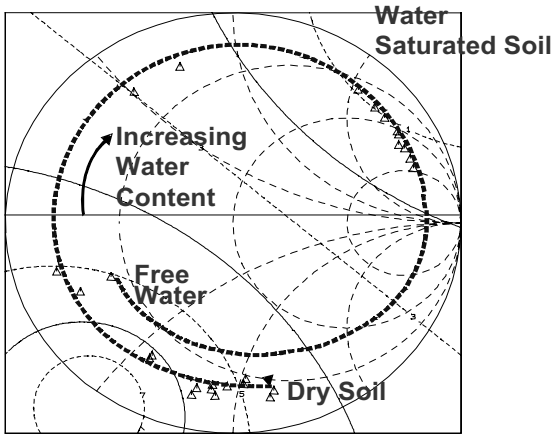


Figure 47. Electrical Reflection Coefficient of the Sensor Electrodes Versus the Water Content.

6. CONCLUSION

Due to their high quality factors, surface acoustic wave devices offer the possibility of passive sensors connected to a read-out unit solely by a radio-frequency link. Based on standard radar systems operating in the VHF/UHF range various wireless SAW sensor systems have been tested successfully. In particular, solutions for the non-contact measurements of temperature, mechanical quantities like pressure, torque, acceleration, or tire-road friction, and impedance sensors for current or water content have been developed. These systems also work in harsh environments and at high temperatures. With appropriate compensation of cross-sensitivities, typical performances of wireless SAW sensor systems are characterized by read-out distances of 3-10 m. *Table 2* summarizes typical resolutions, achievable with wireless SAW sensors.

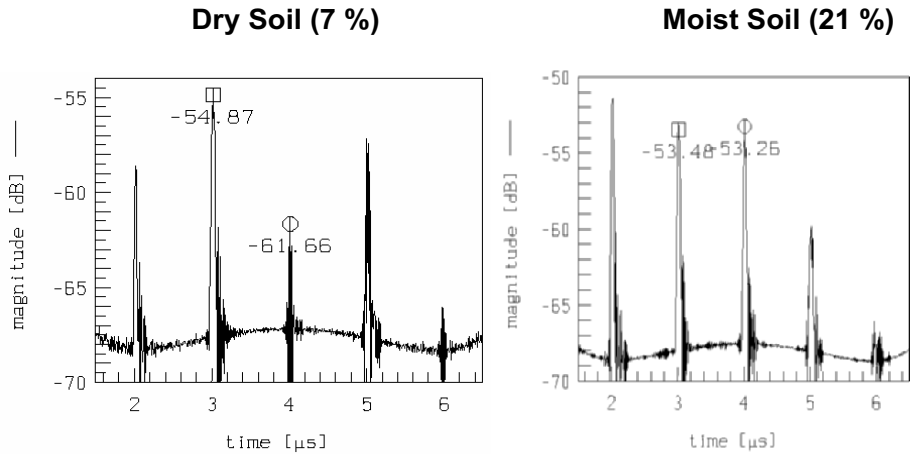


Figure 48. Impulse Response of the Water Content Sensor.

Table 2. Typical Resolution of Wireless SAW Sensors.

Measurand	Physical effect	Resolution
• identitification	analysis of signal	32 Bit
• temperature	variation of SAW velocity	0.1 K
• mechanical measurand (pressure, torque, acceleration, tire-road friction)	variation of elastic constants	1% of full scale
• impedance sensors	variation of amplitude and phase of reflected signal	5% of full scale
• distance	signal delay	20 cm
• relative position	continuous measurement of Doppler phase	2 cm
• angular positioning	measurement of Doppler phase	3 degree

ACKNOWLEDGEMENTS

The results reported in this contribution were obtained by the staff and students of the department ‘ZT MS 1’ of the Siemens AG. The authors

especially want to emphasize the contributions of W.-E. Bulst and M. Guntersdorfer for continuously encouraging this work, and C.C.W. Ruppel, W. Ruile, G. Scholl, F. Schmidt, T. Ostertag, and V. Magori as well as J. Hornsteiner, C. Kordon, O. Sczesny, E. Riha, C. Seisenberger, M. Vossiek, G. Fischerauer, and U. Wolff for their valuable contributions. Special thanks to W. Gawlik, U. Knauer, J. Zimmermann, S. Berek, B. Bienert, H. Zottel, T. Sofronijevic, and K. Riek for the fabrication of the SAW devices.

The authors also gratefully acknowledges the contributions of Prof. F. Seifert, and the staff and the students of the department of Applied Electronics at the University of Technology, Vienna, Austria especially A. Pohl, G. Ostermayer, R. Steindl, M. Brandl, and Ch. Hausleitner.

Very important contributions were developed at the Institute for Communications and Information Engineering of the University of Linz, Austria, by Prof. R. Weigel, and his students H. Scherr, T. Pankratz, and G. Schimetta, and at the Institute for Measurement Systems and Sensor Technology of the Technical University of Munich, Germany, by Prof. E. Schrüfer, and his students K. Pistor, T. Sachs, R. Großmann, and J. Michel. Thanks to A. Kirmayr from the University of Applied Sciences, Munich, for his work on the water content sensor.

REFERENCES

[1] White, R.M.; Voltmer, F.W.: *Direct piezoelectric coupling to surface elastic waves*. In: Appl. Phys. Lett. 1965; 7:314-316.

[2] Auld, B. A.: *Acoustic Fields and Waves in Solids.* Vol. II. New York: Wiley, 1973.

[3] Matthews (ed.), H.: *Surface Wave Filters.* New York: John Wiley & Sons, 1977.

[4] Oliner (ed.), A.: *Acoustic Surface Waves.* Berlin: Springer, 1978.

[5] Morgan, D. P.: *Surface-Wave Devices for Signal Processing.* Amsterdam: Elsevier, 1985.

[6] Datta, S.: *Surface Acoustic Wave Devices.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[7] Feldmann, M.; Hénaff, J.: *Surface Acoustic Waves for Signal Processing.* Boston: Artech House, 1989.

[8] Campbell, C.: *Surface Acoustic Wave Devices and Their Signal Processing Applications.* Boston: Academic Press, 1989.

[9] Nysen, P. A.; Skeie, H.; Armstrong, D.: *System for interrogating a passive transponder carrying phase-encoded information.* US Patents 4 725 841, 4 625 207, 4 625 208, 1983-1986.

[10] Yamanouchi, K.; Shimizu, G.; Morishita, K.: *2.5-GHz SAW propagation and reflection character-istics and application to passive electronic tag and matched filter.* Proceedings of the IEEE Ultrasonics Symp. 1993; 1267-1270.

[11] Reindl, L.; Ruile, W.: *Programmable Reflectors for SAW-ID-Tags*. Proceedings of the IEEE Ultrasonics Symp. 1993; 125-130.

[12] Plessky, V. P.; Kondratiev, S. N., Stierlin, R., Nyffeler, F.: *SAW Tags: New Ideas*. Proceedings of the IEEE Ultrasonics Symp. 1995; 117-120.
[13] Nieuwenhuizen, A.; Venema, M. S.: *Mass-Sensitive Devices.* In: Göpel, W.; Hesse, J.; Zemel, J. N.; Jones, T. A.; Kleitz, M.; Lundström, J.; Seiyama T. (Eds.): *Sensors. A Comprehensive Survey*. Vol. 2. Weinheim: VCH, 1991.
[14] Fischerauer, G.: *Surface Acoustic Wave Devices.* In: Göpel, W.; Hesse, J.; Zemel, J. N.; Meixner, H.; Jones, R. (Eds.): *Sensors. A Comprehensive Survey.* Vol. 8. Weinheim: VCH, 1995.
[15] Ballantine, D. S.; White, R. M.; Martin, S. J.; Ricco, A. J.; Zellers, E. T.; Frye, G. C.; Wohltjen, H.: *Acoustic Wave Sensors: Theory, Design, and Physico-Chemical Applications.* San Diego etc.: Academic Press, 1997.
[16] Dickert, F. L.; Fischerauer; Wolff, G.U.: *SAW Sensors for Harsh Environments.* IEEE Sensors Journal 2001; 1:4-13.
[17] Bao, X. Q.; Burkhard, W.; Varadan, V. V.; Varadan, V. K.: *SAW Temperature Sensor and Remote Reading System.* Proceedings of the IEEE Ultrasonics Symp. 1987; 583-585.
[18] Buff, W.: *SAW sensors*. Sensors Actuators 1992; A42:117-121.
[19] Seifert, F.; Bulst, W.E.; Ruppel, C.C.W.: *Mechanical sensors based on surface acoustic waves.* Sensors Actuators 1994; A44:231-239.
[20] Schmidt, F.; Sczesny, O.; Reindl, L.; Magori, V.: *Remote sensing of physical parameters by means of passive surface acoustic wave devices ('ID TAG').* Proceedings of the IEEE Ultrasonics Symp. 1994; 589-592.
[21] Buff, W.; Plath, F.; Schmeckebier, O.; Rusko, M.; Vandahl, T.; Luck, H.; Möller, F.: *Remote sensor system using passive SAW sensors.* Proceedings of the IEEE Ultrasonics Symp. 1994, 585-588.
[22] Reindl, L.; Scholl, G.; Ostertag, T.; Ruppel, C. C. W.; Bulst, W.-E.; Seifert, F.: *SAW devices as wireless passive sensors.* Proceedings of the IEEE Ultrasonics Symp. 1996, 363-367.
[23] Reindl, L.; Scholl, G.; Ostertag, T.; Scherr, H.; Wolff, U.; Schmidt, F.: *Theory and application of passive SAW radio transponders as sensors.* Transactions on UFFC 1998; Vol 45; 45:5:1281-1292.
[24] Pohl; Steindl, R.; Reindl, L.: *The 'intelligent tire' utilizing passive SAW sensors - measurement of tire friction.* IEEE Transaction on Instrumentation and Measurement 1999, Vol. 48; 6:1041-1046.
[25] Skolnik, M. I.: *Introduction to Radar Systems.* New York etc.: McGraw Hill, 1979.
[26] Cook, C. E.; Bernfeld, M.: *Radar Signals.* Norwood, MA. : Artech House, 1993.
[27] Rao, B.D.; Arun, K.S.: *Model Based Processing of Signals: A State Approach.* Proceedings IEEE 1992, Vol. 80; 2:283 309.
[28] Pohl, Seifert, F.; Reindl, L.; Scholl, G.; Ostertag, T.; Pietsch, W.: *Radio signals for ID tags and sensors in strong electromagnetic interference.* Proceedings of the IEEE Ultrasonics Symp. 1994; 195-198.
[29] Pohl, Ostermayer, G.; Hausleitner, C.; Seifert, F.; Reindl, L.: *Wavelet Transform with SAW Convolver for Sensor Application.* Proceedings of the IEEE Ultrasonics Symp. 1995; 143-146.
[30] Scholl, G.; Schmidt, F.; Ostertag, T.; Reindl, L.; Scherr, H.; Wolff, U.: *Wireless passive SAW sensor systems for industrial and domestic applications.* Proceedings of the IEEE Frequency Control Symp. 1998; 595-601.
[31] Fischerauer, G.; Schmidt, F.; Voss, M.; Bader, R.: *Mechatronic Extension of a Tap Holder for Process Monitoring.* Proceedings of the IECON 1998.

[32] Pohl, Ostermayer, G.; Seifert, F.: *Wireless sensing using oscillator circuits locked to remote high-Q SAW resonators.* IEEE Trans. Ultrason., Ferroelect., Freq. Contr.1998; Vol. 45; 5:1161-1168.
[33] Pohl: *A low cost high definition wireless sensor system utilizing intersymbol interference.* IEEE Trans. Ultrason., Ferroelect., Freq. Contr. 1998; Vol. 45; 5:1355-1362.
[34] Pohl: *A Review of wireless SAW Sensors.* IEEE Transactions on UFFC 2000; Vol. 47; 2:317-332.
[35] Siemens Transportation Group: *Product description A19100-V700-B535-V1-760.*
[36] baumeridenthttp://www.baumerident.com
[37] Reindl, L.; Rösler, U.; Ruppel, C.; Obertreis, R.; Weigel, R.: *Chirped SAW devices for wireless passive sensors.* Proceedings of the IEEE Ultrasonics Symp. 1997; 343-348.
[38] Pankratz, T.; Scherr, H.; Reindl, L.; Ruppel, C.; Weigel, R.: *Low TB radio SAW sensors incorporating chirped transducers and reflectors for wireless pressure sensing applications.* Proceedings of the IEEE MTT-S International Microwave Symp. 1998; 845-848.
[39] Steindl, R.; Pohl, A.; Reindl, L.; Seifert, F.: *SAW delay lines for wirelessly requestable conventional sensors.* Proceedings of the IEEE Ultrasonics Symp. 1998; 351-354.
[40] Steindl, R.; Pohl, A.; Seifert, F.: *Impedance loaded SAW sensors offer a wide range of measurement opportunities.* Proceedings of the IEEE MTT-S 1999; Anaheim, CA. 1453-1456.
[41] Ostermayer, G.; Pohl, A.; Reindl, L.; Seifert, F.: *Multiple Access to SAW Sensors Using Matched Filter Properties.* Proceedings of the IEEE Ultrasonics Symposium 1997; 339-342.
[42] Ostermayer, G.; Pohl, A.; Hausleitner, C.; Reindl, L.; Seifert, F.: *CDMA for Wireless SAW Sensor Applications.* Proceedings of the IEEE International Symposium on Spread Spectrum Techniques & Applications 1996; 795 799.
[43] Ostermayer, G.; Pohl, A.; Steindl, R.; Seifert, F.: *SAW sensors and correlative signal processing – a method providing multiple access capability.* Proceedings of the ISSSTA 1998; South Africa. 902-906.
[44] Pohl, A.; Seifert, F.: *Wireless Interrogable SAW-Sensors for Vehicular Applications.* Proceedings of the IEEE Instrumentation and Measurement Conference 1996; 1465-1468.
[45] Reindl, L.; Steindl, R.; Pohl, A.; Hornsteiner, J.; Riha, E.; Seifert, F.: *Passive SAW sensors for temperature and other measurands.* Proceedings of the Tempmeko 1999; Netherlands. 424-429
[46] Montress, G. K.; Parker, T. E.; Callerame, J.: *A miniature hybrid circuit SAW oscillator using an all quartz packaged resonator.* Proceedings of the IEEE Ultrasonics Symp. 1985; 277-282.
[47] Taziev, R. M.; Kolosovsky, E. A.; Kozlov, A. S.: *Deformation-sensitive cuts for surface acoustic waves in α-quartz.* Proceedings of the IEEE Freq. Contr. Symp. 1993; 660-664.
[48] Scherr, H.; Scholl, G.; Seifert, F.; Weigel, R.: *Quartz pressure sensor based on SAW reflective delay line.* Proceedings of the IEEE Ultrasonics Symp. 1996; 347-30.
[49] Pohl, A.; Ostermayer, G.; Reindl, L.; Seifert, F.: *Monitoring the tire pressure at cars using passive SAW sensors.* Proceedings of the IEEE Ultrason. Symp. 1997; Toronto; 471-474.
[50] Pohl, A.; Springer, A.; Reindl, L.; Seifert, F.; Weigel, R.: *New applications of wirelessly interrogable passive SAW sensors.* Proceedings of the MTT-S 1998; Baltimore. 503-506.
[51] Pohl, A.; Steindl, R.; Reindl, L.: *Measurements of vibration and acceleration utilizing SAW sensors.* SENSOR'99 1999; Germany. Vol. 2; 53-58.

[52] Sachs, T.; Großmann, R.; Michel, J.; Schrüfer, E.: *Remote sensing using quartz sensors.* Proceedings of the SPIE's Symp. on Smart Structures and Materials 1996; Vol. 2718; 47-58.

[53] Wolff, U.; Schmidt, F.; Scholl, G.; Magori, V.: *Radio accessible SAW sensors for simultaneous non-contact measurement of torque and temperature.* Proceedings of the IEEE Ultrasonics Symp. 1996; 359-362.

[54] http://www.sensors.co.uk/

[55] Robbins, W.; Hietala, A.: *A simple phenomenological model of tunable SAW devices using mag-netostrictive thin films.* IEEE Transactions on UFFC 1988; Vol. 35; 6:718-722.

[56] Steindl, R.; Hausleitner, C.; Pohl, A.; Hauser, H.; Nicolics, J.: *Giant magneto-impedance magnetic field sensor with surface acoustic wave technology.* Proceedings of the Eurosensors; 1999; Delft, Netherlands.

[57] Reindl, L.; Ruppel, C.C.W.; Kirmayr, A.; Stockhausen, N.; Hilhorst, M.: *Passive radio requestable SAW water content sensor.* Proceedings of the IEEE Ultrasonics Symp. 1999

[58] Seifert, F.; Weigel, R.: *SAW-Based Wireless Sensor and Communication Techniques.* In: Proceedings of the European Microwave Conference 1997; Jerusalem, Israel. 1323-1346.

[59] Weigel, R.: *Efficient Modeling of Microacoustic Structures in Terms of Electromagnetic Field Concepts.* Proceedings of the International Symposium on Antennas and Propagation (ISAP) 1996; Chiba, Japan. 897-900.

[60] Thomas, R.; Johannes, T. W.; Ruile, W.; Weigel, R.: *Determination of Phase Velocity and Attenuation of Surface Acoustic Waves with Improved Accuracy.* In: Proceedings of the IEEE International Ultrasonics Symposium. (1998), Sendai, Japan. 277-300.

[61] Knauer, U.; Machui, J.; Ruppel, C. C. W.: *Design, Fabrication, and Application of GHz SAW Device.* Proceedings of the IEEE International Microwave Symposium 1997; Denver, USA. pp. 1821-1824.

[62] Reindl, L.; Ruppel, C.C.W.; Berek, S.; Knauer, U.; Vossiek, M.; Heide, P.; Oréans, L.: *Design, Fabrication, and Application of Precise SAW Delay Lines Used in a FMCW Radar System.* Proceedings of the IEEE Transactions on Microwave, Theory and Techniques L. 2001; MTT-49; 787-794.

[63] Mackensen, E.; Kuntz, W.: *Smart Wireless Autonomous Microsystems (SWAMs) for Sensor Actuator Networks.* To be published in: Sensors for industry conference (2004), New Orleans, USA.

[64] Lord Rayleigh: *On waves propagated along the plane surface of an elastic solid.* London Math. Soc. 1885; 17:4-11.

[65] Ruppel, C.C.W.; Ruile, W.; Scholl, G.; Wagner, K.C. Männer, O.: *Review of models for low-loss filter design and applications.* Proceedings of the IEEE Int. Ultrasonics Symp. 1994; pp. 313-324.

[66] Tobolka, G.: *Mixed matrix representation of SAW transducers.* Proceedings of the IEEE Trans. 1979; SU-26; pp. 426-428.

Chapter 6
PRECISE VECTORIAL MAGNETIC SENSORS

Pavel Ripka
Czech Technical University, Faculty of Electrical Engineering
Technicka 2, 166 00 Praha 6, Czech Republic

Abstract: Precise magnetic sensors are required in many fields: besides the traditional geophysical magnetometers they serve for navigation, position tracking, military and security. First a short review of basic practical rules and sensor parameters is given. Next the most important semiconductor magnetic sensors (Hall sensors and semiconductor magnetoresistors) are discussed. Following sections describe the most important precise magnetic sensors: ferromagnetic magnetoresistors (AMR, GMR and SDT) and fluxgate sensors. Other types of magnetic sensors (SQUIDS, Magnetooptic sensors, GMI and Magnetostrictive sensors) are described briefly. Section 6 is devoted to current sensors based on magnetic principles: Instrument current transformers, Rogowski coil (di/dt sensor), Fluxgate DC current sensors, Hall current sensors, and Magnetoresistive current sensor. Other applications of magnetic sensors are reviewed in Section 7. They include position sensors, Compass, Security and military applications, Geophysics and Space research, Biomagnetism and medical applications, and Non-destructive testing. Final section compares the described sensor types. Fluxgates are the most precise vectorial sensors, followed by AMRs, which become great competitors of Hall sensors. Until now, GMR and SDT sensors are rather nonlinear. GMI sensors are recently commercially available, but their present accuracy is not competitive. Common problem of magnetic sensors containing ferromagnetic material is perming (remanence). This was solved only with fluxgates and flipped AMRs.

Keywords: vectorial magnetic sensors, magnetoresistors, fluxgate sensors, Hall sensors, magnetooptic sensors

1. INTRODUCTION

Precise magnetic sensors are required in many fields: besides the traditional geophysical magnetometers they serve for navigation, position tracking, military and security [1]. The most popular precise vectorial magnetic

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 203-228.

sensors are fluxgates [2]; improved AMR magnetoresistors became recently a reasonable alternative for less demanding portable devices [3].

Magnetic field sensors are mainly based on semiconductor materials (Hall sensors and semiconductor magnetoresistors) and ferromagnetic materials (ferromagnetic magnetoresistors and fluxgates). Other magnetic sensors utilize optical, resonant and superconducting properties.

Almost all magnetic sensors measure both DC and AC fields – the only exception are induction coils, which have no DC response.

This chapter summarizes and updates general information on magnetic sensors given in [1].

1.1 Basic Rules

Table 1 gives an overview of the range of magnetic fields to be measured. Obviously no sensor can cover the whole range. In this study we concentrate on sensors suitable to measure fields between 1 nT and 1 mT

Table 1. Typical Magnitudes of Magnetic Field B (1 T = 10 000 Oe).

100 T	Pulse field
10 T	Superconducting magnet
2 T	Electromagnet
0.5 T	Surface of strong perm. magnet (NdFeB)
0.1 T	Surface of cheap magnet (ferrite)
10 mT	Power cable
50 μT	Earth's field in central Europe – total value
	(20 μT horizontal component, 40 μT vertical component)
1 μT	Vehicle in 2 m distance
35 nT	field resolution of 0.1 deg. compass sensor
20 nT	155 mm projectile 1.5 m deep
1 nT	Bomb 6 m deep
10 fT	Human brain

Magnetic field of objects rapidly decreases with distance. In ideal case the dependence follows one of the basic rules:

Dipole field (from small objects)	$B \sim 1/r^3$
Long iron objects (iron pipe)	$B \sim 1/r^2$
Long straight current conductor	$B \sim 1/r$

Figure 1 illustrates the case of long straight current conductor with very distant back conductor.

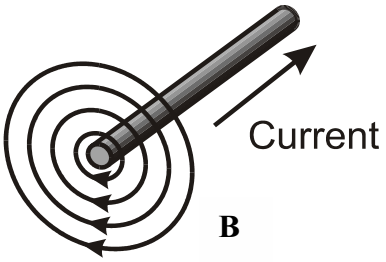


Figure 1. Magnetic Field of a Long Straight Conductor.

1.2 Vectorial Versus Scalar

Scalar sensors

Resonant sensors measure usually only the magnitude of B ("total field B").

$$B = \sqrt{B_x^2 + B_y^2 + B_z^2} \tag{1}$$

Vectorial sensors

Most magnetic sensors measure the projection of **B** into the sensor sensitive axis. They can be single-axis, but very often they are mounted perpendicularly to form tri-axial sensor head.

The rms value of the total field B should often be calculated from the reading of vectorial sensors. The correct way is to calculate rms value from instantaneous values of the total field B(t). To do it properly, all three channels should be sampled synchronously. Some instruments do it in a wrong way, calculating it from rms values of individual sensors. This formula assumes that the field does not change direction, which is often not the case.

$$B_{rms} = \frac{1}{T}\sqrt{\int_0^T \left(B_x^2 + B_y^2 + B_z^2\right) dt} \neq \sqrt{B_{rmsx}^2 + B_{rmsy}^2 + B_{rmsz}^2} \tag{2}$$

1.3 Parameters

Table 2 shows the important parameters of magnetic sensors. Perming is the change of the sensor offset after magnetic shock. Perming is serious problem especially for sensors containing ferromagnetic components, which cannot be demagnetized. This is often the case of sensors utilizing field con-

centrators or shieldings. Crossfield error is due to non-linear response to fields out of the sensing direction. When not suppressed or compensated, this error can degrade the performance of fluxgate and AMR sensors [4].

Table 2. Basic Specifications of Magnetic Sensors.

• FS range, linearity, hysteresis
• TC (temperature coefficient, "tempco") of sensitivity
• Offset, offset tempco and long-term stability
• Perming
• Crossfield sensitivity
• Noise: PSD (power spectrum density), rms or p-p value
• Resistance against environment: temperature, humidity, vibrations

2. SEMICONDUCTOR MAGNETIC SENSORS

Measured by the production volume, by far the most important magnetic sensors are Hall devices [5]. New types of structured semiconductor magnetoresistors were developed for industrial applications. All other semiconductor magnetic sensors are exotic devices such as magnetotransistors, magnetodiodes, rotating current domain sensors and split-current sensors.

2.1 Hall Sensors

The basic equation for Hall devices is

$$V_H = \frac{R_H}{t} IB\,, \tag{3}$$

where V_H is the Hall voltage;
I - measuring current;
R_H - all resistance (material and geometrical constant);
B - measured field;
t - film thickness.

Hall resistance R_H is inversely proportional to the free-carrier density: this explains why Hall effect in metals is much smaller than in semiconductors. Typical current-related sensitivity is $S_I = R_H / t = 100$ V/AT.

For low-power applications, Hall sensors are often supplied by voltage. In this case the voltage-related sensitivity S_V depends on carrier mobility. With high-mobility materials such as InSb, devices with voltage sensitivity of 5V/VT were developed.

Silicon Hall sensors are less sensitive, but using C-MOS process cheap integrated devices can easily be made. Integrated Hall sensors may have only 3 terminals: supply voltage, ground and output. Analog on-chip electronics includes excitation circuits, spinning-current switching for offset reduction, differential amplification and demodulation of the output voltage. Smart Hall sensors include digital circuits for calibration, temperature corrections and other functions. Some programmable Hall devices such as Micronas HAL 800 work normally in analog mode, but they can be switched to digital mode and reprogrammed for a specific magnetic circuit (including temperature compensation of permanent magnet and soft iron parts).

Vertical Hall sensor

Unlike traditional Hall sensors, vertical Hall sensor measures field B which is parallel to the substrate. This allows to make three-axis sensor on one chip by integrating one vertical and two perpendicular horizontal sensors.

The expected advantages of this device are long-term stability and robustness, as the active zone is buried under the chip surface. But due to technological problems these devices have not found many applications.

Permalloy Flux concentrators

are used for Hall sensors and also for magnetoresistors. They increase the sensitivity (typically by the factor of 10 to 50), but bring other problems:

- TC of sensitivity;
- Perming;
- Loss of linearity.

2.2 Semiconductor Magnetoresistors

Lorentz force deflects the electrons from their direct path causing increase of the resistance. The main disadvantages of semiconductor magnetoresistors are:

- low sensitivity for small fields;
- temperature dependence;
- independence of the field sign.

Popular InSb magnetoresistors were produced by Siemens, but their production was discontinued. Murata produces new type of InSb AMR sensor chip which, consists of multiple sensors connected serially [6].

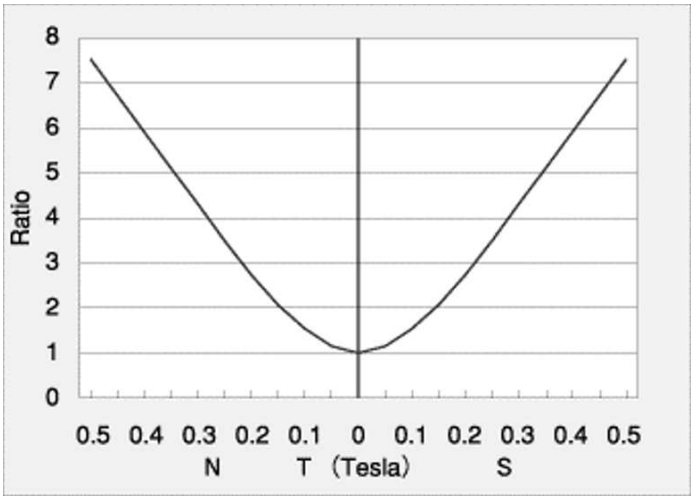


Figure 2. Characteristics of Murata InSb Magnetoresistor (from [6]).

These sensors are used for rotary sensors and proximity switches and position detectors. Another application is recognition of marks made by high coercivity material (e.g. on the banknotes). For this purpose InSb magnetoresistive elements are combined with a permanent magnet.

The temperature coefficient of these magnetoresistors is approx. -0.4%/°C. They can be competitive in temperature stability and noise resistance when used as rotational sensors in automotive industry [7].

3. FERROMAGNETIC MAGNETORESISTORS

3.1 AMR

AMR sensors are based on Anisotropic magnetoresistance effect: the resistivity of ferromagnetic alloys measured in a direction parallel to the magnetisation of e.g. Permalloy film is slightly higher than the resistivity measured perpendicular to the magnetization [8].

AMR sensors are based on Permalloy thin film strip deposited on a silicon wafer (Figure 3). Without external field the strip is magnetized in x direction, in which the measurement current flows. External measured field H_Y rotates magnetisation M and R changes by 2-3%.

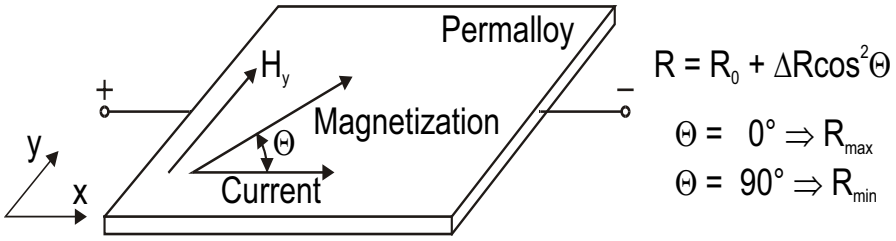


Figure 3. Magnetoresistive Effect in Permalloy: H_y is the Measured Field.

The Anisotropic Magnetoresistors (AMR) are, in general, suitable for measurement of magnetic fields up to 10 mT. The AMR sensors have high sensitivity, wide operating temperature range, offsets which are more stable than those of the Hall sensors and a wide operating frequency range approaching 10 MHz. Another advantage is their low sensitivity to mechanical stress (no piezo-effect, which is a problem for semiconductor devices). The high sensitivity of AMR sensors has led to their application in vehicle detection, contactless measurement of electrical currents, measurements of position and rotational speed in machinery, and in low-end electronic compasses. At present, AMRs have increasing importance in the automotive industry with applications such as pedal-position measurement, wheel-speed sensors for ABS and engine management systems where they are used to measure position and crankshaft angle for electronic ignition timing.

Characteristics of AMR strip is unipolar (Figure 4 a). Shifting the working point to the linear part by bias field is possible, but usually not very practical: temperature stability of the bias field amplitude limits the performance of the whole device. Much better is to use Barber Poles, which deflect the current by 45^0 and make the characteristic linear and sensitive to field sign - Figure 4 b. Barber poles are formed by strips of high-resistive aluminum deposited on top of less-resistive Permalloy (Figure 5).

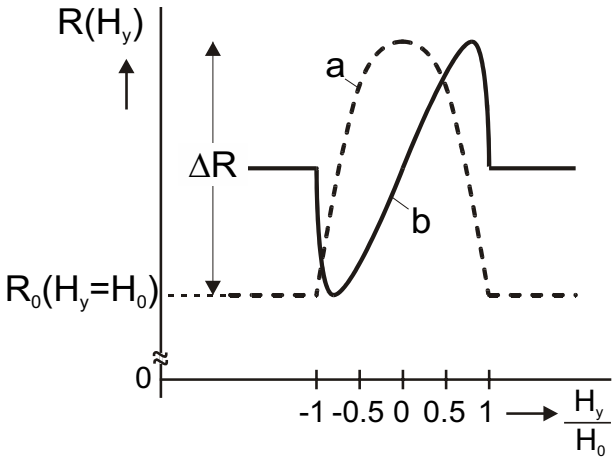


Figure 4. Resistance of AMR Element a) Without Linearization; b)With Barber Poles.

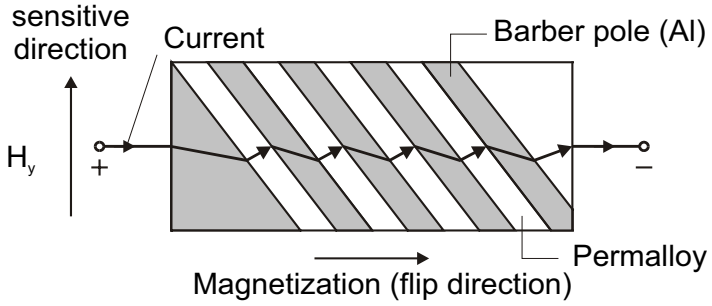


Figure 5. AMR Sensor with Barber Poles.

AMR sensors are usually used in a Wheatstone bridge with opposite sensors having Barber Pole orientations of ± 45° (Figure 6). This arrangement reduces the temperature drift and gives bipolar characteristics. This type of AMR sensor is manufactured by Philips, Honeywell, Sensitech and others.

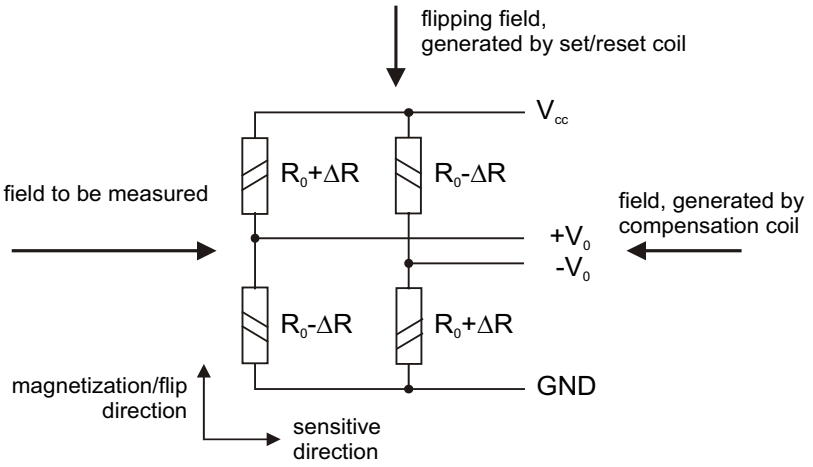


Figure 6. AMR Wheatstone Bridge.

AMR: flipping and feedback compensation

Unwanted change of strip permanent magnetization (e.g. due to exposition to high field) may distort the sensor characteristic. The proper characteristics can be reset by short field "flip" or "reset" pulse.

Typical values of the magnetic fields which are applied in the x direction to cause this magnetization flip are of the order of 300 A/m with 10 μs duration. Such field can be easily achieved by discharging capacitor into the coil. Popular technique is periodical flipping by alternating polarities and processing of the sensor output by synchronous detector. Figure 7 shows that sensor characteristics is reversed after flipping pulse of reversed polarity. The flipping frequency is typically about 200 Hz to give a bandwidth of about 20 Hz. For some applications much lower flipping rate is used to lower the power consumption. Flipping technique reduces the sensor offset, reduces perming and doubles the sensitivity. Flipping also reduces the crossfield error. Sensor parameters depend on the amplitude and shape of the flipping pulse.

Feedback compensation increases the sensor linearity and improves sensitivity tempco, but it cannot improve the offset stability.

Some modern magnetoresistors such as the Honeywell HMR 1000 and the Philips KMZ 41 have both flipping and feedback coils incorporated into the chip.

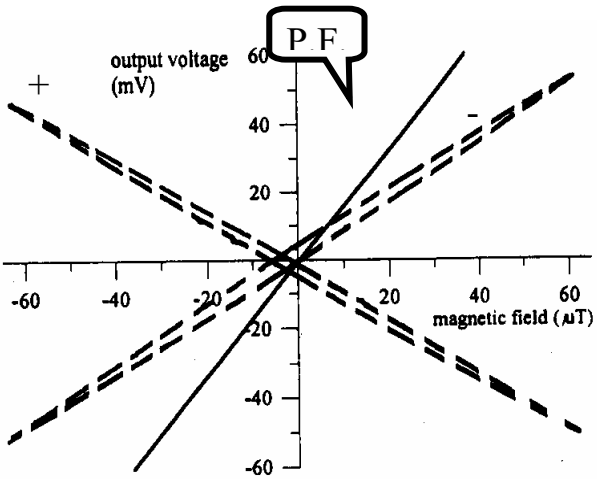


Figure 7. Characteristics of Philips KMZ 10 After Positive [+] and Negative [-] Flip *P.F. is Characteristics for Periodical Flipping* [9].

Figure 8 shows the total error of the precise AMR magnetometer working within the Earth's field range (± 50 μT). The testing field swept several times across the full-scale range deviations from linear response were plotted. The maximum error was ± 50 nT when proper flipping and feedback techniques are used.

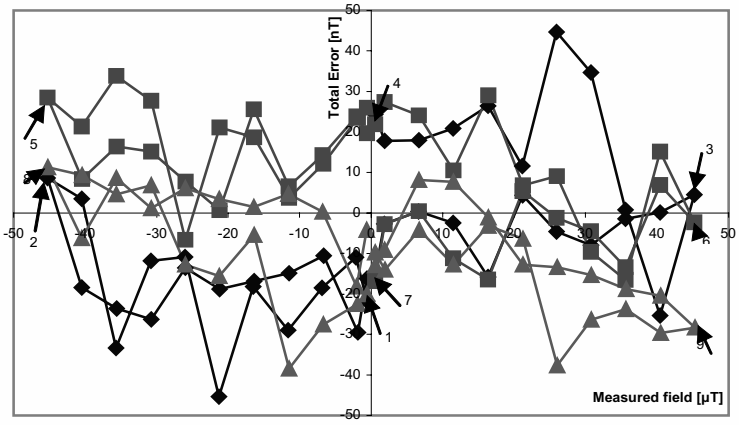


Figure 8. Flipped + Compensated AMR - Overall Accuracy (from [3]).

Figure 9 shows the noise spectrum of AMR sensor. The time-domain noise level is aprox. 15 nT p-p.

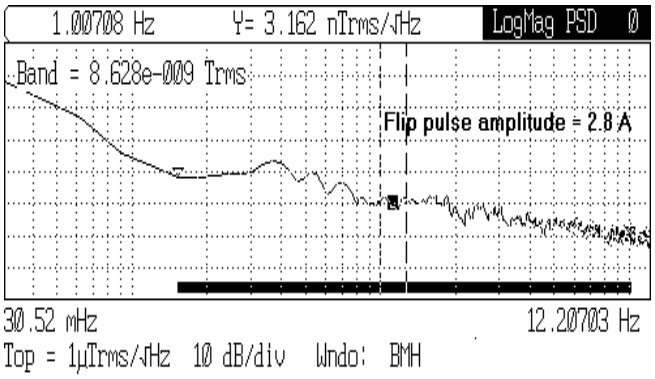


Figure 9. Noise Power Density of Flipped Philips KMZ 51 Sensor. The Peak-to-peak Noise is About 15 nT. (from [3]).

AMRs are more precise than Hall sensors. Their accuracy is already sufficient for some less-demanding applications as vehicle detection, current sensing, position tracking and low-accuracy compass. AMR sensors thus fill the niche between Hall sensors and fluxgates. Precise AMR magnetometer which uses flipping and feedback may have similar complexity and power consumption as fluxgate magnetometer. Using the same sensor size, also the performance will be similar. AMR sensor can even be used in fluxgate mode [10]

Table 3. Advantages of AMR Sensors Compared to Hall Sensors.

high sensitivity
for position sensors: magnet may be cheaper or smaller or airgap higher
for magnetic field sensors: higher accuracy
no piezo effect
higher operational temperatures

3.2 GMR, SDT and Others

These novel devices were developed for reading heads, but they can also serve as linear sensors. Up to now their accuracy is lower than that of AMR sensors, but they can find more sensing applications in the future.

GMR magnetoresistors are based on spin - dependent scattering: resistance of two thin ferromagnetic layers separated by a thin nonmagnetic conducting layer can be altered by changing the orientation of moments of the ferromagnetic layers from parallel to antiparallel as shown in Figure 10.

In a basic magnetic sandwich the antiparallel state is energetically preferred in zero external field. Basic GMR sandwich has large hysteresis and non-linearity and its characteristics is unipolar (it cannot recognize field polarity like semiconductor magnetoresistors). All this can be improved by biasing techniques

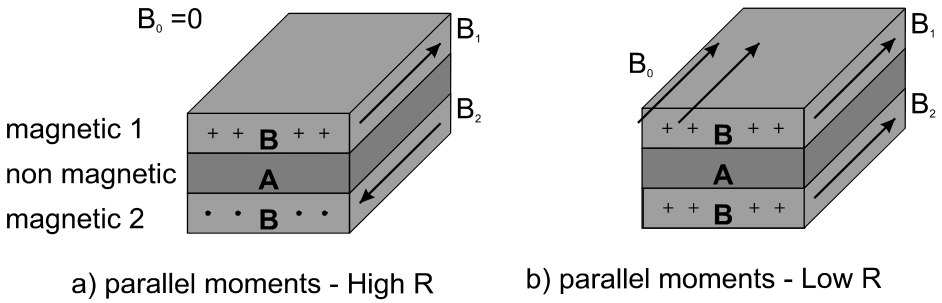


Figure 10. Basic GMR Sandwich (B_0 is the Measured Field) - after [11].

Spin valves

Most of the GMR sensors on the market are spin valves. Spin valves have one of the magnetic layers blocked by another "hard" layer (which is usually artificial antiferromagnet). This gives them characteristics shown in Figure 11. The small hysteresis curve belongs to the rotation of the free (unpinned) layer, while the high-field loop is the characteristics of the hard layer.

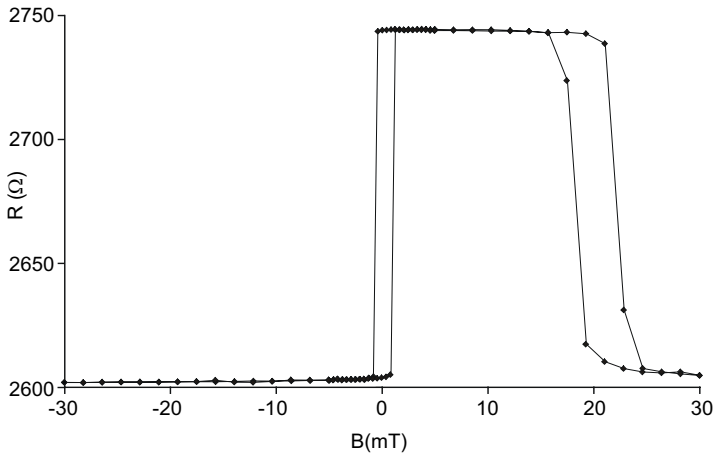


Figure 11. Characteristics of the Spin valve GMR - from [1].

One of the big potential applications of the spin valves is rotational sensor. If the external field (caused e.g. by rotating permanent magnet) is high enough to saturate the soft unpinned layer, the unpinned layer rotates with the external field and the sensor responds only to the field direction, not to its size [12]. Angular characteristics of a (saturated) single spin valve sensor is shown in Figure 12.

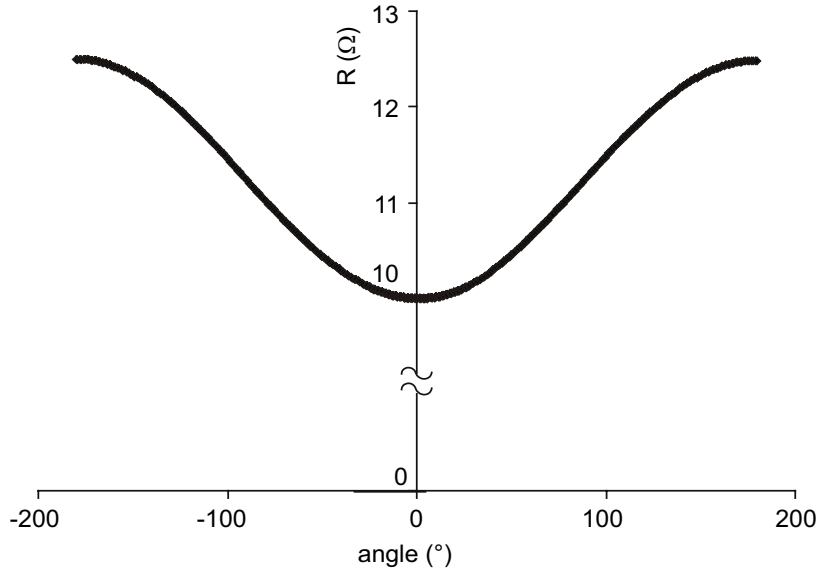


Figure 12. Angular Characteristics of the Spin Valve GMR (for High Field) – from [1].

Also GMR sensors are used in Wheatstone bridges to suppress the offset temperature drift caused by temperature coefficient of the structure resistivity. The problem is, how to give to the individual sensors in the opposite bridge arms different sensitivity to the external field. No elegant solution similar to Barber poles was found. Presently three techniques are being used:

1. biasing by integrated field coil
2. biasing by integrated hard magnet (or antiferromagnet)
3. field concentrators and shields

Integrated flat biasing coil can give required two opposite field directions on the same GMR chip. Contrary to that it is a technological cHallenge to produce two opposite oriented hard magnets on the same layer. Last technique of soft magnetic field concentrators, which increase the external field for two bridge arms and shield the other two is being used in GMR sensors produced by NVE. Example of the characteristics of these sensors is given in

Figure 13. Notice that field concentrators cannot give linear response to GMR sensors bridge.

Biasing by AC field is promising technique for improving the properties of GMR sensors, but the price is increased complexity and power consumption [13].

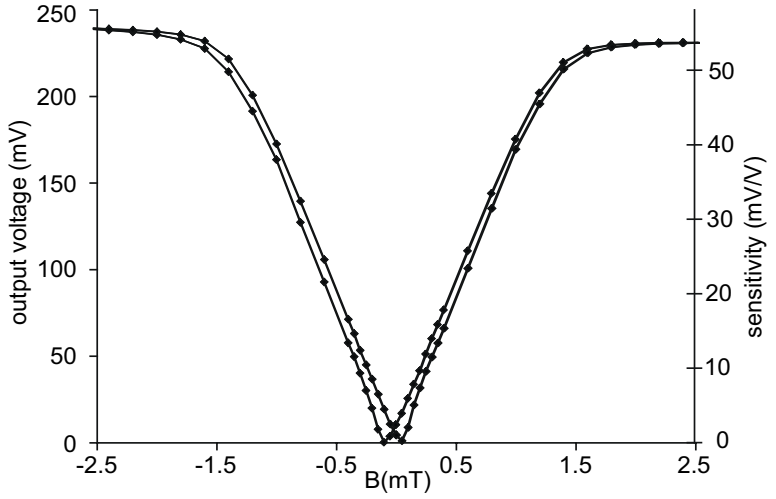


Figure 13. Characteristics of the GMR NVE AA00-02 Sensor – from [14].

SDT (Spin-depending Tunelling) sensors are small, low-power, but also very non-linear [15], [16].

4. FLUXGATES

Fluxgates are the most sensitive room-temperature magnetic sensors [1]. They are based on non-linear magnetization characteristics of ferromagnetic core. Fluxgate can measure fields up to 1 mT with 100 pT resolution. The basic fluxgate sensor consists of ferromagnetic core with excitation and sensing coils (Figure 14). The sensor core is periodically saturated by excitation current I_{exc} so that its permeability μ drops to 1 twice in each period. In the absence of external field, magnetisation is symmetrical and only even harmonics of the excitation frequency are present in the core flux waveform. External measured field B_0 causes assymetry in the magnetisation, which is usually detected as 2nd harmonics in voltage V_{ind} induced into the pick-up coil. The disadvantage of the basic single-core structure is that due to the transformer effect the even components of the induced voltage are high, while the measured 2nd harmonics voltage may be very low.

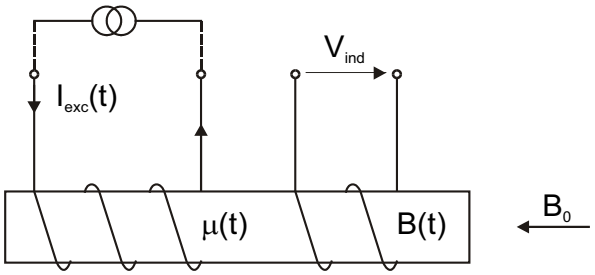


Figure 14. Basic Fluxgate Sensor - from [2].

The most usual fluxgate sensors have ring (toroidal) core, with excitation coil wound around the core and pick-up coil wound over it. There are three advantages of this sensor type:

1. the magnetic part for the excitation field is closed so that the required excitation power is low
2. seen from the pick-up coil the two core halves are excited in the opposite directions so that the net flux without the external field is close to zero.
3. symmetrical structure allows to rotate the core versus the pick-up coil and find the position, in which the effects of the core imperfections are minimized.

Figure 15 shows the typical fluxgate analog feedback magnetometer using ring-core sensor. The generator GEN produces both the excitation frequency f and its second harmonics 2f, which controls the phase-sensitive detector PSD used for processing the sensor output signal. After amplification the feedback loop is closed: the pick-up coil often serves also for the feedback. The magnetometer output V_{out} is derived from the feedback current using differential amplifier DA.

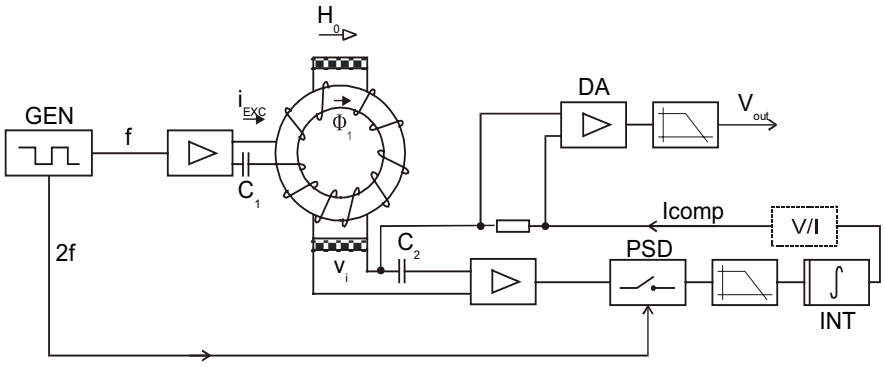


Figure 15. Fluxgate Magnetometer Using Ring-core (Toroidal) Sensor - from [2].

The integrator gives enough feedback gain so that the magnetometer linearity and gain stability are given by the properties of the feedback coil. Resulting linearity error is as low as several ppm. The typical sensitivity tempco is 30 ppm/K, but precise magnetometers can achieve 1 ppm/K. Main problem of commercially available fluxgate magnetometers is the stability of their offset with temperature and time. Typical values are 0.2 nT/K and 1 nT/day, but top fluxgate magnetometers achieved 1 nT long-term stability in the wide temperature range.

Fully digital fluxgate magnetometers have advantages, but they have not reached the performance of analog devices [17].

Micro-fluxgate sensors are still in development. Some types have flat coils [18], [19], other micromachined solenoids. Their cores are electrodeposited or made of amorphous strips. These sensors may have electronics on chip and they are potentially cheap. But at the time being their parameters are comparable with AMR sensors of similar size.

5. OTHER MAGNETIC SENSORS

5.1 SQUIDS

Superconducting magnetometers can measure fields down to fT range. They need liquid helium or at least liquid nitrogen temperatures ("high temperature" SQUIDS) [20]. SQUIDs are rather incremental than absolute sensors, as they measure field increments. Because they have small dynamic range, they can be easily overloaded by moving in the Earth field or by interfering AC fields. To suppress these effects they are mostly made as gradiometers and they often work inside magnetic shieldings.

5.2 Magnetooptic Sensors

They are based on Kerr effect, which is a field-dependent rotation of polarization plane of light in some materials. These sensors are suitable to measure only high fields [21].

5.3 GMI

High-frequency electrical impedance of ferromagnetic metal is affected by eddy currents. The penetration depth depends on the material permeability, which is dependent on the value of the measured field B. Hence the impedance depends on $|B|$. This is the basic explanation of Giant magnetoim-

pedance effect. GMI sensor are made of wires, tapes or thin layers of soft magnetic materials [22]. One of the critical aspects of the first two mentioned types is the reliability and resistance of the electrical contact. Another problem is the perming. Data on temperature offset stability of GMI sensors are rarely available.

Basic GMI sensors have even characteristics – they cannot distinguish the field polarity. "Asymmetrical GMI effect" is based on biasing of GMI sensor by remanent field of magnetically harder material (e.g. created by surface crystallization). First commercially available GMI sensors appeared in 2002.

5.4 Magnetostrictive Sensors

These sensors detect magnetic field by using magnetostrictive material [23]. Its dimensional change can be measured by optical fibre, using capacitive method or by piezoelectric material. Other sensors are based on changing the resonance frequency of free-standing elements manufactured by MEMS technology. In general, the magnetostrictive field sensors have even characteristics.

6. CURRENT SENSORS

The current measurement using shunt resistor is in some cases impractical or impossible. Contactless current sensors may be used for remote conductors at high potentials, underground cables etc. This section is based on overview of current sensors given in [24] and [25]. Traditional DC and AC contactless current sensors are available for ranges from mA to kA with precision from 3% (uncompensated Hall current sensors) to 0.1 % (compensated Hall devices and magnetic amplifiers). Higher precision is easily achievable with Current comparators. Very promising are current sensors based on AMR effect and di/dt sensors.

Besides fulfilling the requirements common to magnetic field sensors, contactless current sensors should be geometrically selective – i.e. sensitive to measured currents, and resistant against interferences from other currents and external fields. The easiest way how to guarantee this, is to use a closed magnetic circuit with a measured conductor inside. This is used in current transformers, fluxgate current sensors and in most of the Hall current sensors. If this is not possible, gradient techniques can be used. Simple integrated current sensors use folded conductor and gradient field sensor, which suppresses response from distant sources, which give low gradient. For large currents the current bar should be kept straight and circular arrays of typically four to eight sensors are being used [26]. Averaging of the sensor output increases sensitivity to the conductor between them and decreases sensi-

tivity to the conductors outside (Figure 16).

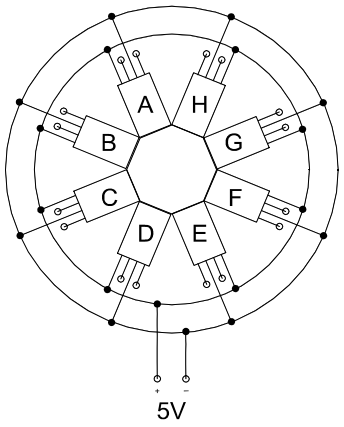


Figure 16. Current Sensor Consisting of Array of 8 Magnetoresistors – from [27].

6.1 Instrument Current Transformers

They have a primary winding with few turns (or a single conductor through the core opening) and a secondary winding, which should be ideally short-circuited, but in reality it has some small burden. The core is usually ring-shaped, either wound of high permeability tape (for precise, low-frequency devices) or made of ferrite (for high-frequency devices). Nanocrystalline materials are very promising for use in instrument current transformers [28]. Electronically enhanced two-stage current transformers show accuracy improvement by two orders of magnitude.

6.2 Rogowski Coil

Circular Rogowski coils (also called di/dt coils) may be used to measure AC or transient currents. The device is extremely linear, as it has air core. It is sensitive to di/dt, so that the output voltage should be integrated. 0.1% error from the measured value in 1000:1 dynamic range is achievable [29]. Rogowski coil with integrator can also be used to measure changes in DC current: however, the limiting factor here is the offset drift of the integrator.

6.3 Fluxgate DC Current Sensors

Fluxgate DC current sensors or “DC transformers” are similar to DC

comparators but of a much simpler design. The accuracy of a typical commercial 40 A module is 0.5 %, linearity 0.1 %, current temperature drift <30μA (-25^{0}C…70^{0}C).

Because of their low offset drift, fluxgate-based "DC current transformers" are superior to the current sensors having Hall sensor in the airgap.

6.4 Hall Current Sensors

Traditional current sensors are based on the Hall element in the airgap of a magnetic yoke. To improve the linearity, the measured current may be compensated. However, open-loop meters are preferred for battery-operated devices. One of the best feedback-compensated small-range Hall current sensors is LTS 25-NP manufactured by LEM. The device has 25 A range and DC to 200 kHz bandwidth, error is 0.02 % and sensitivity TC is 50 ppm/K. The main problem of these sensors is their limited zero stability given by the Hall sensor offset: typical offset drift of a 50 A sensor is 600 μA in the (0^0 C … 70 ^{0}C) range. This parameter is 20-times worse than that of fluxgate-type current sensor modules. Even when using magnetic yoke, Hall current sensors are sensitive to external magnetic fields and close currents due to the airgap (necessary to accommodate the sensor). Another DC error is caused by the hysteresis of the magnetic core – only few Hall current meters have AC demagnetization circuit to erase perming after overrange DC current.

6.5 Magnetoresistive Current Sensor

They are shown in Figure 17. It is based on an AMR bridge, which is made insensitive to an external field, but sensitive to measured current through the primary bus bar [30] The measured current is compensated by feedback current through compensation conductor. Typical application is galvanically isolated current sensing in PWM regulated brushless motor. These sensors are manufactured by F.W. Bell and Sensitech with ranges from 5 to 50 A. Achieved linearity is 0.1 %, temperature coefficient of sensitivity is 100 ppm/K, offset drift in the (–45^{0}C to +85^{0}C) range is 1.4% FS.

While linearization of the AMR sensors is made by using Barber poles, these GMR sensors should be biased by permanent magnet, which is a source of instability.

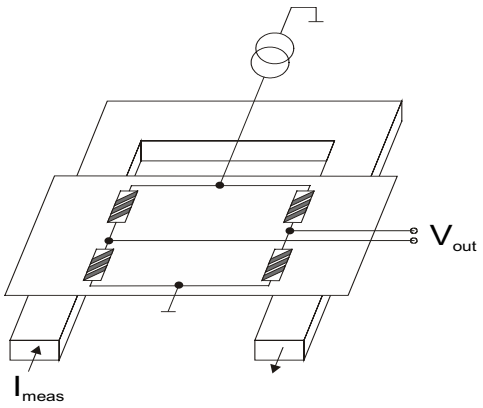


Figure17. Magnetoresistive Current Sensor–after [30].

6.6 Other Current Sensors

Magnetooptical current sensors are suitable for high-voltage high-current applications, but the reported errors are more than 1% even after temperature compensation [21], [31].

AC current clamps are usually based on current transformers with openable core. The measured conductor forms a primary winding, secondary winding is terminated by a small resistor, or connected to current-to voltage converter. Very accurate clamp current transformers use electronic compensation of the magnetization current and achieve error of 0.05% from the measured value in 1 % FS to 100 % FS.

Some of the available **DC current clamps** based on Hall sensor may have 10 mA resolution, but the maximum achieved accuracy is 30 mA, even if they are of the compensated type. Precise DC/AC current clamps based on shielded fluxgate sensor were described in [27] The sensor linearity and hysteresis error is less than 0.3% of the 40 A full-scale. The noise is 10 μA p-p, long-term zero stability is 1 mA. The main advantage is high suppression of the external currents.

Magnetometric measurement of hidden currents

Underground electric conductors can be located and their current can be remotely monitored by measuring the magnetic field in several points supposing that the back conductor is distant. 1 A conductors can be detected from 40 m distance and their position determined with 0.1 m accuracy from 4 m distance [32]. The magnetometer method is also used to measure the currents in constructions such as bridges and in pipelines.

7. APPLICATIONS OF MAGNETIC FIELD SENSORS

Magnetic position switches and rotation counters made of Hall sensors and permanent magnets are cheap devices, which are produced in billion volumes. On the other hand, some high-end applications require analog output with ppm precision.

7.1 Magnetic Position Sensors

Some magnetic position sensors have only bi-state output (position switches), the others have analog output. Magnetic encoders generate digital code. All these sensors can measure linear or angular position or speed.

Some magnetic position sensors and position switches contain permanent magnet. First group of sensors have this magnet attached to the moving object ("target"). The magnetic sensor can be inductive coil (which gives signal amplitude proportional to the speed), pulse wire (Wiegand sensor), magnetoresistor (AMR or GMR) or still most often used Hall sensor. Figure 18 shows linear sensor of this type with the output for two sensor orientations. Figure 19 shows rotation sensor using permanent magnet and two perpendicular Hall sensors.

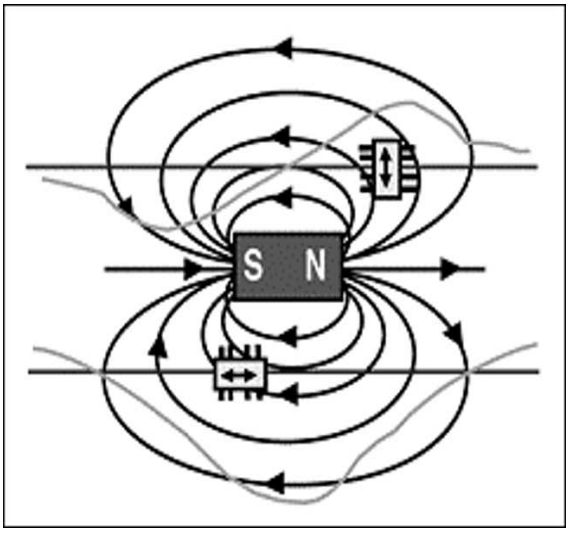


Figure 18. Linear Position Sensor (from NVE brochure [14]).

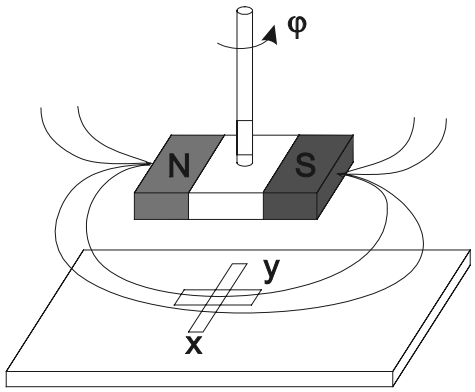


Figure 19. Rotational Position Sensor with Two Orthogonal Hall Elements (after Sentron brochure [33]).

If the target is ferromagnetic, the magnet need not attached to it, but can be a part of the sensor itself. Most important example of this type is the toothwheel sensor used in cars to control the ignition.

Inductive toothwheel sensors use the dependence of the sensor coil impedance on the position of the tooth. AC impedance is affected by eddy currents, so that the so-called eddy current sensors work for all conductive targets.

Magnetic encoders detect small permanent magnets or marks on magnetic tape or disk. Other types of magnetic position sensors such as LVDTs and Inductosyns are based on transformers with movable cores or windings.

7.2 Compass

Many solid-state magnetic compasses use only two-axis magnetic sensors. In such case the compass should be kept in horizontal position (gimballed) in order to measure horizontal part of the Earth's field and calculate the azimuth *a* correctly (Figure 20). One degree out of horizontal position may cause two degrees azimuth error. "Electronically gimballed" compasses can work in arbitrary position. They should use tri-axial magnetic sensor and two-axial inclinometer and some algebra to calculate the correct azimuth.

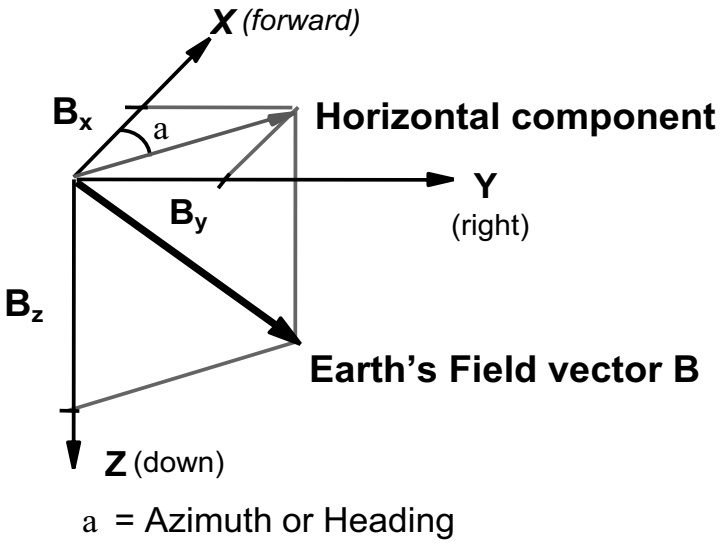


Figure 20. Gimballed Magnetic Compass Measures Two Horizontal Components of the Earth's Field.

Magnetic compasses based on fluxgate sensors may have error well below 0.1 degree, while compass based on AMR sensor can hardly reach 0.5 deg. accuracy.

7.3 Security and Military

These cHallenging applications include finding UXO (unexploded ammunition), target detection and tracking (for missiles and intelligent ammunition), submarine detection, smart navigation (also using local Earth's field anomalies), antitheft systems, and weapon detection. The task is often to find nT field signature in the 50 000 nT Earth's field background, which requires ppm precision. SQUID magnetometers cannot be used because of their small dynamic range. Fluxgates give vectorial information, but they have problem with offset stability. Resonant magnetometers are stable enough but they are scalar instruments, so the interpretation of their measurements is difficult.

7.4 Geophysics and Space Research

Magnetic sensors are used for mapping the local Earth's field anomalies, monitoring its time variations and also to measure magnetic properties of rocks.

Space magnetometers are mostly three-axial fluxgates accompanied by scalar magnetometer, which helps to keep long-term stability [34]. Deep space missions require lower field range, but higher radiation immunity. It is also important to ensure that the magnetic field from the satellite body will not influence the magnetometer on board.

7.5 Biomagnetism and Medical Applications

Tracking devices for monitoring the 3-D position and orientation sensing found potential in virtual and augmented reality systems, orthopaedics and biokinematics and also for monitoring the position of the laparoscopic surgical tools. SQUID sensors have been used for magnetocardiography (MKG) and magnetoencefalography (MEG). The main advantage of these techniques is that their interpretation is easier than in case of EKG and EEG, which register the surface electrical potentials. The reason is that the magnetic fields are more directly connected with their sources, which are electrical currents in the heart or brain. Problem of these techniques is that the measured fields are so small that they require not only sensitive multichannel gradiometer, but also expensive and bulky shielding chamber.

Another diagnostic technique is magnetopneumography: mapping the distribution of ferromagnetic particles in the lungs after they are magnetised by strong DC field. Here the measured fields are in the order of 1 nT, so fluxgate gradiometers can be used in some cases [35].

Magnetic sensors have been used for monitoring the position of magnetic markers, e.g. "magnetic biscuits" for monitoring the digestion system and microparticles for immuno-essay biosensors.

7.6 Non-destructive Testing

Both AC and DC magnetic methods are being used for testing and material evaluation [36]. While for eddy-current methods induction coils are used, DC remanence is usually measured by fluxgate [37] or AMR sensors.

8. CONCLUSIONS

Magnetic sensors are:

- Contactless;
- Robust;
- Resistant against vibrations, temperature and dirt;
- Cheap;
- Low power.

New types include:

- Integrated and vertical Hall sensors;
- Ferromagnetic Magnetoresistors: AMR, GMR and SDT.

Fluxgates are the most precise vectorial sensors, followed by AMRs, which become great competitors of Hall sensors. Until now, GMR and SDT sensors are rather nonlinear. GMI sensors are recently commercially available, but their present accuracy is not competitive.

Common problem of magnetic sensors containing ferromagnetic material is perming (remanence). This was solved only with fluxgates and flipped AMRs.

ACKNOWLEDGMENTS

Section 6 (Current sensors) is based on paper [24], material used with kind permission of the journal editor.

REFERENCES

[1] Ripka, Pavel (ed), *Magnetic sensors and Magnetometers*, Boston –London: Artech, 2001.
[2] Ripka P. Advances in Fluxgate sensors. Sensors & Actuators A 2003; 106: 8-14
[3] Vopalensky M., Ripka P., Platil A. Precise Magnetic Sensors. Sensors & Actuators A 2003; 106: 38-42
[4] Brauer P., et al. Transverse field effect in fluxgate sensors. Sensors and Actuators A 1997; 59: 70-74
[5] Popovic, R. S., *Hall effect devices*, Bristol – Philadelphia, IoP 2004
[6] http://www.murata.com/sensor/index.html
[7] Heremans J. Narrow – gap semiconductor magnetic field sensors and applications. Semiconductor Sci. Technol. 1993; 8: S424-S430
[8] Tumanski, S. *Thin Film Magnetoresistive Sensors*, Bristol – Philadelphia, IoP 2001.
[9] Ripka P. Noise and Stability of Magnetic Sensors. J. Magn. Magn. 1996; 157/158: 424-427
[10] Dimitropoulos P.D., Avaritsiotis J.N., Hristoforou E. A novel micro-Fluxgate sensor based on the AMR effect of ferromagnetic film-resistors, Sensor Actuat A-Phys 2003; 107: 238-247

[11] Smith C. H., Schneider R.W. Low Magnetic Field Sensing with GMR sensors. Sensors Magazine 1999; 9, 10. Also at http://www.sensorsmag.com
[12] Wakiwaka H, Kataoka Y, Shinoura S. Directional sensitivity of GMR element and its industrial applications. Int J Appl Electrom 2001; 15: 89-96
[13] Vopálenský M., Ripka P., Kubík J, Tondra M. Improved GMR Sensor Biasing Design. Sensors and Actuators 2003; 110: 254-258
[14] http://www.nve.com
[15] Jander A., Nordman C.A., Pohm A.V., et al. Chopping techniques for low-frequency nanotesla spin-dependent tunneling field sensors. J Appl Phys 2003; 93: 8382-8384
[16] Vopálenský M., Ripka P., Kubík J., Tondra, M. Alternating Biasing of SDT Sensors. Sensors and Actuators A 2003; 110: 182-186
[17] Magnes W., Pierce D., Valavanoglou A., et al. A sigma-delta fluxgate magnetometer for space applications Meas Sci Technol 2003; 14: 1003-1012
[18] Kawahito S., Cerman A., Aramaki K., et al. A weak magnetic field measurement system using micro-fluxgate sensors and delta-sigma interface. IEEE T Instrum Meas 2003; 52: 103-110
[19] Almazan R.P., Perez L., Aroca C., et al. Magnetometric sensors based on planar spiral coils. J Magn Magn Mater 2003; 254: 630-632
[20] Yamagata K., Hayashi N, Kezuka H, et al. Characteristics of a highly sensitive magnetic sensor made from a thick BPSCCO film Solid State Sci 2003; 5: 441-444
[21] Didosyan Y.S., Hauser H., Nicolics J. Magneto-optical current sensor of high bandwidth. Sensors and Actuators A 2000; 81: 263-267
[22] Knobel M., Pirota K.R. Giant magnetoimpedance: concepts and recent progress. J Magn Magn Mater 2002; 242: 33-40
[23] J.L. Prieto, P. Sanchez C. Aroca, E. Lopez, M.C. Sanchez, O. de Abril, L. Perez Improving the characteristics in magnetostrictive-piezoelectric sensors when the viscous interface is removed. Sensors and Actuators 2000; 84:338–341
[24] Ripka P. Current sensors using magnetic materials, to appear. Journal of Optoelectronics and Advanced Materials 2004
[25] Iwansson K., Sinapius G., Hoornaert W., *Measuring Current, Voltage and Power*, Elsevier, 1999.
[26] Di Rienzo L., Bazzocchi R., Manara A. Circular arrays of magnetic sensors for current measurement. IEEE T Instrum Meas 2001; 50: 1093-1096
[27] Kejík, P, Ripka P, Kašpar K, Draxler, K. Precise DC current sensors, IMTC Conference, Brussels 1996
[28] Draxler, K., Styblikova R. Use of nanocrystalline materials for current transformer construction. Journal of Magn. Magn. 1996; 157/158:447-448
[29] Koon William, Current Sensing for Energy Metering, http://www.analog.com/Analog_Root/static/technology/dataConverters/training/meterBackground/currentSensors/Current_sensing_for_metering.pdf
[30] Drafts B. New Magnetoresistive Current sensor Improves Power Electronics Performance. Sensors, September 1999, also on http://www.sensormag.com/articles/0999/84\
[31] Cruden A. et al. Optical current measurement system for high-voltage applications. Measurement 1998; 24; 97-102
[32] Takagi S., Kojima J., Asakawa K. DC cable sensors for locating underwater telecomunication cables., Proc. IEEE Oceans 96 conf., New York 1996
[33] http://www.Sentron.ch
[34] Acuna M.H. Space-based magnetometers. Rev Sci Instrum 2002; 73: 3717-3736
[35] Ripka P., Navratil P. Fluxgate sensor for magnetopneumometry. Sensors and Actuators A 1997; 60: 76-79.
[36] Vertesy G., Gasparics A. Nondestructive material evaluation by novel electromagnetic methods. Mater Sci Forum 2003; 414-4: 343-352
[37] Gruger H. Array of miniaturized fluxgate sensors for non-destructive testing applications. Sensor Actuat A-Phys 2003; 106: 326-328

Chapter 7

MICRO-HALL MAGNETIC SENSORS: PHYSICS, TECHNOLOGIES, AND APPLICATIONS

Radivoje S. Popovic, Giovanni Boero, and Pierre-Andre Besse
EPFL-Swiss Federal Institute of Technology Lausanne
CH-1015 Lausanne, Switzerland

Abstract: This chapter describes the Hall plates with the planar dimensions from micro-meter to deep sub-micrometer range. In the theoretical part, the physical phenomena in such micro-devices are described. Scaling down the dimensions of a Hall plate, the prevailing mechanism of electron transport changes. In macroscopic devices stationary transport takes place. Depending on the device size and applied voltage, the drift velocity of carriers may be linearly dependent on the voltage or saturated. Scaling down further, the device enters the non-stationary transport range, where velocity overshoot and ballistic transport occur. In each regime magnetic sensitivity and noise of the device are estimated. In the experimental part, a review of the published results on the technologies, performance, and applications of micro-Hall devices is given. In summary, scaling down a Hall sensor brings about a degradation of its basic characteristics. On the other hand, micro-Hall devices open the way to applications where high spatial resolution is necessary, like in scanning Hall probe microscopy.

Keywords: Magnetic sensors, micro-Hall, scaling down, sensitivity, noise, resolution.

1. INTRODUCTION

The standard way of measuring magnetic fields in the range between 0.1 mT (milli-tesla) to several 10ths of Tesla is by using a Hall magnetic sensor. The technology of Hall magnetic sensors is today well established and broadly used both in scientific research and in industry. Several books are devoted to Hall magnetic sensors (see, e.g., the recent book [1]).

The dimensions of conventional Hall magnetic sensors are generally chosen according to the following two criteria: on the one hand, a Hall

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 229-253.

device should be small enough as to not essentially influence the overall production costs of the Hall sensor system; on the other hand, the dimensions of the Hall device should be big enough as to be not sensitive to production tolerances. A compromise of these two criteria currently gives the dimensions of conventional Hall magnetic sensor of the order of 100 μm. The corresponding spatial resolution of the measured magnetic field is then also of this order, which is more than enough for most practical applications.

But there are some magnetic measurement problems in engineering and science that cannot be solved by conventional Hall devices. For example, with a conventional Hall magnetic sensor we cannot map the magnetic field around a micrometer-size magnetic particle. For such an application we need a Hall device of dimensions comparable (or smaller) respect to the dimensions of the source of the magnetic field to be measured.

In this chapter we shall treat such very small Hall magnetic sensors. We call them *micro-Hall magnetic sensors*, where *micro* means – much smaller than conventional, i.e., in the micrometer and sub-micrometer range. We shall first briefly review some elements of the theory of conventional Hall devices that we need as a basis for further analysis. Then we shall try to foresee what happens with the characteristics of a Hall device if we start decreasing its dimensions. Finally, we shall look at several examples of miniaturized Hall magnetic sensors and their applications. Because of the limited space, in the theoretical part we shall avoid the explanation of many details by making reference to the corresponding sections of the book [1].

2. CONVENTIONAL HALL PLATES

A Hall magnetic sensor has usually a form of a Hall plate (&4 of [1]). A Hall plate is a thin semiconductor plate (or a part of a semiconductor layer) fitted with four ohmic contacts at its periphery (Figure 1). If a current-carrying plate is exposed to a perpendicular magnetic field $\boldsymbol{B}$, then across the plate appears a Hall voltage V_{H}. The Hall voltage is the output signal of the sensor, which gives an analog interpretation of the measured magnetic field.

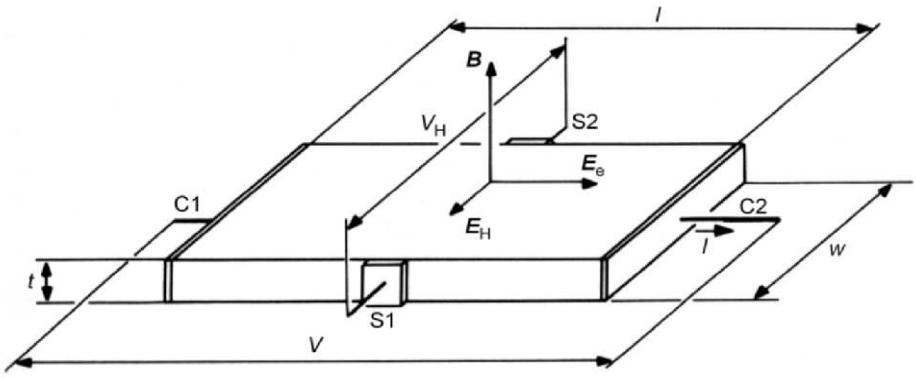


Figure 1. Structure and operation of a Hall plate. A bias voltage V is applied to the plate via the two current contacts C1 and C2. The bias voltage creates an electric field $\boldsymbol{E}_e$ and forces a current $\boldsymbol{I}$. If the plate is exposed to a perpendicular magnetic induction $\boldsymbol{B}$, the Hall electric field $\boldsymbol{E}_H$ occurs in the plate. The Hall electric field gives rise to the appearance of the Hall voltage $\boldsymbol{V}_H$ between the two sense contacts S1 and S2.

The basic effect leading to the appearance of a Hall voltage in a Hall plate is the Hall effect, discovered in 1879 [2]. The fundamental electromotive force generating the Hall voltage is the Lorentz force, which acts on each charge carrier:

$$\mathbf{F} = q\mathbf{E} + q(\mathbf{v} \times \mathbf{B}) \tag{1}$$

Here q denotes the particle charge (for electrons $q=e$ and for holes $q=-e$, where $e \cong -1.6 \times 10^{-19}\,\mathrm{C}$), $\boldsymbol{E}$ is the electric field, $\boldsymbol{v}$ the carrier velocity and $\boldsymbol{B}$ the magnetic induction. The magnetic component of the Lorentz force, $e[\boldsymbol{v} \times \boldsymbol{B}]$, pushes the charge carriers in the lateral direction (along the plate width W). This gives rise to the creation of the Hall electric field,

$$\boldsymbol{E}_H = -(\boldsymbol{v}_\mathrm{d} \times \boldsymbol{B}), \tag{2}$$

where $\boldsymbol{v}_\mathrm{d}$ denotes the drift velocity of carriers. The Hall voltage is then given by

$$V_H = \int_{S1}^{S2} \boldsymbol{E}_H \, d\boldsymbol{w}, \tag{3}$$

where S1 and S2 denote the positions of the sense contacts, see Figure 1.

2.1 Characteristics of Hall Magnetic Sensors

Hall plates are usually described by several "figures of merits". We shall here briefly review some of such characteristics.

Absolute Sensitivity of a Hall magnetic sensor is the ratio:

$$S_A = \left| \frac{V_H}{B_\perp} \right|_C , \tag{4}$$

Here V_H is the Hall voltage, $B_\perp$ is the normal component of the magnetic induction, and C denotes a set of operating conditions. For future reference note that, if we assume that the Hall electric field is constant over the plate width, from (2) – (4) we obtain

$$S_A \cong W \left| v_d \right| , \tag{5}$$

where W denotes the plate width.

To compare different sensors one usually uses the **current related sensitivity** S_I. The current related sensitivity is the absolute sensitivity normalized (divided) by the bias current I.

Offset-Equivalent Magnetic Field is the magnetic field, which would yield at the output of a perfect Hall plate a Hall voltage equal to the measured offset voltage. Using (4), we calculate the offset-equivalent magnetic field as

$$B_{off} = \frac{V_{off}}{S_A} , \tag{6}$$

where V_{off} denotes the offset voltage. The offset voltage is proportional to the input voltage and quasi independent on the magnetic field. Offset is caused by geometrical and electrical asymmetries of the Hall plate.

Noise-Equivalent Magnetic Field Spectral Density is the magnetic field spectral density, which would yield at the output of a noiseless Hall plate the measured voltage noise spectral density. It is given by

$$S_{NB} = \frac{S_{NV}(f)}{S_A^2} , \tag{7}$$

where $S_{NV}(f)$ is the noise voltage spectral density at the sensor output. For the signal-to-noise ratio (SNR) equal to 1, equation (7) gives the **magnetic**

field resolution per hertz. The magnetic field resolution is also often given by average noise-equivalent magnetic field density (in $\mathrm{T}/\sqrt{\mathrm{Hz}}$): $B_{\min} = \sqrt{S_{NB}}$.

At high-enough frequencies, noise in a Hall device is usually dominated by the thermal noise

$$S_{NV} \cong 4kTR_{\text{out}}, \tag{8}$$

Here k is the Botzmann constant, T is the absolute temperature, and R_{out} is the output resistance of the Hall device. At low-enough frequencies, noise is dominated by the $1/f$ noise. It can be shown that

$$S_{NV} \cong \left(\frac{V_{\text{in}}}{g}\right)^2 \frac{\alpha}{N}\frac{1}{f}, \tag{9}$$

Here V_{in} denotes the input voltage of the Hall device, α is the Hooge parameter ($\alpha \leq 10^{-3}$)(&2.5.3 of [1]), g is a geometrical factor (usually, $g \approx 3$), and N is the number of charge carriers in the active region of the Hall plate. If the1/f noise dominates, then the rms noise equivalent magnetic field (resolution for SNR = 1) in a frequency region is given by

$$B_{Nf} \cong \left(\frac{S_1}{S_A^2}\ln\frac{f_2}{f_1}\right)^{1/2}, \tag{10}$$

where $S_1=(V_{\text{in}}/\mathrm{g})(\alpha/N)$, and f_1 and f_2 are the limits of the frequency range of interest.

Cutoff Frequency of a Hall magnetic sensor is the frequency at which its magnetic sensitivity is for 3 dB lower than at low frequencies ($S_{\text{A}}(f_{\text{C}}) = 0.707\ S_{\text{A}}(0)$). The magnetic sensitivity of a Hall plate may depend on frequency due to several effects (&4.6 of [1]). If we neglect the trivial parasitic effects, the cutoff frequency is generally given by

$$f_{\text{C}} = 1/(2\pi\tau), \tag{11}$$

where τ denotes the dielectric relaxation time of the Hall plate. Depending on the plate dimensions and the materials, the dominant dielectric relaxation process may take place either in the plate or around it. If the plate is thick, made of a high-permittivity material, and placed in vacuum,

then the relaxation parameter in (11) is usually determined by the dielectric relaxation time of the plate material, which is given by

$$\tau = \frac{\varepsilon}{\sigma}, \tag{12}$$

where ε and σ denote the permittivity and the conductivity of the plate material, respectively. If, however, the plate is thin and made of a low-permittivity material, but immersed in a high-permittivity dielectric, then the dielectric relaxation is determined by the so-called global capacitance of the plate. Then the relaxation time in (11) is dominated by the so-called *global relaxation time of the plate*, given by

$$\tau_j \cong \frac{\varepsilon_e}{\sigma}\frac{L}{d}\frac{2}{\pi} \tag{13}$$

Here ε_e denotes the external permittivity (of the material surrounding the Hall plate), σ is the conductivity of the plate material, and L and d are the plate length and thickness, respectively.

3. SCALING DOWN A HALL PLATE

Consider now the operation of a Hall plate while reducing all its planar dimensions by a scaling factor K (>1) (&5.5.1 of [1]). While decreasing the dimensions of a Hall plate, we shall see that different physical effects predominantly determine its characteristics at different size ranges. We shall name these size ranges by the usual type of electron transport in a Hall plate of this size range. In *Appendix A* we give a short summary of the carrier transport phenomena in semiconductors.

3.1 Stationary Transport Range

We shall first consider the case when the charge carriers in a Hall plate stay under given conditions (electric field, density of scattering centers, temperature) during enough time to reach steady-state distributions both in terms of momentum and energy. “Enough time” means – enough to make many collisions with scattering centers. Then the charge carriers in average either move with the drift velocity linearly proportional with the driving electric field, or, at higher electric fields, their drift velocity reaches saturation. This transport range is sometimes also called

"diffusive" transport, although it has little to do with the conventional diffusion phenomena.

Linear Drift Range

The linear drift range is the size range appropriate to the conventional Hall devices analyzed up to now. The minimal dimensions of a Hall plate belonging to this size range might be a few micro-meters for a low-mobility material, or several tens of micrometers for a high-mobility material. The essential feature of the charge carrier transport in this size range is that the carrier drift velocity is linearly proportional to the applied electric field $\boldsymbol{E}$:

$$\boldsymbol{v}_{\mathrm{d}} = \mu \boldsymbol{E} \,, \tag{14}$$

where μ denotes the carrier mobility.

We shall make at this point the following approximation for the driving electrical field in the active region of a Hall plate (between the sense contacts):

$$E \cong V/L \,, \tag{15}$$

where V denotes the input voltage and L denotes the plate length. This approximation is for example reasonable for a long perpendicular Hall plate (figure 1) with very small Hall contacts. But note that this is only an approximation: the presence of the sense contacts, due to their short-circuiting effect, brings about a decrease of the local driving electric field – see also S.4 below. This approximation implies in particular that the Hall device equations of our interest can be written without the geometrical factor of Hall voltage G_{H} (&4.2.6 of [1]) (i.e., $G_{\mathrm{H}} \approx 1$).

