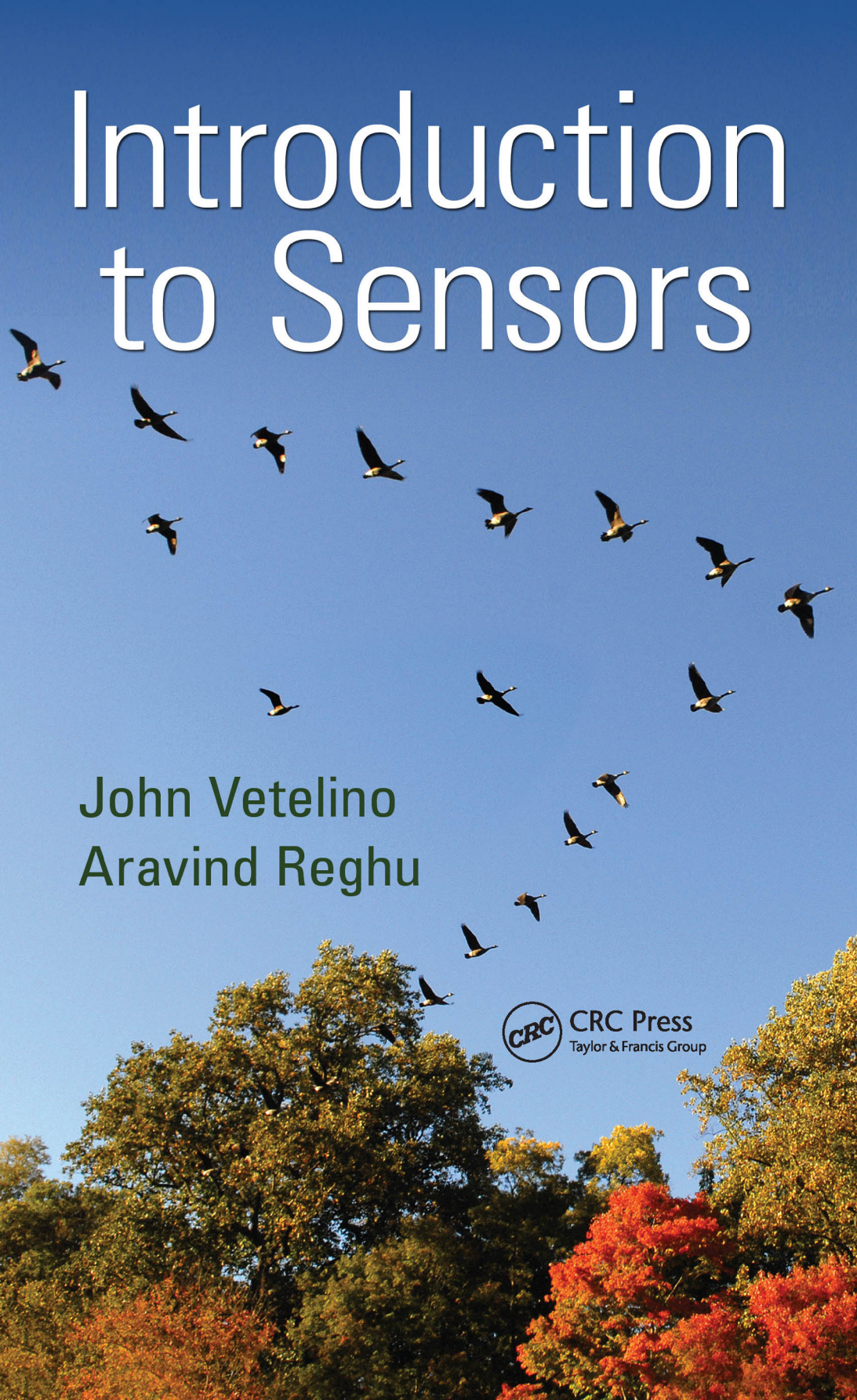
Introduction to Sensors



John Vetelino
Aravind Reghu

CRC Press
Taylor & Francis Group

Introduction to Sensors

Introduction to Sensors

John Vetelino

Aravind Reghu

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

International Standard Book Number: 978-1-4398-0852-8 (Hardback)

Library of Congress Cataloging-in-Publication Data

Vetelino, John.
Introduction to sensors / authors, John Vetelino, Aravind Reghu.
p. cm.
"A CRC title."
Includes bibliographical references and index.
ISBN 978-1-4398-0852-8 (hardcover : alk. paper)
1. Detectors. I. Reghu, Aravind. II. Title.

TK7872.D48V48 2010
681'.2--dc22 2009045582

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

This book came about as a result of a need to present fundamental material to undergraduate and graduate engineering and science students in the general area of sensors. The initial funding for the development of the book came from a National Science Foundation (NSF) grant relating to the integration of emerging technologies such as sensors into engineering and science curricula. Coincident with the NSF grant, sensors became an area of critical need fueled by terrorist actions such as 9/11, national defense, and the health and security of the general public. A course entitled "Introduction to Sensors" was then introduced, and course notes were developed and distributed to students taking the course.

The first chapter introduces students to the fundamentals of sensors independent of the underlying technology. Terms such as *sensor signature*, receiver operating characteristics, and the specific sensor properties are defined and discussed. The human body and the automobile are treated as entities that utilize a multiplicity of sensors. Finally, a complete sensor system that includes the preparation phase, the sensing element and platform, and appropriate electronics resulting in a digital readout is presented.

In the second chapter solid-state electronic sensors whose response relates to electricity and chemistry are presented. The major focus of this chapter is the metal oxide semiconducting sensor commonly referred to as the Taguchi sensor. A brief overview of energy band theory as applied to a metal oxide film is presented. The gas-film redox reactions are discussed relative to gas sensing by the film along with the effect of film structure, dopant, and temperature on the sensor response. The chapter concludes with a brief discussion of solid-state electronic sensors such as the metal oxide semiconductor (MOS) capacitor, the micromachined capacitive polymer, and the Schottky diode sensors.

The third chapter is devoted to piezoelectric sensors. The basic theory associated with piezoelectricity is described using the two-dimensional hexagonal lattice as an example. The fundamental relationship between stress, strain, electric field, and electric displacement is presented for both nonpiezoelectric and piezoelectric crystals. An in-depth design of a passive piezoelectric force/pressure sensor is presented with specific applications given. The principle of operation of an accelerometer is presented along with specific applications. The chapter concludes with material relating to active piezoelectric sensors. In particular, the standard quartz crystal monitor (QCM) is discussed along with the lateral field excited (LFE) acoustic wave sensor developed at the University of Maine. Specific advantages and disadvantages of the QCM and LFE sensors are discussed along with specific sensor applications. Finally, a brief discussion of the surface acoustic wave sensor is presented.

The last three chapters relate to thermal, fiber optic, and magnetic sensors. A brief discussion of the theory underlying each sensor technology is presented followed by specific sensor applications. Relative to thermal sensors, both the resistive thermal device (RTD) and thermistor are discussed. The fiber optic sensor is presented with specific applications such as embedded sensors. The magnetic sensors presented are used to determine measurands such as the magnetic field and semiconductor properties such as carrier concentration and mobility.

Acknowledgments

First and foremost, the authors thank the large number of colleagues and undergraduate and graduate students associated with the Laboratory for Surface Science and Technology and the Electrical and Computer Engineering Department at the University of Maine who made significant contributions. In particular, special thanks are extended to Dean Smith and Zhimin Xu for their valuable contribution relating to the electrical behavior of metal oxide films exposed to the ambient and various target gases. Thanks are also extended to Dr. Todd Mlsna for his input relating to micromachined capacitive polymer gas sensors. Relative to the piezoelectric sensors, the authors owe a debt of gratitude to Dr. Ryszard Lec for his valuable input relating to passive piezoelectric sensors such as the stress/pressure sensor and the accelerometer. Thanks are also extended to Dr. Chao Zhang, Y. Hu, Lester French, Don McCann, Mitchell Wark, Jason McGann, Dr. Jeff Andle, Dr. Robert Falconer, Dr. George Bernhardt, Dr. David Frankel, and Dr. David Kotecki for their valuable input relating to active piezoelectric sensors such as the quartz crystal monitor (QCM), the lateral field excited (LFE) sensor, and the monolithic spiral coil sensor.

The book would not have been possible without the dedicated time and effort expended by Barbara Deschane, Susan Niles, and Laura Hall in its preparation.

Finally, John F. Vetelino extends a special thanks to his wife, Leah, and sons, Kevin and Jay, for their constant moral support, which provided motivation and encouragement in the preparation of this book. Aravind Reghu expresses his gratitude to his wife, Anusha, his parents, Reghu and Geetha, and his brother, Gautham, for their love and support.

The Authors

John F. Vetelino, IEEE Fellow, earned his BS, MS, and PhD degrees in electrical engineering from the University of Rhode Island in 1964, 1966, and 1969, respectively. He has been with the University of Maine since 1969 and is currently professor of electrical and computer engineering. He was one of the founding members of the Laboratory for Surface Science and Technology at the University of Maine and currently is the leader of the solid-state sensor group.

His research group is working on chemical and biological sensors based on acoustic wave and chemiresistive technology. In 1976, he was at the Max Planck Institute Fuer Festkorperforschung in Stuttgart, West Germany, working on solid-state properties of piezoelectric crystals. In 1980, he was awarded the Presidential Research Achievement Award, given annually to the outstanding researcher at the University of Maine. From 1983 to 1987, he was chairman of the Electrical Engineering Department at the University of Maine. In 2008 he received the Distinguished Maine Professor award.

Dr. Vetelino has given invited talks at many universities both in the United States and abroad, at national and international conferences. He also has published several invited papers and received outstanding paper awards at the International Chemical Sensors meeting, the Society of Plastic Engineers Conference, and the IEEE Ultrasonics Symposium. He was a guest editor of the special issue of the IEEE UFFC *Transactions on Sensors and Actuators* in 1998. He currently is a member of the technical program committee for the IEEE Ultrasonics group. Four sensor companies—BIODE Corporation in Westbrook, Maine; Microsensor Conversion Technology in Brookings, South Dakota; Sensor Research and Development Corporation in Orono, Maine; and Mainely Sensors, LLC in Orono, Maine—have been incubated from his research group. He is a member of Phi Kappa Phi, Sigma Xi, Tau Beta Pi, and Eta Kappa Nu. Dr. Vetelino has advised over 40 graduate students to MS and PhD degrees and has published over 200 scientific papers.

Aravind Reghu completed his bachelors degree in electronics and communications engineering from the University of Calicut in India and his masters in electrical engineering from the University of Maine. He is currently pursuing his PhD in electrical engineering from the University of Maine. His area of research is in the field of metal oxide sensors, where he is designing and developing a prototype tool for *in situ* sampling of ice core chemistry utilizing innovative thin-film metal oxide gas sensor technology. The thin-film metal oxide sensor being designed can be used to identify gases like methane and carbon dioxide. The author has also been involved in a number of projects in the fields of microwave engineering, very large scale integrated (VLSI) circuit design, and microelectronics.

1

Introduction

1.1 Background

Sensors dominate the world in which we live. We are awakened in the morning by an alarm clock and proceed to breakfast in our home, which has smoke detectors. Our breakfast is often prepared with the aid of timing devices, and temperature sensors tell us when the food is ready to eat. We then proceed to go to work in our automobile, which may have more than 50 different sensors helping in the operation and control of the vehicle. At our workplace we are surrounded by sensors that can be very simple, such as temperature and humidity sensors. Other workplace sensors are quite complex, such as sensors to detect potential explosives, metal objects, etc. If our work requires travel, then we encounter a wide spectrum of sensors in the detection of a variety of objects that may be on our person or in a suitcase.

In addition to sensors relating directly to the individual, a whole list of sensors exist on local, state, national, and global levels. For example, traffic signals in your town are often controlled by sensors. At the state level, a variety of sensors help us in terms of waste management and water quality. Sensors are present at a national level and are critical to the security of our homeland. On a global level, sensors exist to measure quantities such as ozone and pollutants that may be present in international waters.

In recent years, the complexity of advancing technology, the ever-increasing world population, and the emergence of terrorist-related activities have heightened the need for new types of sensors. Depending upon the particular application, the design, fabrication, testing, and eventual use of the sensors requires a wide variety of both technical and nontechnical expertise. As a result, sensors have become an emerging technology that prevails in the world in which we live.

1.2 The Human Body as a Sensor System

The human body serves as one of the best examples of a complex system that contains a wide variety of sensors capable of sensitively and selectively detecting a wide variety of quantities or **measurands**. The most familiar sensors in the human body are those that relate to vision, hearing, smell, taste, and feel.

The human eye can detect both small and large objects that may be stationary or in motion. The eye may also detect very subtle variations in color or shape. A classic example of this is the recognition of a human face. The recognition of a familiar face in a large crowd truly emphasizes the **selectivity** embodied in the vision system. The human vision system does, however, have a finite **dynamic range,** which is determined by the visible part of the electromagnetic spectrum, namely, from wavelengths of 4,100 Å (violet) to 6,600 Å (red). This represents a frequency range from 7.32×10^{14} cycles per second (cps) to 4.55×10^{14} cps. Considering that the electromagnetic spectrum has a range of over 10^{20} cps, visible light represents a very small part of the spectrum.

The human ear is an example of a very **sensitive** sensor with a limited dynamic range. In particular, the ear is capable of detecting sound levels as low as 10^{-12} Watts/m^2 (0 db) and as high as 1 W/m^2 (120 db), which is the threshold of pain. The frequency range for the ear is limited to 20 Hz to 20 kHz. This represents only a small portion of the sound wave spectrum, which covers a range of over 10^{13} Hz. It is well known that prolonged exposure to high-intensity sound levels may produce serious ear damage. In fact, noise pollution may contribute to anxiety, nervousness, and high blood pressure. Some typical sound level intensities are given in Table 1.1. Prolonged

TABLE 1.1

Typical Sound Level Intensities

Source of Sound	db Level[a]
Jet airplane	150
Jackhammer	130
Rock concert	120
Power mower	100
Vacuum cleaner	70
Normal conversation	50
Whisper	30
Rustling of leaves	10
Threshold of hearing	0

[a] db level = 10 log I/Io, where Io = 10^{-12} W/m^2 (threshold of hearing).

exposure to sound levels over 90 db is considered to be dangerous and ear protection is recommended. This clearly points out the need for sound intensity sensors, particularly in an industrial environment.