According to (5), (14) and (15), the Hall voltage is proportional to the input voltage. But in practice, the maximal input voltage of a Hall device is limited. For example, in certain integrated circuits the voltage is kept constant at 5V. Therefore, we consider the scaling down in the linear drift region under constant voltage. Then the applied electric field as well as the drift velocity of carriers is inversely proportional to the length of the sensor. If we reduce all planar dimensions of a Hall plate working in the linear drift range by the scaling factor K while keeping the input voltage constant, we shall have the following consequences on its main characteristics:

L.1) Sensitivity: does not change, see (5);

L.2) Input current, input and output resistances: do not change;

L.3) Offset-equivalent magnetic field: increases proportionally with K, if we assume that the main cause of offset are the tolerances in the dimensions;
L.4) Noise-equivalent magnetic field (7): (i) in the white (resistive) noise range, it does not change since the resistance and the sensitivity remain constant; but (ii) in the 1/f noise range, noise-equivalent magnetic field increases proportionally with K, because the 1/f noise spectral density is inversely proportional with the number of carriers N (9), which scales as 1/K2. Therefore, 1/f noise may stay dominant until very high frequencies, in the hundreds MHz range.
L.5) Cutoff frequency: since we do not scale down the plate thickness, than fC increases at least proportionally with K (see (11), (13), and Appendix A).

If we keep increasing the scaling factor K, we shall eventually come to a limit of the linear drift range. Depending on the details of the design of the Hall plate under consideration, we may reach first the limit due to overheating or due to velocity saturation.

Overheating. In the above-described scaling approach, we kept the total power dissipated in the Hall plate constant, while we decreased the plate surface by the factor K^2. If we do nothing to improve the heat removal from the Hall device, the difference between the Hall device temperature (of both the crystal lattice and the charge carriers) and the ambient temperature shall increase roughly proportionally with K^2. An increase in temperature brings about a reduction of carrier mobility, and, eventually, onset of intrinsic conditions of the active semiconductor layer.

Velocity saturation. By scaling down a Hall plate while keeping the input voltage constant, we increase the excitation electric field in it. Then the electric field shall eventually reach a value at which the drift velocity of charge carriers reaches saturation. Now if we continue reducing the Hall plate dimensions, the functional relationships between the dimensions and the sensor characteristics become quite different than those valid in the linear drift range.

Saturated Velocity Range

The electron drift velocity reaches a maximal value in GaAs at an electric field of about 3 kV/cm and in Si at about 30 kV/cm (&2.4.2 of [1]). At a bias voltage of 5 V, this corresponds to the lengths of the Hall plate of about 15 μm and 1.5 μm, respectively. If we continue scaling down a Hall plate below these dimensions while keeping the input voltage constant, the carrier drift velocity will not increase any more. Consequently, the Hall electric field shall not increase either, since it is proportional to the carrier velocity (5). So it is not reasonable any more to keep the bias voltage constant. Instead, in this size range, we should better adopt the constant-field scaling rules, developed for MOS transistors [3].

This means: while reducing all planar dimensions of a Hall plate by a scaling factor K, we also reduce the input voltage by the same factor. Then both excitation electric field and Hall electric field stay independent on K. We expect that scaling down in the saturated drift velocity range shall have the following consequences on the main characteristics of a Hall plate working in the velocity saturation range:

S.1) Sensitivity: decreases proportionally with $1/K$. This is so because now the Hall electric field does not depend on K, whereas the integration path S1, S2 for Hall voltage in figure 1 scales down as $1/K$. In other words, in (5), the velocity is constant (the saturation velocity) and the sensitivity becomes proportional to the sensor's width.

S.2) Input current: decreases proportionally with $1/K$. This is so because we assume a constant carrier density, the carrier velocity is saturated, and the cross-section of the plate scales down as $1/K$.

S.3) Input dynamic resistance: tends to infinity. If drift velocity is saturated, current does not depend any more on the applied voltage. So $\mathrm{d}I/\mathrm{d}V \to 0$, and consequently, $R \to \infty$.

S.4) Output resistance: only slightly increases. This is so because in a Hall plate the electric field between the sense contacts is lower than that further away from the sense contacts (see the comment following equation (15) above and figure 5.19 in [1]).

S.5) Noise: (i) the white (resistive) noise slightly increases because of S.4; but (ii) the 1/f noise further increases proportionally with K, much as above in the case (L.4 - ii). The noise is to be translated into the noise-equivalent magnetic field by taking into account that the sensitivity decreases proportionally with $1/K$. Therefore the noise-equivalent magnetic field increases slightly more than linearly with K in the case of white noise and quadratically with K for the 1/f noise.

S.6) Cutoff frequency: increases, as above in the case L.5), as long as we do not reach a limit imposed by the intrinsic dielectric relaxation (12).

The conclusions S.1) and S.5)(ii) imply that scaling a Hall plate beyond the point of drift velocity saturation brings about a rapid degradation of its magnetic field resolution.

3.2 Non-Stationary Transport Range

If we still keep scaling down a Hall plate, then the number of successive collisions that an electron makes on its way between the two current contacts shall also decrease. Eventually, more and more electrons shall make the trip between the current contacts with only a few, or even without any collisions. Then we enter the non-stationary transport range.

Recall the two most important phenomena of non-stationary transport: velocity overshoot and ballistic transport - see Appendix A.

Velocity Overshoot Range

Briefly, velocity overshoot means a swift increase in the drift velocity of charge carriers well above the value of stationary saturation drift velocity. This occurs when a Hall device is so small, and the applied voltage is so high, that electrons experience just a few collisions while passing the distance between the current contacts. For example, this may be the case in a high mobility 2DEG Hall plate, about 5 μm long, working in cryogenic temperature range and biased by a voltage of about 5 V.

Assume that we have a Hall plate working just between the saturated velocity range and the velocity overshoot range. Then a further scaling down the plate planar dimensions results in an increase of the velocity, which may compensate or even overcompensate the decrease of the sensor width – see (5).

Ballistic Transport Range

Finally, let us assume that the dimensions of a Hall plate are so small that we can treat electrons as little charged billiard bolls that can "fly" trough it without any collision. To be definite, let us consider a **ballistic Hall bar** shown in figure 2. We assume that the material of the bar and the operating conditions are such that the length of the plate L is much smaller than the mean path of electrons between collisions. Now the notion of the plate resistance looses its sense and we assume that the device is designed in such a way that the source S limits the device current to a reasonable value. We also assume that there is no space charge effect in the plate so that the external biasing voltage V produces a constant electric field $\boldsymbol{E}$ over the length of the plate.

Hall Voltage. Generally speaking, the mechanism of the generation of a Hall voltage in the device shown in figure 2 is the same as that acting in the device shown in figure 1. With reference to figure 2, the electrons move due to the biasing electric field along the x axis with a velocity $\boldsymbol{v}$. If the device is exposed to a magnetic field $\boldsymbol{B}$ along the y axis, then it appears a Lorentz magnetic force $\boldsymbol{F}_{\mathrm{m}}$, which tends to "press" the electrons against the upper border of the bar. The "accumulated" charges (i.e., flying electrons with a slightly higher density) at the borders create a Hall electric field $\boldsymbol{E}_{\mathrm{H}}$. The Hall electric field acts on the electrons by a Lorentz electric force $\boldsymbol{F}_{\mathrm{e}}$, which exactly compensates the action of the magnetic force $\boldsymbol{F}_{\mathrm{m}}$. The result is that the electrons move parallel with the x axis, as they would without a magnetic field.

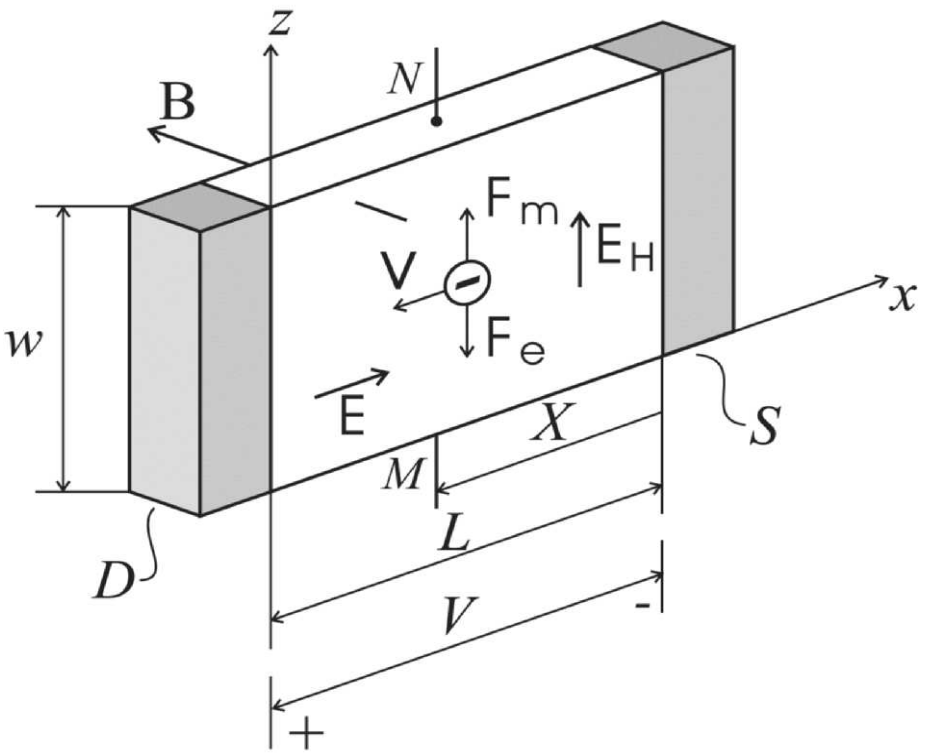


Figure 2. A model of a ballistic Hall plate. Ballistic electrons are emitted the source S and collected by the drain D; M, N are the points for the retrieval of the Hall voltage.

The fundamental difference between the classical Hall effect (Figure 1) and the ballistic Hall effect (figure 2) is the following detail: whereas in the classical (drift) case, the carrier drift velocity $\boldsymbol{v}_d$ is constant along the bar, in the ballistic case, the ballistic electron velocity $\boldsymbol{v}$ increases with the distance from the source X. As a consequence, whereas the classical Hall electric field is essentially constant along the length of the bar, the **ballistic Hall electric field** increases with the distance from the source electrode.

A ballistic Hall device normally operates at a low temperature. Accordingly, it is reasonable to assume that electrons leave the source with zero initial velocity. Then they get the acceleration

$$\boldsymbol{a} = \frac{e}{m^*}\boldsymbol{E} \tag{16}$$

and after a fly-time t achieve the velocity

$$\boldsymbol{a} = \frac{e}{m^*}\boldsymbol{E} \tag{17}$$

Here e and m^* denote the charge and the effective mass of the carriers, respectively, and E is the accelerating electric field (15).

We can now estimate the ballistic Hall electric field by replacing in (5) the drift velocity $\boldsymbol{v}_d$ by the ballistic velocity (17). The result is

$$\boldsymbol{E}_{\mathrm{Hb}} \cong \frac{e}{m^*} t(\boldsymbol{E} \times \boldsymbol{B}) \tag{18}$$

The fly-time can be expressed in terms of the fly distance X:

$$t = \sqrt{\frac{2X}{E}\frac{m^*}{e}} \tag{19}$$

Now (18) becomes

$$\boldsymbol{E}_{\mathrm{Hb}} \cong \sqrt{\frac{e}{m^*}\frac{2X}{E}(\boldsymbol{E} \times \boldsymbol{B})} \tag{20}$$

Suppose that the magnetic field $\boldsymbol{B}$ is normal to the plate and the probe points M, N are situated at the middle of the bare, i.e., $X = L/2$, Then, by integrating the Hall electric field (20) between the probe points (3), we calculate the **ballistic Hall voltage** as

$$V_{\mathrm{Hb}} \cong \sqrt{\frac{e}{m^*}V} \cdot W\ B \tag{21}$$

Here V denotes the bias voltage, $V = E\,L$.

Note that we cannot increase the bias voltage at will: the limit here is given by the occurrence of the impact ionization effect. When the carrier kinetic energy is greater than the band gap energy, electron-hole pairs can be produced by ionization. This implies that the maximum applied voltage is held constant at the value $V=E_g/q$. In this case, the magnetic sensitivity of a ballistic Hall device is proportional with the device width W.

Noise. The good news about a Hall plate working in the ballistic regime is that it may have a small noise. Suppose that the device structure

and technology are so perfect that we may neglect 1/f noise. Then due to the nature of the carrier transport and to the fact that ballistic Hall device normally operates at a low temperature, we now expect to see predominantly *shot noise*.

The bias current I can be decomposed in single electrons flowing through the device at the rate I/q. Each electron produces a voltage difference between the two points M and N for a period corresponding to the transit time T_{trans}. From (19) we obtain the transit time as:

$$T_{trans} = \sqrt{\frac{2\,m^*}{V\,q}}\,L \tag{22}$$

In the time domain, this basic voltage pulses can be approximated by:

$$\partial V(t) = \Delta\,\frac{q}{\varepsilon\,W} \quad \text{for } t \in \left[0, T_{trans}\right] \text{ and } \partial V(t) = 0 \quad \text{elsewhere} \tag{23}$$

The factor Δ accounts for the electron trace position, averaged over all the electron's traces. Namely, the electrons flowing in the center of the device do not produce a voltage unbalance, in contrary to the electrons flowing near the borders. ε is the high-frequency permittivity of the sensor material.

In the frequency domain, the voltage noise spectral density should be given by:

$$S_{NV,out} = 2\cdot\left(\frac{I}{q}\right)\cdot\left(\Delta\frac{q}{\varepsilon\,W}\,T_{trans}\right)^2 = 4\,\Delta^2\,I\,\frac{m^*}{\varepsilon^2\,V}\cdot\left(\frac{L}{W}\right)^2 \tag{24}$$

The first term represents the rate of the events, the second is the square of the Fourier transform of each single event.

Therefore, the voltage shot noise spectral density (24) at the output of a ballistic Hall plate is proportional with the device current. On the other hand, according to the above simple model, the Hall voltage does not depend on the device current at all (21). Therefore, it should be possible to achieve a reasonable magnetic resolution of a ballistic Hall device just by reducing the emissivity of the source S.

Cutoff frequency. In analogy with the case of other very small electron devices, we think the cutoff frequency of a ballistic Hall device may be determined either by the carrier transit time (19) or by the dielectric relaxation time (12) or (13), whichever is longer.

3.3 Summary and Discussion

We may summarize the results of our analysis of scaling down of Hall plates as follows:

- Depending on the predominant nature of carrier transport, the magnetic sensitivity of a Hall plate either stays more or less constant, or decreases proportionally with the plate dimensions, as shown in Figure 3.

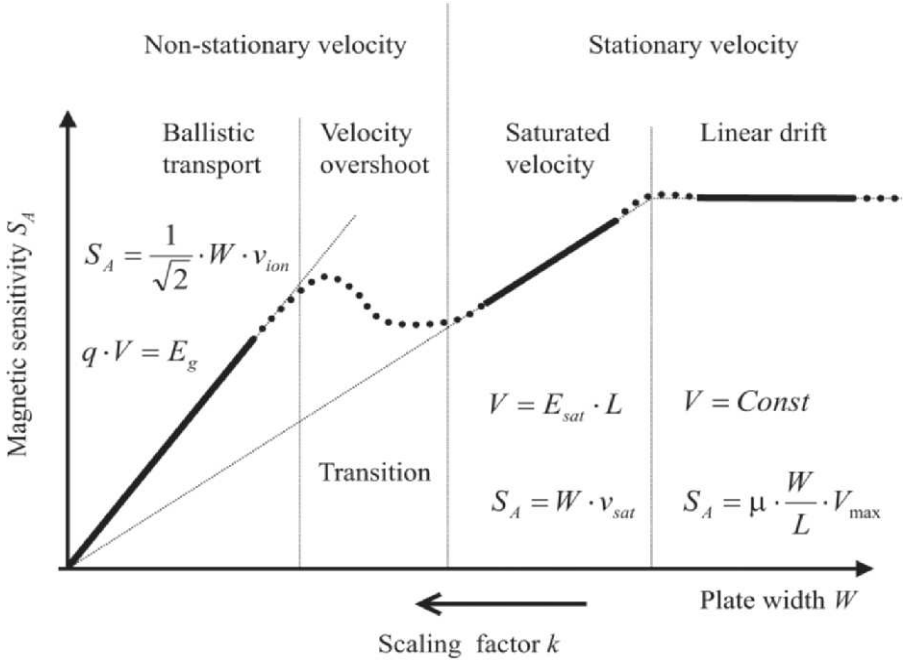


Figure 3. A qualitative piecewise linear approximation of the dependence of Hall plate magnetic field sensitivity on the plate planar dimensions and the scaling factor K. Plate dimensions are represented by the plate width W; $K \propto 1/W$.

- In the stationary transport range, the resistive noise stays almost constant, but the 1/f-noise increase with the scaling factor K. The corner frequency of the 1/f-noise shifts towards higher frequencies by scaling down.
- Therefore, magnetic resolution of a Hall plate in the stationary transport range generally quickly deteriorates with scaling down.
- In the ballistic transport range, we expect that shot noise dominates and the final outcome for magnetic resolution might not be so bad.

By and large, we think that the most important deterioration of the magnetic field resolution of a scaled down Hall magnetic sensor comes about because of the strong increase of the 1/f noise. Although above we neglected 1/f noise in the realm of the ballistic transport, experience teaches us that, for one or another reason, 1/f noise always exists. Therefore, we think that the application of an 1/f noise reduction techniques, such as "spinning current" (&5.6.3), is mandatory while working with micro-Hall magnetic sensors. Thereby the chopping frequency must increase as the size of the Hall devices decreases. Fortunately enough, this requirement is compatible with the fact that (usually) the cutoff frequency increases with the scaling down of a Hall device.

4. EXAMPLES OF MICRO-HALL DEVICES

Hall devices with dimensions down to $(100\ \text{nm})^2$ have been realized with semiconductors (e.g., Si [6], GaAs [7], InSb [8,9]) semiconductors heterostructures (e.g., GaAs/AlGaAs [10], InGaAs/InP [11], Si/SiGe [12], In/(Al,Ga)Sb [13], InAlAs/InGaAs [14]), semimetals (Bi [15,16]), and metals (Al, Au [17]).

Silicon Hall sensors are usually fabricated using low cost IC technologies. GaAs, InSb, and heterostructures sensors are realized structuring thin films, deposited by molecular beam epitaxy (MBE) or metallo-organic chemical vapor deposition (MOCVD), by optical, electron beam and ion beam lithography. Bi thin films are obtained by evaporation [15,16,18] or electrodeposition [19], and structured using techniques similar to those used for semiconductor devices. GaAs devices have been realized also by maskless Si ion implantation into semi-insulating GaAs substrates [7].

Table 1 summarizes the typical key features of micro-Hall sensors. The magnetic field resolution is significantly worse with respect to their larger counterparts, whereas the exceptionally wide frequency, temperature and magnetic field operating ranges are preserved.

Table 2 reports the characteristics of several micro-Hall cross sensors at 300 K estimated from published results. Since the *1/f* noise is linearly dependent on the bias current, the bias current is often set to the value at which the *1/f* noise becomes larger than the thermal noise at the frequency of interest. This current is usually well below the limits imposed by carrier velocity saturation and overheating. At temperatures below 77 K, micro-Hall sensors having magnetic field resolutions better than $10\,\text{nT}/\sqrt{\text{Hz}}$ have been reported [56, 57].

Table 1. Typical Characteristics of Micro-Hall Sensors.

Feature	Value
Active area	$(100\ \text{nm})^2$ to $(10\ \mu\text{m})^2$
Resistance	10 Ω to 100 kΩ
Magnetic field resolution	$1\ \text{nT}/\sqrt{\text{Hz}}$ to $100\ \mu\text{T}/\sqrt{\text{Hz}}$
Bandwidth	DC to 1 GHz
Operating temperature	1 K to 500 K
Operating magnetic field	0 T to 30 T

Table 2. Room Temperature Characteristics of Some Micro-Hall Sensors. The bottom row reports the bias current at which the magnetic field resolution is evaluated. The magnetic field resolution is defined as: $B_{\min} = \sqrt{S_{NV}} / S_A$

Type	**Bi**	**GaAs**	**2DEG**	**2DEG**	**2DEG**	**2DEG**	**2DEG**	**Si**	**GaAs**	**InSb**
Area (μm^2)	$(0.1)^2$	$(0.3)^2$	$(0.4)^2$	$(0.4)^2$	$(0.8)^2$	$(1.5)^2$	$(2.0)^2$	$(2.4)^2$	$(4.0)^2$	$(4.5)^2$
S_I (V/AT)	3.3	30	230	230	300	700	350	87	3100	140
B_{min} ($\text{nT}/\sqrt{\text{Hz}}$)	70000	130	10000	180	4000	300	5	60	10	4
@ I (mA)	0.04	2	0.001	0.3	0.003	0.1	2	2	0.3	3

5. APPLICATIONS OF MICRO-HALL DEVICES

Micro-Hall sensors are employed to perform magnetization measurements of magnetic particles deposited on their active area or as scanning probes to measure the magnetic field on a surface (a method usually referred as scanning Hall probe microscopy). In particular, miniaturized Hall sensors are employed to measure highly inhomogeneous magnetic fields such as those produced by ferromagnetic nanoparticles [22, 24], vortices in superconductors [25, 26], magnetic domains in ferromagnetic materials [27-29] (see also Figure 4 and Figure 5), magnetic force microscopy tips [30], and magnetic recording media [31] (see also Figure 5). In combination with miniaturized magnets, they

represent a new approach to the realization of high density magnetic RAM [32]. Their use for fundamental studies (e.g., the investigation of vortices in superconductors) as well as in industrially oriented projects (e.g., the development of read-heads for magnetic storage media) has proved their versatility and general significance.

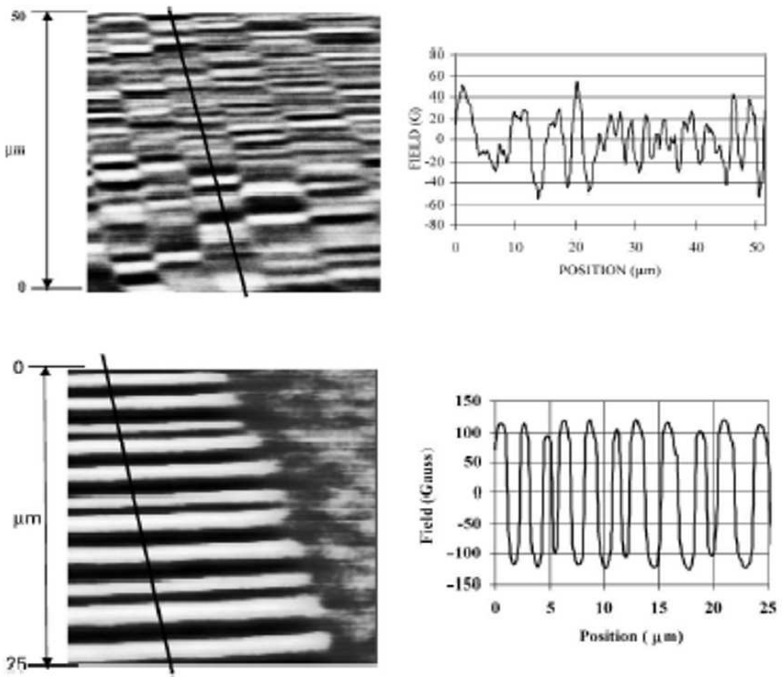


Figure 4. Room-temperature SHPM Images of (a) a 100 MB "Zip" Disk, (b) a 1.4 MB Floppy Disk. Reprinted From [27].

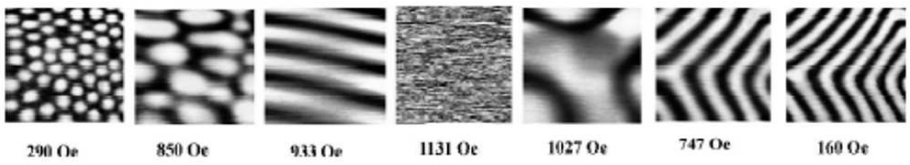


Figure 5. Room Temperature Images (25 µm x 25 µm) of 5.5 µm Thick Single Crystal Bi-Substituted Iron Garnet Epilyer Placed in a Cyclic External Perpendicular Magnetic Field. Reprinted From [27].

In our laboratory, we are developing devices and techniques for magnetometry [33], detection of magnetic microbeads [34], and magnetic resonance spectroscopy [35].

The scanning Hall probe microscopy (SHPM) method is mainly in competition with magnetic force microscopy (MFM) for magnetic imaging with sub-micrometer spatial resolution and sub-millitesla field resolution. For SHPM, the micro-Hall sensors are fabricated at the edge of a planar substrate or on a cantilever tip. The distance between the sensor and the sample surface is controlled by several different techniques: tunneling current [10], shear force, piezorestistive [36, 37] (see Figure 6 and Figure 7), and optical beam deflection [15]. At present, MFM achieves a better spatial resolution (about 10 nm) but a worse field resolution (about $100\,\mu\mathrm{T}/\sqrt{\mathrm{Hz}}$). SHPM presents large magnetic field, temperature, and frequency operating ranges. Moreover, it is a quantitative and almost totally non-invasive method (the biasing current has usually no significant effects on the sample magnetization). Due to the interaction magnetic tip-sample, the MFM technique is more invasive and, hardly, quantitative [38].

Micrometer sized Hall devices have the potential to achieve, at 300 K, magnetic field resolutions of about 10 $\mathrm{nT}/\sqrt{\mathrm{Hz}}$ in the thermal noise frequency region. In order to practically achieve this resolution, efficient *1/f* noise rejection techniques are required [39, 40]. Such magnetic and spatial resolutions are well suited for the measurements of magnetic field produced by micro- and nano-ferromagnetic structures. New applications, e.g. the investigation of naturally present magnetic nanoparticles in living organism [41-44], might become possible. Additionally, this performance might be good enough for 2D susceptibility imaging of diamagnetic and paramagnetic materials at high magnetic fields, as an alternative to SQUID detection at low fields [45].

At low temperatures, thanks to the usually higher mobility and lower thermal noise, the field resolution is significantly better than that at 300 K [46-48]. The possibility to detect the magnetic field produced by a single electron located at a few nanometers from a submicrometer Hall sensor has been anticipated [46], but not practically realized.

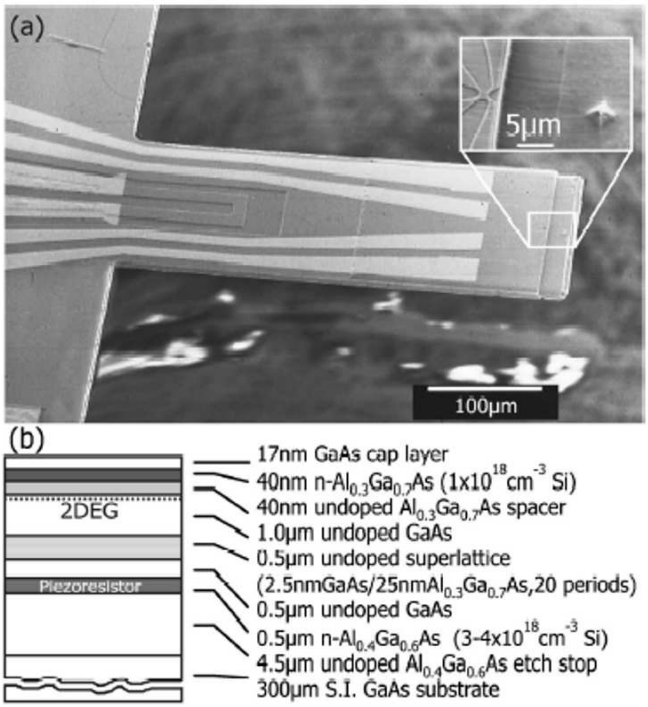


Figure 6. (a) Piezoresistive Cantilever with Hall Probe and Pyramidal AFM Tip. (b) Epilayer Structure of the Cantilever. Reprinted From [37].

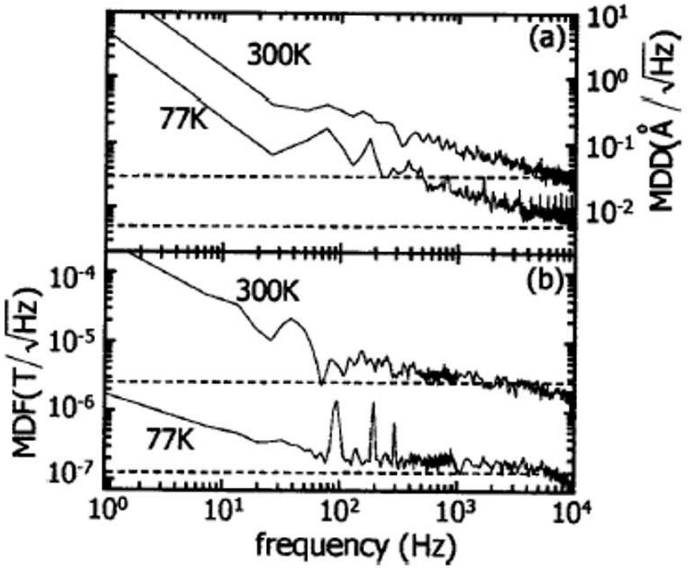


Figure 7. Noise Spectra of (a) the Piezoresistor and (b) the Hall Sensor at 300 K and 77 K. Dotted Lines Indicates the Johnson Noise Limits. Reprinted From [37].

6. CONCLUSIONS

The theory of scaling down Hall devices is described in section 3, whereas sections 4 and 5 summarize the development of micro-Hall sensors. A precise comparison of theory and experimental results from the literature is a difficult task, since the geometry, the material and the operating conditions of published devices vary form one author to the other. In our laboratory we study micro- Hall magnetic sensors made of silicon, InSb and III-V quantum wells at room temperature. From our experience, we can draw the following conclusions:

1) The current related sensitivity remains constant when scaling down sensors made in the same material. In micro-Hall devices, due to overheating or to velocity saturation, one has to decrease the current linearly with the dimensions. This leads to a linear reduction of the absolute sensitivity, as predicted by the theory.
2) The white noise is determined by the thermal noise and stays constant by scaling down, since the output resistance doesn't vary. On the other hand, the 1/f-noise. This has important practical consequences. Methods like spinning current, used to suppress the 1/f-noise, have now to be developed at very high modulation frequencies (many MHz).
3) The offset is increasing with scaling down, since the relative fabrication tolerances are increasing.

For the non-stationary transport range, experimental data are missing. From the theory, the ballistic transport range could be promising from a noise point-of-view.

In conclusion, scaling down a Hall sensor seems always to be leading to a degradation of its magnetic field resolution. On the other hand, such micro-devices open the way to applications where high spatial resolution is necessary like in scanning Hall probe microscopy.

Appendix A

CARRIER TRANSPORT IN SEMICONDUCTORS. A SUMMARY

The velocity of charge carriers is an essential parameter in the Lorentz force (1) and therefore in the Hall effect. The average velocity of carriers is strongly related to the applied electric field and to the scattering processes in the material. We shall briefly review here the theory of the electron transport process in semiconductors underlying our discussion in section 3. More details on the subject can be found for example in [4].

Stationary Transport

In the stationary transport theory, it is assumed that carriers stay under given conditions (electric field, density of scattering centres, temperature) during enough time to reach steady-state distributions both in terms of momentum and energy. "Enough time" means – enough to make many collisions with scattering centres. In other words, it is assumed that the device is so long, that the drift velocity of carriers does not depend on the location in the device. Then the following two cases can occur:

Linear drift range: At low electric field, the drift velocity is linearly related to the electric field by a constant mobility [3]. The mobility is dominantly determined by the scattering in the material, mainly due to acoustic phonons and ionized impurities. The bias current is linearly proportional with the electric field. Joule effects can appear at high current, increasing the lattice temperature in the device and modifying the mobility (**overheating**).

Velocity saturation range: At higher electric fields, carriers cannot dissipate all the energy gained from the electric field. The corresponding temperature of the carriers becomes higher than that of the crystal lattice. Such charge carriers are called hot carriers. The relaxation time of hot carriers is shorter than that of "cold" carriers. This is particularly due to optical phonons scattering, which starts to play an important role. Therefore, the mobility decreases. At sufficiently high fields, the drift velocity tends to saturation and does not increase anymore with the electric field. Remember that we are still in the stationary transport regime; therefore the mean momentum as well as kinetic energy have reached their stationary values.

Non-Stationary Transport (&2.4.4 of [1])

If a device is very small so that electrons experience just a few or even no collisions while passing a region with a high electric field - then electrons cannot settle in the steady state. Then we have non-stationary transport. The most important phenomena of non-stationary transport are velocity overshoot and ballistic transport [4].

Velocity overshoot means a swift temporary increase in the drift velocity of charge carriers well above the value of stationary saturation drift velocity. Velocity overshoot occurs during a very short period of time to electrons, which suddenly enter a high electric field region of a semiconductor device. For example, for electrons in GaAs, this happens after the first 0.1 ps or so in a field of 10 kV/cm at 300K. At these conditions, the mean distance that electrons travel before reaching a steady state is about 100 nm in GaAs and 10 nm in Si. In GaAs, the overshoot velocity may peak at a value 5 times higher than the stationary saturation velocity.

Briefly, velocity overshoot is a consequence of the difference in the momentum and energy relaxation times. It takes many collisions with phonons to reach the steady-state in energy distribution of electrons, while only few collisions are needed for the momentum to relax. And while the electron mean energy is still low, their mean transit time between collisions is that of "cold" electrons, i.e., still relatively long. Consequently, during the first few collisions in a region with very high electric field, the average "mobility" (if we may keep this notion) of electrons is similar as that in a low electric field.

The term **ballistic transport** implies that an electron "flies" trough a region of a solid-state device without any collision, similarly as an electron flies between the two electrodes in a vacuum tube. This may occur in a solid-state device if the length of the region under consideration is (i) shorter than the mean free path between scatterings, but (ii) still much larger than the wavelength of electrons. Then we can treat electrons as little charged billiard bolls of a mass equal to the effective mass of electrons in the material under consideration. We call such electrons **ballistic electrons**.

Let us estimate the length of a device in which we may have ballistic transport. Using (17), we calculate the distance l that an electron travels (without scattering) during the time t after leaving the source:

$$l = \frac{1}{2}\frac{q}{m^*}t^2 E \tag{A.1}$$

Here q and m^* are the carrier charge and effective mass, respectively, and E is the driving electric field. On the other hand, the standard expression for the stationary carrier mobility is

$$\mu = \frac{q}{m^*} < \tau >, \tag{A.2}$$

where $<\tau>$ denotes the average carrier momentum relaxation time. We may now estimate the maximal value of the flight time without scattering t in (A.1) to be $<\tau>$. By replacing $<\tau>$ from (A.2) in (A.1) we obtain an estimation for the maximal length of a ballistic solid-state device:

$$L < \frac{1}{2}\mu^2 \frac{m^*}{q} E, \tag{A.3}$$

This shows that ballistic transport is most probable in high-mobility materials. But note that we now use the notion of mobility merely as a

criterion for judging if in a material we may find ballistic transport or not. Otherwise, the movement of ballistic electrons exposed to electric field is determined by their effective mass, not by mobility.

As an example, let us take the usual value of the mobility of 2DEG at low temperatures: $\mu \approx 10^5$ cm^2/Vs; the effective mass of electrons in InAs $m^* = 0.023\ m$, and the electric field $E = 10^4$ V/cm = 1 V/μm. Then with (A.3) we calculate $L \approx 6.5$ μm. Therefore, under these conditions, a device will be predominantly ballistic if it is smaller than, say, 5 μm.

REFERENCES

[1] Popovic, R. S., *Hall Effect Devices.*Bristol and Philadelphia: IOP Publishing, 2004.
[2] Hall E. H. On a new action of the magnet on electric currents. Am. J. Math. 1879*;* 2287-2292
[3] Sze, S. M., *Semiconductor Devices*, John Wiley and Sons, Inc., 2002
[4] Brennan, K. F., Brown, A. S. *Theory of Modern Electronic Semiconductor Devices.* John Wiley and Sons, Inc., 2002
[5] Demierre M., Pesenti S., Frounchi J., Besse P.-A., Popovic, R.S., Reference magnetic actuator for self-calibration of a very small Hall sensor array. Sensors & Actuators A, 2002; 97:39
[6] Janossy B., Haddab Y., Villiot J.M., Popovic R.S., Hot carrier Hall devices in CMOS technology, Sensors and Actuators A 1998; 71: 172-178
[7] Kanayama T., Hiroshima H., Komuro M. Miniature Hall sensor fabricated with maskless ion implantation. J. Vac. Sci. Technol. B 1988; 6:1010-1013
[8] Sugiyama Y., Kataoka S. S/N studies of micro-Hall sensors made of single crystal InSb and GaAs. Sensors and Actuators 1985; 8:29-38
[9] Oral A., Kaval M., Dede M., Masuda H., Okamoto A., Shibasaki I., Sandhu A., Room-temperature scanning Hall probe microscope (RT-SHPM) imaging of garnet films using new high-performance InSb sensors. IEEE Trans. on Magnetics 2002; 38:2438-2440
[10]Chang A.M., Hallen H.D., Harriott L., Hess H.F., Kao H.L., Kwo J., Miller R.E., Wolfe R., J. van der Ziel, Chang T.Y. Scanning Hall probe microscopy. Appl. Phys. Lett. 1992; 61:1974-1976
[11]Cambel V., Karapetrov G., Elias P., Hasenohrl S., Kwok W.K, Krause J., Manka J. Approaching the pT range with 2DEG InGaAs/InP Hall sensors at 77 K. Microelectronic Engineering 2000; 51-52:333-342
[12]R. G. van Veen, Verbruggen A.H., E. van der Drift, Radelaar S., Anders S., Jaeger H..M. Micron-sized Hall probes on a Si/SiGe heterostructure as a tool to study vortex dynamics in high-temperature superconducting crystals. Rev. Sci. Instrum. 1999; 70:1767-1770
[13]Johnson M., Bennet B.R., Yang M.J., Miller M. M., Shanabrook B.V. Hybrid Hall effect device. Appl. Phys. Lett. 1997; 71: 974-976
[14]Sugiyama Y, Sano K, Sugaya T., Nakagawa T., Micro-Hall devices with high resolution made of d-doped InAlAs/InGaAs pseudomorphic hetorostructures. Proceedings of Transducers '95 & Eurosensors IX, Stockholm, Sweden, 25-29 June 1995.
[15]Chong B.K., Zhou H., Mills G., Donaldson L., Weaver J. M. R. Scanning Hall probe microscopy on an atomic force microscope tip. J. Vac. Sci. Technol 2001; . A 19: 1769-1772
[16]Sandhu A., Masuda H., Kurosawa K., Oral A., Bending S.J., Bismuth nano-Hall

probes fabricated by focused ion beam milling for direct magnetic imaging by room temperature scanning Hall probe microscopy. Electr. Lett. 2001; 37:1335-1336
[17]Novoselev K.S., Morozov S.V., Dubonos S.V., Missous M., Volkov A.O., Christian D.A., Geim A.K. Submicron probes for Hall magnetometry over the extended temperature range from helium to room temperature. J. Appl. Phys. 2003; 93: 10053-10057
[18]Grande W.J., Reznik S., Ertel J., A Hall effect microsensor for micromechanical device characterization. IEEE Trans. Magn. 1997; 33:3394-3396
[19]Yang F.Y., Liu K., Hong K., Reich D. H., Searson P.C., Chien C. L.. Large magnetoresistance of electrodeposited single-crystal bismuth thin films. Science 1999; 284:1335-1337
[20]Geim A.K., Dubonos S.V., Lok J. G. S., Grigorieva I.V., Maan J.C., Hansen L.T., Lindelof P.E., Ballistic Hall micromagnetometry. Appl. Phys. Lett. 1997; 71:2379-2381
[21]Bending S .J., Oral A., Hall effect in a highly inhomogeneous magnetic field distribution. J. Appl. Phys. 1997; 81:3721-3725
[22]L. Theil Kuhn, Geim A.K, Lok J. G. S, Hedegard P, Ylanen K., Jensen J.B., Johnson E, Lindelof P. E., Magnetization of isolated single crystalline Fe-nanoparticles by a ballistic Hall micro-magnetometer. Eur. Phys. J. D 2000; 10:259-263
[23]Li Y., Xiong P., S. von Molnar, Wirth S., Ohno Y., Ohno H., Hall magnetometry on a single iron nanoparticle. Appl. Phys. Lett. 2002; 80:1-3
[24]S. Wirth and S. von Molnar, Hall cross scaling and its application to measurements on nanometer-size iron particle arrays, Appl. Phys. Lett. 2000; 76:3283-3285
[25]Geim A.K., Dubonos S.V., Grigorieva I.V., Novoselov K.S., Peeters F.M.,. Schweigert V.A., Non-quantized penetration of magnetic field in the vortex state of superconductors. Nature 2000; 407:55-57
[26]Zeldov E., Majer D., Konczykowski M., Geshkenbein V.B., Vinokur V.M., Shtrikman H., Thermodynamic observation of first-order vortex-lattice melting transition in $Bi_2Sr_2CaCu_20_8$, Nature 1995; 375:373-376
[27]Sandhu A., Masuda H., Oral A., Bending S.J., Room temperature magnetic imaging of magnetic storage media and garnet epilayers in the presence of external magnetic fields using sub-micron GaAs SHPM. J. Cryst. Growth 2001; 227-228: 899-905
[28]Fukumura T., Sugawara H., Hasegawa T., Tanaka K., Sakaki H., Kimura T., Tokura Y., Spontaneous bubble domain formation in a layered ferromagnetic crystal. Science 1999; 284:1969-1971
[29]Nosolev K.S., Geim A.K., Dubonos S.V., Hill E.W., Grigorieva I.V. Subatomic movements of a domain wall in the Peierls potential. Nature 2003; 426:812-815.
[30]Thiaville A., Belliard L., Majer D., Zeldov E., Miltat J., Measurement of the stray field emanating from magnetic force microscope tips by Hall effect microsensors. J. Appl. Phys. 1997; 82:3182-3191
[31]Solin S. A., Stradling R. A., Thio T., Bennet J. W. Thin, horizontal-plane Hall sensors for read heads in magnetic recording. Meas. Sci. Technol. 1997; 8:1174-1181
[32]Monzon F.G., Johnson M., Roukes M. L., Strong Hall voltage modulation in Hybrid ferromagnet/semiconductor microstructures. Appl. Phys. Lett.1997; 71:3087-3089
[33]Demierre M., Pesenti S., Frounchi J., Besse P. A., Popovic R.S., Reference magnetic actuator for self-calibration of very small Hall sensor array, Sensors and Actuators 2002; A 97-98:39-46
[34]Besse P. A., Boero G., Demierre M., Pott V., Popovic R. Detection of a single magnetic microbead using a miniaturized silicon Hall sensor. Appl. Phys. Lett. 2002; 80: 4199-4201
[35]Boero G., Besse P. A., Popovic R. S., Hall detection of magnetic resonance. Appl. Phys. Lett. 2001; 79:1498-1500
[36]Brook A. J., Bending S. J., Pinto J., Oral A., Ritchie D., Beere H., Springthorpe A., Henini M., Micromachined III-V cantilevers for AFM-tracking scanning Hall probe

microscopy. J. Micromech. Microeng. 2003; 13:124-128
[37]Brook A. J., Bending S. J, Pinto J., Oral A., Ritchie D., Beere H., Henini M., Springthorpe A. Integrated piezoresistive sensors for atomic force-guided scanning Hall probe microscopy. Appl. Phys. Lett. 2003; 82:3538-3540
[38]P. J. A. van Schendel, Hug H.J., Stiefel B., Martin S., Guntherodt H.-J, A method for the calibration of magnetic force microscopy tips. J. Appl. Phys. 2000; 88:435-445
[39]Steiner R., Maier Ch., Hàberli A., Steiner F.-P, Baltes H., Offset reduction in Hall devices by continuous spinning current method, Sensors and Actuators 1998:A 66: 167-172
[40]Bellekom S., CMOS versus bipolar Hall plates regarding offset correction, Sensors and Actuators 1999; A 76:178-182
[41]Blakemore R., Magnetotactic bacteria (in geomagnetic field). Science 1975; 190:377-379
[42]Dunin-Borkowsky R. E., McCartney M. R., Frankel R. B., Bazylinski D. A., Posfai M., Buseck P. R., Magnetic microstructure of magnetotactic bacteria by electron holography. Science 1998; 282:1868-1870
[43]Kirschvink J.L, Homing in on vertebrates. Nature 1997; 390:339-340
[44]Proksch R. B., Schaffer T. E., Moskowitz B. M., Dahlberg E. D., Bazylinski D. A., Frankel R. B., Magnetic force microscopy of the submicron magnetic assembly in a magnetotactic bacterium. Appl. Phys. Lett. 1995; 66 :2582-2584
[45]Thomas I. M., Ma Y. P., Tan S., Wikswo J. P., Spatial resolution and sensitivity of magnetic susceptibility imaging. IEEE Trans. on Appl. Supercond 1993; 3:1937-1940
[46]Bending S. J., Oral A., Hall effect in a highly inhomogeneous magnetic field distribution. J. Appl. Phys. 1997; 81:3721-3725
[47]Liu S., Guillou H., Kent A.D., Stupian G. W. , Leung M. S., Effect of probe geometry on the Hall response in a inhomogeneous magnetic field: A numerical study. J. Appl. Phys. 1998; 83: 6161-6165
[48]Ibrahim I. S., Schweigert V. A., Peeters F. M., Diffusive transport in a Hall junction with microinhomogeneous magnetic field. Phys. Rev. 1998; B 57:15416-15427

Chapter 8
MODERN SILICON-BASED MEMS TECHNOLOGY

Ulrich Mescheder
University of Applied Sciences Furtwangen, Institute for Applied Research
Robert Gerwig Platz 1, 78120 Furtwangen, Germany

Abstract: The fabrication technology for Si-MEMS devices is mainly derived from the processes used for the realization of microelectronic devices. However, due to the specific demands for MEMS, there are special processes needed for MEMS devices. In this chapter, after reviewing general aspects of microtechnology some of the specific MEMS-processes will be discussed, e.g. bulk-micromachining, surface micromachining, DRIE and SOI-technology.

Keywords: MEMS, microsystem technology, Si-technology, Si-micromachining

1. INTRODUCTION

The success of microelectronic is strongly related to fabrication aspects. Using advanced lithography processes and sophisticated physical and chemical methods for pattern transfer minimum feature size well below 0.25 μm is achieved in today's large scale production. The reduction of minimum feature size has improved the performance of the microelectronic devices (power consumption, driving voltage, delay time) and has increased the possible complexity (transistors per chip) while keeping costs per chip almost constant during the last decades. This is achieved by a "giant in one time concept" based on parallel pattern definition processes (e.g. billions of structures are defined with one lithography step) and batch processing of several wafers at one time (e.g. oxidation).

Si-based MEMS, or more generally speaking, Si-based microsystems technology, is benefitting from the achievements in microtechnology. However, it is worth while keeping in mind also the differences between microelectronics and microsystems. First, the demands on miniaturization for microsystems are not as strong as for microelectronic devices. Especially in micromechanical devices (e.g. sensors) minimum feature sizes are often well above one micrometer. Even in monolithically integrated microsystems the corresponding microelectronic part does not demand for sub-micrometer features in most cases. Secondly, the degree of standardization and the

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 255-272.

number of identical devices are for microsystems far less than in the case of microelectronics. Finally, microsystems have some very specific demands such as realization of free standing structures, specific materials with special properties and access of external media to the device (e.g. fluidics in pumping devices). To fulfill the special requirements of microsystems some special microtechnological processes have been developed which will be introduced in the next chapters. Additionally, it is important to notice that for microsystems not only the main device fabrication steps have to be considered. In many cases the packaging of the devices has a very large impact on cost definition and performance of the final devices.

2. MICROFABRICATION TECHNOLOGY: OVERVIEW

Both, the fabrication of microelectronic devices and microsystems can be described as the repeated sequence of the main process steps [1], [2]:

- layer deposition
- lithography
- etching/pattern transfer

The number of these sequences or cycles is depending on the device and ranges from more than 20 (microprocessors) down to less than four (simple micromechanical sensors without integration of electronics). Figure 1 shows parts of process flow charts (simplified) of a p-channel MOS transistor (left) and a of a membrane (e.g. for a micromechanical pump, right).

For standard layer deposition techniques physical and chemical thin film deposition techniques such as PVD (physical vapour deposition: thermal evaporation and sputtering) and CVD (chemical vapour deposition) are used. The layers are deposited all over the wafer. Whereas in microelectronics typical layer thicknesses range from some nanometers to micrometers, some microsystems do require also thicker layers (e.g. 10 µm poly-Si in Bosch's surface micromachined accelerometer). An overview about the most common deposition techniques is shown in Figure 2. Electroplating, even though mainly development for microsystem technology, is used also in microelectronic technology (Cu).

The basic pattern definition is done by mask based lithography which allows a parallel transfer of the patterns on the mask to a certain area (chip or full wafer) either by projection (e.g. 5:1 optics) or proximity/contact (1:1) printing. For feature sizes well below 0.5 µm DUV-lithography using projections steppers is used. For larger feature sizes and lower volume production often found in the case of microsystems even the far more simpler and cheaper proximity printing might be still sufficient[1]. After exposure the

[1] X-ray lithography which is used within the classical LIGA-Process is a special type of proximity printing allowing for high resolution and high aspect ratio.

patterns are developed (positive resist: exposed are developed, non-exposed areas stay, negative resist vice versa). Typical resist thickness is around 1 µm and less (for high resolution demands a thinner resist thickness will reduce the depth of focus problem), however, for special demands in microsystem fabrication very thick resists (> 50 µm) might be needed. In Figure 3 schematic set-ups for contact/proximity printing (left) and projection lithography (centre) are shown. On the right hand side of Figure. 3 the main components of an optical stepper (UV) are shown. In a stepper the different step fields are printed one after another. Steppers are sophisticated, highly automated machines with a typical throughput of around 50 wafers/h.

The resist patterns which are defined within the lithography process are transferred by an etch process into the layer underneath the structured resist.[2] Used are either wet chemical or dry etch processes. Due to the isotropy of the process when etching polycrystalline or amorphous materials wet etching cannot be used for very small features sizes. Therefore, dry etch processes such as RIE (reactive ion etching) are predominately used in modern microelectronic processes. However, when etching crystalline materials such as c-Si, some special etchants will show a pronounced anisotropic etch rate which reflects the anisotropy of the material. In this case even the simpler and cheaper wet etch processes provide anisotropic etch profiles. Some typical etch profiles (cross sections) are shown in Figure 4.

3. SPECIFIC TECHNOLOGY ASPECTS FOR MEMS

3.1 Low Stress Deposition Techniques:

Stress engineering

The need to fabricate structures, which are free to move at least with one degree of freedom is one of the typical demands of microsystem technology. In respect to thin layer deposition techniques this demand is accomplished by deposition techniques, which provide low internal stress (less than 10^8 Pa) of the deposited materials. In general, the total stress should be in the tensile range to avoid buckling of the microstructures.

[2] This holds for subtractive process. In an additive structuring process such as electroplating, the layer will grow only in the areas defined by lithography.

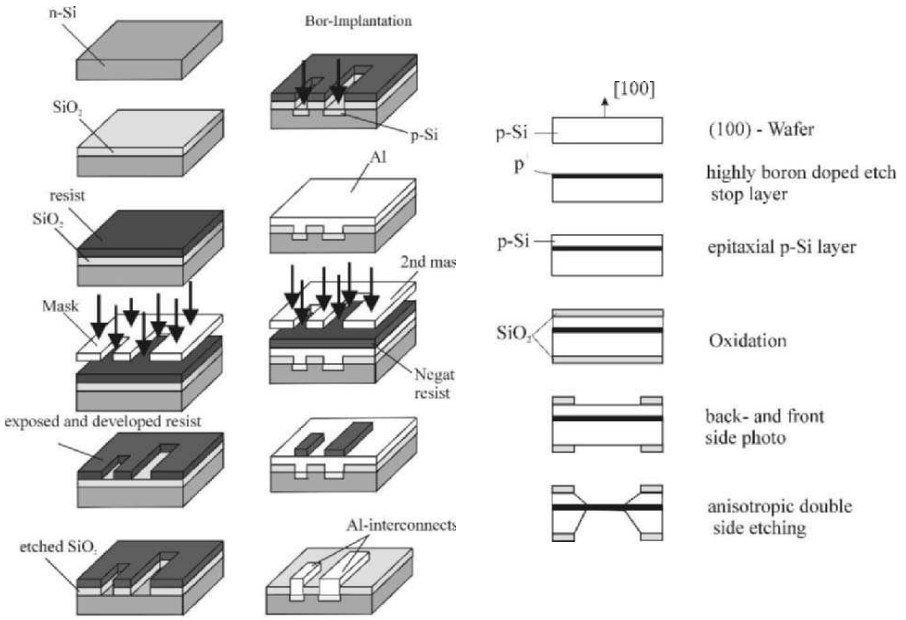


Figure 1. Left: Schematic process flow chart for the realization of a p-channel MOS transistors (first steps only), right: fabrication sequence for the realization of a membrane (lithography steps not shown).

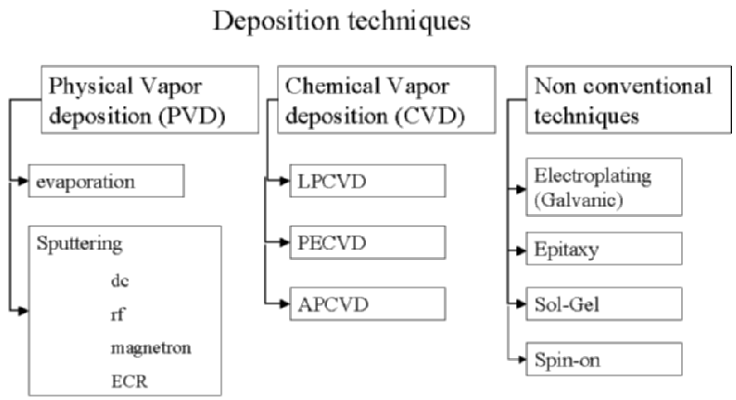


Figure 2. Classification of deposition processes. The processes on the right hand side are especially used during the fabrication of microsystems.

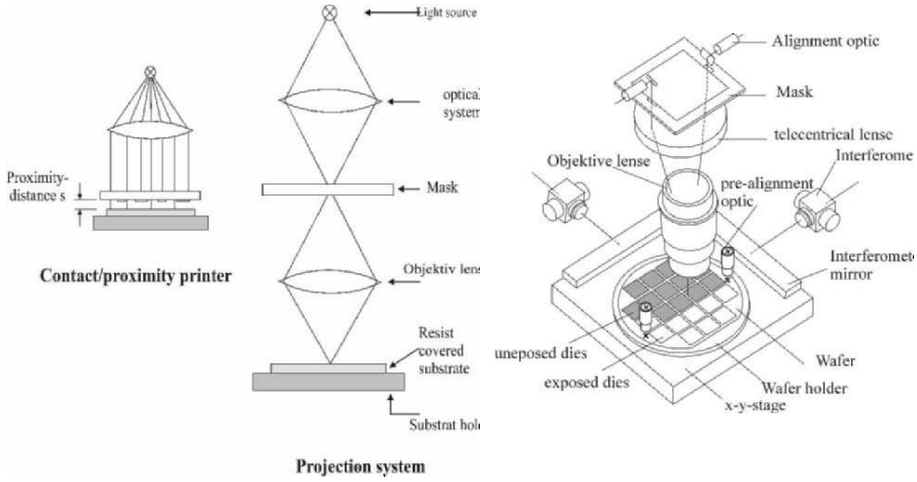


Figure 3. Principle set-up for contact/proximity printing (left) and projection lithography (centre). On the ride hand side the main components of an optical stepper are shown.

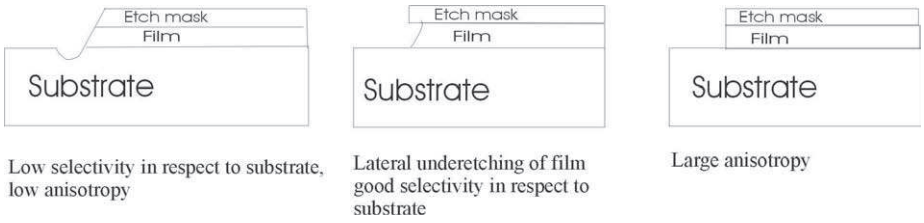


Figure 4. Typical etch profiles for dry etching. Wet etching provides in general isotropic profiles similar to the etch profile in the centre.

Different approaches are used to reduce the internal stress:

Stress compensation
Two layers with opposite stress (tensile and comprehensive) give together a low stress layer system. As an example silicon oxide and silicon nitride provide normally opposite internal stress.

Another example of stress compensation has been shown for highly boron doped epitaxially grown Si which is normally under high tensile stress: by incorporation of Germanium the stress in the Si-layer is considerably reduced [3].

Stress reduction by changing deposition process
Especially for LPCVD and PECVD deposition processes residual stress can be changed by deposition parameters such as deposition temperature and gas

flow. E.g. silicon-rich siliconnitride layers have lower stress than stochiometric siliconnitride films [4].

Stress reduction by annealing
Due to the meta-stability of polycrystalline or amorphous layers post-deposition annealing processes will change the material properties and especially the stress situation. Predominant mechanism for stress reduction are the formation of a suitable grain size distributions by appropriate annealing cycles for a given deposition process of LPCVD polycrystalline films [5] and driving out hydrogen especially in the case of PECVD films.

3.2 High Aspect Ratio Lithography

To meet the requirements for High Aspect Ratio Microsystems (HARMS) special lithography techniques have been developed. For projection lithography (see part 2) the limiting factor is given by the depth of focus (DOF):

$$DOF = \pm 0.3 \frac{\lambda}{NA^2}, \tag{1}$$

where λ is the used wavelength and NA the numerical aperture of the used projection system. Therefore, taking into account that the minimum feature size is depending on λ/NA, the DOF is typically in the order of 1 µm. To overcome this limitation two approaches of contact/proximity printing are used.

X-ray lithography
At a wavelength around nm even with reasonable proximities (distance between mask and wafer) a very good resolution can be obtained [6]. Using light sources with parallel x-ray radiation (synchrotrons) resist pattern with vertical sidewalls and resist thickness up to 100 µm can be generated. Therefore, x-ray lithography is the one of the essential process steps for LIGA (see part 7). Resist patterns printed by synchrotron x-ray lithography are shown in Figure 5.

Special resists (SU-8)
SU-8 is an epoxy-based resist designed specifically for ultra-thick, HARMS-type applications. This resist can be applied in a single-layer coating process at thicknesses of more than 500 µm. Aspect ratios of about 18 for near DUV have been achieved [7].

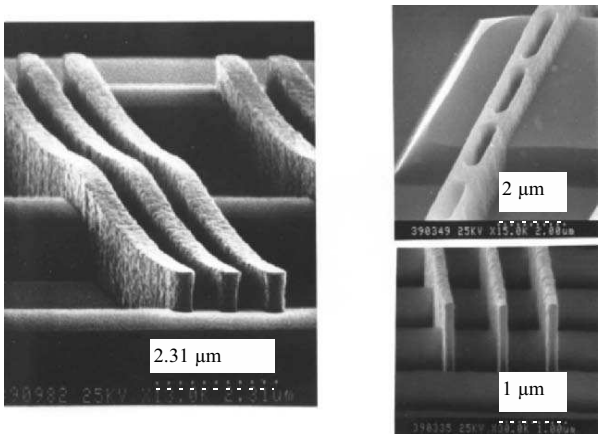


Figure 5. High resolution, vertical resist patterns defined by synchrotron based x-ray lithography.

4. HIGH ASPECT RATIO WET ETCHING: BULK MICROMACHINING

The availability of wet chemical etch processes which allow the anisotropic etching of crystalline silicon has been the basis for the realisation of MEMS devices in the last twenty years or so.

Anisotropic etching is based on electrochemical etch processes provided by special etchants such as ethylene diamine pyrochatechol (EDP), potassium hydroxide (KOH) or tetramethyl ammonium hydroxide (TMAH). In all these etchants the {111} planes are far less etched than any other plane. The most important properties of these anisotropic etchants are summarized in Table 1 [8-10].

Table 1. Etch Solutions for Anisotropic Etching of c-Si.

Property/Etchant	**KOH**	**EDP**	**TMAH**
typical concentration	30%	4mol% Pyrocatechol	2 weight percent
Typical etching temperature (°C)	80	20	80
typical etch rate (µm/h)	60	35	40
Anisotropy ($etchrate_{<100>}/etchrate_{<111>}$)	50	100	50
Masking layer	Si_3N_4	SiO_2	SiO_2, Al, Si_3N_4
comments	cheap, simple process	May cause cancer	Free of metal ions, can be used with Al

Anisotropic etching of c-Si provides a very cheap process for the fabrication of very reliable micromechanical structures (crystalline silicon is an almost perfect mechanical material [11]). However, as the anisotropy is determined by the crystallographic axes, special layout rules have to be considered as shown in Figure 6 for (100)-Si (left) and (110)-Si (right). Whereas on (100)-Si the resulting structures have normally a V-groove shape (angle between {111} and {100}: 54,74°), on {110} vertical sidewalls are possible. Unregular shapes (e.g. circles) will be regular after etching (e.g. rectangular). An example of a bulk micromachining process for the fabrication of a freestanding membrane has been already shown in Figure 1 (right). Typically, bulk micromachining makes use of double side processing and needs therefore also special lithography equipment, which allows double side alignment (e.g. MA6, Suess). As a result, structure size density is rather low for bulk micromachining. However, till today most of Si-pressure sensors are fabricated with Si-bulk-micromachining.

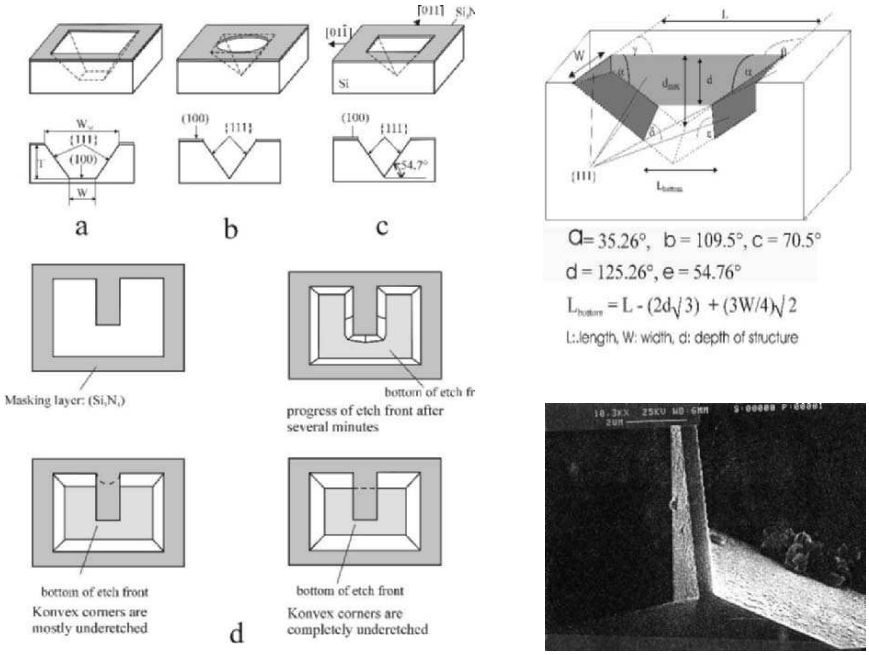


Figure 6. Anisotropic wet etching of c-Si. Left: (100-Si), right: (110)-Si. Top: {111} plane are indicated in grey, bottom: verticalyl etched line as linewidth calibration standards.

Very important demand in bulk-micromachining is the exact thickness control. E.g. for pressure sensors the sensitivity depends on the thickness d of the membrane as $1/d^2$.

With simple time etching the final membrane thickness will depend on the starting wafer thickness and is therefore not well defined. Exact thickness control is possible when using p^{++}-Si: For doping concentration above 10^{20} cm^{-3} boron doped Si provides a reasonable etch stop [12]. However, stress compensation by Ge is needed to get membranes with low internal stress (compare 3.1). An alternative is the electrochemical etch stop provided at a pn-junction with suitable bias voltage [13]. A schematic set-up for an electrochemical etch stop is shown in Figure 7 (left), a resulting structure is shown on the right hand side. The thickness of the beam –used for the realization of a two-axes tilt sensor [14] - is defined by an electrochemical etch stop. Then the beam is etched out of the membrane by RIE. A silicon microphone fabricated with bulk micromachining is shown at the bottom of Figure 7 [15].

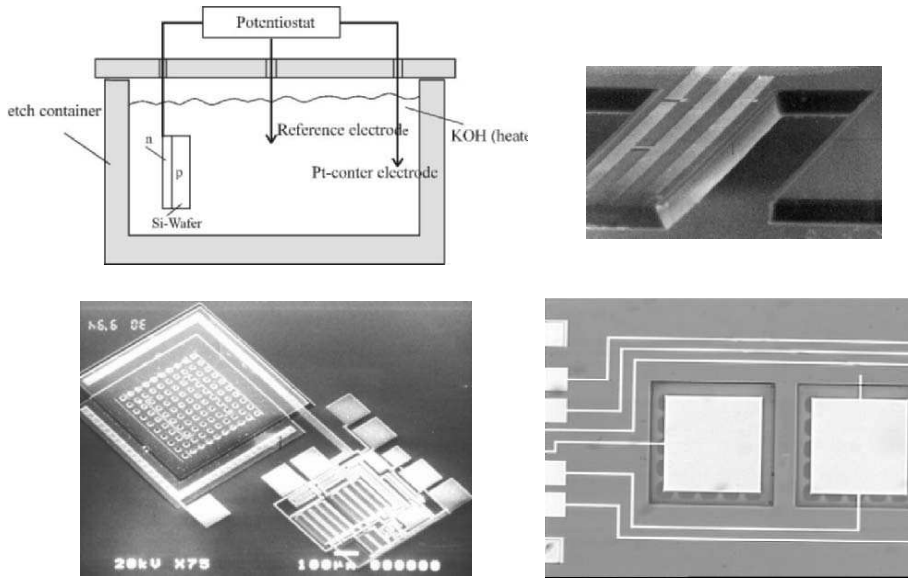


Figure 7. Top: Set-up for electrochemical etch stop (left) and structure defined by this method. The beam is clamped at two sides (the clamp to the seismic mass was removed for this picture). Bottom: Si-microphones produced by bulk micromachining.

A disadvantage of bulk-micromachining is that it is not CMOS compatible as potassium contamination can cause considerable reduction in yield of microelectronic devices. Further on, after bulk etching special care is needed for handling the wafers. Patterning of already bulk-etched Si is not possible with standard lithography.

5. SURFACE MICROMACHINING

Surface micromachining provides a CMOS compatible fabrication technique for free-standing structures and allows a higher structure density than bulk-micromachining. The basic principles of surface micromachining have already been investigated in the 80s [16]. However, first commercial devices fabricated with surface micromachining appeared in the 90s [17].

The basic idea of surface micromachining is the sacrificial layer concept as shown in Figure 8. The functional layer which will define later movable part of the device is deposited on a pre-structured sacrificial layer. After appropriate structuring of the functional layer the sacrificial layer is removed in an isotropic etch process. The thicknesses of the used layers are mostly in the same range as standard thicknesses in a CMOS or BiCMOS processes. For etching fully CMOS compatible processes can be chosen. Due to the sacrificial layer, etching access to the etched area has to be provided by suitable holes in the structure to ensure complete underetching of the structure. The mechanisms during sacrificial layer etching are shown in the bottom of Figure 8, on the right hand side, a typical layout for the etch holes is presented. Main technology requirements of surface micromachining are low stress materials (compare 3.1), prevention of so-called sticking [18] (during rinsing process by capillary forces or during operation by physical adhesion) and selectivity of the used etch processes.

Different combinations of sacrificial and functional materials are summarized in Table 2.

Table 2. Combination of Functional and Sacrificial Layer Materials and Typical Thicknesses.

Functional layer materia	**typ. thickness (µm)**	**Sacrificial layer material**	**typ. thickness (µm)**
Poly-Silicon	(1-10)	Phosphorous-silicatglass PSG	(1 - 7)
		SiO_2	2
		porous silicon	(1 - 30)
SiO_2	2	poly-Si	2
TiNi	8	Polyimid	3
		Au	2
NiFe	2	Al	7
W	3	SiO_2	8

Some examples of surface micromachined devices are shown in the bottom of Figure 8. On the left hand side a detail of Bosch's surface micromachined accelerometer is shown. As can be seen a relative thick functional layer is used. The two other SEM pictures show a surface micromachined gyro. In the close-up a detail of the electrostatic actuator is shown. This structure

provides the primary vibration for the sensor, which is based on the Coriolis force. The holes in the structures are needed for sacrificial layer etching, however, these holes will also reduce air damping.

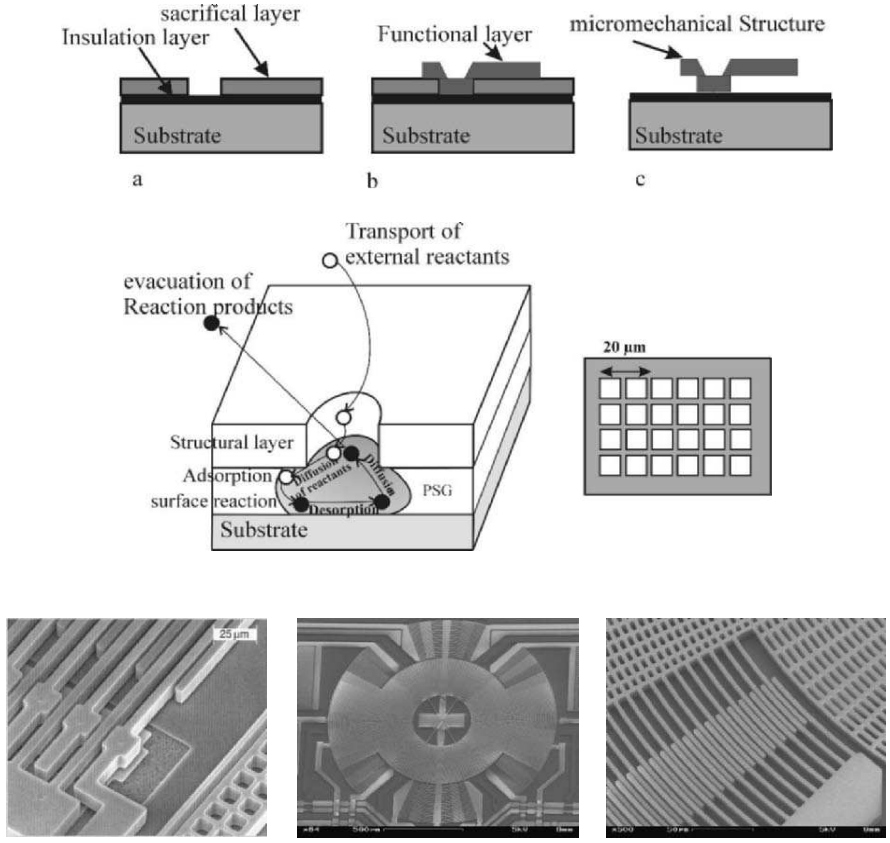


Figure 8. Surface micromachining, top: Basic idea of sacrificial layer etching, centre left: processes involved for sacrificial layer etching, centre right: schematic layout of etch holes. Bottom left: detail of a surface micromachined accelerometer (left), a surface micromachined gyro (centre), and detail of the gyro structure (electrostatic actuation, right), with courtesy of Bosch GmbH, Reutlingen, Germany.

More recently, sealing processes [19] have been used to realize commercial absolute pressure sensors with monolithical integrated electronics [20].

An increased layout flexibility and a better device performance have been provided since so-called DRIE (deep reactive ion etching) processes are available. With the "Bosch-process" [21] and with the so-called cryogenic-process [22] extremely high aspect ratio etching has become possible. With these dry etch techniques it is now possible to etch deeply into the c-Si which was before only possible by anisotropic etching (bulk-

micromachining). Another advantage is that with perfect vertical sidewalls simulated data can be met with a higher accuracy and that cross effects (e.g. torsional components in a bent beam) are negligible.

In the Bosch-process a sidewall passivation is achieved by a cyclic switching between an etching (SF_6) and a passivating (C_4F_8) gas. In the cryogenic-process at very low temperature a sidewall passivating SiO_xF_y, thickness: 10 – 20 nm, is formed which is at these low temperatures not etched by the F-containing radicals of the etching gas.

A special surface micromachining process will be discussed in the next chapter: When using porous Si as sacrificial layer it is possible to fabricate very well defined c-Si free-standing structures in a surface micromachining process [23], thus combining the advantages of surface and bulk-micromachining.

6. SOI TECHNOLOGY

The use of SOI (Silicon On Isolator) technology can simplify the complexity of a fabrication process for free standing structures considerably. The reason is that the buried oxide layer provides an etch stop for both, front side etching (e.g. DRIE) and backside etching (e.g. anisotropic wet etching). Further on, an excellent isolation to the substrate by a high quality buried oxide between device layer (functional Si) and handle wafer is provided. Due to increased industrial use the prices for SOI-wafers have decreased drastically in the last view years. Additionally, a wide range of different thicknesses (handle and device layer) and corresponding doping levels are available. Therefore, fabrication processes based on SOI offer interesting advantages, both, under economical and technological view points. Two different flow charts for the fabrication of freestanding structures using SOI-technology are shown in Figure 9 (top). On the left hand side a combination of bulk micromachining and surface micromachining is used: A membrane is etched from the backside with the buried SiO_2 as etch stop layer. Finally, the exact shape of the freestanding structures is defined by DRIE from the front side and the remaining oxide layer might be removed. On the right hand side, only processing from one side is used. However, in this case there are severe layout restrictions as it is not possible to structure the sacrificial layer (buried oxide) as simple as in a standard surface micromachining process. As can be seen, 1-3 mask levels are sufficient to get excellent freestanding structures.

In the bottom of Fig. 9 some application examples of SOI-technology are shown. On the left hand side, a bi-stable electrostatically driven fibre switch is shown. Extreme geometries can be realized reliably using SOI (minimum width of beams 2 μm, thickness 10 μm, length of freestanding beams up to 800 μm), thus allowing a long range of travel (100 μm) during switching [24].

In the centre, a fibre switch based on a moving mirror is shown [25]. In this case 75 µm deep grooves are etched into the device layer which provide the fibre alignment close to the mirror which will turn the light direction by 90° when driven into the gap between input and output fibres. On the right hand side a tilt sensor is shown which is based on convection. The heating and temperature sensing structures are defined by SOI out of the device layer [26].

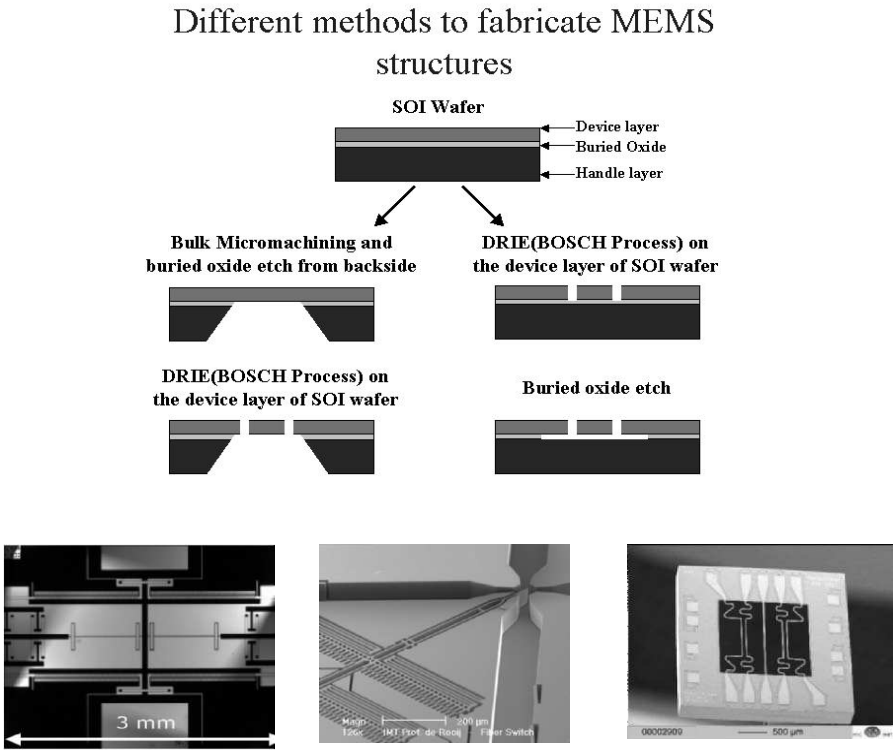


Figure 9. Top: Simplified process flow charts for the fabrication of free-standing structures with SOI-wafers, top left: double side processing, top right: single side processing. Bottom: application examples. Left: bi-stable fibre switch, centre: fibre switch with electrostatically driven mirror, with courtesy of Prof. De Rooji, IMT, University of Neuchatel, right: inclinometer based on convection, with courtesy of Dr. Billat, HSG-IMIT.

Even though a relative young technology it is expected that SOI-technology will have a large impact on microsystem fabrication in the future.

7. LIGA

LIGA (the German acronym for "Lithographie Galvanik Abformung" = Lithography, electroplating and molding) is a non-Si technology. However, it is shortly presented here to provide a complete overview of the different technology concepts for MEMS/microsystems. A schematic representation of the LIGA process is given in Figure 10.

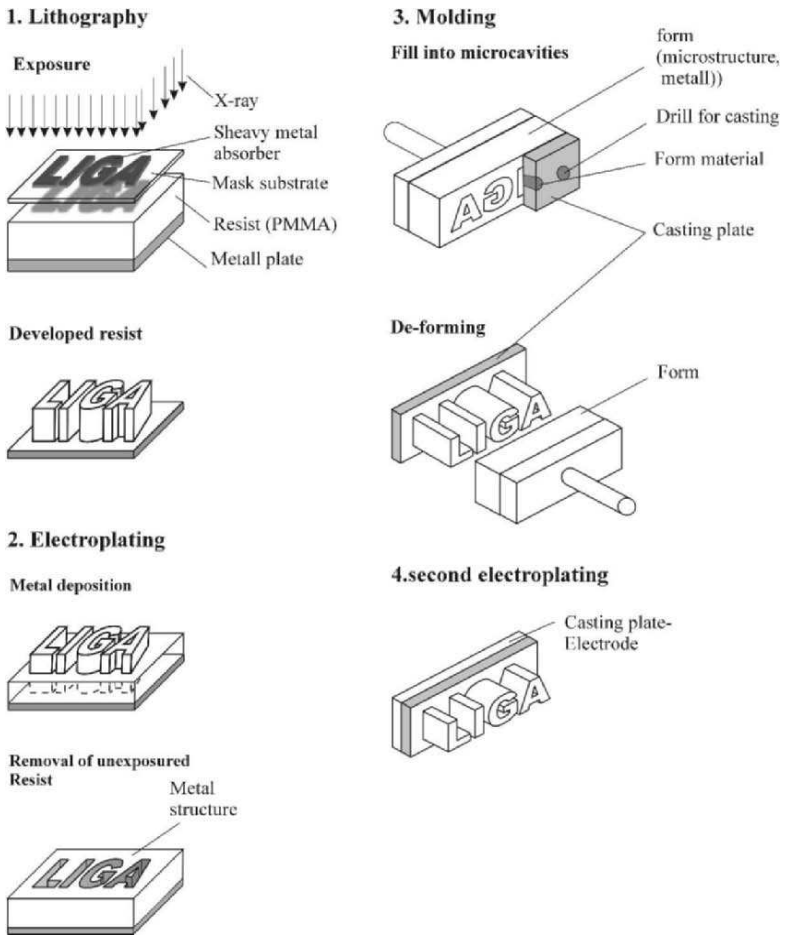


Figure 10. Schematic Process Flow of the LIGA Process.

In the classical LIGA concept [27] x-ray lithography is used which provides a large depth of focus and allows the definition of very thick resist patterns (up to 100 μm thick, step 1). In the following process step the resist

trenches are filled up by an electroplating process (e.g. Au or Ni, step 2). Therefore, this additive structuring process will define the negative metal structure of the resist pattern. The metal mold can than be used for a further plastic molding process (step 3). This plastic mold can be used as final product (using the metal form as master) or it can be used as mold for the second electroplating (step 4), providing a metal structure as final product or as copy of the master form. LIGA technology is especially suited for the production of simple microstructure elements with high aspect ratios. However, in its classical form it cannot provide microsystems. To overcome this limitation LIGA-based processes have been integrated into Si-technology. A further development is the replacement of the very expensive x-ray lithography step by conventional, but high aspect ratio lithography (s. UI.3.2).

8. BONDING PROCESSES

Besides the device fabrication itself the integration of MEMS into the macro world is one of the great challenges in MEMS technology. The first step of this integration is in many cases the bonding of the silicon wafer containing the microsystems to a cap or bottom wafer which might provide additional functions for the final device or may only serve as protection of the fragile MEMS device.

There are a number of different methods available for bonding micromachined silicon wafers together, or to other substrates, to form more complex devices. One of the most popular bonding techniques is anodic bonding [28]. Here, silicon is bonded to glass (pyrex). The silicon wafer and glass substrate are cleaned, pressed together and heated to a high temperature (200°-500°C). In a first step, a large voltage (typically several hundred volts) is applied, therefore driving Na+ of the glass out of the contact zone and forming a depletion zone at the interface. This causes an extremely large electrical field at the depletion zone and thus a large electro static pressure (about 10 bar). Under this large pressure a strong, long-term stable chemical bond is formed between the two materials. Figure 11 shows a set-up of an anodic bonding apparatus.

It is also possible to bond silicon wafers directly together using gentle pressure and high temperature (direct silicon bonding).

Using thin films of Au or Al between the bond partners (at least one of them Si) an eutectic bond can be formed due to the low melting point of the Au-Si or Al-Si alloys. As eutectic bonding needs temperatures, which are not compatible to already deposit metals (e.g. in the case of Al) a local bonding provided by laser can overcome this problem [29].

Other bonding methods include using an adhesive layer, such as a glass, or photoresist. Whilst anodic bonding and direct silicon bonding form very

strong joins they suffer from some disadvantages, including the requirement that the surfaces to be joined are very flat and clean.

Wafer bonding techniques can potentially be combined with some of the basic micromachined structures to form e.g. valves, pumps, etc.

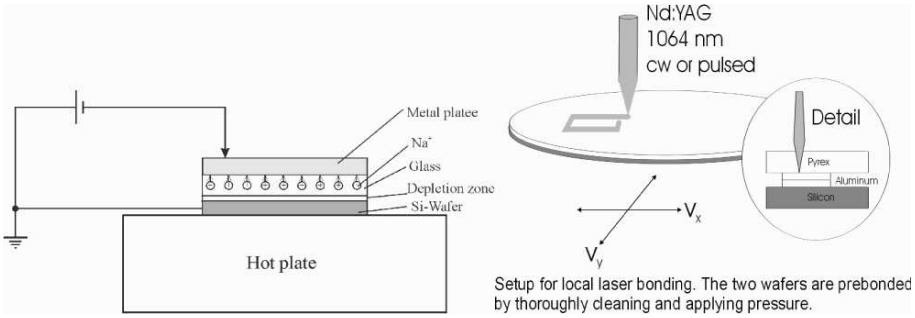


Figure 11. Left: Set-up for Anodic Bonding; Right: Laser Induced Eutectic Bonding.

SUMMARY

Modern Si-based MEMS technology makes use of the full range of processes provided by the microelectronic area. However, different very special processes are needed to meet the specific demands of MEMS/microsystems. In the very first beginning, bulk-micromachining was the only choice to fabricate microstructures, which are free to move. Till to date bulk micromachining is used especially for the fabrication of piezoresistive pressure sensors. However, even though bulk-micromachining is a relative simple technology it has distinctive disadvantage: the widely used KOH-etching is not CMOS compatible and the structure density is rather low for bulk-micomachining. Surface micromachining has been developed as an important alternative. Today, most of the high volume products such as accelerometers, gyros and digital micro mirror devices are fabricated with surface micromachining technology.

A very promising development is the SOI-technology, which simplifies the fabrication process and therefore can reduce process costs and increase yield considerably.

REFERENCES

[1] Sze, S. M., *VLSI-Technology*, New York: McGraw Hill, 1983
[2] Wolf, S., Tauber, R.N, *Silicon Processing for VLSI-Era*, Vol. 1-3, Latt. Press, 2002

[3] Herzog H.-J., Czepregi L., Seidel, H. X-ray investigation of boron- and germanium-doped silicon epitaxial layers. Journal Electrochem. Society 1984; 131: 2696
[4] Stewart, R.A., Kim, J.R. Kim, E.S., White, R.M. and Muller, R.S. Young's Modulus and Residual Stress of LPCVD Silicon-rich Silicon Nitride Determined from Membrane Deflection. Sensors and Materials 1991; 2: 285-98
[5] Guckel H.; Burns D.W., Visser C.C.G, et al. Fine-grained polysilicon films with built-in tensile strain**.** IEEE-transaction el. Dev. 1988; Vol 35 (6): 800-1
[6] Heuberger A., X-ray lithography. Solid State Technology 1986; 93-101
[7] Despont, M., Lorenz, H., Fahrni,.N., Brugger, J., Renaud,P.,Vettiger, P. High-aspect-ratio, ultrathick, negative-tone near-uv photoresist for MEMS applications**.** Proceedings of the Tenth Annual International Workshop on Micro Electro Mechanical Systems 1997; 518–22
[8] Finne, R.M., Klein, D.L. A water soluble amine complexing agent system for etching Si. Journal Electrochemical Society 1967; 114: 965
[9] Seidel, H. Chapt. 3.2 in A. Heuberger (ed.) *Mikromechanik*; Springer-Verlag, 1991.
[10]Schnakenberg, U. et al., TMAHW etchants for silicon micromachining. Proceedings Transducers 1991.
[11]Petersen, K. E. Silicon as Mechanical Material. Proc. IEEE 1982.
[12]Seidel, H. , Czepregi, L. Studies on the anisotropy and selectivity of etchants used for the fabrication of stressfree structures. Proc. Electrochem. Society, Montreal (Canada), 1982.
[13]Waggener, H.A. Electrochemically controlled Thinning of Silicon. Bell Syst. Tech. Journal, 1970; 50: 473
[14] Mescheder U., Majer S. Micromechanical Inclinometer. Sensors&Actuators 1997; A60: 134-8
[15]Kronast, W., Müller, B., Stoffel, A. Miniaturized single-chip silicon microphone with integrated field-effect transistor. Proceedings of Sixth Micromechanics Europe Workshop, MME´95, 3-5 Sept. 1995.
[16]Fan, L.S, Tai, Y.C., Muller, R.S. Pin Joints, Gears, Springs, Cranks and Other Novel Micromechanical Structures. Proceedings Transducers87 1987.
[17]Analog devices, ADXL50, company announcement, 1992.
[18]Legtenberg, R. et al. Stiction of Surface Microstructures after Rinsing and drying: Model and Investigation of Adhesion Mechanisms. Proc. Of Transducers ’93, Yokahama, Japan 1993.
[19]Guckel, H, Burns, DW. Planar Processed Polysilicon Sealed cavities for Pressure Transducers Arrays. Proceedings IEEE International El. Dev.meeting, San Francisco, USA 1984.
[20]Hierold, Ch. et al. CMOS-kompatible Oberflächenmikromechanik: Schlüsseltechnologie für integrierte Mikrosysteme. ITG-Fachbericht Sensoren und Meßsysteme 1998; 148: 15-22
[21]Patent DE4241045, US 5501893 und EP 625285, inventor: Franz Laermer, Andrea Schilp
[22]Jansen, H. V., de Boer, M. J., Legtenberg, R. Elenspoek, M. C. The black silicon method: a universal method for determining the parameter setting of a fluorine based reactive ion etcher in deep silicon trench etching with profile control.. Journal of Micromech. Microeng. 1995; 5: 115-20
[23]Mescheder, U., Kovacs, A. Surface Micromachining Process for C-Si as Active Material. Proc. of Technical Papers of the Transducers ´01, Eurosensors XV, 2001.
[24]Freudenreich, M., Mescheder, U. M, Somogyi, G. Design Considerations and Realization of a novel Micromechanical Bi-stable Switch. Digest of Transducers’03, Boston (USA), June 2003: 1096-9
[25]Marxer, C., de Rooij, N.F. Micro-opto-mechanical 2x2 switch for single-mode fibers based on plasma-etched silicon mirror and electrostatic actuation, J. of Lightwave Technology, 1999; Vol. 17, No. 1: 2-6.
[26]Billat, S. et al. Micromachined inclinometer with high sensitivity and very good stability. Techn. Digest Transducers’01 June 2001: 1488-91

[27]Becker, E.W., Betz, H., Ehrfeld, W., Glashauser, W., Heuberger, A., Michel, H.J., Münchmeyer, D. , Pongratz, S. R.v. Siemens: Naturwissenschaften 1982; 69: 520ff
[28]Köhler, J. et al. Anodisches Bonden - Silizium-Glas-Verbindungen in der Mikromechanik, Sensor Magazin 1992; 2: 6-7
[29]Mescheder, U. et al. Local Laser Bonding for Low Temperature Budget. Sensors and Actuators 2002; A97-98: 422-442

Chapter 9

POROUS SILICON: TECHNOLOGY AND APPLICATIONS FOR MICROMACHINING AND MEMS

Ulrich Mescheder
University of Applied Sciences Furtwangen, Institute for Applied Research
Robert Gerwig Platz 1, 78120 Furtwangen, Germany

Abstract: Porous Si is formed out of single crystalline silicon in a CMOS-compatible electrochemical etch process. This material provides due to its nanoporous structure and high porosity several applications in microsystems/MEMS. Applications of porous Si are found in optics (luminescence, electroluminescence, Bragg-filters), sensing (humidity, gas sensing), chemical reactions (reformers and fuel cells) and microfabrication (sacrificial layer, 3D structuring). Results of research work in our laboratory are presented together with a review of most relevant results of other groups.

Keywords: Porous Silicon, multifunctional material, MEMS, microsystem technology, Si-technology, Si-micromachining, humidity sensor

1. INTRODUCTION

A lot of attentions has been turned to porous Si since the late 80s and beginning 90s of the 20th century when photoluminescence [1] and electroluminescence [2] was observed at this material. A review over the tremendous published work is given in [3]. However, in the last few years porous silicon has been recognized also as an interesting material in different MEMS applications.

As discussed in the previous chapter, a key step in the fabrication process of most microsystems/MEMS devices is the realization of freestanding structures such as beams (accelerometers, gyros) or plates (pressure sensors, pumps). Bulk micromachining and surface micromachining can be used for this purpose. The main advantage of bulk micromachining is the fact that the resulting functional structures are crystalline. Especially in micromechanical applications this provides for almost perfect material properties (excellent reliability, no fatiguing, high quality factor). However, due to the higher layout flexibility and better compatibility to standard CMOS processing

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 273-288.

surface micromachining is used for the fabrication of most of today's commercial microsystems/MEMS. In this chapter an alternative surface micromachining process, which allows the fabrication of freestanding crystalline structures with well-defined thickness down to the sub-µm range is presented. This process makes use of porous Si produced in an electrochemical anodization process as sacrificial layer. Another process discussed in the previous chapter is LIGA, a non-Si technology which provides a replication process e.g. for plastic micro parts. However, LIGA is limited to structures with vertical sidewalls. The use of an electrochemical pore formation process in which crystalline Si is transformed to porous Si provides an etch process with adjustable anisotropy factor A_f. The resulting shapes can be used as forms for micro-moulding or micro-stamping. Therefore, both mentioned processes are based on the use of porous Si as structural material during the fabrication process.

Microsystems are defined as miniaturized multifunctional systems in which electrical and non-electrical functions are integrated in one system. Non-electrical functions are e.g. out of the areas mechanics, optics, chemistry or biology/medicine. Therefore, materials are needed which provide these different functions. An important drawback of system multifunctionality is the increased complexity of the related fabrication processes. Keeping fabrication complexity as low as possible while still achieving high system complexity demands for the use of intelligent materials or multifunctional materials, which take over different tasks in a microsystems. Porous Si is an example of such a multifunctional material.

2. POROUS SI: STRUCTURE, PROPERTIES AND FABRICATION ASPECTS

In most cases porous Si has a sponge like structure with a c-Si skeleton and pores' sizes ranging from nanometer to micrometer, depending of the process parameters used [4]. The material is classified in 3 three groups: microporous (pore sizes less than 2 nm), mesoporous (2-50 nm) and macroporous (pore size > 50 nm). However, also channel like structures can be generated, especially in light assisted processing techniques [5]. High-resolution methods are needed to investigate the structure of the material. In Figure 1, high-resolution TEM (transmission electron microscopy) and AFM (atomic force microscopy) pictures are shown. In the TEM, the sponge like structure with pronounced channels is clearly resolved (top left), at a higher resolution the ellipsoidal pores (2nm/10 nm) are observably. AFM is a surface characterisation method. As discussed in [6] TEM provides a very local insight view of the pore form and distribution, whereas AFM is suitable for the quantitative characterization (e.g. RMS –root mean square-values) of relative large areas of surfaces or interfaces. AFM measurements turned out

to depend critically on surface treatment and measuring conditions. Contact mode provides best resolution but can damage the surface. On the other hand non-contact or tapping mode are more suitable for fragile surfaces, however measuring conditions like pre-load and scan rate can influence the results. Therefore, it is important to clarify the situation by inspection of the interface between porous Si and c-Si after removing the porous Si layer. The AFM pictures in Figure 1 (bottom) were taken at surfaces of two different porous Si layers. In this case RMS values of 1.5 nm and 120 nm reflects well the microporous and mesoporous structure of the investigated materials as has been checked by measurements of the corresponding interfaces to c-Si [6]. However, the higher RMS value of 120 nm is only a rough estimation to the real porous structure, as superposing macrostructures are often observed which might be better characterized by the fractal dimension of the surface.

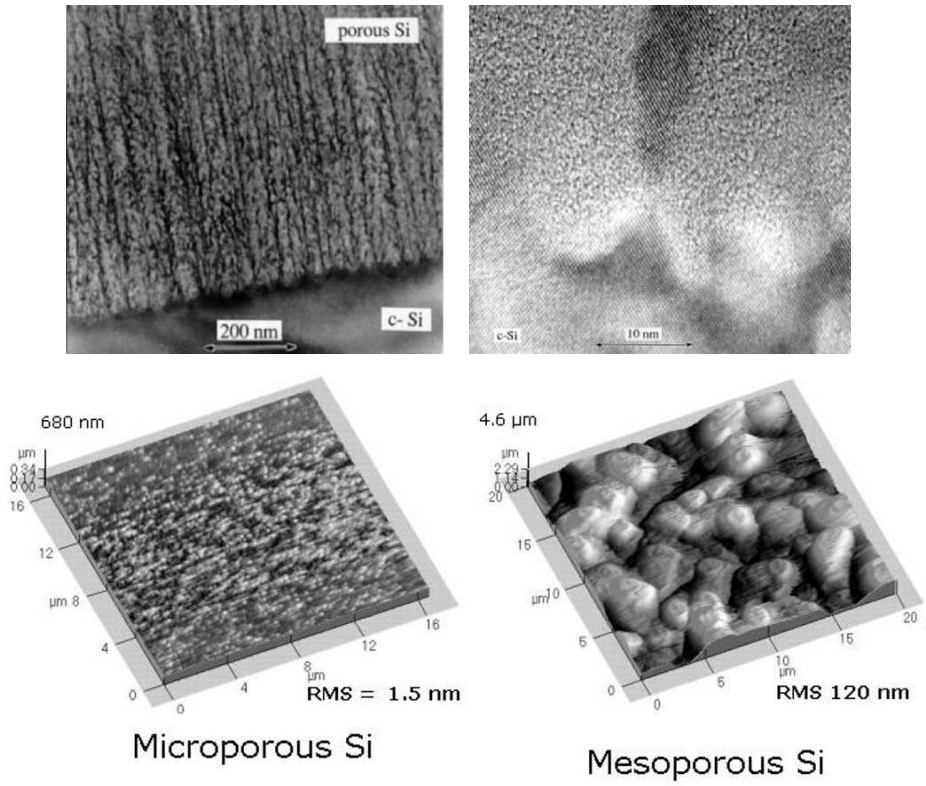


*Figure1.*Top: TEM (left) and HRTEM (right) image of the mesoporous-crystalline Si interface; bottom: AFM surface images of a microporous (left) and a mesoporous (right) [6].

The porosity P of the material is defined as relation of the volume of voids devided by the total volume and can be determined by measuring the

weight of a sample at different stages during the fabrication process using eq. (1):

$$P = \frac{Volume(voids)}{Volume(material)} = \frac{\text{mass}_{\text{origin}} - \text{m}_{\text{anodized}}}{(\text{mass}_{\text{origin}} - \text{m}_{\text{dissolved}})} \quad (1)$$

Here $\text{mass}_{\text{origin}}$ is the starting mass, $\text{m}_{\text{anodized}}$ the mass after anodization and $\text{m}_{\text{dissolved}}$ the mass after dissolving porous Si selectively to c-Si (e.g. in a weakly concentrated KOH). Typical values for the porosity are between 50 – 80 %.

Many material properties can be tailored over a wide range by the pore formation process and thus by the pore size and shape. For MEMS, the most important properties of porous silicon are: low thermal conductivity (down to 1 W/ (m K) [7]), yield strength from 83 GPa at 20 % porosity to 0.87 GPa at 90 % porosity (crystalline Si: 160 GPa) [8]. Additionally, impedance and permittivity is heavily changed by humidity and the refractive index (n = 1.5 – 3.0) and band gap (1.4 – 2.0 eV) is tunable for optical and biological applications.

Porous Si is formed by an anodization process, usually in a mixture of HF/ethanol/H_2O either in a double cell with electrolytical backside contact (allowing additional backside illumination of the wafer during the formation process) or in a single cell arrangement with solid backside contact. The structure of the porous layer, its pore size and porosity, which strongly influence the above mentioned material properties, are controlled by the anodization parameters, especially the current density, as well as by the doping type and doping level quality of the Si crystal. Typical values are 2-20 nm and 50-80 % for the pore size and porosity, respectively. For a flexible choice of the current density as the most critical process parameter it is favourable to use a PC controlled programmable current source allowing dc-operation with pre-selected steps of current density during the formation process or ac-like operation (e.g. continuous switching between low and high current density). This allows the formation of multilayers with different porosity, which is of crucial importance in the formation of optical elements (e.g. filters, reflectors, waveguides). A typical set-up for a two cells arrangement is shown in Figure 2. Suitable masking layers, which will prohibit anodization of c-Si are Si_3N_4, Au or Pt.

It should be noted that pore formation is found only in a certain range of HF-concentration, current density and doping concentration. Generally, pore formation is favoured at low current densities (10 mA/cm²) and high HF concentration (25% H_2O, 50 % ethanol and 25 % HF) whereas at high current densities and low HF concentration electropolishing is observed [9]. In a

medium range of doping n-Si is stable in the anodization process (without additional light) while p-Si is transformed to porous Si.

Rinsing after porous silicon formation is a very critical process step especially for thick porous Si films because cracks can occur in the formed porous Si layer thus damaging the fragile layer. Good results have been obtained with solvents, which reduce the surface tension such as penthane or ethanol. However, the stability of the formed porous silicon can be also increased by a suitable choice of the current density. Especially a two step process with very low current density at the beginning of the pore formation or a process periodically switching between high and low current (e.g. 2 min at 5 mA/cm², 1 min at 15 mA/cm²) will result in very reliable porous layers.

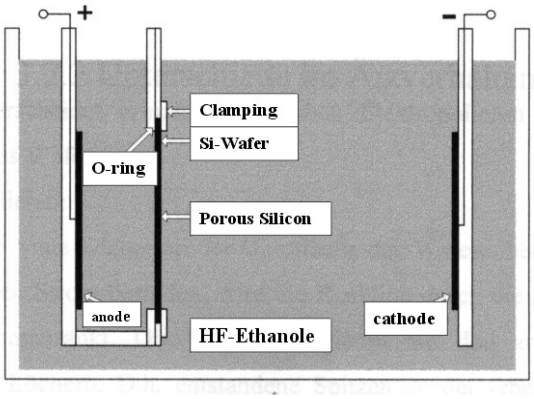


Figure 2. Two-cells arrangement for anodization of crystalline silicon. The Si-wafer is separating the two cells. Therefore, the electrolyte on the left-hand side of the wafer is forming the anode. Special care is needed to avoid any leakage current so that the whole current is flowing uniformly through the wafer.

3. POROUS SI AS STRUCTURAL MATERIAL FOR MICROMACHINING

Surface micromachining using porous Si as sacrifical layer

This process is based on the dependence of porous Si formation on the doping concentration in a standard anodization process. The finally free-standing structures are defined by low doped n-Si regions in the p-substrate. The main idea of this special surface micromachining process is that porous Si is formed in p-Si whereas low doped n-Si is stable. Due to its porous structure the formed porous Si can be removed selectively in respect to c-Si. Different authors have used this approach for the fabrication of c-Si free-standing structures [10], [11]. Here, an approach for definition of well-

defined and thin c-Si free-standing structures is shown [12]. As n-Si within a certain doping concentration range is stable during the anodization process whereas p-Si is transformed at a high rate to porous Si, it is possible to define free-standing n-Si-structures by an exact control of the doping concentration in n-Si. For this purpose, ion implantation is used to define n-doped islands in the p-Si substrate as shown in Figure 3.

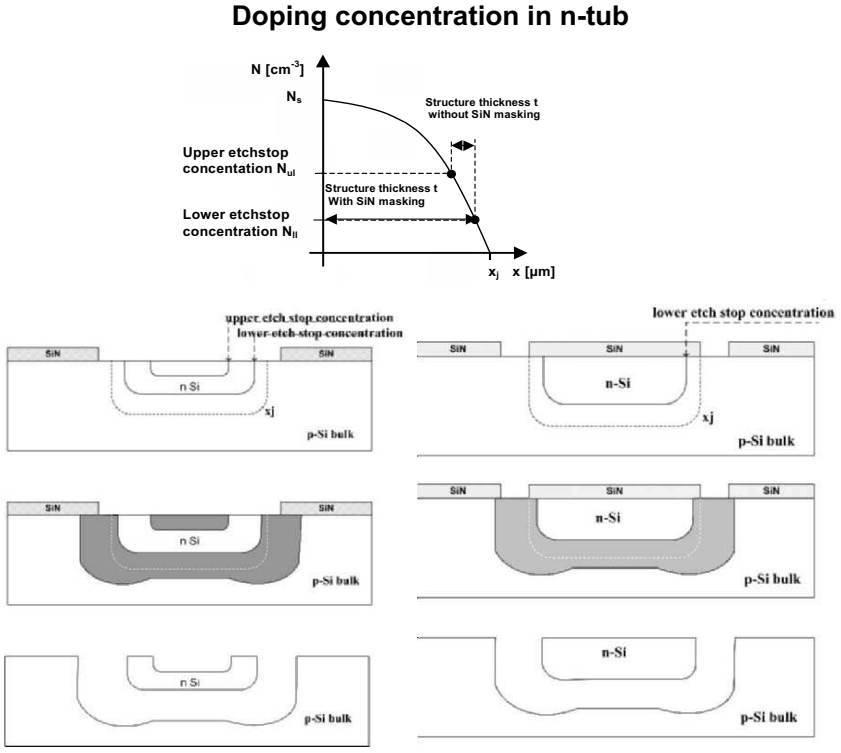


Figure 3. Principle of a surface micromachining process using porous Silicon as sacrificial material for the fabrication of free-standing crystalline Si structures. The structure thickness is defined by the doping profile and the etch stop concentrations N_{ul} and N_{ll}. Without masking layer also anodization of the highly doped n-Si surface occurs, resulting in a very thin structure whereas for a protected n-Si surface anodization takes place only from the bottom of the n-doped tub [12].

In the top part of Figure 3 a typical doping profile is shown. For n-Si two etch stop concentration are found (N_{ul}: upper etch stop concentration, N_{ll}: lower etch stop concentration). In between these concentration n-Si is stable during the anodization process. The final thickness (indicated by the arrows) is either given by the z-difference of this etch stop concentrations (without Si_3N_4 masking layer) or by the depth of the lower etch stop concentration (in the case of a Si_3N_4 masking layer). This effect can be used to define the resulting thickness of the free-standing structure exactly: Without a masking

layer on the n-Si, a limited anodization of the n-Si is found at the surface (very large doping concentration) and at the interface to p-Si (very low doping concentration near the pn-junction). A Si_3N_4 masking layer (Figure 3, bottom right) will prevent the anodization of the n-Si at the surface. Using these principle ideas and ion implantation to define the n-islands, excellent thickness control for free-standing structures is possible as shown in Figure 4. As can be seen on the right hand side of Figure 4, the lower and upper etch stop concentration do not depend on the junction depth. Therefore, it is possible to define the final thickness of the freestanding structures simply by shifting the pn-junction to a certain depth (by implantation energy, anneal time and temperature). As a result a linear dependence of final structure thickness (stable n-Si) on the junction depth is observed (Figure 4, right). The precision can be characterized by a standard deviation of 63 nm, including measurement accuracy.

Therefore, this process provides an accurate definition of free-standing, monocrystalline structures out of low doped n-Si. Some examples of free-standing structures are shown in Figure 4 (bottom).

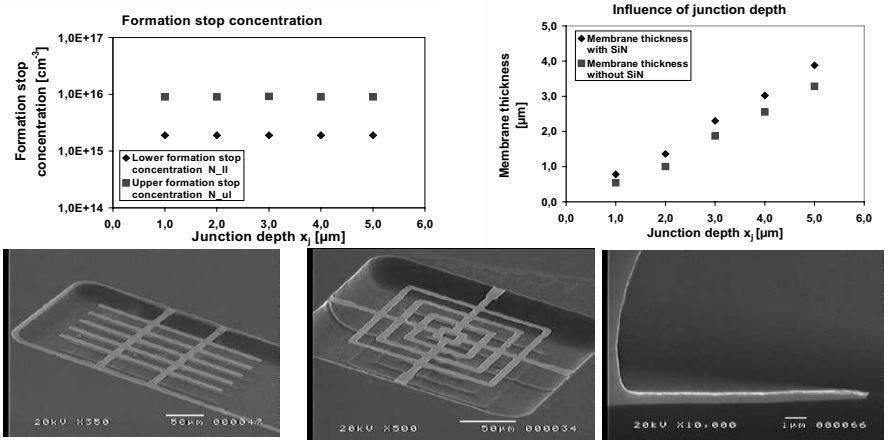


*Figure 4***.** Top left: Etch stop concentration as a function of junction depth x_j. Final thickness of the n-doped crystalline Si structure (top right) and some examples of free-standing structures formed with this process (bottom) [12].

Fabrication of micromoulds by adjustable anisotropy factor

The anodization rate of the pore formation process depends on the used current density. By local adjustment of the current density it is possible to change the resulting thickness of the porous Si layer locally, i.e. the electrochemical anodization process provides the possibility to adjust the anisotropy factor of this process. Basically, this process is based on the idea of a current concentration effect [13], [14]. Making use of the dependence of the anodization rate on current density, the electrochemical anodization of sili-

con can be used for 3D-structuring of crystalline Si and the fabrication of Si based micromoulds. These molds can be used for molding of plastic. The corresponding process is based on the selective pore formation in p-Si in a HF/ethanol/water mixture as already described in the section 2. Using a structured masking layer such as silicon-nitride and a highly doped p^+-Si layer as back electrode of the wafer, the current density provided through the back electrode into the p-Si wafer can be adjusted locally by the size and the areal density of the holes in the insulating silicon-nitride layer on the top side of the wafer (Figure 5). After removal of the produced porous Si in weakly concentrated KOH a mold out of the remaining crystalline p-Si is resolved. In a more sophisticated approach, a further control of the current density is possible by a structured backside contact. The layout in Figure 5 is only a schematic representation of the 2D-hole pattern needed for the final 3D-profile. The local current density is mainly determined by the size and the areal density of the masking patterns in the neighborhood of a specific opening. An additional adjustment of the current density and thus the local anodization rate is achieved by local doping of the backside. By switching the current density at the end of the anodization process to values well above 100 mA/cm², the final interface between porous Si and the p-Si substrate can be electropolished [15] already during the anodization process [16], [17].

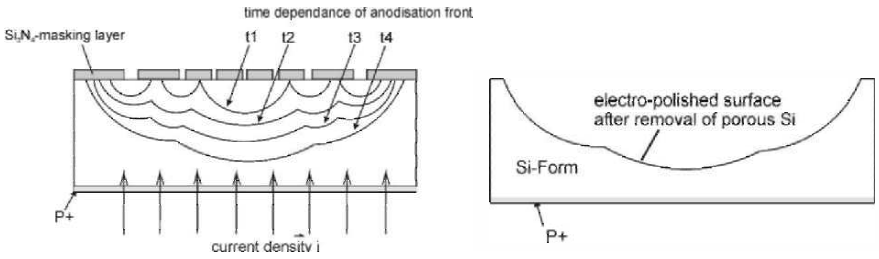


Figure 5. Principle idea of 3D- structuring through a 2D- mask by making use of the current concentration effect during the anodization of crystalline Si.

To investigate the fundamental dependence of anodization depth on layout a simulation model has been developed. Smith et al. [9] have shown that the electrolyte/Si interface can be modelled in a first approximation as Schottky contact and that the porous Si formation process is controlled by the anodization current. Therefore, we have used the device simulator TMA MEDICI (version 4.0) to investigate the layout dependence of the current density. Some typical simulation results are shown in Figure 6. On the right hand side, the current concentration effect is shown for a single opening. A current redistribution is found when many openings are used. Figure 6 (top right) shows as example the situation for two openings. One opening is 50 µm wide, the other 100 µm, the gap in between was chosen as 100 µm. The

current concentration effect is seen as higher current density in the smaller opening. As can be seen, pronounced current concentration effects have to be considered at the edges of the openings. A possibility to solve this problem is to define additional opening in the back-side of the wafer. With a proper set of geometries the edge effect can be considerably reduced.

Experimental results for the anodization depth (corresponding to the depth of the mould after removing the porous Si) are shown on bottom of Figure 6. Here the normalized anodization depth is shown as function of the width of openings. The depth was measured in the centre of an opening. Therefore, the edge effect will not disturb the results. Different depth from 6 µm to 32 µm have been used. The openings were well separated to each other. It should be noted that the current concentration effect is increasing with decreasing width of the structures.

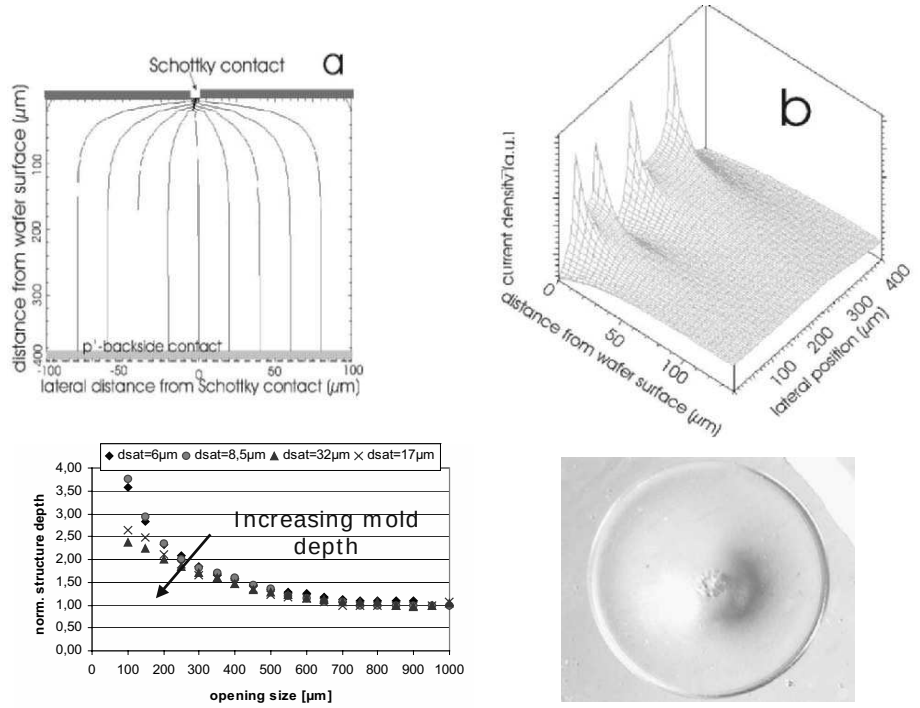


Figure 6. Top: Medici-simulation of the current concentration effect. Top left: single hole, top right: two holes (50 µm and 100 µm wide) in a distance of 100 µm, besides the current concentration effect (larger current density in the centre of the structure increases with decreasing feature size) a pronounced edge effect is observed. Bottom: experimental results. left: normalized etch depth as function of opening size, right: PMMA structure replicated out of a Si-mould.

However, reliable measurements of deep and narrow holes are difficult. Therefore, this results were not considered in Figure 6. The changes of etch rate correspond to a change of the anisotropy factor $A_f = 0 – 0.7$. Therefore, a standard 2D-pattern definition process with mask based lithography and etching of the masking silicon-nitride can be used for structuring of complicated 3D-profiles. These 3D-structures in silicon can be used as molds for a transfer process into plastic using e.g. injection molding. As an example, a PMMA structure in shown in Figure 6, which was replicated out of a Si-mould. The diameter of the shown structure is around 1 mm. As can be seen, this technique can be used for complicated surfaces like aspheric lenses.

4. POROUS SI AS MULTIFUNCTIONAL MATERIAL

4.1 Application in Optics

As already mentioned photo- and electroluminescence was among the first properties, which have made porous Si interesting for R&D investigations. However, after first enthusiastic approaches of “all-in-Si-optoelectronics” the extremely low efficiency of light emitting devices (quantum efficiency < 0,1%) has prohibited commercialization of such devices. In the meantime, other properties have become more interesting: Being the porous layer a mixture of two phases (crystalline Si skeleton and void) the effective refractive index of the layer can be adjusted by selecting appropriate porosity. A large variety of optical elements can be formed by controlled single or multi-step etching process. An example is the application of porous Si as antireflective coating on multi-crystalline Si solar cells. Layers with the ideal parameters (refractive index of 1.94 at $\lambda = 633$ nm and thickness of 74 nm) can be formed by partial etching of the emitter as a final step in the solar cell processing [18]. Another example of an optical application is the realization of Bragg filters by multilayers of porous Si, each layer with a different porousity [19]. Examples of such multilayer optical filters or Bragg-reflectors are shown in Figure 7 [20]. By an appropriate choice of the layer thickness and refractive index, the reflectivity can be turned to almost 100 % in a certain wavelength range (left) or a special wavelength is filtered out of the reflected spectrum (right).

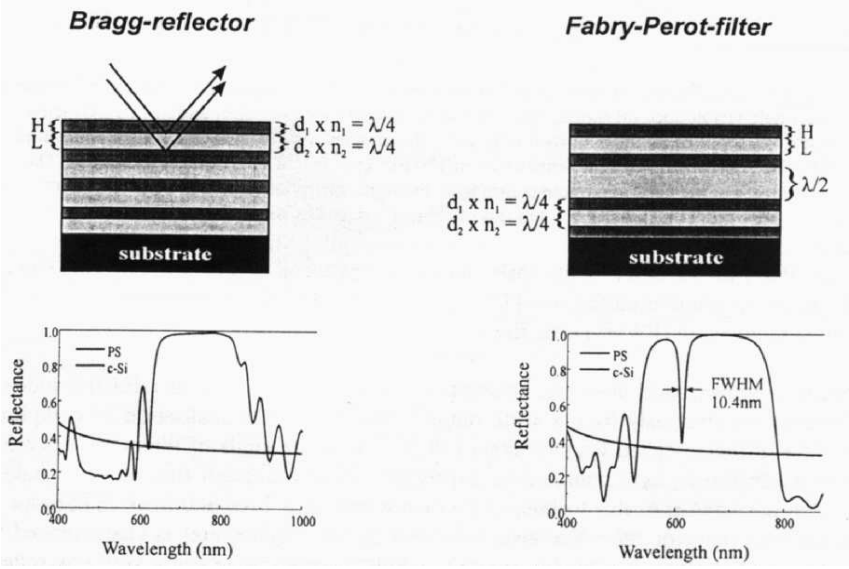


Figure 7. Example of optoelectronic applications of porous Silicon as Bragg-reflector (left) or Fabry-Perot-filter, out of [20].

4.2 Application in Sensors

Due to its extremely high inner surface (1 cm^3 corresponds to about 250-400 m^2), porous Si can be used as sensitive material for different types of measurands. Several groups have realized humidity sensors based on porous Si [21-27]. In principle resistive or capacitive effects can be utilized for measuring the take up of humidity in the porous layer. Besides optimization of pore size and porosity in respect to sensitivity and dynamics of such sensors, reliable contacts on porous Si are essential for such sensors. In a new approach [27] the electrodes to porous Si are realized via n-doped islands in p-doped Si, which is later transformed to porous Si. This contact concept allows for the realization of interdigitated electrodes embedded in the porous layer. Another advantage of this set-up is that additional heating elements can be integrated in the sensing material too. As humidity sensors normally need heating for refresh and porous Si has a very low thermal conductivity, this approach ensures low power consumption.

In Figure 8, SEM-photos of humidity sensors using porous Si are shown [27]. Heater elements can be placed directly within the interdigitated electrodes on the porous Si or on the outer frame around the sensing area. As expected the sensor shows an extremely large sensitivity as can be seen in a semi-logarithmic scale (relative change of capacitance of 5000 %). Hystere-

sis often found in humidity sensors can be avoided by suitable heating of the sensor prior to measurement.

In the cross section of Figure 8 needle type areas beneath the electrodes are observed. This needles (remaining c-Si) are caused by current shortcut which will stop a such a point any anodization of porous Si. A bonded sensor is shown in Figure 9. For hysteresis free operation it is important to use proper heating and refresh procedures during measurements. Good results have been obtained with short (20 s) heating pulses (120 mW) just before each measurement.

Combining the use of porous Si as structural and functional material also new types of highly integrated sensor systems are possible such as comfort index sensors, which can be used e.g. in automotive applications (air conditioning systems in cars) or health monitoring in material science.

Several other types of sensors using porous Si have been presented in the last years. Very interesting applications of porous Si have been shown for biosensors. Both, immobilization of various biomolecules [28] and sensing of biochemical reactions have been shown using porous Si. Combining different properties of porous Si, a biosensor with an optical read-out has been realized [29].

Also chemical sensors based on porous Silicon have been investigated. In [30], a catalytic oxidation reaction for H_2 has been shown with Pd-doped porous Si. Such reactions can be also used for fuel cells, which are discussed in the next section.

4.3 Application in Fuel Cells

A very new application is the use of porous Si for fuel cells. First results were published in 2002 and 2003. Porous Si can help to achieve compact and high efficient energy supply e.g. for laptops and other mobile electronic equipment or for auxiliary power supply in automotive applications.

Porous Si can take up different tasks in a fuel cell: Konle et al. [31] have used porous Si in a micro-steam fuel reformer, which is part of a whole fuel cell system. The reformer is made out of macroporous Si and transfers gasoline into H_2, which is than used in a standard fuel cell. A porous Si reformer can help to overcome a temperature limitation: The main problem of using gasoline for fuel cell supply is the high reaction temperature needed for the H_2-conversion: At around 800 °C, commonly used catalyst templates like aluminum or tempered steel are not suitable. Therefore, a reformer built up by a stack of micro porous Si has been developed. The whole system, a prototype of the reformer and so-called shower-head structure of porous Si prepared with pre-patterned KOH-pits are shown in Figure 10.

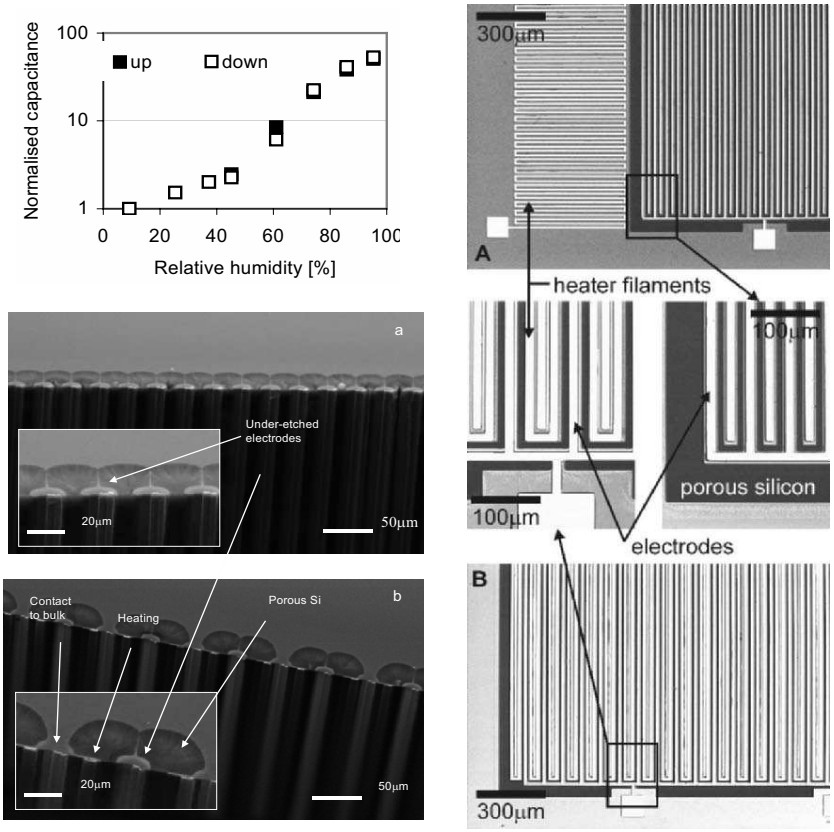


Figure 8. Capacative humidity sensor with porous Silicon as sensitive material [27]. Top left: Normalised capacitance as function of humidity (semi-logarithmic scale, right: Layout of interdigitated electrodes and heater geometry (either outside the sensitive area or within the interdigitated electrodes. Bottom left: SEM-details of the sensor. Both electrodes are formed on the porous layer (top) or one is formed on the substrate to increase reliability (bottom).

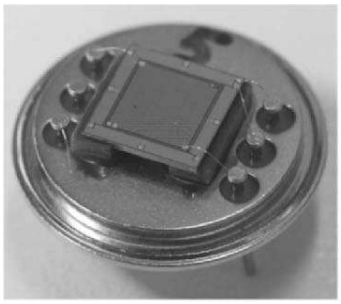


Figure 9. Capacative humidity sensor based on porous Si. The interdigitated sensor electrodes cover the central part. Additional wire connections are provided for the heater. The sensor chip is glued to a bridge type substrate to reduce thermal conductivity to the frame.

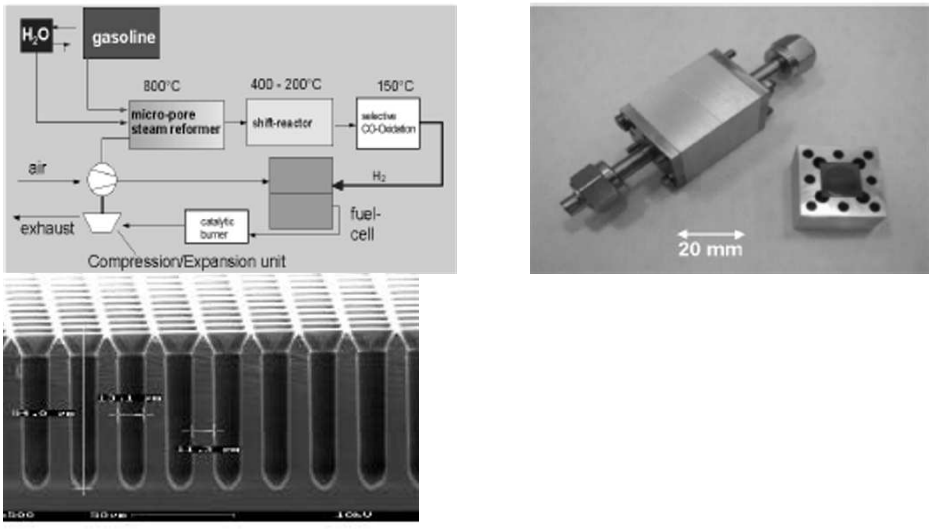


Figure 10. Operation fuel cells with diesel or gasoline, reformers are needed. Left: set-up of such a system, centre: reformer prototype made out of porous Si, right: ordered micro-pores by a pre-patterning KOH etch step prior to the pore formation, out of [31].

As mentioned in the previous section, porous Si can be used as a solid state based electrolytic material in a fuel cell. Neah power systems has just announced a fuel cell based on porous Si [32]. Exchanging the polymer in a PEM-based fuel cell by porous Si they claim a tremendous increase of power density in their new fuel cell. Some features of this fuel cell are shown in Figure 11. The whole wafer is anodized and used as electrolyte (left) instead of a only 10 μm thick polymer membrane in a standard PEM. Porous Silicon has a honeycomb-like structure, the pores are deposited with a catalytic material (right). In the real system, the pores are filled with a liquid electrolyte - similar to water being absorbed into a sponge. This allows for higher levels of exposure of fuel, oxygen, and catalyst.

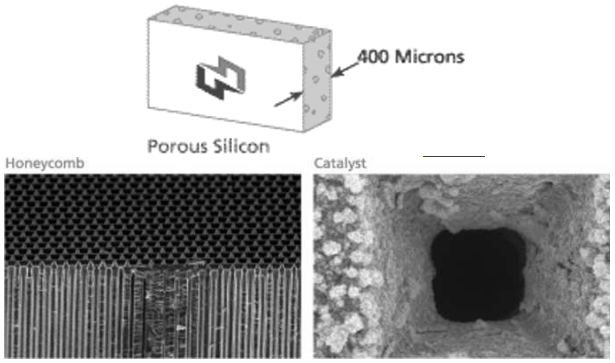


Figure 11. New approach for a porous Si based fuel cell, out of [32].

ACKNOWLEDGEMENTS

The author thanks Dr. Andras Kovacs (FH Furtwangen) for the valuable contributions to the work on porous silicon. Contributions to the presented results by Dr. W. Kronast, B. Müller, Dr. R. Huster, MSc cand. A. Fahad and Valero and MSc. T. Suranyi (all FH Furtwangen) are greatly acknowledged. Thanks also to Dr. I. Bársony, Dr. C. Düscö, P. Fürjes and M. Ádám (all MFA, Budapest) for the support in cooperative projects. The cooperation between the FHF and MFA has been supported by the German-Hungarian bilateral research programme of the BMBF, Germany, and the Education Ministry of Hungary under contracts UNG-005-99 and D-1/99, respectively. The work on micromolds has been supported by the German Ministry for Education and Research (BMBF) under contract 1702601.

REFERENCES

[1] Pickering C., Beale M.I.J., Robbins D.J., Pearson P.J., Greef, R. J., Optical studies of the structure of porous silicon films formed in p-type degenerate and non-degenerate silicon. J. Phys. C 1984; 17:6535-52

[2] Halimaoui, A. et al. Electroluminescence in the visible range during anodic oxidation of porous silicon films. Appl. Phys. Letters 1991; 59:304-6

[3] Canham, L.T. Luminescence bands and their proposed origins in highly porous silicon Phys. Status Solidi B 1995; 190: 9-14

[4] Canham, L.T. "Pore type, shape, size, volume and surface area in porous silicon". In *Properties of Porous Silicon*, L.T. Canham ed., London: INSPEC Publication, 1997.

[5] Lehmann, U. "A new capacitor technology based on porous Silicon". In *Properties of Porous Silicon*, L.T. Canham ed., INSPEC Publication, 1997

[6] Mescheder, U.M., Kovacs, A., Kronast, W., Bársony, I., Ádám, M. and Dücsö, Cs Porous silicon as multifunctional material in MEMS, Proceedings of 1st IEEE-Nanotechnology, Maui (USA), 2001.

[7] Lang, W. "Thermal conductivity of porous Silicon" In *Properties of Porous Silicon*, L.T. Canham ed., London: INSPEC Publication, 1997.

[8] Bellett, D. "Elastic properties of porous silicon" in *Properties of Porous Silicon*L.T. Canham ed., London: INSPEC Publication, 1997.

[9] Smith, R.L, Collins S.D. PS formation mechanism. Journal of Applied Physics 1992; 71(8):R1-R22

[10] Steiner P., Richter A., Lang W. Using porous silicon as a sacrificial layer, J. Micromech. Microeng. 1993; 3:32-6

[11] French, P.J. et al, New silicon micromachining technique for microsystems, Sensors and Actuators 1997; A62: 652-62

[12] Kovacs A., Mescheder U. Surface Micromachining with c-Si as functional layer using porous Si as sacrificial layer, in Digest of Technical Papers of the Transducers ´01, Eurosensors XV, Munich, Germany, June 2001

[13] Hedrich F., Billat S., Lang W., Structuring of membrane sensors using sacrificial porous silicon, Sensors and Actuators 2000; 84: 315 – 23

[14] Steiner P., Lang W. Micromachining applications of porous silicon. Thin Solid Films 1995; 255:52-8

[15] Lammel G., Renaud P. Free-standing, mobile 3D microstructures of porous silicon, in Proceedings Eurosensors XIII, 1999, Den Hague, Netherlands: 535-6

[16] Mescheder, U. et al. 3D Structuring of c-Si using porous Silicon, proceedings 2nd IEEE conference on Nanotechnology, Washington, 2002: 33-6
[17] Verfahren zur Bildung einer Ausnehmung in der Oberfläche eines Werkstücks, insbesondere zur Herstellung von Mikroformen, DE 10234547.3, Deutsches Patentamt
[18] Vazsonyi. E. et al., High efficiency silicon PV cells with surface treatment by anodic etching, Proceedings 13th European Solar Energy Conference, Nice, France, Oct. 23-27, 1995:37-40
[19] Berger M.G. et al, Investigation and design of optical properties of porosity superlattices, Thin Solid Films, 1995; 255:313-6
[20] Thönissen, M., Krüger, M. "Optical application of porous Silicon", p.349-355 in *Properties of Porous Silicon*L.T. Canham ed., London: INSPEC Publication, 1997.
[21] O`Halloran G.M. et al: Porous silicon membrane for humidity sensing applications, in Proc. Eurosensors XII, Southamton, UK, Sept, 1998: 901-4
[22] Rittersma, Z.M. et al: A novel surface micromachined capacitive porous silicon humidity sensor, Sensors and Actuators B 2000; 68: 210-7
[23] O`Halloran G.M. et al: Influence of the Formation Parameters on the Humidity Sensing Characteristics of a Capacitive Humidity Sensor Based on Porous silicon, in Proc. 13 th Eurosensors XIII 1999 The Hague, The Netherlands:55-6
[24] Rittersma, Z. M. *Microsensor Applications of Porous Silicon*, Ph.D. Thesis, Universität Bremen, Deutschland, 1999
[25] Connolly, E.J. et al: Comparison of porous Si, porous poly-Si and porous SiC as material for humidity sensing materials, Sensors and Actuators, A 2002; 99: 25-30
[26] Mai, Le Hoang, Hoa, Pham Thi Mai, Binh, Nguyen Tien, Ha, Nguyen Thi Thu, An, Dao Khac. Some investigation results of the instability of humidity sensors based on alumina and porous silicon materials, Sensors and Actuators B 2000; 66: 63–5
[27] Fürjes P. Kovács A., Dücső Cs., Ádám M, Müller B. and Mescheder U. Porous Silicon Based Humidity Sensor with Interdigitated Electrodes and Internal Heaters, in Proceedings Eurosensors XVI, Prague Sept. 2002:605-6
[28] Laurell T. et al. Enhanced enzyme activity in silicon integrated enzyme reactors utilising porous silicon as coupling matrix, Sensors&Actuators B 1996; 31:161-166
[29] Lin V.S.-Y. et al., A porous silicon-based optical interferometric Biosensor, Science 1997; 278, 840-3
[30] Tsamis, C. et al. Hydrogen Catalytic Oxidation Reaction on Pd-doped porous Silicon, IEEE Sensor Journal 2002; 2:89-95
[31] Konle J. et al. Nano-sized pore formation in p-type silicon for automotive applications, in Proceedings IEEE-Nano 2002.
[32] http://www.neahpower.com/technology/oursolution.shtml, February 2003

Chapter 10
THERMAL DESIGN OF MICROELECTRONICS SYSTEMS

Dmytro Fedasyuk [a,b] and Volodymyr Makar [a]

[a] *CAD Departament, Institute of Computer Sciences and Information Technologies, National University Lviv Polythechnic, Lviv, Ukraine*
[b] *Department of Microelectronics and Information System, Technical University of Lodz, Poland*

Abstract: The mathematical models and methods for thermal analysis of two typical constructions of microelectronics systems have been presented in this chapter. The 3D thermal model of hybrid circuits in the form of multilayer structure with flat and add-on heat sources have been developed. The combination of numerical and analytical methods for investigation of this model has been used which allows to solve both the steady-state and transient thermal problems simultaneously. The mathematical models of the steady-state and transient heat exchange processes in the flip-chip structures, based on the analytical solution of the conjugate transient heat exchange problem for the die and substrate have been described. The appropriate software in the form of CAD tools have been developed which was tested by solving of several test problems. The estimation of accuracy and efficiency of developed models have been done by comparison with the results obtained by another authors with use of numerical methods. These comparisons show some advantages of analytical approach, namely considerably less computational cost, no influence of discretization errors and possibility to enlarge the sphere of thermal design problems.

Keywords: multichip modules, hybrid circuits, flip-chip, thermal resistance, steady-state and transient thermal problems, Fourier series.

1. INTRODUCTION

In general process of designing of microelectronics systems (MES) considerable attention should be paid to the thermal design of those, because it is one of the limitations how small and dense electronics can be produced. Thermal design is the process by which engineers use simulation software tools to predict and uncover potential thermally associated risk areas and to obtain desirable and correct solution as fast as possible in the product design cycle. The ultimate goal of these efforts is to provide the optimal design that meet design requirements and reliability qualifications.

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 289-329.

Many tools exist to help thermal design user during this process [1]. However, on the one hand, some of them are too universal, that is, are not specifically developed for application in the microelectronics, on the other hand, some of them are particular special-purpose and can be only used to solve of narrow range of problems. So, thermal design engineers have been still experienced some difficulties. One of the main reasons for this is the enormous variety of sorts of MES differing by package types, kinds of heat exchange, ways of chip settings to carrier, quantity of pins etc.

2. REVIEW OF MES

Here we just intend to present some common results of the analysis of the MES structure types and features of their thermal regimes. The different types of chips settings to carrier are shown in Figure 1.

MES's can have a cavity up configuration or a cavity down configuration. Some structures use special methods of heat source attach which decrease the thermal resistance of junction and heat sink for increasing the heat flow from the IC chip through chip carriers and seal lid to ambient.

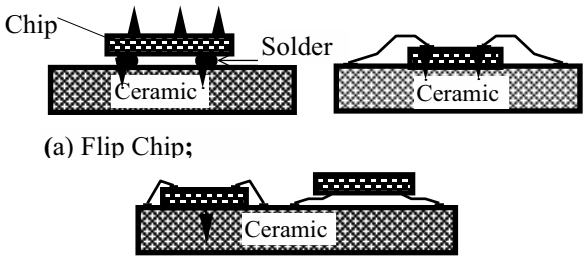


Figure 1. Heat Removal Depends on the Chip Bonding.

The structure of multichip modules with seal lid and heat sink attach is shown on Figure 2. The heat flow paths for the structure are shown on Figure 3.

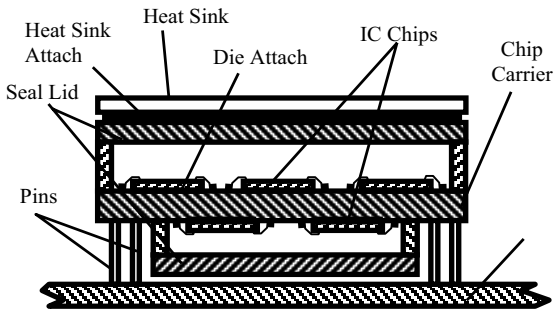


Figure 2. Multichip Modules with Bilateral Attach of IC Chips and Seal Lid.

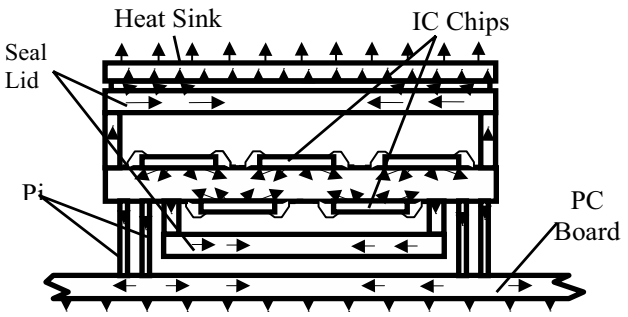


Figure 3. Heat Flow Paths in Multichip Modules.

The analysis of the MCM's structure types and the heat exchange process allows to distinguish three kinds of heat exchange: conduction, convection and heat radiation. From IC chip to substrate the heat is removed by conduction. From the substrate heat is removed by conduction to pins and to PC board. Parallel heat flow is through conduction to seal lid and to the ambient from the package by convection and radiation. In this article the heat transfer arising from the radiation - because in the examined temperature range it has no significant contribution - is neglected.

There are kinds of MCM's structure, in which the heat exchange is realized both from the bottom to the heat sink and from the upper seal lid to the heat sink by conduction and convection. The input data for thermal field simulation have to include: the values of chip power dissipation; the geometrical parameters; the thermal-physical parameters of structure's materials. For the analysis of thermal fields in MCM's one has to take into account the heat exchange through pins to PC board and from the package through seal lid to environment. The values of heat transfers through each element of a structure are correlated among them and depend on the thermal conductivity and thickness of the substrate, type of attach, number of pins and their position, size of seal lid. For example, if the thermal conductivity coefficient of the substrate and package is small then the heat flow value through the pins will be large. Therefore we have to take into account every kind of thermal removal in the process of the thermal model creating [2].

3. THERMAL DESIGN PROBLEMS

Thermal design includes two interrelated tasks:

- analysis of the systems design under the influence of temperature field;
- synthesis of systems under the influence of temperature field.

The analysis of design solutions requires the modeling and investigation of the temperature fields of systems and their influence on functional parameter. The synthesis of design solutions requires optimum functional

and construction parameters, which provide certain distribution of temperature fields within a systems [3].

The thermal simulation problem includes the following tasks:

- analysis of steady-state temperature fields;
- analysis of transient temperature fields;
- analysis of electronic circuits being influenced by temperature fields;
- analysis of electronics circuits being under external thermal influences;
- analysis of stability and sensitivity of functional parameters under the influence of temperature fields;
- statistical analysis of temperature fields;
- simulation of electronic system reliability under the influence of existing temperature fields.

The methodology for thermal simulation includes these steps:

- qualitative analysis of electrical circuit, structure and environment of the electronic system;
- development of a thermal model of the system;
- formalization of mathematical model of temperature field with initial and boundary conditions;
- development or selection of the analysis method for a mathematical model of temperature fields;
- software development;
- the temperature field simulation and analysis.

4. THERMAL SIMULATION OF HYBRID CIRCUITS

Formalization of the problem of the MES's thermal field simulation is based on the description of the thermal conduction process in the general thermal model (Figure 4) within the assumptions about the heat exchange with the environment by the system of initial and boundary conditions. (A_x is the model dimension along X axis; B_y is the model dimension along Y axis; X_c is the heat source center coordinate along X axis; Y_c is the heat source center coordinate along axis; a_x is the dimension of the heat source element along X axis; b_y is the dimension of the heat source element along Y axis; h_j is the j-th layer thickness; λ_j is the j-th layer thermal conductivity coefficient; C_j is the j-th layer specific heat; ρ_j is the j-th layer specific density).

The fundamental principle of the proposed solution method is based on the decomposition of the original problem into a series of subproblems, which allow to obtain analytical solutions at the lowest level and then combination of these solutions into a general one by using the numerical approach.

To formalize the boundary conditions of various types we introduce them in the following form [4]:

$$\Phi[1,S] = T\,; \quad \Phi[2,S] = \frac{\partial T}{\partial S} \tag{4.1}$$

$$\Phi[3,S] = \lambda \frac{\partial T}{\partial S} + \alpha_S (T - T_a)\,, \tag{4.2}$$

where 1, 2, 3 are denote of boundary conditions types; S is the value of coordinates, $S \in (x, y, z)$; T is the temperature; λ is the thermal conductivity coefficient; α_S is the convective heat transfer coefficient; T_a is the ambient temperature.

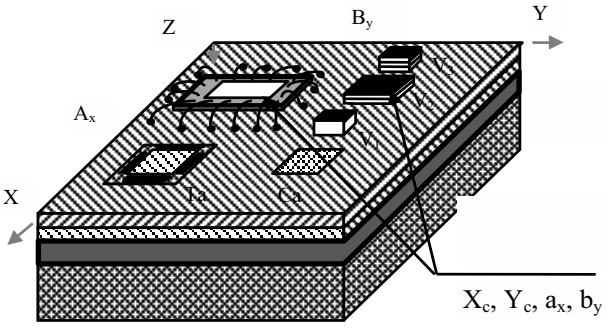


Figure 4. General Model of the MES's Structure.

4.1 Mathematical Description of Thermal Model

Mathematically, the process of transient thermal conductance can be described by the Fourier differential equation for the *j*-th layer:

$$c_j \rho_j \frac{\partial T^{(j)}}{\partial t} = \lambda_j \left(\frac{\partial^2 T^{(j)}}{\partial x^2} + \frac{\partial^2 T^{(j)}}{\partial y^2} + \frac{\partial^2 T^{(j)}}{\partial z^2} \right) \tag{4.3}$$

The initial condition has the form:

$$T(x, y, z, 0) = T_a(x, y, z) \tag{4.4}$$

To obtain unique solution for the entire structure, the particular solutions per layer are combined into a general solution by using the boundary conditions for the adjacent layer for $T(x, y, z)$ and flows $\lambda \frac{\partial T}{\partial z}$

$$T^{(j)}(x, y, z_j, t) = T^{(j+1)}(x, y, z_j, t) \tag{4.5}$$

$$\lambda_j \left.\frac{\partial T^{(j)}}{\partial z}\right|_{z=z_j} = \lambda_{j+1} \left.\frac{\partial T^{(j+1)}}{\partial z}\right|_{z=z_j} \tag{4.6}$$

4.2 Solution of the Thermal Problem by a Combined Method

The solution of the problem, which has been formalized above is realized with the help of a specially developed method which is a combination of numerical and analytical methods. These techniques allow us to obtain a generalized solution which makes it possible to analyze both steady-state and transient thermal fields in a 3-D multilayer structure with various boundary conditions. The sequence of steps is the following:

A. Reduction of a transient equation to a quasi-steady-state form

To decrease the dimension of the differential problem and to obtain similar solutions for both transient and steady-state cases we use the method of planes for the initial equation transformation. In particular we use the transverse scheme of the method for the parabolic equation having substituted the time derivative by the left-hand difference relation

$$\frac{\partial T^{(j)}}{\partial t} = \frac{T_{i+1}^{(j)} - T_i^{(j)}}{\tau}, \tag{4.7}$$

where $\tau = t_{i+1} - t_i$, resulting in the following equation

$$\frac{c_j \rho_j}{\lambda_j \tau} T_{i+1}^{(j)}(x, y, z, \tau) = \left(\frac{\partial^2 T_{i+1}^{(j)}}{\partial x^2} + \frac{\partial^2 T_{i+1}^{(j)}}{\partial y^2} + \frac{\partial^2 T_{i+1}^{(j)}}{\partial z^2} \right) + Q^{(j)}(x, y, z, \tau), \tag{4.8}$$

where

$$Q^{(j)}(x, y, z, \tau) = T_i^{(j)}(x, y, z, \tau) \frac{c_j \rho_j}{\lambda_j \tau} \tag{4.9}$$

is the nonhomogeneous part of equation (4.8). Thus, the original problem falls into k boundary problems for the differential equation (4.8). As the function $T_a(x, y, z)$ is known from the initial condition, the unknown functions $T_k(x, y, z, \tau),\ k = i + 1$. K can be found sequentially by solving

for every value of $k = 1, i$ one differential equation of the form (4.8) with the same boundary conditions. For a steady-state case when $T_{i+1}^{j} - T_{i}^{j} = 0$ we get well-known Laplace's equation.

B. Obtaining the solution in the form of twofold series of eigenfunctions

The solution to a nonhomogeneous equation (4.8), according to Steklov theorem, may be presented in the form of a twofold Fourier series in terms of eigenfunctions of the corresponding homogeneous problem:

$$\mathrm{T_k^{(j)}(x,y,z,\tau) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} Z_{nmk}^{(j)}(z,\tau)\Psi_n(x)\Psi_m(y)} \tag{4.10}$$

The form of eigenfunctions $\Psi_n(x)\Psi_m(y)$ and eigenvalues α_1, α_2 is obtained by solving ordinary differential equations, which have been obtained as a result of separation of variables with the account of the correspondingly assumed boundary conditions

$$\ddot{X} + \alpha_1^2 X = 0 \tag{4.11}$$

$$\ddot{Y} + \alpha_2^2 Y = 0 \tag{4.12}$$

The eigenfunctions and eigenvalues obtained as a result of solving the equations (4.11) and (4.12). have the following form:

C. Obtaining the recurrence relation for a transient problem

To find $Z_{nmk}^{(j)}(z,i)$ we make a substitution of (4.10) into (4.8) and having expanded $Q(x,y,z,\tau)$ into a series analogous to (4.10), we find its coefficients. We obtain a nonhomogeneous differential equation of the second order:

$$\mathrm{\ddot{Z}_{nmk}^{(j)} - \gamma^2 Z_{nmk}^{(j)} = a_{nmk}^{(j)}(z)} \tag{4.13}$$

Having solved the equation (4.13) by the constants variation method for every n and m at k-th time step we get the recurrence relation for a transient problem. $Z_{nmi}^{(j)}$ is the solution of the equation (4.13) obtained in the previous time step. At the first time step ($k = 1$, $i = 0$), $Z_{nmk}^{(j)}$ is determined from the initial condition (4.4) by its expansion into a series of the form (4.10) and by finding the coefficients:

$$Z_{nm0}^{(j)}(z,\tau)=\frac{1}{\left\|\Psi_n\right\|^2\left\|\Psi_m\right\|^2}\times\int_0^{A_x}\int_0^{B_y}T_a(x,y,z)\Psi_n(x)\Psi_m(y)dxdy \quad (4.14)$$

D. Per layer formation of the system of equations

The general solution of the mathematical model for the thermal conduction in a multilayer structure is obtained in the matrix form, which is formed and solved, proceeding from boundary conditions along Z coordinate. Consider now the formation of the system of equations. At the upper and lower surfaces of the structure the conditions of the 1-st, 2-nd or 3-rd type (homogeneous and nonhomogeneous) (4.1) and (4.2) are assigned. Using *2N* boundary conditions (4.5) and (4.6) along Z coordinate (*N* -the number of layers), we get the 2N x 2N system of equations.

Matrices reflect the heat exchange between the layers and with the medium along Z coordinate in accordance with the assigned boundary conditions. The first row of the matrix shows the type of heat exchange between the upper side and the medium. The second and the third rows and elements show the heat transfer between next structure element according to the hierarchy, or with the lower side and the medium-heat sink. Filling the matrices according to the boundary conditions and solving the system of equations, we form the mathematical model along Z coordinate.

E. The development of the integration procedure

Now we will describe the integration method along Z coordinate in the expression (4.13). In the methods presented here it is necessary to have fixed nodes (for example, the trapezoidal formula). The appearance of additional intermediate points results in the increase of the number of calculations. The method of spline quadratures allows for increase in the accuracy of integration and in decrease of the memory storage. Then

$$\int_a^b f(x)dx=\sum_{i=1}^{n}\Delta i\frac{f_i+f_{i+1}}{2}-\Delta^3 i\frac{\upsilon_i+\upsilon_{i+1}}{2}, \quad (4.15)$$

where υ_i coefficients at near second member of spline; f_i, f_{i+1} are the function value; is the integration step. Using this integration method we obtain the expression of integration coefficient $Z_{nmk}^{(j)}(z,i)$ along Z coordinate.

F. Approximation and eigenfunctions expansion

Now consider the way of approximating the function $T(x,y,z_N,t)$ to integrate it along X and Y and its further use in the mathematical model. The function of thermal field (distribution) at one of the sides is presented in the

form of the two-dimensional finite-element approximation. The domain is discretized into K rectangular finite elements, each being considered in a nornmalized system of coordinates. Four-node finite elements are used. Then for the i-th element, the function $E_l(x, y, t)$ is of the form:

$$\mathrm{E_l(x,y,t)} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{N_\alpha E_l^\alpha(t)}, \tag{4.16}$$

where E_l^α is the value of the function of a node α of the finite element.

The solution of the problem of thermal conduction in a multilayer structure by analytical methods has been obtained in the form of multiple trigonometric series. Consequently, it is necessary to get the Fourier series expansion of the function $E_l(x, y, t)$. The coefficients of the series $E_{nmk}^{(j)}(t)$; $n, m = 0, \infty$ are determined from

$$\mathrm{E_{nmk}^{(j)}(t)} = \frac{1}{\left\|\Psi_\mathrm{n}\right\|^2 \left\|\Psi_\mathrm{m}\right\|^2} \times \int_0^{\mathrm{A_x}} \int_0^{\mathrm{B_y}} \mathrm{E(x,y,t)\Psi_n(x)\Psi_m(y)dxdy} \tag{4.17}$$

The function $E_l(x, y, t)$ is represented in the discrete form (4.16). The formulas for the calculation of integrals in the local system of coordinates are obtained in an analytical form. The above methods of approximating of $T(x, y, z_N, t)$ in 3-D thermal conduction analytical models allow to take into account the irregularity (non-uniformity) of the function distribution in the Derichlet's boundary condition, to combine analytical and numerical solutions of the thermal conduction problems.

4.3 Thermal Resistance and Heat Capacitance of the Add-on Elements

The use of a thermal model as a multilayer 3-D structure with a flat heat source for transient thermal field calculation in microelectronic systems structures (MCM's or hybrid packages) with add-on elements leads to big errors, as there is no accounting for initial heat accumulation. It is known, that for heating of the die from temperature T_1 to T_2 the quantity of heat energy needed equals to [9]:

$$\mathrm{\Delta Q_e} = \mathrm{c_{ve} V_e \rho_e (T_2 - T_1)}, \tag{4.18}$$

where c_{ve} is the heat capasitance constant; V_e and ρ_e is the volume and density of add-on element.

When turning on for the first time, the heat energy, which is released in the element spends for add-on element heating (is accumulated in it) ands partially penetrates in the substrate or heat sink.

We will obtain the formula:

$$Q = \Delta Q_e + S_e \cdot \alpha_e \cdot \Delta T_e \cdot t \tag{4.19}$$

Here S_k is the square of heat flow into the (construction), α_e is the heat transfer coefficient from the element into the construction.

If we assume $T_1 - T_2 = \Delta T$, then after the differention we will obtain the formula for defining the temperature change velocity in the element relative to the substrate temperature:

$$c_{ve} \cdot V_e \cdot \rho_e \cdot \frac{d(\Delta T_e)}{dt} = P_0 - S_e \cdot \alpha_e \cdot \Delta T_e, \tag{4.20}$$

where P_0 is the heat dissipation power in the element. Heat dissipation energy, which is conducted from the add-on element to the substrate (multilayer structure) heats it and is partially conducted to the heat sink or the substrate (Figure 5).

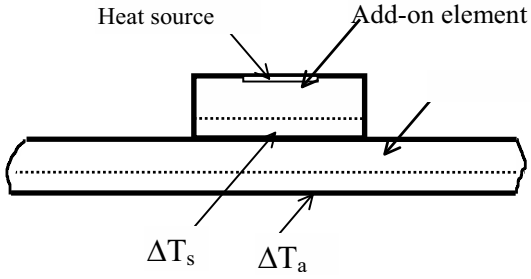


Figure 5. Cross-section of the Add-on Element with the Substrate.

Thus, the heat energy balance can be expressed as the system of equations:

$$\left.\begin{aligned} P_0 \cdot dt &= c_{ve} V_e \rho_e \cdot d(\Delta T_e) + S_e \alpha_e \cdot \Delta T_e \cdot dt, \\ S_e \alpha_e \cdot \Delta T_e \cdot dT &= c_{vs} V_s \rho_s d(\Delta T_s) + S_s \alpha_s \cdot \Delta T_s \cdot dt \end{aligned}\right\} \tag{4.21}$$

If we assume

$$C_e = c_{ve}V_e\rho_e, C_s = c_{vs}V_s\rho_s,$$
$$R_e = \frac{1}{S_e\alpha_e}, R_s = \frac{1}{S_s\alpha_s},$$

then we will obtain the system of equations

$$\left.\begin{aligned} \tau_e \frac{d(\Delta T_e)}{dt} + \Delta T_e &= P_0 R_e, \\ \tau_s \frac{d(\Delta T_s)}{dt} + \Delta T_s &= \frac{R_s}{R_e} R_, \end{aligned}\right\} \tag{4.22}$$

where $\tau_e = C_e R_e, \tau_e = C_s R_s$.

To solve the system of differential equations (4.22) means to solve the second non-homogenous differential equation taking into account the solution of the first one. The solution of the first equation will be:

$$\Delta T_e(t) = P_0 R_e\left[1 - \exp(-\frac{t}{\tau_e})\right] \tag{4.23}$$

Then we can expess the second equation from the system as:

$$\frac{d(\Delta T_s)}{dt} + \frac{1}{\tau_s}\Delta T_s = \frac{R_s P_0}{\tau_s}(1 - e^{-\frac{t}{\tau_e}}) \tag{4.24}$$

If we assume

$$a_s = \tfrac{1}{\tau_s}, F(t) = \tfrac{R_s \cdot P_0}{\tau_s}(1 - e^{-\frac{t}{\tau_e}}),$$

the solution will be

$$\Delta T_s(t) = e^{-a_s \cdot t}(c_1 + \int F(t) \cdot e^{a_s t} dt\ ) \tag{4.25}$$

Having accomplished corresponding transformations we will finally obtain the formula:

$$\Delta T_s(t) = P_0 R_s\left[1 - e^{-a_s t} - \frac{\tau_e}{\tau_e - \tau_s} e^{-a_s t}(e^{\frac{\tau_e - \tau_s}{\tau_s \tau_e}} - 1)\right], \tag{4.26}$$

which defines the temperature change in time on the substrate under the add-on element.

Transient temperature in the add-on element is defined from the formula:

$$T_e(t) = T_a + \Delta T_e + \Delta T_s,$$

or

$$T_e(t) = P_{-}\left\{R_s\left[1 - e^{-a_s t} - \frac{\tau_e}{\tau_e - \tau_s} e^{-a_s t} \times \left(e^{\frac{\tau_e - \tau_s}{\tau_s \tau_e}} - 1\right)\right] + R_e(1 - e^{-a_e t})\right\} + T_a \quad (4.27)$$

Based on the obtained formula (4.27) one can calculate the temperature difference only in the heat release area. For us to be able to calculate influenced temperature difference in another element on the substrate we shall assume that the heat power from the add-on element is transferred only through the substrate. Then the heat release power *P(t),* which is transferred from the element to the substrate is approximated with the expression:

$$P(t) = \frac{T_s(t)}{R_s} = P_0\left[1 - e^{-\frac{t}{\tau_s}} - \frac{\tau_s}{\tau_e - \tau_s} \times e^{-\frac{t}{\tau_s}}\left(e^{-\frac{\tau_e - \tau_s}{\tau_e \tau_s}} - 1\right)\right] \quad (4.28)$$

Having defined the change of heat release power in time in the way described above and exploiting the methodology described in part 4.2, one can define the temperature in any point of multilayer structure.

4.4 Software Implementation

The above solution techniques of the heat conduction problem were implemented in the steady-state and transient thermal field simulation system MONSTR (MOdelling of Non steady-state and Steady-state Thermal Regimes). The system includes a database, means of interactive communication with the user during the preparation of the input data, software modules for carrying out computations, for visualization and the analysis of the achieved results. A database is designed and realized in a specially developed database manager, which makes it possible not only to describe and handle the data, but also to develop information interfaces providing interaction of the subsystem with other CAD tools. Visualization of the computational results is realized by displaying one-dimensional plots and 3-D surfaces in color. To increase the information contained in the surfaces being displayed the object rotation and re-coloring have been implemented.

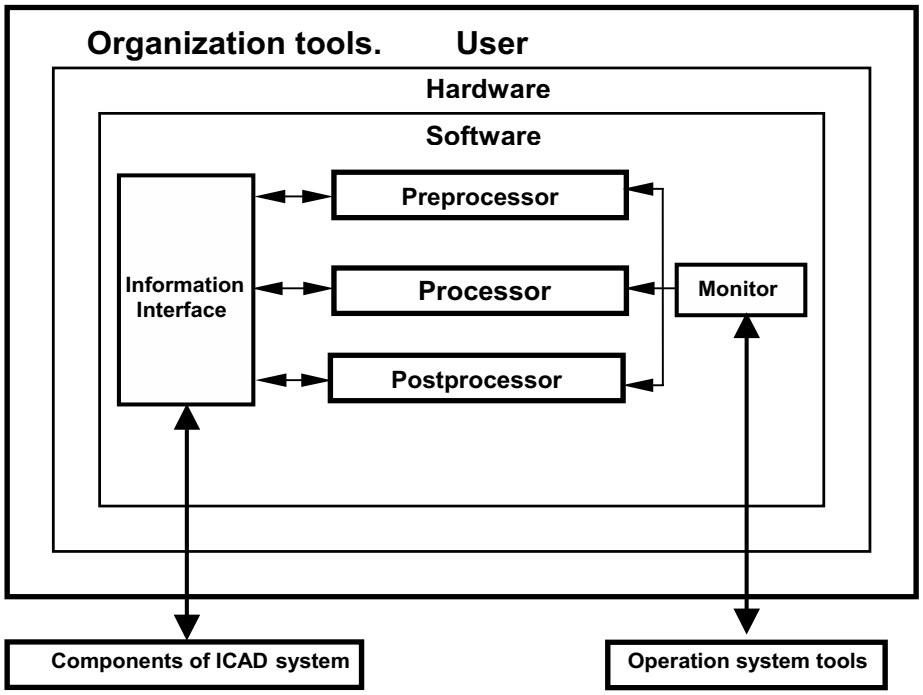


Figure 6. The Structure of Thermal Simulation System MONSTR.

One of the main requirements for developing the thermal simulation system is the possibility to adapt it to the real process of MES's designing, in which a lot of another tools, of course, are used. Thus, the system includes three main components:

1. Software and methodology tools, which include the following - preprocessor, processor, postprocessor, information interface and display driver.
2. Hardware, which includes standard configuration of tools.
3. Organization tools.

The main components of the system MONSTR are shown in Figure 6.

The adaptation to the conditions of real designing provides for following tools:

- tools of information interaction with circuits, physical and technological CAD system;
- the extending tools of interaction with users;
- providing the operational possibility to change structure input and output data in database;
- the information and linguistic tools developed for more convenient data processing by pre- and postprocessor.

The above mentioned properties define the structure and components of preprocessor, processor, and postprocessor.

Preprocessor consists of the following components:

- control module is responsible for input and processing of user's commands and executes a role of "main menu" in preprocessor;
- problem-oriented editor of input data allows to input, modify, remove and add the information which describes objects of designing and electro-thermal problems;
- the blocks of semantic and syntax's control;
- the block of formation and transformation of the input information in a form necessary for the processor;
- the information and program interfaces to the circuits, physical and technological CAD systems;
- modules of access to database.

While using a preprocessor it is possible to formulate the set of the separate problems for system's processor, which can be used by several computers simultaneously, that allows the essential decrease of simulation time. Thus using of a preprocessor in the electro-thermal design gives the designer of ES's a possibility to raise an effectiveness of using the system MONSTR in the environment of an integrated CAD system.

Processor of the system contains next components:

- the control module responsible for interaction with the user;
- the block of choosing the name of a project already formed;
- the block of control and correction of the information model parameters according to infological model of ES, to realize semantic control by automatic and interactive regime and changing input data automatically;
- the processing block and block of processing control;
- the block of computing results checking.

The running processor allows to repeat the process of simulation of the same problems several times changing and correcting only the parameters of an information model. Thus the processor is responsible for simulation and investigation of thermal and mathematical models.

Postprocessor has to realize next functions:

- to visualize and analyze obtained results chosen from the catalogue on a text and graphical display;
- to process extra simulations;
- to form files of the simulation results for circuit, physical and technological CAD systems;
- interactions with the user while visualizing and analyzing the results obtained;
- to form final results of electro-thermal simulation.

The postprocessor consists of: the 1-D, 2-D, 3-D functions graphic interpreter; the forming tables, printer and plotter drivers module; the arithmetic functions interpreter; the system of database management. The

postprocessor is running independently from the processor that allows to use it in parallel with simulation. And it has a possibility to 1-D functions in any scale and 3-D color picture of temperature field distribution within the substrate. For displaying a text the special module is developed.

Let's consider the components, which join the preprocessor, processor and postprocessor in the integrated system: monitor and information interface. Monitor realizes the interaction between the users and subsystem's components and joins the autonomous modules in accordance with the logical structure. It consists of the command interpreter and tools of calling the objective modules; as extra functions the monitor realizes the calling of operation system commands and extra modules, which are not included in the system menu. The information interaction is realized by interface, which connects the local database and external files. The database management system executes the next functions: creating a database; operating with subcatalogues; operating with files; interconnecting with outside system for files and data exchange. The database has relational structure at a conceptual level. The MONSTR system utilizing methods described above has been applied to performance simulation of semiconductor IC's [6], hybrid packages [3] and multichip modules [2].

4.5 Simulation Tests

The transient temperature with and without taking into account thermal resistance and heat capacitance of the add-on element are investigated. The input data and test structure are shown in Figure 7.

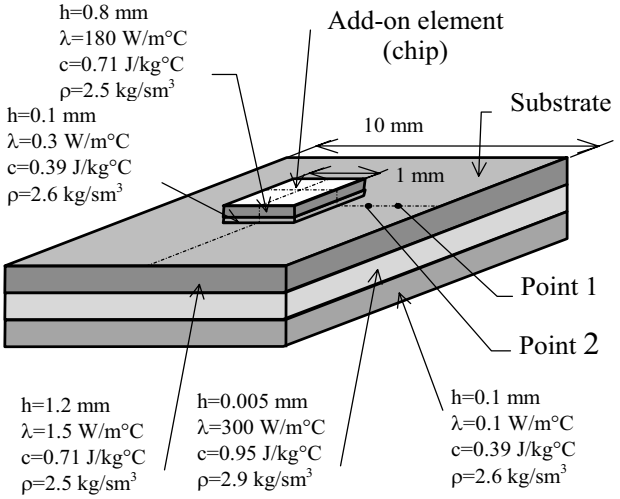


Figure 7. Thermal Test Structure with Add-on Element.

The effects of the thermal resistance and heat capacitance of the add-on element on the transient temperature of the substrate are show in Figure 8. The curves 2 and 4 have been obtained with taking into account thermal resistance and heat capacitance of the add-on element. The changing of the transient temperature in point 1 is shown by curves 1 and 2.

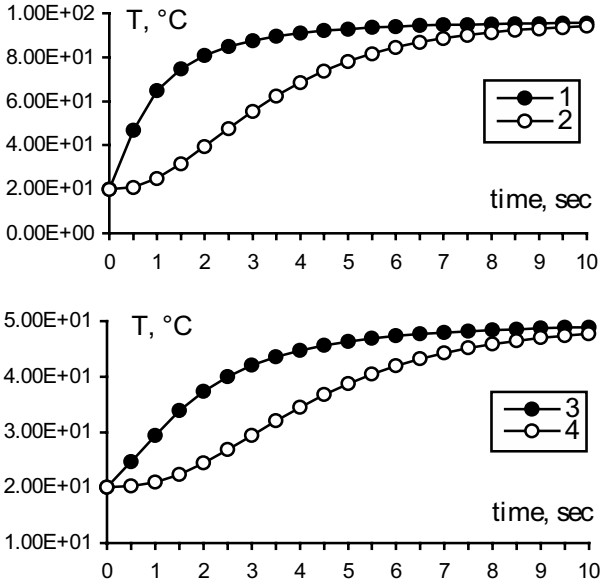


Figure 8. Time Historyof Temperature on the Substrate.

The next consideration is given to the comparison of simulation results obtained for a real hybrid circuit using three different kinds of thermal simulators. The first simulator is above mentioned system MONSTR. The second one is NAPOM-PYRTHERM which uses a semi-analytical approach employing Bessel functions. The third numerical simulator STEA is based on the principle of analogy between thermal and electrical phenomena [6, 7]. The NAPOM-PYRTHERM program was developed in the Department of Microelectronics & Computer Science at the Technical University of Lodz. It should be mentioned that a simulator similar to the MONSTR was also developed independently at the TU of Lodz.

The analysed hybrid circuit contains six thin-film resistors placed on alumina substrate attached to a copper heat sink by a thin joint layer. The geometrical data on resistors as well as the amount of power dissipated in each of them are given in Table 1. The circuit layout and its cross-sectional view together with the data assumed for the thermal model used in the simulations are presented in Figure 9. The heat exchange coefficient α at

the bottom surface was experimentally set to 30 000 W/(m^2K), which corresponds to forced liquid cooling.

Table 1. Resistor Parameters.

Resistor	Power, *W*	Dimensions, *mm*		Centre co-ordinates, *mm*	
		x	*y*	*x*	*y*
R1	*2,50*	*3,6*	*0,8*	*19,9*	*1,1*
R2	*4,12*	*1,7*	*2,0*	*20,85*	*3,5*
R3	*2,44*	*1,4*	*0,6*	*7,9*	*1,0*
R4	*2,38*	*0,8*	*1,4*	*5,3*	*6,0*
R5	*2,76*	*3,6*	*0,8*	*4,8*	*1,7*
R6	*4,28*	*1,7*	*2,0*	*3,25*	*5,1*

For the load conditions specified in Table 1, the temperature distribution on the top surface was captured by the Hughes TVS 4100 infrared camera. The measurement accuracy is ±3 K. Then, the circuit was simulated with all three earlier described simulators. The measured and simulated temperature values in the middle of each resistor are presented in Table 2.

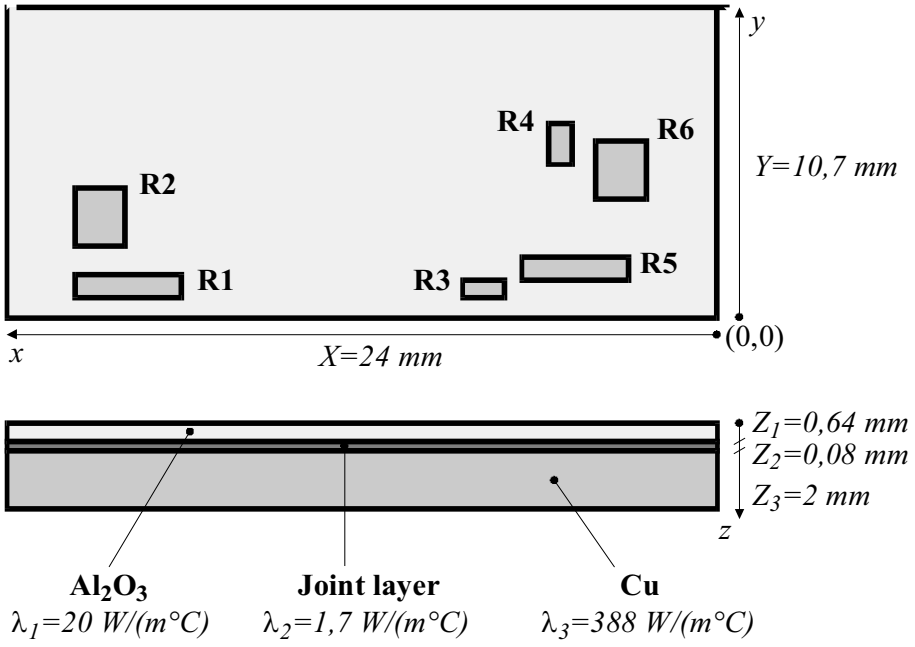


Figure 9. The Test Hybrid Circuit.

Table 2. Simulation Results.

Resistor	Temperature rise in the heater center, *C*				
	Measurement	MONSTR 30 components	MONSTR 500 components	NAPOM/ PYRTTHERM	STEA
R1	*37,0*	*36,9*	*37,3*	*38,0*	*40,6*
R2	*57,0*	*59,0*	*57,7*	*60,0*	*58,1*
R3	*81,5*	*78,2*	*73,4*	*74,0*	*83,2*
R4	*64,0*	*58,0*	*63,9*	*65,0*	*68,8*
R5	*40,0*	*37,7*	*38,0*	*40,0*	*41,3*
R6	*64,0*	*63,3*	*60,3*	*65,0*	*63,7*
Maximal deviation from measurement		*10 %*	*10 %*	*10 %*	*10 %*
Average deviation from measurement		*4 %*	*4 %*	*4 %*	*4 %*

The temperature distribution maps on circuit surface obtained from measurement and simulations are given in Figures 10-12 respectively. The reference ambient temperature was assumed equal to 0 C, thus the computed values represent temperature rise over ambient temperature. In order to investigate the influence of Fourier series truncation, the simulations with MONSTR program were performed for different number of series components.

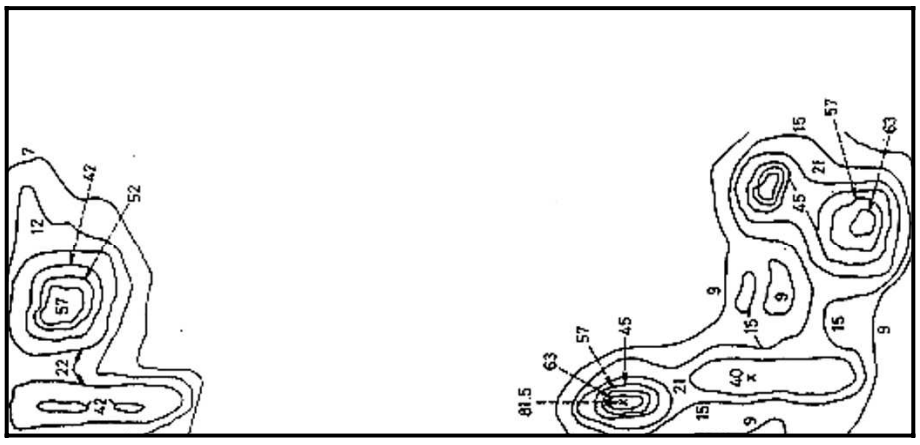


Figure 10. Temperature Map of the Upper Surface - Thermographical Measurement.

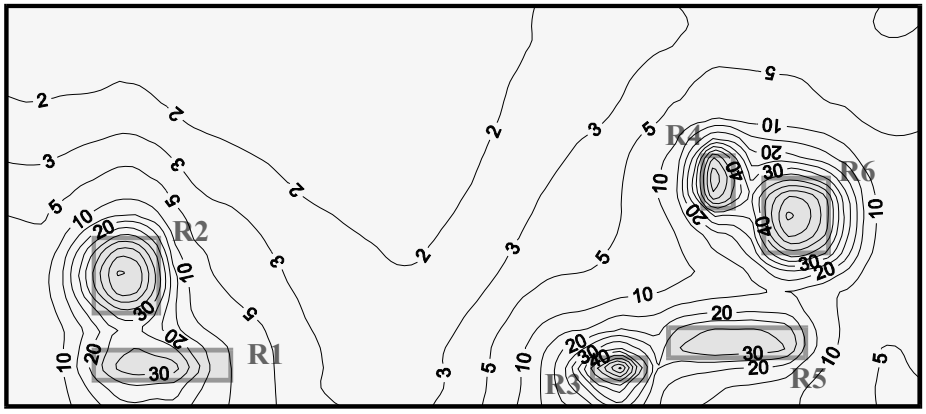


Figure 11. Temperature Map of the Upper Surface – STEA Program.