The human nose represents perhaps one of the most sensitive and selective sensors in the human body. The nose is capable of differentiating subtle differences in odor such as may occur in different types of wine. The nose can also detect various degrees of "freshness" that may occur in fish. The limits of sensitivity of the nose are in the low parts per billion (ppb) to the parts per million (ppm) level in air for a target gas. This sensitivity level is better than that of many of the best commercially available gas sensors.

Taste is not as sensitive or selective as the other familiar body sensors. Although differences in levels of sweetness, sourness, and saltiness can be detected in taste, people often confuse taste with odor. The true measure of the taste sensor may be experienced if one has a cold. In this case, the interference from odor is eliminated.

The so-called feel sensor is usually associated with the hand, but can in fact be located anywhere in the body. This sensor enables one to determine such physical features as an object's size, shape, roughness, and weight. In recent years advances in robotics have resulted in the design and fabrication of artificial feel sensors, such as robotic arms, hands, and fingers, which have found widespread applications in industry.

In addition to the five common body sensors, other abstract sensors exist in the human body that can affect the individual in a variety of different ways. For example, the death of a loved one evokes a feeling of sadness in an individual. Danger often evokes excitement, and overcoming an obstacle may evoke pleasure or satisfaction.

Finally, the human body has a built-in defense or immune system that is often triggered by sensors different from those already discussed. For example, the invasion of the human body by antigens associated with a particular disease or infection triggers the formation of antibodies, which then fight the disease or infection.

The sensors within the human body, particularly the five principal sensors, can be looked upon as real-time control systems. It is essential that these control systems communicate accurately and effectively so the human body can avoid dangers and perform satisfactorily. Ideally, we would like our sensors to respond quickly, sensitively, and selectively to a particular measurand. However, with a degradation in sensor operating accuracy or function caused by factors such as misuse, age, or disease, it is then necessary to make certain modifications to restore the sensor performance. For example, deteriorating eyesight or hearing can be corrected with use of glasses or hearing aids. These corrective measures then simply appear in the feedback loop of the real-time control system associated with the sensor.

1.3 Sensors in an Automobile

In a real-world system, such as an automobile, the ability of an electronic control system to communicate accurately and effectively to the automobile operator is critical to the operation of the automobile. Since the automobile has to be operating in concert with an outside world of significant complexity, the need for reliable, effective, and accurate sensors is extremely high.

In order to appreciate the need, diversity, and complexity of sensors required in an automobile, it is interesting to briefly examine the systems in an automobile in which sensors are used. In Figure 1.1, several potential areas of the automobile in which sensors are used are highlighted. It can be seen that the sensor functions may range from a simple sensing of oil pressure, water temperature, and fuel level to the intensive control of the engine and transmission to optimize economy and performance while minimizing potentially dangerous emission effluents. In order to appreciate the exacting specification and complexity of some automobile sensors, it is interesting to examine the sensors appropriate for the engine and transmission, or what is commonly referred to as the power train. Various power train sensors and their required specifications are presented in Table 1.2. These sensors are critical to the automobile performance and relate to engine

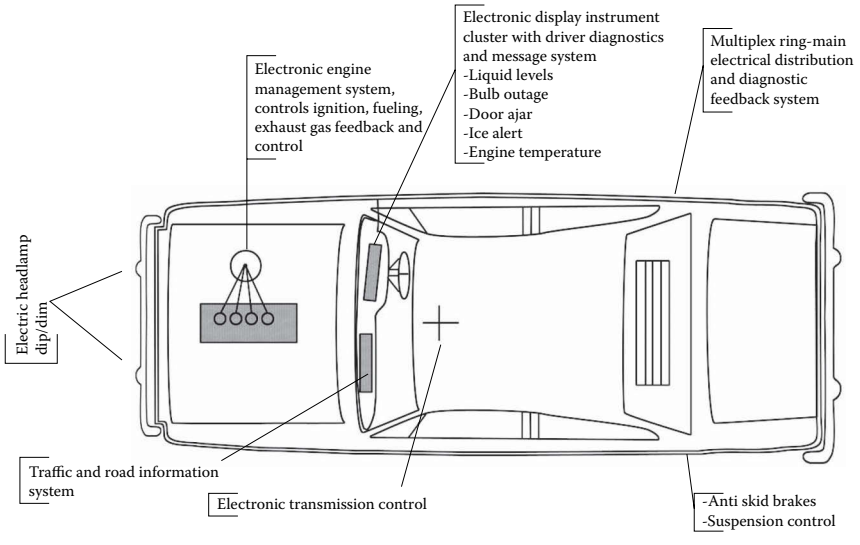


FIGURE 1.1
(Please see color insert following page 146) Areas where sensors can be utilized in the automobile. (From Westbrook, M. H., and Turner, J. D., *Automotive Sensors,* London: Institute of Physics Publishing, 1994, 9. With permission.)

TABLE 1.2

Optimized Specifications for Some Common Automotive Power Train Sensors

Sensor/Type	Proposed Sensing Method	Range	Accuracy	Temperature Operating Range	Response Time
Inlet manifold absolute or differential pressure sensor (petrol engines)	Piezoresistive silicon strain gauged diaphragm or capacitive diaphragm	0–105 kPa	±1% at 25°C	–40 to + 125 °C	1 ms
Inlet and exhaust manifold pressure sensor (diesel engines)	As above	20–200 kPa	±3%	As above	10 ms
Transmission oil pressure sensor	Differential transformer diaphragm or capacitive diaphragm	0–2,000 kPa	±1%	40 to +160°C	10 ms
Inlet manifold air temperature sensor	Metal film or semiconductor film	–40 to 150°C	±2% or ±5%	–40 to 150°C	20 ms
Coolant temperature sensor	Thermistor	40 to +200°C	±2%	As above	10 s
Ambient air temperature	Thermistor	40 to +100°C	As above	–40 to +100°C	As above
Distributor mounted timing/trigger/speed sensor	Hall effect or optical digitizer or eddy current	Zero to maximum engine speed	±1%	–40 to +125°C	N/A
Road speed sensor (speedometer cable fitting)	Optical digitizer or reed switch or Hall effect	As above	±5%	–40 to +125°C	N/A
Inlet manifold air mass flow (bidirectional)	Ultrasonic or corona discharge or ion flow	±200 kg/h	±2%	As above	1 ms
Throttle position sensor	Potentiometer	0–4 Ω from closed to open throttle	±3%	–40 to +125°C	N/A
Gear selector position sensor	Cam-operated switch or potentiometer	8-position selection or 0–5 kΩ	N/A or ±1	–40 to + 1 50 0°C	N/A
Gear selector hydraulic valve position sensor	Optical encoder	As above	±2%	–40 to 100°C	N/A

(*Continued*)

TABLE 1.2

Optimized Specifications for Some Common Automotive Power Train Sensors (Continued)

Sensor/Type	Proposed Sensing Method	Range	Accuracy	Temperature Operating Range	Response Time
Engine knock sensor	Piezoelectric accelerometer	5–10 kHz g range TBE	N/A	40 to 125°C	Depends on resonant frequency
Exhaust gas oxygen sensor for stoichiometric operation	Zirconium dioxide ceramic with platinum surface electrodes or titanium discs in aluminum	Less than one A/F ratio (used as a switch between lean and rich A/F ratios)	Not known	300 to 850°C (tip operating temperature)	
Exhaust gas oxygen sensor for lean burn operation	Zirconium dioxide oxygen pumping device with heater	14:1 to 30:1 A/F ratio	TBE	As above	5.0 ms

Source: Westbrook, M. H., and Turner, J. D. 1994. *Automotive sensors*, 9. London: Institute of Physics Publishing. With permission.

timing, manifold vacuum pressure and mass airflow, exhaust gas oxygen level, transmission control valve position, transmission input and output speed, and throttle and accelerator position. It can be seen that the requirements for parameters such as accuracy and operating temperature range are exacting. In addition to meeting technical specifications, these sensors must also meet space and weight requirements and be of minimal cost and high reliability. To discuss each of the power train sensors in detail requires background in such diverse areas as physics, chemistry, engineering, economics, and politics.

The multidisciplinary nature of sensors can best be illustrated by describing the development of gas sensors relating to the control of the combustion mixtures in car engines. The goal of these sensors is to decrease atmospheric pollution while increasing fuel economy. The initiative in developing the combustion control gas sensors was prompted by a political decision. In the early 1970s the Environmental Protection Agency (EPA) of the United States legislated the Clean Air Act, which required automobile manufacturers to reduce exhaust gases such as carbon monoxide (CO), hydrocarbons (CH_x) and oxides of nitrogen (NO_x) by about 90%. This requirement had to be satisfied by cars manufactured after the 1976 models. Several approaches were tried to satisfy this regulation. The initial approaches, which tried to optimize the fuel supply and the ignition system, were not successful. The second approach involved converting the polluting gases into inert species. This was done using thermal reactors and catalytic converters. The objective was to oxidize CO and CH_x and reduce NO_x.

This was achieved by recycling NO_x through the engine. The economic impact of this approach was negative in that it increased automobile production costs significantly and the fuel economy decreased.

Material problems relating to the difficulties encountered in disposing of reliable and inexpensive refractory materials in thermal reactors caused the industry to look toward catalytic converters. The approach using catalytic converters solves the oxidation of CO and CH_x and reduces NO_x to transform all polluting constituents into harmless byproducts. However, the three-way catalytic converter is effective only if the engine is fed with near-stoichiometric air/fuel (A/F) mixtures. A stoichiometric A/F mixture is one in which the combustion of the gasoline is ideal. Too much air results in a lean mixture, while too much gasoline results in a rich mixture. This A/F mixture requirement thus created an immediate market for an A/F sensor and appropriate feedback control. Since the A/F sensor lifetime was projected to be about 30,000 automobile miles and the U.S. car production was several tens of million each year, a great sensor market was realized.

The A/F sensor currently in use is based upon a zirconia oxygen sensor. This type of sensor has been used in the past for sensing the partial pressure of oxygen in a gas. These sensors are basically galvanic cells and

consist of a closed tube of calcia or yttria stabilized zirconia (CSZ or YSZ). The tube is covered on both sides with electrodes (normally platinum). Since CSZ and YSZ are ionic conductors of oxygen, when the two sides of the tube are exposed to different oxygen pressures a voltage appears between the electrodes. If one side of the tube is exposed to air for which the oxygen partial pressure is known, the partial pressure of the other side can be obtained by measuring the voltage across the cell. The output of the galvanic cell is fed to a signal conditioning block, which then controls the fuel metering system so as to maintain a stoichiometric A/F mixture. A block diagram of the complete system currently being used to meet EPA regulations is shown in Figure 1.2.

The description of the evolution of the A/F sensor clearly illustrates the many disciplines that determine the need for sensors and influence their design and manufacturing. In the case of the A/F sensor, the initial motivation was environmental, namely, a concern for air quality. This concern resulted in the Clean Air Act, which was political. Then science and engineering entered the picture in the approaches used to transform polluting constituents into inert byproducts. Both approaches had to be studied carefully to minimize economic-related factors such as cost and fuel economy. Material problems relating to waste disposal eliminated one approach. Principles of electrical engineering, physics, mechanical engineering, and chemistry were then used to realize the zirconia oxygen sensor and the signal conditioner.

Similar scenarios can be given for the development of other sensors. For example, in automated control of industrial operations ranging from fabrication of integrated circuits to the assembly of automobiles, a wide variety of chemical and physical sensors are used. In medicine, sensors are used in the detection and control of a wide range of potentially life-threatening diseases. Sensors also find use in studies dealing with plants, animals, and the environment.

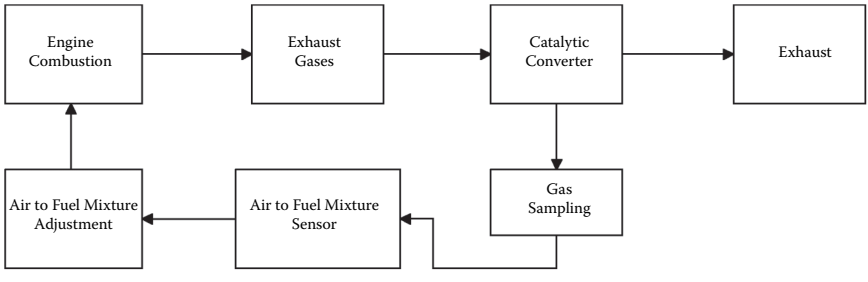


FIGURE 1.2
Complete feedback control system currently in use to meet EPA regulations.

1.4 Classification of Sensors

Due to the multitude of sensors that may exist, at first glance it may appear to be a formidable task to classify sensors in some logical fashion. One possible way to classify sensors may be according to their evolution. Inevitably, the design, fabrication, and utilization of a sensor are driven by a need. For example, in the case of the A/F sensor for the automobile, the need was dictated by EPA via the Clean Air Act. Sensors in various manufacturing processes may be driven by the need to increase production or decrease cost. A particular biosensor may be driven by the need to rapidly detect a life-threatening disease. A possible evolutionary process associated with the development of a sensor is shown in Figure 1.3.

Although it is impossible to list all the potential needs that may result in the development of a particular sensor, the needs may be grouped into the various areas shown in Table 1.3.

The measurand either is identified concurrently with the need or results as a consequence of the need. For example, in the study of ozone depletion, which falls into the area of environment, the measurand is obviously ozone. On the other hand, in the need dictated by the Clean Air Act, namely, to

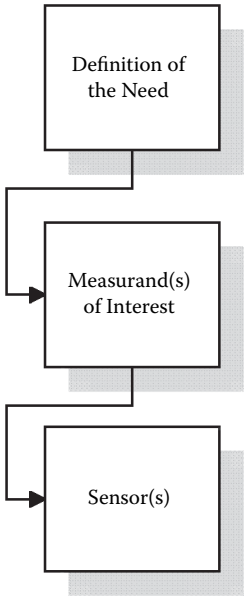


FIGURE 1.3
An evolutionary process associated with sensor development.

TABLE 1.3

Possible Need Areas for Sensors

Agriculture	Manufacturing
Automotive	Marine
Construction	Medical care
Consumer products	Military
Energy	Oceanography
Environment	Security systems
Fishery	Space
Food technology	Transportation
Forestry	Waste management
Health	Other

reduce automobile exhaust gases such as CO, CH_x, and NO_x by 90%, the measurand is not obvious. However, in attempting to satisfy this need, the identification of a sensor to maintain a stoichiometric A/F mixture in engine combustion resulted. Various groups of measurands possible and examples of particular measurands for which a sensor is required are presented in Table 1.4.