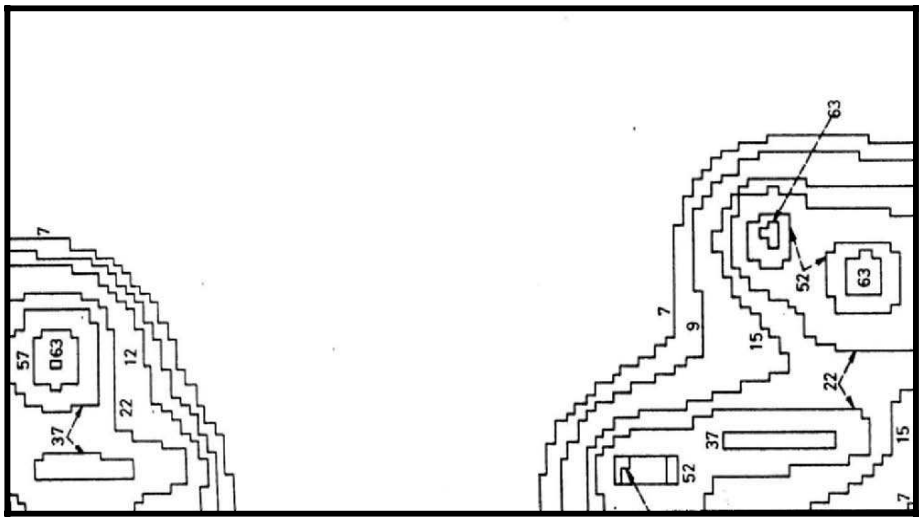


Figure 12. Temperature Map of the Upper Surface – NAPOM Program.

As one can see all simulators give similar results, which are in good accordance with the thermographical measurements. The simulation results are within the measurement error margin, except for the resistor R3, however this phenomenon the authors explain by the possible presence of an air cushion in the solder layer under this resistor. For the hot spots, the maximal error does not exceed 10% for all four simulators and the average error is less than 4%.

The maximal number of Fourier series elements taken into account was experimentally set to 500, so that further increase of the number of components does not change the numerical result at 8 byte word length.

However, it should be underlined that similar accuracy can be attained taking into account only 30 series components. For the STEA program a non-uniform division mesh comprised 5500 nodes in the final model.

5. THERMAL ANALYSIS AND MODELLING OF FLIP-CHIP STRUCTURES

In recent years the flip-chip bonding technique has become widely used in the microstructures. The method of flip-chip bonding is used for automating the labour-consuming processes because of the improvements in integrated circuits and semiconductor chips bonding technology. One of the key issues in the flip-chip structure design is to provide heat removal and temperature field analysis in three-dimensional structures.

Figure 13 shows the schematic flip-chip structure. The flip-chip basically consists of the substrate and chip attached to the substrate by solder bumps [8]. Heat sources are located on the lower surface of the chip. In practice there are various types of solder bumps with different functions: for electrical coupling, for mechanical stability, for heat removal, etc [9, 10].

There are the methods for thermal fields modeling in MCM with flip-chip bonding described in literature, which utilize the finite elements approach. Other papers solve this problem using partially isothropical areas with the components: die, outlets and substrate. Still, different from the above mentioned researches, the new method is developed for die thermal field simulation in the complex structure, which allows for obtaining analytically and calculating the thermal field both in the flip-chip die and in the substrate.

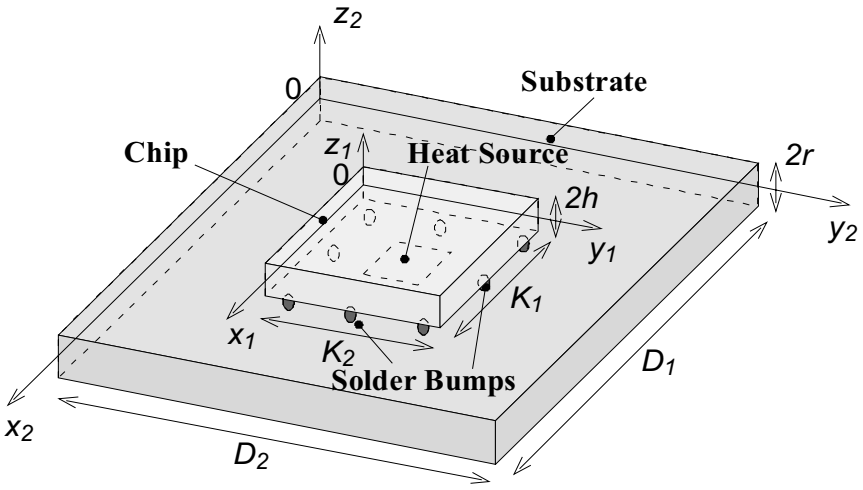


Figure 13. Schematic Flip-chip Structure.

5.1 Steady-state Thermal Problem

Let us consider a multi-layer die (N layers) of certain size (K_1, K_2, h_k) with heat sources (their number is ks), attached to multi-layer (M layers) substrate (P_1, P_2, h_p) by solder bumps [2, 3]. Each die layer with height $h_1^{(i)}$ is made of a material with thermal conductivity $\lambda_1^{(i)}$ ($i = \overline{1, N}$); accordingly, for each substrate layer we have $h_2^{(j)}$ and $\lambda_2^{(i)}$ ($j = \overline{1, M}$). The solder bumps are: signal (their number is kp_1) and constructive (their number is kp_2). The interaction of the passive side of the die with the ambient at temperature T_{amb} is specified by the heat exchange coefficient α. Similarly, for the lower and upper surfaces of the substrate we specify heat exchange coefficients with the ambient α_2 and β_2.

The solution of the problem of steady-state heat exchange in flip-chip bonding is sought in the die domain D_1 (co-ordinate system X_1, Y_1, Z_1) and substrate domain D_2 (X_2, Y_2 Z_2). The employed specifications and different types of bumps are shown on Figure 14.

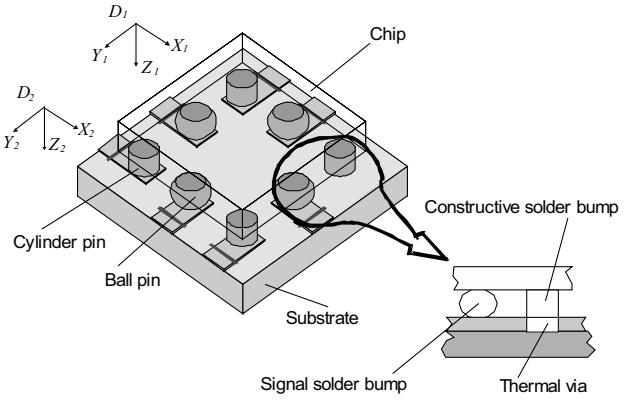


Figure 14. Different Types of Solder Bumps.

The die and the substrate are attached by solder bumps; the area of cross-section of i-th lead is Sp_i, its heat source power is Pp_i. The heat sources are placed on the insulated lower die surface; the cross-section area of j-th heat source is Ss_j, its power is Ps_j.

The heat exchange problem for these two conjugate multi-layer structures for each domain (D_1 and D_2) is formulated by the system of differential equations, the solutions of which are the temperature in i-th die layer $T_1^{(i)}(x_1, y_1, z_1)$ and the temperature in j-th substrate layer $T_2^{(i)}(x_2, y_2, z_2)$.

For $D_1 = \bigcup_{i=1}^{N} D_1^{(i)}$:

$$\frac{\partial^2 T_1^{(i)}}{\partial x_1^2}+\frac{\partial^2 T_1^{(i)}}{\partial y_1^2}+\frac{\partial^2 T_1^{(i)}}{\partial z_1^2}=0, \tag{5.1}$$

$$\frac{\partial T_1^{(i)}}{\partial x_1}\Big|_{\substack{x_1=0\\ x_1=K_1}}=\frac{\partial T_1^{(i)}}{\partial y_1}\Big|_{\substack{y_1=0\\ y_1=K_2}}=0 \tag{5.2}$$

$$-\lambda_1^{(1)}\frac{\partial T_1^{(1)}}{\partial z_1}\Big|_{z=0}+\alpha_{T1}\left(T_1^{(1)}-T_{amb}\right)\Big|_{z=0}=0 \tag{5.3}$$

$$T_1^{(i)}(x_1,y_1,z_1)\big|_{z_1=z_{1i}}=T_1^{(i+1)}(x_1,y_1,z_1)\big|_{z_1=z_{1i}}$$

$$\lambda_1^{(i)}\frac{\partial T_1^{(i)}}{\partial z_1}\Big|_{z_1=z_{1i}}=\lambda_1^{(i+1)}\frac{\partial T_1^{(i+1)}}{\partial z_1}\Big|_{z_1=z_{1i}},\ i=\overline{1,N-1} \tag{5.4}$$

$$\lambda_1^{(N)}\frac{\partial T_1^{(N)}}{\partial z_1}\Big|_{z_1=z_{1N}}=\begin{cases}\dfrac{P_{s_m}}{S_{s_m}},\text{if}\,(x_1,y_1)\in S_{s_m},\ m=\overline{1,ks}\\ -\dfrac{P_{p1_k}}{S_{p1_k}},\text{if}\,(x_1,y_1)\in S_{p1_k},\ k=\overline{1,kp1}\\ -\dfrac{P_{p2_l}}{S_{p2_l}},\text{if}\,(x_1,y)_1\in S_{p2_l},\ l=\overline{1,kp2}\\ 0,\text{if}\,(x_1,y_1)\notin S_{s_j}\wedge(x_1,y_1)\notin S_{p1_k}\wedge(x_1,y_1)\notin S_{p2_l}\end{cases} \tag{5.5}$$

For $D_2=\bigcup_{j=1}^{M}D_2^{(j)}$:

$$\frac{\partial^2 T_2^{(j)}}{\partial x_2^2}+\frac{\partial^2 T_2^{(j)}}{\partial y_2^2}+\frac{\partial^2 T_2^{(j)}}{\partial z_2^2}=0 \tag{5.6}$$

$$\frac{\partial T_2^{(j)}}{\partial x_2}\Big|_{\substack{x_2=0\\ x_2=D_1}}=\frac{\partial T_2^{(j)}}{\partial y_2}\Big|_{\substack{y_2=0\\ y_2=D_2}}=0 \tag{5.7}$$

$$\left(-\lambda_2^{(1)}\frac{\partial T_2^{(1)}}{\partial z_2}+\alpha_{T2}(T_2^{(1)}-T_{ambl})\right)\Big|_{z_2=0}=\begin{cases}\dfrac{P_{p1_i}}{S_{p1_i}},\text{if}\,(x_2,y_2)\in S_{p1_i}\\ 0,\text{if}\,(x_2,y_2)\notin S_{p1_j}\end{cases} \tag{5.8}$$

$$\lambda_2^{(l)}\frac{\partial T_2^{(l)}(x_2,y_2,z_2)}{\partial z_2}\Big|_{z_2=z_{21}}=\lambda_2^{(2)}\frac{\partial T_2^{(2)}(x_2,y_2,z_2)}{\partial z_2}\Big|_{z_2=z_{21}}+\begin{cases}\dfrac{P_{p2i}}{S_{p2i}},\text{if}\,(x_2,y_2)\in S_{p2i}\\ 0,\text{if}\,(x_2,y_2)\notin S_{p2i}\end{cases} \tag{5.9}$$

$$\lambda_2^{(j)} \left.\frac{\partial T_2^{(j)}(x_2, y_2, z_2)}{\partial z_2}\right|_{z_2=z_{2j}} = \lambda_2^{(j+1)} \left.\frac{\partial T_2^{(j+1)}(x_2, y_2, z_2)}{\partial z_2}\right|_{z_2=z_{2j}}, \tag{5.10}$$

$$T_2^{(j)}(x_2, y_2, z_2)\big|_{z_2=z_{2j}} = T_2^{(j+1)}(x_2, y_2, z_2)\big|_{z_2=z_{2j}},\ j=\overline{2,M-1}$$

$$\left.\left(\lambda_2^{(M)} \frac{\partial T_2^{(M)}}{\partial z_2} + \beta_{T2}\left(T_2^{(M)} - T_{amb\,2}\right)\right)\right|_{z_2=z_{2M}} = 0 \tag{5.11}$$

The solutions for die and substrate domains are found by the Fourier method in the form of trigonometric series:

$$T_1(x_1, y_1, z_1) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} Z_{1nm}(z_1)\cos\frac{n\pi x_1}{K_1}\cos\frac{m\pi y_1}{K_2} \tag{5.12}$$

$$T_2(x_2, y_2, z_2) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} Z_{2kl}(z_2)\cos\frac{k\pi x_2}{D_1}\cos\frac{l\pi y_2}{D_2} \tag{5.13}$$

The unknown coefficients of Fourier series $Z_{1nm}(z_1)$ and $Z_{2kl}(z_2)$ are found analytically with the use of boundary conditions and conjugation conditions. As the result, we obtain for each n, m a system of linear algebraic equations that is solved by the method of exclusion of unknowns.

For determining the unknown powers of heat flow via the leads of two types (P_{lead1i} and P_{lead2j}), we use the temperature equality conditions on the junction of the die and the substrate, as two conjugate domains

$$\iint_{S_1} T_1^{(N)}(x_1, y_1, z_1^{(N)})dx_1 dy_1 = \iint_{S_2} T_2^{(1)}(x_2, y_2, 0)dx_2 dy_2 \tag{5.14}$$

$$\iint_{S_1} T_1^{(N)}(x_1, y_1, z_1^{(N)})dx_1 dy_1 = \iint_{S_2} T_2^{(1)}(x_2, y_2, z_2^{(1)})dx_2 dy_2 \tag{5.15}$$

where S_1, S_2 are the areas of junction of the die and the substrate in co-ordinate systems X_1, Y_1, Z_1 and X_2, Y_2, Z_2, accordingly.

For all leads of the first type we employ the condition (5.14), for all leads of the second type - the condition (5.15). After necessary substitutions and simplifications we get a system of linear heterogeneous equations for

determination of heat flow in all leads, which contain *kp* equations and *kp* unknown variables. The solution is found by a numerical solving method [4].

The obtained solutions allow to calculate the temperature field in an arbitrary structure point, dependent on the influence of heat sources, their placing, and constructive and topological characteristics of solder bumps, as well as thermal-physical parameters of the die and the substrate.

5.2 Transient Thermal Problem

Let us consider the die of certain size K_1, K_2, $2h$ with heat sources and solder bumps. Mathematically the transient conductivity process in the die can be described by the Fourier differential equation:

$$\frac{\partial^2 T}{\partial x_1^2} + \frac{\partial^2 T}{\partial y_1^2} + \frac{\partial^2 T}{\partial z_1^2} - \frac{c_1 \rho_1}{\lambda_1} \frac{\partial T}{\partial t} = 0, \tag{5.16}$$

where $T=T(x_1,y_1,z_1,t)$ is the temperature in the die, λ_1 is the thermal conductivity, c_1 is the specific heat capacity, ρ_1 is the specific die density.

Write the boundary condition on the bottom surface taking into account the heat source placement (P^i_{source}, S^i_{source} are the heat dissipation power and the area of i-th heat source) and solder bump location (P^j_{pin}, S^j_{pin} are the heat removal power and the area of j-th pin):

$$\frac{\partial T}{\partial z_1}_{|z_1=-h} = \begin{cases} \dfrac{P_{pin}{}^j(t)}{\lambda_1 S_{pin}{}^j}, \text{if } (x_1, y_1) \in S_{pin}{}^j, \\ -\dfrac{P_{source}{}^i(t)}{\lambda_1 S_{source}{}^i}, \text{if } (x_1, y_1) \in S_{source}{}^i, \\ 0, \qquad \text{else} \end{cases} \tag{5.17}$$

The boundary condition for the top surface can be described as:

$$\frac{\partial T}{\partial z_1}_{|z_1=h} = -\frac{\alpha^+}{\lambda_1}(T^+ - \Theta^+), \tag{5.18}$$

where $^+$ and T^- are $T(x_1, y_1, z_1, t)$ function values on lower and top borders accordingly; α^- and α^+ are the heat transfer coefficients with environment,

temperatures of which are Θ^{+} and Θ^{--} accordingly. The thermal removal into the ambient from the side surface of the die can be neglected due to small dimension. Thus boundary condition for the side surfaces of the die can be described as:

$$\frac{\partial \mathrm{T}}{\partial \mathrm{x}_1}\Big|_{\substack{\mathrm{x}=0\\ \mathrm{x}=\mathrm{K}_1}} = \frac{\partial \mathrm{T}}{\partial \mathrm{y}_1}\Big|_{\substack{\mathrm{y}=0\\ \mathrm{y}=\mathrm{K}_2}} = 0 \tag{5.19}$$

The initial condition has the following form

$$\mathrm{T}(\mathrm{x}_1, \mathrm{y}_1, \mathrm{z}_1, 0) = \mathrm{T}_0 \tag{5.20}$$

Thus a temperature in any die point can be described by the differential heat conductivity equation (5.16) with boundary conditions (5.17)-(5.19) and the initial condition (5.20).

Similarly, we describe the heat exchange problem for the substrate with dimensions $D_1 \times D_2$ and thickness $2r$

$$\frac{\partial^2 \mathrm{W}}{\partial \mathrm{x}_2^{\,2}} + \frac{\partial^2 \mathrm{W}}{\partial \mathrm{y}_2^{\,2}} + \frac{\partial^2 \mathrm{W}}{\partial \mathrm{z}_2^{\,2}} - \frac{\mathrm{c}_2 \rho_2}{\lambda_2}\frac{\partial \mathrm{W}}{\partial \mathrm{t}} = 0, \tag{5.21}$$

where $W(x_2, y_2, z_2, t)$ is the temperature in the $(x_2, y_2, z_2,)$ substrate's point at the time t; λ_2 is the thermal conductivity, c_2 is the specific heat capacity, ρ_2 is the specific density of the substrate.

$$\frac{\partial \mathrm{W}}{\partial \mathrm{z}_2}\Big|_{\mathrm{z}_2=-\mathrm{r}} = \frac{\beta^{-}}{\lambda_2}(\mathrm{W}^{-} - \Omega^{-}), \tag{5.22}$$

$$\frac{\partial \mathrm{W}}{\partial \mathrm{z}_2}\Big|_{\mathrm{z}_2=\mathrm{r}} = \begin{cases} -\dfrac{\mathrm{P_{pin}}^{\mathrm{j}}(\mathrm{t})}{\lambda_2 \mathrm{S_p}^{\mathrm{j}}}, \text{if } (\mathrm{x}_2, \mathrm{y}_2) \in \mathrm{S_{pin}}^{\mathrm{j}} \\ -\frac{\beta^{+}}{\lambda_2}(\mathrm{W}^{+} - \Omega^{+}), \text{else} \end{cases}, \tag{5.23}$$

here β^{-} and β^{+} are heat transfer coefficients where the temperatures of environment are Ω^{-} and Ω^{+} accordingly.

$$\frac{\partial \mathrm{W}}{\partial \mathrm{x}_2}\Big|_{\substack{\mathrm{x}_2=0\\ \mathrm{x}_2=\mathrm{D}_1}} = \frac{\partial \mathrm{W}}{\partial \mathrm{y}_2}\Big|_{\substack{\mathrm{y}_2=0\\ \mathrm{y}_2=\mathrm{D}_2}} = 0, \tag{5.24}$$

$$W(x_2, y_2, z_2, 0) = W_0 \tag{5.25}$$

The conjugated heat exchange problem for the die and the substrate conjugate is solved analytically.

The essence of the method proposed is the presentation of search function of four variables in the linear expansion form along z coordinate and finding two functions with three variables. The method consists of following steps:

- transforming the differential equation of second order to the system of two differential equations with the exclusion of *z* variable and satisfying the boundary conditions on *z* axis;
- obtaining the solution in the form of two Fourier series in terms of eigenfunctions as the result of separation of *x* and *y* variables;
- solving the system of differential equations of the first order for finding the Fourier coefficients of two functions of linear decomposition with the application of Laplaces's transformation;
- obtaining heat removal powers through the solder bumps taking into account the condition of conjugate heat exchange in the die and the substrate.

The last step requires forming and solving the system of linear integral equations with the dimension equal to the number of solder bumps. Here we apply one of the methods for approximate solution of integral equations, the method of successive approximations. Three first scheme steps are applied to the substrate in a similar way.

Because of the complexity of all analytical expressions, we have presented only some die solution schemes.

Thus we are searching for the approximative heat exchange solution problem by Kovalenko [10]. The temperature distribution along *z* coordinate can be represented by polynomial

$$T(x_1, y_1, z_1, t) = \sum_{j=0}^{m} T_{j+1}(x_1, y_1, t) \cdot z_1^{\,j}, \tag{5.26}$$

that reduces thermal problem to two dimensional one.

To format the equation that must satisfy the function $T_j\ (x_1,\ y_1,\ t)$, we multiply the initial equation on z_1^p $(p=0,1,\ldots,\ m)$ and integrate the result by coordinate $z_1$on interval $[-h_1,\ h_1]$.

At this we take into account the next identify

$$z^p \frac{\partial^2 T}{\partial z_1^{\,2}} = \frac{\partial}{\partial z_1}\left(z_1^{\,p} \frac{\partial T}{\partial z_1} - p z_1^{\,p-1} T \right) + p(p-1) z_1^{\,p-2} \tag{5.27}$$

and boundary conditions (5.17), (5.18).

Let us assume the linear decomposition of the temperature along the thickness of the die (using m=1) and then we shall have

$$T(x_1, y_1, z_1, t) = T_1(x_1, y_1, t) + z_1 \cdot T_2(x_1, y_1, t) \qquad (5.28)$$

We integrate the initial equation by coordinate z_1 on interval $[-h_1, h_1]$.

$$\int_{-h}^{h}\left(\frac{\partial^2 T}{\partial x_1^{\,2}} + \frac{\partial^2 T}{\partial y_1^{\,2}} + \frac{\partial}{\partial z_1}\left(\frac{\partial T}{\partial z_1}\right) - \frac{c_1\rho_1}{\lambda_1}\frac{\partial T}{\partial t}\right)dz_1 = 0 \qquad (5.29)$$

If designated $\Delta_{xy} = \frac{\partial^2}{\partial x_1^{\,2}} + \frac{\partial^2}{\partial y_1^{\,2}}$, we obtain the expression

$$\Delta_{xy}\int_{-h}^{h} T dz_1 + \frac{\partial T}{\partial z_1}\Bigg|_{-h}^{h} - \frac{c_1\rho_1}{\lambda_1}\frac{\partial}{\partial t}\int_{-h}^{h} T dz_1 = 0 \qquad (5.30)$$

Using dependence (5.28) for expression (4.30) we receive

$$\Delta_{xy}\left(T_1 \cdot z_1 + T_2 \cdot \frac{z_1^{\,2}}{2}\right)\Bigg|_{-h}^{h} + Q^{+} - Q^{-} - \frac{c_1\rho_1}{\lambda_1}\frac{\partial}{\partial t}\left(T_1 \cdot z_1 + T_2 \cdot \frac{z_1^{\,2}}{2}\right)\Bigg|_{-h}^{h} = 0 \qquad (5.31)$$

that in the final form is written as follows

$$\Delta_{xy} T_1 - \frac{a^{+}}{h^2} T_1 - \frac{a^{+}}{h} T_2 + \frac{a^{+}}{h^2}\Theta^{+} - \frac{c_1\rho_1}{\lambda_1}\frac{\partial T_1}{\partial t} = \frac{Q^{-}}{2h}, \qquad (5.32)$$

where $a^{+} = \frac{\alpha^{+} h}{2\lambda_1}$.

To create the second equation we multiply the initial equation on z1 and integrate the result by coordinate z1 on interval $[-h_1, h_1]$.

$$\int_{-h}^{h}\left(\left(\frac{\partial^2 T}{\partial x_1^{\,2}} + \frac{\partial^2 T}{\partial y_1^{\,2}}\right)z_1 + \frac{\partial}{\partial z_1}\left(z_1\frac{\partial T}{\partial z_1} - T\right) - \frac{c_1\rho_1}{\lambda_1} z_1 \frac{\partial T}{\partial t}\right)dz_1 = 0, \qquad (5.33)$$

after due transformations and taking into account (5.28) we may write

$$\Delta_{xy}\left(T_1\cdot\frac{z_1^{\,2}}{2}+T_2\cdot\frac{z_1^{\,3}}{3}\right)\Bigg|_{-h}^{h}+\left(hQ^{+}-T_1-hT_2-(hQ^{-}-T_1+hT_2)\right)-$$
$$-\frac{c_1\rho_1}{\lambda_1}\frac{\partial}{\partial t}\left(T_1\cdot\frac{z_1^{\,2}}{2}+T_2\cdot\frac{z_1^{\,3}}{3}\right)\Bigg|_{-h}^{h}=0 \tag{5.34}$$

In the final form we obtain

$$\Delta_{xy}T_2-\frac{3a_1^{\,+}}{h^2}T_2-\frac{3a^{+}}{h^3}T_1-\frac{c_1\rho_1}{\lambda_1}\frac{\partial T_2}{\partial t}=-\frac{3Q^{-}}{2h^2}-\frac{3a^{+}}{h^3}\Theta^{+}, \tag{5.35}$$

where $a_1^{\,+}=a^{+}+1$

Performing analogous transformations with boundary conditions (5.19), we receive boundary conditions for functions $T_1(x_1, y_1, t)$ and $T_2(x_1, y_1, t)$:

$$\left.\frac{\partial T_1}{\partial x_1}\right|_{\substack{x_1=0,\\ x_1=K_1}}=\left.\frac{\partial T_1}{\partial y_1}\right|_{\substack{y_1=0,\\ y_1=K_2}}=0, \tag{5.36}$$

$$\left.\frac{\partial T_2}{\partial x_1}\right|_{\substack{x_1=0,\\ x_1=K_1}}=\left.\frac{\partial T_2}{\partial y_1}\right|_{\substack{y_1=0,\\ y_1=K_2}}=0 \tag{5.37}$$

Thus for the determining the unknown functions $T_1(x_1, y_1, t)$ and $T_2(x_1, y_1, t)$: we receive the system of differential equations (5.32) and (5.35) with boundary conditions (5.36), (5.37).

We seek unknown functions in the form of trigonometric series and perform their decomposition by eigenfunctions of variable separation on x_1 and y_1 coordinates.

$$T_1(x_1,y_1,t)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}T_{1nm}(t)\cos\frac{n\pi}{K_1}x_1\cos\frac{m\pi}{K_2}y_1 \tag{5.38}$$

$$T_2(x_1,y_1,t)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}T_{2nm}(t)\cos\frac{n\pi}{K_1}x_1\cos\frac{m\pi}{K_2}y_1 \tag{5.39}$$

In a similar way we decompose the Q^{-} function, which still contains the unknown heat removal powers through the solder bumps. Then, we designate the coefficient of the Fourier decomposition by qsnm and take into account the decomposition of the lower surface of the die into N cells. Each cell corresponds to either the solder bump or the heat source or a free area.

Now we appoint the function $Q^s(t)$ to the *s*-th cell (*s=1,...,N*) and obtain the formula:

$$Q^{-}(x_1, y_1, t) = \sum_{s=1}^{N} Q^s(t) \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} q_{nm}^s \cos\frac{n\pi}{K_1} x_1 \cos\frac{m\pi}{K_2} y_1 \quad (4.40)$$

Thus, we have obtained the system of differential equations on variable t with the boundary conditions for the determination of Fourier series coefficients and the quantity of the sought-for functions $T_1(x_1, y_1, t)$ and $T_2(x_1, y_1, t)$.

$$\begin{cases} A_{nm} T_{1nm} + \dfrac{a+}{h} T_{2nm} + \dfrac{c_1\rho_1}{\lambda_1} \dfrac{dT_{1nm}}{dt} = C_{nm}(t); \\ \dfrac{3a+}{h^3} T_{1nm} + B_{nm} T_{2nm} + \dfrac{c_1\rho_1}{\lambda_1} \dfrac{dT_{2nm}}{dt} = D_{nm}(t); \end{cases} \quad (5.41)$$

$$T_{1nm}(0) = \begin{cases} T_0, \text{if } n = m = 0, \\ 0, \text{if } n, m \neq 0. \end{cases} \quad (5.42)$$

$$T_{2nm}(0) = 0,$$

here we used the following designations:

$$A_{nm} = \left(\frac{\pi n}{K_1}\right)^2 + \left(\frac{\pi m}{K_2}\right)^2 + \frac{a^+}{h^2}, \; B_{nm} = \left(\frac{\pi n}{K_1}\right)^2 + \left(\frac{\pi m}{K_2}\right)^2 + \frac{3a_1^+}{h^2},$$

$$C_{nm} = \frac{a^+}{h^2} \Theta^+_{nm} - \frac{1}{2h} \sum_{s=1}^{N} q_{nm}^s Q^s(t),$$

$$D_{nm} = \frac{3a^+}{h^3} \Theta^+_{nm} + \frac{3}{2h^2} \sum_{s=1}^{N} q_{nm}^s Q^s(t)$$

Now we solve system (5.41) with the condition (5.42) using the Laplace's transformation

$$T_{1nm}(t) = \frac{\lambda_1}{4c_1\rho_1(\gamma_{nm} - \omega_{nm})h} \sum_{s=1}^{N} q_{nm}^s \int_0^t Q^s(\tau) \left[\frac{\lambda_1}{c_1\rho_1} (B_{nm} + \frac{3a^+}{h^2}) g_{nm}(\tau - t) - f_{nm}(\tau - t) \right] d\tau \quad (5.43)$$

$$T_{2nm}(t)=\frac{3\lambda_1}{4c_1\rho_1(\gamma_{nm}-\omega_{nm})h^2}\sum_{s=1}^{N}q_{nm}^{s}\int_0^t Q^s(\tau)\left[f_{nm}(\tau-t)-\frac{\lambda_1}{c_1\rho_1}(A_{nm}+\frac{a^+}{h^2})g_{nm}(\tau-t)\right]d\tau \quad (5.44)$$

where

$$f_{nm}(\tau-t)=\gamma_{nm}e^{\gamma_{nm}(\tau-t)}-\omega_{nm}e^{\omega_{nm}(\tau-t)};$$

$$\mathrm{g_{nm}}(\tau-t)=e^{\gamma_{nm}(\tau-t)}-e^{\omega_{nm}(\tau-t)}$$

$$\gamma_{nm}=\frac{\lambda_1}{c_1\rho_1}\left(\frac{1}{h^2}\sqrt{\left(3a^+ +4(a^+)^2+\frac{9}{4}\right)+\left(\frac{\pi n}{K_1}\right)^2+\left(\frac{\pi m}{K_2}\right)^2}+\frac{2a^+}{h^2}+\frac{3}{2h^2}\right)$$

$$\omega_{nm}=\frac{\lambda_1}{c_1\rho_1}\left(\frac{1}{h^2}\sqrt{\left(3a^+ +(4a^+)^2+\frac{9}{4}\right)-\left(\frac{\pi n}{K_1}\right)^2-\left(\frac{\pi m}{K_2}\right)^2}-\frac{2a^+}{h^2}-\frac{3}{2h^2}\right) \quad (5.45)$$

For determination of the unknowns $P^1_{pin}(t),...P^k_{pin}(t)$ for each j-th solder bump we apply the conjugate condition (5.46) that conjugates the die with the substrate via the solder bumps. Here $S1^j_{pin}$ and $S2^j_{pin}$ are the areas of the solder bumps junction with the die and substrate in accordance with coordinate transformation for each pin in the substrate domain: $x_{2pin_j}=A-\frac{K_1}{2}+x_{1pin_j}$, $y_{2pin_j}=B-\frac{K_2}{2}+y_{1pin_j}$, where A and B are coordinates of the chip center in the substrate domain; h_{pin}, λ_{pin} are the height and the thermal conductivity of the solder bumps, respectively. $W(x_2, y_2, t)$ is the substrate temperature, the coefficients of which are found similar to the die.

$$\iint_{\mathrm{S1_{pin}^{\,j}}}\mathrm{T(x_1,y_1,-h,t)dx_1dy_1}=\iint_{\mathrm{S2_{pin}^{\,j}}}\mathrm{W(x_2,y_2,r,t)dx_2dy_2}+\frac{\mathrm{h_{pin}}}{\lambda_{\mathrm{pin}}}\mathrm{P_{pin}^{j}(t)} \quad (5.46)$$

Then we obtain the system of Walter's integral first form equations:

$$\begin{cases} \sigma_{11}\int\limits_0^t P_{pin}^1(\tau)\varphi_{11}(t,\tau)d\tau + \sigma_{12}\int\limits_0^t P_{pin}^2(\tau)\varphi_{12}(t,\tau)d\tau + ... + \\ + \sigma_{1k}\int\limits_0^t P_{pin}^k(\tau)\varphi_{1k}(t,\tau)d\tau = \delta_1(t) \\ \sigma_{21}\int\limits_0^t P_{pin}^1(\tau)\varphi_{21}(t,\tau)d\tau + \sigma_{12}\int\limits_0^t P_{pin}^2(\tau)\varphi_{22}(t,\tau)d\tau + ... + \\ + \sigma_{2k}\int\limits_0^t P_{pin}^k(\tau)\varphi_{2k}(t,\tau)d\tau = \delta_2(t) \\ \qquad ... \qquad\qquad ... \qquad\qquad ... \\ \sigma_{k1}\int\limits_0^t P_{pin}^1(\tau)\varphi_{k1}(t,\tau)d\tau + \sigma_{12}\int\limits_0^t P_{pin}^2(\tau)\varphi_{k2}(t,\tau)d\tau + ... + \\ + \sigma_{kk}\int\limits_0^t P_{pin}^k(\tau)\varphi_{kk}(t,\tau)d\tau = \delta_k(t) \end{cases} \tag{5.47}$$

where $P^1_{pin}(t),...P^k_{pin}(t)$ are unknown heat removal flows through solder bumps, $\sigma_{ij}$is some coefficients, $\delta_i(t)$, $\varphi_{ij}(t,\tau)$ are known functions determined from the condition (5.36), $i, j = \overline{1,k}$. Applying the method of successive approximations to the (5.37), we will receive the values of the sought-for bump power functions on the interval pieces and put them into formulas (5.34), (5.35). Thus, we have found the coefficients of the function decomposition, which linearly represent a solution to the heat exchange problem.

5.3 Software Implementation

The above solution techniques of the thermal problems were implemented in the Web-based system called iTSim. iTSim is a novel network CAD tool that is intended for the thermal design of microelectronic components of various embodiments and production techniques. The system is built on the basis of CORBA technology that is an industrial standard for distributed applications.

iTSim system is based on the following conceptual principles:

1. Simulation of microelectronic systems of various embodiments and production techniques.
2. Implementation of different thermal modeling methods.
3. Distributed architecture with the centralized information fund.
4. Integration with the Internet.

5. Facilities for the integration with external CAD/CAM systems.
6. Implementation of the functions proper to existing thermal simulators.
7. Support of different hardware platforms and operating systems.
8. Usage of the modern computer technologies and development tools.

The system architecture developed in accordance with the key principles described above is presented in Figure 15. The architecture combines a few well-known architectural styles such as preprocessor-processor-postprocessor architecture, client-server architecture, object-oriented architecture, and modular architecture.

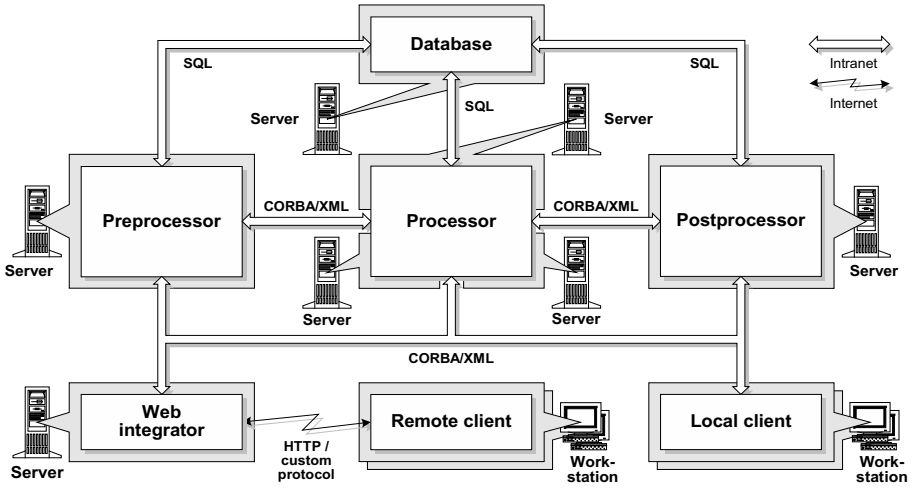


Figure 15. Architecture of iTSim system.

The iTSim system consists of seven components with strongly defined functional purposes: preprocessor, processor, postprocessor, database, local client, web integrator, and remote client. Basically, each system component should be installed at separate or several networked computers.

Interaction among the preprocessor, the processor, the postprocessor, the local client and the web integrator is performed by means of CORBA interfaces implemented in the preprocessor, the processor and the postprocessor (Figure 16). Each CORBA interface offers the access to a certain predefined functionality of the system component and provides a set of methods to reach this functionality. Except internal interaction of the system components, these CORBA interfaces can be also used for the integration with external CAD/CAM systems that support CORBA mechanisms or by means of corresponding bridges.

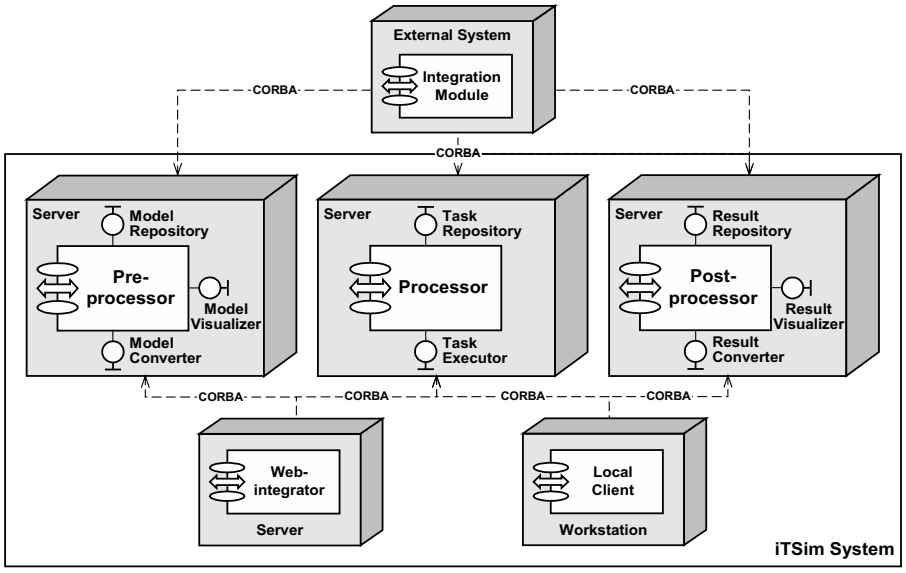


Figure 16. System Components Interaction Via CORBA Interfaces.

During storing in the database and passing across the system components all project related data (thermal models, simulation tasks, and simulation results) are presented in flexible and extendable XML format. Such choice allows to pass project data among different platforms and operating systems. Interaction between the database and the simulation engine (the preprocessor, the processor and the postprocessor) is performed via standard SQL queries. In the database all project data are stored in several relational tables with implicit associations on the level of XML content. The remote client interacts with the web integrator via Internet in accordance with HTTP protocol. If needed it is possible to introduce an additional custom protocol based on TCP/IP commands that provides the remote client with advanced functionality.

The local client provides a user with the access to the system simulation engine from Win32 workstations connected to local network. The local client is implemented in C++ with the usage of OpenGL graphic libraries.

The local client operates with the thermal models of microelectronic components of various embodiments and production techniques, and offers the following functional facilities:

1. Visual building of the thermal models of microelectronic components in 3-D representation (Figure 3).
2. Definition of the thermal simulation tasks to perform steady-state temperature analysis, non steady-state temperature analysis, and constructive-parametrical search.

3. Visualization of the thermal simulation results in the form of color temperature maps laid on the 3D skeletons of the thermal models (Figure 17).

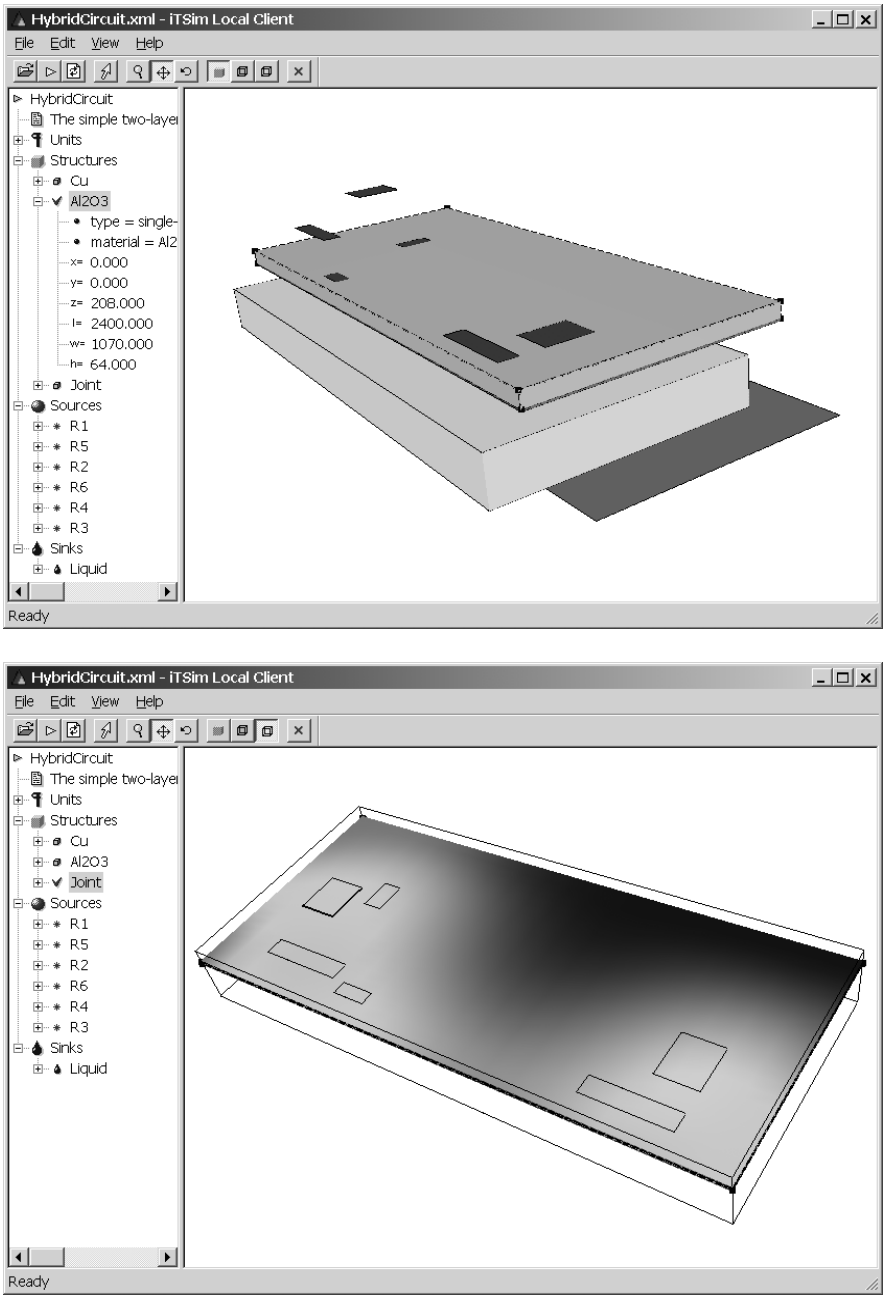


Figure 17. iTSim User Interfaces for Thermal Design Within Intranet.

During thermal modeling the local client interacts with the preprocessor, the processor and the postprocessor via CORBA services.In real working configurations the iTSim system can have any number of local clients installed on separate workstations that are connected to local network. In this case all thermal designers work in common information space and use common system facilities to perform design tasks.

Remote access to the iTSim system is provided by means of the web integrator built on the basis of a JSP-server, and the remote client. The remote client works under usual web browser like Internet Explorer or Netscape Navigator and connects to the JSP server in order to obtain a sequence of the web pages with dynamic project related information and forms for data input. Like the local client the web integrator communicates with the iTSim simulation engine via CORBA mechanisms.

Existing implementation of the web integrator allows to calculate steady-state temperature fields into multi-layer hybrid circuits with flat heat sources on the surface. For the lower surface of the modeling object it is possible to assign one of two boundary conditions: convective heat exchange or uniform temperature distribution with predefined temperature. The upper surface of the modeling object can have only convective heat exchange with the external environment. Simulation process for a remote user consists of seven sequential steps from definition of the hybrid circuit configuration to visualization of the temperature map on the upper circuit surface (Figure 18).

It is worth to point that the modeling object and the simulation tasks accepted for the remote thermal analysis can be changed or extended by means of updating the web-integrator and the remote client only, without any influence on the other iTSim components.

5.4 Simulation Tests

The flip-chip structure with D_1=8 mm; D_2=16 mm; r=0.5 mm dimensions substrate made of ceramics (λ=155W/°Cm, c=1004J/kg°C, ρ=3500kg/m^3) and K_1=3.6 mm; K_2=4 mm; h=0.3 mm dimensions chip made of silicon (λ=125W/°Cm, c=703J/kg°C, ρ=1330kg/m^3) was used in the model experiment. The chip is established to the substrate by means of eight 0.2 mm diameter solder bumps made of a solder alloy (λ=50 W/°Cm, c=376J/kg°C, ρ=8850kg/m^3) and placed along the chip perimeter on the 0.1mm distance from the chip edge. The flat heat source with 1.8x2 mm dimensions is placed in the center of the lower chip surface. The lower surface of the substrate is isothermal with the temperature of 20°C. Other structure surfaces have free convection exchange with the environment, the coefficient of this exchange being α=15W/(°Cm^2). The ambient temperature

is 20°C and the heat source power equals to 1 W. The thermal test structure is shown in Figure 19.

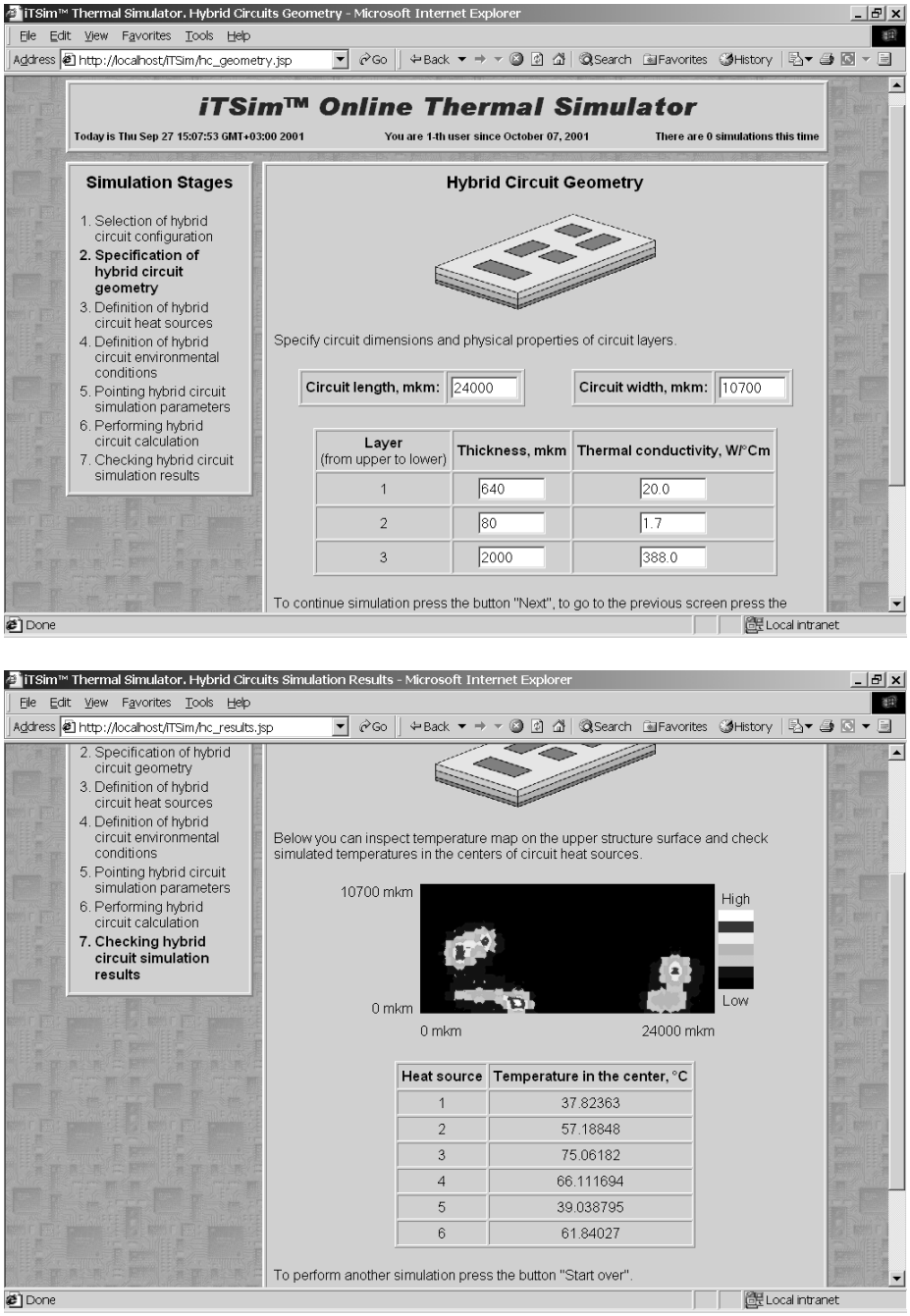


Figure 18. iTSim User Interfaces for Remote Thermal Simulation Via Internet.

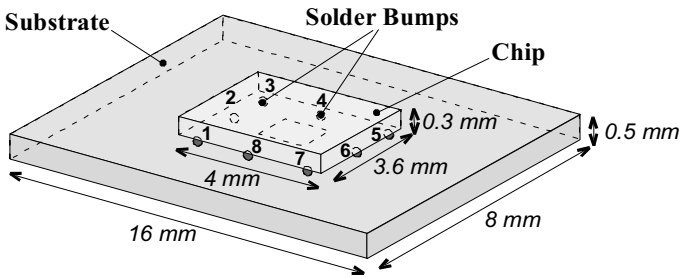


Figure 19. Thermal Test Structure.

To approve accuracy of the method described in this paper the test structure was modeled by two other systems: STEA simulator and WebTAFC simulator. Both simulators were developed by the authors of this article. The STEA simulator is based on the thermal and electrical analogy and intended for thermal simulation of wide range of microelectronic systems including flip-chip structures [15]. The WebTAFC simulator is oriented to Web-based steady state thermal analysis of flip-chip structures by the analytical method [12]. A reader can simulate the test structure in the WebTAFC simulator and compare the results with ones presented below by visiting the WebTAFC Internet server:
http://www.neo.dmcs.p.lodz.pl/WebTAFC.html.

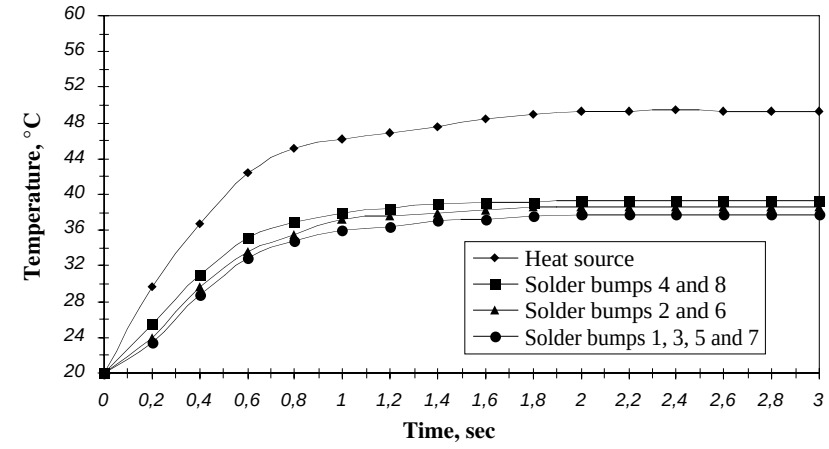
Table 3 contains the temperatures in the heat source and on the solder bumps of the test structure that were obtained by the original method described in the paper (denoted as "the method") and two simulators: STEA and WebTAFC.

Table 3. The Results of Thermal Simulation of the Test Structure.

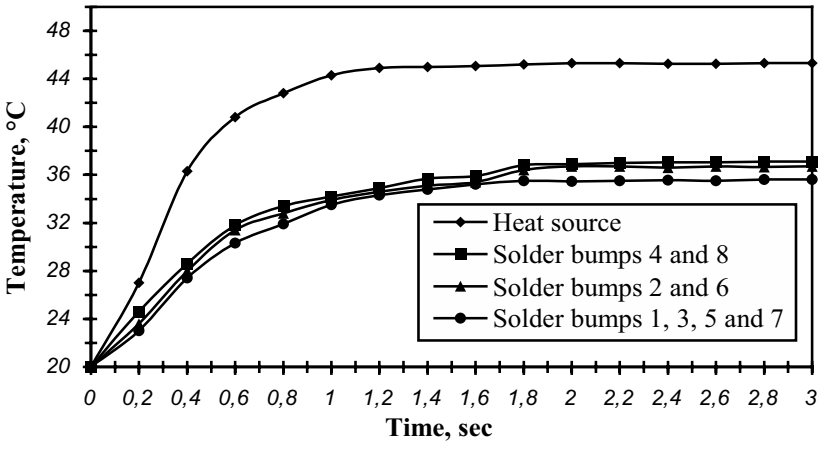
	Time, sec										
	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
Heat Source											
Method	20.0	29.7	36.8	42.5	45.1	46.3	46.9	47.6	48.5	49.0	49.2
STEA	20.0	27.0	36.3	40.8	42.8	44.3	44.9	45.0	45.1	45.2	45.3
WebTAFC	48.5 (in steady-state regime)										
Solder Bumps 4 and 8											
Method	20.0	25.5	31.0	35.1	36.9	38.0	38.4	39.0	39.2	39.2	39.3
STEA	20.0	24.6	28.6	31.8	33.4	34.2	34.9	35.7	35.9	36.8	36.9
WebTAFC	38.8 (in steady-state regime)										
Solder Bumps 2 and 6											
Method	20.0	23.9	29.6	33.6	35.6	37.2	37.6	38.0	38.3	38.6	38.7
STEA	20.0	23.6	28.0	31.4	32.8	33.9	34.6	35.1	35.4	36.4	36.7
WebTAFC	37.9 (in steady-state regime)										
Solder Bumps 1, 3, 5 and 7											
Method	20.0	23.4	28.8	33.0	34.9	36.1	36.4	37.0	37.2	37.6	37.7
STEA	20.0	23.0	27.4	30.3	31.9	33.5	34.3	34.8	35.2	35.5	35.5
WebTAFC	36.2 (in steady-state regime)										

One can see that the transient process simulated by the method presented above is similar to the one modeled by STEA simulator. However there is an essential divergence in numeric results obtained by these two methods. The deviation of the temperatures doesn't exceed 10-15% in any time point for both transient and steady state temperatures. The difference between the results obtained by the original method and WebTAFC simulator is much smaller and takes 1-2% for steady-state temperatures.

Figure 20 depicts the transient temperatures obtained for the test structure by application of the method described above (a) and STEA simulator (b). The figures show the temperatures in the heat source and on the solder bumps.



a)



b)

Figure 20. Transient temperatures obtained by the original method (a) and STEA simulator (b).

Both pictures show the similarity of the transient processes simulated by two different methods. One can notice a slight delay in pins heating simulated by STEA simulator of the beginning of the transient process. This delay is due to heat expansion across the chip. The results obtained by the original method don't reflect such delay expressively. Both figures show that the transient thermal process takes approximately 1.5 sec; after this time the temperature regime of the test structure can be considered as a steady state one.

6. CONCLUSIONS

High power densities and significant resulting temperature values encountered in modern microelectronic systems induced still increasing need for CAD tools, which would allow the consideration of thermal phenomena in the process of electronic system design. In order to compute the temperature distribution in a structure, the heat equation must be solved. There are two main groups of methods for solving the equation: numerical and analytical ones. Most commercial thermal simulators are based on numerical methods. Although these methods render possible analysis of complex shapes, solutions depend strongly on the choice of structure meshing. Another disadvantage of numerical methods is that, in order to obtain acceptable accuracy, it is necessary to use dense discretization mesh, which is computationally inefficient and time consuming. Therefore, since microelectronic circuits have usually simple geometry, it is worth considering the possibility to model such structures analytically.

This paper present modeles, methods and software for thermal simulation of two typical constructions of MES, namely, hybrid circuits with add-on elements and flip-chip structures. Thermal models of those microelectronics systems in the form of multilayer structures with heat sources and analytical approach for investigation of the models have been developed. The main results and conclusions can be summarised as follows.

1. The problem of MES's thermal field analysis has been formalized. Our method of the initial and boundary problem formalization makes it possible to take into account various boundary conditions and apply this analysis in system design. The combination of numerical and analytical methods has been developed for the solution of 3-D parabolic equations. This combination allows for both the steady-state and non-steady-state solutions as well as approximations in boundary conditions along the Z coordinate of the nonhomogeneous functions and their derivatives along the X and Y coordinates.

2. The mathematical models of the steady-state and transient heat exchange processes in the flip-chip structures are described. The models are based on the analytical solution of the conjugate transient heat exchange

problem for the die and substrate. The essence of the new method is in the presentation of the sought function of four variables in the form of linear expansion along the z coordinate and finding two functions with three variables. The accuracy of the model is validated by comparison with the calculation results of the transient temperatures in flip-chip structure by the thermal-electrical analogy method. The calculation error obtained doesn't exceed 10-15% in any time point. Also the same test structure is modelled by the analytical method in steady-state regime. The error is up to 2%. The mathematical model developed is to be used for investigation and analysis of the transient thermal fields in the flip-chip structures in microelectronics designing.

3. The system "MONSTR" for MES's thermal design has been developed and applied to several practical applications; the system allows for the reduction in the development time and for the improvement in the quality of ES's.The practical value of the electro-thermal simulation methodology presented above lies in the following possibilities: to determine real temperature values at any point of the 3-D structure MES's; to test MES's simulation for conditions of external and internal thermal fields influence and thermal shocks (which replaces the expensive physical experiments); to perform the optimization of parameters by taking into account local temperature values; to calculate temperature tolerances of element parameters; to perform statistical analysis of thermal fields with structural and electro-thermal parameter variations.

4. The iTSim network CAD tool is developed and is intended for the distributed thermal design of microelectronic components of different embodiments and production techniques that are described in the paper. iTSim solution is built on the basis of CORBA technology and provides thermal designers with the access to the simulation engine via Internet. iTSim thermal design tool provides a thermal designer with a lot of facilities that absent in exiting thermal simulation software. Among these facilities everyone can emphasize the implementation of various modeling methods, distributing the calculation tasks among servers in local network, multi-user working mode, common information data fund, the access to the system via Internet, standard mechanisms for the integration with CAD/CAM systems, and others. Due these facilities of the iTSim system it is possible to increase the efficiency of designers' work, improve the quality of thermal design solutions, and reduce time to the market for new microelectronic products.

5. The accuracy of simulation is comparable with the one obtained using numerical methods but the computation time is much less because there is no need to solve a large set of difference equations. An additional advantage of the analytical approach is that the results do not depend on the choice of structure discretisation mesh. The fundamental limitation is that those analytical methods cannot be applied for structures having complex shapes or boundary conditions because then it is extremely difficult to obtain the

solution formula and the model must be simplified, which may introduce excessive simulation errors.

REFERENCES

[1]. Fedasyuk D, Petrov D, Levus E. Survey and evaluation of modern thermal simulators. In Proceedings of International Conference on Modern Problems of Telecommunications, Computer Science and Engineers Training; 2000; Slavsko, Lviv, Ukraine.

[2]. Koval V, Fedasyuk D. Multilevel Thermal Simulation of MCM's by System 'MONSTR-M'. In Proceedings of the European Design and Test Conference (ED&TC 1995); 1995, March 6-9; Paris, France.

[3]. Koval VA, Fedasyuk DV, Kazymyra IY, Blyzniuk NB. Providing of electro-thermal compatibility of hybrid microcircuits in CAD environment. Collection of papers presented at the International Workshop on Thermal Investigations of ICs and Microstructure, 1997 Cannes (France).

[4]. Koval VA., Fedasyuk DV. 3-D Transient thermal simulation of microelectronics systems. Collection of papers presented at 2nd THERMINIC Workshop, 1996, Budapest (Hungary).

[5]. Koval VA, Fedasyuk DV, Farmaga IW, Mikchalchuk MM. MONSTR: The Simulator of the Thermal Design of Electronic System. In Proceedings of the 2nd Advanced Training Course on Mixed Design of VLSI Circuits- Education of Computer Aided Design of Modern VLSI Circuit, Mix VLSI'95; 1995, 29-31 May; Krakow, Poland.

[6]. Koval V.A.,. Fedasyuk D.V Thermal Simulation of the Semiconductor IC's: General and Practical Approach. Microelectronics Journal, 1997; 28:221-227

[7]. Janicki M., Napieralski A., Fedasyuk D., Petrov D. Thermal modelling of hybrid circuits: simulation method comparison. Microelectronics Reliability Journal, 2000; 40:541-546

[8]. Koval VA., Fedasyuk DV. Thermal analysis and modeling of the flip-chip bonding. The analytical approach. In Proceedings of the 3-rd Advanced Training Course on Mixed design of integrated circuits and system - Education of computer aided design of modern IC's and devices; 1996 30 May - 1 June; Lodz, Poland.

[9]. Fedasyuk. DV. Thermal studies of the flip-chip bonding. In Proceedings of the 4-th International Workshop MIXED DESIGN OF INTEGRATIOED CIRCUITS AND SYSTEM, MIXDEZ'97; 1997 12-14 June; Poznan, Poland.

[10].Fedasyuk D, Levus E., Mykhalchuk M., Petrov D. Modelling and Analysis of Methods of Providing Thermal Performance of Flip-Chip Structure. Collection of papers presented at International 5-th THERMINIC Workshop; 1999 3-6 October; Rome, Italy.

[11].Kovalenko A.D. *Thermoelesticity*. Kiev: Vishcha Shcola, 1975 (in Russian).

[12].Fedasyuk D., Levus E., Petrov D. Flip-chip structure transient model. Microelectronics Reliability, 2001; 41:1965-1970.

[13].Fedasyuk D., Petrov D. Thermal Modelling on the Basis of Thermal-Electrical Analogy. Proceedings of International Conference on Modern Problems of Telecommunications, Computer Science and Engineers Training; 1998 23-28 February; Lviv, Ukraine.

[14].Petrov D., Fedasyuk D. WebTAFC – Web-Based Thermal Simulator for Flip-Chip Structures. In Proceedings of the 6-th International Conference CADSM-2001; 2001 12-17 February; Lviv-Slavsko, Ukraine.

[15].Fedasyuk D., Petrov D. ITSim – New Generation of the CAD Tools for the Distributied Thermal Design of Microelectronic Systems. In Proceedings of the 9th International Conference on Mixed Design of Integrated Circuits and Systems (MIXDES'2002); 2002 June 20-22; Wroclaw, Poland.

Chapter 11
CHEMICAL AND BIOLOGICAL SENSORS BASED ON MICROCANTILEVERS

P.G. Datskos [a,b], N.V. Lavrik [a,b] and M. J. Sepaniak [c]

[a] *Oak Ridge National Laboratory*
Engineering Science and Technology Division
Oak Ridge, Tennessee, U.S.A

[b] *University of Tennessee*
Department of Physics and Astronomy
Knoxville, Tennessee, U.S.A

[c] *University of Tennessee*
Department of Chemistry
Knoxville, Tennessee, U.S.A

Abstract: The recent developments in micro-mechanical or micro-electro-mechanical (MEMS) systems have enabled the use of new transduction modes that involve mechanical energy and are based primarily on mechanical phenomena. As a result, an innovative family of chemical and biological sensors has emerged. In this work we discuss the recent advances in cantilever transducers. While MEMS represents a diverse family of designs, devices with simple cantilever configurations are especially attractive as transducers for chemical and biological sensors. We discuss four important aspects of cantilever transducers, which include operation principles and models, figures of merit, and applications. We also provide a brief analysis of historical predecessors of the modern cantilever sensors.

Keywords: microcantilever, MEMS, micromechanical, resonance, deformation, transducer, stress, chemical sensor, biosensor, analyte, response mechanism, noise, thermomechanical, figures of merit.

1. INTRODUCTION

1.1 Definitions and Concepts

New concepts of chemical sensors have been a subject of extensive research efforts in recent decades. According to established definitions, a chemical sensor consists of a physical transducer (*i.e.* a transducer of physical quantities into convenient output signals) and a chemically selective layer (see Figure 1) so that measurable output signals can be produced in

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 331-379.

response to chemical stimuli [1]. Specific binding sites present in chemically selective layers provide affinity of targeted analytes to the sensor active area. Highly selective receptor layers can be designed using concepts of molecular and biomolecular recognition (Figure 1B).

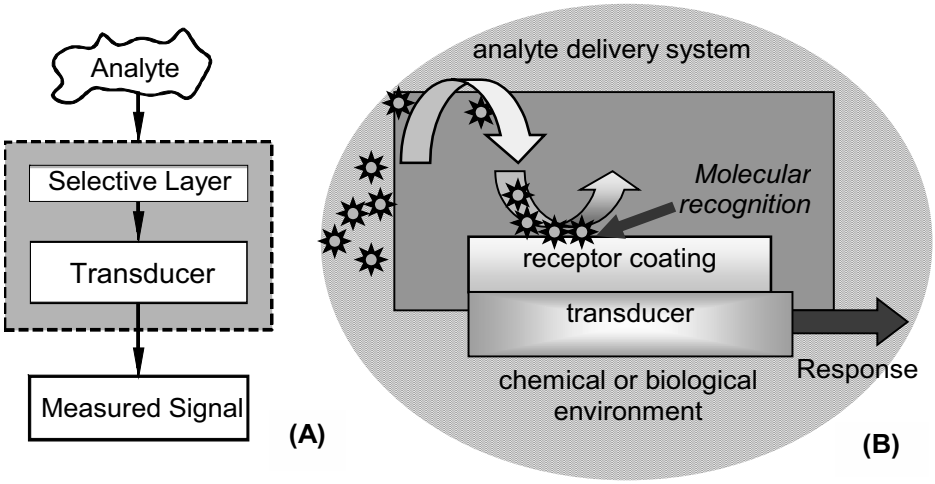


Figure 1. Generalized structure a chemical sensor. (A) Schematic representation of a chemical sensor that produces an output signal in response to the presence of a target analyte. (B) A chemical sensor with a receptor layer that provides a highly selective response in biological sensors.

The main fundamental transduction modes used in chemical sensors are categorized as [2, 3]: (a) thermal, (b) mass, (c) electrochemical, and (d) optical. Because each of these detection modes is associated with features that are complementary rather than competitive with each other, the search for an "ideal transducer" continued. During the last two decades, advances in micro-electro-mechanical systems (MEMS) have facilitated development of sensors that involve transduction of mechanical energy and rely primarily on mechanical phenomena [4-11]. Development of microfabricated cantilevers for atomic force microscopy (AFM) [12] marked an important milestone in establishing efficient technological approaches to MEMS sensors. However, the key concepts [13-18] as well as early experimental studies [19-24] related to mechanical transducers for chemical sensors can be traced back to the pre-MEMS era.

Functionality of MEMS sensors is based on mechanical movements and deformations of their micromachined components, such as single-clamped suspended beams (cantilevers), double-clamped suspended beams ("bridges") or suspended diaphragms. Since the advent of scanning probe microscopy (SPM), fabrication and characterization of microscale cantilevers useful as AFM probes [12-25] has been a subject of extensive research efforts [26-28]. As a result of more recent advances in several converging areas of science and technology, not only a variety of sophisticated probes became available for SPM but also an innovative family of physical, chemical and biological sensors based on cantilever technology was shaped out [6, 9-11, 29-33]. While MEMS transducers span a great variety of designs [4], devices with very simple cantilever-type configurations appeared to be especially suitable as transducers of physical, chemical and biological stimuli into readily measured signals [29-34]. In Figure 2 (A-C), we show examples of different cantilever devices. Figure 2A illustrates sizes and shape of typical commercially available AFM cantilevers. Designs of cantilever devices vary substantially depending on their readout and desired parameters. For example, piezoresistive cantilevers are often U–shaped (Figure 2B) in order to provide a

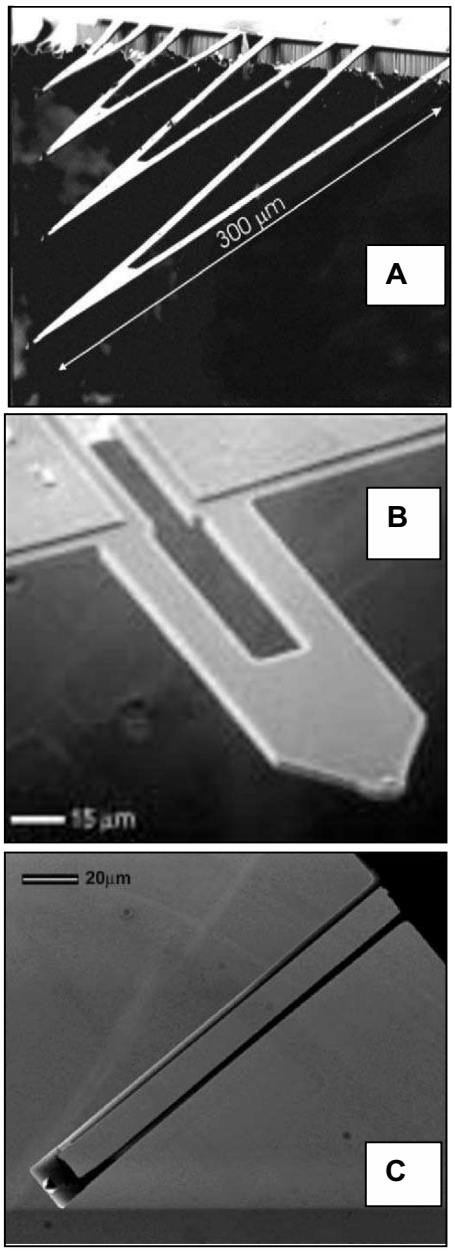


Figure 2. Examples of microcantilever devices. (A) Commercially available triangular microcantilevers used in AFM. (B) Piezoresistive single crystal Si microcantilever. (C) Cantilever with improved thermal isolation optimized for calorimetry.

proper path for electrical current. Very long and thin legs of the cantilever shown in Figure 2C improve thermal isolation of the device.

The general idea behind MEMS sensors is that physical, chemical or biological stimuli can affect mechanical characteristics of the micro-mechanical transducer in such a way that the resulting change can be measured using electronic, optical or other means [35]. In particular, microfabricated cantilevers together with read-out means that are capable of measuring 10^{-12} to 10^{-6} m displacements can operate as detectors of surface stresses [5, 36-42], extremely small mechanical forces [43-46], charges [47-51] heat fluxes [5, 52, 53], and IR photons [54-61]. As device sizes approach the nanoscale, their mechanical behavior starts resembling vibrational modes of molecules and atoms (Figure 3). Scaling of cantilevers is associated with respective changes in their mass, frequency and energy content. In the nanomechanical regime, it is possible to attain extremely high fundamental frequencies approaching those of vibrational molecular modes. Ultimately, very small nanomechanical transducers can be envisioned as human-tailored molecules that interact controllably with both their molecular environment and readout components. Nanomechanical resonators with mass of 2.34×10^{-18} g and resonance frequency of 115 MHz were fabricated and displacement of 2×10^{-15} m $Hz^{-1/2}$ were measured [62]. Mass sensitivity of only a few femtograms was reported recently using nanoscale resonators [63].

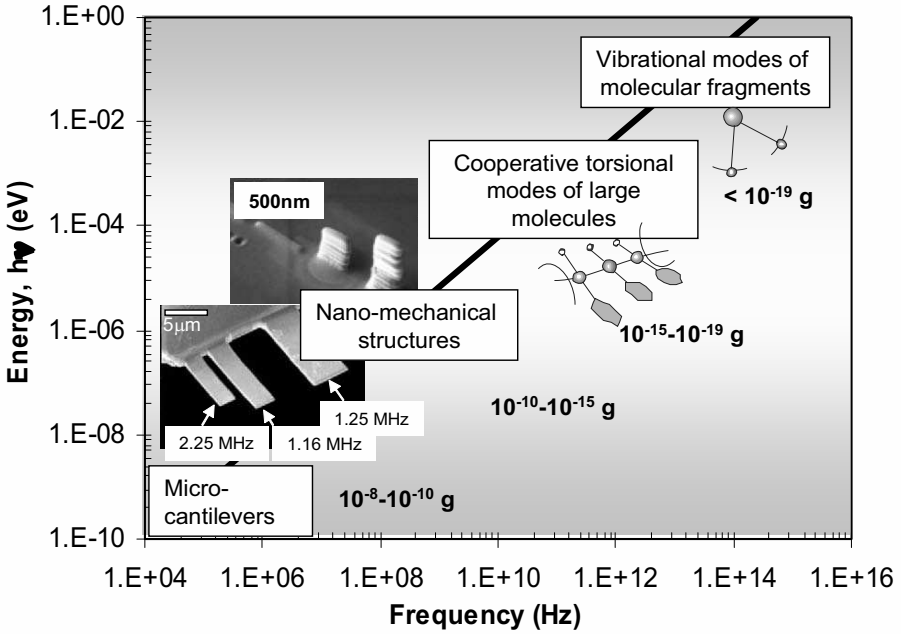


Figure 3. Spatial scaling of mechanical resonator is associated with respective scaling of masses, frequencies and energies. In the nanomechanical regime, it is possible to attain extremely high fundamental frequencies approaching those of vibrational molecular modes.

In this work we use the terms “cantilever”, “microcantilever” and “bridge” to denote devices such as single-clamped and double-clamped suspended beams of various sizes and shapes. For simplicity, we will mainly use the term “MEMS”, although derived terms, such as micro-opto-electro-mechanical systems (MOEMS), and biological micro-electro-mechanical systems (Bio-MEMS) could also be justified in this content to emphasize specific features of certain sensors based on micromechanical transducers.

1.2 Early Work

Studies of mechanical phenomena associated with changes in the chemical environment have a substantial historical background. In 1924, Palmer [19] studied coherence of loosely contacting thin filaments induced by electromagnetic waves in presence of different gases and analyzed correlation between the observed responses and the heat of gas absorption. Almost at the same time, Meehan [21] observed adsorption-induced expansion of yellow pine charcoal exposed to carbon dioxide vapors and showed that this was a reversible process. Later, Yates [64] conducted similar studies on the expansion of porous glass exposed to non-polar gases such as Ar, N_2, O_2, H_2 and Kr. One of the first descriptions of a chemical detector based on a cantilever mechanical transducer can be found in the patent issued to Norton as early as in 1943 [14]. A further refinement of this principle described by Norton was described by Shaver in 1969 [15].

Mechanical stresses and deformations produced in response to a changing chemical environment have also drawn attention as a principle of powering miniature mechanical devices. For example, in the 1960s, Steinberg *et al.* [17] developed the concept of devices that provide direct conversion of chemical stimuli into mechanical energy [18, 65]. Work on such devices referred to as *mechano-chemical engines* has not been pursued since then primarily because of the difficulty associated with microfabrication. Furthermore, practical implications of these devices were limited until advances in microtechnology and, more recently, in MEMS opened up an opportunity to fabricate miniaturized mechanical components routinely.

1.3 From Macro- to Micro-Mechanical Transducers

Long before the advent of AFM, macroscopic cantilever devices and mechanical resonators were used by many researchers as very sensitive transducers and highly precise oscillators. One of the earliest examples is an electronic circuit that provided a precision time standard by using a macroscopic tuning fork [66]. The use of macroscopic cantilevers in chemical sensors can be traced back to the 1940s [14]; Norton proposed a hydrogen detector based

on a macroscopic bimetallic plate. Norton's work was revisited almost three decades later by Shaver [15] who also used large 100 mm long, 125 μm thick bimaterial cantilevers in a hydrogen sensor. A decade later, Taylor and co-workers from Oak Ridge National Laboratory studied bending induced by molecular adsorption of He, H_2, NH_3 and H_2S on large nickel cantilevers (100 mm long) coated with 80 nm of gold [67]. The operation of sensors described by Norton and Shaver was based on very high solubility of hydrogen gas in palladium and concomitant expansion of the metal, *i.e.* the same phenomenon that was used recently in an integrated on-chip MEMS-based hydrogen sensors [68]. Cantilever deflections observed in the studies of Taylor and co-workers were associated with adsorbate-induced stresses (*i.e.* a surface rather than bulk phenomenon) and indicated another fundamental mechanism [69] that was later utilized in a variety of cantilever-based chemical and biological sensors. Macroscopic cantilever transducers were also demonstrated to be rather sensitive calorimetric devices useful, for instance, as IR detectors. In 1957 Jones used a thin, few millimeter long metallic strip in order to detect IR radiation due to thermal expansion of the strip [70].

Optical means for measuring mechanical responses of macroscopic mechanical transducers with sub-micrometer accuracy have existed since the 1920s [20]. A difficulty associated with the use of macroscale cantilever transducers for practical applications was their extremely high susceptibility to external vibrations stemming from their large suspended masses and respectively low resonance frequencies. Cantilever transducers had attained little practical appeal until both microscopic cantilevers and more precise means for their readout became widely available. Significant technological challenges of microfabricated cantilevers prevented them from becoming widespread. Except for very few studies [71, 72], the idea of microfabricated transducers based on suspended resonating or deformable structures [16, 24, 73-77] remained almost abandoned until the advent of AFM [12].

2. FUNDAMENTAL MODELS

Cantilever-based sensors involve measurements of cantilever deflections, cantilever resonance frequencies and their damping characteristics. The variety of transduction mechanisms that are involved in the functioning of cantilever sensors is depicted in Figure 4. Depending on the measured parameter - cantilever deflection or resonance frequency - the mode of cantilever operation can be referred to as either static or dynamic. Each of these modes, in turn, can be associated with different transduction scenarios (Figure 4). Static cantilever deflections can be caused by either external forces exerted on the cantilever (as in AFM) or intrinsic stresses generated on the cantilever surface or within the cantilever. While cantilever microfabrication technol-

ogy is capable of producing nearly stress-free suspended beams, additional intrinsic stresses may subsequently originate from thermal expansion, interfacial processes and physicochemical changes. Cantilever sensors operating in the dynamic mode are essentially mechanical oscillators, resonance characteristics of which depend upon the attached mass as well as visco-elastic properties of the medium. For instance, adsorption of analyte molecules on a resonating cantilever results in lowering of its resonance frequency due to the increased suspended mass of the resonator.

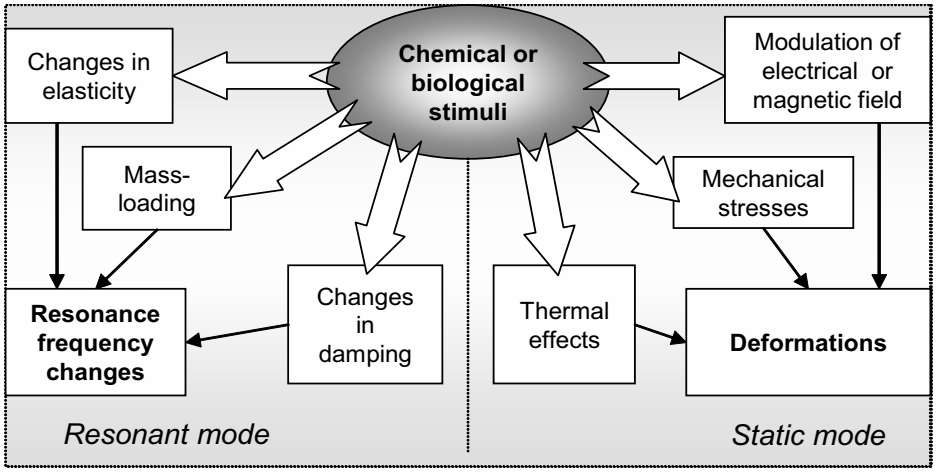


Figure 4. Conversion of input stimuli into output signals by cantilever transducers is associated with a number of transduction mechanisms. Depending on the measured parameter - structural deformations or resonance frequency changes - the mode of sensor operation can be refereed to as either *static* or *resonant*. Each of these modes, in turn, can be associated with different transduction scenarios.

Depending on the nature of the input stimuli, microcantilever sensors can be referred to as *physical, chemical or biological* sensors. The variety of transduction modes (Figure 4) stems from the fact that a stimulus of each type may affect the mechanical state of the transducer directly or may undergo one or several transformations before the measured mechanical parameter of the transducer is affected. Cantilever transducers can detect chemical and biochemical species more directly due to adsorption-induced stresses [36, 38, 42, 78-81] or mass loading effects [63, 82, 83].

Modern MEMS sensors have much in common with their predecessors, such as resonant gate transistors [16, 24, 73-75, 84], acousto-mechanical resonance sensors [85-90] as well as macroscopic cantilever devices [13, 15, 67, 91, 92]. In fact, more than 50 years separated some of the earliest experimental [93] and theoretical [13, 94] studies on cantilever systems and the recognition of cantilevers as a platform for chemical sensors by the broader

research community [11, 34, 53, 95-99]. Due to this remarkable time span, well-established analytical models of cantilevers are now available and can be used to design and evaluate microcantilever transducers or analogous MEMS for chemical analytical applications. Furthermore, the classical models of cantilever devices have been frequently revisited, refined and compared against the results of numerical computational methods that become widespread in recent years [41, 100]. Nevertheless, simple classical models remain very useful for understanding basic principles of MEMS sensors. Depending on the operation mode of a MEMS sensor, static or dynamic models are applicable.

In the beginning of the 20th century, development of the elasticity theory by Timoshenko [13, 94] and experimental studies of thin films by Stoney [93] become major milestones in developing fundamental analytical models that would subsequently find wide applications in MEMS analysis. These classical models [13, 93] focused primarily on static deformations of purely elastic beams and plates have been revisited more recently [8, 39-41, 101-104]. As applied to various cantilever sensors operating in the static mode, the expression for strain-induced deformations of bimaterial plates derived by Timoshenko [13] in the 1920s appeared to be of particular significance.

2.1 Static Deformations

In the absence of external gravitational, magnetic and electrostatic forces, cantilever deformation is unambiguously related to a gradient of mechanical stress generated in the device. Depending on a particular origin of this stress, analytical models suitable for quantitative analysis of microcantilever responses may or may not be available. For instance, simple models are applicable to thermally induced stresses and concomitant deformations of microcantilevers made of two layered materials with different coefficients of thermal expansion. Theoretical evaluation of bimetal thermostats reported by Timoshenko [13] provided an analytical expression for the radius of curvature of a bimaterial cantilever as a function of a temperature change. This deformation resulting from unequal thermal expansion of each layer has been used extensively as an operation principle of thermostats and often referred to as the bimetallic effect. Taking into account the length of the cantilever, l, the respective defection of the tip, Δz [$= l^2 / (2R)$], can be expressed as

$$\Delta z = \frac{3l^2}{t_1 + t_2} \left[\frac{\left(1 + \frac{t_1}{t_2}\right)^2}{3\left(1 + \frac{t_1}{t_2}\right)^2 + \left(1 + \frac{t_1 E_1}{t_2 E_2}\right)\left(\frac{t_1^2}{t_2^2} + \frac{t_2 E_2}{t_1 E_1}\right)} \right] (\alpha_1 - \alpha_2)\Delta T \qquad (1)$$

where t_1 and t_2 is the thickness of the two layers of the bimaterial plate, E_1 and E_2 are the Young's moduli, and α_1, and α_2 are the thermal expansion coefficients for the materials of these layers, respectively. While the objective of Timoshenko's original work was evaluation of bimaterial thermostats, a strain induced deformation is also an important response mechanism of cantilever based chemical sensors in which a chemically selective layer undergoes expansion upon interaction with its chemical medium [104].

Femto-Joule level calorimetric sensitivity of conventional AFM cantilevers demonstrated experimentally by Barnes *et al.* is consistent with the theoretical predictions made using Equation (1) [52]. As applied to chemical and biological sensors, cantilever based calorimetry enables two transduction scenarios (Figure 4). First, the presence of analyte species can be detected due to the heat associated with their adsorption on the transducer. Second, the heat produced in the course of a subsequent chemical reaction on the cantilever surface can be characteristic of the analyte presence. However, molecular adsorption processes and interfacial chemical reactions may also affect mechanical stresses in thin plates more directly and independently of the thermal effects. Apart from fundamental interest in the direct conversion of chemical energy into mechanical energy [18], this mechanism means that cantilever transducers are compatible with many responsive phases and can function in both gas and liquid environments.

It has been known since the 1960s that molecular and atomic adsorbates on atomically pure faces of single crystals tend to induce significant surface stress changes. Long before the first microfabricated cantilevers were created, changes in surface stresses in these systems had been studied by measuring minute deformations of relatively thin (up to 1 mm) plates. Using this method, often referred to as the beam-bending technique [38, 91, 92], Kosch *et al.* studied [105, 106] surface stress changes induced by adsorption of atoms on atomically pure surfaces in vacuum. Using the Shuttleworth equation [69], the surface stress, σ, and surface free energy, γ, can be interrelated

$$\sigma = \gamma + \left(\frac{\partial \gamma}{\partial \varepsilon} \right) \tag{2}$$

where σ *i*s the surface stress. The surface strain $d\varepsilon$ is defined as ratio of change in surface area, $\partial\varepsilon=\partial A/A$. In many cases, the contribution from the surface strain term can be neglected and the free energy change approximately equals the change in surface stress.

Adsorbate and chemically induced surface stresses have also been extensively studied with regard to their role in colloidal systems. Important examples of colloidal phenomena associated with the surface stress changes include swelling of hydrogels upon hydration or formation of surfactant monolayers at the air-water interface [107]. Fundamental studies of adsorp-

tion and absorption induced mechanical phenomena, however, had limited implications for chemical sensors until mass produced microcantilevers for AFM became widely available. As compared with their macroscopic predecessors, microcantilevers coupled with the optical lever readout greatly simplified real-time measurements of surface stress changes in the low mN m^{-1} range.