Once the measurand has been identified, the next step in the evolutionary scale is to identify the sensor appropriate to the measurand. The realization of the sensor can be thought of as the two-step process shown in Figure 1.4. The measurand must undergo a biological, chemical, or physical interaction in order that the measurand may be converted to an entity suitable for creating a sensor output. Possible types of measurand interactions are presented in Table 1.5.

Once the measurand interaction has taken place, a means of detecting this interaction must be identified. Often the nature of the interaction will define the detection means. In other cases, the means of detection will be entirely different than the nature of the interaction. Examples of various means of detection are given in Table 1.6.

Concurrent with the identification of the measurand interaction and means of detection is the identification of the material(s) used in the sensor. The candidate materials for a sensor cover a very wide range and are presented in Table 1.7.

1.5 Example of a Gas Sensor: The Taguchi Sensor

In order to understand and appreciate the steps one must undertake in the realization of a sensor, it is useful to consider an example. The Taguchi sensor [3] is probably one of the most famous gas sensors currently available.

TABLE 1.4

Examples of Measurands for Which Sensors Are Required

Acoustic
- Intensity
- Spectrum
- Velocity

Biological
- Glucose
- Cholesterol
- Oxygen in blood
- Urea
- Immunosensor
- Biomass
- Other

Chemical
- Gas type and concentration
- Ion type and concentration
- Components dissolved or suspended in solution
- Other

Electric
- Charge, current
- Potential, potential difference
- Electric field (amplitude, phase polarization, spectrum)
- Conductivity
- Permittivity
- Other

Magnetic
- Magnetic field (amplitude, phase polarization, spectrum)
- Magnetic flux
- Permeability
- Other

Mechanical
- Position (linear, angular)
- Velocity
- Acceleration
- Force
- Stress, pressure
- Strain
- Mass, density
- Moment, torque
- Speed of flow, rate of mass transport
- Shape, roughness, orientation
- Stiffness, compliance
- Viscosity
- Crystallinity, structural integrity
- Other

Optical
- Intensity, spectrum
- Velocity
- Other

Radiation
- Type
- Energy
- Intensity
- Other

Thermal
- Temperature
- Flux
- Specific heat
- Thermal conductivity
- Other

Other
- (Specify)

This sensor is shown in Figure 1.5. The measurand, which is a gas, interacts with a doped semiconducting metal oxide film such as SnO_2, changing the resistance or conductivity of the film. The resistance of the film, R_s, then becomes an indicator of the gas concentration. The equivalent circuit for the sensor is shown in Figure 1.6. By measuring the direct current (DC) in the circuit, which is determined by R_s for a fixed R and V, one is able to deduce the gas concentration.

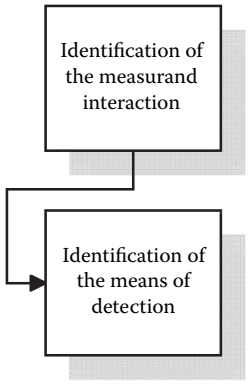


FIGURE 1.4
Two-step process in the realization of a sensor.

TABLE 1.5

Possible Types of Measurand Interactions

Biological	Physical
• Biochemical transformation	• Thermoelectric
• Physical transformation	• Photoelectric
• Effect on test organism	• Photomagnetic
• Spectroscopy	• Magnetoelectric
• Other (specify)	• Elastomagnetic
Chemical	• Thermoelastic
• Chemical transformation	• Elastoelectric
• Physical transformation	• Thermomagnetic
• Electrochemical process	• Thermooptic
• Spectroscopy	• Photoelastic
• Other (specify)	• Other (specify)

TABLE 1.6

Means of Detection Used in Sensors

Biological
Chemical
Electric, magnetic, or electromagnetic wave
Heat, temperature
Mechanical displacement or wave
Radioactivity, radiation
Other (specify)

TABLE 1.7

Sensor Materials
Inorganic
Organic
Conductor
Insulator
Semiconductor
Liquid, gas, or plasma
Biological substance
Other (specify)

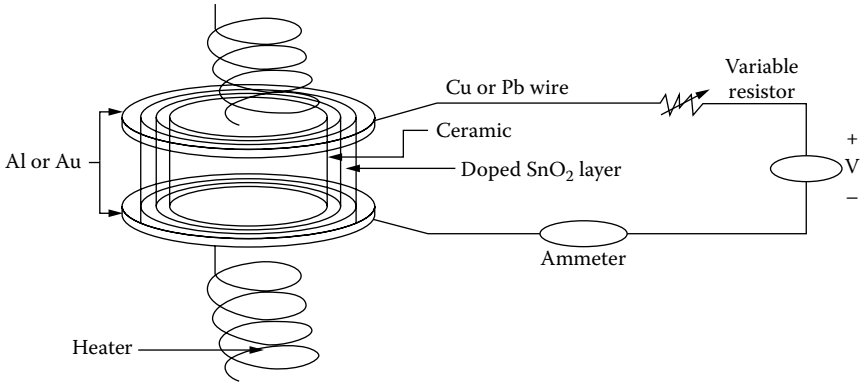


FIGURE 1.5
Taguchi sensor.

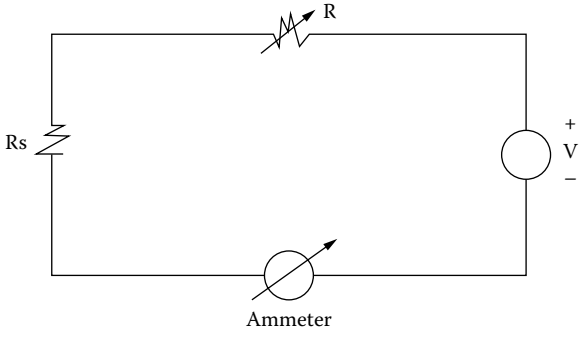


FIGURE 1.6
Equivalent circuit for the Taguchi sensor.

The need for this sensor is driven by the desire to determine the concentration of a particular target gas (measurand). Depending upon the dopant in the SnO_2 film and the temperature of the heater, this type of sensor can detect a wide variety of target gases. The appropriate area of application for this sensor is the environment. The measurand reacts chemically with the doped SnO_2 film; therefore, the measurand interaction is chemical. Since the current is monitored to detect the measurand, the means of detection is electrical. The sensor materials used in this sensor include the metal electrodes (Al or Au), the wire (Cu or Pb), the selective film (doped SnO_2), the substrate on which the film is deposited (ceramic alumina), and the heater (tungsten or nichrome).

1.6 The Sensor as a Passive or Active Element

The sensor itself may be a passive or an active device. In a passive sensor, the actual sensor doesn't have an external signal applied to it. For example, in the galvanic cell used as an A/F sensor, two sides of the zirconia tube are exposed to different oxygen pressures, which creates a voltage. When the oxygen pressure is known on one side, one can deduce from the voltage the oxygen pressure on the other side. Another example of a passive sensor is the thermocouple. In this device, the thermoelectric effect plays the major role. In particular, current flow is initiated by virtue of a temperature difference across the device. In an active sensor the sensor has an external signal applied to it. For example, in the case of the Taguchi sensor the external signal is an electric current that is supplied by a voltage source (see Figure 1.5). The measurand then causes changes in the external signal that are monitored. Other active sensors include acoustic wave (AW) sensors and fiber optic (FO) sensors. In an AW sensor, an acoustic wave is the external signal, while in an FO sensor, it is an allowed electromagnetic mode of propagation.

Inherent to both the passive and the active sensor is a sensing element. This is the entity that the measurand of interest interacts directly with. In the case of the Taguchi sensor the sensing element is the SnO_2 film, while in the A/F sensor the sensing element is either CSZ or YSZ.

A passive or an active sensor may also have several transducers. A transducer is a structure that converts a signal from one form to another. In a passive sensor, transducers are used to measure a signal that is induced in a sensor by the measurand. The metal electrodes at the end of the zirconia tube in the A/F sensor are examples of passive transducers. In this case, ionic conductivity is transformed to voltage. In an active sensor, transducers are used to excite a signal in the sensor. For example, in the case of the Taguchi sensor, the metal electrodes cause an electric current to flow through

the doped SnO_2 film. In AW and FO sensors, transducers are used to excite acoustic modes and electromagnetic modes, respectively.

1.7 The Sensor as Part of a Measurement System

The sensor is by its nature either an active or passive device that serves as the heart of a measurement system. The roles that a sensor may play in a measurement system are best illustrated by the block diagrams shown in Figures 1.7 and 1.8.

When the sensor is used in a stand-alone measuring device as shown in Figure 1.7, the desired output is usually in the form of a digital readout. For example, in the case of the Taguchi sensor, the sensor is as shown in Figure 1.5 and the signal conditioner is used to translate a measured current to a digital readout.

If the sensor is used in optimizing the performance of a particular system as shown in Figure 1.8, it is one element of a larger system. In this case the sensor output is fed to an appropriate signal conditioner, which modifies the signal in a form suitable to interface to the feedback control system. This feedback control system then appropriately modifies one or more subsystems in the system under study so that an "optimized" measurand may

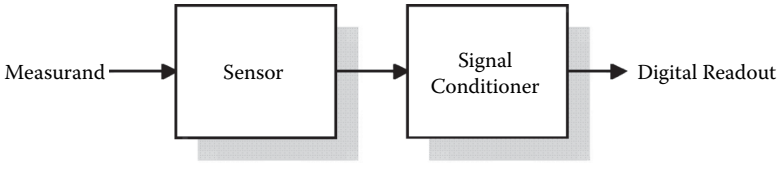


FIGURE 1.7
The sensor used as a stand-alone measuring device.

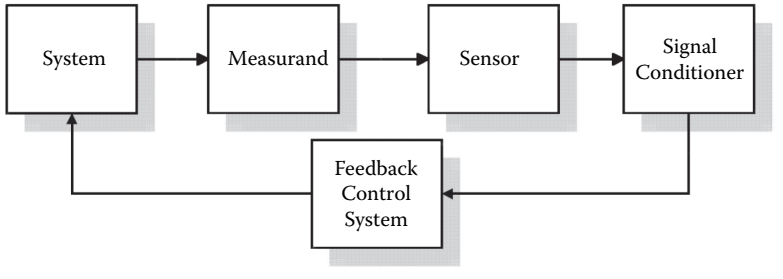


FIGURE 1.8
The sensor used as a measuring device to optimize the performance of a particular system.

be produced. An example of the sensor being used to optimize the performance of a system is the case of the A/F sensor in the automobile (see Figure 1.2).

1.8 Sensor Properties

There are several properties associated with a sensor that are critical to the sensor performance. Some of the more important properties are listed in Table 1.8.

Several of the sensor properties can best be described by referring to a sequence of typical sensor responses as shown in Figure 1.9. The sensor response or signature that is associated with one of the means of detection shown in Table 1.6 is plotted as a function of time. For example, in the case of the Taguchi sensor the means of detection is electric and the output variable, which is measured as the sensor response, is the resistivity (resistance) or conductivity. In the case of the A/F sensor discussed earlier the

TABLE 1.8

Important Sensor Properties Critical to the Sensor Performance

Response time	Resolution
Recovery time	Dynamic range
Reproducibility	Selectivity
Aging	Size and weight
Stability (short term, long term)	Cost
Sensitivity	

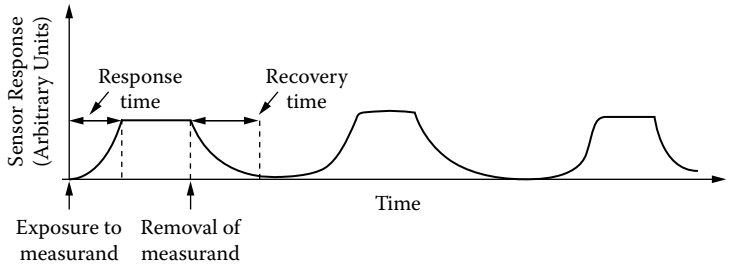


FIGURE 1.9
A sequence of typical sensor signatures.

means of detection is again electrical with the sensor response being the voltage. In the AW sensor the means of detection is mechanical and the sensor response that is measured may be the amplitude, frequency, or velocity of the acoustic wave.

The response and recovery times are shown in Figure 1.9. The **response time** is the time it takes the sensor to reach 90% of its steady-state value after the introduction of the measurand. The **recovery time** is the time that it takes the sensor to be within 10% of the value it had before exposure to the measurand. Typically the response times are much shorter than the recovery times. For example, in a gas sensor a typical response time is of the order of a minute or less, whereas the recovery time can be 30 minutes or longer. The response time is basically the real time it takes for a measurand to interact with the sensing element. For example, in the Taguchi sensor it is the time it takes for the chemical reaction between the target gas (measurand) and the film (sensing element) to occur. The recovery time is longer and more involved than the response time. Once the measurand has been removed from the sensor other steps may have to be taken in order to return the sensor to its equilibrium value. For example, in the Taguchi sensor the film needs to be heated in order to induce a reversible chemical reaction to return the film to its original state.

Reproducibility relates to the ability of the sensor to produce the same signature upon repeated exposure to a particular measurand. A sensor that exhibits excellent reproducibility would have the same response and recovery time along with the same response level. After repeated use of a sensor over a long period of time it is only natural to expect some degradation in the sensor signature. This might be due to the accumulation of trace impurities or imperfections in the sensing element or the nature of the environment that the sensor must operate in. This degradation, the time for which varies from one type of sensor to another, is commonly referred to as **aging**.

The **stability** of a sensor can be deduced by examining the general shape of the sensor signature. In the short term, one would like to minimize or eliminate any drift, ripple, or sudden variations in a single sensor response. In other words, the sensor should respond in a well-behaved fashion, as shown in Figure 1.9. Once the sensor has reached its saturation value, it should maintain that value until the measurand has been removed. In the long-term sensor signature parameters such as the response time, saturation level and recovery time should be reproducible from exposure to exposure of exactly the same concentration of the measurand. Any significant systematic or sudden changes in signature parameters would indicate that the sensor is becoming unstable in the long term.

Sensitivity and **resolution** are sensor properties that are extremely critical when one is working with precise control systems or the sensing of potentially dangerous measurands. Mathematically, the sensitivity, S, of a sensor is

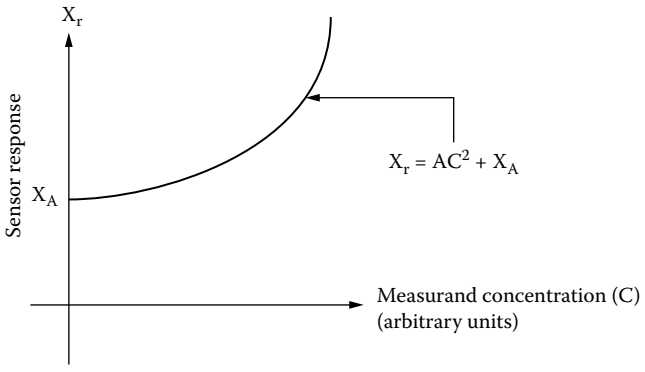


FIGURE 1.10
An example of a nonlinear sensor dose-response as a function of measurand concentration.

defined as the slope of the dose-response curve, which is the sensor response vs. the measurand concentration curve. Ideally, one would like to have the sensitivity be a constant over the dynamic range of the sensor. This would imply a linear dose-response curve.

The lower limit of sensitivity is determined by the minimum amount of measurand capable of producing a measurable response.

Resolution is how accurately one can measure the measurand concentration and relates directly to the error associated with the experimental measurement. It is important to realize that the minimum measurable amount of measurand, or lower limit of sensitivity, may in fact not be equivalent to the sensitivity. This is due to the fact that the measurand response may not vary linearly with measurand concentration. As an example, consider the nonlinear variation of the sensor response (X_r) with the measurand concentration (C) shown in Figure 1.10. It is obvious in this case that the sensitivity increases with the measurand concentration.

Obviously in designing a sensor, one should try to obtain a response that is linear with concentration, or at least operate the sensor in the linear region of the dose-response curve.

The sensor's **dynamic range** is defined as the range of measurand values for which a sensor response occurs. Ideally, this dynamic range should be associated with the measurand concentration range for which the variation of the sensor response, X_r with concentration C is linear. In this case, the signal conditioner design and implementation is straightforward. The dynamic range, however, could extend to regions where the X_r vs. C variation is nonlinear, in which case the corresponding signal conditioner would be more complex. The dynamic range can best be illustrated by referring to the example of an X_r vs. C variation, given in Figure 1.11.

For the example shown in Figure 1.11, specific sensor response regions may be defined as follows:

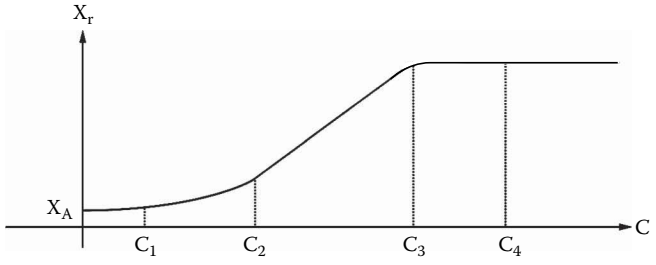


FIGURE 1.11
An example of a complete sensor response versus concentration.

$0 \le C \le C_1$: below the sensor response threshold level

$C_1 \le C \le C_4$: complete sensor dynamic range

$C_2 \le C \le C_3$: ideal sensor dynamic range (linear region of X_r vs. C curve)

$C_1 \le C \le C_2$ and $C_3 \le C \le C_4$: nonlinear region of the sensor dynamic range

$C \ge C_4$: saturation region of sensor

The sensitivity parameter, *S*, may be used when evaluating a single sensor or comparing the performance of several sensors. In the single sensor case often the sensor response will vary depending upon sensor operating conditions. For example, in the case of the Taguchi sensor, the sensor response is strongly temperature dependent. Therefore, it is necessary to operate the sensor at a single temperature that is maintained by adding a heater to the sensor. This single temperature is normally the temperature for which S is maximum.

Another very important sensor property is **selectivity**. Ideally, the sensor should respond only to the measurand of interest and not respond to any other measurand. If a sensor responds to measurands other than the target, then various signal processing techniques need to be employed in the signal conditioner in order to eliminate nontarget measurand signals. In the design of the signal conditioner the sensor signature is an important input parameter. A sensor may respond to a wide variety of measurands, but the signature for each measurand may be uniquely different. The signal conditioner could range from a very simple signal processing scheme, such as threshold detection, to a more complex scheme, such as one employing neural networks.

Three other properties also need to be considered in the realization of a sensor: namely, **size**, **weight**, and **cost**. The location of a sensor, either isolated or within another system, determines the limits on the size and

weight of the sensor. For example, sensors associated with spacecraft and guided missiles have very small size and weight limitations. Finally, the question of economics or cost needs to be addressed. This is often dependent on the scope and need for the sensor. For example, in the case of A/F sensors in the automobile, the number of sensors required is extremely large; therefore producing the cheapest sensor becomes a driving force. In the case of an application relating to monitoring the concentration of hydrogen in the rockets of a spacecraft, the need is essential and cost is of secondary concern.

It is important to realize that the particular sensor application determines which sensor property is the most important. In the detection of very highly toxic gas, sensitivity becomes one of the most important sensor properties. In applications where the sensor must respond numerous times to a measurand such as in "on-line" control systems, properties such as reproducibility and aging become paramount. In applications relating to the implantation of biosensors in humans or animals, weight and size become very important properties.

1.9 Historical Development of Sensors

One of the earliest sensors known to mankind is the compass, which was commonplace even before Roman times. The compass basically consists of a magnetic material, such as iron, which, when constructed as a thin needle and allowed to rotate, will constantly point in a northerly direction. Early travelers and explorers on land and sea relied on the compass for navigation.

In more recent times, the need for sensors relative to safety became recognized. For example, in early mining operations it was necessary to be aware of potentially dangerous gases such as methane and hydrogen sulfide (H_2S). Miners resorted to the use of creatures such as canaries and cats to detect these gases. An indication of the presence of potentially dangerous gases occurred when the canary or the cat died. This is particularly interesting in the case of H_2S, a dangerous gas that is denser than air. Miners, when approaching a section of a mine that might have H_2S gas present, often examined this area with one or more cats. Since the higher concentrations of H_2S were located near the mine floor, the cats were the first to react to this gas. If the cat behaved strangely or died, the miner knew that he had only a short time to vacate the area.

In the early 1900s rapid advancement in technology prompted the development of various types of sensors. This is illustrated in Table 1.9, where the history of the development of some important chemical sensors is outlined.

The disciplines involved in the development of a particular sensor can cover a wide range. For example, in the area of biosensors one can have microbiologists, medical doctors, chemists, physicists, or engineers working

TABLE 1.9

History of the Development of Some Important Chemical Sensors

1923	Catalytic combustion type sensor
1930	Practical use of glass electrode for pH measurement
1938	Humidity sensor using LiC1 film
1952	Galvanic cell type gas sensor
1961	Solid electrolyte type sensor
1961	Ion electrode sensor
1962	Biosensors—basic concept
1962	Oxide semiconductor type gas sensor (Taguchi sensor)
1964	Piezoelectric quartz crystal sensor
1966	Glucose sensor
1970	Ion-sensitive field effect transistor (ISFET) sensor
1970	Optical fiber gas sensor
1975	Pd gate field effect transistor (FET) hydrogen sensor
1976	Practical use of oxygen sensors for automobiles
1977	Enzyme FET
1977	Surface acoustic wave sensor concept
1981	Surface acoustic wave gas sensor
1989	Acoustic plate mode sensor
2004	Lateral field excited acoustic wave sensor

on a single sensor. Often there is little or no collaboration between different investigators, and hence sensor development has not proceeded in a cohesive fashion but rather a haphazard fashion. This is best illustrated by following the development of the Taguchi gas sensor. The idea that a sorbed gas can affect the electrical properties of a material was first pointed out by Bardeen [4] in the late 1940s. However, it was not until 1962 [5] that a deliberate effort was made to use this phenomenon in gas sensing. The first gas sensor, which did not appear until 1968, was the result of a tremendous amount of work by Naoyoshi Taguchi. When Taguchi started his work he had no knowledge of gas absorption or even metal oxide semiconductors. His motivation for the development of the gas sensor was triggered by a liquefied petroleum gas explosion in Japan that killed 10 people. Taguchi worked 7 days a week and holidays for 2,000 days (over 6 years) developing the Taguchi gas sensor. He went from the beginning to the end of the periodic table before he finally settled on SnO_2 as the sensing element. He made an average 20 samples a day resulting in roughly 40,000 samples. After developing the commercial product [3], Taguchi had to fight other, much larger companies in the marketplace. Finally, Taguchi succeeded and formed Figaro Engineering, which is the leading producer of gas sensors in the world today.

Obviously a better, more efficient and focused way to develop a gas sensor would be to form a research team comprised of people with different

backgrounds working in a coordinated fashion. Unfortunately, this has not been the case in the area of gas sensors. Investigators have either performed excruciating searches for a particular sensing element or stumbled upon one by accident. Also, when a sensing element was identified, in many cases, the microscopic and macroscopic behavior of the sensing element was not understood. For example, in the case of the detection of H_2S gas (target measurand) by tungsten trioxide (WO_3) (sensing element) commercial sensors have been available [6] since the mid-1970s. However, only recently has microscopic and macroscopic work [7, 8] been done to understand the operation of these sensors and optimize their performance.

Even though sensor technology has evolved in a rather random fashion, the advances in materia1s and microfabrication technology have had a positive effect on sensors. This is illustrated in Figure 1.12, where an old methane gas sensor, the Davey lamp, is compared to two examples of modern gas sensors, a Taguchi sensor and a sensor developed in Great Britain. Microprocessor technology has also had a positive influence in the interfacing of the sensor with the real world via the signal conditioner and also interfacing the sensor to a control system.

An ideal procedure for the research and development of a prototype sensor is summarized in Figure 1.13.

Once a need has been identified, a research team comprising of researchers in different disciplines should be formed. Working together and using

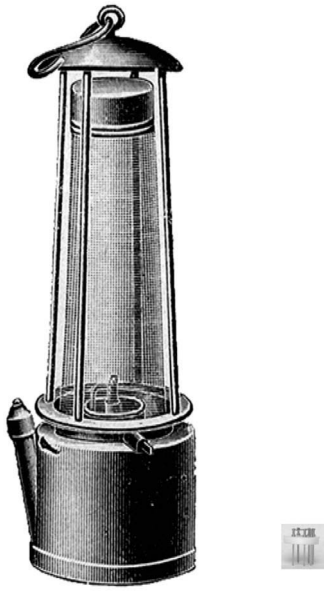


FIGURE 1.12
Old and new gas sensors. The Davey lamp (left) and a modern Taguchi and British gas sensor (right) for the detection of methane.

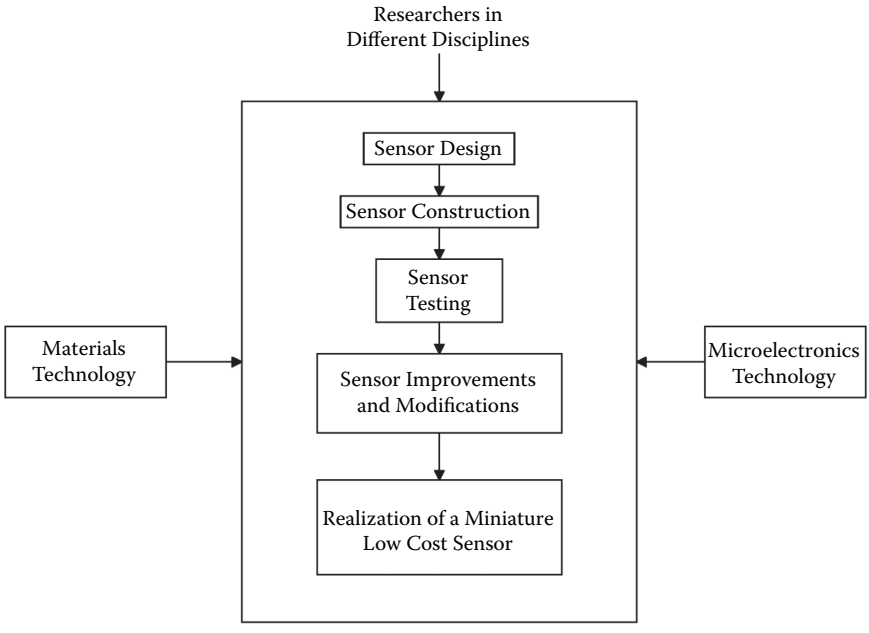


FIGURE 1.13
Ideal process for sensor research and development.

recent developments in materials and microelectronics technology, these researchers potentially should be able to develop a miniature low-cost sensor to satisfy their need.

1.10 Sensor System

The exact definition of a sensor is often interpreted differently depending upon the context in which it is used. Perhaps it is best viewed as the sensor system shown schematically in Figure 1.14. The system consists of three distinct components: (1) sample preparation, (2) sensing element, and (3) sensor electronics.

The sample preparation phase of a sensor system is a very critical component of the system. Often dangerous target measurands such as bacteria, viruses, or pathogens may occur in minute concentrations. For example, in the case of a toxic gas, small volumes of the gas may not be sufficient to result in detection by the sensing element. Rapid and efficient pumping units in which a large volume of gas is collected and sampled are necessary. The sample is then delivered usually with appropriate mass flow controllers and appropriate tubing to a prefiltering component. This component

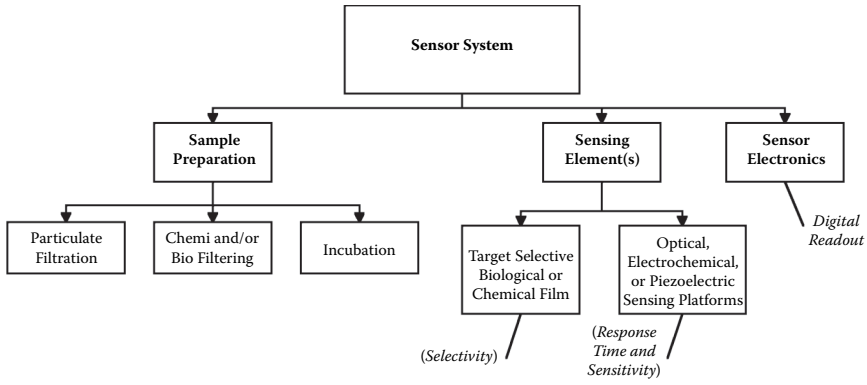


FIGURE 1.14
Sensor system.

provides the first level of selectivity to the sensor system. In particular, the primary function of this component is to filter those gaseous components that may result in false alarms. The filtering may be done using micron to nanoscale structures or molecular sieves such as zeolites to eliminate potential interferents. Once the prefiltering is accomplished, the filtered gaseous sample is then delivered to the sensing element(s). In the case of bacteria, a very low concentration is often present and the bacteria need to be incubated so that a sufficient concentration of bacteria is presented to the sensing element(s).