Cantilevers intended for use as chemical sensors are typically modified so that one of the sides is relatively passive while the other side exhibits high affinity to the targeted analyte. In order to understand how different modifying coatings provide responses of cantilever sensors in the static bending mode, it is useful to consider the three distinctive models. The first model is most adequate when interactions between the cantilever and its environment are predominantly surface phenomena.

In general, changes in surface stresses can be largely attributed to changes in Gibbs free energy associated with adsorption processes. An example of this situation is given in Figure 5, where chemisorption of straight-chain thiol molecules on a gold coated cantilever is schematically depicted. Since spontaneous adsorption processes are driven by an excess of the interfacial free energy, they are typically accompanied by reduction of the interfacial stress. In other words, surfaces usually tend to expand (see Figure 5) as a result of adsorptive processes. This type of surface stress change is defined as compressive, referring to a possibility of return of the surface into the original compressed state. The larger the initial surface free energy of the substrate, the greater the possible change in surface stress results from spontaneous adsorption processes. Compressive surface stresses were experimentally observed on the gold side of gold coated cantilevers exposed to vapor-phase alkylthiols [80, 81].

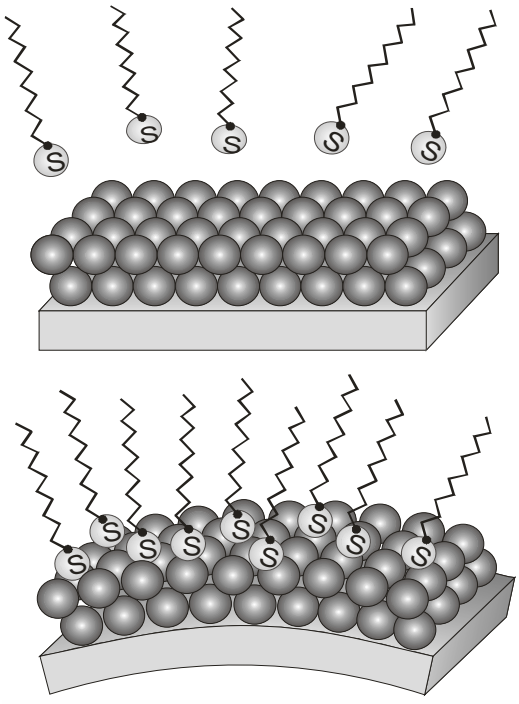


Figure 5. Schematic depiction of chemisorption of straight-chain thiol molecules on a gold coated cantilever. Spontaneous adsorption processes are driven by an excess in the interfacial free energy, and accompanied by reduction of the interfacial stress.

In many cases, adsorbate-induced deformations of thin plates can be accurately using a

modification of the relationships originally derived by Stoney [93, 101]:

$$\frac{1}{R} = \frac{6(1-\upsilon)}{Et^2}\Delta\sigma \tag{3}$$

where R is the radius of microcantilever curvature, υ and E are Poisson's ratio and Young's modulus for the substrate, respectively, t is the thickness of the cantilever, and $\delta\sigma$ is the differential surface stress. Knowledge of the radius of curvature, R, allows the tip displacement of a microcantilever with length l tip to be determined by

$$\Delta z = \frac{1}{2}\frac{l^2}{R} = \frac{3l^2(1-\upsilon)}{Et^2}\Delta\sigma \tag{4}$$

When adsorbate-induced stresses are generated on ideal smooth surfaces or within coatings that are very thin in comparison to the cantilever, the analysis according to Equation (4) is rather straightforward. Using Equation (4), the predictions for the cantilever bending can be based on the expected surface stress change. Alternatively, responses of cantilever sensors converted into surface stress changes can be analyzed as the measure of the coating efficiency independently of the transducer geometry.

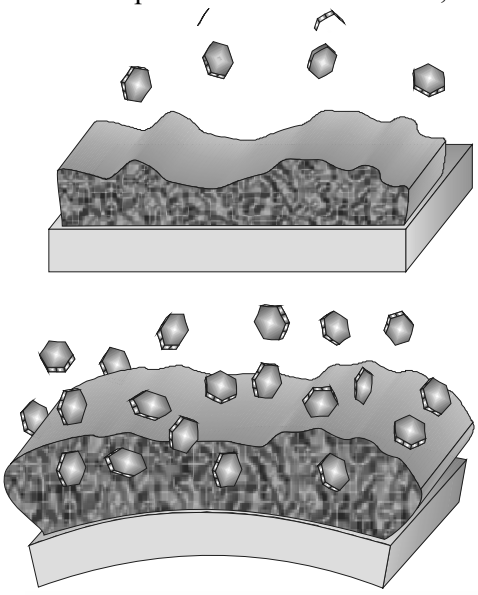


Figure 6. Schematic depiction of analyte-induced cantilever deformation when the surface is modified with a thicker analyte-permeable coating that undergoes swelling in presence of analyte molecules.

The second model of analyte-induced stresses (Figure 6) is applicable for a cantilever modified with a much thicker than a monolayer analyte-permeable coating [108, 109]. Taking into account interactions of the analyte molecules with the bulk of the responsive phase, a predominant mechanism of cantilever deflection can be described as deformation due to analyte-induced swelling of the coating (Figure 6). Such swelling processes can be quantified using approaches developed in colloidal and polymer science, *i.e.* by evaluating molecular forces acting in the coating

and between the coating and the analyte species. In general, dispersion, electrostatic, steric, osmotic and solvation forces [107], acting within the coating can be altered by absorbed analytes. Depending on whether it is more appropriate to describe the responsive phase as solid or gel-like, these altered forces can be put into accordance with, respectively, stress or pressure changes inside the coating. An in-plane component of this change multiplied by the coating thickness yields an apparent surface stress change that can be used in Stoney's model [Equation (4)] in order to estimate deflections of a cantilever coated with thin, soft, responsive films. It is important to note that the magnitude of apparent surface stress scales up in proportion with the thickness of the responsive phase.

The third model (Figure 7) is most relevant to nanostructured interfaces and coatings, such as surface-immobilized colloids, that have been recently recognized as a very promising class of chemically responsive phases for cantilever sensors and actuators [63, 110-114]. It is worthy to note that grain boundaries, voids, and impurities have been long known as being responsible for high intrinsic stresses in disordered, amorphous and polycrystalline films [105].

Analyte-induced deflections of cantilevers with structured phases (Figure 7) combine mechanisms of bulk, surface, and inter-surface interactions [107]. A combination of these mechanisms facilitates efficient conversion of the energy of receptor-analyte interactions into mechanical energy of cantilever bending. Recent studies demonstrated that up to two orders of magnitude increases in cantilever responses can be obtained when receptor molecules are immobilized on nanostructured instead of smooth gold surfaces [113, 115, 116]. Furthermore, nanostructured responsive phases offer an approach to substantially increase the number of binding sites per cantilever without compromising their accessibility for the analyte. In fact, many of these nanostructured phases exhibit behaviors of *molecular sponges*. Although deflections of cantilevers with nanostructured coatings or

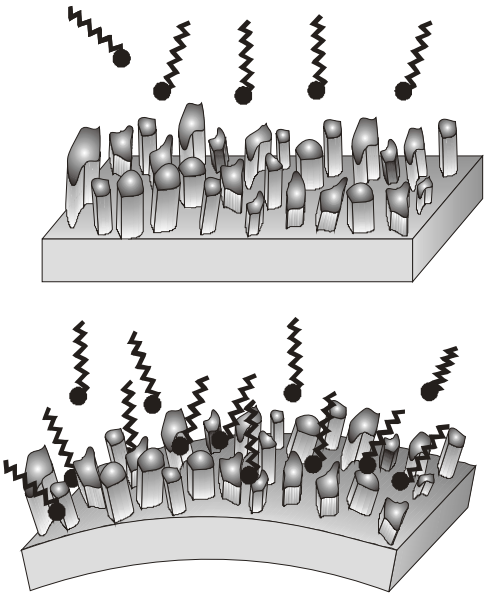


Figure 7. Schematic depiction of analyte-induced cantilever deformation in the case of a structured responsive phase. Bulk, surface, and inter-surface interactions are involved in the response mechanism.

thicker hydrogel layers cannot be accurately predicted using the analytical models mentioned above, estimates for the upper limit of the mechanical energy produced by any cantilever transducer can always be based on simple energy conservation. This upper limit in available energy is given by the product of the energy associated with the binding site-analyte interaction and the number of such interactions on the cantilever surface.

2.2 Resonant Operation

Cantilever transducers operating in gases or in vacuum can be treated as weakly damped mechanical oscillators. Their resonant behavior can be readily observed using excitation in alternated electric, electromagnetic or acoustic fields. Furthermore, minute sizes and mass of microfabricated cantilevers makes them susceptible to thermally induced noise, which has the same origin as Brownian motion of small particles in liquids. Therefore, cantilever sensors may operate in the resonant mode either with or without external excitation.

As a first approximation, it can be assumed that the cantilever tip displacement is directly proportional to the force exerted on the cantilever tip. Then, a simplified model of a resonating cantilever transducer can be based on the Hook's law applied to a rectangular leaf spring with an effective suspended mass m_0 and a spring constant k. The effective suspended mass of a cantilever can be related to the total mass of the suspended portion of the beam, m_b, through the relationship: $m_0 = n\, m_b$, where n is a geometric parameter. For a rectangular cantilever, n has a typical value of 0.24 and the spring constant k is given by [35]:

$$k = \frac{Ewt^3}{4l^3} \tag{5}$$

where E is the modulus of elasticity for the material composing the cantilever and w, t, l are the width, thickness and length of the cantilever, respectively. Assuming a spring constant k and an effective suspended mass m_0, which consists of both a concentrated and a distributed mass, the microcantilever fundamental resonance frequency, f_0, in absence of damping can be approximated as [35]

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m_0}} \tag{6}$$

Equation (6) is often used as a starting point in estimating the mass sensitivity of resonating cantilever sensors of various shapes and sizes [63, 82, 83, 95, 117, 118]. As a rule, Equation (6) gives a good approximation of the

resonance frequency for a weakly damped mechanical resonator, such as a microscopic cantilever in air. However, more accurate calculations of the microcantilever resonance frequencies require the dissipation of the resonator energy through various mechanisms to be taken into account. This can be done by introducing a mechanical quality factor (Q-factor). In the case of damping force being proportional to the cantilever velocity (viscous damping), the resonance frequency of the mechanical resonator is [35, 119].

$$f_{0,Q} = \frac{1}{2^{3/2}\pi}\sqrt{\frac{k}{m_0}}\frac{\sqrt{2Q-1}}{Q} \tag{7}$$

When Equations (6) and (7) are used to analyze gravimetric responses of resonating microcantilever transducers, it is reasonable to assume that the spring constant, k, remains unaffected.

In summary, the resonant operation of cantilever transducers encompasses three mechanisms: (i) adsorbate-induced mass-loading, (ii) chemically induced changes in the cantilever stiffness, and (iii) mechanical damping by the viscous medium. A more detailed discussion of various damping mechanisms in microcantilever transducers related to external as well as intrinsic dissipation phenomena is given in the next section.

2.3 Energy Dissipation in Microcantilevers

Mechanical deformations in MEMS always involve appreciable dissipation of mechanical energy into thermal energy. Mechanisms that determine this dissipation are related to inelastic phenomena in solids and viscous properties of fluids [120]. In analogy to other types of resonators, the mechanical quality factor (or Q-factor) is commonly used to quantify energy dissipation in MEMS. The Q-factor is inversely proportional to the damping coefficient [121] or total energy lost per cycle of vibration in a microcantilever transducer and can be defined as

$$Q = \frac{2\pi W_0}{\Delta W} \tag{8}$$

where W_0 and ΔW_0 are, respectively, the mechanical energy accumulated and dissipated in the device per vibration cycle.

It is important to emphasize that both the resonance behavior of any microcantilever and its off-resonance thermal noise are critically dependent on the Q-factor [58, 121-126]. Therefore, the Q-factor, is one of the important characteristic of MEMS sensors operating in both resonance and static re-

gimes. Based on the spectral analysis, the *Q*-factor can be calculated as a ratio of the resonance frequency, f_0, to the width of the resonance peak at its half amplitude. Hence, *Q*-factor is frequently used to characterize the degree of the resonance peak sharpness. Alternatively, *Q*-factors of mechanical oscillators can be related to time constants of exponentially decaying oscillator amplitude during a ring-down process [125]

$$Q = \pi \tau_d f_0 \tag{9}$$

where τ_d is the time constant of the exponentially decaying ring-down amplitude.

The *Q*-factor of microcantilevers depends on a number of parameters, such as cantilever material, geometrical shape, and the viscosity of the medium. Obviously, increased damping of a microcantilever oscillator by the medium translates into lower *Q*-factor values [127] as compared to the same oscillator in vacuum. Models of drag forces exerted on solid bodies in fluids [120, 128, 129] can be used to evaluate viscous damping effects. A very important distinctive feature of viscous damping is that the damping force is proportional to the linear velocity of the vibrating cantilever. The other damping mechanisms, involving clamping loss and internal friction within the microcantilever, were reviewed in recent studies by Yasumura *et al.*[125]. As a rule, these dissipation mechanisms are associated with damping forces independent of the linear cantilever velocity. Clamping loss has an insignificant contribution to the total dissipation in the case of longer microcantilever with high length-to-width and width-to-thickness ratios. However, ultimate minimization of clamping loss can be achieved in oscillators with double-paddle or "butterfly" geometries [130] rather than single-clamped cantilevers or double-clamped bridges. Hence, fundamental studies of intrinsic friction effects in MEMS often rely on measurements of resonances in double-paddle resonators. *Q*-factors as high as 10^5 were reported for torsional butterfly-shaped resonators fabricated from a single-crystal silicon [131-133].

Internal friction can be linked to a variety of physical phenomena, in particular, thermo-elastic dissipation (TED) [128] motion of lattice defects, phonon-phonon scattering, and surface effects [134, 135]. The TED limit and phonon-phonon scattering mechanisms correspond to very high *Q*-factors (10^6 to 10^8), which can hardly be observed experimentally due to the contribution from other dissipation mechanisms present in real MEMS. The fact that surface effects may limit *Q*-factors of MEMS oscillators in vacuum can be verified by annealing the device and controllably changing its surface [125]. As the thickness of the oscillator decreases, TED becomes even a less significant mechanism of dissipation [125].

As mentioned previously *Q*-factors of MEMS resonators in vacuum can be very high [133, 136]. However, *Q*-factors of rectangular microcantilevers

in air are typically in the range of 10 to 1000 while cantilever transducers in aqueous solutions rarely have *Q*-factors above 10. Very strong viscous damping in liquids makes resonant operation of microcantilevers, and, in turn, measurements of adsorbed mass using microcantilever sensors, rather challenging.

3. COMMON READOUT SCHEMES

Operation of any cantilever sensor relies on real-time measurements of cantilever deflections with at least nanometer accuracy. Therefore, an important part of any cantilever sensor is a read-out system capable of monitoring changes in one of the parameters directly related to the cantilever deflection. Such parameters include cantilever tip position, spatial orientation, radius of curvature, and intrinsic stress. Specific requirements for the read-out of cantilever sensors can be dictated by the operation mode (either static or dynamic), cantilever design and materials used as well as the magnitude of expected responses. In this section, we discuss means of cantilever readout that can be broadly classified as optical and electrical [35]. Using optical, piezoresistive, piezoelectric, capacitance, or electron tunneling methods [35], deformations and resonance frequency shifts of cantilever transducers can be measured with sufficient precision. All these methods are compatible with array formats.

In order to insure the best possible performance of cantilever sensors, inherent advantages and disadvantages of different readout techniques were analyzed in recent studies. The optical beam deflection method was shown to have excellent readout efficiency in the case of cantilevers with a reflecting area of at least a10x10 μm^2. Optical readout techniques may, however, be inefficient when applied to nanocantilevers. The shortcomings of some optical techniques, in particular the optical deflection method, are related to loss of intensity and directionality of optical beams reflected (scattered) by nanosize cantilevers. By contrast, electron transfer methods can be used with cantilevers that are only a few hundred nanometers long. An important issue still to be addressed is the readout of nanocantilevers arranged in dense arrays. Among already explored readout methods, a charge shuttling method is one of well suited for nanocantilevers arrays. The basic concept of this method is similar to electrostatic charge shuttling demonstrated by Tuominen *et al.* [137]. However, implementation of electron transfer signal transduction in aqueous environments is challenging. Electroactive ions in electrolytes tend to cause parasitic currents that overwhelm electron transfer signal. These leakage currents, however, can be significantly reduced by using proper insulation, reducing bias voltage, and controlling electroactive ions in the solution.

3.1 Optical Methods

It is noteworthy that cantilever sensors inherited not only the unique advantages of a microfabricated AFM probe, but also the elegant "optical lever" read-out scheme commonly used in modern AFM instruments. Optical methods most extensively used for measurements of cantilever deflections in AFM [138-142] include optical beam deflection (also referred to as the "optical lever" method) [139] and optical interferometry [138-140]. In the optical beam deflection technique, a laser diode is focused at the free end of the cantilever. The optical lever method proposed for the use in AFM by Amer *et al.* [139] appeared to be simpler and at least as sensitive as more complex interferometric schemes. In Amer's studies [139], a small mirror was attached to a cantilever (made then out of a tungsten wire) so that a position of a laser beam bounced off this mirror could be monitored using a position sensitive photodetector (PSD). This optical detection scheme (Figure 8) can discern extremely small changes in the cantilever bending; measurements of 10^{-14}m displacements were reported. A most common type of PSD is based on a quadrant photodiode that consists of four cells: A, B, C, and D. Each of the cells is coupled to the input of a transimpedance amplifier the output voltages of which, V_A, V_B, V_C , and V_D, are proportional to the illumination of the respective quadrant. The normalized differential output, $V_{out} = [(V_A + V_C) - (V_B + V_D) / (V_A + V_B + V_C + V_D)]$, depends linearly on the vertical displacement of the weighted center of the light spot projected by the cantilever. Absence of electrical connections to the cantilever, linear response, simplicity and reliability are important advantages of the optical lever method. As this method has been used in vast majority of the work on cantilever sensors, its limitations are well recognized. For instance, changes in the optical properties of the medium surrounding the cantilever may interfere with the output signal. This interference can largely be avoided by using a proper orientation of the cantilever relative to the optical components as discussed in the recent paper of the authors [116]. The effect of the refractive index change as well as other interfering factors can be further suppressed by using differential pairs or arrays of cantilevers. However, applications of cantilever sensors with the optical lever readout are limited to analysis of low opacity, low turbidity me-

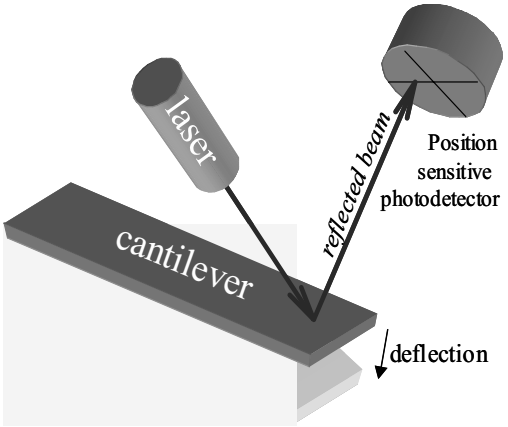


Figure 8. The "optical lever" readout commonly used to measure deflections of microfabricated cantilever probes in AFM.

dia. Another limitation of the optical lever method is related to the bandwidth of PSDs, which typically on the order of several hundred kHz.

As the requirements of the high bandwidth become more critical in the case of smaller and stiffer cantilevers that operate in the resonant mode, alternatives to the optical lever readout were explored. For instance, motion of a microscopic structure, such as a cantilever, illuminated with a tightly focused laser beam, produces change in the spatial distribution of the reflected and/or scattered light. A simple spot photodetector alone or in combination with a knife-edge obstacle can be used to monitor these intensity fluctuations [143]. The readout bandwidth of this method can be extended into the GHz range by using a small area, high-speed avalanche photodiode. Approaches based on a single photodetector and light scattering, however, suffer from the interference with ambient light, nonlinear response, and a poorly controllable optical gain. More accurate high-band width optical measurements of cantilever deflections can be carried out using interferometric schemes. Notably, interferometry was the first optical technique used for measurements of cantilever deflections in AFM. Interferometry was revisited as a MEMS readout and characterization tool more recently because of its potential for high bandwidth high-resolution mapping of nanometer scale motions of small cantilevers [45, 138] arranged in large 2-D arrays. Notably, Rugar *et al*. [44, 138] used interefrometry to measure sub-nanometer deflections of the ultrasensitive cantilevers designed for ultrasensitive force measurements that could ultimately permit single-spin magnetic resonance magnetic resonance microscopy.

Optical detection techniques were also used in order to simultaneously read out large arrays of bimaterial cantilevers that found applications in IR imaging [59, 61]. In the case of a cantilever array, a light source illuminates the whole array so that the reflected light can be either directly analyzed by a CCD-camera or combined with the reference beam to produce an interferogram.

3.2 Piezoresistance Method

Piezoresistivity is the phenomenon of changes in the bulk resistivity with applied stress. When a silicon microcantilever with an appropriately shaped doped region is deformed, the change in the resistance of the doped region reflects the degree of the deformation. The most common materials that exhibit a strong piezoresistive effects are doped single crystal silicon [54, 144-146] and doped polysilicon [97, 147].

The variation in resistance is typically measured by including the cantilever into a dc-biased Wheatstone bridge.

Typical resistance of a single crystal silicon microcantilever with a boron doped channel is a few kΩ . When a bias, *V,* is applied to the Wheatstone bridge with resistors of identical initial resistance, *R*, the differential voltage, ΔV, across the bridge can be expressed as $\Delta V = V\,(\Delta R/4R)$.

Piezoresistive cantilevers are often designed to have two identical "legs" so that the resistance of the boron channel can be measured by applying contacts to the cantilever base next to the legs (see Figure 2b). Figure 9 shows an example of a single crystal Si piezoresistive cantilever that is used in AFM and has a more common rectangular shape.

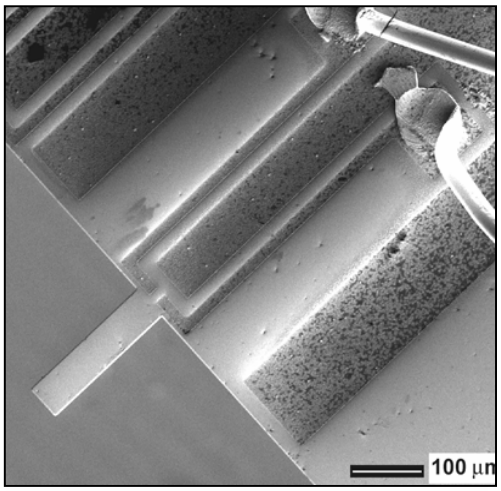


Figure 9. An example of a piezoresistive cantilever that can be used in both AFM and cantilever sensors.

Typical piezoresistive cantilevers have a piezoresistive layer that is at least one hundred nanometers thick which imposes a constrain on the minimum thickness of the microcantilever. Another disadvantage of the piezoresistive technique is that it requires current to flow through the cantilever. This results in additional dissipation of heat and associated thermal drifts. When the cantilever is heated appreciably above the ambient temperature, any changes in the thermal conductivity of the environment will result in fluctuations of the cantilever temperature that, in turn, may lead to parasitic cantilever deflection and piezoresistance changes.

3.3 Piezoelectric Method

Piezoelectric readout technique requires deposition of piezoelectric material, such as ZnO, on the cantilever. Due to a piezoelectric effect, transient charges are induced in the piezoelectric layer when a cantilever is deformed [7, 148-150]. Lee and White [151] created piezoelectric cantilevers with resonances in the acoustic frequency range. The piezoelectric cantilever used in these studies had a zinc oxide (ZnO) piezoelectric thin film sandwiched between two aluminum layers on a supporting layer of silicon nitride. A few years later, Lee *et al.* [152] micromachined piezoelectric cantilevers using PZT films. More recently, Adams *et al.* [153] demonstrated a microcantilever chemical detection platform based on an array of piezoelectric microcantilevers with a power consumption in the nanowatts range. Lee *et al.* [152] reported that micromachined piezoelectric cantilevers 100 μm wide and

200 μm long and 2.1 μm thick had a gravimetric sensitivity of 300 cm^2/g, which enabled detection of 5 ng mass. Characterization of the gravimetric sensitivities in this study was conducted by depositing known amount of gold on the backside of the cantilevers.

The main disadvantage of the piezoelectric as well as piezoresistance readout is that they require electrical connections to the cantilever. An additional disadvantage of the piezoelectric technique is that in order to obtain large output signals it requires the thickness of the piezoelectric film to be well above the values that correspond to optimal mechanical characteristics. Furthermore, the piezoelectric readout is inefficient when slowly changing cantilever deflections need to be measured. Because of the aforementioned disadvantages, application of piezoelectric readout to MEMS sensors is somewhat limited.

3.4 Capacitance Method

Capacitance readout is based on measuring the capacitance between a conductor on the cantilever and another fixed conductor on the substrate that is separated from the cantilever by a small gap [57, 98, 154]. Changes in the gap due to cantilever deformation result in changes in the capacitance between two conductor plates. Since the capacitance of a flat condenser is inversely proportional to the separation distance, sensitivity of this method relies on a very small gap between the cantilever and the substrate. Capacitance readout suffers from interference with variations in the dielectric constant of the medium [155]. While differential schemes may eliminate this interference, electrically conductive media, such as electrolytes, make capacitance readout more challenging. One of the main advantages of capacitance readout is that it can be used in integrated MEMS devices that are fully compliant with standard CMOS technology. An interesting variation of the capacitance methods is the "electron shuttling" regime that is especially promising for nano-electro-mechanical systems. For instance, Erbe *et al.* reported on the "quantum bell" [156, 157] that consists of five metal-coated cantilever structures and operates in the radio frequency range.

3.5 Electron Tunneling Method

Electron tunneling was utilized to measure the deflection of cantilevers in AFM [35]. The tunneling occurs between a conducting tip and the cantilever separated by a subnanometer gap. The tunneling current is very sensitive to the size of this gap and, therefore, deformations of the cantilever. This current can be described as [158-159].

$$I \propto Ve^{-a\sqrt{\Phi s}} \tag{10}$$

where V is the bias voltage, Φ is the height of the tunneling barrier, s is the tunneling gap distance, and a is a conversion factor with a value of 1.025 $Å^{-1}$ $eV^{-1/2}$. For typical values of Φ and s, the tunneling current increases by one order of magnitude for each 0.1 nm change in s [158]. Using an electron tunneling readout technique, cantilever displacements as small as 10^{-4} nm have been measured [158]. Extremely high sensitivity of the tunneling readout is associated with its nonlinear response and a limited dynamic range. It is also worthy to note that tunneling processes are sensitive to the nature of materials between which the tunneling process occurs, which often translated into challenging requirements to device implementation. Despite its well known limitations, electron tunneling readout was successfully used in accelerometers [158], infrared sensors [160] and magnetic field sensors [161].

4. FIGURES OF MERIT OF CANTILEVER TRANSDUCERS

Some of the most important figures of merit of any chemical or biological sensor are responsivity, limit of detection (LOD), specificity, and reproducibility. Chemical specificity of MEMS sensors may rely on the use of certain responsive phases, such as polymers, self-assembled monolayers, or biological receptors that exhibit higher affinity to the targeted analytes. This reliance on the selective properties of responsive coatings is a common feature of microcantilever sensors operating in both static and resonant regimes. However, transduction efficiency of the two regimes is dictated by very distinct models and mechanisms. Transduction efficiency of the static mode increases when the stiffness of the cantilever is reduced. Therefore, longer cantilevers with very small spring constants are preferable for the use in the static mode. On the other hand, the sensitivity of the resonant mode increases progressively with the operation frequency. As a rule, static and resonant operation of the same microcantilever sensor is characterized by the same specificity but different responsivities and LODs. The fundamental limits of microcantilever chemical transducers are determined by ratios of their responsivities to the levels of intrinsic noise. The next two sections focus on the models specific to optimization of microcantilever sensors based on, respectively, measurements of resonance frequency variations and adsorption-induced deformations. In particular, noise sources that influence the smallest detectable displacements and resonance frequency shifts are discussed.

4.1 Responsivity of the Resonance-Based Transducers

Although adsorption-induced stresses were extensively explored as a transduction principle in many cantilever sensors, the advantage of the resonant operation is that it can potentially provide mass detection at the single molecule level. The resonance frequency of a cantilever beam depends on its geometry as well as the elastic modulus and density of its material. By changing cantilever dimensions, its resonance frequency can be varied from hundreds of Hz to a few GHz (see Figure 3). For a given cantilevers mass, higher spring constants correspond to higher resonance frequencies. For a given cantilever thickness, shorter cantilevers have higher spring constants. Depending on the cantilever material, GHz resonance frequencies can be achieved, when the cantilever length is less than a few microns [62].

The dependence of the fundamental frequency on the cantilever parameters for a rectangular cantilever with dimension, *l*, *w*, and *t*, respectively, in length, width and thickness can be expressed as [35]

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{E\,w\,t^3}{4l^3\left(m_c + 0.24\,w\,t\,l\rho\right)}} \tag{11}$$

where ρ is the density of the adsorbate material and m_c is the concentrated mass. The mass of the adsorbed material can be determined from the initial and final resonance frequency and the initial mass of the cantilever [162]

$$\frac{f_0^2 - f_1^2}{f_0^2} \approx \frac{\Delta m}{m} \quad \text{or} \quad \frac{f_0 - f_1}{f_0} \approx \frac{1}{2}\frac{\Delta m}{m} \tag{12}$$

where f_0 and f_1 are the initial and final frequency respectively and Δm and m are adsorbed mass and initial mass of the cantilever, respectively. If the adsorption is not confined to the very end of the cantilever [82], Equation (12) should be modified in order to take into account the effective mass of the cantilever. For any geometry of a microcantilever sensor, the mass responsivity can be defined as

$$S_m = \lim_{\Delta m \to 0} \frac{1}{f_0}\frac{\Delta f}{\Delta \Gamma} = \frac{1}{f_0}\frac{df}{d\Gamma} \tag{13}$$

where $\Delta\Gamma$ is normalized per active area of the device ($\Delta\Gamma = \Delta m/A$, where A is the active area of the cantilever). Equation (13) shows that the mass responsivity is the fractional change of the resonant frequency of the structure with addition of mass to the sensor.

4.2 Responsivity in the Static Deformation Regime

Adsorption-induced cantilever bending enabled some of the most sensitive detection of trace-level analytes in gases and is a preferable mode of cantilever operation in liquids. A distinctive feature of microcantilever sensors operating in the static mode is that they convert a sum of weak intermolecular forces involved in analyte-sensor interactions into readily measured displacements. This means, that the sensor may respond differently to the same amount of different analytes depending on the sensor-analyte affinity. Furthermore, adsorbate-induced stresses and associated deformations can be distinguished from the bulk effects, such as changes in volume of thicker polymer films, which also lead to cantilever deformations. Currently, rigorous quantitative evaluation of the sensor responsivity according to these mechanisms is lacking. Semi-quantitative analysis of the static mode responsivity can be based on Equations (1)-(4) as reported in the literature [36, 104, 115, 122]. In particular, a linear relationship between the cantilever tip displacement and the differential surface stress is given by Equation (4). According to Equation (4), there is a quadratic dependency of surface stress induced deflections on the microcantilever length. It is worthy to note that a quadratic dependency of microcantilever deflection responses on the microcantilever length is also predicted by Equation (1) for the case of stresses in the cantilever bulk. For microcantilever operating in the static mode, an increased length is, therefore, a prerequisite of high responsivity. This conclusion is consistent with the fact that microcantilevers 400 to 1500 μm long were successfully used in chemical sensors operating in the static mode [113, 115, 163]. Equations (1) and (4) also indicate a strong effect of the thickness of a microcantilever on its deflection responses. In fact, in the case of very thin responsive phases or purely surface interactions, Equation (4) predicts that deflection responses are inversely proportional to the total thickness of the microcantilever. However, more complex dependencies follow from Equation (1) in the case of expanding coatings with the thicknesses comparable to the cantilever thickness.

4.3 Intrinsic Noise Sources

Noise processes in microcantilever sensors can be divided into processes intrinsic to the device and those related to interactions with its environment (for instance, adsorption-desorption noise) or originated from the readout. Here we focus on the intrinsic noise mechanisms since they determine ultimate fundamental limits of the microcantilever sensors performance. One of the essential features of microcantilevers is that they are mechanical devices (oscillators) that can accumulate and store mechanical energy. Over the past decades there have been extensive efforts to identify the fundamental intrin-

sic sources of noise in mechanical systems and identify the relationships between parameters of the mechanical system and its noise level [123].

When a microcantilever detector is equilibrated with the ambient thermal environment (a thermal bath), there is a continuous exchange of the mechanical energy accumulated in the device and thermal energy of the environment. This exchange dictated by the fluctuation dissipation theorem results in spontaneous oscillation of the microcantilever so that the average mechanical energy per mode of cantilever oscillation is defined by thermal energy, $k_B T$. Sarid [35] referred to this type of noise as "thermally induced lever noise". In other words, any cantilever in equilibrium with its thermal environment has a "built-in" source of white thermal noise [164]. The amplitude of the resulting thermally induced oscillation of a cantilever beam is proportional to the square root of the thermal energy. As a result of the dynamic exchange between cantilever mechanical energy and the ambient thermal energy, the actual frequency, f, of thermally induced cantilever oscillations at any given moment can also noticeably deviate from the resonance frequency, f_0. The amplitude of such frequency fluctuations, δf_0, due to the exchange between mechanical and thermal energy is [164]

$$\delta f_0 = \frac{1}{z_{max}} \sqrt{\frac{2\pi f_0 k_B T B}{k Q}} \tag{14}$$

where z_{max} is the amplitude of the cantilever oscillations, k_B is the Boltzmann constant (1.38×10^{-23} J/K), T is the absolute temperature (300 K at room temperature), B is the bandwidth of measurement, and k is the cantilever force constant. Equation (14) predicts increased absolute fluctuations of the resonance frequency, δf_0, as the resonance frequency f_0 increases. However, relative frequency instability, $\delta f_0 / f_0$, decreases in the case of higher frequency oscillators

$$\frac{\delta f_0}{f_0} = \frac{1}{z_{max}} \sqrt{\frac{2\pi k_B T B}{k Q f_0}} \tag{15}$$

Although Equations (14) and (15) are valid for thermally excited cantilevers, they can also be used to evaluate effects of thermal noise on the frequency instability of any externally driven cantilever [164]. As applied to cantilever sensors operating in the resonance mode, an important implication of Equations (14) and (15) is that frequency instability due to effects of thermal noise can be minimized by driving the transducer with the highest possible amplitude. In the case of self-oscillating systems with a positive feedback, however, an intrinsic Q-factor should always be used in Equation (15)

since the amplitude of cantilever oscillation, z_{max}, is already explicitly taken into account in this analysis.

By changing the physical dimension of a cantilever, its mass detection limit can be affected by many orders of magnitude The smallest (thermal noise limited) detectable mass Δm_{th} can be found by combining Equations (6) and (15)

$$\Delta m_{th} = \frac{8G}{z_{max}} \frac{m_0^{5/4}}{k^{3/4}} \sqrt{\frac{k_B T B}{Q}} \tag{16}$$

where G is a geometrical factor that is close to unity and accounts for distribution of the added mass.

Thermomechanical noise of microcantilevers is also known to be a fundamental noise source in AFM [35, 122]. Similar analysis is applicable to microcantilever sensors operating in the static regime. The analysis provided by Sarid [35] involves the Q-factor of a vibrating microcantilever, its angular resonance frequency, ω_0 and stiffness, k. While Q-factor can be defined empirically as the ratio of the resonance frequency to the resonance peak width, knowing the exact mechanisms of cantilever damping is important for evaluation of the thermo-mechanical noise spectrum. The model evaluated by Sarid [35] assumes that damping of the cantilever is of viscous nature. Assumption of predominantly viscous damping is valid for microcantilevers in air or water and, therefore, justified for micromechanical devices used as force probes in SPM. In the case of a cantilever in a viscous medium, such as air or water, the damping force is proportional to the cantilever linear velocity. The resulting noise density spectrum can be expressed as [121]

$$< \delta z_{TM}^{\ 2} >^{1/2} = \sqrt{\frac{4k_B T B}{Q k} \left[\frac{\omega_0^{\ 3}}{\left(\omega_0^{\ 2} - \omega^2\right)^2 + \omega^2 \omega_0^{\ 2} / Q^2} \right]} \tag{17}$$

The expression given by Equation (17) predicts a frequency independent noise density for the frequencies well below the mechanical resonance frequency, ω_0, (*i.e.*, $\omega << \omega_0$). At these low frequencies, the rms of the cantilever tip displacement due to thermo-mechanical noise is

$$< \delta z_{TM}^{\ 2} >^{1/2} = \sqrt{\frac{4k_B T B}{Q k \omega_0}} \tag{18}$$

However, at the resonance (*i.e.* $\omega = \omega_0$) [35]

$$< \delta z_{TM}{}^{2} >^{1/2} = \sqrt{\frac{4 k_B T B Q}{k \omega_0}} \tag{19}$$

An alternative model for the thermomechanical model has been described elsewere [121]. Majorama *et al*. [121] showed that the density of thermo-mechanical noise follows a $1/f^{1/2}$ dependence below the resonance when the damping is due to intrinsic friction processes (see Figure 10, TM$'$ curve). Our analysis shows that, regardless of the dissipation mechanism, the off-resonance thermo-mechanical noise is lower in the case of microcantilevers with higher Q-factor and higher k. It should be emphasized that, while predictions based on Equation (17) are often reported in the literature [124, 165], the noise density calculated according to the two alternative models may substantially deviate from each other at low frequencies [121, 166, 167]. Furthermore, the intrinsic friction model predicts the low frequency noise to be independent of the cantilever resonance frequency provided that the stiffness, k, is constant. By contrast, the viscous damping model predicts that the low frequency noise of a microcantilever detector can be decreased by increasing its resonance frequency even without changes in its stiffness. Therefore, knowing the actual mechanisms of mechanical dissipation in the microcantilever detector can be critical in analyzing thermo-mechanical noise of microcantilever sensors.

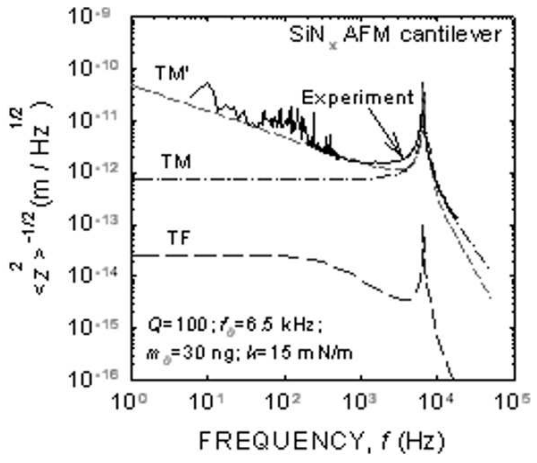


Figure 10. Spectral densities of thermo-mechanical and temperature fluctuation noise Theoretical plots calculated for a 600 nm thick silicon nitride AFM cantilever coated with 50 nm of gold are shown along with the experimental data obtained for this cantilever.

In the case of layered (for instance bimaterial) microcantilever transducers an additional noise source need to be considered. This noise mechanism is related to the temperature sensitivity of bimaterial cantilevers [13, 53, 94] and the fact that local temperature undergoes appreciable fluctuation at the microscopic scale. As discussed previously by Kruse [168], the mean square magnitude of these temperature fluctuations is

$$< \delta T^2 >= \frac{k_B T^2}{C} \tag{20}$$

where C is the total heat capacity of the microcantilever. The temperature fluctuation $<\delta T^2>$ of Equation (20) is the integration over all frequencies f, where $f = \omega/2\pi$. The spectral density of the root mean square (rms) temperature fluctuations is given by [168]

$$< \delta T^2 >^{1/2} = \frac{2\sqrt{k_B B}\, T}{G^{1/2}\sqrt{1+\omega^2\tau^2}} \tag{21}$$

where G is the thermal conductance of the principal heat loss mechanism.

Temperature fluctuation noise of the cantilever detector manifests itself as a fluctuation in z and exhibits a frequency dependence influenced by both the thermal response and the mechanical response of the microcantilever. The spontaneous fluctuations in displacement of the microcantilever caused by temperature fluctuations are given by

$$< \delta z_{TF}^{\ 2} >^{1/2} = \frac{T\sqrt{4 k_B B}}{G^{1/2}\sqrt{1+\omega^2\tau^2}} \frac{1}{\sqrt{\left(1-\frac{\omega^2}{\omega_0^{\ 2}}\right)^2 + \frac{\omega^2}{\omega_0^{\ 2} Q^2}}} \tag{22}$$

Figure 10 illustrates the spectral densities of both thermo-mechanical and temperature fluctuation noise and compares these theoretical dependencies with the experimentally measured noise behavior of a standard silicon nitride AFM cantilever.

5. DEMONSTRATED APPLICATIONS

5.1 Gas Phase Analytes

Detection of mercury vapors reported by Thundat *et al*. [95] was one of the first gas sensor applications of microscopic cantilevers. Commercially available delta-shaped silicon nitride AFM cantilevers were used in those studies. The typical length, thickness and force constant of these cantilevers were, respectively, 180 μm, 0.6 μm and 0.06 N/m. An evaporated 50 nm gold coating provided affinity of one side of the cantilevers to mercury. It was found that both resonance frequencies and static deflections of the gold coated cantilevers underwent changes in presence of mercury vapor (30 μg m^{-3}) added to a nitrogen carrier gas. When one side of the cantilevers was completely coated with gold, the resonance frequency of the cantilevers increased as a result of exposure to mercury vapors. This rather unexpected result was explained by competing effects of the absorbed mercury on the cantilever force constant and on the cantilever suspended mass. It was concluded that interaction of mercury with the gold coating led to lowering of the cantilever resonance frequency as a result of a relatively small increase in the cantilever effective mass and a more significant increase in the cantilever force constant. This model was confirmed by the fact that exposure to mercury vapors did lower the resonance frequency when only an end portion of the cantilever was coated with gold. In the latter experiment, the region close to the clamping point (which largely defines the cantilever force constant) remained without gold and, therefore, was unaffected in presence of mercury.

Ferrari and co-workers [7] used ceramic cantilevers actuated by piezoelectric excitation to measure resonant frequency changes due to adsorption of water. Ferrari *et al.* [7] used poly(N-vinylpyrrolidinone) and poly(ethyleneglycol) as hydrophilic coating materials and obtained frequency shifts of about 500 and 1400 Hz for relative humidity changes from 12 to 85%, respectively. Dual (static/dynamic) mode responses of gold coated were reported for several other gaseous phase analytes, in particular, 2-mercaptoethanol [118]. In the case of 2-mercaptoethanol, analyte-induced deflections rather than changes in the resonance frequency of gold-coated AFM cantilevers was found to be a preferable mode of sensor operation. Measurements of cantilever deflections permitted detection of mercaptoethanol vapor at concentrations down to 50 parts per billion (ppb). The calibration curve obtained in the static deflection mode had a slope of 0.432 nm per ppb in the concentration range of 0 - 400 ppb.

Fairly high sensitivity and selectivity demonstrated in the early studies on cantilever sensors relied on properties of some metals used as active coatings. For instance, gold is a very chemically inert metal that,

nevertheless has very high reactivity towards mercaptans (or thiols), *i.e.* compounds with one or more sulfohydryl (-SH) groups. High solubility of hydrogen in palladium and palladium based alloys is another mechanism that leads to selective interaction of metal coatings with gas-phase analytes. Good sensitivity of Au and Pd coated cantilevers to, respectively, mercury and hydrogen was subsequently used to implement a palm-sized, self-contained sensor module with spread-spectrum-telemetry reporting [98]. The device utilized polysilicon cantilevers operating in the static deflection mode and integrated with CMOS circuitry that provided their capacitive read-out as well as radio-frequency output for telemetry. The implemented prototype provided reversible, real-time hydrogen sensing and dosimetric (cumulative) mercury-vapor sensing. It was shown in a separate study [169] that a dosimetric mode of hydrogen sensing can also be realized using cantilevers transducers. For this purpose, alpha platinum oxide was used as a coating that undergoes reduction and, therefore, irreversible mass loss in presence of hydrogen. Commercially available AFM cantilevers were coated with 20-50 nm platinum oxide films using reactive sputtering of platinum. Exposure of the sensor to 4% hydrogen in argon resulted in both static cantilever and resonance frequency changes. These changes reached saturation in about 30 minutes, thus indicating complete reduction of platinum oxide. Overall changes in the coating mass were estimated to be less than a nanogram. In addition to steady the state responses associated with irreversible chemical and physical changes in the coating, the observed transient features in the bending response were consistent with the expected thermal response due to the exothermic oxidation of hydrogen.

As inorganic coatings alone can not provide the selectivity patterns sought in many applications, modification of cantilevers with chemically selective organic layers has been a subject of more recent studied. One of the first cantilever sensors with organic coatings was a humidity sensor described by Thundat *et al*. [170]. Another design of a cantilever humidity sensor employed cantilevers with integrated piezoresistive read-out [97]. The design included both humidity sensitive and reference cantilevers as a part of a Wheatstone bridge. The layered silicon/silicon oxide cantilevers were 200 μm long, 50 μm wide 1.5 μm thick with deflection sensitivity, z^{-1} ($\Delta R/R$), of approximately 10^{-6} nm^{-1}. The active (humidity sensitive) cantilever was additionally coated with 10 μm photoresist. Swelling of the photoresist layer in presence of water vapor provided sensor responses that were nearly proportional to RH% in the range of 2 to 60%. The reference cantilever provided temperature compensation and could also be used for temperature measurements [97]. The reference channel is also useful to minimize effects of other noise sources. It should be noted, however, that the photoresist coating used in these studies was not highly specific to water vapor and responded to alcohol vapor as well [97]. In the case of ethanol, concentration levels as low as 10 ppm could be easily detected. Somewhat

different responsivities were obtained in the case of methanol, ethanol and 2-propanol.

In analogy to chemical sensors based on SAW transducers, cantilevers coated with various commercially available polymers were proposed for distinguishing between different volatile organic compounds (VOCs) in air. Lang *et al.* [30, 99] reported on a multi-cantilever sensor, in which signals were collected in a quasi-simultaneous (time-multiplexing) manner from eight individual cantilever transducers, each modified with a different coating [163, 171-173]. This design allowed the researchers to transfer the concept of an "electronic nose" from more conventional transduction principles [2] to innovative nanomechanical devices. Poly-methylmethacrylate (PMMA) as well as Pt metal coatings were used in some of these studies in order to demonstrate versatility of the cantilever arrays. Using a cantilever sensor with a PMMA coating, responses to a series of alcohols were obtained in both resonance and static deflection mode. Based on the differences in the shapes of response curves (either static deflection or resonance frequency change plotted as a function of time), the presence of different alcohols could be differentiated. In this case, the observed selectivity was primarily related to the fact that alcohols with different molecular weight and/or molecular structure have different diffusion rates in the PMMA coating. Therefore, the use of a multi-cantilever array with different polymeric coatings was the next logical step in developing a "chemical nose" based upon the cantilever platform. It was shown that cantilevers coated with several readily available "generic" polymers, such as polymethylmethacrylate, polystyrene, polyurethane and their blends or co-polymers, respond differently to various VOCs [161, 171]. By applying principal component or artificial neuron network analysis to response patterns from arrays of such polymer-modified cantilevers, the concept of an "electronic nose" was implemented.

Selectivity of cantilever gas sensors can also be controlled by coating the transducer surface with stationary phases developed for chemical separation applications, such as gas-chromatography (GC). Such coatings were previously used to impart chemical selectivity to sensors based on electroacoustic SAW transduces [85]. More recently, this approach has been extended to cantilever sensors [108]. In one study, thin films of commercially available GC polysilane phases, SP2340 (polycyano-phase) and OV25 (polymethyl-phenyl-phase) were deposited on one side of the Si cantilevers using a combination of spin-coating and FIB milling [108]. The thicknesses of both the polysilane film and the cantilever material were varied by adjusting the conditions of spin-coating and FIB milling. The selectivity pattern of polysilane coated Si cantilevers was assessed by measuring bending responses to a series of VOCs with different polarities. For 300 nm thick silicon cantilevers, maximal responses were observed with

*ca.*100 nm of the polysilane modifying coatings. Selectivity was shown to be consistent with a common GC phase classification scheme [108].

In order to create cantilever sensors with even more distinctive selectivity patterns with regard to different classes of VOCs, sol-gel coatings as well as covalently attached or evaporated films of synthetic receptors were found to be useful [10, 109, 174]. Thin films of sol-gels were formed on one side of 600 nm thick silicon cantilevers using aqueous solutions of organosilane precursors and spin coating procedures [109]. The cantilevers with sol-gel coatings exhibited strong bending in response to vapors of polar VOCs, in particular ethanol, while sensitivity to less polar compounds was relatively low. Additional chemical modification of the sol-gel coated cantilevers with a hydrophobic organosilane (hexamethyldisilazane) resulted in partial reversal of this trend and an 8-fold increase in the response to n-pentane (the least polar compound among the screened analytes). Despite these promising results, obtaining uniform coatings on cantilevers using spin coating procedures was challenging. Alternative methods of physical and chemical modification of microcantilever transducers for gas sensors were investigated by the authors of this review in the recent studies [174]. An ultrathin layer of thiol-modified receptors of the cyclodextrin family was formed preferentially on one side of gold coated cantilevers using self-assembling procedures similar to those used previously with other types of transducers with noble metal surfaces. The coatings of this type can be expected to provide the sensor specificity based on molecular recognition mechanisms as well as size exclusion effects. For instance, sensor responsivities varied for trichloroethylene, tetrachloroethylene and 2,7-dimethylnaphthalene [174]. Covalent attachment of the reactive receptor molecules onto nanostructured noble metal surfaces was found to be particularly suitable for creating highly efficient cantilever sensors. In the case of cantilevers with smooth gold surfaces, the magnitude of cantilever responses was only moderate. However, dramatic enhancements of responses were achieved when the same receptor layers were deposited on cantilevers with nanostructured (granular) gold films [115]. Using cantilever sensors with nanostructured surfaces, LODs as low as 0.17 ppb and 0.28 ppm were obtained for, respectively, 2,7-dimethylnaphtalene and tetrachloroethylene.

Microcantilever transducers were successfully used for detection of explosives present at extremely low concentrations [114, 175]. Pinnaduwage *et al.* reported detection of 10 to 30 ppt levels of pentaerythritol tetranitrate (PETN) and hexahydro-1,3,5-triazine (RDX) using commercially available cantilevers coated with a gold layer that was functionalized with a self-assembled monolayer of 4-mercaptobenzoic acid. The reported data [175] indicated the limit of detection of just a few femto-grams. Lavrik *et al.* [114] used trimaterial microcantilevers modified with a tert-butylcalix [6] arene coating and achieved large bending responses in presence of vapor phase

TNT and its analogs, mononitrotoluene (MNT) and 2,4-dinitrotoluene (2,4-DNT). Lavrik *et al.* estimated that the noise limited TNT detection threshold was 520 ppt [114].

Bimaterial cantilevers can also be used to detect local temperature changes associated with a chemical reaction that involves analyte molecules and is catalyzed by a catalyst on the cantilever surface. One of the first implementations of this detection scheme was reported by Barnes *et al.* [52]. By measuring deflections of a 1.5 μm thick Si cantilever with a 0.4 μm Al coating heat flux generated by a gas-phase catalytic reaction between O_2 and H_2 over a Pt surface was detected. For a standard AFM bi-material cantilever and AFM optical cantilever read-out, the limits of detection were estimated to be 1 pJ of thermal energy and 10^{-5} K of local temperature differences [52]. Even higher sensitivity of this method can be achieved using modified silicon or silicon nitride cantilevers with increased thermal isolation between their active regions (catalytic areas) and supporting bases ("heat sinks"). When heat escape through the surrounding environment becomes the principal path for the heat exchange, the sensitivity of cantilever based calorimetric sensors reaches its fundamental limit.

While practical applications of cantilever based calorimetric detectors that involve catalytic reactions can be somewhat limited, much more versatile detection and identification of species adsorbed on cantilevers can be achieved using photothermal spectroscopy methods [176-178]. For instance, we have shown that analytes present on cantilevers in a form of thin coatings (about 100 nm average thickness) can be detected in a calorimetric spectroscopy mode [177, 178]. Such cantilever based spectroscopic instruments offer excellent portability combined with the differentiating power inherent to vibrational spectroscopy.

In our very recent studies, a five-element microcantilever array was used to detect and differentiate vapor-phase analytes [179]. Four silicon microcantilevers in the array were coated with four different organic coatings and the fifth one was uncoated. Poly(diphenoxyphosphazene) (PDPP), hydrophobicaly derivatized β-cyclodextrin (HDA-β-CD), tetra-tertbutylcalix[6]arene (Cal-6) were deposited on nanostructured dealloyed gold (DA) [113] using physical vapor deposition in vacuum as described previously [180]. The fifth microcantilever in the array had DA coating only and was used as a reference. It was found that both the magnitude and the kinetics of the deflection responses were unique for each coating-analyte pair. Figure 11 shows typical data obtained in the case of HDA-β-CD and Cal-6 coatings and two analytes, tetrachloroethylene (TCE) and ethanol (EtOH). As can be seen in Figure 13a, the responses of each cantilever underwent different changes in magnitude and shape when the composition of the injected mixture was gradually varied from a 100% tetrachloroethylene (TCE) headspace sample to a 100% ethanol (EtOH) headspace sample. These observations can be explained by the fact that

analyte-induced changes in the responsive phase (generally swelling, although we have observed contraction as well) involve analyte adsorption, diffusion, dissolvation as well as subsequent molecular restructuring associated with these phenomena. Each of these factors depends on the analyte-coating pair and influence the response kinetics (i.e. the shape of the response curves). Figure 11b demonstrates a few possible ways of extracting response patterns from the acquired kinetic data.

We should point out that the static bending mode was used extensively in the majority of studies on cantilever sensors. However, resonating cantilevers scaled down to the nanoscale offer uniquely high mass sensitivity [63]. Resonating cantilever devices were used to measure femto-gram level mass sensitivity. This mass sensitivity is sufficient for single-cell detection (Figure 12). Lavrik *et al* [63] used cantilever with a resonance frequency of 2.2 MHz and measured a mass change of 5.5 fg due to chemisorption of 11-mercaptoundecanoic acid.

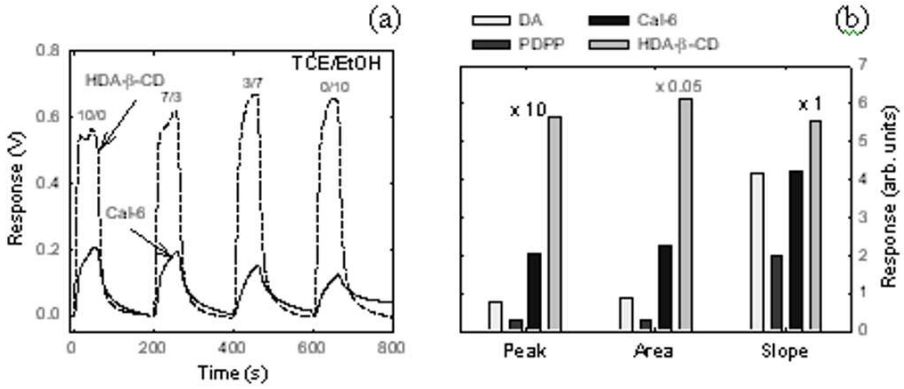


Figure 11. (a) Responses of cantilevers in an array to TCE/EtOH mixtures; (b) three of many possible methods (peak height, peak area, initial slope) of extracting distinctive data from these responses [114].

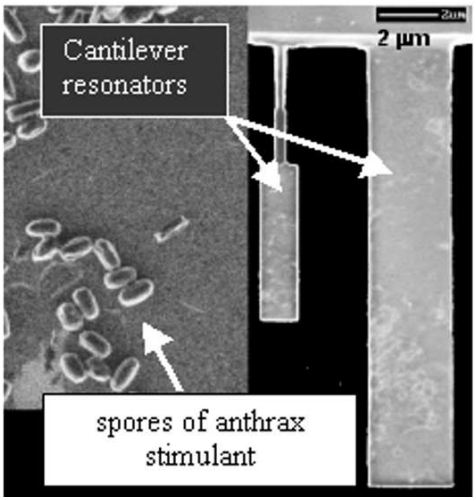


Figure 12. Resonating cantilever devices that provide femto-gram level mass sensitivity [63] detection.

5.2 Liquid Phase Analytes

Early works on cantilever based chemical detection in liquids involved standard AFM cantilevers and AFM heads for their read-out. Butt *et al.* [181] studied responses of 190 μm long 0.6 μm thick, gold coated, silicon nitride AFM cantilevers to various chemical factors and found that the steady state deflections depend upon both pH and ionic strength of the aqueous medium. The ionic strength was varied by adjusting concentration of KNO_3 varied from 0 to 1 M. The pH calibration plot acquired in 0.1M KNO_3 was represented by two roughly linear regions corresponding to pH of 2-5.5 and pH of 8-11, respectively. This broad calibration plot with a relatively flat region at pH of 5.5-8 is qualitatively consistent with existing models of silicon nitride surface dissociation [181]. The average slope of the calibration curve was approximately 9 nm per pH (at 0.1 M KNO_3). Further research in this direction focused on surface modifications of cantilevers that can lead to improved or modified pH sensitivity. Alkylthiols terminated with different chemical groups were most extensively used as modifying agents for gold-coated cantilevers. The other modification procedures involved silane-oxide chemistry and spontaneous oxidation of evaporated aluminum films. When pH responses of cantilevers modified with carboxylic acid, hydroxyl and amino groups were analyzed in several independent studies [78, 182], reasonable correlation between the experimental calibration plots and expected protonation-deprotonation behavior of the cantilever surfaces

was found. The reported pH responsivities varied from 15 to 50 nm/pH and were dependent on the surface treatment, cantilever type, and pH range.

Some of the most impressive figures of merit demonstrated with cantilever sensors are related detection of heavy metal ions. In particular, Ji *et al.* [183] reported on highly sensitive and selective detection of Cs^+ ions using a cantilever sensor with a self-assembled responsive layer of the molecular recognition type. The responsive layer for this sensor was formed using a newly synthesized receptor compound which combined calixarene and crown-ether macrocycles and had a reactive -SH group that provided its covalent attachment to gold surfaces. Using a microcantilever transducer with this responsive layer, Cs^+ ions could be detected in the concentration range of 10^{-11} to 10^{-7} M. The observed cantilever deflections in response to Cs^+ ions reached a steady state in less than 5 min. By modifying gold coated cantilever transducers with another self-assembled responsive monolayer, triethyl-12-mercaptododecylammonium bromide, a detector of trace amounts of CrO_4^{2-} was implemented [184]. As an extension of this approach, detection of trace levels of Ca^{2+} was also achieved using cantilever transducers modified with Ca^{2+} selective self-assembled responsive layers [185]. Two alternative chemical functionalization procedures were used in this study and resulted in the two kinds of self-assembled monolayers terminated with, respectively, with phosphate and N,N-diethyl-acetamide moieties.

Cantilevers modified with synthetic receptor compounds of the molecular recognition type were also found to be useful for detection of various neutral aromatic compounds in aqueous solutions [116, 180]. In particular, self-assembled monolayers of a thiolated beta- cyclodextrin derivative and evaporated thin films of heptakis (2,3-O-diacetyl-6-O-tertbutyl-dimethylsilyl)-beta- cyclodextrin were studied on smooth and nanostructured gold-coated microcantilever surfaces. In those studies, micro and nano structural modifications of microcantilevers chemical sensors were shown to improve the stress transduction between the chemical coating and the transducer. Structural modifications of microcantilever surfaces were achieved using either chemical dealloying [116] or focused ion beam milling [180]. The dealloyed surfaces [113, 115, 116, 174, 180] contained nanometer-sized features that enhanced the transduction of molecular recognition events into cantilever response, as well as increased coatings stability in the case of thicker films. The observed response factors for the analytes studied varied from 0.02-604 nm/ppm. Calibration plots obtained for 2,3-dihydroxynaphthalene and several volatile organic compounds revealed proportionality between the analyte concentrations and cantilever deflections in the range of up to several hundred nanometers. By manipulating surface morphologies and film thicknesses, improvements in the limits of detection of as great as 2 orders of magnitude were demonstrated [116, 180].

5.3 Biosensors

One of the first attempts to combine the biosensor concept and a cantilever transducer took advantage of the ultrahigh calorimetric sensitivity of a bimaterial microcantilever [10, 176]. By immobilizing glucose oxidize on the surface of 320-μm long, gold coated silicon nitride, Subramanian *et al.* created a glucose sensor that responded to presence of glucose in the aqueous medium due to the enzyme-induced exothermic processes [186]. This sensor exhibited a good linear calibration curve for glucose concentrations in the range of 5-40 mM.

Another indirect method of detecting biological species using micromachined cantilevers was proposed by Colton and co-workers [187, 188]. Their force amplified biological sensor (FABS) utilized a micromachined cantilever placed in a strong magnetic field. Similar to many conventional biological assays, such as the enzyme-linked immunosorbent assay, the FABS method relied on labeled biological material, however, magnetic beads rather than enzymes or fluorophores were used as a label. It was shown that an important advantage of the FABS method over existing bioassays is its capability to detect trace amounts of extremely dilute biological samples.

More recently, significant attention has been drawn to direct conversion of various biological receptor-ligand interactions into mechanical responses using cantilever transducers [31, 33, 79, 189-196]. Butt *et al.* explored high sensitivity of cantilever transducers to interfacial stress changes [189] in their work on a biosensor for a herbicide. The researchers reported bending responses of microfabricated cantilevers associated with interaction between the surface-immobilized herbicide (2,4-dichlorophenoxyacetic acid) and the monoclonal anti-herbicide antibody in an aqueous solution. The cantilever with immobilized herbicide exhibited partially reversible bending in response to the anti-herbicide antibodies at the concentrations of 5 and 25 μg/mL in a phosphate buffer saline (PBS).

Despite the excellent sensitivity of cantilever transducers and their capability to detect receptor-ligand interactions directly, the static deflection mode is not free from long-term drifts and instabilities inherent to other types of biosensors. In addition to temperature-induced drifts, it was also established that both specific binding and nonspecific adsorption of proteins on various surfaces are accompanied with very slow surface stress changes [192]. Moulin *et al.* [192] used microfabricated cantilevers to measure surface stress changes associated with nonspecific adsorption of immunoglobulin G (IgG) and bovine serum albumin (BSA) on gold surfaces. Compressive and tensile surface stress changes were observed upon adsorption of, respectively, IgG and BSA. It was also concluded that biological assays based on surface stress measurements are sensitive to subtle differences in preparation and purification of proteins that are otherwise identical and can

not be differentiated using other techniques [192]. Taking into account extremely high sensitivity of cantilever bending to interfacial biomolecular binding events, Moulin *et al*. [31] proposed a clinically relevant cantilever biosensor for differentiation of low density lipoproteins (LPL) and their oxidized form (oxLDL). For this purpose, silicon nitride cantilevers with freshly evaporated gold were modified with heparin. The modification procedure included successive incubation of the cantilevers in 2-aminoethanethiol hydrochloride and heparin solutions, and saturation of nonspecific binding sites with BSA (each incubation was followed by rinsing in purified water). The adsorption induced surface stress changes measured using the cantilever biosensor were notably slower in comparison to the binding kinetics observed in the control surface plasmon resonance measurements.

A significant milestone in developing cantilever based biosensors was demonstration of their applicability to DNA analysis [79, 193, 195, 196]. Fritz *et al.* reported on sensitive and specific monitoring of oligonucleotide hybridization using arrays of functionalized cantilevers and optical read-out of their deflections [79]. A thin layer of gold on one side of the cantilevers permitted controllable immobilization of thio-modified oligonucleotides. When 12-mer oligo-nucleotides with different degree of complementarity were used in the hybridization assay, a single base pair mismatch was clearly detectable. The use of a differential pair of cantilever transducers, *i.e.* functionalized and "blank", and analysis of the differential deflections was an important refinement that minimized interfering effects of temperature, mechanical vibrations and fluid flow in the cell and, therefore, provided more reliable differentiation of the responses that accompanied specific biomolecular interactions. In addition to oligonucleotide hybridization assays, differential cantilever deflection method was shown to be very promising for monitoring a wide range of biological affinity interactions. In particular, irreversible differential responses were observed in the course of protein A-IgG interactions [31]. In a similar study reported by Raiteri *et al.* [190], 85 ng/mL myoglobin in an aqueous solution was detected using a differential pair of cantilevers, one of which was functionalized with monoclonal anti-myoglobin antibodies.

Under carefully controlled experimental conditions (temperature, pH, ionic strength, etc.), even a single cantilever transducer provides a sensitive means for detection of various biomolecular interactions. For instance, Thundat and coworkers differentiated a single-nucleotide mismatch using a cantilever transducer placed in a thermally stabilized flow cell [193, 195]. Both the rate and the magnitude of deflection response were dependent on the length and sequence mismatch of the analyzed oligonucleotide. The same group of researchers also reported on detection of ultra-low concentrations (0.2 ng/ml) of prostate specific antigen (PSA) using a similar

thermoelectrically stabilized cell housing a single cantilever transducer [195].

Biotin-streptavidin is another example of high-affinity biomolecular interactions that was successfully monitored using cantilever transducers. Raiteri *et al.* used biotin functionalized silicon nitride cantilevers and measured their deflection responses in presence of 100 nM streptavidin [190]. These responses that reached approximately 50 nm in magnitude within 10 min were largely reversible. In the case of high-affinity streptavidin-biotin interactions, the reversible nature of the responses is especially unexpected and apparently indicates a nontrivial relationship between the surface coverage of streptavidin molecules on the cantilever surface and the associated surface stress change.

It is important to note that the deflection responses of cantilever based biosensors rarely exceed 100 nm regardless of the chosen biological affinity system. This implies that signal-to-noise ratios and sensitivities achievable with conventionally designed cantilever biosensors may be limited by common fundamental mechanisms. A question arises as to whether such limitations can be surmounted. Recent studies conduced by the authors of this review addressed this question by exploring cantilevers with asymmetrically nanostructured surfaces and demonstrated feasibility of cantilever biosensors with dramatically enhanced responsivities. It was found that interfacial biomolecular recognition events were converted into mechanical responses much more efficiently when high density of nano-size features were present on one side of the cantilever transducers. Creating of such interfaces may rely on surface immobilization of gold nanospheres or dealloying of co-evaporated Au:Ag films. The most efficient transduction was achieved when the cantilevers were modified with 50 to 75 nm thick dealloyed Au:Ag films. Unlike conventional cantilevers with smooth surfaces, these nanostructured transducers exhibited up to several micron deflections upon adsorption of protein A and biotin-labeled albumin on nanostructured gold surfaces. Additional micrometer scale deflections of the cantilevers were observed upon interaction of the surface immobilized receptors with, respectively, immunoglobulin G and avidin [10].

A promising approach to ultrasensitive detection of biological species in air was demonstrated by Ilic *et al.* using smaller cantilevers operating in the resonance mode [83, 197]. The silicon nitride cantilevers used in these studies were about 5 μm long, had resonance frequencies in the MHz range and permitted detection of minute amounts of biological molecules or cells. Studies by Ilic *et al.* [84] have shown that the mass sensitivity of a cantilever sensor was sufficient to detect mass changes due to attachment of a single microbial cell while modification of the cantilever surface with appropriate biological receptors, for instance antibodies, provided high detection specificity.

6. CONCLUSIONS

Cantilever transducers are recognized as promising platform for the next generation of chemical and biological sensors. It is anticipated that microfabricated cantilevers can provide a versatile platform for real-time, in-situ measurements of physical, chemical, and biochemical properties of physiological fluids. In general, the MEMS platform offers an unparalleled capability for the development and mass production of extremely sensitive, low-cost sensors suitable for rapid analysis of many chemical and biological species. Compared with more conventional sensors, cantilever sensors offer improved dynamic response, greatly reduced size, high precision, and increased reliability. A plethora of physical, chemical, and biological sensors based on the micromachined cantilever transducers have already been demonstrated [7, 30, 34, 52, 55, 95, 99, 116, 117, 163, 176].

An important advantage of microcantilever sensors is that they can operate in vacuum, gases, and liquids. A compelling feature of the cantilever-based sensors operating in the resonant mode is that four response parameters (resonance frequency, phase, amplitude, and *Q*-factor, measured simultaneously) may provide complementary information about the interactions between the sensor and the environment. The damping effects of a liquid medium, however, reduce resonance responses of cantilever devices. In most common liquids, such as aqueous solutions, the amplitude of the cantilever oscillations at the resonance can be orders of magnitude lower as compared to the same resonating cantilever operating in air. On the other hand, operation in the static mode is unaffected by viscous properties of the medium. Therefore, microcantilever sensors operating in the static mode are especially attractive as a platform for nanomechanical biochemical assays and other biomedical applications.

Micro- and nano-scale cantilevers have extremely small thermal masses and can be heated and cooled with thermal time constants of less than a millisecond. This is advantageous for rapid reversal of molecular absorption processes and regeneration purposes. Both static and dynamic responses of cantilever sensors can be measured with very high precision using several readout techniques [35] based on optical beam deflection, interferometry, electron transfer, piezoresistance, capacitance, and piezoelectric properties. Other resonance parameters of cantilever transducers, such as amplitude and *Q*-factor, can be extracted from these responses measured in an appropriate frequency or time domain. In many instances, deformations and resonance frequency shifts measured simultaneously may provide complementary information about the interactions between the transducers and the environment.

Cantilevers operating in the static mode often surpass in performance cantilevers operating in the resonant operation. However, the mass sensitivity of cantilever transducers operating in the resonance mode

increases as their dimensions are reduced. Therefore, cantilever sensors with progressively increased mass sensitivity can be fabricated by simply reducing the transducer dimensions [63, 126, 198, 199]. As the technology of nanosize mechanical structures advances, nanomechanical devices approach the GHz frequency domain that is already widely used in electronic data processing and optical telecommunication.

In order to achieve fundamentally limited performance of resonating cantilever sensors, it is necessary to delineate the noise present in micromechanical systems. The type of noise in cantilever devices referred to as "thermo-mechanical" [35, 123] or "mechanical-thermal noise" [200, 201] arises from the dynamic equilibrium between mechanical energy of the device and thermal energy of the surrounding environment at nonzero temperatures. This type of noise imposes the ultimate limits on the performance of microcantilever sensors operating in the resonance mode.

The advantages of cantilever sensors can be further expanded by arranging individual cantilever transducers into large multi-sensor arrays integrated with on-chip electronic circuitry [202]. One-dimensional and two-dimensional arrays of cantilever transducers offer additional advantages that cannot be overlooked. In particular, such arrays provide a viable platform for the development of high-performance "electronic noses".

Acknowledgments: We like to acknowledge support from the U.S. Department of Energy, the Defense Advanced Research Projects Agency, and the National Science Foundation and. This work was performed in part at the Cornell Nanofabrication Facility (a member of the National Nanofabrication Users Network) which is supported by the National Science Foundation under Grant ECS-9731293, Cornell University and industrial affiliates. This work was partially supported by the Laboratory Director's Research and Development Program of Oak Ridge National Laboratory. Oak Ridge National Laboratory is operated for the U.S. Department of Energy by UT-Battelle under contract DE-AC05-96OR22464.

7. REFERENCES

[1] Janata J., *Principles of Chemical sensors.* New York: Plenum, 1989.
[2] Janata J., Josowicz M., Devaney D. M. Chemical sensors. Anal. Chem 1994; 66: R207.
[3] Janata J. Chemical sensors. Anal. Chem. 1990; 62: 33R-44R.
[4] Kovacs G. T. A., *Micromachined Transducers.* New York: McGraw-Hill, 1998.
[5] Berger R., Lang H. P., Gerber C., Gimzewski J. K., Fabian J. H., Scandella L., Meyer E., Guntherodt H.-J. Micromechanical Thermogravimetry. Chemical Physics Letters 1998; 294: 363.
[6] Wachter E. A., Thundat T. Micromechanical Sensors for Chemical and Physical Sensing. Rev. Sci. Instrum 1995; 66: 3662.

[7] Ferrari V., Marioli D., Taroni A., Ranucci E., Ferruti P. Development and Application of Mass Sensors Based on Flexural Resonances in Alumina Beams. IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control 1996; 43: 601.
[8] Moulin A. M., Stephenson R. J., Welland M. E. Micromechanical Thermal Sensors: Comparison of Experimental Results and Simulations. J. Vac. Sci. Technol. B 1997; 15: 590-596.
[9] Tuller H. L., Mlcak R. Inorganic Sensors Utilizing MEMS and Microelectronic Technologies. Curr. Opin. Solid State Mat. Sci. 1998; 3: 501.
[10] Datskos P. G., Sepaniak M. J., Tipple C. A., Lavrik N. Photomechanical Chemical Microsensors. Sens. Actuator B-Chem. 2001; 76: 393.
[11] Sepaniak M., Datskos P., Lavrik N., Tipple C. Microcantilever Transducers: A New Approach in Sensor Technology. Analytical Chemistry 2002; 74: 568A-575A.
[12] Binnig G., Quate C. F., Gerber C. Atomic Force Microscope. Phys. Rev. Lett. 1986; 56: 930-933.
[13] Timoshenko S. P. Analysis of Bi-Metal Thermostats. J. Opt. Soc. Am. 1925; 11: 233-255.
[14] Norton F. J. Gas Analyzer US patent 2,307,800, 1943
[15] Shaver P. J. Bimetal Strip Hydrogen Gas Detectors. Rev. Sci. Instrum. 1969; 40: 901-&.
[16] Newell W. E. Miniaturization of Tuning Forks. Science 1968; 161: 1320-&.
[17] Steinberg I., Oplatka A., Katchalsky A. Mechanochemical Engines. Nature 1966; 210: 568.
[18] Katchalsky A., Steinberg I., Oplatka A., Kam A.Mechanochemical Engine US Patent 3,321,908, 1967
[19] Palmer W. G. The Use of the Coherer to Investigate Adsorption Films. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 1924; 106: 55-68.
[20] Sears J. E. The Principles of Fine Measurement. The Scientific Monthly 1925; 20: 427-435.
[21] Meehan F. T. The Expansion of Charcoal on Sorption of Carbin Dioxide. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 1927; 115: 199.
[22] Bangham D. H., Fakhoury N. The Swelling of Charcoal. Part I. Preliminary Experiments with Water Vapour, Carbon Dioxide, Ammonia, and Sulphur Dioxide. Proceedings of the Royal Society of London. Series A 1930; 130: 81.
[23] Jones R. V., McCombie C. W. Brownian Fluctuations in Galvanometers and Galvanometer Amplifiers. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 1952; 244: 205-230.
[24] Nathanson H. C., Newell W. E., Wickstrom R. A. Tuning Forks' Sound a Hopwful Note. Electronics 1965; 38: 84.
[25] Nonnenmacher M., Greschner J., Wolter O., Kassing R. Scanning Force Microscopy with Micromachined Silicon Sensors. Journal of Vacuum Science & Technology B 1991; 9: 1358-1362.
[26] Walters D. A., Cleveland J. P., Thomson N. H., Hansma P. K., Wendman M. A., Gurley G., Elings V. Short Cantilevers for Atomic Force Microscopy. Rev. Sci. Instrum. 1996; 67: 3583-3590.
[27] Cleland A. N., Roukes M. L. Fabrication of High Frequency Nanometer Scale Mechanical Resonators from Bulk Si Crystals. Applied Physics Letters 1996; 69: 2653-2655.
[28] Giles R., Cleveland J. P., Manne S., Hansma P. K., Drake B., Maivald P., Boles C., Gurley J., Elings V. Noncontact Force Microscopy in Liquids. Applied Physics Letters 1993; 63: 617-618.
[29] Thundat T., Oden P. I., Warmack R. J. Microcantilever Sensors. Microscale Thermophysical Engineering 1997; 1: 185-199.

[30] Berger R., Gerber C., Lang H. P., Gimzewski J. K. Micromechanics: A Toolbox for Femtoscale Science: "Towards a Laboratory on a Tip". Microelectron. Eng. 1997; 35: 373-379.
[31] Moulin A. M., O'Shea S. J., Welland M. E. Microcantilever-Based Biosensors. Ultramicroscopy 2000; 82: 23-31.
[32] Craighead H. G. Nanoelectromechanical Systems. Science 2000; 290: 1532-1535.
[33] Pereira R. D. Atomic Force Microscopy as a Novel Pharmacological Tool. Biochem. Pharmacol. 2001; 62: 975-983.
[34] Thundat T., Oden P. I., Datskos P. G., Chen G. Y., Warmack R. J. Microcantilever Sensors. The 16th Warner Brandt Workshop on Charged Particle Penetration Phenomena, ed. Oak Ridge Tennessee, 1996.
[35] Sarid D., *Scanning Force Microscopy.* New York: Oxford University Press, 1991.
[36] Butt H. J. A Sensitive Method to Measure Changes in the Surface Stress of Solids. J. Colloid Interface Sci. 1996; 180: 251-260.
[37] Samuel J., Brinker C. J., Frink L. J. D., van Swol F. Direct Measurement of Solvation Forces in Complex Microporous Media: A New Characterization Tool. Langmuir 1998; 14: 2602-2605.
[38] Frink L. J. D., van Swol F. A Common Theoretical Basis for Surface Forces Apparatus, Osmotic Stress, and Beam Bending Measurements of Surface Forces. Colloids & Surfaces A Physicochemical & Engineering Aspects 2000; 162: 25-36.
[39] Sader J. E. Surface Stress Induced Deflections of Cantilever Plates with Applications to the Atomic Force Microscope: V-Shaped Plates. Journal of Applied Physics 2002; 91: 9354-9361.
[40] Sader J. E. Surface Stress Induced Deflections of Cantilever Plates with Applications to the Atomic Force Microscope: Rectangular Plates. Journal of Applied Physics 2001; 89: 2911-2921.
[41] Godin M., Tabard-Cossa V., Grutter P., Williams P. Quantitative Surface Stress Measurements Using a Microcantilever. Applied Physics Letters 2001; 79: 551.
[42] Raiteri R., Butt H. J., Grattarola M. Changes in Surface Stress at the Liquid/Solid Interface Measured with a Microcantilever. Electrochim. Acta 2000; 46: 157-163.
[43] Bruland K. J., Garbini J. L., Dougherty W. M., Sidles J. A. Optimal Control of Ultrasoft Cantilevers for Force Microscopy. Journal of Applied Physics 1998; 83: 3972-3977.
[44] Rugar D., Yannoni C. S., Sidles J. A. Mechanical Detection of Magnetic-Resonance. Nature 1992; 360: 563-566.
[45] Mamin H. J., Rugar D. Sub-Attonewton Force Detection at Millikelvin Temperatures. Applied Physics Letters 2001; 79: 3358-3360.
[46] Streckeisen P., Rast S., Wattinger C., Meyer E., Vettiger P., Gerber C., Guntherodt H. J. Instrumental Aspects of Magnetic Resonance Force Microscopy. Appl. Phys. A-Mater. Sci. Process. 1998; 66: S341-S344.
[47] Cleland A. N., Roukes M. L. A Nanometre-Scale Mechanical Electrometer. Nature 1998; 392: 160-162.
[48] Stephan A. C., Gaulden T., Brown A. D., Smith M., Miller L. F., Thundat T. Microcantilever Charged-Particle Flux Detector. Rev. Sci. Instrum. 2002; 73: 36-41.
[49] Erbe A., Blick R. H., Tilke A., Kriele A., Kotthaus J. P. A Mechanically Flexible Tunneling Contact Operating at Radio Frequencies. Applied Physics Letters 1998; 73: 3751-3753.
[50] Scheible D. V., Erbe A., Blick R. H. Tunable Coupled Nanomechanical Resonators for Single-Electron Transport. New Journal of Physics 2002; 4: 86.1.
[51] Datskos P. G., Thundat T. Nanocantilever Signal Transduction by Electron Transfer. J. Nanosci. Nanotechnol. 2002; 2: 369-373.
[52] Barnes J. R., Stephenson R. J., Welland M. E., Gerber C., Gimzewski J. K. Photothermal Spectroscopy with Femtojoule Sensitivity Using a Micromechanical Device. Nature 1994; 372: 79-81.

[53] Berger R., Gerber C., Gimzewski J. K., Meyer E., Guntherodt H. J. Thermal Analysis Using a Micromechanical Calorimeter. Applied Physics Letters 1996; 69: 40-42.
[54] Oden P. I., Datskos P. G., Thundat T., Warmack R. J. Uncooled Thermal Imaging Using a Piezoresistive Microcantilever. Applied Physics Letters 1996; 69: 3277-3279.
[55] Wachter E. A., Thundat T., Oden P. I., Warmack R. J., Datskos P. G., Sharp S. L. Remote Optical Detection Using Microcantilevers. Rev. Sci. Instrum. 1996; 67: 3434-3439.
[56] Datskos P. G., Rajic S., Datskou I. Photoinduced and Thermal Stress in Silicon Microcantilevers. Applied Physics Letters 1998; 73: 2319.
[57] Amantea R., Goodman L. A., Pantuso F., Sauer D. J., Varhese M., Villianni T. S., White L. K. Progress Towards an Uncooled IR Imager with 5 mK NETD. Infrared Technology and Applications XXIV 1998; 3436: 647.
[58] Perazzo T., Mao M., Kwon O., Majumdar A., Varesi J. B., Norton P. Infrared Vision Using Uncooled Micro-Optomechanical Camera. Applied Physics Letters 1999; 74: 3567-3569.
[59] Mao M., Perazzo T., Kwon O., Zhao Y., Majumdar A., Varesi J., Norton P. Infrared Vision Using an Uncooled Thermo-Opto-Mechanical Camera; Design, Microfabrication, and Performance. Microelectromechanical Systems, Lee, Y. C., K. Goodson, R.S. Keynton, A. Lee, L. Lin, F.K. Forster, ed. Nashville, TN, The American Society of Mechanical Engineers, 1999.
[60] Datskos P. G., Rajic S., Datskou I. Detection of Infrared Photons Using the Electronic Stress in Metal-Semiconductor Cantilever Interfaces. Ultramicroscopy 2000; 82: 49-56.
[61] Senesac L. R., Corbeil J. L., Lavrik N. V., Rajic S., Datskos P. G. IR Imaging Using Uncooled Cantilever Sensors. Ultramicroscopy 2003; 97: 451.
[62] Knobel R. G., Cleland A. N. Nanometre-Scale Displacement Sensing Using a Single Electron Transistor. Nature 2003; 424: 291.
[63] Lavrik N. V., Datskos P. G. Femtogram Mass Detection Using Photothermally Actuated Nanomechanical Resonators. Applied Physics Letters 2003; 82.
[64] Yates D. J. C. The Expansion of Porous Glass on the Adsorption of Non-Polar Gases. Proceedings of the Royal Society of London. Series A 1954; 224: 526.
[65] Sussman M. V., Katchalsky A. Mechanochemical Turbine: A New Power Cycle. Science 1970; 167: 45.
[66] Dye D. The Valve-Maintained Tuning-Fork as a Precision Time-Standard. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 1923; 103: 240-260.
[67] Taylor E. H., Waggener W. C. Measurement of Adsorption Forces. J. Phys.Chem. 1979; 83: 1361.
[68] Baselt D. R., Fruhberger B., Klaassen E., Cemalovic S., C.L. Britton, S.V. Patel, Mlsna T. E., McCorkle D., Warmack B. Design and Performance of a Microcantilever-Based Hydrogen Sensor. Sensors and Actuators B-Chemical 2003; 88: 120.
[69] Shuttleworth R. Proc. Phys. Soc. London 1950; 63A: 444.
[70] Jones R. V. The Detection of Thermal Radiation Using Linear Expansion. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 1959; 249: 100.
[71] Burov J., Hinkov V. Photoelectrical Resonant Gate Transistor (Prgt). Optical and Quantum Electronics 1975; 7: 474-480.
[72] Borisov M. I., Hinkov V. P., Burov J. I., Strashilov V., Bransalov K. P. Resonant Gate Transistor with an Active Vibrator (RGTAV). International Journal of Electronics 1975; 39: 377-384.
[73] Nathanson H. C., Newell W. E., Wickstrom R. A., Davis J. R. Resonant Gate Transistor. IEEE Trans. Electron Devices 1967; ED14: 117-&.
[74] Newell W. E. Yuned Integrated Circuits. Proc. IEEE 1967; 52: 1603.