The sensing elements, which may be a single element or an array of elements, typically consist of a measurand selective film and a sensing platform. The purpose of the selective film is to sorb the target measurand. Once this takes place, mechanical and electrical property changes occur in the film. This may include changes in mass, elasticity, viscosity, conductivity, or dielectric constant. The sensing platform is the entity upon which the measurand selective film is deposited. Typically it consists of a substance on which the film may be attached and an appropriate metallic electrode configuration deposited. Popular sensing elements examined to date include the acoustic wave and metal oxide semiconductor platforms. In the case of the former, an acoustic wave is excited in a piezoelectric material and the frequency of the acoustic wave changes as the electric and mechanical properties of the measurand selective film change. In the case of the latter, the platform is an insulator, which may be single crystal or amorphous. Once the sensing elements have responded to the target measurand(s) and the output has been transduced into an electrical signal (frequency, resistance, or impedance), the signals must then be processed. This component can be eliminated if there is only one sensing element and the sensing film is selective to only the target measurand. However, in reality this seldom occurs and one must employ multiple

sensing elements and intelligently monitor the sensing element output. This is usually done using pattern recognition and neural network techniques. Finally, a digital readout indicates the presence (absence) and concentration level of the target measurand. This component would contain the necessary signal processing and conditioning to convert the analog signal to a digital output that can be easily read by the end user.

Clearly the exact structure of the sensor system is dependent on the types of sensing elements used. Although certain components, such as the sample collection, delivery, and prefiltering, are independent of the sensing element, processing of the sensor element output, and signal conditioning are sensing element dependent.

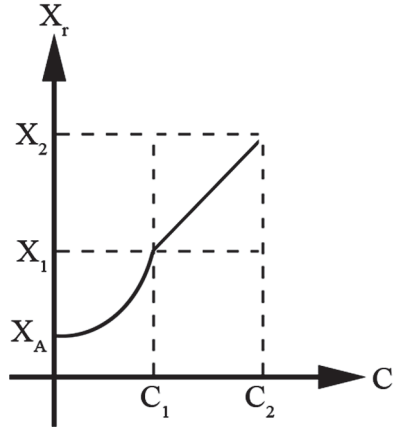
References

1. Westbrook, M. H., and Turner, J. D. 1994. *Automotive sensors*, 7. London: Institute of Physics Publishing.
2. Westbrook, M. H., and Turner, J. D. 1994. *Automotive sensors*, 9. London: Institute of Physics Publishing.
3. Taguchi, N. 1971. U.S. Patent 3,625,756, December 7.
4. Bardeen, J. 1947. *Phys. Rev.* 71:717.
5. Seiyama, T., Kato, A., Fujiishi, K., and Negatani, M. 1962. *Anal. Chem.* 34:1502.
6. Detector Electronics, Minneapolis, MN. Private communication.
7. Xu, Z., Vetelino, J. F., Lec, R., and Parker, D. C. 1990. *J. Vac. Sci. Technol.* A8:3634.
8. Smith, D. J., Vetelino, J. F., Falconer, R. S., and Wittman, E. L. 1993. *Sensors Actuators B* 13–14:264.

Questions

1. Plot a graph to show efficiency vs. A/F mixture ratio in a catalytic convertor and explain why air to fuel mixture is very difficult to design.
2. Referring to published literature, plot a sensor response vs. time for an *E. coli* (biological) sensor defining each segment of the response.
3. Explain the function and operation of an oxygen sensor in an automobile.
4. NO_x, detection where x is either 1 or 2, can be used in at least two distinctly different applications. What are those applications and the concentration levels of interest?

5. The response of a sensor X_r vs. concentration, C, is parabolic for $C \leq C_1$, linear for $C1 \leq C \leq C2$ and constant for $C \geq C2$, as shown below. Assume a measurement of accuracy of x%.



   a. Obtain the minimum concentration observable.
   b. Obtain an expression for the resolution in the following ranges: $C \leq C1$ and $C1 \leq C \leq C2$.
   c. What is the operating range?
   d. What is the saturation region?
   e. What is the ideal dynamic range?

2

Electrochemical Sensors

2.1 Background

Electrochemical sensors form a group of very important sensors whose response is a result of an interaction between electricity and chemistry—hence the nomenclature *electrochemical sensors*. Basically, the electrochemical sensors can be divided into three general types: **conductimetric**, **potentiometric**, and **amperometric**. The conductimetric sensors are those whose response relates to the measurement of electrical resistance, impedance, or admittance. The potentiometric sensor response relates to the measurement of voltage, while the amperometric sensor response relates to the measurement of electrical current. These sensors basically respond to a chemical environment. This chemical environment may include a measurand, which may be in a liquid, gas, or solid phase. Typical measurands may include chemical species such as gases (O_2, CO, CO_2, H_2S, NO, NO_2, etc.), protons (pH), humidity, or biological quantities such as glucose, cholesterol, enzymes, antibodies, proteins, etc. It is interesting to note that since there are a large number of biological applications for these sensors in humans, animals, and plants, these sensors are often referred to as bioelectrochemical sensors by researchers working in the fields of biology and medicine.

Work in the area of electrochemical sensors has been ongoing for almost 100 years. One of the earliest works relates to the use of oxygen-ion-conducting solid electrolytes as fuel and concentration cells. However, the advent of microfabrication technology along with rapid improvements in materials technology in the last 40 years has resulted in the movement of sensors from the macro to the micro domain. Amperometric sensors such as the air/fuel (A/F) sensor in the automobile, conductimetric sensors such as the Taguchi sensor, and potentiometric sensors such as the chemical field effect transistor (CHEMFET) are examples of microsensors. Through the cooperative efforts of engineers and scientists, further breakthroughs in reducing not only the size but also the cost of electrochemical sensors may be expected in the future.

2.2 Conductimetric Sensors

Conductimetric sensors basically consist of an electrically insulating element, such as a sensing platform, upon which one deposits a sensing semiconducting film or metal and metal electrodes. This type of sensor is an active sensor and requires an excitation source such as a direct current (DC) or alternating current (AC) voltage to cause current to flow in the sensing element. The measurand interacts with the sensing film and causes changes in the sensing element resistance (DC case) or impedance (AC case). Several configurations have been used to model the conductiometric sensor. Basically these configurations are of the type shown in Figure 2.1.

In the more general case of an AC excitation the sensor can be electrically modeled as an impedance, $z(\omega)$, whose value can be determined by the following relation:

$$z(\omega) = \frac{v(\omega)}{I(\omega)} \tag{2.1}$$

where $v(\omega)$ = voltage between the electrodes, and $I(\omega)$ = current through the sensor.

When the sensor is exposed to a measurand (normally a gas), the sensing element's impedance changes, and this change is directly related to the concentration of the measurand. A general equivalent circuit for the conductimetric sensor is shown in Figure 2.2.

One can clearly see that the equivalent circuit for the conductimetric sensors is a rather complex arrangement of capacitors and resistors. The significant interaction, of course, should occur between the sensing element and the measurand, and hence the circuit elements, C_s and R_s, should be the focus of attention. However, one must be aware of the effect of the other elements. If the effects are significant, means to account for them must be explored. In

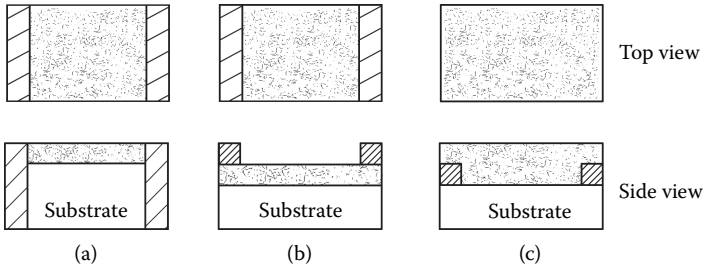


FIGURE 2.1
Different configurations used to model the conductimetric sensor: cross-hatched region, electrodes; dotted region, sensing element.

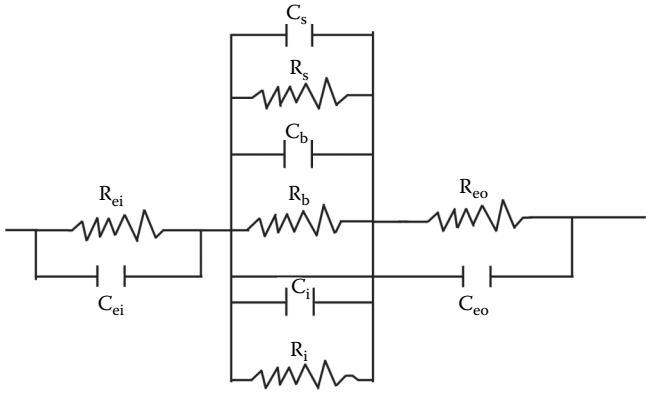


FIGURE 2.2
Equivalent circuit of the conductimetric sensor: R_{ei}, C_{ei}, R_{eo}, and C_{eo}, circuit elements of the input and output electrodes; C_s and R_s, circuit elements for the sensing element; C_i and R_i, circuit elements for the sensing element/substrate interface; C_b and R_b, circuit elements for the substrate.

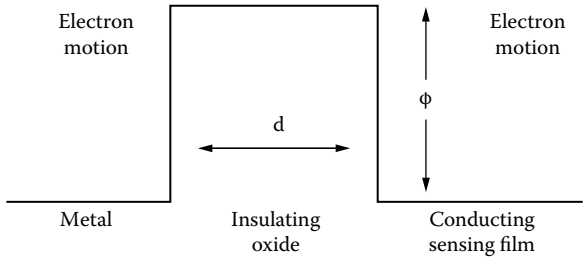


FIGURE 2.3
The potential diagram associated with a metal-insulating oxide-conducting sensing film junction.

this case, hermetically sealing the electrode or substrate from the measurand should reduce the effect of some of the circuit elements.

The electrodes used in conductimetric sensors are metals. Hence, the type of contacts that one must consider depends upon the nature of the sensing film. For conducting sensing films the two possible types of contacts are metal-conductor and metal-insulating oxide-conductor. In the case of the former, the contact is ohmic, and hence the contact resistance is zero. In the second case, an oxide layer impedes the flow of current and presents a potential barrier, as shown in Figure 2.3. This may result in a tunneling current at the junction.

The tunneling current is given by the Fowler Nordheim equation [1], which is a nonlinear current vs. voltage relationship in which the corresponding contact resistance depends upon the barrier height and width

in a complicated fashion. If aluminum is used as the contact metal, a nonlinear contact resistance results. This is due to the fact that aluminum is prone to oxidize. If, however, one uses noble metals such as gold, platinum, or palladium, the contact can be prepared oxide-free. However, in this case, one must realize that noble metals are far more costly than aluminum.

In the case of semiconducting sensing films, again two types of contacts arise: a metal-semiconductor and metal-insulating oxide-semiconductor. The metal-semiconductor junction is the familiar Schottky barrier, where the current may be a nonlinear function of the voltage. The contact resistance associated with this junction is given by the following relation [1]:

$$R = \frac{k}{qA_rT} e^{q\varphi/kT} \tag{2.2}$$

where k = Boltzman's constant, φ = Potential barrier height, q = electron charge, and T = temperature.

A_r is the Richardson constant for thermionic emission of electrons given by

$$A_r = \frac{4\pi mqk^2}{h^2} \tag{2.3}$$

where h = Planck's constant, and m = electron mass.

In the case where the contact metal is not a noble metal, an oxide layer appears between the metal and semiconductor. The current-voltage relationship is then given by the following equation [1]:

$$J = A_R T^2 \exp\left[\frac{-q\left[\phi - \left(qv/4\pi\varepsilon d\right)^{1/2}\right]}{kT}\right] \tag{2.4}$$

where v = applied voltage, d = oxide thickness, and ε = oxide dielectric constant.

In some cases the current transport is not due to electrons or holes, but rather ions. In this case an electrochemical reaction will take place at the contact and the mass transport of the electroactive ions must be considered. The contact resistance then becomes a strong nonlinear function of the measurand concentration.

The previous discussion has been restricted to the DC case. However, in the AC case one must consider the capacitors in the equivalent circuit. Since capacitance is related to charge distribution and dielectric constant, any accumulation of charge at the contact will give rise to capacitive effects that must be considered.

In the DC case, the value of the resistance of the sensing element depends upon whether the layer is a semiconductor or a conductor. If the layer is a semiconductor, the contact resistance is lower than the selective layer resistance and can be either used to modulate the overall sensor resistance or neglected. The resistance of the layer can be expressed in terms of the layer conductivity, σ, or resistivity, $\Re$, as follows:

$$R = \frac{\ell}{\sigma A} = \frac{\ell \Re}{A} \tag{2.5}$$

where ℓ = length of the layer between the electrodes, and A = layer cross-sectional area.

The layer conductivity depends upon the concentration (n and p) and mobility (μ_n and μ_p) of the electrons and holes, respectively, in the semiconductor layer, as follows:

$$\mu = q(n\mu_n + p\mu_p) \tag{2.6}$$

As the target gas interacts with the layer it will modify the carrier concentrations and mobilities accordingly. The conductivity will either increase or decrease depending on the nature of the reaction between the measurand and layer. In the case where the sensing layer is a conductor, the contact resistance is larger than the layer resistance before exposure to the measurand. However, after exposure to the measurand, the layer resistance increases quickly and soon exceeds the contact resistance. Obviously for this case, the contact resistance cannot be neglected and needs to be properly accounted for.

At first glance, it may appear that the conductimetric sensors are reasonably simple and their support electronics is not complicated. However, most conductimetric sensors are treated only macroscopically. Therefore, the relationship between the output signal and measurand concentration is given without a complete understanding of the microscopic behavior. The microscopic understanding is determined to a large extent by the understanding of the measurand-layer interaction. In the case of the metal oxide semiconductor layer, the understanding of the microscopic interactions is fair at best, while for biological layers very little is known about the measurand-layer interaction.

Depending upon the type of sensing film and whether the impedance measurement is a DC or AC type, conductimetric sensors can be divided into four groups.

The first group is called the semiconductor metal oxide sensors. The sensing film is a metal oxide semiconductor and the measurement is DC. The most famous of the semiconductor metal oxide sensors is the Taguchi sensor.

The second group of conductimetric sensors is called the chemiresistors. The sensing film in this case may be inorganic conductor or organic semiconductor, and the measurement is DC.

The third group is called a conductor sensor. The sensing film in this case is an electrical conductor.

The fourth and final conductimetric type sensor is called a dielectrometer or capacitor sensor. It differs from the other conductimetric sensors in that it uses an AC signal and, in many cases, the sensing film is an electrical insulator. In this case, the change in sensing layer capacitance is monitored upon exposure to the measurand.

2.3 Semiconducting Metal Oxide Sensors

2.3.1 Background

Since the early work of Bardeen [2] in the late 1940s, it has been well known that the adsorption of a foreign species such as gas on a semiconductor can create surface states. These states can modify [3–10] the carrier concentrations in a semiconductor, causing a change in the semiconductor's resistivity or conductivity. This so-called active surface created problems with early work on semiconductors and several investigations were made to pacify the surface.

The first deliberate effort to use the active surface on the semiconductor as a gas sensor appeared in 1962. Seiyama et al. [11] monitored the change in the conductivity of ZnO films when they were exposed to gases such as CO_2, benzene, ethyl alcohol, and oxygen. For the first three gases the conductivity increased, since these gases behaved as electron donors, while in the case of oxygen the conductivity decreased, which leads one to conclude that oxygen behaves as an electron acceptor. In a follow-up paper, Seiyama and Kagawa [12] further examined the characteristics of this detector. They also performed measurements on a host of other semiconducting films, such as Cr_2O_3, TiO_2, CuO, and SnO_2.

Although there has been some work [13] on organic semiconductors and elemental or semimetal semiconductors, the most popular semiconductors are the metal oxide variety. Among the metal oxide semiconductors, the one that has received the most attention has been SnO_2. This resulted in the development of the so-called Taguchi gas sensor (TGS) [14].

Since the sensing element of the TGS is a SnO_2 film, a large amount of work has been done on SnO_2. The physical and electrical properties of SnO_2 films have been examined in detail by Jarzebski and Martin [15]. Other studies have been concerned with the effect various gases have on the SnO_2 film. In particular, the gases most frequently studied [16–36] included CO, H_2O vapor, propane, H_2, NO_x, O_2, H_2S, ethanol, C_3H_8, CO_2, methane, SO_2, and benzene. Several of these studies were directed toward obtaining a physically realistic and practically useful gas sensor. Properties studied included film durability, sensor size, weight and portability, sensitivity, recovery time, response time, stability, selectivity, and reproducibility. In particular, the sensor recovery times for pure undoped SnO_2 films were observed [30] to be in excess of 10 minutes, while the addition of a gold dopant caused the recovery time to decrease to less than 3 minutes. Similarly, sensitization of the surface by adding dopants such as Nb, V, Ti, Mo, Au, Pt, ThO_2, Pb, or Sb_2O_3 [17,18,24,27,28,30] decreased the sensor response time significantly. Also, the selectivity of the sensor could be improved by heating the film in environments such as SO_2 [32]. By varying the operating temperature of the sensor, it was shown [32] that the sensor response to certain gases was maximum at a particular temperature. This information could be used in the design of a selective sensor. Finally, it has been pointed out by Vetelino and coworkers [37,38] that heat treating the film was found to stabilize the film, decrease the response time, and increase the sensitivity.

Although there have been numerous papers reporting the characteristics of SnO_2 and other semiconductor metal oxide-based sensors, much less work has been devoted to a complete understanding or modeling of what is actually taking place in the film. This is a little surprising when one realizes that as early as the 1950s [3] it was observed that the resistance of semiconductors is sensitive to the sorption of gases. The fact that semiconductor sensors exhibit hysteresis and are sensitive to large numbers of gases has contributed in part to the lack of a complete understanding of the physical or chemical processes taking place in the film. The form of the gas interactions with semiconductor surfaces has been grouped into four types [39]: (1) reduction/oxidation (redox) of the semiconductor, (2) ion exchange, (3) adsorption, and (4) surface states. The interactions present when SnO_2 is the film of interest cover all four types. For example, Pink et al. [26] suggested a redox type reaction is involved in the sensitivity of their SnO_2 film to hydrocarbons. The H_2S sensor based on SnO_2 has been suggested [39] to be an ion exchange case. The simple adsorption case seems to be the most popular for semiconducting oxide sensors. If the adsorbing gas is a reducing agent, it injects electrons into the semiconductor, and if it is an oxidizing agent, it extracts electrons from the semiconductor. An example of the latter is the NO_x-based SnO_2 sensor [19]. Finally, the creation of a surface state case has been suggested [16,23,26] to be the primary mechanism in some of the Taguchi sensors.

The above work seems to suggest that the appropriate microscopic theoretical description is dependent upon what type of gas is interacting with

SnO_2. Clearly, there is a need for a good understanding of all the physical and chemical processes involved in SnO_2-based sensors.

In addition to SnO_2 there has been work on other semiconducting metal oxide films for sensor applications. These films include ZnO [11,12,23,31,32,40], TiO_2 [13,35,39], WO_3 [37,38], Co_3O_4 [13], CeO_2 [32], Mn_2O_3 [32], Cr_2O_3 [12,32], CuO [12], and Ag_2O [39]. Clearly next to SnO_2 the most work has been done on ZnO. In fact, the early work on metal oxide semiconductor films for sensor applications was done [11] on ZnO. Interest, however, turned away from ZnO after the introduction of the TGS sensor. Recently, however, there has been some interesting work [40,41] on ZnO in the polycrystalline, single crystal, and compressed disc form. The most recent work [41] indicates that ZnO single crystal sensors have good long-term stability and are easy to fabricate. Recently there has also been some work done on WO_3 [37,38] at the University of Maine in order to try to optimize these films for gas sensing applications. The work on the other films is limited; however, there is reason to believe that much work is unpublished due to proprietary reasons. This is evidenced by the commercial WO_3-based H_2S detector available from Detector Electronics Industries in Minneapolis [42].

Although the macroscopic theory of operation of the semiconducting metal oxide sensor is straightforward, the microscopic theory is involved and in many cases not understood. In order to formulate a complete microscopic description, one must be able to completely describe the role gas-film-ambient chemistry and film structure play in determining film electrical properties. Some important questions that need to be addressed include:

- Why are certain films selectively sorbent to certain gases?
- How can the sensitivity of a film be maximized?
- What steps are necessary in film preparation to ensure both short- and long-term stability in the sensor response?
- How can the sensor response and recovery times be minimized?
- What role do ambient air parameters such as humidity and temperature play?
- How can the reproducibility of sensor properties from sensor to sensor be maintained?

Before attempting to answer these questions, the metal oxide semiconductor electrical properties will be discussed.

2.3.2 Electrical Properties of the Metal Oxide Semiconductor

2.3.2.1 Intrinsic and Extrinsic Metal Oxide Semiconductor

The intrinsic metal oxide semiconductor is defined as a single crystal metal oxide that has no imperfections, vacancies, interstitial atoms, or dislocations.

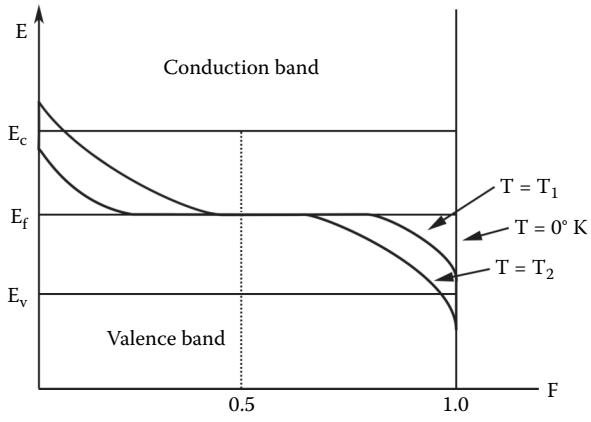


FIGURE 2.4
The energy band structure for an intrinsic semiconductor with the Fermi-Dirac probability function superimposed upon it. $0^{\circ}\,K < T_1 < T_2$.

The energy band diagram for an intrinsic semiconductor with the Fermi-Dirac probability functions superimposed upon it is shown in Figure 2.4. In an actual metal oxide semiconductor imperfections are present, causing it to become extrinsic. The imperfections can be either donors or acceptors. In the case of the former they donate electrons to the semiconductor, causing it to become *n*-type, while in the latter case holes are created, causing it to become *p*-type. The energy band structure for the *n*- and *p*-type metal oxide semiconductors and the appropriate location of the Fermi level are shown in Figures 2.5 and 2.6. In contrast to silicon, which has an energy gap of about 1 eV, metal oxide semiconductors have an energy gap of about 3 eV. The excess carriers have caused the Fermi level to be either slightly below the conduction band (*n*-type) or slightly above the valence band (*p*-type). The appropriate carrier concentrations can be obtained by invoking charge neutrality in the semiconductor and applying the mass-action law.

Metal oxide semiconductors typically have an excess amount of metal, which creates oxygen vacancies. The extra electrons associated with the metal, which are normally bound to oxygen, are loosely bound and readily enter the conduction band. The oxygen vacancies are therefore said to act as donors. Metal ions located in interstitial locations or acting as substitutional impurities for the host metal ion can also donate electrons to the conduction band. However, the predominant donor is the oxygen vacancies. It is therefore obvious that metal oxide semiconductors are *n*-type. To achieve *p*-type metal oxide one would have an excess of oxygen present. The arguments as to why excess oxygen would produce holes would parallel the discussion on why oxygen deficiencies produce electrons. However, to date there has been no consistent method to make *p*-type metal oxides.

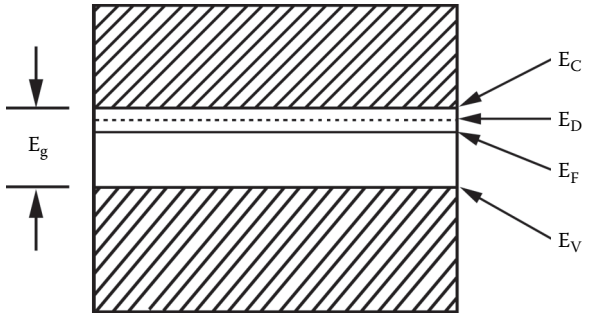


FIGURE 2.5
The energy band structure of a typical *n*-type metal oxide semiconductor with Eg ≈ 3 eV. The donor level, E_D, is located .01 eV below E_C.

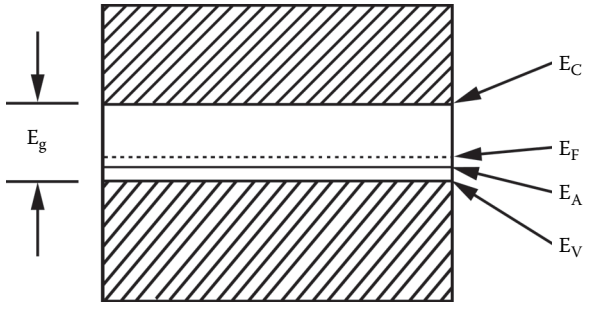


FIGURE 2.6
The energy band structure of a typical *p*-type metal oxide semiconductor with Eg ≈ 3 eV. The acceptor level, E_A, is located .01 eV above E_V.

The conductivity of the metal oxide is given by Equation 2.6, where the mobility of the carriers is defined as the drift velocity per unit electric field. As the carriers begin to move their motion is dominated by scattering with lattice vibrations or phonons. Since lattice vibrations become more vigorous with increasing temperature, the mobility decreases with increasing temperature, *T*, according to the following relation:

$$\mu \propto T^{-3/2} \tag{2.7}$$

The carrier concentrations in metal oxides typically increase exponentially with temperature. This increase is much greater than the decrease in mobility with temperature. Hence, overall the conductivity increases with increasing temperature. Examples of typical variations of the conductivity, electron carrier concentrations, and mobility [43] are given in Figures 2.7 to 2.9 for radio frequency (RF) sputtered tungsten trioxide (WO_3) films (500 Å thick) on a $LiNbO_3$ substrate.

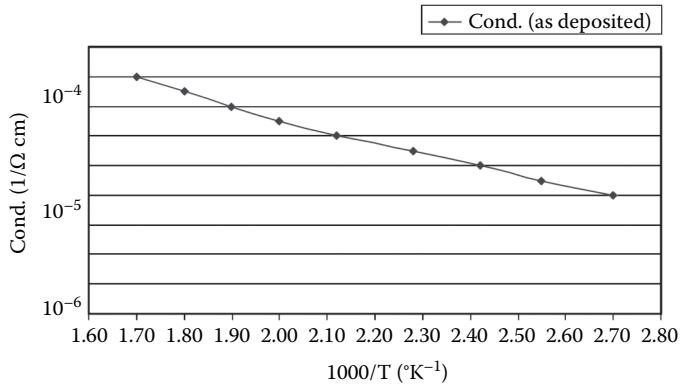


FIGURE 2.7
(Please see color insert following page 146) Typical conductivity results vs. 1,000/temperature (K^{-1}) for (as deposited)[2] 500 Å thick WO_3 film on $LiNbO_3$ in ambient air (humidity 50%). (Smith, D. J. 1992. Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications. MS thesis, Electrical Engineering, University of Maine.)