[75] Newell W. E., Wickstrom R. A., Page D. J. Tunistors - Mechanical Resonators for Microcircuits. Proceedings of the IEEE International Meeting on Electron Devices 1967: 24.
[76] Newell W. E., Wickstrom R. A., Page D. J. Tunistors-Mechanical Resonators for Microcircuits. IEEE Trans. Electron Devices 1968; ED15: 411-&.
[77] Newell W. E., Wickstrom R. A. Tunistor - a Mechanical Resonator for Microcircuits. IEEE Trans. Electron Devices 1969; ED16: 781-&.
[78] Fritz J., Baller M. K., Lang H. P., Strunz T., Meyer E., Guntherodt H. J., Delamarche E., Gerber C., Gimzewski J. K. Stress at the Solid-Liquid Interface of Self-Assembled Monolayers on Gold Investigated with a Nanomechanical Sensor. Langmuir 2001; 16: 9694.
[79] Fritz J., Baller M. K., Lang H. P., Rothuizen H., Vettiger P., Meyer E., Guntherodt H. J., Gerber C., Gimzewski J. K. Translating Biomolecular Recognition into Nanomechanics. Science 2000; 288: 316-318.
[80] Datskos P. G., Egert C. M., Sauers I. 2-Mercaptoethanol Vapor Sensing with Microcantilevers. Sensors and Actuators B 2000; 61: 75.
[81] Berger R., Delamarche E., Lang H. P., Gerber C., Gimzewski J. K., Meyer E., Guntherodt H. J. Surface Stress in the Self-Assembly of Alkanethiols on Gold Probed by a Force Microscopy Technique. Applied Physics a (Materials Science Processing) 1998; 66: S 55-S 59.
[82] Oden P. I. Gravimetric Sensing of Metallic Deposits Using an End-Loaded Icrofabricated Beam Structure. Sensors and Actuators B 1998; 53: 191.
[83] Ilic B., Czaplewski D., Zalalutdinov M., Craighead H. G., Neuzil P., Campagnolo C., Batt C. Single Cell Detection with Micromechanical Oscillators. Journal of Vacuum Science & Technology B 2001; 19: 2825-2828.
[84] Nathanson H. C., Wickstrom R. A. A Resonant Gate Transistor with High-Q Bandpass Properties. Appl. Phys. Lett. 1965; 7: 84.
[85] Grate J. W., Klusky M. Surface Acoustic-Wave Vapor Sensors Based on Resonator Devices. Anal. Chem. 1991; 63: 1719.
[86] Grate J. W., Wenzel S. W., White R. M. Flexural Plate Wave Devices for Chemical-Analysis. Anal. Chem. 1991; 63: 1552.
[87] Grate J. W., Martin S. J., White R. M. Acoustic Wave Microsensors Part I. Anal. Chem. 1993; 65: 940A.
[88] Grate J. W., Martin S. J., White R. M. Acoustic-Wave Microsensors Part II. Anal. Chem. 1993; 65: A987.
[89] Grate J. W., McGill R. A. Dewetting Effects on Polymer-Coated Surface-Acoustic-Wave Vapor Sensors. Anal. Chem. 1995; 67: 4015.
[90] Grate J. W., *Surface-Acoustic Wave Array Detectors.* ed. 1995.
[91] Martinez R. E., Augustyniak W. M., Golovchenko J. A. Direct Measurement of Crystal-Surface Stress. Phys. Rev. Lett. 1990; 64: 1035-1038.
[92] Ibach H. Adsorbate-Induced Surface Stress. J. Vac. Sci. Technol. A-Vac. Surf. Films 1994; 12: 2240-2243.
[93] Stoney G. G. The Tension of Metallic Films Deposited by Electrolysis. Proceedings of the Royal Society of London, A 1909; 82: 172-177.
[94] Timoshenko S., *Theory of Plates and Shells.* McGraw-Hill, 1940.
[95] Thundat T., Wachter E. A., Sharp S. L., Warmack R. J. Detection of Mercury Vapor Using Resonating Microcantilevers. Applied Physics Letters 1995; 66: 1695-1697.
[96] Brugger J., Beljakovic G., Despont M., deRooij N. F., Vettiger P. Silicon Micro/Nanomechanical Device Fabrication Based on Focused Ion Beam Surface Modification and Koh Etching. Microelectron. Eng. 1997; 35: 401-404.
[97] Boisen A., Thaysen J., Jensenius H., Hansen O. Environmental Sensors Based on Micromachined Cantilevers with Integrated Read-Out. Ultramicroscopy 2000; 82: 11-16.

[98] Britton C. L., Jones R. L., Oden P. I., Hu Z., Warmack R. J., Smith S. F., Bryan W. L., Rochelle J. M. Multiple-Input Microcantilever Sensors. Ultramicroscopy 2000; 82: 17-21.
[99] Baller M. K., Lang H. P., Fritz J., Gerber C., Gimzewski J. K., Drechsler U., Rothuizen H., Despont M., Vettiger P., Battiston F. M., Ramseyer J. P., Fornaro P., Meyer E., Guntherodt H. J. A Cantilever Array-Based Artificial Nose. Ultramicroscopy 2000; 82: 1-9.
[100] Dahmen K., Lehwald S., Ibach H. Bending of Crystalline Plates under the Influence of Surface Stress - a Finite Element Analysis. Surf. Sci. 2000; 446: 161-173.
[101] von Preissig F. J. Applicability of the Classical Curvature-Stress Relation for Thin Films on Plate Substrates. J. Appl. Phys. 1989; 66: 4262.
[102] Miyatani T., Fujihira M. Calibration of Surface Stress Measurements with Atomic Force Microscopy. Journal of Applied Physics 1997; 81: 7099-7115.
[103] Fang W. L. Determination of the Elastic Modulus of Thin Film Materials Using Self-Deformed Micromachined Cantilevers. J. Micromech. Microeng. 1999; 9: 230-235.
[104] Hu Z. Y., Thundat T., Warmack R. J. Investigation of Adsorption and Absorption-Induced Stresses Using Microcantilever Sensors. Journal of Applied Physics 2001; 90: 427-431.
[105] Koch R. The Intrinsic Stress of Polycrystalline and Epitaxial Thin Metal-Films. J. Phys.-Condes. Matter 1994; 6: 9519-9550.
[106] Koch R. Intrinsic Stress of Ultrathin Epitaxial Films. Appl. Phys. A-Mater. Sci. Process. 1999; 69: 529-536.
[107] Israelachvili J., *Intermolecular and Surface Forces.* San Diego: Academic Press, 1991.
[108] Betts T. A., Tipple C. A., Sepaniak M. J., Datskos P. G. Selectivity of Chemical sensors Based on Micro-Cantilevers Coated with Thin Polymer Films. Analytica Chimica Acta 2000; 422: 89-99.
[109] Fagan B. C., Tipple C. A., Xue Z. L., Sepaniak M. J., Datskos P. G. Modification of Micro-Cantilever Sensors with Sol-Gels to Enhance Performance and Immobilize Chemically Selective Phases. Talanta 2000; 53: 599-608.
[110] Raguse B., Muller K. H., Wieczorek L. Nanoparticle Actuators. Advanced Materials 2003; 15: 922.
[111] Kaneko D., Gong J. P., Osada Y. Polymer Gels as Soft and Wet Chemomechanical Systems - an Approach to Artificial Muscles. Journal of Materials Chemistry 2002; 12: 2169-+.
[112] Baughman R. H., Cui C. X., Zakhidov A. A., Iqbal Z., Barisci J. N., Spinks G. M., Wallace G. G., Mazzoldi A., De Rossi D., Rinzler A. G., Jaschinski O., Roth S., Kertesz M. Carbon Nanotube Actuators. Science 1999; 284: 1340-1344.
[113] Lavrik N. V., Tipple C. A., Sepaniak M. J., Datskos P. G. Enhanced Chemi-Mechanical Transduction at Nanostructured Interfaces. Chemical Physics Letters 2001; 336: 371.
[114] Lavrik N. V., Sepaniak M. J., Datskos P. G. Detection of Explosive Compounds Using Microcantilevers with Nanoporous Coatings. Sensors Letters 2003; 1: 25-32.
[115] Lavrik N. V., Tipple C. A., Sepaniak M. J., Datskos P. G. Gold Nano-Structures for Transduction of Biomolecular Interactions into Micrometer Scale Movements. Biomedical Microdevices 2001; 3: 33-44.
[116] Tipple C. A., Lavrik N. V., Culha M., Headrick J., Datskos P., Sepaniak M. J. Nanostructured Microcantilevers with Functionalized Cyclodextrin Receptor Phases: Self-Assembled Monolayers and Vapor-Deposited Films. Analytical Chemistry 2002; 74: 3118-3126.
[117] Davis Z. J., Abadal G., Kuhn O., Hansen O., Grey F., Boisen A. Fabrication and Characterization of Nanoresonating Devices for Mass Detection. Journal of Vacuum Science & Technology B 2000; 18: 612-616.
[118] Datskos P. G., Sauers I. Detection of 2-Mercaptoethanol Using Gold-Coated Micromachined Cantilevers. Sensors & Actuators B Chemical 1999; 61: 75-82.

[119] Martin Y., Williams C. C., Wickramasinghe H. K. Atomic Force Microscope–Force Mapping and Profiling on a Sub 100-Å Scale. Journal of Applied Physics 1987; 61: 4723.
[120] Blom F. R., Bouwstra S., Elwenspoek M., Fluitman J. H. J. Dependence of the Quality Factor of Micromachined Silicon Beam Resonators on Pressure and Geometry. Journal of Vacuum Science & Technology B 1992; 10: 19-26.
[121] Majorana E., Ogawa Y. Mechanical Thermal Noise in Coupled Oscillators. Physics Letters 1997; A233.
[122] Butt H. J., Jaschke M. Calculation of Thermal Noise in Atomic-Force Microscopy. Nanotechnology 1995; 6: 1-7.
[123] Cleland A. N., Roukes M. L. Noise Processes in Nanomechanical Resonators. Journal of Applied Physics 2002; 92: 2758-2769.
[124] Djuric Z. Mechanisms of Noise Sources in Microelectromechanical Systems. Microelectron. Reliab. 2000; 40: 919-932.
[125] Yasumura K. Y., Stowe T. D., Chow E. M., Pfafman T., Kenny T. W., Stipe B. C., Rugar D. Quality Factors in Micron- and Submicron-Thick Cantilevers. J. Microelectromech. Syst. 2000; 9: 117-125.
[126] Kawakatsu H., Kawai S., Saya D., Nagashio M., Kobayashi D., Toshiyoshi H., Fujita H. Towards Atomic Force Microscopy up to 100 Mhz. Rev. Sci. Instrum. 2002; 73: 2317-2320.
[127] Chen G. Y., Warmack R. J., Thundat T., Allison D. P., Huang A. Resonance Response of Scanning Force Microscopy Cantilevers. Review of Scientific Instruments 1994; 65: 2532.
[128] Bishop J. E., Kinra V. K. Elastothermodynamic Damping in Laminated Composites. International Journal of Solids and Structures 1997; 34.
[129] Buser R. A., de Rooij N. F. Tuning Forks in Silicon. IEEE Workshop Micro Electro Mechanical Systems, ed. Salt Lake City, UT, USA, 1989.
[130] Kleiman R. N., Kaminsky G. K., Reppy J. D., Pindak R., Bishop D. J. Single-Crystal Silicon High-Q Torsional Oscillators. Review of Scientific Instruments 1985; 56: 2088-2091.
[131] Mihailovich R. E., MacDonald N. C. Dissipation Measurements of Vacuum-Operated Single-Crystal Silicon Microresonators. Sensors and Actuators a-Physical 1995; 50: 199-207.
[132] Mihailovich R. E., Parpia J. M. Low Temperature Mechanical Properties of Boron-Doped Silicon. Physical Review Letters 1992; 68: 3052.
[133] Houston B. H., Photiadis D. M., Marcus M. H., Bucaro J. A., Xiao Liu, Vignola a. J. F. Thermoelastic Loss in Microscale Oscillators. Applied Physics Letters 2002; 80: 1300.
[134] Lifshitz R., Roukes M. L. Thermoelastic Damping in Micro- and Nanomechanical Systems. Physical Review B 2000; 61: 5600.
[135] Lifshitz R. Phonon-Mediated Dissipation in Micro- and Nano-Mechanical Systems. Physica B-Condensed Matter 2002; 316: 397-399.
[136] Milanovich F. P., *Multi-Analyte, Single Fiber, Optical Sensors.* ed. 1995.
[137] Tuominen M. T., Krotkov R. V., Breuer M. L. Stepwise and Hysteretic Transport Behavior of an Electromechanical Charge Shuttle. Phys. Rev. Lett. 1999; 83: 3025.
[138] Rugar D., Mamin H. J., Guethner P. Improved Fiber-Optic Interferometer for Atomic Force Microscopy. Applied Physics Letters 1989; 55: 2588-2590.
[139] Meyer G., Amer N. M. Novel Optical Approach to Atomic Force Microscopy. Applied Physics Letters 1988; 53: 1045-1047.
[140] Hoummady M., Farnault E., Yahiro T., Kawakatsu H. Simultaneous Optical Detection Techniques, Interferometry, and Optical Beam Deflection for Dynamic Mode Control of Scanning Force Microscopy. Journal of Vacuum Science & Technology B 1997; 15: 1539.
[141] Kikukawa A., Hosaka S., Honda Y., Imura R. Phase-Controlled Scanning Force Microscope. Jpn. J. Appl. Phys. Part 2 - Lett. 1994; 33: L1286-L1288.

[142] Hadjar Y., Cohadon P. F., Aminoff C. G., Pinard M., Heidmann A. High-Sensitivity Optical Measurement of Mechanical Brownian Motion. Europhys. Lett. 1999; 47: 545-551.
[143] Carr D. W., Evoy S., Sekaric L., Craighead H. G., Parpia J. M. Measurement of Mechanical Resonance and Losses in Nanometer Scale Silicon Wires. Applied Physics Letters 1999; 75: 920-922.
[144] Tortonese M., Yamada H., Barrett R. C., Quate C. F. Atomic Force Microscopy Using a Piezoresistive Cantilever. The Proceedings of Transducers '91, IEEE, ed. Pennington, New Jersey, 1991.
[145] Tortonese M., Barrett R. C., Quate C. F. Atomic Resolution with an Atomic Force Microscope Using Piezoresistive Detection. Appl. Phys. Lett. 1996; 62: 834.
[146] Abedinov N., Grabiec P., Gotszalk T., Ivanov T., Voigt J., Rangelow I. W. Micromachined Piezoresistive Cantilever Array with Integrated Resistive Microheater for Calorimetry and Mass Detection. Journal of Vacuum Science & Technology a-Vacuum Surfaces and Films 2001; 19: 2884-2888.
[147] Thaysen J., Boisen A., Hansen O., Bouwstra S. Atomic Force Microscopy Probe with Piezoresistive Read-out and a Highly Symmetrical Wheatstone Bridge Arrangement. Sens. Actuator A-Phys. 2000; 83: 47-53.
[148] Wang Q. M., Cross L. E. Ferroelectrics 1998; 215: 187.
[149] Zurn S., M. Hseih, G. Smith, D. Markus, M. Zang, G. Hughes, Y. Nam, Arik M., Polla D. Smart materials and structures 2001; 10: 252.
[150] DeVoe D. L., Pisano A. P. Modeling and Optimal Design of Piezoelectric Cantilever Microactuators. Joyrnal of Microelectromechanical Systems 1997; 6: 266.
[151] Lee S. S., White R. M. Self-Excited Piezoelectric Cantilever Oscillators. Sensors & Actuators A Physical 1996; 52: 41.
[152] Lee J. H., Yoon K. H., Kim T. S. Characterization of Resonant Behavior and Sensitivity Using Micromachined Pzt Cantilever. Integrated Ferroelectrics 2002; 50: 43.
[153] Adams J. D., Parrott G., Bauer C., Sant T., Manning L., Jones M., Rogers B., McCorkle D. L., Ferrell T. L. Nanowatt Chemical Vapor Detection with a Self-Sensing, Piezoelectric Microcantilever Array. Appl. Phys. Lett. 2003; 83: 3428.
[154] Amantea R., Knoedler C. M., Pantuso F. P., Patel V. K., Sauer D. J., Tower J. R. An Uncooled IR Imager with 5 mK NETD. SPIE, ed. Orlando, FL, 1997.
[155] Baxter L. K., *Capactive Sensors, Design and Applications.* New York: IEEE Press, 1997.
[156] Erbe A., Blick R. H. Nanomechanical Resonators Operating at Radio Frequencies. Physica B 1999; 272: 575-577.
[157] Erbe A., Weiss C., Zwerger W., Blick R. H. Nanomechanical Resonator Shuttling Single Electrons at Radio Frequencies. Phys. Rev. Lett. 2001; 84: 96106.
[158] Kenny T. W., Kaiser W. J., Podosek J. A., Rockstad H. K., Reynolds J. K., Vote E. C. Micromachined Tunneling Displacement Transducers for Physical Sensors. J. Vac. Sci. Technol. A 1993; 11: 797.
[159] Binnig G., Rohrer H. IBM J. Res. Dev. 1986; 30: 355.
[160] Kenny T. W., Kaiser W. J., Waltman S. B., Reynolds J. K. Novel Infrared Detector Based on a Tunneling Displacement Transducer. Appl. Phys. Lett. 1991; 59: 1820.
[161] DiLella D., Whitman L. J., Colton R. J., Kenny T. W., Kaiser W. J., Vote E. C., Podosek J. A., Miller L. M. A Micromachined Magnetic-Field Sensor Based on an Electron Tunnelingdisplacement Transducer. Sens. Actuator A-Phys. 2000; 86: 8.
[162] Zhang J., O'Shea S. Tuning Forks as Micromechanical Mass Sensitive Sensors for Bio- or Liquid Detection. Sensors & Actuators B Chemical 2003; 94: 65.
[163] Lang H. P., Berger R., Battiston F., Ramseyer J. P., Meyer E., Andreoli C., Brugger J., Vettiger P., Despont M., Mezzacasa T., Scandella L., Guntherodt H. J., Gerber C., Gimzewski J. K. A Chemical sensor Based on a Micromechanical Cantilever Array for the Identification of Gases and Vapors. Applied Physics a (Materials Science Processing) 1998; 66: S 61-S 64.

[164] Albrecht T. R., Grutter P., Horne D., Rugar D. Frequency-Modulation Detection Using High-Q Cantilevers for Enhanced Force Microscope Sensitivity. Journal of Applied Physics 1991; 69: 668-673.
[165] Savran S. A., Sparks A. W., J. Sihler, Li J., Wu W.-C., Berlin D. E., Burg T. P., Fritz J., Schmidt M. A., Manalis S. R. Fabrication and Characterization of a Micromechanical Sensor for Differential Detection of Nanoscale Motions. Journal of Microelectromechanical Systems 2002; 11: 703.
[166] Saulson P. R. Thermal Noise in Mechanical Experiments. Physical Review D 1990; 42: 2437.
[167] Kajima M., Kusumi N., Moriwaki S., Mio N. Wide-Band Measurement of Mechanical Thermal Noise Using a Laser Interferometer. Physics Letters a 1999; 264: 251-256.
[168] Kruse P. W. "Principles of Uncooled Infrared Focal Plane Arrays." In *Uncooled Infrared Imaging Arrays and Systems, Semiconductors and Semimetals, 47*, Kruse, P. W., and D.D. Skatrud, ed. San Diego: Academic Press, 1997.
[169] Thundat T., Maya L. Monitoring Chemical and Physical Changes on Sub-Nanogram Quantities of Platinum Dioxide. Surface Science. 1999; 430: L546-L552.
[170] Thundat T., Chen G. Y., Warmack R. J., Allison D. P., Wachter E. A. Vapor Detection Using Resonating Microcantilevers. Anal. Chem. 1995; 67: 519.
[171] Lang H. P., Berger R., Andreoli C., Brugger J., Despont M., Vettiger P., C.Gerber, Gimzewski J. K., Ramseyer J. P., Meyer E., Guntherodt H. J. Sequential Position Readout from Arrays of Micromechanical Cantilever Sensors. Applied Physics Letters 1998; 72: 383.
[172] Lang H. P., Baller M. K., Berger R., Gerber C., Gimzewski J. K., Battison F. M., Fornaro P., Ramseyer J. P., Meyer E., GŸntherodt H. J. An Artificial Nose Based on a Micromechanical Cantilever Array. Anal. Chim. Acta 1999; 393: 59.
[173] Lang H. P., Baller M. K., Berger R., Gerber C., Gimzewski J. K., Battiston F. M., Fornaro P., Ramseyer J. P., Meyer E., Guntherodt H. J. An Artificial Nose Based on a Micromechanical Cantilever Array. Analytica Chimica Acta. 1999; 393: 59-65.
[174] Lavrik N. V., Tipple C. A., Sepaniak M. J., Datskos P. G. Sensing and Actuating Functionality of Hybrid MEMS Combining Enhanced Chemi-Mechanical Transduction with Surface-Enhanced Raman Spectroscopy. Proc. SPIE 2001; 4560: 152-161.
[175] Pinnaduwage L. A., Boiadjiev V., Hawk J. E., Thundat T. Sensitive Detection of Plastic Explosives with Self-Assembled Monolayer-Coated Microcantilevers. Appl. Phys. Lett. 2003; 83: 1471.
[176] Datskos P. G., Rajic S., Sepaniak M. J., Lavrik N., Tipple C. A., Senesac L. R., Datskou I. Chemical Detection Based on Adsorption-Induced and Photoinduced Stresses in Microelectromechanical Systems Devices. Journal of Vacuum Science & Technology B 2001; 19: 1173.
[177] Arakawa E. T., Lavrik N. V., Datskos P. G. Detection and Differentiation of Biological Species Using Microcalorimetric Spectroscopy. Ultramicroscopy 2003; 97: 459.
[178] Arakawa E. T., Lavrik N. V., Datskos P. G. Detection of Anthrax Simulants Using Microcalorimetric Spectroscopy: Bacillus Subtilis and Bacillus Cereus Spores. Applied Optics 2003; 42: 1757.
[179] Pampa, L.R. Senesac, Sepaniak M. J., Datskos P. G. in preparation.
[180] Headrick J., Lavrik N. V., Sepaniak M. J., Datskos P. G. Enhancing Chemi-Mechanical Transduction in Microcantilever Chemical Sensing by Surface Modification. Ultramicroscopy 2003; 97: 417.
[181] Butt H. J., Jaschke M., Ducker W. Measuring Surface Forces in Aqueous-Electrolyte Solution with the Atomic-Force Microscope. Bioelectrochem. Bioenerg. 1995; 38: 191-201.
[182] Ji H. F., Hansen K. M., Hu Z., Thundat T. Detection of pH Variation Using Modified Microcantilever Sensors. Sens. Actuator B-Chem. 2001; 72: 233-238.

[183] Ji H. F., Finot E., Dabestani R., Thundat T., Brown G. M., Britt P. F. A Novel Self-Assembled Monolayer (Sam) Coated Microcantilever for Low Level Caesium Detection. Chemical Communications 2000; 6: 457-458.
[184] Ji H. F., Thundat T., Dabestani R., Brown G. M., Britt P. F., Bonnesen P. V. Ultrasensitive Detection of Cro42- Using a Microcantilever Sensor. Analytical Chemistry 2001; 73: 1572-1576.
[185] Cherian S., Mehta A., Thundat T. Investigating the Mechanical Effects of Adsorption of Ca2+ Ions on a Silicon Nitride Microcantilever Surface. Langmuir 2002; 18: 6935-6939.
[186] Subramanian A., Oden P. I., Kennel S. J., Jacobson K. B., Warmack R. J., Thundat T., Doktycz M. J. Glucose Biosensing Using an Enzyme-Coated Microcantilever. Applied Physics Letters 2002; 81: 385-387.
[187] Baselt D. R., Lee G. U., Colton R. J. Biosensor Based on Force Microscope Technology. Journal of Vacuum Science & Technology B 1996; 14: 789-793.
[188] Baselt D. R., Lee G. U., Hansen K. M., Chrisey L. A., Colton R. J. A High-Sensitivity Micromachined Biosensor. Proc. IEEE 1997; 85: 672-680.
[189] Raiteri R., Nelles G., Butt H. J., Knoll W., Skladal P. Sensing of Biological Substances Based on the Bending of Microfabricated Cantilevers. Sens. Actuator B-Chem. 1999; 61: 213-217.
[190] Raiteri R., Grattarola M., Butt H. J., Skladal P. Micromechanical Cantilever-Based Biosensors. Sens. Actuator B-Chem. 2001; 79: 115-126.
[191] Porwal A., Narsude M., Rao V. R., Mukherji S. Microcantilever Based Biosensors. IETE Tech. Rev. 2002; 19: 257-267.
[192] Moulin A. M., O'Shea S. J., Badley R. A., Doyle P., Welland M. E. Measuring Surface-Induced Conformational Changes in Proteins. Langmuir 1999; 15: 8776-8779.
[193] Hansen K. M., Ji H. F., Wu G. H., Datar R., Cote R., Majumdar A., Thundat T. Cantilever-Based Optical Deflection Assay for Discrimination of DNA Single-Nucleotide Mismatches. Analytical Chemistry 2001; 73: 1567.
[194] Wu G. H., Datar R. H., Hansen K. M., Thundat T., Cote R. J., Majumdar A. Bioassay of Prostate-Specific Antigen (PSA) Using Microcantilevers. Nat. Biotechnol. 2001; 19: 856-860.
[195] Wu G. H., Ji H. F., Hansen K., Thundat T., Datar R., Cote R., Hagan M. F., Chakraborty A. K., Majumdar A. Origin of Nanomechanical Cantilever Motion Generated from Biomolecular Interactions. Proc. Natl. Acad. Sci. U. S. A. 2001; 98: 1560-1564.
[196] Marie R., Jensenius H., Thaysen J., Christensen C. B., Boisen A. Adsorption Kinetics and Mechanical Properties of Thiol-Modified DNA-Oligos on Gold Investigated by Microcantilever Sensors. Ultramicroscopy 2002; 91: 29-36.
[197] Ilic B., Czaplewski D., Craighead H. G., Neuzil P., Campagnolo C., Batt C. Mechanical Resonant Immunospecific Biological Detector. Appl. Phys. Lett. 2000; 77: 450.
[198] Olkhovets A. S., Evoy S., Carr D. W., Parpia J. M., Craighead H. G. Actuation and Internal Friction of Torsional Nanomechanical Silicon Resonators. J. Vac. Sci. Technology 2000; B18: 3549.
[199] Yang Y. T., Ekinci K. L., Huang X. M. H., Schiavone L. M., Roukes M. L. Monocrystalline Silicon Carbide Nanoelectromechanical Systems. Applied Physics Letters 2001; 78: 162.
[200] Gabrielson T. B. Mechanical-Thermal Noise in Micromachined Acoustic and Vibration Sensors. IEEE Transactions on Electron Devices 1993; 40: 903-909.
[201] Gabrielson T. B. Fundamental Noise Limits for Miniature Acoustic and Vibration Sensors. Journal of Vibration and Acoustics 1995; 117: 405-410.
[202] Lang H. P., Hegner M., Meyer E., Gerber C. Nanomechanics from Atomic Resolution to Molecular Recognition Based on Atomic Force Microscopy Technology. Nanotechnology 2002; 13: R29-R36.

Chapter 12
UNCOOLED INFRARED MEMS DETECTORS

P.G. Datskos [a,b] and N.V. Lavrik [a,b]

[a] *Oak Ridge National Laboratory*
Engineering Science and Technology Division
Oak Ridge, Tennessee, U.S.A

[b] *University of Tennessee*
Department of Physics and Astronomy
Knoxville, Tennessee, U.S.A

Abstract: Bimaterial microcantilevers arranged into focal plane arrays (FPAs) can function as uncooled IR imaging devices. In order to analyze the performance of such devices and compare various FPAs, it is essential to have an in-depth understanding of their operation, figures of merit, and fundamental limitations. We give an overview of figures of merit that are applicable to both cooled and uncooled IR detectors. Specific focus of this chapter is a performance analysis for microcantilever IR detectors with an optical readout. We discuss responsivity of microcantilever IR detectors and analyze the different sources (and mechanisms) of noise present in them. A model SiN_x microcantilever device with an Al layer in the bimaterial region was fabricated and its performance as an IR detector was analyzed.

Keywords: IR detector, thermal detector, focal place array, MEMS, microcantilever, bolometer, uncooled, figures of merit, responsivity, *NETD*, response time, noise, temperature fluctuation, thermo-mechanical, thermal isolation

1. INTRODUCTION

The detection of infrared (IR) radiation is very important for a variety of activities in both commercial and defense areas. In particular, the wavelength regions from 3 to 5 :m and 8 to 14 :m are of importance since atmospheric absorption in these regions is especially low [1,2]. Presently, however, the high cost of IR cameras has limited their use to primarily military applications. A much greater number of commercial applications [3] will be come possible as a result of recent advances in uncooled thermal detectors.

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 381-419.

The development of uncooled detector arrays started in the early 1980s and was pursued mostly in the United States. By the end of the 1990s, large focal plane arrays (FPAs) of resistive bolometers and ferroelectric devices with 320x240 pixels were available [3]. These FPAs exhibited noise equivalent temperature difference (*NETD*) values as low as 40 mK. Microbolometer FPAs developed by Boeing exhibited an *NETD* of 23 mK at 60 Hz frame rate [4]. Murphy *et. al*. [5] has recently reported a 320x240 microbolometer FPA with 25 μm pitch pixels. The reported average *NETD* value for these FPAs is about 35 mK with an f/1 aperture and operating at 30 Hz frame rates [6]. The challenge facing existing (and future) uncooled IR images is to achieve *NETD* values of only a few mK (*i.e.* performance equal to the best cooled photonic devices) while reaching the resolution of high definition television [6]. It has recently been shown that bimaterial microcantilevers arranged into FPAs can function as uncooled IR imaging devices. Using microcantilever FPAs, IR imaging of objects close to room temperature has already been demonstrated by several research groups.

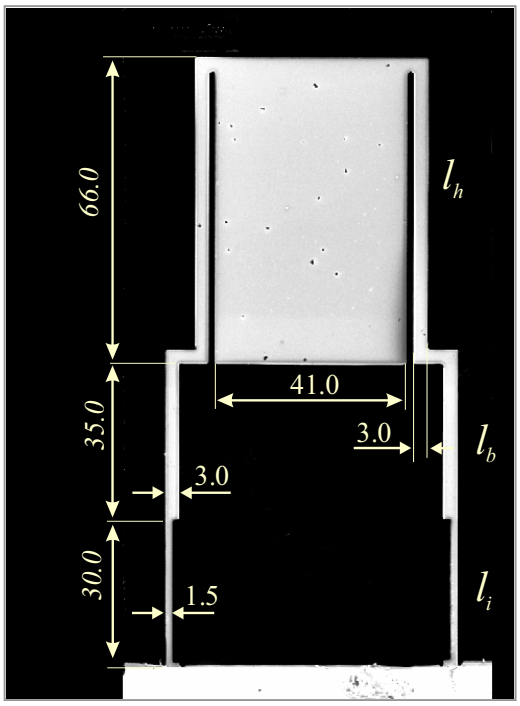


Figure 1. Ion micrograph and dimensions (in micrometers) of the cantilever device fabricated and used in our studies.

We fabricated a model microcantilever thermal detector (Figure 1) and investigated its performance both theoretically and experimentally. The geometry shown in Figure 1 was selected because pixels of this shape can be arranged (tiled) into an FPA with an acceptably high (>80%) fill factor. After evaluating general expressions for IR responses and noise, we applied them to the selected microcantilever design (Figure 1). Similar analysis can be applied to a variety of microcantilever designs and used to guide their optimization for IR imaging applications.

2. PHOTON DETECTION MECHANISMS

According to the transduction principles, IR detectors [7, 8] can be classified broadly as quantum (opto-electronic) [9] and thermal. Thermal IR detectors, in turn, can be based on pyroelectric [10], thermoelectric, thermoresistive (bolometers) [6, 11-15], and micromechanical transducers [16-25]. Quantum IR detectors use photon adsorption in semiconductor materials with narrow bandgaps, $\gamma_g < hc/\lambda$, or metal-semiconductor structures (Schottky barriers) with appropriately small energy barriers, $\Delta\gamma < hc/\lambda$. Because of the nature of photo-excitation and thermal-excitation processes in semiconductors, photonic IR detectors exhibit strong wavelength dependence and only operate efficiently when $k_BT < \hbar/\lambda$, where k_B is the Boltzmann constant, T is the detector temperature, $\hbar$ is the Plank constant, and λ is the wavelength of radiation to be detected. Although quantum IR detectors may have short (sub-nanosecond) response times and very high detectivities (see the definitions latter in this paper) approaching fundamental limits, they require deep cooling in order to reduce thermal generation of charge carriers and thermal noise that varies as $\exp[-\gamma_g/k_BT]$. Cooling of quantum IR detectors down to or below liquid nitrogen temperatures is commonly used for sensitive imaging in the mid- to far-infrared regions using IR photodetectors. Thermal IR detectors are based on measuring the amount of heat produced in the detector upon absorption of IR radiation and can operate at or even above room temperatures. Spectral characteristics that are flat and extended into the far-IR range are typical for thermal IR detectors. A spectral response of a thermal IR detector is primarily defined by the absorbance spectrum of the detector active region, which, in principle, can be close to unity for radiation ranging from the visible to the far-IR. When thermal IR detectors are arranged into FPAs (see Figure 2), the long-wavelength roll-off of the spectral characteristics is typically affected by diffraction phenomena (associated with smaller detector sizes) rather than materials properties. Microbolometers and FPAs of microbolometers operating at the room temperature have already demonstrated sufficient performance, which makes them quite competitive with respect to more traditional cooled IR detectors based on narrow-band semiconductors and Schottky barriers. However, thermal IR detectors tend to have slower response times ($> 10^{-3}$ sec), and somewhat lower detectivities, which are limited by relatively high background temperature fluctuations present in any uncooled IR detector.

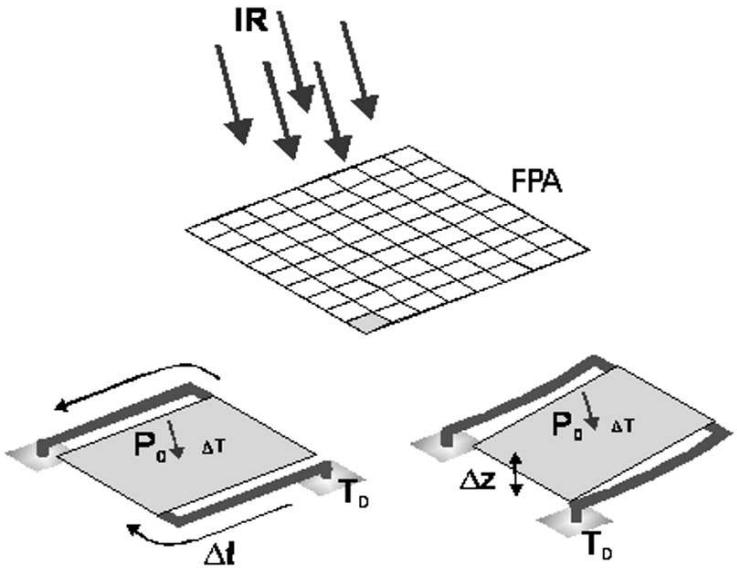


Figure 2. IR detectors can be tiled to form large focal plane arrays (FPAs) with high fill factors.

2.1 Transduction Mechanisms in MEMS

More recently, a new type of uncooled thermal detectors based on thermal expansion phenomena in micromechanical structures has been introduced and studied [1, 16-18, 20, 21, 23, 25]. Suspended microstructures (microcantilevers) with bimaterial regions provide direct conversion of absorbed heat into a mechanical response and can be referred to as thermo-mechanical detectors. An important advantage of thermo-mechanical detectors is that they are essentially free of intrinsic electronic noise and can be combined with a number of different readout techniques. The readout techniques demonstrated to date include capacitive [19], piezoresistive [24], electron tunneling [26] and optical [1, 27].

A means to measure the deformation (bending) of the thermo-mechanical detector in response to the absorbed incident radiation is essential in order for the microcantilever FPA to be an imaging IR detector. Earlier studies have shown that micromechanical deformations can readily be determined by a number of means, including optical, capacitive, piezoresistive, and electron tunneling with extremely high sensitivity [28-31].

Metal-coated microcantilevers that are commonly employed in atomic force microscopy (AFM) can be probed with sub-Angstrom ($<10^{-10}$ m) deflection accuracy. For example, Hansma [31] and Binnig [30] have demonstrated AFM sensitivities of 10^{-11} N, corresponding to deflection magnitudes of 5×10^{-11} m. Measurements of even smaller deflections are associated with detection of atto-Newton (10^{-18} N) level forces [32].

Although there are several ways to measure deformations of microcantilevers, the "optical lever" approach is very attractive [29] for this purpose. According to this approach, a beam of light, usually from a laser, is focused on the end of a microcantilever [29]. Microcantilever devices are normally designed so that the probing laser light is mostly reflected without absorption. The light reflected off the cantilever is directed onto a two-element (or four-element) photodiode, the outputs of which are electronically amplified and their difference signal is measured [27, 29]. When the microcantilever deforms in response to absorbed IR radiation, the spatial distribution of the light projected onto the photodiode changes. As a result, integral illumination of one photodiode increases whereas that of the other photodiode decreases and the difference in their relative signal changes. The change in the measured signal is related to the degree of microcantilever bending and is a function of the absorbed thermal radiation.

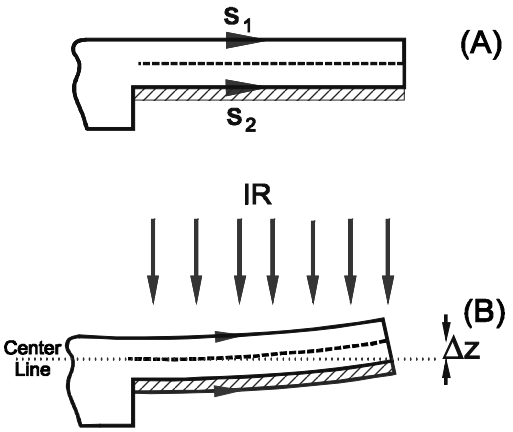


Figure 3. Schematic of a bimaterial cantilever device showing the mechanical response to absorbed infrared radiation. Initial state (A) of the bimaterial cantilever changes when absorbed photons heat up the device, which leads to (B) bending due to the difference in thermal expansion of the two materials. The parameter Δz denotes displacement of the cantilever end.

2.2 Response of Thermomechanical Detectors

When the temperature of the system changes, microcantilevers undergo bending which is due to difference in thermal expansion of the two materials

composing the microcantilever (Figure 3). This basic phenomenon is frequently referred to as the "bimetallic effect" [33, 34]. We previously showed that silicon nitride cantilevers with a thin gold film on one side undergo measurable bending in response to minute temperature changes [23]. In that particular case, the differential stress in the cantilever was created due to dissimilar thermal expansion of the silicon nitride substrate and the gold coating. Barnes *et al.* [35] estimated that the smallest temperature changes that can be detected using bimaterial microcantilevers are in the range of 10^{-5} to 10^{-4} K.

The steady state deflection of the tip of a bimaterial cantilever tip in response to a temperature change, ΔT, is given by [34]

$$\Delta z = \frac{3l_b^2}{t_1 + t_2}\left[\frac{\left(1+\frac{t_1}{t_2}\right)}{3+\left(1+\frac{t_1}{t_2}\right)^2+\left(1+\frac{t_1}{t_2}\frac{E_1}{E_2}\right)\left(\frac{{t_1}^2}{{t_2}^2}+\frac{t_2}{t_1}\frac{E_2}{E_1}\right)}\right](\alpha_1-\alpha_2)\Delta T \quad (1)$$

where l_b is the microcantilever bimaterial length, t_1 and t_2 are the thickness of the coating and the microcantilever substrate, α_1 and α_2 are the thermal expansion coefficients of the coating and microcantilever, and E_1 and E_2 are the Young's moduli of the coating and microcantilever. As shown in Figure 3, Δz refers to the vertical deflection of the centerline of the microcantilever, at its outmost (right) end due to thermally induced stress. It can be concluded from Equation (1) that the displacement, $\Delta z(\omega)$, can be maximized by designing detectors with proper geometry and selecting appropriate bimaterial layers. Although the thickness of each layer, as well as the overall length of the cantilever, have a dramatic effect on the displacement magnitude, optimization of the cantilever design can not be achieved based on the condition of the maximum displacement alone. For example, the amplitude of cantilever deflection increases as the square of the bimaterial length, l_b. However, as the cantilever length increases so does the thermo-mechanical noise. We should point out that Equation (1) reverts to an equivalent form which has been used extensively by a number of groups [1, 35, 36] *viz.*

$$\Delta z = 3(\alpha_1 - \alpha_2)\frac{\left(1+\frac{t_1}{t_2}\right)}{t_2^2\left(4+6\frac{t_1}{t_2}+4\frac{t_1^2}{t_2^2}+\frac{E_1}{E_2}\frac{t_1^3}{t_2^3}+\frac{E_2}{E_1}\frac{t_2}{t_1}\right)} l_b \Delta T \quad (2)$$

The response of microcantilever detectors can also be expressed as a function of incident radiation power. The temperature rise, ΔT, of the detector as a result of photon absorption is given by the following solution of the heat flow equation [6]

$$\Delta T = \frac{\eta P_0}{G\sqrt{1+\omega^2 {\tau_{th}}^2}} \quad (3)$$

where P_0 is the amplitude of the radiant power falling on the cantilever, η is the absorbance (absorbed fraction) of the radiant power, G, is the thermal conductance of the principal heat loss mechanism, ω is the angular frequency of modulation of the radiation and τ_{th} is the thermal response time. Another frequency-dependent term is related to the fact that microcantilever detectors are mechanical resonators and amplitude of their responses is a resonant function of input stimuli (temperature changes in the present case) [29]

$$\Delta z(\omega) = \frac{\Delta z}{\sqrt{\left(1-\frac{\omega^2}{{\omega_0}^2}\right)^2 + \frac{\omega^2}{{\omega_0}^2 Q^2}}} \quad (4)$$

where Δz is the steady state deflection given by Equation (1), ω and ω_0 are angular frequencies of, respectively, temperature changes and microcantilever mechanical resonance, and Q is the quality factor. We can obtain the final expression for $\Delta z(\omega)$ by combining Equations (1), (3) and (4), we obtain

$$\Delta z(\omega) = \eta \frac{3}{2} \frac{l_b^2}{t(_1 + t_2)} \left[\frac{\left(1 + \frac{t_1}{t_2}\right)^2}{3\left(1 + \frac{t_1}{t_2}\right)^2 + \left(1 + \frac{t_1}{t_2}\frac{E_1}{E_2}\right)\left(\frac{t_1^2}{t_2^2} + \frac{t_2}{t_1}\frac{E_2}{E_1}\right)} \right] \times \tag{5}$$

$$(\alpha_1 - \alpha_2)\frac{P_0}{G\sqrt{1 + \omega^2 \tau_{th}^2}} \frac{1}{\sqrt{\left(1 - \frac{\omega^2}{\omega_0^2}\right)^2 + \frac{\omega^2}{\omega_0^2 Q^2}}}$$

3. FIGURES OF MERIT OF IR DETECTORS

Because of the complex multi-parameter relationships that determine the performance of any IR detector, rigorous analytical definitions of figures of merit and their understanding are of crucial importance. As we will show later, interpretation of these definitions in the framework of quantitative expressions for cantilever responses and noise is a key to rational design of microcantilever IR detectors. During the last several decades, a number of different figures of merit have been introduced in order to characterize IR detectors [6, 37-43]. Although many of these parameters are still in use, evolution of the IR detectors has been accompanied by evolution of characterization methods and respective parameters. Some of the previously introduced figures of merit become outdated merely because of the changes in units used to characterize IR detectors. Here we only discuss the figures of merit that are currently accepted and used by the IR community [39, 44]. It is important to keep in mind that different assumptions are sometimes made in defining and measuring these parameters.

When evaluating the performance of various, especially uncooled, IR detectors, the parameters of the major importance [39, 43, 44] are: (i) responsivity, *R*; (ii) noise equivalent power (*NEP*); (iii) normalized detectivities, *D** (and *D***); (iv) noise equivalent temperature difference (*NETD*); (v) minimum resolvable temperature difference (*MRTD*), and (vi) response time τ. The definitions of these parameters and their fundamental limits in the case of uncooled thermal IR detectors are provided below. There are a number of additional parameters that can be used for a more detailed and comprehensive characterization of IR detectors. These include: linearity of response, cross-talk between detector elements in an FPA, dynamic range, and

modulation transfer function (*MTF*). The linearity of response, cross-talk and dynamic range are basic parameters amenable to a whole variety of analog devices and transducers and their definitions are readily available from a number of sources [37, 45]. *MTF* is traditionally used in testing the performance of lenses, imaging systems and their components and describes how the output contrast changes as a series of incrementally smaller features are imaged [45].

3.1 Responsivity

The responsivity parameter R (applicable to all detectors) reflects the gain of the detector and is defined as the output signal (typically voltage or current) of the detector produced in response to a given incident radiant power falling on the detector [39, 43].

$$R_V = \frac{\Delta V}{P_0} \qquad or \qquad R_I = \frac{\Delta I}{P_0} \tag{6}$$

where P_0 is the radiant input power (Watts); ΔV and ΔI are, respectively, the output voltage (Volts) and current (Amperes) of the detector. When the definition of the responsivity is expanded to include the frequency dependence and the wavelength (spectral) dependence [39], then, the responsivity is known as the spectral responsivity, $R(\lambda, \omega)$. It is worthy to emphasize that very distinct factors define characteristic features of spectral responsivities in the case of quantum and thermal IR detectors. Quantum IR detectors exhibit a cut-off in the spectral responsivity above a certain characteristic wavelength that is related to the photon energy sufficient to generate additional charge carriers (free electrons or electron-hole pairs). Hence, R (λ, ω) has a long wavelength cut-off defined by the bandgap energy of the semiconductor or the energy barrier at the metal-semiconductor interface used in the device. In the case of thermal IR detectors, however, the far-IR range is readily accessible simply by using appropriate detector absorbing areas and materials with high-absorptivity (either direct or resonant absorption) in this region.

Another derivative of *responsivity* is known as *blackbody responsivity*, R (T, ω) and is defined to include the dependence of the detector output signal on the temperature, T, of the blackbody-type source. The responsivity is a useful technical parameter that allows the prediction of the detector signal levels caused by an IR radiation of the given power and wavelength or as a

result of thermal radiation from an object at a given temperature and emissivity. Although responsivity provides a good indication about the performance of an IR detector, it does not take into account the level of any intrinsic noise in the device, and, therefore, provides little or no information about the threshold sensitivity of the detector. In other words, an IR detector with high responsivity is not necessarily able to detect low-level IR radiation or to distinguish IR sources of nearly the same intensity. It can be concluded that knowing the detectors responsivity is important at the stage of designing an IR detection system while comparative evaluation of different detectors should rely on other figures of merit.

In the case of thermo-mechanical IR detectors, such as microcantilever bimorphs, the intrinsic thermo-mechanical responsivity of the detector can be defined in terms of the mechanical response of the device (in units of meters per Watt), i.e. displacement, Δz, per absorbed power, P_0

$$R_z = \frac{\Delta z}{P_0} \tag{7}$$

Similarly as before, a spectral responsivity $R_z(\lambda,\omega)$ and a blackbody responsivity $R_z(T,\omega)$ can be defined. Combining Equations (5) and (7), we obtain the frequency dependent responsivity, $R_z(\omega)$, of a bimaterial microcantilever IR detector

$$\begin{aligned} R_z(\omega) &= \frac{\Delta z(\omega)}{P_0} \\ &= \eta \frac{3}{2} \frac{l_b^2}{t(_1+t_2)} \left[\frac{\left(1+\frac{t_1}{t_2}\right)^2}{3\left(1+\frac{t_1}{t_2}\right)^2 + \left(1+\frac{t_1}{t_2}\frac{E_1}{E_2}\right)\left(\frac{t_1^{\,2}}{t_2^{\,2}} + \frac{t_2}{t_1}\frac{E_2}{E_1}\right)} \right] \times \\ &\qquad (\alpha_1 - \alpha_2) \frac{1}{G\sqrt{1+\omega^2 \tau_{th}^{\,2}} \sqrt{\left(1-\frac{\omega^2}{\omega_0^{\,2}}\right)^2 + \frac{\omega^2}{\omega_0^{\,2} Q^2}}} \\ &= \frac{R_0(\omega)}{\sqrt{\left(1-\frac{\omega^2}{\omega_0^{\,2}}\right)^2 + \frac{\omega^2}{\omega_0^{\,2} Q^2}}} \end{aligned} \tag{8}$$

As follows from Equations (4) and (8), the responsivity, $R_z(\omega)$ can be used to express deflection responses of a microcantilever detector as a function of changes in its temperature, ΔT

$$\Delta z(\omega) = \frac{R_z(\omega) G \sqrt{1 + \omega^2 \tau_{th}{}^2}}{\eta} \Delta T \tag{9}$$

Equation 9 is used later in our analysis of the performance limits imposed by temperature fluctuation noise in microcantilevers.

3.2. Noise Equivalent Power

The minimum radiant flux that can be measured by different IR detectors with the same responsivity is inversely proportional to the level of total intrinsic noise in the detector. A convenient way to characterize the sensitivity of an IR detector is to specify its *noise equivalent power (NEP)*, a parameter defined as the radiant power incident on the detector that produces a signal equal to the rms detector noise [39, 43]. By this definition, *NEP* takes into account both gain and noise parameters of the detector and can be related to the detector responsivity, R_V, R_I or R_z and the rms detector noise [39]

$$NEP = \frac{\left\langle \delta V^2 \right\rangle^{1/2}}{R_V} \quad \text{or} \quad NEP = \frac{\left\langle \delta I^2 \right\rangle^{1/2}}{R_I} \quad \text{or} \quad NEP = \frac{\left\langle \delta z^2 \right\rangle^{1/2}}{R_z} \tag{10}$$

where $<\delta V^2>^{1/2}$, $<\delta I^2>^{1/2}$ and $<\delta z^2>^{1/2}$ are, respectively, the rms noise voltage, current and displacement measured within the whole operation bandwidth. Since *NEP* depends on *R*, it also depends on the photon wavelength as well as on the modulation frequency of the IR power and, therefore, can be regarded to as *NEP* (λ ,ω). *NEP* can also be specified a function of the detector temperature, i.e. *NEP* (T_D, ω) [39]. Frequency dependence of *NEP* is determined by the detector thermal response time, τ, and by the spectral density (i.e. frequency dependency) of the detector noise. *NEP* without specifying the measurement bandwidth may have ambiguous interpretation. The units of *NEP* may imply either a full operational bandwidth or a 1 Hz bandwidth. As a rule, *NEP* (λ ,ω) and *NEP* (T_D, ω) refer to a 1 Hz bandwidth and have units of W $Hz^{-1/2}$.

3.3 Normalized Detectivity

The parameter *NEP* is generally sufficient to evaluate and compare the performance of single (spot) IR detectors by predicting the minimum detectable power. However, a figure of merit that is directly proportional to the detector performance is more convenient. Starting with a parameter known as detectivity, D, which is defined as the inverse value of *NEP* and taking into account the detector absorbing (active) area, A_d and the signal bandwidth, B, one can define specific (or normalized) detectivity, D^* as [39]

$$D^* = \frac{(A_d B)^{1/2}}{NEP} \tag{11}$$

Normalized (or specific) detectivity D^* is, therefore, the detector output signal-to-noise ratio at 1 watt of input IR radiation normalized to a detector with a unit active area and a unit bandwidth. The units of D^* are Jones; Jones = (cm $Hz^{1/2}$ W^{-1}). It should be noted that the definition of specific detectivity, D^*, was originally proposed for quantum detectors, in which the noise power is always proportional to the detector area and noise signal (ΔV_n or ΔI_n) proportional to the square root of the area. However, the noise in thermal IR detectors does not always obey this scaling trend. In fact, neither temperature fluctuations nor thermo-mechanical noise (see the next section) scale up with the detector area. Therefore, D^* should be very cautiously interpreted when applied to thermal IR detector. In fact, D^* tends to overestimate the performance of larger absorbing area thermal detectors and underestimates the performance of smaller ones. In general (even in the case of quantum detectors), D^* ignores the significance of smaller detector elements for high-resolution FPAs.

In order to take into account possible variability in the efficiency of the optics used for characterization of IR detectors, the focal ratio F, (reciprocal of twice the half angle of the marginal ray from the edge of the optics to the focal point, is included into a modified definition of the normalized detectivity, D^{**}

$$D^{**} = \frac{F(A_d B)^{1/2}}{NEP} \tag{12}$$

A reasonable approximation of F assumes that the target remains at infinity so that $F = f_0 / d$, where f_0 is a focal length of the optics and d is a diameter of the optics.

3.4 Noise Equivalent Temperature Difference

The noise equivalent temperature difference (*NETD*) is a parameter characterizing the low-signal performance of thermal imaging systems and is more applicable to FPAs of IR detectors. *NETD* is defined as the temperature of a target that is different from the background temperature and results in a detector output signal equal to the detector rms noise [39, 43, 46]. *NETD* can be specified for a single detector element or can be averaged of all detector elements in an array. Alternatively, *NETD* can be defined as the difference in temperature between two blackbodies, which corresponds to a signal-to-noise ratio of unity [43]. When an IR imager produces images, it actually maps the detected temperature variation across a scene or an object. However, the resulting images are also affected by the emissivity of the objects in the scene.

Small values of *NETD* reflect the ability of an IR imager to distinguish slight temperature or emissivity differences of the objects. A number of relationships for predicting *NETD* has been derived [6, 39, 41, 46]. *NETD* can also be calculated by knowing the rms noise and a signal for a given difference between target and background temperatures [46]

$$NETD = \frac{\left\langle \delta V^2 \right\rangle^{1/2}}{\Delta V_{T_t}} \left(T_t - T_B \right)$$

or

$$NETD = \frac{\left\langle \delta I^2 \right\rangle^{1/2}}{\Delta I_{T_t}} \left(T_t - T_B \right) \tag{13}$$

or

$$NETD = \frac{\left\langle \delta z^2 \right\rangle^{1/2}}{\Delta z_{T_t}} \left(T_t - T_B \right)$$

where ΔV_{T_t}, ΔI_{T_t} and Δz_T are, respectively, changes in the output voltage, current and deflection of the detector in response to a change of the target temperature from the background temperature, T_b, to T_t. Using the responsivity definition given by Equation (7), we can substitute Equation (13) Δz_t in with $R(\omega)\times\Delta P_0$ we get

$$NETD = \frac{\left\langle \delta z^2 \right\rangle^{1/2}}{R_z(\omega)\,\Delta P_0}(T_t - T_b) \tag{14}$$

where ΔP_0 is a change in the incident radiation power when temperature of the target changes from T_b, to T_t.

In Section 4, we will discuss several distinct noise mechanisms in MEMS detectors and show how the respective fundamental limits of *NETD* depend on the detector design and its operation regime.

3.5 Minimum Resolvable Temperature Difference

The minimum resolvable temperature difference (*MRTD*) is widely accepted by the IR community as a measure of the total performance [39, 44] of IR imaging systems. The rigorous definition of *MRTD* involves both the temperature sensitivity and the spatial resolution of a thermal imaging system [44, 46]. *MRTD* is a more informative parameter than *NETD* when a combination of spatial resolution and temperature sensitivity need to be are taken into account. Although well-established methods have been developed to measure *MRTD*, it is still one of the most difficult figures of merit to determine.

3.6 Response Time

A general definition of the *response time*, τ, for any sensor systems is the time required for a transient output signal to reach 0.707 ($2^{-1/2}$) of its steady state change. Similarly to other sensor or transducer systems, any IR detector exhibits a characteristic transient behavior when the input IR power changes abruptly. In many cases, transient behavior of an IR detector accounted for by considering a first order low-pass filter. In these cases, the detector responsivity in the time domain is given by [45]

$$R(t) = R_{t=\infty}\left[1 - e^{-t/\tau}\right] \tag{15}$$

where, τ is the detector response time, and ω is the angular modulation frequency of the radiation to be detected (angular frequency, $\omega = f/2\pi$, is commonly used in theoretical evaluations of thermal detectors). However, the transient behavior of thermomechanical detectors can be more complex than that predicted by Equation (15) because of the mechanical resonances present in mechanical devices.

When transduction of the absorbed IR energy is based on photoelectronic excitation (quantum detectors), the intrinsic response time can be less than a nanosecond. Although the response time of quantum IR detectors is often limited by high-impedance electronic readout, their overall response times are commonly shorter than 10^{-3} s [15]. This satisfies the requirements of the majority real-time IR imaging applications. By contrast, much longer response times (typically in the range of 10^{-3} to 10^{-1} s) of thermal IR detectors are associated with the required accumulation of heat in the detectors active area and are directly related to their transduction mechanism. The response time of a thermal IR detector, τ_{th}, can be calculated as the ratio of the heat capacity of the detector to its effective thermal conductance, *viz*.,

$$\tau_{th} = \frac{C}{G} \tag{16}$$

where C is the heat capacity of the detector active area and G is the total thermal conductance between the active area of the detector and the support structure (i.e. a heat sink). Equation 16 provides a convenient way to estimate the response time of a thermal IR detector, including thermomechanical devices. In Equation 16, the heat capacitance, C, is total capacitance which takes into account the individual capacitances for each of the materials comprising the detector active area given by the products of the specific heat capacitances and their respective masses. The value of G should take into account all the heat loss mechanisms in the detector and in the case of evacuated thermal detectors these are conductive and radiative losses.

4. NOISE SOURCES IN THERMAL IR DETECTORS

Noise sources may exist within the IR detector itself, as it interacts with its environment or can be introduced by the detector readout system. While significantly reduced sizes and heat capacitances improve sensitivity of micromachined thermal and calorimetric detectors, the noise mechanisms intrinsic to the microscale, impose several limitations on their performance. Two of such limitations (background limited and temperature fluctuation limited) are applicable to all types of thermal IR detectors and stem merely from the fact that every object, depending on its thermal mass and the degree of heat exchange with the environment, undergoes certain temperature fluctuations. These spontaneous temperature fluctuations are negligible for macroscopic objects, however, may become a significant source of noise in the case of highly thermally isolated microscopic detectors, such as microbolometers and micromachined suspended microcantilevers.

In the case of thermo-mechanical IR detectors, there is an additional fundamental limitation that is related to spontaneous microscopic mechanical motion (oscillation) of any suspended microstructure due to its thermal energy. For the majority of the readout means, these thermo-mechanical oscillations are indistinguishable from the temperature-induced bending and, therefore, directly contribute to the detector noise.

4.1 Temperature Fluctuation Noise

All IR detectors that are based on conversion of IR radiant power into heat are affected by temperature fluctuation noise [6, 11, 47]. The magnitude of spontaneous temperature fluctuations of the detector can be derived from the fluctuation-dissipation theorem (FDT) [6, 11]

$$\left\langle \delta T_{TF}{}^{2} \right\rangle = \frac{k_B T_D{}^{2}}{C} \tag{17}$$

where $<\delta T^2>$ is the mean square fluctuations in temperature of the detector, and T_D is the detector temperature. The temperature fluctuations $<\delta T_{TF}{}^2>$ in Equation (17) is a result of the integration over all frequencies. Taking into account the thermal time constant, τ_{th}, the spectral density of the root mean square (rms) temperature fluctuations is given by [6]

$$\left\langle \delta T_{TF}{}^{2} \right\rangle^{1/2} = \frac{2T_D\sqrt{k_B B}}{\sqrt{G\left(1+\omega^2\tau_{th}{}^{2}\right)}} \tag{18}$$

where k_B is the Boltzmann constant and B is the measurements bandwidth. Equation (18) shows that thermal conductance G of the principal heat loss mechanism is the key design parameter that affects the temperature fluctuation noise. Figure 4 shows examples of temperature fluctuation spectra calculated for a typical IR sensitive micro-mechanical detector using Equation (18).

When a thermal detector operates in vacuum or a gas environment at reduced pressures, heat conduction through the supporting microstructure of the device is usually a dominant heat loss mechanism [15]. In the case of an extremely good thermal isolation, however, the principal heat loss mechanism can be reduced to only radiative heat exchange between the detector and its surroundings which is given by G=4 ε A_d σ_t T^4 assuming that the emissivity ε=1; σ_t is the Stefan-Boltzmann constant. For all micromechanical IR detectors in atmospheric environment, the heat conduction through air is likely to be dominant heat dissipation mechanism. The thermal conductivity of air at standard temperature and pressure is 2.4×10^{-2} W m^{-1} K^{-1} [15], which yields the respective thermal conductance larger than the thermal conductance through the supporting beams of a typical microcantilever or microbolometer detector.

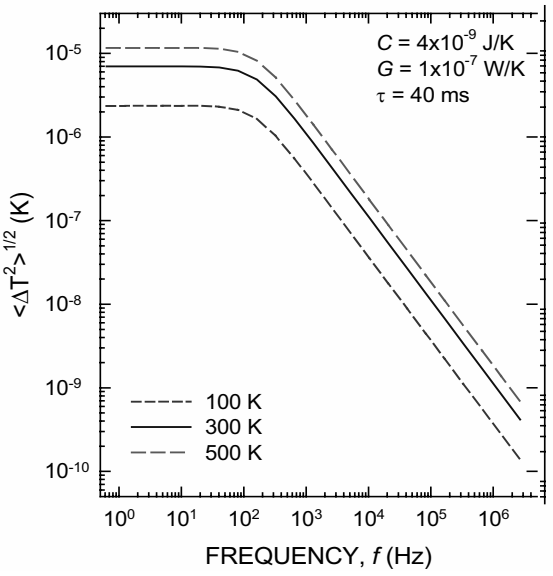


Figure 4. Spectral density of temperature fluctuation noise (rms values of temperature fluctuation) calculated for a typical thermal IR detector using Equation 18.

In the case of temperature sensitive bimaterial microcantilever devices,

the temperature fluctuation noise can manifest itself as spontaneous displacements of the cantilever tip, δz_{TF}. By substituting Equation (18) into Equation (9), we obtain the rms of δz_{TF}

$$\left\langle \delta z_{TF}^2 \right\rangle^{1/2} = \frac{R_z(\omega)\sqrt{4k_B B G}\, T_D}{\eta} \tag{19}$$

4.2 Thermo-Mechanical Noise

In addition to temperature fluctuation noise and background fluctuation noise, there are noise sources that are specific to microcantilever IR detectors. One of the essential features of microcantilever detectors is that, unlike other types of uncooled IR detectors, they are mechanical devices (oscillators) that can accumulate and store mechanical energy. When a microcantilever detector is in equilibrium in a thermal bath, there is a continuous exchange of the mechanical energy accumulated in the device and thermal energy of the environment. This exchange is dictated by the fluctuation dissipation theorem [48, 49] and results in spontaneous oscillation of the microcantilever so that the average mechanical energy per mode of cantilever oscillation is defined by thermal energy $k_B T$. Sarid [29] described this noise source as "thermally induced lever noise".

The analysis provided by Sarid [29] involves the Q-factor of a vibrating microcantilever, its resonance frequency, ω_0 and stiffness, k. While the Q-factor can be defined empirically as the ratio of the resonance frequency to the resonance peak width, knowing the exact mechanisms of microcantilever damping is important for evaluation of the thermo-mechanical noise spectrum. The thermo-mechanical model described by [29] assumes that damping of the microcantilever is of viscous nature. Assumption of predominantly viscous damping is valid for microcantilevers in air or water and, therefore, justified for micromechanical devices used as force probes in scanning probe microscopy. In the case of a microcantilever in a viscous medium, such as air or water, the damping force is proportional to the microcantilever linear velocity. The resulting noise spectrum is given by [29, 50].

$$< \delta z_{TM}{}^2 >^{1/2} = \sqrt{\frac{4k_B T B}{k Q \omega_0} \frac{1}{\left(1 - \frac{\omega^2}{\omega_0^2}\right)^2 + \frac{\omega^2}{Q^2 \omega_0^2}}} \tag{20}$$

The expression given by Equation (20) results in a frequency independent noise density for the frequencies well below the mechanical resonance frequency, ω_0, (*i.e.*, $\omega << \omega_0$). At these low frequencies, the rms of the microcantilever tip displacements due to thermo-mechanical noise is [29]

$$\left\langle \delta z_{TM}^2 \right\rangle_{\omega<<\omega_0}^{1/2} = \sqrt{\frac{4k_B TB}{Qk\omega_0}} \tag{21}$$

where k is the spring constant, defined as the ratio of the force applied to the microcantilever divided by the displacement of the tip. At the resonance (*i.e.* $\omega = \omega_0$) [29]

$$\left\langle \delta z_{TM}^2 \right\rangle_{\omega=\omega_0}^{1/2} = \sqrt{\frac{4k_B TBQ}{k\omega_0}} \tag{22}$$

The latter condition may only become important for a thermo-mechanical IR detector operating at or near one of its resonance frequencies. It should be pointed out that alternative models exist to adequately describe thermo-mechanical noise of a mechanical oscillator below its resonance [50]. Majorana *et al.* [50] discussed an alternative model of thermo-mechanical noise that accounts for internal friction processes rather than viscous damping. According to this model, the thermo-mechanical noise spectrum can be expressed as [50, 51]

$$\left\langle \delta z_{TM'}^2 \right\rangle^{1/2} == \sqrt{\frac{4k_B TB}{kQ\omega} \frac{1}{\left(1-\frac{\omega^2}{\omega_0^2}\right)^2 + \frac{\omega^2}{Q^2\omega_0^2}}} \tag{23}$$

Equation (23) indicates that the density of thermo-mechanical noise follows a $1/f^{1/2}$ trend below the resonance when the damping is due to intrinsic friction processes rather than viscous interactions with the medium. Analysis of Equations (20) and (23) shows that, regardless of the dissipation mechanism, the off-resonance thermo-mechanical noise is lower in the case of microcantilevers with higher Q-factor and higher k. It should be emphasized that, while predictions based on Equation (20) are often reported in the literature [1, 52, 53], the noise density calculated according to the two alternative models may substantially deviate from each other at low frequencies

[49-51]. Furthermore, the intrinsic friction model predicts the low frequency noise to be independent of the microcantilever resonance frequency provided that the stiffness, k, is constant. By contrast, the viscous damping model predicts that the low frequency noise of a microcantilever detector can be decreased by increasing its resonance frequency even without changes in its stiffness. Therefore, knowing the actual mechanisms of mechanical dissipation in the microcantilever detector can be critical in analyzing thermomechanical noise of a microcantilever detector in the frequency range relevant to real-time IR imaging.

5. FUNDAMENTAL LIMITS

The fundamental limits to the performance of the microcantilevers are related to the properties of the thermal detectors themselves and, therefore, independent of the readout means. The primary fundamental limits are the background fluctuation limit and the temperature fluctuation limit. These limits exist due to the fluctuations in the detector temperature that result from the dynamic nature of the heat exchange between the detector and its surroundings. An ideal, noiseless readout amplifies and displays both the signal and the noise inherent to the microcantilevers without distorting them or changing the signal-to-noise ratio. In practice, there are no ideal, noiseless readout methods; at best, the readout decreases the inherent signal-to-noise ratio of the microcantilevers by only a small amount.

Effects of distinct noise mechanisms on the performance of microcantilever IR detectors can be understood by analyzing expressions for signal-to noise ratios or *NETD* that account for the respective noise mechanism. In order to obtain analytical expressions for *NETD* for a general case, we will take into account efficiency of the imaging optics and assume that the emissivity of the target, ε_t=1. Then, the incident radiation power, P_0, used in Equation (14) can be expressed as a function of the target temperature, T_t as

$$P_0 = \frac{\tau_0 A_d}{4F^2}\left(\frac{dP}{dT}\right)_{\lambda_1-\lambda_2} T_t \tag{24}$$

where, F and τ_o are, respectively, the focal ratio and the absorption of the optics, A_d is the detector absorbing area, and $(dP/dT)_{\lambda_1-\lambda_2}$ is the slope of the function $P= f(T_t)$ that describes power radiated by a blackbody target within

the spectral band from λ_1 to λ_2 per unit area of the detector. By substituting Equation (24) into Equation (14), we obtain

$$NETD = \frac{4F^2 \left\langle \delta z^2 \right\rangle^{1/2}}{\tau_0 A_d R(\omega) \left(\frac{dP}{dT} \right)_{\lambda_1 - \lambda_2}} \tag{25}$$

The generalized expression for *NETD* of a microcantilever IR detector given by Equation (25) will be used in the following sections in order to evaluate the limit imposed by a specific noise mechanism.

5.1 Background Fluctuation Noise Limit

The highest possible degree of thermal isolation is always limited by radiative heat exchange between the detector and its surroundings. As in the case of any other heat loss mechanisms, the temporal fluctuations in the radiative heat exchange give rise to temporal fluctuations in the detector temperature. These fluctuations correspond to the noise level commonly referred to as the background fluctuation limit. Uncooled detectors available today can only remotely approach the background fluctuation noise limit.

Since the background fluctuation noise is the manifestation of temperature fluctuation noise when the radiation exchange is the dominant mode of heat exchange, Equation (25) can be rewritten for the background fluctuation limit, $NETD_{BF}$, as [6, 11, 42].

$$NETD_{BF} = \frac{8F^2}{\eta \tau_0 \left(\frac{dP}{dT} \right)_{\lambda_1 - \lambda_2}} \sqrt{\frac{2k_B \sigma_T B (T_D^5 + T_B^5)}{A_d}} \tag{26}$$

where T_D is the detector temperature and B is the measurements bandwidth.

5.2 Temperature Fluctuation Noise Limit

If the microcantilevers are in an evacuated package, the dominant heat loss mechanism will be due to thermal conduction along the legs of the microcantilever. Even if the microcantilever and the substrate to which it is

mounted are at exactly the same temperature, there is a dynamic heat exchange in both directions along the legs which will cause the microcantilever temperature to fluctuate. Only the temperature fluctuation noise should be taken into account when the radiative heat exchange between the detector and its surroundings is negligible compared to the thermal conduction. By combining Equations (19) and (24), we find that the temperature fluctuation noise limit, $NETD_{TF}$, is [43]

$$NETD_{TF} = \frac{8F^2\sqrt{2k_B BGT_D}}{\tau_0 \eta A_d \left(\frac{dP}{dT}\right)_{\lambda_1-\lambda_2}} \tag{27}$$

As can be concluded from Equation (27), it is important to minimize the heat exchange between the microcantilever and its surroundings in order to decrease the effect of the temperature fluctuation noise and improve the performance of a microcantilever IR detector.

5.3 Thermo-mechanical Noise Limit

As we have shown earlier the thermo-mechanical noise is one of the factors that can limit the performance of micromechanical IR detectors. By substituting the expression for rms of thermo-mechanical noise [Equation (15)] into Equation (25), we find that the thermo-mechanical noise limit, $NETD_{TM}$, is

$$NETD_{TM} = \frac{8F^2}{\tau_0 \eta A_d R_0 \left(\frac{dP}{dT}\right)_{\lambda_1-\lambda_2}} \sqrt{\frac{k_B T_D B}{kQ\omega_0}} \tag{28}$$

Another expression for the thermo-mechanical noise limit, $NETD_{TM'}$, can be found in analogous way by assuming the alternative model of the thermomechanical noise (Equation 23) that accounts for the damping caused by intrinsic friction

$$NETD_{TM'} = \frac{8F^2}{\tau_0 \eta A_d R_0 \left(\frac{dP}{dT}\right)_{\lambda_1-\lambda_2}} \sqrt{\frac{k_B T_D B}{kQ\omega}} \tag{29}$$

From the practical standpoint, it would be very important to determine which of the two models of the thermo-mechanical noise, Equations (20) and (23), is applicable to a particular type of a micromechanical thermal detector. Nevertheless, the importance of stiffer microcantilevers (higher k) and higher Q-factors is obvious despite the differences between the two models.

6. THE MODEL MICROCANTILEVER DETECTOR

The microcantilever fabricated for the present studies (see Figure 1), consists of a wider, outmost region ("head") connected by two narrow beams ("legs") to the base. Each of the legs consists of a bimaterial part and a thermal isolation region. The "head" and bimaterial parts of the legs of the microcantilever have a total thickness of 650 nm and are made of silicon nitride with an aluminum layer t_1 = 100 nm thick. The length and the width of the bimaterial parts are, respectively, l_b = 101 μm and w_b=3 μm. The thermal isolation regions consist of SiN_x. The thickness of SiN_x in the thermal isolation regions is t_3 = 250 nm. The length and the width of these regions is, respectively, l_i = 30 μm and w_i = 1.5 μm. Taking into account the design described above, t_1 and t_2 used in the Equation (1) are 100 x 10^{-9} m and 550 x 10^{-9} m, respectively. Tables 1 and 2 summarize parameters of the fabricated device and materials used for its fabrication. By substituting these geometrical parameters and the materials properties into Equation (1), we obtain a deflection of 42.5 nm per Kelvin. This parameter can be interpreted as the temperature responsivity of the detector, $Y_T = \Delta z/\Delta T$. By optimizing the relative thicknesses of SiN_x and Al the value of Y_T can be increased substantially. For example Y_T can be increased almost three-fold by creating a cantilever with t_1 =217 x 10^{-9} m and t_2 =433 x 10^{-9} m. The length and the area of the IR absorbing head is l_d = 66x10^{-6} m and A_d = 2.706x10^{-9} m^2, respectively.

Figure 5 shows experimentally measured and calculated noise spectral densities for the microcantilever IR detector shown in Figure 1. Our calculations were conducted using Equations (19), (20) and (23) and parameters mentioned above.

Table 1. Thermophysical properties of SiN_x, Al and SiO_2

	Young's modulus E (x 10^9 Pa)	Expansion Coefficient α (x 10^{-6} K^{-1})	Thermal Conductivity g (W m^{-1} K^{-1})	Heat Capacity c (x 10^6 J m^{-3} K^{-1})	Density ρ(x 10^{-3} kg m^{-3})
SiN_x	385	2.1	19	2.2	2.4
Al	70	25	237	2.43	2.7
SiO_2	68	0.75	1.1	1.47	2.2

Table 2. Mechanical parameters of the microcantilever thermal detector

	K (N/m)	M (kg)	Q	ω_0(rad/s)	f_0(Hz)
Air	0.01	$4x10^{-12}$	100	58,400	9,300
Vacuum	0.01	$4x10^{-12}$	2000	75,400	12,000

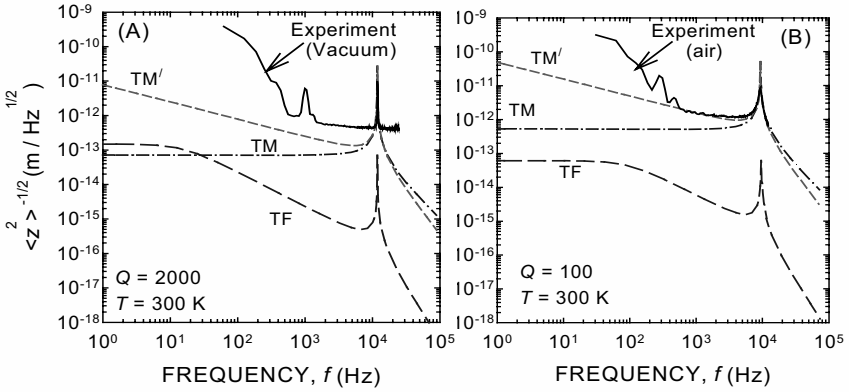


Figure 5. Experimentally measured and calculated noise spectra for the cantilever shown in Figure 1. Three series of calculations were conduced using, respectively, the model of temperature fluctuation noise and two alternative models of thermo-mechanical noise. Each series of these calculations was performed using the cantilever parameters for both (A) vacuum and (B) air. Note the following artifacts present in the experimental data. The peak around 120 Hz in the data for air results form interference with fluorescent ambient lighting. The peak at 1 kHz in the data for vacuum resulted from mechanical resonances of the mount in the vacuum chamber.

6.1 Calculations of *NETD* Limits

Equations (26), (27), (28) and (29) were used in order to evaluate the limits of *NETD* imposed by different noise mechanisms for the microcantilever IR detector used in the present studies. For this purpose, we established the values of the relevant parameters. The focal ratio F is defined as the reciprocal of twice the half angle of the marginal ray from the edge of the optics to the focal point. However, the focal point moves from the focal length of the optics when the target is at infinity, away from the optics as the target becomes nearer. Thus, the angle of the marginal ray decreases as the target approaches the imager.

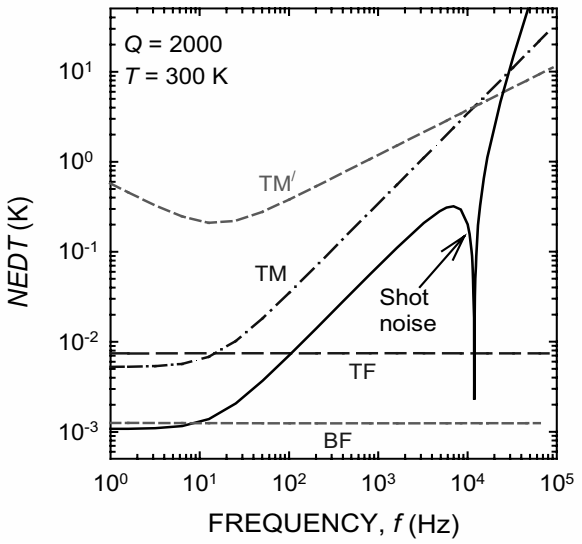


Figure 6. The values of NETD calculated as a function of frequency for different noise mechanisms. Parameters used in these calculations are summarized in Table 2. TM, TF, and BF denote curves corresponding to, respectively, thermo-mechanical, temperature fluctuation, and background fluctuation mechanisms.

A reasonable approximation is to assume that the target remains at infinity and that F (= f / D) is the *f/no*. of the optics, where f is focal length of the optics and D is diameter of the optics. The following values of parameters for our microcantilever are assumed: F=1, B=30Hz, T_D=T_B=300K, τ_o=0.90, η=0.6. It is also assumed that, similarly to many of other types of uncooled thermal detectors, the microcantilevers operate in the 8 μm to 14 μm atmospheric window, where $(dP / dT)_{8\_14:m} = 2.62$ W m^{-2} K^{-1}. We obtained $NETD_{BF}$ = 1.26 mK $NETD_{TF}$ = 7.4 mK, and $NETD_{TM}$ = 5.3 mK. Using the alternative model for the thermo-mechanical noise leads to an $NETD_{TM}$ that depends on the frequency and exhibits $f^{1/2}$ dependence. Assuming a frequency of 1 Hz to be the weighted center of the 30 Hz modulation frequency bandwidth, we determined $NETD_{TM}$,= 510 mK.

Table 3 summarizes the parameters obtained from in these calculations. The values of *NETD* are calculated as a function of frequency for different noise mechanisms and are plotted in Figure 6. Figure 7 shows the temperature fluctuation noise limit, $NETD_{TF}$, calculated for the detector shown in Figure 1 and plotted as a function of thermal conductance for 200, 300, and 400 K.

Table 3. Performance limits of the microcantilever infrared detector.

Figure of Merit	Value
$NETD_{BF}$	1.26×10^{-3} K
$NETD_{TF}$	7.4×10^{-3} K
$NETD_{TM}$	5.3×10^{-3} K
$NETD_{TM}^{/}$	510×10^{-3} K *
τ	10.2×10^{-3} s

* calculated for a 30 Hz bandwidth centered at 15 Hz

The plots in Figure 7 illustrate the importance of low values of G_T. The values of G_{rad} that are plotted in Figure 7 correspond to the radiative heat loss mechanism at the three temperatures of 200, 300 and 400 K. These values of G_{rad} represent the ultimate limit for the thermal isolation (for the corresponding temperatures) and are shown in Figure 7. On the other hand, heat loss dominated by conductance through air G_{air} is characteristic of microcantilever detectors suspended above the substrate at a microscopic distance in air. As can be seen in Figure 7 high values of G_{air} correspond to respectively high values of $NETD_{TF}$, which are higher at higher temperatures.

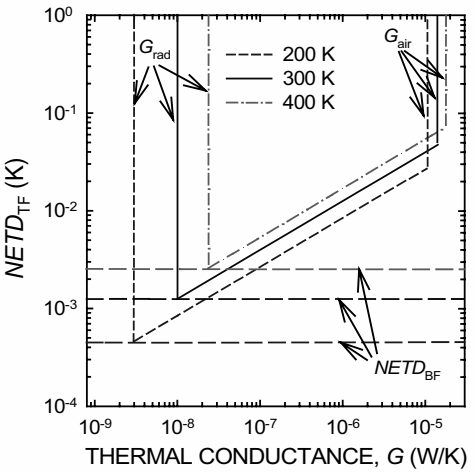


Figure 7. Temperature fluctuation limited $NETD_{TF}$ as a function of thermal conductance for temperatures of 200, 300 and 400 K. Also plotted are (i) the radiation limit of thermal conductance, (ii) the conductance through air only, and (iii) the values of $NETD_{BF}$ for the three temperatures.

6.2 Evaluation of "Optical Lever" Readout

The "optical lever" readout [54] is used extensively in AFM to detect changes in deflection or oscillation of a microcantilever probe [29]. When the laser beam is focused on the microcantilever, the reflected beam deflects in accordance to microcantilever bending. The amount of microcantilever bending can be measured by directing the reflected spot onto a position-sensitive (bi-cell) photodetector that consists of two photodiodes, PD_1 and PD_2. Our further evaluations will consider displacement, Δz, of the tip a bi-material microcantilever with length l due to thermal IR radiation falling upon it (Figure 3). The input parameters of this evaluation also include power of the probing laser beam, P_1, distance, L, between the microcantilever and the photodetector. Deflection of the microcantilever causes redistribution of the laser beam power fallen upon each of the two photodiodes. By assuming a square (*i.e.*, non-Gaussian) spatial distribution of power within

the laser beam, the difference between the power of light fallen upon each of the two halves of the photocell can be approximated as [29]

$$\Delta P_l = P_l \frac{4L}{ld} \Delta z \,, \tag{30}$$

where d is the diameter of the light spot projected on the photocell. This difference in photodiode illumination leads to a difference, Δi, between photocurrents of the two halves of the photocell [29]

$$\Delta i = \gamma P_l \frac{4L}{ld} \Delta z \tag{31}$$

where γ is the photocell responsivity in Amperes per Watts. Equation (31) shows that the gain of the optical read-out is proportional to the laser power, P_l, as well as the geometric factor $\beta = 4\ L\ /\ l$, and inversely proportional to the diameter, d, of the light spot projected on the photocell. A differential amplifier connected to the two halves of the bi-cell produces the output current, i_s, proportional to Δz [29]

$$i_s = \gamma P_l \frac{4L}{ld} \Delta z = \gamma P_l \frac{4L}{ld} P_0 R_z \tag{32}$$

Therefore, the thermal infrared power P_0 falling on the microcantilever can be calculated by measuring the photo-current i_s, if the photocell responsivity γ, the laser power P_l the distance L, the diameter of the light spot projected on the photocell, and the microcantilever length, l, the product of the microcantilever responsivity are known. If the responsivity, R_z is determined independently using a calibrated source, the incident infrared power, P_0, is

$$P_0 = \frac{i_s ld}{\gamma L P_l R_z} \tag{33}$$

The total responsivity of a microcantilever IR detector with the optical readout can be defined as the current measured at the output of the differential photocell divided by the incident infrared radiation power: $R_{ol} = i_s\ /\ P_0$. Then, there is a relationship between R_{ol} and R_z given by

$$R_{ol} = \frac{\gamma P_l \beta}{d} R_z \tag{34}$$

Equation (34) shows how the total responsivity, R_{ol}, is defined by the parameters of the optical readout. If the power of the probing beam is constant, higher responsivity, R_{ol}, can be achieved by decreasing the divergence of the readout optical beam.

6.3 Noise Contribution from the Optical Readout

As pointed out earlier, optical readout inevitably can add to the inherent noise of the microcantilever, although the added noise can be very small. Sarid [29] identified five main sources of noise associated with the optical readout process. Four of these are: (i) optical pathlength drift, (ii) Johnson noise in the photodetector, (iii) laser phase noise, and (iv) shot noise in the photodetectors due to the statistical nature of the rate of arrival at the photodetectors of the laser photons. Furthermore, both microcantilever bending in response to the measured IR radiation and temperature fluctuation or thermo-mechanical noise are amplified by the optical read out by a factor of β. According to Sarid [29], low frequency components of laser phase noise are reduced greatly at the differential amplifier due to good common mode rejection. Likewise, optical pathlength changes are not a significant noise source since they do not affect the differential signal in the direction perpendicular to the laser beam. Therefore, Johnson noise and shot noise in the photodetectors are the most important mechanisms of noise added by the optical readout. The mean square noise current $<\delta i^2_J>$ from each photodetector is [29]

$$< \delta i_J^{\,2} > = \frac{4 k_B T B}{r} \tag{35}$$

where r is the photodetector resistance. The differential amplifier adds in quadrature the mean square noise from the two photocells which doubles the noise. According to Sarid [29], however, the Johnson noise contribution is much smaller as compared to photodetector shot noise. Photodetector shot noise is sometimes referred to as light source shot noise, since it reflects an intrinsic feature of any light. Photonic shot noise is related to the statistical

nature of the rate of arrival at the photodetectors of the laser photons. It is a well-established phenomenon that manifests itself as fluctuation of the current output in the photodetector [29]

$$< \delta i_S{}^2 > = 2q\gamma B \frac{P_I}{2} \tag{36}$$

where $<\delta i^2{}_S>$ is the mean square fluctuation of photo-current in the differential photodetector due to photonic shot noise and q is the charge of an electron. Taking into account the relationship between a photodetector output current and a microcantilever tip displacement given by Equation (32), the shot noise can be expressed in terms of the microcantilever rms displacement

$$< \delta z_S{}^2 >^{1/2} = \frac{d}{\gamma}\sqrt{\frac{B}{\beta P_I}} \tag{37}$$

Equation (37) shows how the value of rms displacement equivalent to photon shot noise is influenced by the parameters of the optical readout. *NETD NETD NETD.*

6.4 Total Noise

Taking into account all the fundamental noise sources described above, we obtain a generalized expression for the total $NETD_T$

$$NETD = \frac{4F^2\left\langle \delta z_{TF}{}^2 + \delta z_{TM}{}^2 + \delta z_{sh}{}^2 \right\rangle^{1/2}}{\tau_0 \eta A_d R(\omega)\left(\frac{dP}{dT}\right)_{\lambda_1-\lambda_2}} = \sqrt{NETD_{TF} + NETD_{TM} + NETD_{NS}} \tag{38}$$

In order to evaluate Equations (36) and (38), appropriate values of the parameters need to be selected. For this purpose we use the same values for F, τ_o, η, B, T, and $(dP / dT)_{8\_14\ \mu m}$ as before. In addition, the following parameters of the optical readout are assumed: β = 6000, γ = 0.6 A/W,

$q = 1.6 \times 10^{-19}$ Cb, $d = 2.5 \times 10^{-3}$ m, and a laser power $P_l = 10^{-4}$ W. Using these values, the microcantilever responsivity, $R_{ol,}$ was calculated to be 950 A/W.