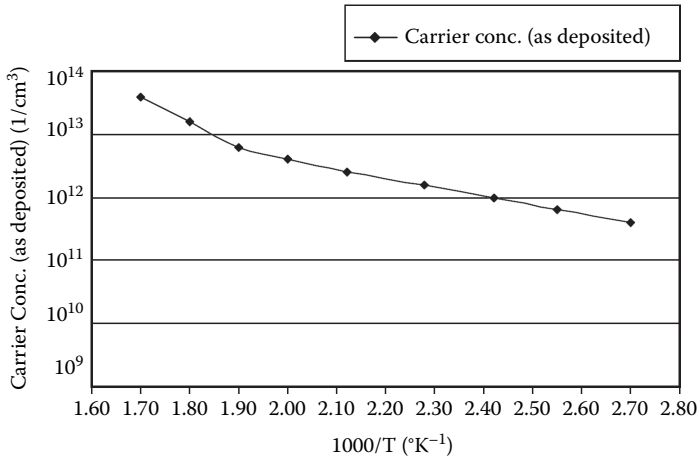


FIGURE 2.8
(Please see color insert following page 146) Typical electron carrier concentration results vs. 1,000/temperature (K^{-1}) for as deposited 500 Å thick WO_3 film on $LiNbO_3$ in ambient air (humidity 50%). (Smith, D. J. 1992. Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications. MS thesis, Electrical Engineering, University of Maine.)

2.3.2.2 Metal Oxide Surface

Since metal oxide semiconductors are used as films, it is important to characterize the film surface due to the fact that the target measurand interacts with the film at or near the film surface. The surface, even though it has no contaminants or impurities, is still imperfect. In particular, the geometry of the film is severely interrupted at the surface. This can then lead to

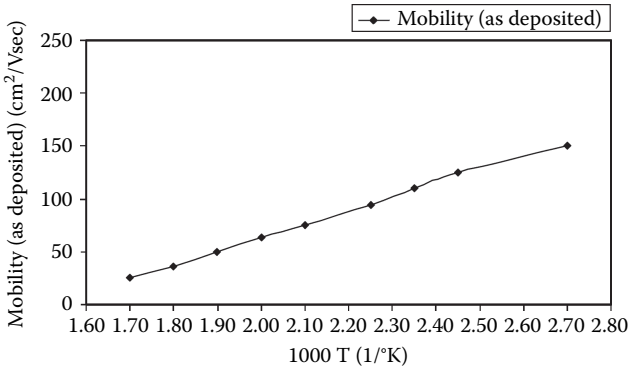


FIGURE 2.9
(Please see color insert following page 146) Typical carrier mobility results vs. 1.000/temperature (K^{-1}) for as deposited 500 Å thick WO_3 film on $LiNbO_3$ in ambient air (humidity 50%). (Smith, D. J. 1992. Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications. MS thesis, Electrical Engineering, University of Maine.)

localized allowed energy levels, which may occur in the energy gap. Since metal oxides are more ionic than covalent in nature, the metal ion is referred to as the cation and oxygen is the anion. The surface metal ions or cations are positively charged, and hence tend to capture extra electrons and therefore act as acceptors. On the other hand, the surface oxygen ions and anions are negatively charged and therefore tend to give up their electrons and act as donors. The band structure for the metal oxide semiconductor surface prior to any charge motion is shown in Figure 2.10. In this case there is no net surface charge, and it is called the flat-band case. The surface states are shown as narrow bands with the donor states occupied and the acceptor states unoccupied. It is interesting to point out that the donor and acceptor states form bands. This is due to the fact that the oxide surface may contain different crystal faces and surface discontinuities or steps.

Once the charges begin to move, the corresponding band structure changes. Since the allowed energy level for the acceptor states is less than the conduction band electron energy, the electrons from the donor sites in the *n*-type metal oxide move down into the acceptor states, creating partially occupied acceptor states. A so-called double layer is then formed, which has positively charged donor ions in the semiconductor and negatively charged surface acceptor states as a sheet of charges on the surface. This double layer is shown in Figure 2.11.

The region $0 \le x \le x_0$ is called the space charge layer or depletion layer since it is depleted of mobile charges. The charge density in this region is given as

$$\rho = qN_D; \quad 0 \le x \le x_0 \tag{2.8}$$

and

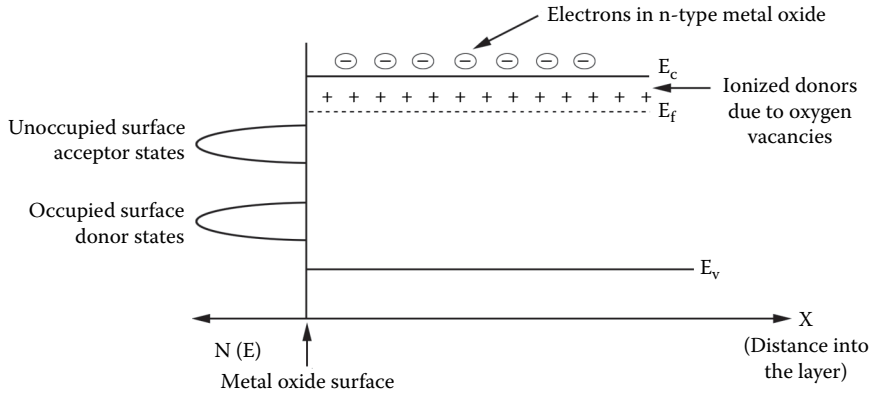


FIGURE 2.10
Band structure for *n*-type metal oxide. The acceptor surface states due to the surface metal ions and the donor surface states due to the surface oxygen ions are shown.

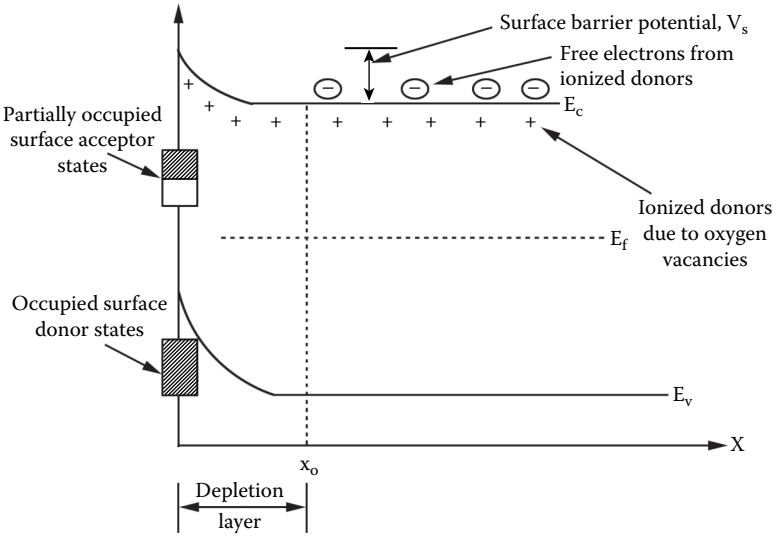


FIGURE 2.11
The double layer near a metal oxide surface. The depletion region contains fixed ionized donors (positive charge) and fixed occupied acceptors. This causes a surface barrier potential, V_s, to form.

$$\rho = -qN_D; \quad x = 0 \tag{2.9}$$

where N_D = donor density due to oxygen vacancies in the metal oxide.

The separation of charge in the depletion region causes an electric field and hence a surface barrier potential to appear. The number of electrons extracted from the depletion region is the number of electrons moved to the

surface. Notice also that the Fermi level is now located between the partially occupied acceptor states and the occupied donor states.

In order to determine the variation of the potential, ϕ, across the depletion region, one must solve the one-dimensional Poisson equation:

$$\frac{d^2\varphi}{dx^2} = -\frac{qN_D}{\varepsilon} \tag{2.10}$$

where ε = the metal oxide dielectric constant.

It is assumed that the doping is homogeneous; therefore, N_D does not vary with x.

It is convenient to define a parameter, V, so as to be able to relate more directly to the band structure, as follows:

$$V(x) = \phi_b - \phi(x) \tag{2.11}$$

where ϕ_b = potential in the metal oxide bulk.

Solving Equation 2.11 for $\phi(x)$ and substituting into Equation 2.10 and integrating, one obtains

$$-\frac{dV}{dx}\bigg|_x^{x_o} = -\int_x^{x_o} \frac{qN_D}{\varepsilon} dx \tag{2.12}$$

Noting that at $x = x_0$, the space charge region contains a net charge of zero or all the surface charge is compensated, therefore dV/dX = 0 at $x = x_0$. Equation 2.12 now becomes

$$\frac{dV}{dx} = \frac{qN_D}{\varepsilon}(x - x_0) \tag{2.13}$$

The integration of Equation 2.13 from x to x_0 yields

$$V = \frac{qN_D}{2\varepsilon}[x - x_0]^2 \tag{2.14}$$

The surface barrier potential, V_s, then becomes

$$V_S = V\big|_{x=0} = \frac{qN_D}{2\varepsilon} x_0^2 \tag{2.15}$$

The variation of the potential across the depletion region is shown in Figure 2.12. In order for the electrons to reach the surface states, they must overcome the potential barrier shown in Figure 2.12.

For a typical value of N_D of about $10^{23}/m^3$, $\varepsilon = 10\varepsilon_o$ (ε_o = permittivity of a vacuum) and $V_s = 1$ V, x_0 is about 100 nm.

2.3.2.3 Surface or Space Charge Capacitance

Since a separation of charge occurs near the surface of the metal oxide, a space charge capacitance exists. The resulting capacitor can be approximated as a parallel plate capacitor and represented as

$$C = \frac{\varepsilon A}{x_0} \tag{2.16}$$

where A = parallel plate area.

Solving Equation 2.15 for x_o and substituting into Equation 2.16, one obtains

$$1/C^2 = \frac{2V_s}{qN_D\varepsilon A^2} \tag{2.17}$$

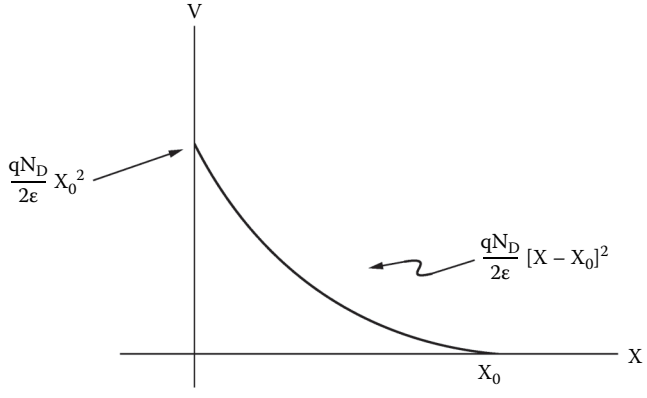


FIGURE 2.12
Variation of the potential, $v(x)$, across the depletion region.

If a base voltage, V, is applied across the depletion region, the surface or space charge capacitance becomes a function of voltage. Two important parameters can then be determined—the charge density in the region, N_D, and the surface barrier potential, V_s—as follows:

$$1/C^2 = \frac{2(V_s - V)}{qN_D\varepsilon A^2} \tag{2.18}$$

A plot of $1/C^2$ vs. the bias voltage appears in Figure 2.13.

The charge density, N_D, can be deduced from the slope of the $1/C^2$ curve, while the surface barrier potential is obtained when $1/C^2$ is zero.

Experimentally the measurement of the $1/C^2$ vs. bias voltage curve is rather straightforward. A possible measurement system is shown in Figure 2.14. The

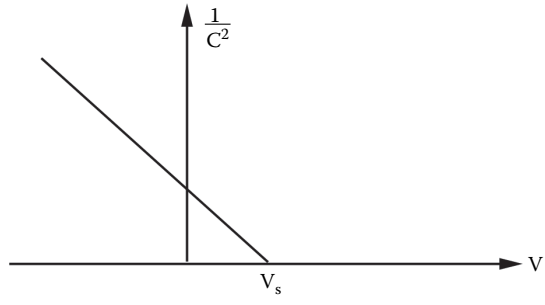


FIGURE 2.13
The variation of $1/C^2$ as a function of the bias voltage, V.

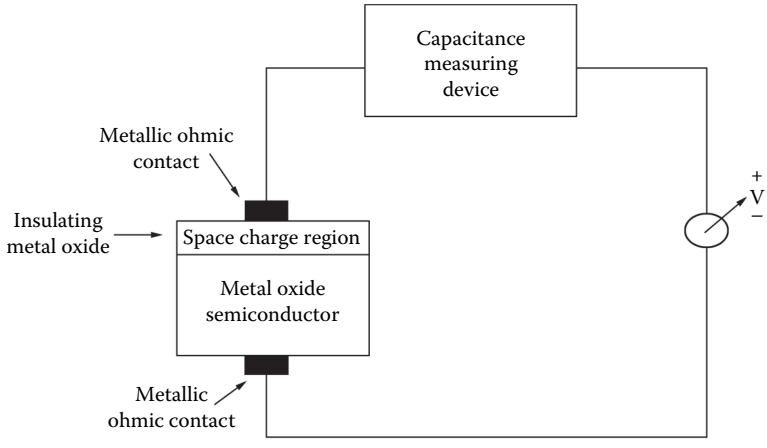


FIGURE 2.14
Surface capacitance measurement system.

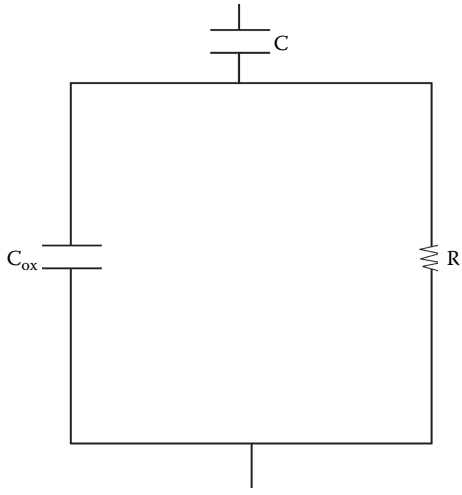


FIGURE 2.15
Equivalent circuit for the space charge region and the metal oxide semiconductor.

equivalent circuit is shown in Figure 2.15. *C* is the capacitance of the space charge region between the ohmic metallic contact and metal oxide semiconductor. C_{ox} is the metal oxide capacitance, and *R* is the resistance of the metal oxide semiconductor. By varying frequency and analyzing the equivalent circuit, *C* can be determined.

2.3.2.4 Accumulation and Inversion Layer

In the previous discussion the majority of the carriers in the metal oxide semiconductor were electrons and they migrated to the surface, giving rise to a so-called double layer. This is the normal case for semiconductor metal oxide sensors. There are, however, two other cases. One relates to the situation when there is an excess of majority carriers on the surface or an accumulation layer. The other case has an excess of minority carriers on the surface or what is called an inversion layer.