Using Equation (36), we found the contribution of shot noise at the differential output of the photocell to be $<\delta i^2_{S}>^{1/2} = 1.7 \times 10^{-12}$ A. The equivalent rms of microcantilever displacement calculated according to Equation (31) is $<\delta z^2_{N S}>^{1/2} = 1.13 \times 10^{-13}$ m. These results are summarized in Table 4. The $NETD_{NS}$ limited by shot noise is shown in Figure 7. Unlike the intrinsic mechanisms of microcantilever noise, the shot noise is frequency independent. As a result, $R(\omega)$ is the only frequency-dependent term in the expression for $NETD_{NS}$. As it follows from Equation (25), $NETD_{NS}$ is inversely proportional to $R(\omega)$. This explains a specific shape of the dependency of $NETD_{NS}$ on the modulation frequency shown in Figure 7. At low modulation frequencies values of $NETD_{NS}$ are well below *NETD* values defined by other noise mechanisms. As a result, there is a very small contribution of the optical readout into the total *NETD*.

Table 4. Performance of the microcantilever with optical readout, measured at the output of optical lever.

Figure of Merit	Value
Rol	950 A / W
$<\delta i^2_{N}>^{1/2}$	1.7×10^{-12} A
$<\delta z^2_{N}>^{1/2}$	1.13×10^{-13} m
$NETD_{NS}$	1.1×10^{-3} K
$NETD_{T}$	9.2×10^{-3} K

7. DISCUSSION

Several fundamental noise sources were identified as highly relevant to the performance of MEMS based IR detectors. The dynamic thermal equilibrium between the detector and its surroundings is one of these fundamental noise sources that manifests itself as small, but measurable, temporal fluctuations in the detector temperature. In a microcantilever IR detector, the temperature fluctuations manifest themselves as nanoscale motions of the

microcantilever end, *i.e.*, as fluctuations in displacement Δz. The amplitude of these oscillations increases when the detector becomes better thermally isolated from its surroundings with an inverse square root dependence on G_T. However, the signal response varies as G_T^{-1}, and therefore the ratio of the output signal to the temperature fluctuation noise improves as the thermal conductance between the detector and its surroundings decreases. This trend reaches its fundamental limit when the conductance is reduced to radiative heat exchange only. It should be emphasized that this regime (known as the background fluctuation limit) is not accessible by any of the uncooled IR detectors existing today.

Optimized thermal IR detectors can closely approach temperature fluctuation noise limit determined by heat loss due to thermal conduction through detector structural parts. Microbolometers developed by Raytheon with FPAs with individual elements of 25 μm^2 exhibited an *NETD* of 30 mK. Since *NETD* scales inversely proportional with the detector area, this *NETD* value corresponds to almost 8 mK for a detector with an area of 50 μm^2 [55].

If, however, the thermal detector is not in an evacuated package, the dominant heat loss mechanism is likely to be thermal conduction through the air surrounding the detector. The temperature fluctuation noise associated with this relatively high conduction through the air could limit the performance of any IR thermal detector below acceptable levels.

A noise mechanism specific to micro-mechanical devices is termed broadly as “thermo-mechanical” [29, 56] or “mechanical-thermal noise”[57]. While this type of noise is absent in the other types of thermal detectors, it plays a very important role in microcantilevers. The importance of thermomechanical noise was recognized in early studies of mechanical systems. Early attempts to study thermo-mechanical noise [58] can be traced back to the late 1920's when Gerlach [59] investigated the rotational Brownian motion of a small mirror suspended by a thin wire. Thermo-mechanical noise attracted more attention as the goal to reach performance limits in sensor technology was set [57, 60]. Gabrielson [57] cited examples that ranged from the mirrored galvanometer [59], condenser microphones, and gravitational-wave detectors [60].

More recently, thermo-mechanical models of noise have been discussed by a number of researchers [49-51, 53, 61]. Temperature fluctuations and thermomechanical vibrations in microcantilever devices is a major fundamental noise source in AFM [29]. Stowe *et. al*. [32] used a 60 nm thick microcantilever with a resonance frequency of 1.7 kHz and demonstrated atto-Newton force resolution that was only limited by intrinsic thermo-mechanical noise. Currently, however, there is no model that would unambi-

guously predict the frequency dependence of thermo-mechanical noise at frequencies well below the microcantilever mechanical resonances. For example, Savran et. al. [61] measured the spectral density of microcantilever motion in air and found that it is frequency-independent at frequencies below the resonance. However, Kajima et. al. [51] measured a noise spectrum showing a $f^{-1/2}$ dependence of the noise below resonance.

Our experimental results shown in Figure 5 seem to indicate that there is a noise component proportional to $f^{-1/2}$ at low frequencies. In fact, experimentally measured spectra tend be on the upper side of theoretical predictions for low frequencies. Although the observed frequency dependencies could be in part due to a superposition of ambient sources of noise below 1000 Hz, the "$f^{-1/2}$ dependence" in the frequency range of 1 to 6 kHz is less ambiguous and, most likely, related to thermo-mechanical noise.

In general, noise in micromechanical devices can be related to noise sources intrinsic to the device and noise sources arising from interactions with its environments (heat exchange, adsorption-desorption, and readout noise [56, 57, 62]. Determining applicability of different noise models is particularly important for designing more efficient micro and nanoscale mechanical systems [63].

It is interesting to note that the contribution of shot noise (from the optical readout) to the output signal [third term in Equation (38)], can be negligible compared to the temperature fluctuation noise and thermomechanical noise *i.e.*, the first and second terms in Equation (38). The rms noise current $<\delta i^2_S>^{1/2}$ is proportional to $P1^{1/2}$. However, the microcantilever responsivity R_{ol} increases linearly with P_1. Therefore, increasing the "read" laser beam power, P_1, above the chosen value of 0.1 mW would further improve the signal-to-noise ratio. For the calculations conducted in the present studies, we assumed that the diameter of the light spot projected on the photocell is 2.5×10^{-3} m. Reducing this size would increase the gain of the optical readout. Since this would affect both the intrinsic microcantilever noise and useful responses to the same degree, the resulting $NETD_{NS}$ would remain unaffected. Similarly, increasing the photocell responsivity γ or the displacement amplification β would have no effect on the $NETD_{NS}$. The value of the total $NETD_T$ listed in Table 4 is comparable to the one listed as the $NETD_{TF}$ and $NETD_{TM}$ values of Table 3. This is consistent with the fact that the contribution of the shot noise in the output signal of the optical readout is small as compared to the temperature fluctuation noise. In effect, the optical is almost a noiseless readout means. Therefore, the overall system comprised of the microcantilever and the optical can be temperature fluctuation noise limited and/or thermo-mechanical noise limited.

The microcantilever detector fabricated according to the selected design utilized SiN_x as the base material and Al as a bimaterial layer. We found that the intrinsic responsivity of the implemented microcantilever detector was $\Re$ = 0.066 m/W assuming the detector absorbance of 0.6. The background fluctuation noise limited $NETD_{BF}$ for the bimaterial microcantilever was found to be 1.3×10^{-3} K assuming f/1 optics and a 30 Hz frame rate. The temperature fluctuation noise limited $NETD_{TF}$ and the thermo-mechanical noise limited $NETD_{TM}$ were found to be, respectively, 7.4×10^{-2} K and 5.3×10^{-2} K. These values are attributable to the microcantilever itself and the IR focusing optics; they do not account for any noise introduced by the optical readout process. The contribution from the optical readout appeared to be quite insignificant compared to the intrinsic microcantilever noise. There is approximately a 2 % difference between the values of *NETD* calculated with and without taking into account the optical readout (shot) noise.

The performance of microcantilever thermal detectors with the parameters described in this work are limited primarily by temperature fluctuation and/or thermo-mechanical noise. However, the thermal isolation of the bimaterial microcantilever studied here is not fully optimized, *i.e.*, the value of conductance $G = 3.85 \times 10^{-7}$ W/K is large compared to the respective parameter of the advanced microbridge bolometers. Therefore, the performance of micromechanical IR detectors can be further improved by using fabrication materials and designs that provide stiffer (higher k) microcantilevers and better thermal isolation. For example, using SiO_2 as the material for thermal isolation will reduce the value of conductance of the detector G_T, to about 2×10^{-7} W/K, *i.e.* about twenty times smaller compared to the detector of the same geometry made from SiN_x.

An important question is whether the *NETD* of microcantilever IR detectors are comparable with those of silicon microbridge bolometers employing VO_x as the resistive material. Because *NETD* is inversely proportional to the sensitive (absorbing) area, it is important to compare devices with similar areas. The area of the bimaterial microcantilever studied here is 2.71×10^{-9} m^2, which is close to that of pixels in a typical bolometer FPA (2.5×10^{-9} m^2). Over the past few years the reported *NETD*s of bolometer FPAs with a 46 μm pitch have decreased from over 50 mK down to only 25 mK [64]. More recently, a value of 35 mK was reported for a 320x240 FPA of microbolometers with a detector pitch size of only 25 μm [5]. This was achieved primarily due to advances in the resistive readout and optimization in the detector thermal isolation. Similarly, these design parameters should be optimized in order to improve the performance of microcantilever IR detectors and FPAs based on them.

Further optimization of microcantilever IR detectors requires minimization of both temperature fluctuation and thermo-mechanical noise. In particular, thermo-mechanical noise in microcantilever IR detectors can be greatly reduced by increasing the microcantilever stiffness and Q-factor. The adverse effect of the temperature fluctuation noise can be reduced by increasing the detector thermal isolation. This could be achieved by changing the geometry (lengthening, narrowing or thinning) of the thermal isolation region or using materials with lower thermal conductivity, such as SiO_2. It is important to emphasize, however, that the highest practically useful thermal isolation is likely to be limited by the requirement of fast response times in the case of real-time imaging. In fact, the thermal response time of the detector is inversely proportional to the detector thermal conductance and directly proportional to the detector heat capacity. The thermal response time of the microcantilever thermal detector investigated in this work is 10.2 ms which is satisfactory since 30 frames per second operation typically requires the response time to be shorter than 12 ms.

8. CONCLUSIONS

With the development of thermal detectors, a challenge arises in the efforts to define parameters that both applicable to uncooled IR detectors and consistent with parameters used to describe cooled photon detectors. In this work we provided a detailed analysis of figures of merit that apply to all IR detectors and evaluated them in more detail for the case of thermo-mechanical detectors based on MEMS. Our analysis identified several factors that determine the performance of uncooled IR detectors based on bimaterial microcantilevers. The common factors that limit the performance of microcantilever and other thermal detectors include temperature fluctuation noise and its manifestation, background fluctuation noise. Taking into account the mechanical nature of microcantilever detectors, we also evaluated the limitations imposed by the thermo-mechanical noise. In addition to generalized theoretical evaluations, we conducted both theoretical and experimental studies that focused on one specific design of a bimaterial microcantilever with the optical readout.

The performance of uncooled IR systems is limited by the detector intrinsic noise. Fundamental mechanisms of heat exchange and dissipation induce spontaneous temperature fluctuations of all thermal IR detectors and impose two important fundamental limits to their performance, referred to as the background fluctuation noise limit and the temperature fluctuation noise

limit. The performance of a thermal IR detector can reach its absolute theoretical limit when the thermal isolation between the detector and its surrounding is so high that the dominant heat loss mechanism is radiation exchange between the detector and its surrounding. Although the thermal conductance of the support structure can be almost infinitely reduced, this will also result in a much longer thermal response time, which is inversely proportional to the thermal conductance. Optimization of thermal IR detectors may, therefore, involve a tradeoff between an acceptable time-constant and their responsivity. Since the response time is directly proportional to the heat capacity, the limitations of this tradeoff can be eliminated in part by reducing the heat capacity of the detector. Although the fundamental limits are independent of readout means, they do depend on the intrinsic gain of the detector. Theoretical relationships between the figures of merit and the detector parameters can be used to rationally design thermal IR detectors with optimized performance.

As can be concluded from the present work, uncooled microcantilever thermal detectors have the potential to achieve *NETD* values already demonstrated by other uncooled thermal detectors and achieve sensitivity levels of resistive microbridge bolometers. The models discussed in the present work provide a clear guidance for designing and fabricating microcantilever thermal detectors that are limited mainly by temperature fluctuation noise. More specifically, this guidance dictates that the proper selection of the materials and geometries should provide a combination of high thermal isolation, high stiffness (spring constant) and finesse (Q-factor) of the microcantilever structure. In general, the analyzed figures of merit indicate that, in the nearest future, microcantilever thermal detectors will likely approach the performance of currently available resistive microbridge bolometers with *NETD* values in the range of 40 to 60 mK.

ACKNOWLEDGMENTS

We would like to acknowledge support from the Defense Advanced Research Projects Agency, the National Science Foundation and U.S. Department of Energy. This work was partially supported by the Laboratory Director's Research and Development Program of Oak Ridge National Laboratory. Oak Ridge National Laboratory is operated for the U.S. Department of Energy by UT-Battelle under contract DE-AC05-96OR22464.

REFERENCES

[1] Mao M., Perazzo T., Kwon O., Zhao Y., Majumdar A., Varesi J., Norton P. Infrared Vision Using an Uncooled Thermo-Opto-Mechanical Camera; Design, Microfabrication, and Performance. Microelectromechanical Systems, Lee, Y. C., K. Goodson, R.S. Keynton, A. Lee, L. Lin, F.K. Forster, ed. Nashville, TN, The American Society of Mechanical Engineers, 1999.

[2] Datskos P. G., Rajic S., Senesac L. R., I. Datskou. Quantum Well Microcantilever Photon Detectors. Ultramicroscopy 2001; 86: 191.

[3] Balcerak R. S. Uncooled IR Imaging: Technology for the Next Generation. Infrared Technology and Applications XXV, SPIE 1999; 3698: 110.

[4] Howard P. E., Clarke J. E., Bradley M. G., Ionescu A. C., Li C. Progress in Uncooled Focal Plane Sensor Technology at Boeing. SPIE - Infrared Technology and Applications XXVI, Andresen, B., Fulop, G., Strojnik, M., ed. SPIE, 2000.

[5] Murphy D. F., Ray M., Wyles R., Asbrock J. F., Lum N. A., Wyles J., Hewitt C., Kennedy A., Van Lue D., Anderson J. S., Bradley D., Chin R., Kostrzewa T. High-Sensitivity 25-Micron Microbolometer Fpas. SPIE-Infrared Detectors and Focal Plane Arrays VII, Dereniak, E. L. and Sampson, R. E., ed. SPIE, 2002.

[6] Kruse P. W. "Principles of Uncooled Infrared Focal Plane Arrays." In *Uncooled Infrared Imaging Arrays and Systems, Semiconductors and Semimetals, 47*, Kruse, P. W., and D.D. Skatrud, ed. San Diego: Academic Press, 1997.

[7] Kayes R. J., *Optical and Infrared Detectors.* Berlin: Springer-Verlag, 1977.

[8] Miller J. L., *Principles of Infrared Technology.* New York: Van Nostram Reinhold, 1994.

[9] Rogalski A. New Trends in Infrared Detector Technology. Infrared Phys. Technol. 1994; 35: 1.

[10] Hanson C. Uncooled Thermal Imaging at Texas Instruments. Infrared Technology XXI, SPIE 1993; 2020: 330.

[11] Kruse P. W. A Comparison of the Limits to the Performance of Thermal and Photon Detector Imaging Arrays. Infrared Phys. Technol. 1995; 36: 869.

[12] Butler N., Blackwell R., Murphy R., Silva R., Marshall C. Low-Cost Uncooled Microbolometer Imaging System for Dual Use. Infrared Technology XXI, SPIE 1995; 2552: 583.

[13] Wood R. A. Uncooled Thermal Imaging with Monolithic Silicon Focal Planes. Infrared Technology XXI, SPIE 1993; 2020: 322.

[14] Wood R. A., Foss N. A. Micromechanical Bolometer Arrays Achieve Low-Cost Imaging. Laser Focus World 1993; June: 101.

[15] Wood R. A. "Monolithic Silicon Microbolometer Arrays." In *Uncooled Infrared Imaging Arrays and Systems, Semiconductors and Semimetals*, Kruse, P. W. and Skatrud, D. D., ed. San Diego: Academic Press, 1997.

[16] Datskos P. G., Oden P. I., Thundat T., Wachter E. A., Warmack R. J., Hunter S. R. Remote Infrared Detection Using Piezoresistive Microcantilevers. Appl. Phys. Lett. 1996; 69: 2986.

[17] Datskos P. G., Rajic S., Datskou I. Photoinduced and Thermal Stress in Silicon Microcantilevers. Applied Physics Letters 1998; 73: 2319.

[18] Amantea R., Knoedler C. M., Pantuso F. P., Patel V. K., Sauer D. J., Tower J. R. An Uncooled IR Imager with 5 mK NETD. Infrared Technology and Applications XXIII, Andresen, B. F. and Scholl, M. S., ed. Proc. SPIE, 1997.

[19] Amantea R., Goodman L. A., Pantuso F., Sauer D. J., Varhese M., Villianni T. S., White L. K. Progress Towards an Uncooled IR Imager with 5 mK NETD. Infrared Technology and Applications XXIV, Andresen, B. F. and Scholl, M. S., ed. Proc. SPIE, 1998.
[20] Lai J., Perazzo T., Shi Z., Majumdar A. Optimization and Performance of High-Resolution Micro-Optomechanical Thermal Sensors. Sensors and Actuators 1997; 58: 113.
[21] Perazzo T., Mao M., Kwon O., Majumdar A., Varesi J. B., Norton P. Infrared Vision Using Uncooled Micro-Optomechanical Camera. Appl. Phys. Lett. 1999; 74: 3567.
[22] Senesac L. R., Corbeil J. L., Rajic S., Lavrik N. V., Datskos P. G. IR Imaging Using Uncooled Cantilever Sensors. Ultramicroscopy 2003; 97: 451.
[23] Wachter E. A., Thundat T., Datskos P. G., Oden P. I., Sharp S. L., Warmack R. J. Remote Optical Detection Using Microcantilevers. Rev. Sci. Instrum. 1996; 67: 3434.
[24] Oden P. I., Datskos P. G., Thundat T., Warmack R. J. Uncooled Thermal Imaging Using a Piezoresistive Microcantilevers. Appl. Phys. Lett. 1996; 69: 3277.
[25] Oden P. I., E.A. Wachter, P.G. Datskos, T. Thundat, and R.J. Warmack. Optical and Infrared Detection Using Microcantilevers. Infrared Technology XXII, SPIE 1996; 2744: 345.
[26] Kenny T. W., Reynolds J. K., Podosek J. A., Vote E. C., Miller L. M., Rockstad H. K., Kaiser W. J. Micromachined Infrared Sensors Using Tunneling Displacement Transducers. Rev. Sci. Instrum. 1996; 67: 112.
[27] Datskos P. G. Micromechanical Uncooled Photon Detectors. Photodetector Materials and Devices V, SPIE 2000; 3948: 80.
[28] Tortonese M., Barrett R. C., Quate C. F. Atomic Resolution with an Atomic Force Microscope Using Piezoresistive Detection. Appl. Phys. Lett. 1996; 62: 834.
[29] Sarid D., *Scanning Force Microscopy.* New York: Oxford University Press, 1991.
[30] Ohnesorge F., Binning G. True Atomic Resolution by Atomic Force Microscopy through Repulsive and Attractive Forces. Science 1993; 260: 1451.
[31] Hoh J. H., Cleveland J. P., Prater J.-P., Hansma P. K. Quantized Adhesion Detected with Atomic Force Microscope. J. Am. Chem. Soc. 1992; 114: 4917.
[32] Stowe T. D., Yasumura K., Kenny T. W., Botkin D., Wago K., Rugar D. Attonewton Force Detection Using Ultrathin Silicon Cantilevers. Appl. Phys. Lett. 1997; 71: 288.
[33] Shaver P. J. Bimetal Strip Hydrogen Gas Sensor. Rev. Sci. Instrum. 1969; 40: 901.
[34] Timoshenko S. P. Analysis of Bi-Metal Thermostats. J. Opt. Soc. Am. 1925; 11: 233-255.
[35] Barnes J. R., Stephenson R. J., Woodburn C. N., O'Shea S. J., Welland M. E., Rayment T., Gimzewski J. K., Gerber C. Femtojoule Calorimeter Using Micromechanical Sensors. Rev. Sci. Instrum. 1994; 65: 3793.
[36] Hu Z. Y., Thundat T., Warmack R. J. Investigation of Adsorption and Absorption-Induced Stresses Using Microcantilever Sensors. J. Appl. Phys. 2001; 90: 427-431.
[37] Williams C. S., Becklund O. A., *Optics: A Short Course for Engineers & Scientists.* New York: Wiley-Interscience, 1972.
[38] Dereniak E. L., Crowe D. G., *Optical Radiation Detectors.* John Wiley & Sons, 1984.
[39] Dereniak E. L., Boreman G. D., *Infrared Detectors and Systems.* New York: Wiley and Sons, 1996.
[40] Kruse P. W., McGlauchlin L. D., McQuistan R. B. "Photon Noise Limitations of Thermal Detectors." In *Elements of Infrared Technology*, ed. New York: John Wiley and Sons, Inc., 1962.

[41] Kruse P. W. A Comparison of the Limits to the Performance of Thermal and Photon Detector Imaging Arrays. Infrared Phys. Technol. 1994; 34: 25.
[42] Kruse P. W. Uncooled IR Focal Plane Arrays. Infrared Technology XXI, SPIE 1995; 2552: 556.
[43] Kruse P. W., *Uncooled Thermal Imaging, Arrays Systems and Applications.* Bellingham, WA: SPIE Press, 2001.
44. Holst G. C., *Testing and Evaluation of Infrared Imaging Systems.* Bellighasm WA: SPIE, 1998.
[45] Vincent J. D., *Fundamentals of Infrared Detector Operation and Testing.* New York: John Wiley & Sons, 1989.
[46] Lloyd J. M., *Thermal Imaging Systems.* New York: Plenum Press, 1975.
[47] Kruse P. W., McGlauchlin L. D., McQuistan R. B., *Elements of Infrared Technology.* New York: John Wiley and Sons, 1962.
[48] Callen H. B., Welton T. A. Irreversibility and Generalized Noise. Physical Review 1951; 83: 34.
[49] Saulson P. R. Thermal Noise in Mechanical Experiments. Physical Review D 1990; 42: 2437.
[50] Majorana E., Ogawa Y. Mechanical Thermal Nnoise in Coupled Oscillators. Physics Letters 1997; A233.
[51] Kajima M., Kusumi N., Moriwaki S., Mio N. Wide-Band Measurement of Mechanical Thermal Noise Using a Laser Interferometer. Phys. Lett. A 1999; 264: 251-256.
[52] Zhao Y., Mao M. Y., Horowitz R., Majumdar A., Varesi J., Norton P., Kitching J. Optomechanical Uncooled Infrared Imaging System: Design, Microfabrication, and Performance. J. Microelectromech. Syst. 2002; 11: 136-146.
[53] Djuric Z. Mechanisms of Noise Sources in Microelectromechanical Systems. Microelectronics Reliability 2000; 40: 919.
[54] Meyer G., Amer N. M. Novel Optical Approach to Atomic Force Microscopy. Appl. Phys. Lett. 1988; 53: 1045.
[55] Kruse P. W., *Private Comminication.* ed. 2003.
[56] Cleland A. N., Roukes M. L. Noise Processes in Nanomechanical Resonators. J. Appl. Phys. 2002; 92: 2758-2769.
[57] Gabrielson T. B. Mechanical-Thermal Noise in Micromachined Acoustic and Vibration Sensors. IEEE Trans. Electron Devices 1993; 40: 903-909.
[58] Uhlenbeck G. E., Goudsmit S. A Problem in Brownian Motion. Physical Review 1929; 34: 145.
[59] Gerlach W. Naturwissenschaften 1927; 15: 15.
[60] Cocia E., Fafone V. Noise Attenuators for Gravitational Wave Experiments. Nuclear Instruments and Methods in Physics Research A 1995; 366: 395.
[61] Savran S. A., Sparks A. W., J. Sihler, Li J., Wu W.-C., Berlin D. E., Burg T. P., Fritz J., Schmidt M. A., Manalis S. R. Fabrication and Characterization of a Micromechanical Sensor for Differential Detection of Nanoscale Motions. J. Microelectromech. Syst. 2002; 11: 703.
[62] Gabrielson T. B. Fundamental Noise Limits for Miniature Acoustic and Vibration Sensors. J. Vib. Acoust.-Trans. ASME 1995; 117: 405-410.
[63] Masoller C. Noise-Induced Resonance in Delayed Feedback Systems. Phys. Rev. Lett. 2002; 88: 34102.
[64] Murphy R., Kohin M., Backer B. S., Butler N. R., Blackwell R., Allen T. Recent Developments in Uncooled IR Technology. Infrared Detectors and Focal Plane Arrays VI, Dereniak, E. L. and Sampson, R. E., ed. SPIE, 2000.

Chapter 13
BULK ACOUSTIC WAVE SENSORS IN CHEMICAL ANALYSIS

Maria Teresa S. R. Gomes
University of Aveiro, Portugal

Abstract: Bulk acoustic wave devices have been developed for the detection and quantification of a large number of compounds. Most of those works do not include an application to real samples and are poor characterized from an analytical chemistry point of view. An overview of the analytical requirements for a real application is reviewed and usual problems frequently faced by the analyst using thickness shear mode devices are presented, as well as some solutions and illustrative practical examples taken from the literature.

Keywords: bulk acoustic wave devices, thickness shear mode resonators, quartz crystal microbalance, chemical analysis.

1. INTRODUCTION

The applications and demands for chemical sensors reach all areas of human activity, as well as the earth and out of space environment. Reliable, simple, low cost instruments prone to be used by non-specialists would be the ultimate goal for any company or researcher. Unfortunately, there are not so many cases of success in the market of chemical sensors. However, a few examples clearly show that it is possible to develop reliable sensors to be used in a large scale: One million of Tagushi sensors are fabricated every year, and there is not a single chemist that had never used a pH electrode.

The number of publications on mass sensors shows their importance and how the investigators believe that they can be an answer to solve analytical problems. There are unfortunately not so many papers reporting applications analysing real samples, and most of the work deal only with standards. Validity of calibration curves is also ignored and the real performance of the new reported sensors is not evaluated. The lack of chemical evaluation comes often from the fact that the authors give no value to it. Often it is said that the aim of the sensor is just to detect changes and trends and not to do chemical analysis. This argument has some weakness from the analytical chemistry point of view, and renders its practical utility questionable. A few

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 421-445.

questions remain unanswered: Are those trends related to real changes in the sensor chemical environment, or are just reflexes of baseline drift? Is the sensor sensitive and fast enough to detect a change in concentration significant for the purpose in study? How long does calibration stand? Are there species other than the analyte to which the sensor responds?

The development of the sensor must take in account those questions, and strategies to improve the signal to noise ratio, detection limit, accuracy and precision need to be seek. This chapter is not an overview of the problems, which are multidisciplinary and would benefit with the point of view of physicists, chemists, mathematicians and engineers, and is also not a list of recipes, but a modest reflection around these problems based on the author own experience.

2. BULK ACOUSTIC WAVE DEVICES

A variety of acoustic wave devices can be manufactured with piezoelectric substrates or oriented voltage pooled films or materials. Different crystal orientation, thickness and geometry of the electrodes allow the design of a variety of devices using several acoustic wave types. This chapter will deal just with bulk acoustic wave devices, but surface wave and plate wave sensing devices can also be produced.

Most of the bulk acoustic wave devices are made of a piezoelectric substrate, usually quartz, onto which metal electrodes are deposited on both sides. When a voltage is impressed between both electrodes, the quartz crystal deforms. Connecting the quartz crystal to an appropriate oscillator [1,2] makes it to oscillate with a frequency, which is dependent on the thickness of the quartz wafer. The resonant frequency of the quartz crystal is remarkable constant and is the reason for its use in high accurate and precise instruments as clocks and counters.

Figure 1 shows a schematic diagram of the quartz wafer, wave motion and particle displacement.

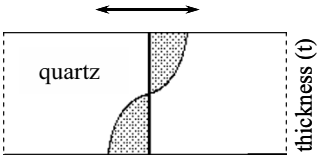


Figure 1. Quartz Wafer, Wave Motion and Particle Displacement.

Sauerbrey [1] and Lostis [2] found that whenever a small mass was attached to the electrodes, a proportional frequency decrease was observed. The principle that allowed the crystals to be used as very precise balances

was found. Sauerbrey derived an equation that relates the frequency decrease (ΔF) with the mass of a uniformly distributed deposited film (ΔM_f):

$$\frac{\Delta F}{F} = -\frac{\Delta M_f}{tA\rho_q} \tag{1}$$

for a quartz crystal with density ρ_q, thickness t, fundamental resonance frequency F, and electrode area A.

Sauerbrey equation is still used to calculate the mass deposited onto the crystal electrodes, in spite of the fact that it is only valid for thin rigid layers, within a relation for the mass to unit area of the film to quartz not larger than 0.02 [5-7]. These restrictions come from the assumptions made during its derivation, namely that the added mass behaves like quartz and does not experience a shear deformation.

After Sauerbrey, several investigators attempted to derive other equations which could be used with larger mass loads and that took into consideration the properties of the deposited film.

Behrndt [8] showed that measuring the period of the crystal oscillation, instead of the frequency, allowed to increase the working areal density load range up to 10% of the one of quartz [9-11].

Miller and Bolef [12] followed by Lu and Lewis [13] treated, for the first time, the film and the crystal as a composite resonator structure. Lu and Lewis [13] derived the following equation:

$$\tan\left(\frac{\pi F_c}{F_q}\right) = -\frac{\rho_f v_f}{\rho_q v_q}\tan\left(\frac{\pi F_c}{F_f}\right), \tag{2}$$

where F_c is the frequency of the composite resonator, F_q the resonant frequency of quartz, F_f the resonant frequency of the deposited film, ρ_f and ρ_q the density of the film and the quartz, respectively, v_f and v_q are the shear-wave velocities in the film and quartz, respectively. This equation shows that the frequency depends on the ratio

$$z = \frac{\rho_f v_f}{\rho_q v_q}, \tag{3}$$

which is the ratio between the shear – mode acoustic impedance of the deposited material and quartz. The "z-match" devices based on the last equation are more exact than the previous ones and, according to Benes, experimentally proved to be able to handle areal density ratio for film over quartz of 0.7 [9, 11, 14]. However, practical application is not simple as both

density and shear modulus of the deposit must be known. The situation is even more complex as, often, the z ratio for the thin film is different from that of the bulk material [9]. Benes [9, 11, 14] has made attempts to overcome this problem, proposing an auto-z-match method of estimating the z ratio by measuring two quasiharmonically related resonant frequencies, but there are left a few theoretical as well as practical problems [6, 10].

Mecea and Bucur [5, 6] developed a more complex theory, based on a compound resonator idea formed by the quartz crystal and the deposited film, which is known by the energy transfer model. The main novelty comes from the idea that the adherent deposit film on the quartz does not have its own vibrational frequency, as in Miller and Bolef model, but vibrates synchronously with the quartz. This model predicts that the quartz resonator is strongly affected by the viscosity of the deposited film, and can be applied not just for sensors operating in gas phase but also for crystals in liquid phase.

The possibility of extending the use of quartz crystal sensors to liquid media was only shown to be possible in 1980 [15, 16]. Density (d), conductivity, and viscosity (η) of the solutions were shown to affect the frequency of the crystal [15, 17] and Yao and Zhou [18] verified that for mixtures of water / organic solvents, the frequency depended also on the dielectric constant of the liquid. The first physical models were developed in 1985 by Bruckenstein and Shay [19], and Kanasawa and Gordon [20]. Bruckenstein and Shay [19] used dimensional analysis to develop the following equation:

$$\Delta F = -2.26 \times 10^{6} n f^{\frac{3}{2}} (\eta_l d_l)^{\frac{1}{2}}, \tag{4}$$

where n = 1 or 2, depending on whether one or two faces of the crystal are in contact with the solution. Kanasawa and Gordon [20] derived an expression similar to the one of Bruckenstein and Shay, where the relevant parameters influencing the frequency were also shown to be density and viscosity.

These early studies clearly showed that, besides attached mass, the frequency depended on the properties of the solution. This opened new opportunities as it could be easily used to detect contamination of starting materials in pharmaceutical formulations [21]. Also the monitoring of bacteria growth can be made recording the frequency of a quartz crystal while nutrients are consumed and metabolites are produced [22, 23], with application in toxicology studies [23].

Unfortunately, the application of a simple equation, capable of predicting the frequency shift produced by known density and viscosity changes, or the inverse, the prediction of the properties of the solution after frequency recording, is not possible. The problem is even more complex as Kurosawa

et al. [24] showed that the circuit influenced the frequency changes measured in solutions and that the equations of Bruckenstein and Shay and Kanasawa and Gordon did not hold for electrolyte or polymer solutions.

Other factors have been recognised to influence frequency. Schumacher et al. [25-27] introduced surface roughness, and frequency became also a function of the mass per unit area of the liquid confined in the cavities of the roughness surface. Heusler at al. [28] considered surface stress influence on the resonant frequency, and Thompson et al. [29, 30] considered the effect of changes in interfacial surface structure and interfacial properties such as free energy and slippage in the behaviour of a crystal in contact with a liquid.

Other equations have been developed and an overview of the main theories can be found in the work of Thompson et al. [30]. Besides the assumptions and approximations made on the several theoretical models, other factors such as mechanical clamping, clamping in the electrical circuit, and temperature, also affect the absolute accuracy. For these reasons it is necessary to use calibration curves in quantitative work.

3. CHEMICAL MASS SENSORS

A chemical sensor is a device able to continuously monitor specific chemical species. Janata [31] clearly distinguishes between a chemical sensor and a chemical sensing system, which performs analysis by discrete steps.

Any chemical specie that interacts with the electrodes of the piezoelectric crystal, changing its mass, to speak of the best known sensing mechanism, can be detected. A good example is the amalgamation of mercury vapour onto gold electrodes [32, 33]. Whenever the analyte does not directly interact with the metal of the electrodes, it is sometimes possible to quantitatively convert it into a detectable substance. This was successfully attempted by Guilbault et al., and Gomes et al. in the quantification of SO_2 [34, 35], and cyanide [36]. Those compounds were indirectly quantified as mercury, as the formation of Hg (II) complexes promotes mercury disproportionation with the consequent amalgamation of the mercury vapour onto the gold electrodes.

Whenever no detectable changes are produced on the electrodes, the analyst is forced to coat the electrode with a chemical substance, which does interact with the analyte. For instance, amines are known to interact by an acid-base reaction with CO_2, [37] and host-guest molecules and zeolites possess cavities capable of accommodate certain ions or molecules [38, 39].

Figure 2 shows schematically the changes in the frequencies of the crystal due to coating and analyte interaction.

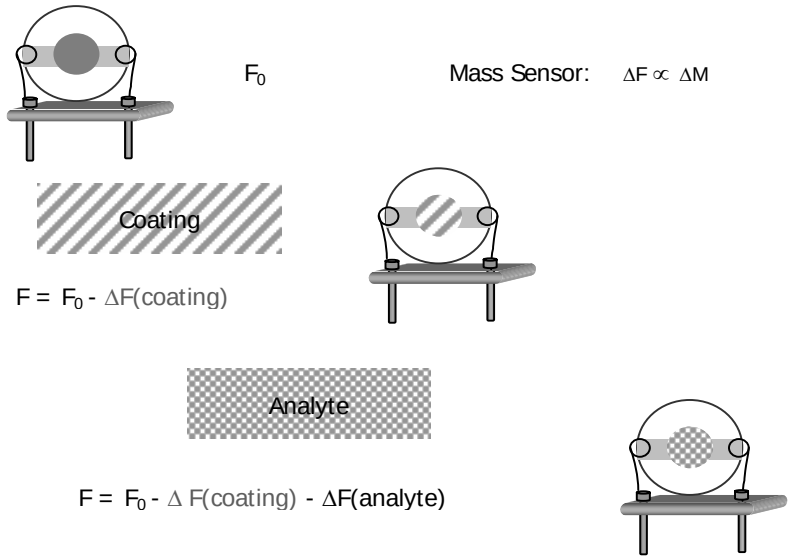


Figure 2. Changes in the Frequency of the Quartz Crystal Due to Coating and Analyte Interactions.

As a true sensor must provide continuous information, the interaction between the analyte and the electrode, coated or not, needs to be reversible. As Edmonds says [40], reversibility, as beauty, lies on the eye of the beholder, and in many situations recovery of the sensor to baseline is slower than the time needed to obtain the response signal [41], although when both processes are governed by diffusion both rates can be identical [42]. The time required for response and complete recover must be shorter than the process control time constant. Anyway, it is out of question to use a sensing element, which is not reversible within the time of the analysis. In analytical applications, where confidence of the analysis needs to be known, individual calibration of each sensor cannot be avoided, especially whenever irreproducible coating methodologies are applied and sensitivity of each sensor is unknown.

Weak interactions between the analyte and the electrode are desirable from the point of view of reversibility. On the other hand, chemical reactions, where molecular bonds need to be formed and broken, and where energies involved are so high (300 kJ mol^{-1} per bond) that in practice we are in the presence of irreversible interactions, tend to be more specific [43]. Examples of the first case are easy to find, and there are in the literature a large number of sensors employing gas chromatographic stationary phases as sensitive layers. The analytical signal has its origin in absorption phenomena and selectivity is governed by polarity similarity. It is well known in organic chemistry the rule “like dissolves like”, but differences between partition

coefficients between different gases and the coating tend to be slight and selectivity of these sensors are poor. Weak interactions governed by van der Waals forces are examples of reversible interactions. In the case of adsorption, the energies involved can go up to 40 kJ mol^{-1}, ranging from van der Waals' to acid-base interactions. Energies involved in coordination chemistry are between adsorption and covalent bond formation and could be regarded as a trade off between reversibility and selectivity [43]. Examples of charge transfer complex formation based sensors are those involving the interaction of gases as NO_2 and a metal ion surrounded by organic ligands, as in metallophtalocyanines [44]. Selectivity can be tailored by choosing the central metal, usually based on hard and soft acids and bases concept, and choosing the ligands, based on electronegativity or steric effects. Di Natale et al. presented a set of different metalloporphyrins with different selectivity degree to several organic compounds [45-47].

The search for a suitable coating is one of the most important and sometimes difficult jobs during the development of a sensor. Besides reversibility, the coating must be stable, it must not be volatile, or swallow, decompose, change its shape or spread along time. Very promising coatings can sometimes cause unexpected problems if they are very viscous and are applied for instance with an airbrush, as small drops spread along time causing tremendous changes in sensitivity [48].

Figure 3 shows schematically the basic requirements for the sensing layer.

As acoustic wave sensors do not respond only to mass, the terms mass sensors, and quartz crystal microbalances, although in use, can be very restrictive regarding the sensing possibilities. Other terms as bulk acoustic wave and thickness shear wave devices are more general.

Besides detecting changes in both density and viscosity of solutions it also detects changes in the morphology of the film and phase changes [48-51]. All phenomena that can produce a detectable and reproducible signal can be valuable for analytical purposes. Combined phenomena, for instance mass and phase changes [52], density and viscosity changes [23, 53] can be a gold mine for analytical chemists. As long as linearity between signal magnitude and concentration of the analyte can be assured for any of the contributions, there is no need to attempt to separate them. In fact, the analyst will be pleased to have reproducible phenomena that allow enhancing and increasing the gravimetric analytical signal. The sign of the response is also not important from the analytical chemistry point of view. It can be a frequency decrease, increase, or even have both negative and positive contribution. The analyst will gladly use the largest signal he can get, as long as it can be correlated with the analyte concentration and is reproducible.

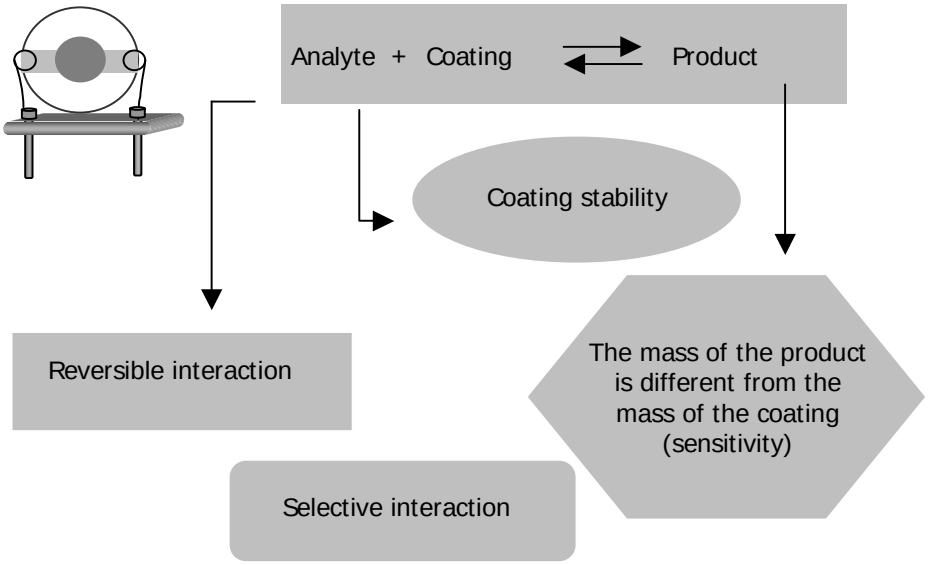


Figure 3. Main Requirements Governing Coating Selection.

Table 1 shows examples of sensors based on different sensing mechanisms.

Table 1. Examples of Thickness Shear Wave Sensors Based on Several Transducing Physical Properties.

Transducing Property	Analyte	Ref.
mass	carbon dioxide amine	54 55
density and viscosity	ethanol bacteria growth	53 23
mass and phase transition	carbon dioxide	52

4. EXPERIMENTAL LAYOUT

Experiments can be performed in batch or in a flowing injection system. Operation in batch implies that after each sample, or standard introduction, stationary stage must be attained before frequency recording takes place. In flow injection analysis (FIA) the minimum frequency is read and reproducibility is guaranteed by operating at a constant flow. Both

methodologies need baseline stability, but the batch experiments need longer stability of baseline, because one must be sure that reading is constant before recording it, or that it changes with the same rate as baseline. FIA offers the advantage of simple and easy recording of the frequency with precision, as there is no doubt in pointing out the minimum frequency, even if baseline frequency experiences a steady frequency increase, as far as its changing rate is slower enough to assure it did not change during the time needed to record the frequency shift. FIA experiments tend to be faster and easy to perform than experiments in batch, and, for this reason, only experimental layouts for FIA methods will be present.

Figure 4 shows the experimental layout used for gas analysis. A constant nitrogen flow is maintained in the system. This stream of nitrogen carries both the standards and samples, injected through a Teflon septum, to the cell housing the quartz crystal. The cell divides the stream in two parts, each one directed to the centre of the quartz crystal. Both faces of the quartz crystal are coated in order to obtain the highest sensitivity. In case of harmful gases, and for security reasons, a trap must be added to the cell outlet.

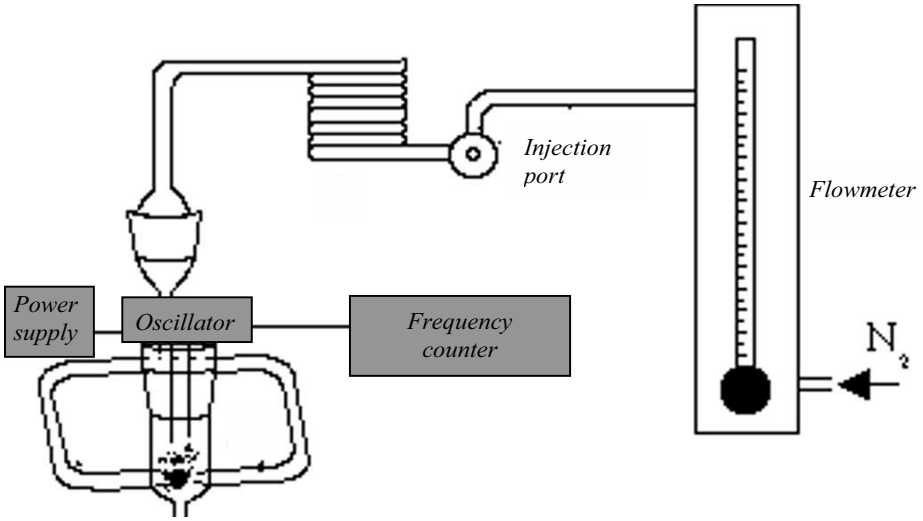


Figure 4. Experimental Layout Used for Flow Injection Gas Analysis.

This is the simplest layout. The only purpose of the coil after the injection port is to delay the analyte in order that any pressure pulse becomes well separated from the signal. The coil should be kept as small as possible in order to decrease dispersion [56]. It is possible to carry out chemical reactions to transform the analyte into a detectable compound, for which a reaction cell with the reagent is assembled after the injection port, and a

dessicant tube inserted before the quartz crystal cell. Figure 5 shows an example of such an experimental layout.

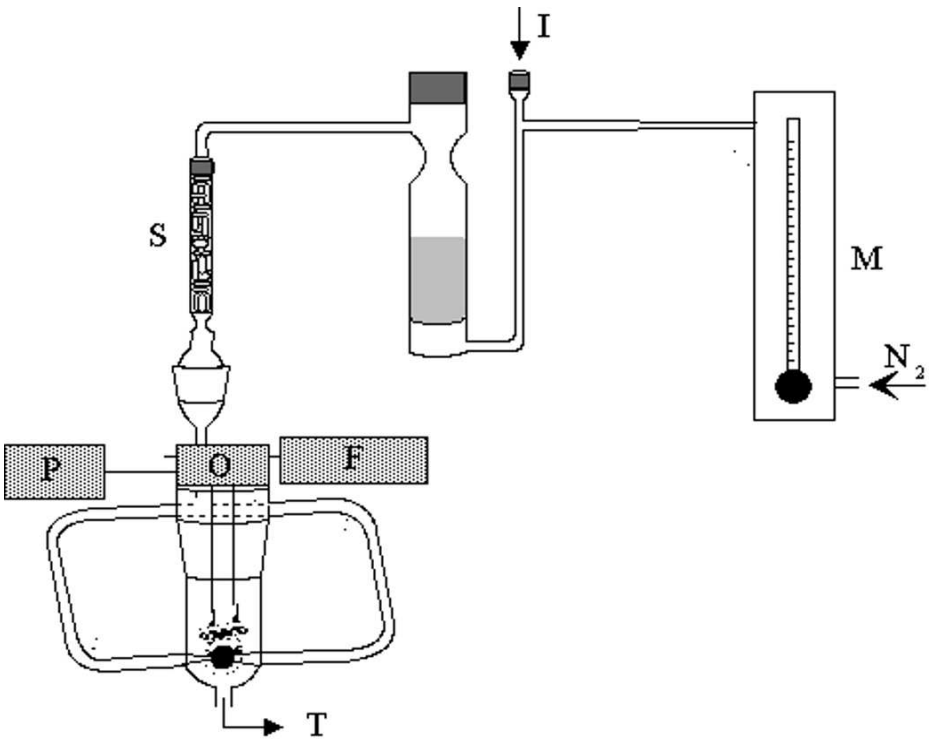


Figure 5. Experimental layout for an experiment where the gas injected in I passes through a reagent and is converted into a detectable compound (P: power supply, O: oscillator, F: frequency meter, T:trap, S: dessicant; I: injection port; M: flowmeter).

With most oscillators, the quartz crystal can only vibrate with one face in contact with a liquid, and a special cell with a sealant to prevent wetting one of the faces of the crystal must be used. Only one face of the crystal will be used for sensing purposes and only the electrode in contact with the liquid must be coated. It is important to remember that there is no advantage in coating both faces as some coating compounds, for instance some commercial metal ionophores can be very expensive (5 mg of a sodium ionophore costs 284 Euros).

Figure 6 shows one possible experimental layout used for the analysis of solutions. A constant stream of an inert carrying liquid, for instance Milli-Q water, is flowing through the crystal cell. A constant pressure is maintained inside the bottle of water, and this pressure is the driving force that keeps the liquid flowing. The loop of the injection valve was filled with the analyte, or

standard solution, and is carried to the crystal cell after the opening of the valve.

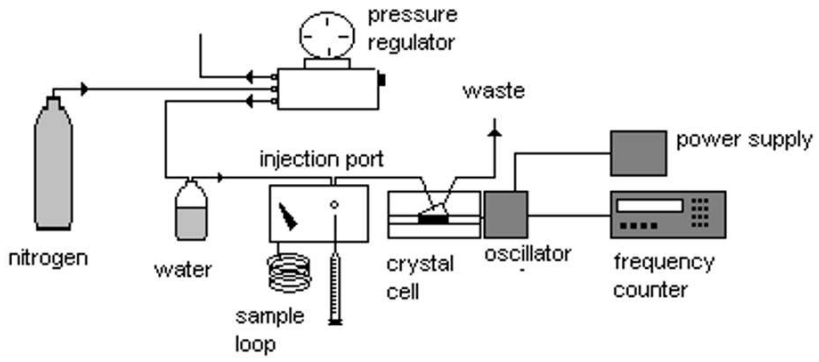


Figure 6. Experimental Layout Used for the Analysis of Solutions.

Once again it is possible to conduct reactions in the system if water is replaced by a reagent or other reagent streams merge with the analyte stream. If a reaction is taking place inside the system, the introduction of a coil of tube, between the injection port and the crystal cell, can increase the reaction time and the quantity of the product formed. These FIA systems are quite versatile and allow easy changes of configuration.

Some authors include a reference crystal while others don't. This reference should correct for baseline drift. Most of the drift is however related to changes that flow, temperature, and humidity produce on the coating, as well as to coating aging. These effects on coating could only be corrected if the reference crystal was coated with a compound that would physically and chemically behave in the same way as the sensing layer, but that remained insensitive to the analyte. In practice, it is impossible to find such a compound and this is the reason why some investigators do not include a reference.

It is very important to remember that most investigators achieved higher frequency stability when the oscillating circuits were placed very near, or directly attached to the crystal. Shielding of the device is also important but there are no general rules to follow. It is important to emphasise that increasing signal to noise ration is more art than science [57].

5. DEVELOPING A SPECIFIC SENSOR

Sensors are designed to solve specific analytical problems and it is mandatory to start by clearly defining the analytical problem. It is important to know the physical and chemical properties of the sample. Is it a gas or a liquid? If it is a liquid, the coating needs to be insoluble and case it is a solid, it needs to be dissolved. The composition of the sample needs also to be known, as the analyst needs to know if there are other components that can cause interference and select the coating that minimizes it. Quartz crystals have reduced physical dimensions and the amount of sample is generally not a problem, but in case of low concentrations, as the sensor responds to mass and not to concentration, the volumes injected, for instance in flow injection methodologies, must be higher than for the samples of higher concentration, in order to display a response close to the centroid of the calibration line. In the limit, pre-concentration steps need to be considered. Because of what was said, the range of the concentration of the analyte on samples needs also to be known, and sensitivity requirements need to be achieved and the range of concentrations to be accommodated within the linear range of the calibration. Accuracy demands must be defined, as stability and recalibration requirements need to be established. The number of samples to be analysed is also a question, as the larger the number, the larger the time and money that can ordinarily be spent on the method development and on attempts to reduce the operator time per sample.

Chemical and physical ruggedness is important for all sensors, and especially if the sensors are intended to be used in the field and in hostile and harsh environments or by less trained operators. *In vivo* sensors have special requirements of biocompatibility, miniaturization and electrical shock hazard minimization and bulk acoustic wave sensors do not seem to be the best option available.

6. MASS SENSITIVITY

Sensitivity is a measure of the ability to discriminate between small differences in concentration [57]. High sensitivity is attained both with a good precision and a steep calibration curve. Usually, sensitivity is considered to be synonymous of calibration sensitivity and it is quantitatively defined by the slope of the calibration curve.

Devices with high frequencies of oscillation were predicted to be more sensitive to mass changes, but the higher the frequency, the thinner the quartz wafer. As very thin plates are difficult to handle, 5 to 10 MHz AT cut quartz crystals are generally used. A few authors tried to increase sensitivity working at higher harmonics or with crystals with unusual high fundamental frequencies [58, 59].

Sensitivity is mainly defined by the choice of the sensitive layer, although most of the choices are empiric. Besides the experiments carried out with commercial available gas chromatographic stationary phases [60, 61], coatings with undefined structure, as Ucon 75-H-90,000 and Ucon-LB-300X [62] and acetone extract of soots resulting from burning of various organo chlorine compounds [63, 64] have been employed. There are also examples where families of compounds, known to interact with some de degree of selectivity with the analyte, and by a known mechanism, have been used [37], as well as selective receptor molecules as crownethers, cryptands, cavitands, porphyrins, calixarenes and cyclodextrines [38, 39, 45]. A word must be said about the investigations on modelling of solubility interactions between vapour molecules and polymeric phases using linear solvation energy relationships (LSERs), where the partition coefficient of a vapour in a polymer is correlated with parameters related to polarizability, dipolarity, hydrogen bond acidity or basicity and dispersion interactions [62]. A very complete review paper written by Grate and Abraham [63] can be a very good guide for coating selection, where the investigators can find a summary of the desirable coating properties, an interesting overview of the interactions between the analyte and the coating and also practical recommendations of specific functional groups that should be included on the coating compound in order to maximize particular interactions.

Sensitivity is usually improved by increasing the coating amount and by carefully design the quartz crystal cell. While coating mass governs sensitivity in absorption phenomena [42], better sensitivities, if responses are governed by adsorption, can be obtained with larger coated areas. Therefore, roughness of the active surface can play an important role, and a significant increase in sensitivity to amines has been achieved rubbing the gold electrode of the quartz crystal with alumina [55].

Swelling of polymers, due to gas absorption, produces an extra frequency decrease. Therefore, if the same mass of a gas would be sorbed on two sensors, one coated with a hard porous coating and the other with a polymer, a larger response would be observed in the last one [67].

Binary systems like Au-Hg, for which several phases are known to occur for different compositions, show a complex calibration curve, and it was shown that the sensitivity of the gold electrode quartz crystal to mercury is not constant and changes with the standard concentration [51]. It is preferred to work within the range of concentrations where sensitivity is higher, which can be achieved diluting the sample or injecting an appropriate amount, whenever this is possible [35].

It is also possible a phase transition to occur during a gas absorption process and the capture of CO_2 by a quartz crystal coated with tetramethylammonium fluoride tetrahydrate is a good example. Besides the usual negative shift in the frequency, a sudden frequency increase is also

present, due to the liquefaction of the CO_2 absorption product [52]. The frequency difference between the minimum and maximum values is the analytical signal, which was verified to be proportional to the CO_2 concentration. This is an example where sensitivity can be improved if the analyst does not reject what seemed to be, at first sight, an aberration of the sensor signal. Table 2 shows a few examples where different solutions were applied to enhance the analytical signal.

Table 2. Examples of Applications where Different Measures Capable of Enhancing the Analytical Signal were Applied.

Analyte/coating	**Action to increase the analytical signal**	**Ref.**
Amine on a bare quartz crystal	Increasing roughness	55
CO_2 on an amine coatin	Decreasing temperature	48
SO_2 as mercury vapour on gold electrodes	Adequate the sample amount to the linear highest sensitivity portion of the calibration curve	35
CO_2 on TMAF	Selecting the working temperature where phase transition takes place	52
CO_2 in amine	Decreasing flow	48

Langmuir shape calibrations are common with adsorption interactions. Sensitivity is the highest at the lowest concentrations and calibration curves level off when the surface becomes saturated. As the sensitivity begins to decrease with concentration, the uncertainty in the concentration becomes higher. It is always desirable to adequate the sample to match the centroid of the lowest linear portion of the calibration curve.

Duplicating devices could only give equal responses if the sensing layer could be perfectly reproduced. Although there are coating methods superior to others [68] this is very difficult, not only because deposition techniques need to apply coating with identical thickness and morphology, which is practically impossible with the most used coating methodologies, smearing, spray or dip coating, but because the bare surface needs to be perfectly cleaned and to possess equal roughness. Therefore, each sensor needs to be calibrated and its sensitivity evaluated. However, this irreproducibility does not mean than an experienced investigator is not capable of preparing a duplicate device capable of attaining a detection limit on the desired range.

It is not unusual to observe a degradation of sensitivity along time. Morphological changes and aging can be responsible for these changes. Changes in morphology can be particularly important when gas flowing streams impinge onto the crystal [48]. The smoothing of the coating along time can be perceptible looking at microscopic images of the surface [48]. It is also necessary to be careful in avoiding solid particles in the stream along with dust particles, as they can damage the sensing layer [69]. Aging can also be caused by irreversible sorption of small amounts of water or other compounds.

In flow injection analysis, high dispersion, associated with long and large diameter tubes or with the presence of reaction chambers, can decrease peak height and decrease sensitivity.

The interaction between coating and analyte, and therefore sensitivity, is temperature dependent. Spontaneous adsorption processes are exothermic and for this reason the responses of the quartz crystal are higher at low temperatures [48]. Other processes as phase transition are also temperature dependent and if temperature is too low they do not occur [52].

Sometimes, in order to achieve a certain detection limit a pre-concentration step is necessary. Once again, we are no longer talking of a true sensor but of a sensing system.

7. SELECTIVITY

An ideal sensor would respond only to the analyte. Selectivity is the degree to which the response is free of interferences. Selectivity can only be obtained from a criterions choice of the coating, as frequency shifts do not reveal their source.

Most of the works appearing in the literature use non-specific coatings, as polymers where the gases dissolve, and can hardly be used in the real world. It is however true that there is no such thing as a universal sensor, and that the development of each sensor must be pursued with its purpose in mind. The investigator needs to know the composition of the sample for which the sensor is designed. It is unthinkable that a lot of effort is spent eliminating an interference if that substance is absent from the material which is going to be analysed. Analysis of volatiles in headspace can also eliminate some problems, as non-volatile compounds are no longer of concern. For those reasons, we can hardly speak of a sensor for CO_2, SO_2, or other compound. It is mandatory to specify that the sensor measures for instance CO_2 or SO_2 in wine.

Although not very often reported, except for ion selective electrodes, numeric selective coefficients are very useful to evaluate interferences and

can help on the choice of a particular membrane composition in the presence of a complex matrix sample [39].

Selectivity requirements are less demanding when an array of sensors is employed, as in electronic noses or tongues, and complex systems can be analysed with the aid of chemometrics [70].

The analyst can also get rid of interfering compounds using molecular scrubbers or separation methods. Although sensitivity is reduced in the pre-separation steps, reliable analytical systems can be designed. Inteference from SO_2 and NO_x in the analysis of CO_2 emitted by a car engine, were efficiently removed passing the sample through a zeolite [71]. Figure 7 shows the experimental arrangement.

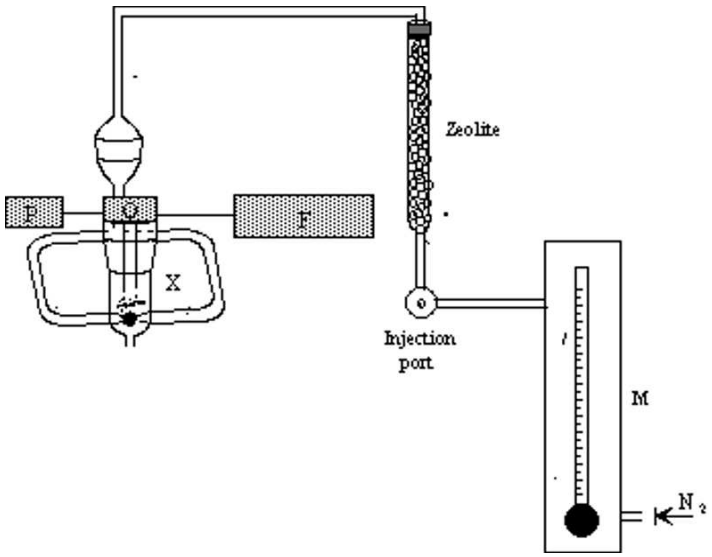


Figure 7. Experimental arrangement including a zeolite trap, after the injection port (P: power supply, O: oscillator, X: crystal, F: frequency meter, M: flowmeter).

8. CALIBRATION

Calibration curves are usually temperature dependent, and for adsorption processes responses increase with temperature decrease [48]. Usually, adsorption isotherms show Langmuir shape with an initial steep slope, which levels off due to surface coverage and shortage of vacant sites.

Although long lasting calibration curves are desired by every analyst, frequent recalibrations are often needed. There are a huge number of reasons responsible for limited calibration validity, and a complete list is not going to be given. Only a few facts prone to be at least partially controlled by the analyst will be mentioned.

Sensitivity is not uniform on the crystal surface and is higher at the centre of the electrodes, decreasing through the edges according to a Gaussian function [3]. A change in the position of the coating, due to spreading or to the formation of holes, leads to changes in sensitivity and recalibration is mandatory [48].

Especially for crystals with easily bending connecting wires, the chances that their position inside the cell differs each time they are taken out and reinserted into the cell are high. For this reason it would be preferred not to touch the crystal during the analysis of both standards and samples.

Aging can be a result of coating degradation, but also of the crystal itself. In experiments of mercury determination, mercury amalgamates onto the gold electrodes of the piezoelectric crystal and crystals are cleaned by heating them at 180-200 ^{o}C. After a large number of temperature cycles it is observed an increase in the frequency of the clean crystal that can be attributed to the detachment of some of the gold segments.

9. VARIABILITY

Variability can lead to frequent changes of sensitivity and calibration curve, as discussed before. Besides, there are variability sources that can affect the response of individual samples or standards, either making calibration impossible or leading to erroneous results. It is difficult to separate permanent changes from momentary changes. Probably some of the variability sources already discussed can turn to be permanent and affect the whole calibration curve, while others affect one single sample of a whole set.

Besides being sensitive not only to the analyte but to other matrix compounds, there are many factors prone to affect the response of a thickness shear mode sensor. Being aware of some of those factors can contribute to alert the analyst against misleading signals.

Temperature can cause changes in the response signals for several reasons. It is true that the frequency of a bare quartz crystal changes with temperature, but as AT-cut quartz crystals are used, this is not expected to be a major problem when working at room temperature. Therefore, most experiments are not run at controlled temperatures, except when working with liquids, as density and viscosity do change significantly with temperature shifts of just a few degrees. The main problem with temperature slight shifts arises from the changes occurring within the coating. Polymer coatings are especially prone to modify with temperature and modulus decrease with thermal expansion leading to a decrease in the frequency. This frequency decrease observation does not correspond to any mass change.

The polymer expands when a vapour is absorbed, and this volume increase cannot also be neglected as it can contribute significantly to the total observed frequency decrease. Once again this volume increase reduces

the modulus of the sorbent polymer and multiplies the frequency decrease that would be observed in a pure gravimetric process [67]. Analytical chemists can use the total frequency shift regardless of its gravimetric or non-gravimetric origin, possibly with a gain in terms of sensitivity. However, a difficulty can arise due to the fact that the decrease in the modulus can be enlarged for higher vapour concentrations, which would be translated in nonlinear sensor responses relative to the vapour concentrations [67].

Temperature range within which experiments are run is also important with polymers as a discontinuity in the slope of the velocity-temperature plot would be observed crossing the glass transition temperature T_g. This discontinuity arises due to the change in the polymer thermal expansion rates at T_g [67]. Chain relaxation processes, occurring after perturbation of chain segments by acoustic waves, can also influence polymer modulus and frequency [72]. A more puzzling effect, known as film resonance, where the upper surface of the polymer film lags behind the motion of the device, is characterized by an increase in wave velocity [72, 73]. Film resonance depends on film thickness, acoustic wave frequency, polymer modulus and density [72].

10. RELIABILITY

The performance of the sensors can only be evaluated if reference samples are analysed or the results obtained analysing real samples are compared with the ones obtained by an official method. The results should be compared both in terms of accuracy and precision [74].

The sensors are usually developed and tested in the laboratory and said to be useful for field, and at real time, monitoring. Most of the sensors are however very fragile, prone to interferences from environmental noise, change their frequency with temperature, etc, and would benefit from further development that would involve shielding and increasing their resistance to mechanical shock. This is a hard job and it is usually not fully done because, as sensors need to be developed for a specific application, the market is very small and does not justify the investment.

11. SENSOR ARRAYS AND HIGHER ORDER SYSTEMS

Single selective sensors are not feasible for some applications with complex matrices, while in other situations they are not even desirable.

An array of sensors can transfer much of the developing effort and selectivity requirements to the mathematical and computational field.

Mathematical methods are used to select the optimal sensing elements to integrate the array and to interpret the results of the set of sensors.

Arrays of piezoelectric quartz crystals were used for classification and quantification purposes [75]. Examples are the arrays developed to detect odorants and hazardous organic vapours [76-81], fish storage time [82], organic contaminants in water [83], to discriminate between aromas of different kinds of dripped coffee or wine and other beverages [84, 85], perfume and flavour identification [86], and monitoring of the drying process of organic samples [87].

Tatsuma et al. [88] reported an array of independent quartz resonators fabricated on a single quartz wafer.

A hybrid sensor system, containing sensors based on different transducer principles was proposed by Göpel et al. [89] and applied to different samples, found in the car, food and aroma industries. The authors claimed that the use of different transducer principles was essential for the unequivocal identification of odours and flavours. Di Natale et al. [90] also found advantages integrating QCM and electrochemical sensors for fish freshness evaluation, as QCM showed to be capable of higher resolution in the first stages of the process and the electrochemical sensors showed higher resolution at the end.

Di Natale et al. [91] merged together data obtained from headspace QCM measurements (electronic nose) and liquid potentiometric measurement (electronic tongue). In spite of using metalloporphyrins as sensitive material both for gas and liquid sensors, a net increase in information was clearly obtained with the integration of both systems.

12. HYPHENATION WITH OTHER TECHNIQUES

Selectivity and sensitivity can be improved by hyphenating the quartz crystal microbalance with separation techniques or pre-concentration devices.

In spite of the fact that we are no longer talking of true sensors, and the fact that such hyphenations are usually not cheap and instrumentation becomes bulky, they can be of interest for chemical analysis and deserve to be mentioned.

Powerful separation techniques as gas or liquid chromatography would allow separating the components of a sample before they reach the quartz crystal. In spite of the fact that we are no longer talking of mass sensors, these sensitive mass detectors can be useful in adding some extra selectivity to complex and difficult analysis [92]. However, as the quartz crystal is now used as a chromatographic detector, it is probably no longer desirable to have an extreme selective device and non-specific coatings have also been

used [16, 93-95] as well as a few compounds with some specificity [92, 96-99].

The hyphenation with electrochemical techniques gave rise to a new acronym and instrument: the electrochemical quartz crystal microbalance (EQCM) which success comes from the fact that it provides new information to the study of electrochemical processes, as mass deposited or stripped from the work electrode, which is a piezoelectric quartz crystal, can now be monitored simultaneously with the current vs. potential measurements. The mass information is however not restricted to electrochemical processes and adsorption or ion and solvent movements in redox polymer films can be detected and studied. For quantitative analysis, the main advantage comes from the fact that electrochemical pre-concentration of the species on the crystal followed by anodic stripping allows detecting and quantifying electrochemical active species at very low concentrations. Trace metals in the environment which would be impossible to detect by a quartz crystal microbalance can now be quantified [23, 100].

13. NETWORK ANALYSER

A sole frequency, usually the series frequency, is measured with a frequency counter. This frequency ceases to exist under high loads or viscous losses.

An impedance /network analyser measures the voltage applied across the crystal at several frequencies, and the current flowing through it. This is known as a passive method in opposition to the oscillator method, known as active method, where the crystal is part of the oscillating circuit. With the passive method, other frequencies as the frequency at minimum or highest impedance can always be read, even in the absence of series and parallel frequency. Besides allowing a more complete understanding of the sensor, as more information, including the electric parameters of the equivalent circuit could be available, it makes possible to use the sensor in analytical situations where the active method fails.

The network analyser is very useful for material characterization and can give an insight into the transduction mechanism, as for instance an increase in mass produce a shift in the admittance peak but no changes in its height, which does experience a decrease in contact with a viscous liquid. However, the network analyser is not adequate for sensing application, as it is bulkier than an oscillator and expensive. Oscillators are generally suitable for sensing applications but must be used carefully as changes in attenuation can produce frequency changes [72].

REFERENCES

1. Barnes C. Development of quartz crystal oscillators for under-liquid sensing. Sens. Actuators A 1991; 29:59-69.
2. Eichelbaum F., Borgräber R., Schröder J., Lucklum R., Hauptman P. Interface circuits for quartz-crystal-microbalance sensors. Rev. Sci. Instrum. 1999; 70:2537-45.
3. Sauerbrey G. Verwendung von Schwingquartzen zur Wägung dünner Schichten und zur Mikrowägung., Z. Physik 1959; 155: 206-22.
4. Lostis, P., *Etude, realisation et controle de lames minces interférentielles introduisant une différance de marche quelconque entre deux vibrations rectangulaires*, PhD thesis, Paris: University of Paris, 1958
5. Mecea V.M. Loading vibrating quartz sensors., Sens. Actuators A, 1994; 40:1-27.
6. Mecea V.M., Bucur R.V., Indrea E. On the possibility of thin film structure study with a quartz crystal microbalance. Thin Solid Films 1989; 171:367-75.
7. Pulker H.K., Benes E., Hammer D., Söllner E. Progress in monitoring thin film thickness with quartz crystal resonators. Thin Solid Films 1976; 32:27-33.
8. Behrndt K.H. Long term operation of crystal oscillators in thin-film deposition., J. Vac. Sci. Technol., 1971; 8:622-6.
9. Benes E., Harms K.-C., Thorn G. Enhanced composite resonator analysis and its application to the quartz crystal microbalance. 29-31 May 1985, Proceedings 39th Annual Symposium on Frequency Control.
10. Wajid A. Improving the accuracy of a quartz crystal microbalance with automatic determination of acoustic impedance ratio. Rev. Sci. Instrum. 1991; 62:2026-33.
11. Benes E. Improved quartz crystal microbalance technique. J. Appl. Phys. 1984; 56:608-26.
12. Miller J.G., Bolef D.I. Acoustic wave analysis of the operation of quartz-crystal film-thickness monitors. J. Appl. Phys. 1968; 39:5815-17.
13. Lu C.-S., Lewis O. Investigation of film-thickness determination by oscillating quartz resonators with large mass load. J. Appl. Phys. 1972; 43:4385-90.
14. Benes E., Schmid M., Thorn, G. Progress in monitoring thin film thickness by use of quartz crystals, Thin Solid Films 1989; 174:307-14.
15. Nomura T., Minemura A. Behavior of a piezoelectric quartz crystal in an aqueous solution and the application to the determination of minute amount of cyanide. Nippon Kagaku Kaishi 1980; 10:1621-5.
16. Konash P.L., Bastiaans G. J. Piezoelectric crystals as detectors in liquid chromatography. Anal. Chem. 1980; 52:1229-31.
17. Nomura T., Okuhara M. Frequency shifts of piezoelectric quartz crystals immersed in organic liquids. Anal. Chim. Acta 1982; 142:281-4.
18. Yao S.-Z, Zhou T.-A. Dependence of the oscillation frequency of a piezoelectric crystal on the physical parameters of liquids. Anal. Chim. Acta 1988; 212:61-72.
19. Bruckenstein S., Shay M. Experimental aspects of use of the quartz crystal microbalance in solution. Electrochim. Acta, 1985; 30:1295-300.
20. Kanasawa K.K., Gordon II, J.G. The oscillation frequency of a quartz resonator in contact with a liquid. Anal. Chim. Acta 1985; 175:99-105.
21. Gomes M.T.S.R., Veríssimo M.I S., Oliveira J.A.B.P. A New method for monitoring the contamination of glycerol with diethylene glycol. J. Pharm. Pharmacol. 1999: 51:233-6.
22. Chen K., Le D., Zhang H., Nie L., Yao S. Model of quartz crystal microbe growth sensor and its application to estimation of microbial populations in mineral waters. Anal. Chim. Acta 1996; 329:83-9.
23. Yamasaki A., Cunha A., Oliveira, J.A.B.P, Duarte C.A.,. Gomes M.T.S.R Assessment of copper toxicity using an acoustic wave sensor. Biosens. Bioelectr. 2004; 19:1203-8.
24. Kurosawa S., Tawara E., Kama N., Kobatake, Y. Oscillating frequency of piezoelectric quartz crystal in solutions. Anal. Chim. Acta 1990; 230:41-9.

25. Schumacher R., Borges G., Kanazawa K. K. The quartz microbalance: a sensitive tool to probe surface reconstructions on gold electrodes in liquid. Surface Sci. 1985*;* 163:L621-26.
26. Stöckel W., Schumacher R. In situ microweighing at the junction metal/electrolyte. Ber. Bunsen-Ges. Phys. Chem. 1987; 91:345-9.
27. Schumacher R. The quartz microbalance: a novel approach to the in-situ investigation of interfacial phenomena at the solid/liquid junction. Angew. Chem. 1990; 29:329-38.
28. Heusler K. E., Grzegorzewski A., Jäckel L., Pietrucha, J. Measurement of mass and surface stress at one electrode of a quartz oscillator 1988, Ber. Bunsenges. Phys. Chem, 92: 1218-1225.
29. Thompson M., Arthur C. L., Dhaliwal G. K. Liquid-phase piezoelectric and acoustic transmission studies of interfacial immunochemistry. Anal. Chem. 1986; 58:1206-9.
30. Thompson M., Kipling A.L., Duncan- Hewitt W. C., Rajakovic L.V., Cavic-Vlasak B. A. Thickness-shear-mode acoustic wave sensors in liquid phase: a review. Analyst 1991; 116:881-90.
31. Janata, J. Centennial retrospective on chemical sensors. Anal. Chem. 2001; 73:150A-63A.
32. Ho, M.H., Guilbault, G.G. Determination of nanogram quantities of mercury in water with a gold-plated piezoelectric crystal detector. Anal. Chim. Acta 1981; 130:141-7.
33. Gomes M.T.S.R., Morgado E.V., Oliveira J.A.B.P. Optimisation of a flow injection system with a piezoelectric quartz crystal detector for the determination of inorganic mercury. Anal. Lett. 1999; 32:2715-23.
34. Suleiman A.A., Guilbault D, Mercury displacement in the determination of sulfur dioxide with a piezoelectric crystal detector. Anal. Chem. 1984; 56:2964-66.
35. Gomes, M.T., Rocha T.A., Duarte A.C, Oliveira, J.P. Determination of sulfur dioxide in wine using a quartz crystal microbalance. Anal. Chem. 1986; 68:1561-64.
36. Gomes, M.T.S.R., Silva A.A.F., Duarte A.C., Oliveira J.A.B.P. Determination of cyanide in waste waters using a quartz crystal microbalance. Sens. Actuators B 1998; 48:383-6.
37. Gomes, M.T.S.R., Barros C.M.F., Santana Marques M.G., Oliveira J.A.B.P. The adsorption of carbon dioxide by tertiary alkanolamines. Can. J. Chem. 1999; 77:401-8.
38 Gomes, M.T.S.R., Tavares K.S., Oliveira J.A.B.P. The quantification of potassium using a quartz crystal microbalance. Analyst. 2000, 125:1983-6.
39. Gomes, M. T. S. R.; Costa, J. R. M. L.; Oliveira, J. A. B. P. The quantification of sodium in mineral waters using a quartz crystal microbalance. Talanta 2003; 59: 247-52.
40. Edmonds T.E. "Voltametric and amperometric transducers." In *Chemical Sensors* T.E Edmonds, ed. Glasgow: Blackie and Son Ltd., 1988.
41. Gomes, M.T., Duarte A.C., Oliveira J.P. Detection of CO_2 using a quartz crystal microbalance. Sens. Actuators B 1995; 26-27:191-4.
42. Milanko O., Milinković S.A., Rajaković L.V. Evaluation of coating material used on piezoelectric sensors for the detection of organophosphorous compounds in the vapour phase. Anal. Chim. Acta 1992; 269:289-300.
43. Nieuwenhuizen M.S., Barendsz, A.W. Processes involved at the chemical interface of a SAW chemosensor. Sens. Actuators B 1987; 11:45-63.
44. Kim, S.-R., Kim, J.-D., Choi, K. H., Chang, Y. H. NO_2 sensing properties of octa(2-ethylhexyloxy) methallophthalocyanine LB films using quartz-crystal microbalance. Sens. Actuators B 1997; 40:39-45.
45. Brunink J.A., Di Natale C., Bungaro F., Davide F.A.M., D'Amico A., Paolesse R., Boschi T., Faccio M., Ferri G. The application of metalloporphyrins as coating material for quartz crystal microbalance-based chemical sensors. Anal. Chim. Acta 1996; 325:53-64.
46. Di Natale C., Paolesse R, Macagnano A., Troitski V.I., Berzina, T.S., D'Amico A. Characterization and design of porphyrins-based chemical sensors for electronic nose applications. Sens. Actuators B 1998; 52:162-68.
47. Di Natale C., Paolesse R, Macagnano A., Troitsky V.I., Berzina, T.S., D'Amico A. Pattern recognition approach to the study of the interactions between metalloporphyrin

Langmuir-Blodgett films and volatile organic compounds. Anal. Chim. Acta 1999; 384:249-59.
48. Gomes M.T.S.R., Duarte A.C., Oliveira J.A.B.P. Critical assessment of the parameters that affect the selection of coating compounds for piezoelectric quartz crystal microbalances. Talanta 1999; 48:81-89.
49. Okahata Y, Ebato H. Application of a quartz-crystal microbalance for detection of phase transitions in liquid crystals and lipid multibilayers. Anal. Chem. 1989; 61:2185-8.
50. Kremer F.J.B., Ringsdorf H., Schuster, A., Seitz M., Weberskirch R. Detection of phase transitions in thin films with a quartz crystal microbalance. Thin Solid Films 1996; 284-285:436-8.
51. Gomes M.T.S.R., Oliveira M.O., Oliveira J.A.B.P. Utilization of a quartz crystal microbalance to obtain Au-Hg phase diagram. Langmuir 1999; 15:8780-2.
52. Gomes M.T.S.R., Rocha T.A.P., Duarte A.C., Oliveira J.A.B.P. Performance of a tetramethylammonium fluoride tetrahydrate coated piezoelectric crystal for carbon dioxide detection. Anal. Chim. Acta 1996; 335:235-8.
53. Gomes M.T.S.R., Veríssimo M.I.S., Oliveira J.A.B.P. Alcohol determination using an acoustic wave sensor. Fresenius J. Anal. Chem. 2001; 369:613-15.
54. Gomes, M.T.; Duarte, A.C.; Oliveira, J.P. The Utilisation of a quartz crystal microbalance for measuring carbon dioxide in wine. Anal. Chim. Acta 1996, 327:95-100.
55. Gomes, M.T.S.R.; Veríssimo, M I.S.; Oliveira, J.A B.P. Detection of volatile amines using a quartz crystal with gold electrodes. Sens. Actuators B 1999, 57:261-67.
56. Ruzička J., Hansen E.H. *Flow injection analysis.* New York: John Wiley & Sons, 1981.
57. Skoog D.A., Holler, F.J., Nieman, T.A. *Principles of Instrumental Analysis.* USA: Hartcourt Brace College Publishers, 5th ed, 1998.
58. Lin Z., Yip M., Joseph S., Ward M.D. Operation of an ultrasensitive 30-MHz quartz crystal microbalance in liquids. Anal. Chem. 2993; 65:1546-51
59. Kurosawa S., Aizawa H., Tozuka M., Nakamura M., Park J.-W. Immunosensors using quartz crystal microbalance. Meas. Sci. Technol. 2003; 14:1882-7.
60. Lopez-Roman A., Guilbault G.G. Use of piezoelectric crystals as sensitive and specific detectors for SO_2. Anal. Lett. 1972; 5:225-35.
61. Edmonds T., West T.S. A quartz crystal piezoelectric device for monitoring gaseous pollutants. Anal. Chim. Acta 1980; 117:147-57.
62. Karmarkar K.H., Guilbault G.G. The detection of ammonia and nitrogen at the parts per billion level with coated piezoelectric crystal detector. Anal. Chim. Acta 1975; 75:111-7.
63. Webber L.M., Hlavay J., Guilbault G.G. Piezoelectric detectors for specific detection of environmental pollutants. Mikrochim. Acta 1978; 1:351-8.
64. Webber L.M., Karmarkar K.H.., Guilbault G.G. A coated piezoelectric crystal detector for the selective detection and determination of hydrogen sulphide in the atmosphere. Anal. Chim. Acta 1978; 97:29-35.
65. Grate, J.W., Wise B.M., Abraham M.H. Method for unknown vapor characterization and classification using a multivariate sorption detector. Initial derivation and modelling based on polymer-coated acoustic wave sensor arrays and linear solvation energy relationships. Anal. Chem. 1999; 71:4544-53.
66. Grate, J.W., Abraham M.H. Solubility interactions and the design of chemically selective sorbent coatings for chemical sensors and arrays. Sens Actuators B 1991; 3:85-111.
67. Grate, J.W., Klusky, M., McGill R.A., Abraham M.H, Whiting G.P., Andonian-Haftvan J. The predominant role of swelling-induced modulus changes of the sorbent phase in determining the responses of polymer-coated surface acoustic wave vapour sensors. Anal. Chem. 1992; 64:610-24.
68. Gomes, M.T., Duarte A.C., Oliveira J.P. Comparison of two methods for coating piezoelectric crystals. Anal. Chim. Acta 1995; 300: 329-34.
69. Barkó G., Némeh, R., Hlavay, J. Investigation of the reliability of piezoelectric chemical sensors. Anal. Chim. Acta 2003; 480:307-16.

70. Carey W.P., Beebe R., Kowalski B.R., Illman D. L., Hirschfeld T. Section of adsorbates for chemical sensor arrays by pattern recognition. Anal. Chem. 1988; 58:149-53.
71. Gomes, M.T.S.R., Nogueira P.S.T., Duarte A.C., Oliveira J.A.B.P. Development of a methodology for the analysis of carbon monoxide using a quartz crystal microbalance. Analyst. 1999; 124:1449-53.
72. Grate J.W., Frye G.C. "Acoustic wave sensors." In *Sensors update vol.2*. H. Bates, W. Göpel, J. Hesse, ed. Weinheim: VCH Verlagsgesellschaft mbH, 1996.
73. Lucklum R, Rösler S, Hartmann J., Hauptmann P. On-line detection of organic pollutants in water by thickness shear mode resonators. Sens. Actuators B 1996; 35-36:103-11.
74. Gomes, M.T.S.R., Rocha T.A.P., Duarte A.C., Oliveira J.A.B.P. Quantification of CO_2 in wines with piezoelectric crystals coated with tetramethylammonium fluoride and comparison with other methods. Analusis 1998; 26:179-82.
75. Grate J.W. Acoustic wave microsensor arrays for vapour sensing. Chem. Rev. 2000; 100:2627-48.
76. Muramatsu H., Tamiya, E., Karube, I. Detection of odorants using lipid-coated piezoelectric crystal resonators. Anal. Chim. Acta 1989; 225:399-408.
77. Chang S.-M., Iwasaki, Y., Tamiya E., Karube, I. Detection of odorants using an array of piezoelectric crystals and neural-network pattern recognition. Anal. Chim. Acta 1991; 249:323-9.
78. Hierlemann, A. Weimar U., Kraus G.; Schwezer-Berberich M., Göpel W. Polymer-based sensor arrays and multicomponent analysis for the detection of hazardous organic vapours in the environment. Sens. Actuators B 1995; 26-27:126-34.
79. Barkó G., Papp B., Hlavay, J. Application of pattern recognition and piezoelectric sensor array for the detection of organic compounds. Talanta 1995; 42:475-82.
80. Kasai N., Sugimoto I., Nakamura M. Discrimination of odorants of definite concentrations by using plasma-organic-film-coated QCR sensors. Sens. Actuators B 2000; 65:114-9.
81. Auge, J. Haupmann P., Hartmann J.; Rösler S., Lucklum R. Versatile microcontrolled gas sensor array system using the quartz microbalance principle and pattern recognitin methods. Sens. Actuators B 1995; 26-27:181-6.
82. Di Natale C., Brunink J.A., Bungaro, F., Davide F., D'Amico A., Paolesse, R., Boschi T., Faccio M, Ferri G. Recognition of fish storage time by a metalloporphyrins-coated QMB sensor array. Sens. Meas. Sci. Technol. 1996; 7:1103-14.
83. Menon, A. Zhou R., Josse F. Coated-quartz crystal resonator (QCR) sensors for on-line detection of organic contaminants in water. IEEE Trans. Ultrason., Ferroelect. Freq. Contr. 1998; 45:1416-26.
84. Nanto H., Kawai, T., Sokooshi H., Usuda, T. Aroma identification using a quartz-resonator sensor in conjunction with pattern recognition. Sens. Actuators B 1993; 13-14:718-720.
85. Cao Z., Lin H.-G., Wang B.-F., Chen Z.-Z., Ma F.-L., Wang K.-M., Yu R.-Q. Discrimination of vapours of alcohols and beverage samples using piezoelectric crystal sensor array. Anal. Lett. 1995; 28:451-66.
86. Nakamoto T., Fukuda, A., Moriizumi T. Perfume and flavour identification by odour-sensing system using quartz-resonator sensor array and neural network pattern recognition. Sens. Actuators B 1993; 10:85-90.
87. Carey W. P., Kowalski, B.R. Monitoring a dryer operation using an array of piezoelectric crystals. Anal. Chem. 1988; 60:541-544.
88. Tatsuma T., Watanabe Y., Oyama, N., Kitakizaki, K.; Haba, M. Multichannel quartz crystal microbalance. Anal. Chem. 1999; 71:3632-36.
89. Ulmer H., Mitrovics, J., Noetzel G., Weimar U., Göpel, W. Odours and flavours identified with hybrid modular sensor systems. Sens. Actuators B 1997; 43:24-33.
90. Di Natale C., Olafsdottir G., Einarsson S., Martinelli E., Paolesse R., D'Amico A. Comparison and integration of different electronic noses for freshness evaluation of cod-fish fillets. Sens. Actuators B 2001; 77:572-8.

91. Di Natale C., Paolesse R., Macagnano A., Mantini A., D'Amico A., Legin A., Lvova L., Rudnitskaya A., Vlasov Y. Electronic nose and electronic tongue integration for improved classification of clinical and food samples. Sens. Actuators B 2000; 64:15-21.
92. Gomes, M.T.S.R., Tavares K.S., Oliveira J.A.B.P. Development of a sensor for calcium based on a quartz crystal microbalance. Fresenius J. Anal. Chem. 2001; 369: 616-9.
93. King W.H. Piezoelectric sorption detector. Anal. Chem. 1964; 36: 1735-39.
94. Karasek F.W., Gibbins K. R. A gas chromatograph based on the piezoelectric detector. J. Chrom. Sci. 1971; 9: 535-40.
95. Konasharasek F.W., Gibbins K.R. A gas chromatograph based on the piezoelectric detector. J. Chrom. Sci. 1971; 9:535-40.
96. Chang P., Shih J.-S. Preparation and application of cryptand-coated piezoelectric crystal gas-chromatographic detector. Anal. Chim. Acta 1998; 360:61-68.
97. Chang P., Shih J.-S. Application of piezoelectric Ru(III)/cryptand-coated quartz crystal gas chromatographic detector for olefins. Anal. Chim. Acta 1999; 380:55-63.
98. Jane Y.-S., Shih J.-S. Application of crown ether coated piezoelectric crystal as a detector for ion chromatography. Analyst 1995; 120:517-22.
99. Chiou C.-S., Shih J.-S. Piezoelectric cryptand-coated quartz crystal liquid-chromatographic detector for cations/anions and polar organic molecules. Anal. Chim. Acta 1999; 392:125-33.
100. Gomes, M.T.S.R. Application of the piezoelectric quartz crystals to the analysis of trace metals in solution: a review. IEEE Sensors J. 2001; 1:109-18.

Chapter 14

ELECTRONIC BIOSENSORS BASED ON BIOMATERIAL-NANOPARTICLE HYBRID SYSTEMS

Eugenii Katz
Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Abstract: Nanomaterials such as metal or semiconductor nanoparticles and nanorods exhibit similar dimensions to those of biomolecules, such as proteins (enzymes, antigens, antibodies) or DNA. The integration of the nanoparticles, exhibiting unique electronic, photonic and catalytic properties, with the biomaterials, revealing unique recognition features or catalytic properties, yields novel hybrid nanobiomaterials of synergetic properties and functions. The present review article describes recent advances in the use of biomolecule-nanoparticle (metallic or semiconductive) assemblies for bioanalytical applications. In these systems the unique electronic, photonic, optical and catalytic properties of the nanoparticles are used to transduce the specific biorecognition and biocatalytic functions of the biomaterial components.

Keywords: biosensors, biomaterials, nanomaterials, enzyme, antigent, antibodie, nanoparticle, nanorods, DNA

1. INTRODUCTION

The unique electronic, optical and catalytic properties of metal and semiconductor nanoparticles (1 nm – 200 nm), together with the different methods available for the preparation of nanoparticles of controlled shape and size, provide exciting building blocks for nanoscale assemblies, structures and devices. A variety of synthetic methodologies for the preparation of nanoparticles within a narrow size distribution are available [1]. Often, the nanoparticles are prepared by "wet chemistry" procedures where the clustering of the metal atoms or semiconductor molecules proceeds in the presence of a surface capping ligand. This capping ligand binds to the metal/semiconductor clusters, prevents aggregation of the particles into bulk material, and controls the final dimensions of the

S. Y. Yurish and M.T.S.R. Gomes (eds.), Smart Sensors and MEMS, 447-474.

nanoparticles. Many capping systems are available including hydrophobic monolayers [2], positively- or negatively-charged hydrophilic monolayers [3] and polymer layers [4]. Association of molecular units to the nanoparticles introduces chemical functionalities that can provide recognition or affinity interactions between different appropriately modified particles, and thereby dictate the structure when aggregation occurs [5]. New collective properties of aggregated nanoparticles [6] such as the coupled plasmon absorbance, interparticle energy transfer, and electron transfer or conductivity may be observed in the clustered assemblies.

The chemical functionalities associated with nanoparticles enable the assembly of 2D- and 3D-nanoparticle architectures on surfaces [6]. Composite layered or aggregated structures of molecule- or macromolecule-crosslinked nanoparticles on surfaces have been prepared, and the specific sensing of substrates [7], tunable electroluminescence [8], and enhanced photoelectrochemistry [9], have been accomplished. The assembly of nanoparticle architectures on surfaces has also led to the fabrication of nanoscale devices such as single electron transistors [10], nanoparticle-based molecular switches [11], metal-insulator-nanoparticle-insulator-metal (MINIM) capacitors [12], and others. Several reviews have addressed recent advances in the synthesis and properties of nanoparticles [6, 13] and the progress in the integration of composite nanoparticle systems with surfaces [6, 14].

Recent trends involving the convergence of biotechnology and nanotechnology accelerate the development of hybrid nanomaterials incorporating the highly selective catalytic and recognition properties of biomaterials such as proteins/enzymes and DNA with the unique electronic, photonic and catalytic features of nanoparticles. The conjugation of nanoparticles and other nano-objects (e.g. nanorods, carbon nanotubes) with biomaterials is a tempting research project that provides a route into nanobiotechnology [15, 16]. Evolution has optimized fascinating macromolecular structures exhibiting unique recognition, transport and catalytic properties. The conjugation of nanoparticles with biomaterials could provide electronic or optical transduction of the biological phenomena, resulting in the development of novel biosensors [17]. Enzymes, antigens/antibodies, and biomolecular receptors have dimensions in the range of 2–20 nm, comparable to those of nanoparticles, and thus the two have structural compatibility. Several fundamental features show biomaterials to be important future building blocks for nanoparticle architectures: (*i*) Biomaterials reveal specific and strong complementary recognition interactions, e.g. antigen-antibody, nucleic-acid-DNA, hormone-receptor. The functionalization of a single kind of nanoparticle or different kinds of nanoparticles with biomaterials could lead to biomaterial-nanoparticle recognition and thus to self-assembly. (*ii*) Various biomaterials include several binding sites, e.g. the two Fab-sites of antibodies, the four

binding domains of streptavidin or concanavalin A. This allows the multidirectional growth of nanoparticle structures. (*iii*) Proteins may be genetically engineered and modified with specific anchoring groups. This facilitates the aligned binding to nanoparticles, or the site-specific linkage of the biomaterial to surfaces. Consequently, directional growth of nanoparticle structures may be dictated. Furthermore, other biomaterials, such as double-stranded DNA, may be synthetically prepared in complex rigidified structures that act as templates for the assembly of nanoparticles by intercalation, electrostatic binding to phosphate groups, or by association to functionalities tethered to the DNA. (*iv*) Enzymes provide catalytic tools for the manipulation of biomaterials. For example, the ligation of nucleic acids, or the endonuclease scission processes of nucleic acids, provide effective tools for controlling the shape and structure of biomaterial-nanoparticle hybrid systems. In this context, it is important to note that Mother Nature has developed unique biocatalytic replication processes. The use of biocatalysts for the replication of biomaterial-nanoparticle conjugates may provide an effective system for the formation of nanostructures of pre-designed shapes and compositions.

The unique properties of nanoparticles make nanoparticle-biomaterial conjugates attractive labels for sensing applications. The electronic sensing of biomaterials on surfaces is a common practice in analytical biochemistry. Thus, the immobilization of nanoparticle-biomaterial conjugates on surfaces provides a general route for the development of electronic biosensors. It is the aim of this paper to review the recent advances in the application of biomaterial-nanoparticle hybrid systems for designing novel biosensors.

2. BIOELECTRONIC SYSTEMS FOR SENSING BASED ON NANOPARTICLE-ENZYME HYBRIDS

Electrical contacting of redox-enzymes with electrodes is a key process in the tailoring of enzyme-electrodes for bioelectronic applications such as biosensors [18-20] or biofuel cell elements [21]. Redox-enzymes usually lack direct electrical communication with electrodes and electrical contacting of redox proteins with electrodes was achieved by the application of diffusional electron mediators [22], the tethering of redox-relay groups to the protein [23], or the immobilization of the enzymes in redox-active polymers [24, 25]. Nonetheless, relatively inefficient electrical contacting of the enzymes with the electrode is achieved due to the non-optimal modification of the enzymes by the redox units or the lack of appropriate alignment of the enzymes in respect to the electrode. Very efficient electrical communication between redox-proteins and electrodes was achieved by the reconstitution of apo-enzymes on relay-cofactor monolayers associated with electrodes [26-29]. For example, apo-glucose oxidase was reconstituted on a

relay-FAD layer [26, 27], and apo-glucose dehydrogenase was reconstituted on a pyrroloquinoline quinone (PQQ)-modified polyaniline film associated with an electrode [28]. Effective electrical communication between the redox-centers of the biocatalysts and the different electrodes was observed and reflected by high turnover electron transfer rates from the redox-sites to the electrode. The effective electrical contacting of these redox enzymes was attributed to the alignment of the proteins on the electrodes and to optimal positioning of the intermediary electron-relay units between the enzyme redox centers and the electrode.

Application of conductive nanoparticles for electrical wiring of redox enzymes paves the way to the novel bioelectronic sensors. A few biocatalytic electrodes have been prepared for biosensor applications by co-deposition of redox enzymes and Au nanoparticles on electrode supports [30,31]. For example, direct electron transfer between hemoglobin and a glassy carbon electrode was facilitated by lipid-protected gold nanoparticles [31]. The biocatalytic electrodes were reported to operate without electron transfer mediators, but the random and non-optimized positioning of the redox proteins on the conductive nanoparticles did not allow the efficient electron transfer between the enzyme active sites and the electrode support. Highly efficient electrical contacting of the redox enzyme glucose oxidase (GOx) through a single Au nanoparticle was accomplished by the reconstitution of the apo-flavoenzyme, apo-glucose oxidase, (apo-GOx) on a 1.4 nm Au_{55}-nanoparticle functionalized with N^6-(2-aminoethyl)-flavin adenine dinucleotide (FAD cofactor amino-derivative) (**1**). The conjugate produced was assembled on a thiolated monolayer using different dithiols (**2-4**) as linkers, Figure 1(A) [32].

Alternatively, the FAD-functionalized Au nanoparticle could be assembled on a thiolated monolayer associated with an electrode, and subsequently apo-GOx was reconstituted on the functional nanoparticles, Figure 1(B). The enzyme-electrodes prepared by these two routes reveal similar protein surface coverages of ca. 1×10^{-12} mole·cm^{-2}. The nanoparticle-reconstituted glucose oxidase layer was found to be electrically contacted with the electrode without any additional mediators, and the enzyme assembly stimulates the bioelectrocatalyzed oxidation of glucose (Figure 1(C)). The resulting nanoparticle-reconstituted enzyme electrodes revealed unprecedented efficient electrical communication with the electrode (electron transfer turnover rate ca. 5000 s^{-1}). This electrical contacting makes the enzyme-electrode insensitive to oxygen or to common oxidizable interferants such as ascorbic acid. The electron transfer from the enzyme active center through the Au nanoparticle is rate-limited by the structure of the dithiol molecular linker that bridges the particle to the electrode. The conjugated benzene dithiol (**4**) was found as the most efficient electron transported unit among the linkers (**2**) - (**4**). The future application of

effective molecular wires such as oligophenylacetylene units could further improve the electrical contacting efficiency.

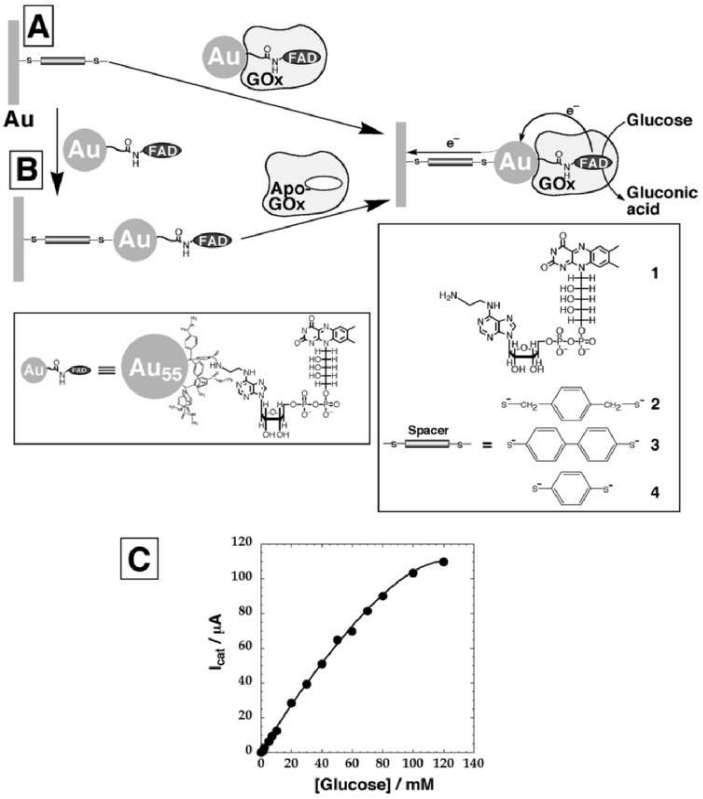


Figure 1. Electrical 'wiring' of glucose oxidase (GOx) by the apo-enzyme reconstitution with a Au nanoparticle functionalized with a single FAD cofactor unit. (A) Reconstitution process performed in a solution followed by the assembly adsorption onto a dithiol-modified Au electrode. (B) A Au-FAD conjugate adsorption onto a dithiol-modified Au electrode followed by the reconstitution of the apo-GOx at the interface. (C) Calibration plots of the electrocatalytic current developed by the reconstituted GOx electrode in the presence of different concentrations of glucose.

While the previous system employed the metal nanoparticle as a nano-electrode that electronically communicates the enzyme redox-site with the macroscopic electrode, one may use enzyme-nanoparticle hybrid systems where the product generated by the biocatalytic process activates the functions of the nanoparticle. This has recently been demonstrated by tailoring an acetylcholine esterase (AChE)-CdS nanoparticle hybrid monolayer on a Au electrode, and the activation of the photoelectrochemical functions of the nanoparticles by the biocatalytic process [33]. The

CdS-AChE hybrid interface was assembled on the Au electrode by the stepwise coupling of cystamine-functionalized CdS to the electrode, and the secondary covalent linkage of the enzyme AChE to the particles (Figure 2(A)). In the presence of acetylthiocholine (**5**) as substrate, the enzyme catalyzes the hydrolysis of **5** to thiocholine (**6**) and acetate. Photoexcitation of the CdS semiconductor yields the electron-hole pair in the conduction-band and the valence-band, respectively. The enzyme-generated thiocholine (**6**) acts as an electron donor for valence-band holes. The scavenging of the valence-band holes results in the accumulation of the electrons in the conduction-band and their transfer to the electrode with the generation of a photocurrent (Figure 2(B)).

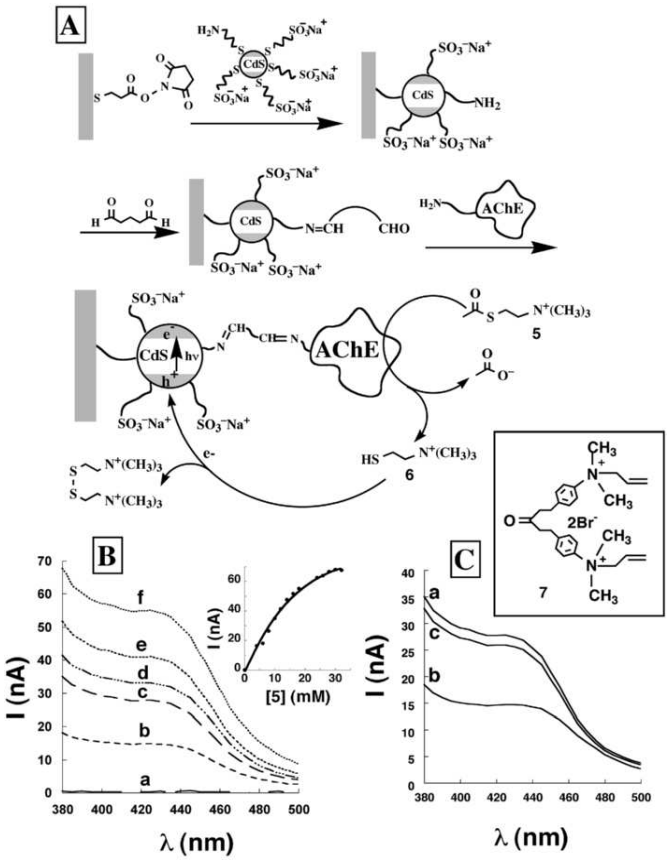


Figure 2. (A) Assembly of the CdS-nanoparticle/AChE hybrid system used for the photoelectrochemical detection of the enzyme activity. (B) Photocurrent action spectra observed in the presence of acetylthiocholine (**5**): (a) 0 mM, (b) 6 mM, (c) 10 mM, (d) 12 mM, (e) 16 mM, (f) 30 mM. Inset: Calibration curve corresponding to the photocurrent at λ = 380 nm at variable concentrations of **5**. (C) Photocurrent spectra corresponding to the CdS/AChE system in the presence of **5**, 10 mM, (a) without the inhibitor, (b) upon addition of the inhibitor (**7**), 1×10^{-6} M, (c) after rinsing the system and excluding of the inhibitor.

The addition of enzyme inhibitors such as 1,5-*bis* (4-allyldimethylammoniumphenyl)pentane-3-one dibromide (**7**) blocks the biocatalytic functions of the enzyme and as a result inhibits the photocurrent formation in the system (Figure 2(C)). Thus, the hybrid CdS/AChE system provides a functional interface for sensing of the AChE inhibitors (e.g. chemical warfare) by means of photocurrent measurements. A similar system composed of a photoactivated CdS nanoparticles and co-immobilized formaldehyde dehydrogenase that utilizes formaldehyde as an electron donor has been reported [34]. In this hybrid system the direct electron transfer from the enzyme active center to the CdS photogenerated holes was achieved and the steady-state photocurrent signal in the system was reported to be directly related to the substrate concentration.

3. BIOELECTRONIC SYSTEMS FOR SENSING OF BIORECOGNITION EVENTS BASED ON NANOPARTICLES

The unique optical [35], photophysical [6], electronic [36, 37] and catalytic [38] properties of metal and semiconductor nanoparticles turn them into ideal labels for biorecognition and biosensing processes. For example, the unique plasmon absorbance features of Au nanoparticles and specifically the interparticle-coupled plasmon absorbance of conjugated particles have been widely used for DNA [39] and antibody-antigen [40] analyses. Similarly, the tunable fluorescence properties of semiconductor nanoparticles were used for the photonic detection of biorecognition processes [41, 42]. Also electrochemical methods were used to follow nanoparticle labels bound to biomaterials. For example, differential pulse voltammetry signals (ca. 1.2 V) corresponding to the oxidation of Au nanoparticle labels connected to the target DNA molecules were applied for electrochemical DNA analysis [43].

Metal and semiconductor nanoparticles coupled to biomaterials generate solubilized entities. Nonetheless, even nanoscale particulate clustered systems include many atoms/molecules in the clusters. The solubility of the nanoparticle-biomaterial structures allows the application of washing procedures on surfaces that include a sensing interface, and thus non-specific adsorption processes are eliminated. On the other hand, the specific capturing of biomaterial-nanoparticles on the respective sensing interfaces allows the secondary dissolution of the captured nanoparticles, and thus enables the amplified detection of the respective analyte by the release of many ions/molecules as a result of a single recognition event.

Most of the used detection schemes have commonly relied on a highly sensitive electrochemical stripping transduction/measurement of the metal

tracer. Stripping voltammetry is a powerful electroanalytical technique for trace metal measurements [44]. Its remarkable sensitivity is attributed to the 'built-in' preconcentration step, during which the target metals are accumulated (plated) onto the working electrode. The detection limits are thus lowered by 3-4 orders of magnitude, compared to pulse-voltammetric techniques, used earlier for monitoring DNA hybridization. Such ultrasensitive electrical detection of metal tags has been accomplished in connection to a variety of novel DNA-linked particle nanostructure networks. The clustered systems could be loaded with additional markers, including redox-active moieties. For example, silica nanoparticles loaded with *tris*(2,2'-bipyridyl)cobalt(III) [45] or Au nanoparticles functionalized with tethered ferrocene units [46] were applied for labeling of DNA and further electrochemical DNA detection based on the redox process of the redox-active complex units.

Powerful nanoparticle-based electrochemical DNA hybridization assays were developed using Au, Ag, Cu or In metal tracers [47-51]. Such protocols have relied on capturing the gold [47, 48], silver [49], Cu_{core}-Au_{shell} [50] nanoparticles, or In nanorods [51] to the hybridized target, followed by anodic-stripping electrochemical measurement of the metal tracer. The probe or target immobilization has been accomplished directly on carbon or indium-tin oxide (ITO) electrodes [52, 53]. Alternatively, the DNA probe was linked to streptavidin-coated magnetic beads [47] or adsorbed onto the walls of polystyrene microwells [48]. The DNA-functionalized beads were collected on an electrode surface, and picomolar levels of the DNA target have thus been electrochemically detected. For example, an electrochemical method was employed for the Au nanoparticle-based quantitative detection of the 406-base human cytomegalovirus DNA sequence (HCMV DNA) [48]. The HCMV DNA was immobilized on a microwell surface and hybridized with the complementary oligonucleotide-modified Au nanoparticle. The resulting surface-immobilized Au nanoparticle/double-stranded DNA assembly was treated with HBr/Br_2 resulting in the oxidative dissolution of the gold particles. The solubilized Au^{3+}-ions were then electrochemically reduced and accumulated on the electrode and subsequently determined by anodic stripping voltammetry using a sandwich-type screen-printed microband electrode (SPMBE). The combination of the sensitive detection of Au^{3+}-ions at the SPMBE due to non-linear mass transport of the ions, and the release of a large number of Au^{3+} ions upon the dissolution of the particle associated with a single recognition event provides an amplification path that enabled the detection of the HCMV DNA analyte at a concentration of 5×10^{-12} M. Further sensitivity enhancement was achieved by the catalytic enlargement of the gold tracer associated with the double-stranded DNA assembly by the deposition of gold [47] or silver [54]. Combining such enlargement of the metal-particle tags, with the effective 'built-in'

amplification of electrochemical stripping analysis paved the way to sub-picomolar detection limits.

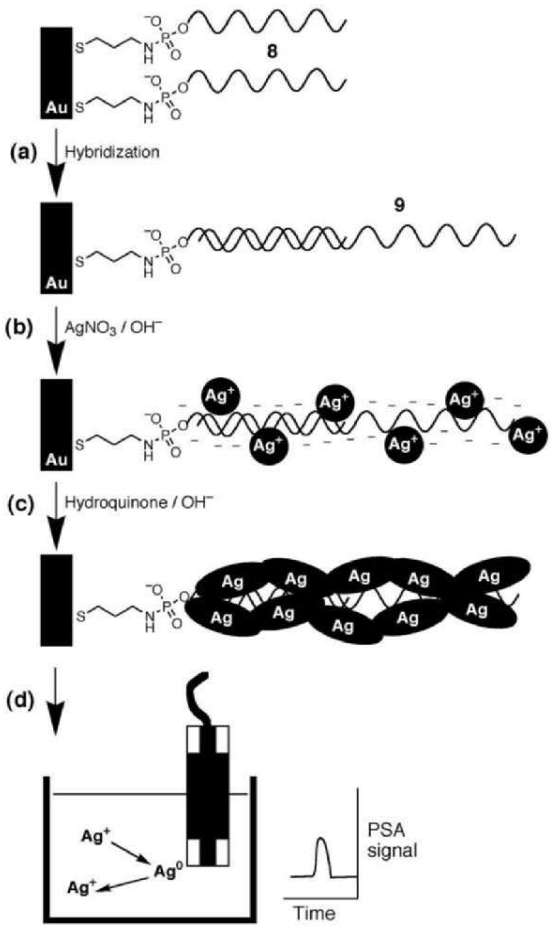


Figure 3. Schematic outline of the steps involved in the amplified electrochemical detection of DNA by the catalytic silver cluster deposition on the DNA strand: (a) Hybridization of the complementary target DNA (**9**) with the DNA probe (**8**) covalently linked to the electrode surface through a cystamine monolayer; (b) Loading of the Ag^+ ions onto the immobilized DNA; (c) Reduction of Ag^+ ions by hydroquinone to form silver aggregates on the DNA backbone; (d) Dissolution of the silver aggregates in an acid solution and transfer to the detection cell followed by stripping potentiometric detection (PSA = potentiometric stripping analysis).

An electrochemical protocol for detecting DNA hybridization based on the deposition of the metal clusters along the DNA backbone (instead of capturing it at the end of the duplex) was described recently [55]. This protocol relies on DNA-template induced generation of conducting nanowires as a result of DNA hybridization. The use of DNA as a

metallization template [56] has evoked substantial research activity directed to the generation of conductive nanowires and the construction of functional circuits [57-60]. Such approach was applied to grow silver [58, 59], palladium [57] or platinum [60] clusters on DNA templates. Elements from the methods used for the generation of metal nanocircuitry based on DNA templates were adapted to develop DNA detection schemes as outlined in Figure 3. The short DNA primer (**8**) attached to the electrode hybridizes with the target DNA (**9**) (step a). The phosphate groups associated with the long target DNA (**9**) collect Ag^+-ions from the solution by electrostatic interaction (step b). The bound Ag^+-ions are then reduced by hydroquinone, resulting in the formation of metallic silver aggregates along the DNA (step c). The subsequent dissolution and stripping electrochemical detection of the nanoscale silver clusters (step d) provides then the route to detect the hybridized DNA. It should be noted, however, that the short DNA primer might also bind some Ag^+ ions that yield a background response. The background signal could be avoided and the sensitivity provided by this method could be improved upon application of peptide nucleic acids (PNA) that lack phosphate groups and thus do not bind Ag^+ ions as the primer for hybridization of the target DNA.

The catalytic features of metal nanoparticles that enable the electroless deposition of metals on the nanoparticle clusters allow the enlargement of the particles to conductive interparticle-connected entities. The formation of conductive domains as a result of biorecognition events provides then an alternative path for the electrical transduction of biorecognition events. This was exemplified by the design of a miniaturized immunosensor based on Au nanoparticles and their catalytic properties [61] (Figure 4(A)). Latex particles stabilized by an anionic protective layer were attracted to a gap between micron-sized Au electrodes by the application of a non-uniform alternating electric field between the electrodes (dielectrophoresis). Removal of the protective layer from the latex particles by an oppositely charged polyelectrolyte resulted in the aggregation of the latex particles and their fixation in the gap domain. Adsorption of protein A on the latex surface yielded a sensing interface for the specific association of the human immunoglobulin (IgG) antigen. The association of the human immunoglobulin on the surface was probed by the binding of the secondary Au-labeled anti-human IgG antibodies to the surface, followed by the catalytic deposition of a silver layer on the Au nanoparticles. The silver layer bridged the gap between the two microelectrodes, resulting in a conductive 'wire'. Typical resistances between the microelectrodes were 50-70 Ω, whereas control experiments that lack the specific catalytic enlargement of the domain by the Au nanoparticle-antibody conjugate yielded resistances $>10^3$ Ω. The method enabled the analysis of human IgG with a detection limit of ca. 2×10^{-13} M.

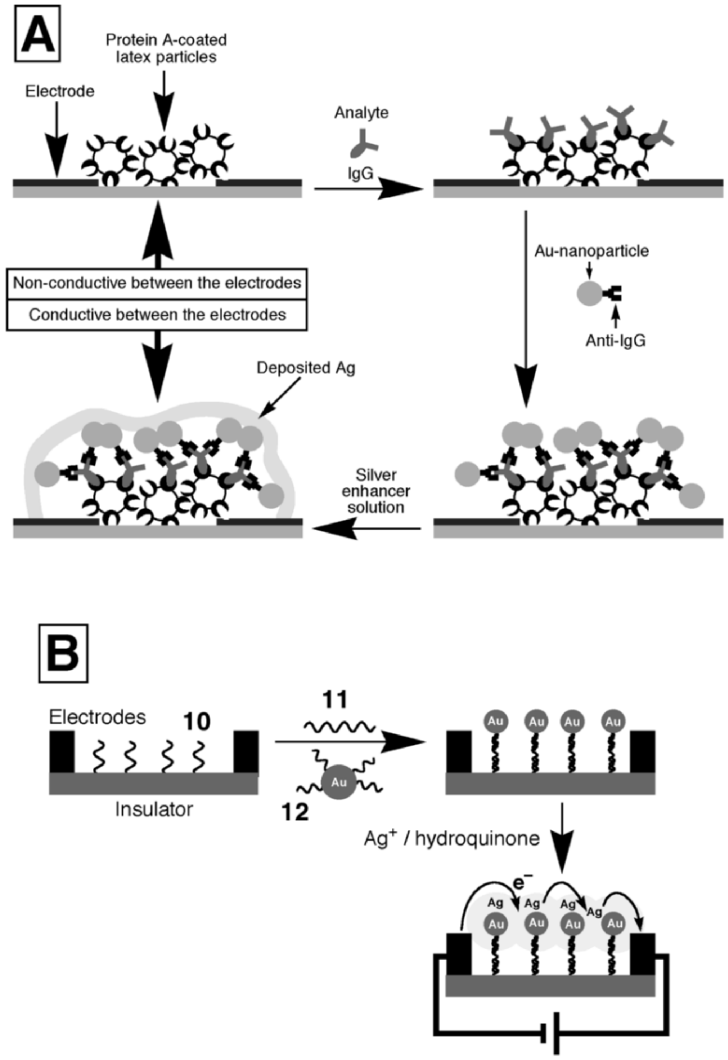


Figure 4. (A) Immunosensing at micro-sized Au electrodes based on the change of conductivity between the Au strips upon binding of Au nanoparticles followed by silver deposition. (B) The use of a DNA-nanoparticle conjugate and subsequent silver deposition to connect two microelectrodes, as a means to sense a DNA analyte. (Part A is adapted from ref. 6, Scheme 15(A), with permission).

A related DNA detection scheme was developed using microelectrodes fabricated on a silicon chip [62] (Figure 4(B)). A probe nucleic acid (**10**) was immobilized on the SiO_2 interface in the gap separating the microelectrodes. The target 27-mer-nucleotide (**11**) was then hybridized with the probe interface, and subsequently a nucleic acid (**12**)-functionalized Au

nanoparticles were hybridized with the free 3'-end of the target DNA. The Au nanoparticle catalyzed reduction of Ag^+-ions by hydroquine, and this resulted in the deposition of silver on the Au nanoparticle assembly and the decrease of the resistance between the electrodes. Single-base mutants of the analyte oligonucleotide (**11**) were washed off from the capture-nucleic acid (**10**) by the use of a buffer with the appropriate ionic strength. A difference of 10^6 in the gap resistance was observed upon analyzing by this method the target DNA and its mutant. The low resistances between the microelectrodes were found to be controlled by the concentration of the target DNA, and the detection limit for the analysis was estimated to be ca. 5×10^{-13} M. This sensitivity translates to ca. 1 $\mu g\cdot\mu L^{-1}$ of human genomic DNA or ca. 0.3 $ng\cdot\mu L^{-1}$ of DNA from a small bacterium. The resulting sensitivity suggests that the DNA extracted from pathogens may be analyzed with no pre-PCR amplification. The simultaneous analysis of a collection of DNA targets was accomplished with a chip socket that included 42 electrode gaps, and appropriate different nucleic acid sensing probes between the electrode gaps [63].

The immobilization of nanoparticles on surfaces may also be used to yield high surface area electrodes [64] and hence for increasing the hybridization capacity of the surface [65]. Such use of nanopartice supporting films relied on the self-assembly of 16-nm diameter Au nanoparticles onto a cystamine-modified gold electrode and resulted in surface densities of oligonucleotides as high as 4×10^{14} $molecules\cdot cm^{-2}$. The detection of the ferrocenecarboxaldehyde tag (conjugated to the target DNA) resulted in a detection limit of 500 pM. The roughening of a Au-quartz crystal with a monolayer consisting of Au nanoparticles was also employed for the enhanced microgravimetric analysis of DNA [66]. Similarly, electrode surfaces roughened by the Au nanoparticles deposition were employed as platforms for the enhanced immunoassay using impedimetric or amperometric read-out signals [67].

Efficient methods for the preparation of semiconductor nanoparticles (e.g. CdS, CdSe, PbS, ZnS) and their functionalization with biomaterials were recently developed [68]. These nanoparticles were applied for labeling of biomaterials in biorecognition processes (e.g. DNA sensing). For example, CdS-semiconductor nanoparticles modified with nucleic acid were employed as tags for the detection of hybridization events of DNA [69]. Dissolution of the CdS (in the presence of 1 M HNO_3) followed by the electrochemical reduction of the Cd^{2+} to Cd^o that accumulates on the electrode, and the stripping-off of the generated Cd^o (to Cd^{2+}) provided the electrical signal for the DNA analysis. Figure 5(A) shows the chronopotentiograms resulting in the analysis of different concentrations of the complementary target DNA using the CdS nanoparticles as tags. A further development has included the use of magnetic particles functionalized with the sensing nucleic acid and CdS nanoparticles

functionalized with the complementary oligonucleotide as labels. The hybridization of the complementary nucleic acids allowed the separation of the duplex from the reaction mixture and its concentration on the electrode surface followed by the electrochemical detection. Electroless deposition of Cd^o onto CdS nanoparticles allowed their enlargement and further amplification of the electrochemical signal resulting in the sensitivity level as low as 100 fmol of the analyte DNA.

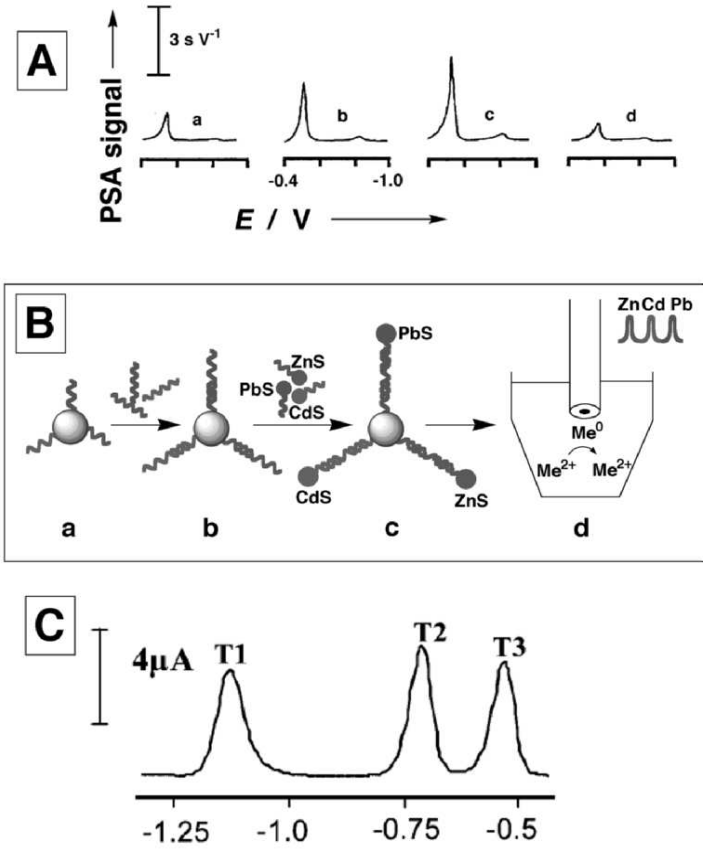


Figure 5. (A) Stripping potentiograms measured upon the sensing of different concentrations of DNA, which is bound to magnetic particles and labeled with CdS-nanoparticles: (a) 0.2 $mg \cdot L^{-1}$. (b) 0.4 $mg \cdot L^{-1}$. (c) 0.6 $mg \cdot L^{-1}$ (d) Control experiment with non-complementary DNA, 0.6 $mg \cdot L^{-1}$. (B) Multi-target electrochemical DNA detection based on different nanocrystal labels: (a) Introduction of probe-modified magnetic beads. (b) Hybridization with the DNA targets. (c) Second hybridization with the nanoparticle-labeled probes. (d) Dissolution of the nanoparticles and the electrochemical detection. (C) Stripping voltammogram recorded upon the simultaneous analysis of three different 60-mer DNA targets related to the BRCA1 breast-cancer gene (54 nM each) labeled: (T1) with ZnS nanoparticles, (T2) with CdS nanoparticles, and (T3) with PbS nanoparticles. (Part A is adapted from ref. 69, Figure 2; Parts B and C are adapted from ref. 70, Figures 1 and 2(E), with permission).

An interesting aspect of these systems is, however, the future possibility of using a combination of different metal or semiconductor tags linked to different nucleic acids for the simultaneous high throughput analysis of different DNA targets (a library) linked to different magnetic beads. By this approach [70], different nucleic acid probes complementary to a series of DNA targets are linked to different magnetic particles. Similarly, different semiconductor or metal nanoparticle tags complementary to segments of the series of the target DNAs were used as amplifying detection units for the primary hybridization process. The hybridization of the nucleic acid-functionalized semiconductor or metal particle to the specific DNA targets, followed by the dissolution of the nanoparticles and the electrochemical accumulation and stripping off of the metals, enables the determination of the specific DNA targets present in the sample. That is, the characteristic potentials needed to strip off the metal provide electrochemical indicators for the nature of the analyzed DNA. A model system that follows this principle was developed [70]. Three kinds of magnetic particles modified with three different nucleic acids were hybridized with the complementary analyte nucleic acids, and subsequently hybridized with three different kinds of semiconductor nanoparticles, ZnS, CdS, PbS, that were functionalized with nucleic acids complementary to the free ends of the analyte DNA (Figure 5(B)).

The magnetic particles allow easy transportation and purification of the analyte sample; whereas the semiconductor particles provide non-overlapping electrochemical signals that transduce the specific kind of hybridized DNA. Stripping voltammetry of the metals originating from the semiconductor nanoparticles yielded well-defined and resolved stripping peaks, e.g. at –1.12 V (Zn), -0.68 V (Cd), and –0.53 V (Pb) (vs. Ag/AgCl reference), thus allowing the simultaneous electrochemical analysis of several DNA analytes tagged with the labeling semiconductor nanoparticles. For example, Figure 5(C) depicts stripping voltammograms for a solution containing three DNA samples labeled with the ZnS, CdS and PbS nanoparticle tracers. The functionalization of the nanocrystal tags with thiolated oligonucleotide probes offered the voltammetric signature with distinct electrical hybridization signals for the corresponding DNA targets. The position and size of the resulting stripping peaks provided the desired identification and quantitative information, respectively, on a given target DNA. The multi-target DNA detection capability was coupled to the amplification feature of stripping voltammetry (to yield fmol detection limits) and with an efficient magnetic removal of non-hybridized nucleic acids to offer high sensitivity and selectivity. Up to 5-6 targets can thus be measured simultaneously in a single run in connection to ZnS, PbS, CdS, InAs, and GaAs semiconductor particles. Conducting massively parallel assays (in microwells of microtiter plates or using multi-channel microchips,

with each microwell or channel carrying out multiple measurements) could thus lead to a high-throughput analysis of DNA.

The amplification paths for electrochemical analyzing DNA that were discussed in the previous sections have employed a single reporter unit, e.g. a metal or semiconductor nanoparticle, per one hybridization event. For further enhancing the sensitivity of DNA detection it is possible to load multiple tags per binding event [71, 72]. This can be accomplished by linking the biorecognition units to polymeric microbeads carrying multiple redox tracers in external positions (on their surface) or internal positions (via encapsulation). A triple-amplification bioassay that couples the carrier-sphere amplifying units (loaded with numerous gold nanoparticles tags) with the 'built-in' preconcentration feature of the electrochemical stripping detection and the catalytic enlargement of the multiple gold-particle tags was demonstrated [71] (Figure 6(A)).

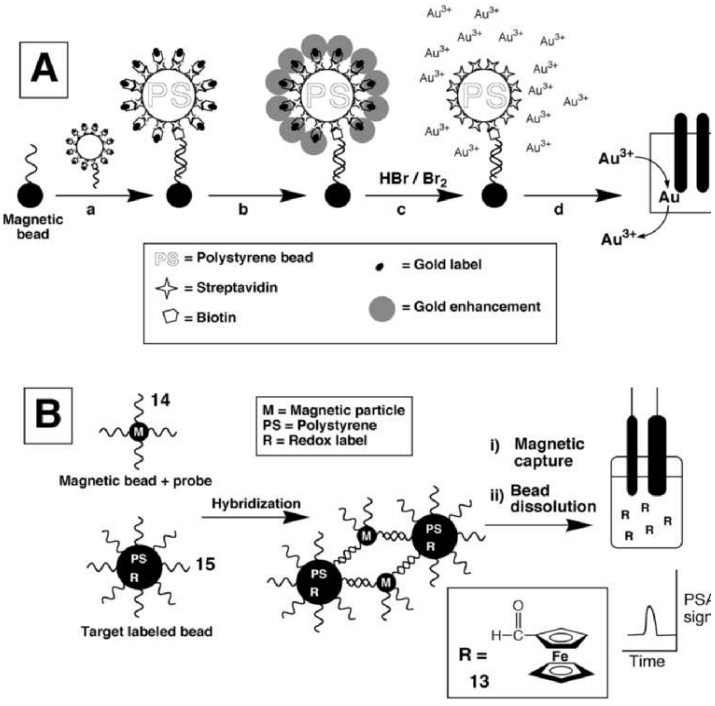


Figure 6. (A) The amplified DNA detection using nucleic acid/Au nanoparticle-functionalized beads as labels and electroless catalytic deposition of gold on the nanoparticles as amplification path: (a) Hybridization of the nucleic acid/Au nanoparticle-functionalized beads with the target DNA associated with a magnetic bead. (b) The enhanced catalytic deposition of gold on the nanoparticles. (c) Dissolution of the gold clusters. (d) The detection of the Au^{3+} ions by stripping voltammetry. (B) The amplified DNA detection based on polystyrene beads loaded with a ferrocene redox marker.

The gold-tagged beads were prepared by binding biotinylated Au nanoparticles to streptavidin-coated polystyrene spheres. These beads were functionalized with a single-stranded oligonucleotide, which was further hybridized with a complementary oligonucleotide linked to a magnetic particle (Figure 6(A), step "a"). The numerous Au nanoparticle labels associated with one ds-oligonucleotide pair were enlarged by electroless deposition of gold (Figure 6(A), step "b") and transported to the electrode array with the use of the magnetic particle. Then the Au assembly was dissolved upon the reaction with HBr/Br_2 and electrochemically analyzed using electrochemical deposition/stripping procedure (Figure 6(A), steps "c" and "d"). Such triple-amplification route offered a dramatic enhancement of the sensitivity. In another approach, carbon nanotubes loaded with many CdS nanoparticles were employed as labels for DNA hybridization [71]. Dissolution of the bound CdS nanoparticles in 1 M HNO_3 followed by the electrochemical detection of the released Cd^{2+} ions provided an amplified signal for the hybridization event.

Internal encapsulation of electroactive tags within carrier beads offers an alternative means to label the probes, and it might reveal some advantages as compared with the external labeling of the probes by nanoparticles. For example, ultrasensitive electrical DNA detection was recently reported based on polystyrene beads impregnated with ferrocenecarboxaldehyde as a redox marker (**13**) [72] (Figure 6(B)). The capturing DNA (**14**) was linked to magnetic particles and the polystyrene beads functionalized with the complementary nucleic acid (**15**) were hybridized with the nucleic acid-modified magnetic particles. Collection of the hybridized system and the dissolution of the beads in an organic solvent released the beads-immobilized redox label (**13**). This allowed the chronopotentiometric detection of the target DNA with a sensitivity that corresponds to the 5.1×10^{-21} mol (~31,000 molecules) under experimental conditions that involved 20 minutes hybridization and the 'release' of the marker by dissolution of the modified beads in an organic medium. The amplified electrochemical readout signal was observed with the remarkable discrimination of a large excess (10^7-fold) of non-complementary nucleic acids revealing the analytical advantages of this sensing process. Further efforts should be directed to encapsulate different redox markers in different polystyrene host beads that could allow parallel multi-target DNA detection. Other marker encapsulation routes hold great promise for electrical DNA detection. Particularly attractive are the recently developed nanoencapsulated microcrystalline particles, prepared by the layer-by-layer technique, that offer large marker/biomolecule ratios and superamplified bioassays [73]. Related analytical procedures that combined multiple amplification pathways based on enzyme-functionalized liposomes and the

accumulation of the biocatalytic-reaction product were reported for the ultrasensitive DNA assays [74]. Such bioassay relied on the large surface area of the liposomes that carry a large number of enzyme molecules. Sensing of the accumulated product was accomplished by means of chronopotentiometry.

Photoelectrochemical transduction of DNA recognition processes has been demonstrated by using semiconductor (CdS) nanoparticles modified with nucleic acids [75]. Semiconductor CdS nanoparticles (2.6±0.4 nm) were functionalized with one of the two thiolated nucleic acids (**16**) or (**17**) that are complementary to the 5' and 3' ends of a target DNA (**18**). An array of CdS nanoparticle layers was then constructed on a Au electrode by a layer-by-layer hybridization process (Figure 7(A)).

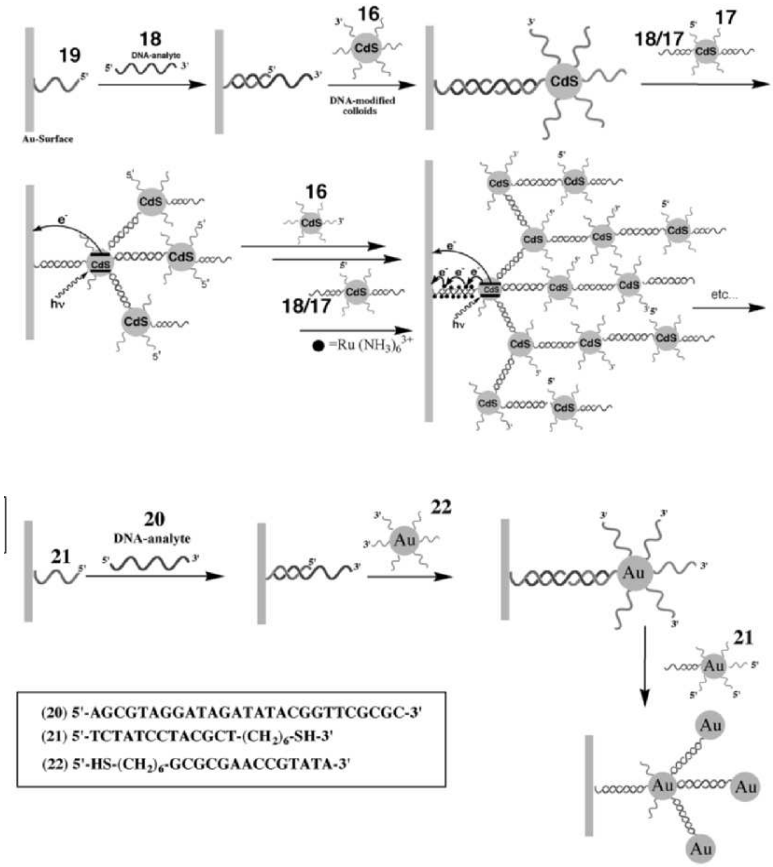


Figure 7. (A) The construction of CdS nanoparticle/DNA superstructures, and their use for generation of photocurrents. (B) Dendritic amplified DNA sensing using oligonucleotide-functionalized Au nanoparticles assembled on a quartz crystal microbalance (QCM) electrode.

A primary thiolated DNA monolayer of **19** was assembled on a Au electrode and the target DNA (**18**) acted as a crosslinking unit for the association of the **16**-modified CdS nanoparticles to the electrode by the hybridization of the ends of **18** to the **19**-modified surface and the **16**-functionalized CdS particles, respectively. The subsequent association of the second type of **17**-modified CdS particles hybridized to the first generation of the CdS particles resulted in the second generation of CdS particles. By the stepwise application of the two different kinds of nucleic acid-functionalized CdS nanoparticles hybridized with **18**, an array with a controlled number of nanoparticle generations could be assembled on the electrode. This array was characterized by spectroscopic means (absorption, fluorescence) upon the assembly of the array on glass supports, and by microgravimetric quartz crystal microbalance analyses on Au-quartz piezoelectric crystals. Illumination of the array resulted in the generation of a photocurrent. The photocurrents increased with the number of CdS nanoparticle generations associated with the electrode, and the photocurrent action spectra followed the absorbance features of the CdS nanoparticles, implying that the photocurrents originated from the photoexcitation of the CdS nanoparticles. That is, photoexcitation of the semiconductor induced the transfer of electrons to the conduction-band and the formation of an electron-hole pair. Transfer of the conduction band electrons to the bulk electrode, and the concomitant transfer of electrons from a sacrificial electron donor to the valence-band holes, yielded the steady-state photocurrent in the system. The ejection of the conduction-band electrons into the electrode occurred from nanoparticles in intimate contact with the electrode support. This was supported by the fact that $Ru(NH_3)_3^{6+}$ units (E^o = -0.16 V vs. SCE), which are electrostatically bound to the DNA, enhanced the photocurrent from the DNA-CdS array. That is, the $Ru(NH_3)_6^{3+}$ units acted as electron wiring elements that facilitated electron hopping of conduction-band electrons from CdS particles that lack contact with the electrode through the DNA tether. The system is important not only because it demonstrates the use of photoelectrochemistry as a transduction method for DNA sensing, but also since the system reveals the nano-engineering of organized DNA-tethered semiconductor nanoparticles on conductive supports. These latter nano-engineered structures are the first step towards electronic nano-circuitry.

Nanoparticles as components of metal-nanoparticle-nucleic acid hybrids represent high molecular weight units that make these conjugates ideal labels for microgravimetric quartz-crystal-analyses of biorecognition processes on the surfaces of piezoelectric crystals. Furthermore, as nanoparticles act as catalysts for the deposition of metals, even higher mass changes may be stimulated and thus the amplified microgravimetric detection of biorecognition processes may be accomplished. For a quartz piezoelectric crystal (AT-cut) the crystal resonance frequency changes by Δf when a mass change Δm occurs on the crystal according to equation 1 (the Sauerbrey

equation [76]), where f_o is the fundamental frequency of the quartz crystal, Δm is the mass change, A is the piezoelectrically active area, ρ_q is the density of quartz (2.648 $g \cdot cm^{-3}$), and μ_q is the shear modulus (2.947×10^{11} $dyn \cdot cm^{-2}$ for AT-cut quartz).

$$\Delta f = -2 \cdot f_o^2 \frac{\Delta m}{A \cdot (\mu_q \cdot \rho_q)^{1/2}} \quad (1)$$

Microgravimetric (QCM) DNA detection using nucleic acid-functionalized Au nanoparticles as "nano-weights" was accomplished by the hybridization of a target DNA (**20**) to an Au-quartz crystal modified with a probe oligonucleotide (**21**), followed by the hybridization of the interface with Au nanoparticles functionalized with DNA (**22**) that is complementary to the free 3'-segment of the target DNA (**20**) [77] (Figure 7(B)). Further amplification of the response was reported by the use of a secondary Au nanoparticle that is functionalized with the nucleic acid (**21**) that is complementary to the 5'-segment of the target DNA (**20**) and enables a layer-by-layer deposition of the Au nanoparticles. The hybridization of the **21**-modified Au nanoparticles with the analyzed DNA (**20**) followed by hybridization of the complex to the primary nanoparticle layer, yielded a 'second generation' of Au nanoparticles reminiscent of the growth of dendrimers [77,78]. Concentrations as low as 1×10^{-10} M of DNA could be sensed by the amplification of the target DNA by the nucleic acid-functionalized Au nanoparticle labels. It has been shown that the increase of the size of the Au nanoparticles labels from 10 nm up to ca. 40-50 nm results in an enhanced Δf signal, thus increasing the amplification factor in the DNA analysis [79]. Further increase of the Au nanoparticle size, however, resulted in smaller changes in the microgravimetric signal because of incomplete hybridization of the DNA analyte due to too large size of the labeling particles.

A further method for the amplified microgravimetric quartz-crystal-microbalance analysis of DNA utilized the catalytic metal deposition on the nanoparticle labels [80]. Figure 8(A) depicts the amplified detection of the 7249-base M13mp18 DNA using the catalytic deposition of gold on a Au nanoparticle conjugate [81]. The DNA primer (**23**) was assembled on a Au/quartz crystal. After hybridization with M13mp18 DNA (**24**), the double-stranded assembly was replicated in the presence of the mixture of nucleotides (dNTP-mix) dATP, dGTP, dUTP, biotinylated-dCTP (B-dCTP) and polymerase (Klenow fragment). The resulting biotin-labeled replica was then reacted with a streptavidin-Au-nanoparticle conjugate (**25**), and the resulting Au-labeled replica was subjected to the Au nanoparticle catalyzed deposition of gold by the NH_2OH stimulated reduction of $AuCl_4^-$.

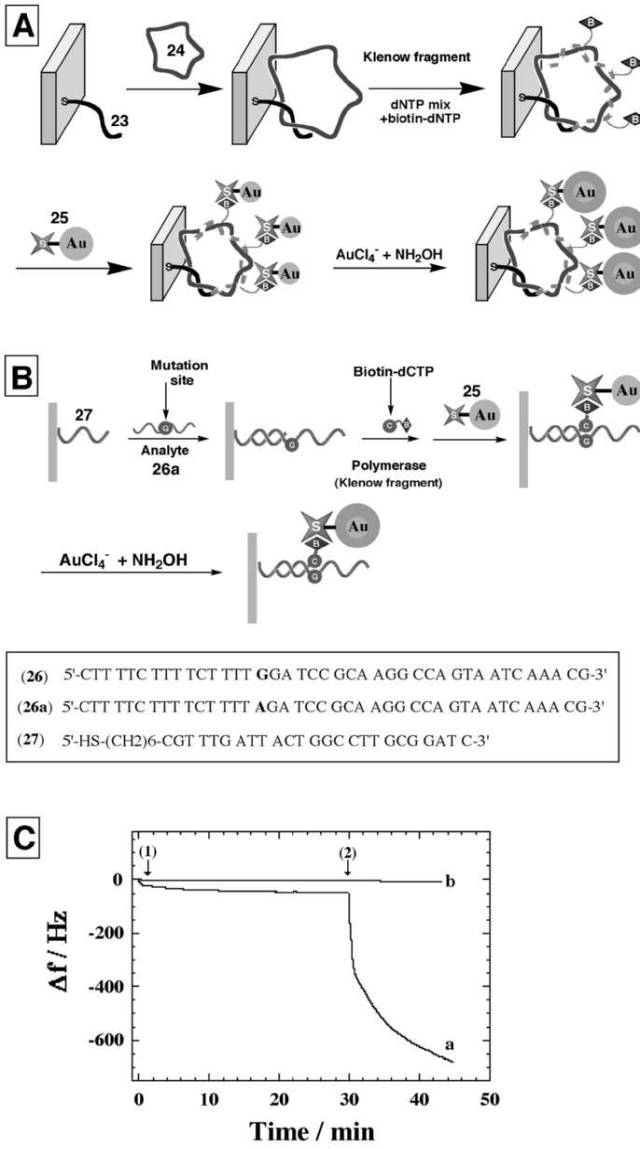


Figure 8. (A) Amplified detection of the 7249-base M13mp18 DNA using the catalytic deposition of gold on a Au nanoparticle conjugate. (B) Analysis of a single-base mismatch in DNA using the catalytic deposition of gold on a Au nanoparticle conjugate. (C) Microgravimetric detection of a single-base mutant enhanced by the catalytic deposition of gold on a Au nanoparticle conjugate: The frequency responses observed with a mutant (a) and with a normal DNA sequence (b). Arrow (1) shows the attachment of the Sav-Au conjugate. Arrow (2) shows catalytic deposition of gold on the Au nanoparticles. (Adapted from ref. 81, Scheme 1 and Figure 2, with permission).

The replication process represents the primary amplification since it increases the mass associated with the crystal, and simultaneously generates a high number of biotin labels for the association of the streptavidin-Au nanoparticle conjugate.

The binding of the nanoparticle conjugate represents the secondary amplification step for the analysis of the M13mp18 DNA. The third step, involving the catalyzed precipitation of the metal, led to the highest amplification in the sensing process due to the increase in the mass of the Au nanoparticle. The M13mp18 DNA could be sensed by this method with a lower detection limit of ca. 1×10^{-15} M. This amplification route was also applied for the analysis of a single base mismatch in DNA as depicted in Figure 8(B). This is exemplified by the analysis of the DNA mutant (**26a**) that includes the single base substitution of the A-base in the normal gene (**26**) with a G-base. The analysis of the mutant was performed by the immobilization of the probe DNA (**27**) that is complementary to the normal gene (**26**) as well as to the mutant (**26a**) up to one base prior to the mutation site, on the Au-quartz crystal. Hybridization of the normal gene or the mutant with this probe interface, followed by the reaction of the hybridized surfaces with biotinylated-dCTP (B-dCTP) in the presence of polymerase (Klenow fragment) incorporated the biotin-labeled base only into the assembly that included the mutant (**26a**). The subsequent association of the streptavidin-Au conjugate (**73**), followed by the catalyzed deposition of gold on the Au nanoparticles, amplified the analysis of the single base mismatch in (**26a**). Figure 8(C), curve (a), shows the microgravimetric detection of the mutant (**26a**) by this method while the normal gene (**26**) does not alter the frequency of the crystal (Figure 8(C), curve (b)). Using this method, the mutant could be detected with a detection limit of 5×10^{-13} M.

4. CONCLUSIONS

This article has summarized recent advances in the rapidly developing area of functional biomaterial-nanoparticle hybrid systems. This topic represents an interdisciplinary effort to combine the unique electronic and catalytic properties of nano-objects with the naturally optimized recognition and reactivity functions of biomaterials. The fact that nanoparticles and biomaterials such as enzymes, antibodies or nucleic acids are of similar dimensions turns the hybrid systems into attractive nanoelements or building blocks of nanostructures and devices.

The analytical applications of nanoparticle-biomaterial systems have advanced tremendously in the last decade. The catalytic properties of biomaterials have enabled the amplification of biorecognition events. Upon the assembly of nanoparticle-biomaterial hybrid systems on surfaces, the electronic detection of biorecognition events can become feasible by

electrical conductivity, piezoelectric or photoelectrochemical transduction means.

The electronic triggering of redox proteins by the incorporation of nanoparticles represents a novel strategy for the electrical contacting of redox enzymes with their macroscopic environment. The use of other nano-objects, such as metal or semiconductor nanorods or carbon nanotubes, for the electrical contacting of redox enzymes may be envisaged.

REFERENCES

[1] *Metal Nanoparticles. Synthesis, Characterization and Application.* D.L. Feldheim, C.A. Foss, Jr., eds. New York: Marcel Dekker, 2002.

[2] a) Badia A., Singh S., Demers L., Cuccia L., Brown G.R., Lennox R.B. Self-assembled monolayers on gold nanoparticles. Chem. Eur. J. 1996; 2:359-363; b) Brust M., Fink J., Bethell D., Schiffrin D.J., Kiely C. Synthesis and reactions of functionalized gold nanoparticles. J. Chem. Soc., Chem. Commun. 1995; 1655-1666.

[3] a) Yao H., Momozawa O., Hamatani T., Kimura K. Stepwise size-selective extraction of carboxylate-modified gold nanoparticles from an aqueous suspension into toluene with tetraoctylammonium cations. Chem. Mater. 2001; 13:4692-4697; b) Hussain N., Singh B., Sakthivel T., Florence A.T. Formulation and stability of surface-tethered DNA–gold–dendron nanoparticles. Int. J. Pharm. 2003; 254:27-31; c) Templeton A.C., Chen S., Gross S.M., Murray R.W. Water-soluble, isolable gold clusters protected by tiopronin and coenzyme A monolayers. Langmuir 1999; 15:66-76.

[4] a) Miyazaki A., Nakano Y. Morphology of platinum nanoparticles protected by poly(*N*-isopropylacrylamide). Langmuir 2000; 16:7109-7111; b) Wuelfing W.P., Gross S.M., Miles D.T., Murray R.W. Nanometer gold clusters protected by surface-bound monolayers of thiolated poly(ethylene glycol) polymer electrolyte. J. Am. Chem. Soc. 1998; 120:12696-12697; c) Teranishi T., Kiyokawa I., Miyake M. Synthesis of monodisperse gold nanoparticles using linear polymers as protective agents. Adv. Mater. 1998; 10:596-599.

[5] Valina-Saba M., Bauer G., Stich N., Pittner F., Schalkhammer T. A self assembled shell of 11-mercaptoundecanoic aminophenylboronic acids on gold nanoclusters. Mater. Sci. Eng. C 1999; 8-9:205-209.

[6] Shipway A.N., Katz E., Willner I. Nanoparticle arrays on surfaces for electronic, optical, and sensor applications. ChemPhysChem 2000; 1:18-52.

[7] Lahav M., Shipway A.N., Willner I. Au-nanoparticle–bis-bipyridinium cyclophane superstructures: assembly, characterization and sensoric applications. J. Chem. Soc. Perkin Trans. 2 1999; 1925-1931.

[8] Chen W., Grouquist D., Roark J. Voltage tunable electroluminescence of CdTe nanoparticle light-emitting diodes. J. Nanosci. Nanotechnol. 2002; 2:47-53.

[9] Shipway A.N., Willner I. Nanoparticles as structural and functional units in surface-confined architectures. Chem. Comm. 2001; 2035-2045.

[10] a) Feldheim D.L., Keating C.D. Self-assembly of single electron transistors and related devices. Chem. Soc. Rev. 1998; 27:1-12; b) Kim T.W., Choo D.C., Shim J.H., Kang S.O. Single-electron transistors operating at room temperature, fabricated utilizing nanocrystals created by focused-ion beam. Appl. Phys. Lett. 2002; 80:2168-2170.

[11] a) Long B., Nikitin K., Fitzmaurice D. Assembly of an electronically switchable rotaxand on the surface of a titanium dioxide nanoparticle. J. Am. Chem. Soc. 2003; 125:15490-15498; b) Willner I, Willner B. Functional nanoparticle architectures for

sensoric, optoelectronic, and bioelectronic applications. Pure Appl. Chem. 2002; 74:1773-1783.

[12] Feldheim D.L., Grabar K.C., Natan M.J., Mallouk T.E. Electron transfer in self-assembled inorganic polyelectrolyte/metal nanoparticle heterostructures. J. Am. Chem. Soc. 1996; 118:7640-7641.

[13] a) Trindade T., O'Brien P., Pickett N.L. Nanocrystalline semiconductors: synthesis, properties, and perspectives. Chem. Mater. 2001; 13:3843-3858; b) Lue J.-T. A review of characterization and physical property studies of metallic nanoparticles. J. Phys. Chem. Solids 2001; 62:1599-1612; c) Grieve K., Mulvaney P., Grieser F. Synthesis and electronic properties of semiconductor nanoparticles/quantum dots. Cur. Opin. Colloid Interface Sci. 2000; 5:168-172; d) Schwerdtfeger P. Gold goes nano—from small clusters to low-dimensional assemblies. Angew. Chem. Int. Ed. 2003; 42:1892-1895.

[14] a) Brust M., Kiely C.J. Some recent advances in nanostructure preparation from gold and silver particles: a short topical review. Colloid Surf. A 2002; 202:175-186; b) McConnell W.P., Novak J.P., Brousseau III L.C., Fuierer R.R., Tenent R.C., Feldheim D.L. Electronic and optical properties of chemically modified metal nanoparticles and molecularly bridged nanoparticle arrays. J. Phys. Chem. B 2000; 104:8925-8930; c) Gangopadhyay R., De A. Conducting polymer nanocomposites: a brief overview. Chem. Mater. 2000; 12:608-622.

[15] Katz E., Shipway A.N., Willner I. "Biomaterial-nanoparticle hybrid systems: synthesis, properties and applications." In *Nanoparticles – From Theory to Applications*, G. Schmid. ed. Weinheim: Wiley-VCH, Chapter 6, pp. 368-421, 2003.

[16] a) Niemeyer C.M. Nanoparticles, proteins, and nucleic acids: biotechnology meets materials science. Angew. Chem. Int. Ed. 2001; 40:4128-4158; b) Niemeyer C.M. Functional hybrid devices of proteins and inorganic nanoparticles. Angew. Chem. Int. Ed. 2003; 42:5796-5800; c) Parak W.J., Gerion D., Pellegrino T., Zanchet D., Micheel C., Williams S.C., Boudreau R., Le Gros M.A., Larabell C.A., Alivisatos A.P. Biological applications of colloidal nanocrystals. Nanotechnology 2003; 14:R15-R27; d) Csaki A., Maubach G., Born D., Reichert J., Fritzsche W. DNA-based molecular nanotechnology. Single Molecules 2002; 3:275-280.

[17] a) Penn S.G., Hey L., Natan M.J. Nanoparticles for bioanalysis. Curr. Opin. Chem. Biol. 2003; 7:609-615; b) West J.L., Halas N.J. Engineered nanomaterials for biophotonics applications: improving sensing, imaging, and therapeutics. Annu. Rev. Biomed. Eng. 2003; 5:285-292.

[18] Armstrong F.A., Wilson G.S. Recent developments in faradaic bioelectrochemistry. Electrochim. Acta 2000; 45:2623-2645.

[19] Willner I., Katz E. Integration of layered redox-proteins and conductive supports for bioelectronic applications. Angew. Chem. Int. Ed. 2000; 39:1180-1218.

[20] a) Willner I., Katz E., Willner B. "Layered functionalized electrodes for electrochemical biosensor applications." In: *Biosensors and Their Applications*, V.C. Yang, T.T. Ngo, eds. New York: Kluwer Academic Publishers, Chapter 4, pp. 47-98, 2000; b) Willner I., Willner B., Katz E. Functional biosensor systems via surface-nanoengineering of electronic elements. Rev. Molec. Biotechnol. 2002; 82:325-355; c) Habermüller L., Mosbach M., Schuhmann W. Electron-transfer mechanisms in amperometric biosensors. Fresenius J. Anal. Chem. 2000; 366:560-568; d) Armstrong F.A., Heering H.A., Hirst J. Reactions of complex metalloproteins studied by protein-film voltammetry. Chem. Soc. Rev. 1997; 26:169-179.

[21] a) Willner I., Arad G., Katz E. A biofuel cell based on pyrroloquinoline quinone and microperoxidase-11 monolayer-functionalized electrodes. Bioelectrochem. Bioenerg. 1998; 44:209-214; b) Willner I., Katz E., Patolsky F., Bückmann A.F. A biofuel cell based on glucose oxidase and microperoxidase-11 monolayer-functionalized-electrodes. J. Chem. Soc. Perkin. Trans. 2 1998; 1817-1822; c) Katz E., Filanovsky B., Willner I. A biofuel cell based on two immiscible solvents and glucose oxidase and microperoxidase-11 monolayer-functionalized electrodes. New J. Chem. 1999; 23:481-487; d) Katz E.,

Willner I., Kotlyar A.B. A non-compartmentalized glucose-O_2 biofuel cell by bioengineered electrode surfaces. J. Electroanal. Chem. 1999; 479:64-68; e) Barton S.C., Kim H.-H., Binyamin G., Zhang Y., Heller A. The "wired" laccase cathode: high current density electroreduction of O_2 to water at +0.7 V (NHE) at pH 5. J. Am. Chem. Soc. 2001; 123:5802-5803; f) Chen T., Barton S.C., Binyamin G., Gao Z., Zhang Y., Kim H.-H., Heller A. A miniature biofuel cell. J. Am. Chem. Soc. 2001; 123:8630-8631; g) Katz E., Shipway A.N., Willner I. "Biofuel cells: Functional design and operation." In *Handbook of Fuel Cells - Fundamentals, Technology, Applications*, W. Vielstich, H. Gasteiger, A. Lamm, eds. London: Wiley, Vol. 1, Part 4, Chapter 21, pp. 355-381, 2003.

[22] Katz E., Shipway A.N., Willner I. "Mediated electron-transfer between redox-enzymes and electrode supports." In *Encyclopedia of Electrochemistry*, Vol. 9: Bioelectrochemistry, G.S. Wilson, ed. Weinheim: Wiley-VCH GmbH, 2002, Chapter 17, pp. 559-626.

[23] a) Degani Y., Heller A. Direct electrical communication between chemically modified enzymes and metal electrodes. 1. Electron transfer from glucose oxidase to metal electrodes via electron relays bound covalently to the enzyme. J. Phys. Chem. 1987; 91:1285-1289; b) Schuhmann W., Ohara T.J., Schmidt H.-L., Heller A. Electron transfer between glucose oxidase and electrodes via redox mediators bound with flexible chains to the enzyme surface. J. Am. Chem. Soc. 1991; 113:1394-1397; c) Degani Y., Heller A. Direct electrical communication between chemically modified enzymes and metal electrodes. 2. Methods for bonding electron-transfer relays to glucose oxidase and D-amino-acid oxidase. J. Am. Chem. Soc. 1988; 110:2615-2620; d) Willner I., Riklin A., Shoham B., Rivenzon D., Katz E. Development of novel biosensor enzyme-electrodes: glucose oxidase multilayer arrays immobilized onto self-assembled monolayers on electrodes. Adv. Mater. 1993; 5:912-915; e) Willner I., Katz E., Riklin A., Kasher R. Electron transfer communication in glutathione reductase assemblies: electrocatalytic, photocatalytic and catalytic systems for the reduction of oxidized glutathione. J. Am. Chem. Soc. 1992; 114:10965-10966; f) Badia A., Carlini R., Fernandez A., Battaglini F., Mikkelsen S.R., English A.M. Intramolecular electron-transfer rates in ferrocene-derivatized glucose oxidase. J. Am. Chem. Soc. 1993; 115:7053-7060.

[24] Heller A. Electrical wiring of redox enzymes. Acc. Chem. Res. 1990; 23:128-134.

[25] Emr S.A., Yacynych A.M. Use of polymer-films in amperometric biosensors. Electroanalysis 1995; 7:913-923.

[26] a) Willner I., Heleg-Shabtai V., Blonder R., Katz E., Tao G., Bückmann A.F., Heller A. Electrical wiring of glucose oxidase by reconstitution of FAD-modified monolayers assembled onto Au-electrodes. J. Am. Chem. Soc. 1996; 118:10321-10322; b) Katz E., Riklin A., Heleg-Shabtai V., Willner I., Bückmann A.F. Glucose oxidase electrodes via reconstitution of the apo-enzyme: tailoring of novel glucose biosensors. Anal. Chim. Acta 1999; 385:45-58.

[27] Raitman O.A., Katz E., Bückmann A.F., Willner I. Integration of polyaniline/poly(acrylic acid) films and redox enzymes on electrode supports: an *in situ* electrochemical / surface plasmon resonance study of the bioelectrocatalyzed oxidation of glucose or lactate in the integrated bioelectrocatalytic systems. J. Am. Chem. Soc. 2002; 124:6487-6496.

[28] Raitman O.A., Patolsky F., Katz E., Willner I. Electrical contacting of glucose dehydrogenase by the reconstitution of a pyrroloquinoline quinone-functionalized polyaniline film associated with an Au-electrode: An in situ SPR-electrochemical study. Chem. Commun. 2002; 1936-1937.

[29] a) Guo L.-H., McLendon G., Razafitrimo H., Gao Y. Photo-active and electro-active protein films prepared by reconstitution with metalloporphyrins self-assembled on gold. J. Mater. Chem. 1996; 6:369-374; b) Zimmermann H., Lindgren A., Schuhmann W., Gorton L. Anisotropic orientation of horseradish peroxidase by reconstitution on a thiol-modified gold electrode. Chem. Eur. J. 2000; 6:592-599.

[30] a) Zhao J., Henkens R.W., Stonehurner J., O'Daly J.P., Crumbliss A.L. Direct electron-transfer at horseradish-peroxidase colloidal gold modified electrodes. J. Electroanal. Chem. 1992; 327:109-119; b) Crumbliss A.L., Perine S.C., Stonehurner J., Tubergen K.R., Zhao J., Henkens R.W., O'Daly J.P. Colloidal gold as a biocompatible immobilization matrix suitable for the fabrication of enzyme electrodes by electrodeposition. Biotech. Bioeng. 1992; 40:483-490; c) Zhao J., O'Daly J.P., Henkens R.W., Stonehurner J., Crumbliss A.L. A xanthine oxidase/colloidal gold enzyme electrode for amperometric biosensor applications. Biosens. Bioelectron. 1996; 11:493-502; d) Bharathi S., Nogami M. A glucose biosensor based on electrodeposited biocomposites of gold nanoparticles and glucose oxidase enzyme. Analyst 2001; 126:1919-1922; e) Liu S., Ju H. Electrocatalysis via direct electrochemistry of myoglobin immobilized on colloidal gold nanoparticles. Electroanalysis 2003; 15:1488-1493; f) Wang X.-Y., Zhong H., Lv Y., Chen H.-Y. Fabrication of nanoelectrode ensembles of porous gold nanoshells and direct electrochemistry of horseradish peroxidase immobilized on the electrode. Chem. Lett. 2003; 32:1054-1055; g) Liu S., Ju H. Reagentless glucose biosensor based on direct electron transfer of glucose oxidase immobilized on colloidal gold modified carbon paste electrode. Biosens. Bioelectron. 2003; 19:177-183; h) Liu S., Ju H. Nitrite reduction and detection at a carbon paste electrode containing hemoglobin and colloidal gold. Analyst 2003; 128:1420-1424; i) Gu H.-Y., Sa R.-X., Yuan S.-S., Chen H.-Y., Yu A.-M. The self-assembly, characterization of hepatocytes on nano-sized gold colloid and construction of cellular biosensor. Chem. Lett. 2003; 32:934-935.

[31] Han X., Cheng W., Zhang Z., Dong S., Wang E. Direct electron transfer between hemoglobin and a glassy carbon electrode facilitated by lipid-protected gold nanoparticles. Biochim. Biophysica Acta 2002; 1556:273-277.

[32] Xiao Y., Patolsky F., Katz E., Hainfeld J.F., Willner I. 'Plugging into enzymes': Nanowiring of redox-enzymes by a gold nanoparticle. Science 2003; 299:1877-1881.

[33] Pardo-Yissar V, Katz E., Wasserman J., Willner I. Acetylcholine esterase-labeled CdS nanoparticles on electrodes: photoelectrochemical sensing of the enzyme inhibitors. J. Am. Chem. Soc. 2003; 125:622-623.

[34] Curri M.L., Agostiano A., Leo G., Mallardi A., Cosma P., Monica M.D. Development of a novel enzyme/semiconductor nanoparticles system for biosensor application. Mater. Sci. Eng. C 2002; 22:449-452.

[35] a) Mulvaney P. Surface plasmon spectroscopy of nanosized metal particles. Langmuir 1996; 12:788-800; b) Alvarez M.M., Khoury J.T., Schaaff T.G., Shafigullin M.N., Vezmar I., Whetten R.L. Optical absorption spectra of nanocrystal gold molecules. J. Phys. Chem. B 1997; 101:3706-3712; c) Alivisatos A.P. Perspectives on the physical chemistry of semiconductor nanocrystals. J. Phys. Chem. 1996; 100:13226-13329; d) Brus L. Quantum crystallites and nonlinear optics. Appl. Phys. A 1991; 53:465-474.

[36] a) Hicks J.F., Miles D.T., Murray R.W. Quantized double-layer charging of highly monodisperse metal nanoparticles. J. Am. Chem. Soc. 2002; 124:13322-13328; b) Hicks J.F., Zamborini F.P., Osisek A.J., Murray R.W. The dynamics of electron self-exchange between nanoparticles. J. Am. Chem. Soc. 2001; 123:7048-7053; c) Chen S., Murray R.W. Electrochemical quantized capacitance charging of surface ensembles of gold nanoparticles. J. Phys. Chem. B 1999; 103:9996-10000; d) Hicks J.F., Zamborini F.P., Murray R.W. Dynamics of electron transfers between electrodes and monolayers of nanoparticles. J. Phys. Chem. B 2002; 106:7751-7757; e) Chen S., Murray R.W., Feldberg S.W. Quantized capacitance charging of monolayer-protected Au clusters. J. Phys. Chem. B 1998; 102:9898-9907.

[37] Khairutdinov R.F. Physical chemistry of nanocrystalline semiconductors Colloid J. 1997; 59:535-548.

[38] a) Lewis L.N. Chemical catalysis by colloids and clusters. Chem. Rev. 1993; 93:2693-2730; b) Kesavan V., Sivanand P.S., Chandrasekaran S., Koltypin Y., Gedankin

A. Catalytic aerobic oxidation of cycloalkanes with nanostructured amorphous metals and alloys. Angew. Chem. Int. Ed. 1999; 38:3521-3523.
[39] He L., Musick M.D., Nicewarner S.R., Salinas F.G., Benkovic S.J., Natan M.J., Keating C.D. Colloidal Au-enhanced surface plasmon resonance for ultrasensitive detection of DNA hybridization. J. Am. Chem. Soc. 2000; 122:9071-9077.
[40] a) Kubitschko S., Spinke J., Brückner T., Pohl S., Oranth N. Sensitivity enhancement of optical immunosensors with nanoparticles. Anal. Biochem. 1997; 253:112-122; b) Lyon L.A., Musick M.D., Natan M.J. Colloidal Au-enhanced surface plasmon resonance immunosensing. Anal. Chem. 1998; 70:5177-5183; c) Englebienne P., Hoonacker A.V., Verhas M. High-throughput screening using the surface plasmon resonance effect of colloidal gold nanoparticles. Analyst 2001; 126:1645-1651.
[41] Niemeyer, C.M. Nanoparticles, proteins, and nucleic acids: biotechnology meets materials science. Angew. Chem. Int. Ed. 2001; 40:4128-4158.
[42] Bruchez M., Jr., Moronne M., Gin P., Weiss S., Alivisatos A.P. Semiconductor nanocrystals as fluorescent biological labels. Science 1998; 281:2013-2015.
[43] Ozsoz M., Erdem A., Kerman K., Ozkan D., Tugrul B., Topcuoglu N., Ekren H., Taylan M. Electrochemical genosensor based on colloidal gold nanoparticles for the detection of factor V Leiden mutation using disposable pencil graphite electrodes. Anal. Chem. 2003, 75:2181-2187.
[44] Wang, Joseph, *Stripping Analysis*. New York: VCH, 1985.
[45] Zhu N., Cai H., He P., Fang Y. Tris(2,2'-bipyridyl)cobalt(III)-doped silica nanoparticle DNA probe for the electrochemical detection of DNA hybridization. Anal. Chim. Acta 2003; 481:181-189.
[46] Wang J., Li J., Baca A.J., Hu J., Zhou F., Yan W., Pang D.-W. Amplified voltammetric detection of DNA hybridization via oxidation of ferrocene caps on gold nanoparticle/streptavidin conjugates. Anal. Chem. 2003; 75:3941-3945.
[47] Wang J., Xu D., Kawde A.-N., Polsky R. Metal nanoparticle-based electrochemical stripping potentiometric detection of DNA hybridization. Anal. Chem. 2001; 73:5576-5581.
[48] Authier L., Grossiord C., Brossier P., Limoges B. Gold nanoparticle-based quantitative electrochemical detection of amplified human cytomegalovirus DNA using disposable microband electrodes. Anal. Chem. 2001; 73:4450-4456.
[49] Cai H., Xu Y., Zhu N., He P., Fang Y. An electrochemical DNA hybridization detection assay based on a silver nanoparticle label. Analyst 2002; 127:803-808.
[50] Cai H., Zhu N., Jiang Y., He P., Fang Y. Cu@Au alloy nanoparticle as oligonucleotides labels for electrochemical stripping detection of DNA hybridization. Biosens. Bioelectron. 2003; 18:1311-1319.
[51] Wang J., Liu G., Zhu Q. Indium microrod tags for electrochemical detection of DNA hybridization. Anal. Chem. 2003; 75:6218-6222.
[52] Drummond T.G., Hill M.G., Barton J.K. Electrochemical DNA sensors. Nature Biotechnol. 2003; 21:1192-1199.
[53] Lee T.M.-H., Li L.-L., Hsing I.-M. Enhanced electrochemical detection of DNA hybridization based on electrode-surface modification. Langmuir 2003; 19:4338-4343.
[54] Wang J., Polsky R., Xu D. Silver-enhanced colloidal gold electrochemical stripping detection of DNA hybridization. Langmuir 2001; 17:5739-5741.
[55] Wang J., Rincón O., Polsky R., Dominguez E. Electrochemical detection of DNA hybridization based on DNA-templated assembly of silver cluster. Electrochem. Commun. 2003; 5:83-86.
[56] Richter J. Metallization of DNA. Physica E 2003; 16:157-173.
[57] Richter J., Seidel R., Kirsch R., Mertig M., Pompe W., Plaschke J., Schackert H.K. Nanoscale palladium metallization of DNA. Adv. Mater. 2000; 12:507-510.
[58] Braun E., Eichen Y., Sivan U., Ben-Yoseph G. DNA-templated assembly and electrode attachment of a conducting silver wire. Nature 1998; 391:775-778.

[59] Eichen Y., Braun E., Sivan U., Ben-Yoseph G. Self-assembly of nanoelectronic components and circuits using biological templates. Acta Polymerica 1998; 49:663-670.
[60] Mertig M., Ciacchi L.C., Seidel R., Pompe W., De Vita A. DNA as a selective metallization template. Nano Lett. 2002; 2:841-844.
[61] Velev O.D., Kaler E.W. In situ assembly of colloidal particles into miniaturized biosensors. Langmuir 1999; 15:3693-3698.
[62] Park S.-J., Taton T.A., Mirkin C.A. Array-based electrical detection of DNA with nanoparticle probes. Science 2002; 295:1503-1506.
[63] Urban M., Möller R., Fritzsche W. A paralleled readout system for an electrical DNA-hybridization assay based on a microstructured electrode array. Rev. Sci. Instr. 2003; 74:1077-1081.
[64] Doron A., Katz E., Willner I. Organization of Au colloids as monolayer films onto ITO glass surfaces: application of the metal colloid films as base interfaces to construct redox-active monolayers. Langmuir 1995; 11:1313-1317.
[65] Cai H., Xu C., He P., Fang Y. Colloid Au-enhanced DNA immobilization for the electrochemical detection of sequence-specific DNA. J. Electroanal. Chem. 2001; 510:78-85.
[66] Lin H., Zhao H., Li J., Tang J., Duan M., Jiang L. Study on colloidal Au-enhanced DNA sensing by quartz crystal microbalance. Biochem. Biophys. Res. Comm. 2000; 274:817-820.
[67] a) Wang M., Wang L., Wang G., Ji X., Bai Y., Li T., Gongb S., Li J. Application of impedance spectroscopy for monitoring colloid Au-enhanced antibody immobilization and antibody–antigen reactions. Biosens. Bioelectron. 2004; 19:575-582; b) Lei C.-X., Gonga F.-C., Shen G.-L., Yu R.-Q. Amperometric immunosensor for *Schistosoma japonicum* antigen using antibodies loaded on a nano-Au monolayer modified chitosan-entrapped carbon paste electrode. Sens. Actuat. B 2003; 96:582-588; c) Hua S.-Q., Xie J.-W., Xu Q.-H., Rong K.-T., Shen G.-L., Yu R.-Q. A label-free electrochemical immunosensor based on gold nanoparticles for detection of paraoxon. Talanta 2003; 61:769-777.
[68] Katz E., Shipway A.N., Willner I. "Chemically functionalized metal nanoparticles: synthesis, properties and applications." In *Nanoscale Materials*, L.M. Liz-Marzan, P. Kamat, eds. New York: Kluwer, 2003, Chapter 2, pp. 5-78.
[69] Wang J., Liu G., Polsky R., Merkoçi A. Electrochemical stripping detection of DNA hybridization based on cadmium sulfide nanoparticle tags. Electrochem. Commun. 2002; 4:722-726.
[70] Wang J., Liu G., Merkoçi A. Electrochemical coding technology for simultaneous detection of multiple DNA targets. J. Am. Chem. Soc. 2003; 125:3214-3215.
[71] a) Wang, J., Kawde A. Amplified electrical transduction of DNA hybridization based on polymeric beads loaded with multiple gold nanoparticle tags. Electroanalysis 2004; 16: in press; b) Wang J., Liu G., Jan M.R., Zhu Q. Electrochemical detection of DNA hybridization based on carbon-nanotubes loaded with CdS tags. Electrochem. Commun. 2003; 5:1000-1004.
[72] Wang J., Polsky R., Merkoçi A., Turner K. "Electroactive beads" for ultrasensitive DNA detection. Langmuir 2003; 19:989-991.
[73] Trau D., Yang W.J., Seydack M., Carusu F., Yu N.-T., Renneberg R. Nanoencapsulated microcrystalline particles for superamplified biochemical assays. Anal. Chem. 2002; 74:5480-5486.
[74] Alfonta L., Singh A., Willner I. Liposomes labeled with biotin and horseradish peroxidase: a probe for the enhanced amplification of antigen-antibody or oligonucleotide-DNA sensing processes by the precipitation of an insoluble product on electrodes. Anal. Chem. 2001; 73:91-102.
[75] Willner I., Patolsky F., Wasserman J. Photoelectrochemistry with controlled DNA-crosslinked CdS nanoparticle arrays. Angew. Chem. Int. Ed. 2001; 40:1861-1864.

[76] Buttry D.A., Ward M.D. Measurement of interfacial processes at electrode surfaces with the electrochemical quartz crystal microbalance. Chem. Rev. 1992; 92:1355-1379.
[77] a) Zhou X.C., O'Shea S.J., Li S.F.Y. Amplified microgravimetric gene sensor using Au nanoparticle modified oligonucleotides. Chem. Commun. 2000; 953-954; b) Patolsky F., Ranjit K.T., Lichtenstein A., Willner I. Dendritic amplification of DNA analysis by oligonucleotide-functionalized Au-nanoparticles. Chem. Commun. 2000; 1025-1026.
[78] Han S., Lin J., Satjapipat M., Baca A.J., Zhou F. A three-dimensional heterogeneous DNA sensing surface formed by attaching oligodeoxynucleotide-capped gold nanoparticles onto a gold-coated quartz crystal. Chem. Commun. 2001; 609-610.
[79] Liu T., Tang J., Zhao H., Deng Y., Jiang L. Particle size effect of the DNA sensor amplified with gold nanoparticles. Langmuir 2002; 18:5624-5626.
[80] Willner I., Patolsky F., Weizmann Y., Willner B. Amplified detection of single-base mismatches in DNA using microgravimetric quartz-crystal-microbalance transduction. Talanta 2002; 56:847-856.
[81] Weizmann Y., Patolsky F., Willner I. Amplified detection of DNA and analysis of single-base mismatches by the catalyzed deposition of gold on Au-nanoparticles. Analyst 2001; 126:1502-1504.

Index