The accumulation layer is caused by occupied surface donor states that are located near the conduction band of the semiconductor, as shown in Figure 2.16. The donor electrons are injected into the conduction band and an excess of electrons appear on and near the surface, and positively ionized donor states appear near the surface.

A surface inversion layer is formed if the unoccupied surface acceptor states are located near the top of the valence band, as shown in Figure 2.17. In this case, electrons move from the valence band to occupy surface acceptor states. In this case, the surface has negatively charged ionized acceptors, positively charged oxygen vacancy donors whose electrons now occupy surface acceptor states, and holes. This is called an inversion layer since it has

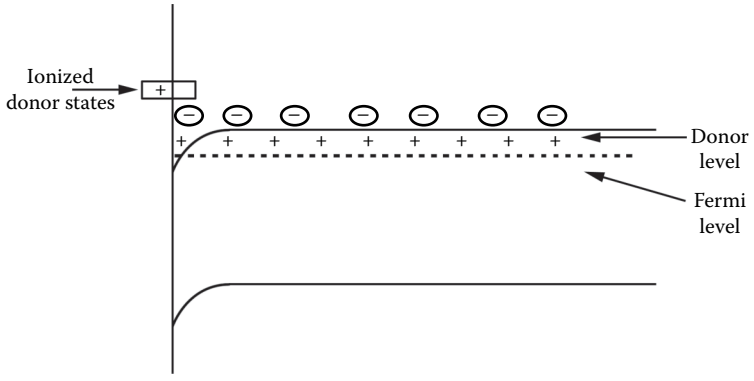


FIGURE 2.16
An accumulation layer in *n*-type metal oxide whose surface donor states are in the conduction band.

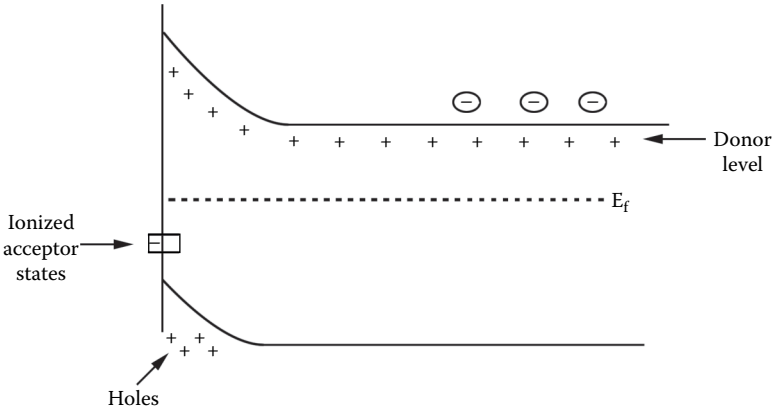


FIGURE 2.17
An inversion layer in an *n*-type metal oxide that has ionized surface acceptors, ionized oxygen vacancy donors, and holes on the surface.

excess holes on the surface while the metal oxide is *n*-type. It is interesting to point out that even though the majority of carriers in the metal oxide semiconductor bulk are electrons, the positively charged holes near the surface form a *p*-type channel. Holes then move through this channel and increase the conductivity. This is in fact the primary conducting channel in the silicon field effect transistor (FET).

2.3.2.5 Surface States and Surface Conductivity

Basically two types of conductivity can exist in the metal oxide semiconductor, the bulk conductivity, which is given as

$$\sigma_b = qN_D\mu_n \tag{2.19}$$

and the surface conductivity, which is due to the surface states located at the semiconductor metal oxide surface. The surface conductivity, σ_s, may be defined as follows:

$$\sigma_s = qN_s\mu_n \tag{2.20}$$

where N_s = number of carriers per unit area on the semiconductor metal oxide surface.

If one assumes a double-layer model as shown in Figure 2.11, N_s may be represented as follows:

$$N_s = N_D t \tag{2.21}$$

where t = semiconducting metal oxide film thickness.

Therefore, using Equations 2.19 to 2.21, the following relationship exists between surface and bulk conductivity:

$$\sigma_s = \sigma_b t \tag{2.22}$$

The semiconductor metal oxide can be modeled as shown in Figure 2.18.

2.3.2.6 Metal Oxide Semiconductor Film Structure

The previous discussion applied to the case of a single crystal film. Unfortunately, it is very difficult to grow single crystal films, and therefore they are seldom used in actual metal oxide semiconducting sensors. However, the semiconductor

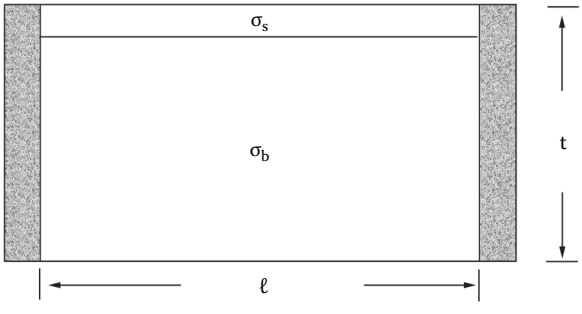


FIGURE 2.18
Model of the metal oxide semiconductor film with surface and bulk film conductivities. The gray regions are electrodes.

theory previously discussed can be extended in an attempt to describe the electrical properties of the other film structures. Basically, the semiconductor metal oxide film may occur in one or more of the following forms:

1. Single crystal
2. Amorphous
3. Polycrystalline
4. Compressed powders

With the exception of the compressed powders, most films are deposited using either vacuum evaporation or sputtering techniques. Vacuum deposition does not offer a high degree of control in terms of film deposition parameters. As a result, reproducibility is poor and it is not uncommon for as deposited films to have conductivities that may vary by several orders of magnitude. Sputter techniques, on the other hand, offer a high degree of control in terms of deposition parameters, resulting in better film reproducibility from deposition to deposition. Sputtering parameters such as argon/oxygen mixture, RF or DC power, and chamber pressure can be maintained very accurately. As a result, sputtering is the preferred technique for film deposition.

Most metal oxide films are deposited at or near room temperature and the film structure is amorphous. This structure is not ideal for sensing applications. First, the structure is not stable, causing film properties to change with time. Second, these films are prone to sorb and desorb water significantly, which causes variations in film electrical properties. Third, the film oxidation is not complete, which also causes electrical properties to change with time. Fortunately, these films can be stabilized to a reasonable extent by heat treating the film. This process, which may be called curing, rids the film of water, stabilizes the film structure, and hastens the oxidation process. In regard to the film structure, the film undergoes an amorphous-to-polycrystalline phase transition. This is clearly seen by referring to Figure 2.19, which shows the variation of the experimentally measured conductivity of WO_3 as a function of inverse temperature [44]. At about 315°C a discontinuity in the conductivity is observed, indicating the amorphous-to-polycrystalline phase transition. This has also been substantiated with transmission electron micrographs (TEMs). Relative to oxidation, the as deposited films have oxygen vacancies on the surface and the bulk. As the film is heated, the oxidation of the film due to ambient oxygen accelerates and then stabilizes. The stability of the film is clearly evidenced in Figure 2.20, where the conductivity of cured and as deposited films are plotted as a function of time [43].

In order to eliminate the curing process, it is also possible to sputter deposit the metal oxide films at elevated temperatures. However, one must be aware of the transition temperature between the amorphous and polycrystalline phases since different metal oxides have different transition temperatures.

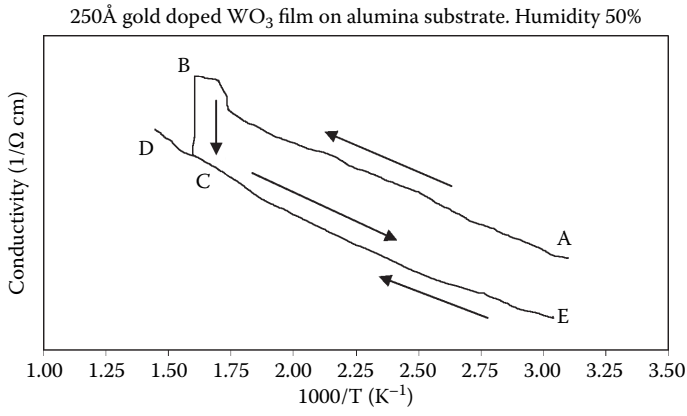


FIGURE 2.19
(Please see color insert following page 146) The electrical conductivity as a function of inverse temperature for a gold-doped 250 Å thick WO_3 film deposited on an alumina substrate. Arrows indicate the direction of increasing and decreasing temperature. The measurements were performed in air at 50% relative humidity. The reproducible measurement error was about 10%. (From Xu, Z., Electrical Properties of Thin Tungsten Trioxide Films, MS thesis, Electrical Engineering, University of Maine, 1990. With permission.)

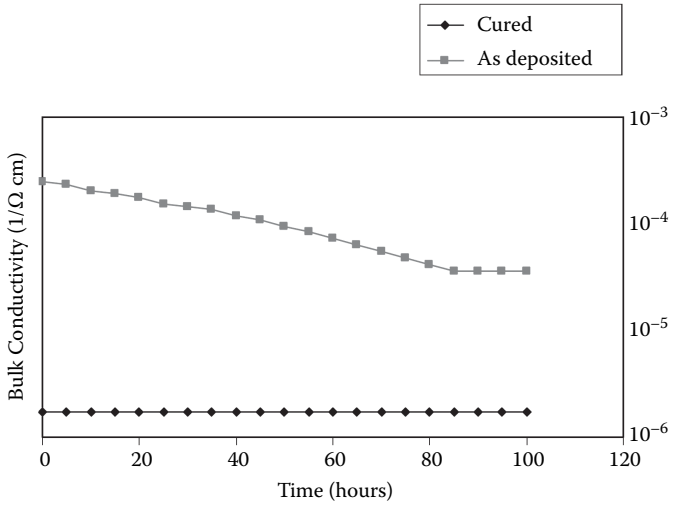


FIGURE 2.20
(Please see color insert following page 146) As deposited and cured WO_3 film conductivity variation vs. extended time at 200°C and relative humidity 20% in ambient air. (From Smith, D. J., Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications, MS thesis, Electrical Engineering, University of Maine, 1992. With permission.)

Compressing powders of metal oxide semiconductors has been another popular way of preparing films. The approach is rather simple and does not require expensive equipment such as an RF sputtering system. Despite the fact that the method is relatively crude, the reproducibility of the film's electrical properties is reasonably good. Many sensors available on the market have in fact been prepared in this fashion.

Surface processes dominate the film resistance in compressed powders. The resistance is due primarily to the intergranular boundaries where the depletion layer is located. This is obvious by referring to Figure 2.21 [45]. The single grain can be thought of as being equivalent to a small single crystal. It is interesting to note that the reproducibility in the compressed powders is problematic. Also, it is seen that the grain boundary increases the film resistance. This is also observed in polycrystalline films (see Figure 2.19) where the conductivity decreases or film resistance increases as the film becomes polycrystalline.

Since the intergranular boundaries appear in both compressed powders and polycrystalline films, it is worthwhile to examine these boundaries more closely. Typical crystallites that appear in a compressed powder or a polycrystalline sample are shown in Figure 2.22. Since the depletion region and the associated double layer, as shown in Figure 2.11, normally occur in *n*-type metal oxide semiconductors, the surface of the crystallites

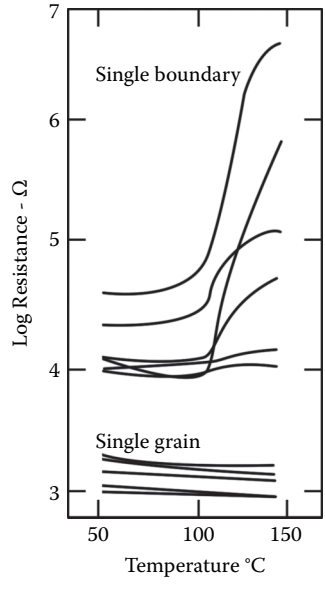


FIGURE 2.21
The variation of the resistance vs. temperature in a single grain and across a grain boundary on a compressed powder ceramic. (From Madou, M. J. and Morrison, S. R., *Chemical Sensing with Solid State Devices*, Academic Press, 1989, p. 36. With permission.)

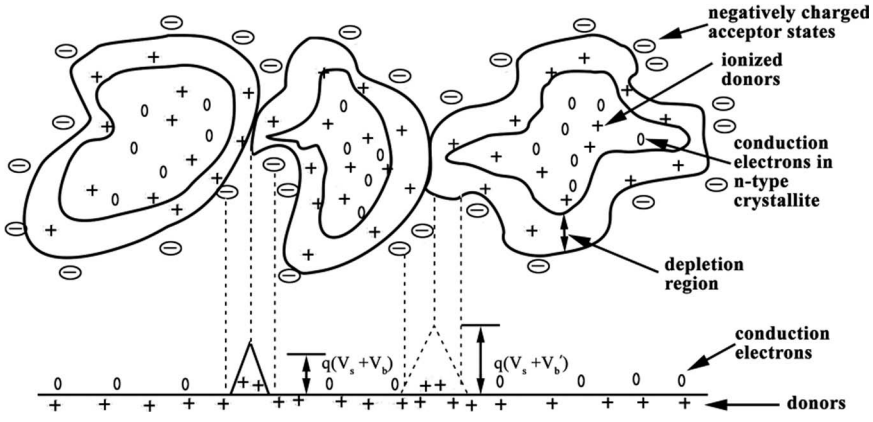


FIGURE 2.22
Physical structure and energy band model of crystallites in polycrystalline materials or powders. Note that the intercrystallite potential, qV_b, is usually different between different crystallites.

is negatively charged with partially occupied surface acceptor states due to the surface metal ions. The charge region space is depleted of mobile carriers, and hence most of the resistance appears at the depletion region and at the intercrystallite boundaries. Also shown in Figure 2.22 is the appropriate energy band structure. It is clearly seen that the electrons must overcome the potential barrier and the intercrystallite potential, qV_b, in order to move from one crystallite to another. The intercrystallite potential is due to the fact that neighboring crystallites have different crystallographic orientations, and hence different energy band structures. In addition to the films occurring in single crystal, amorphous, polycrystalline, and powder form, they can also exist in a state that may be a combination of two or more of the above forms. Two common forms are the sintered and calcinated forms. In the former, heat is applied to the constituent materials to help fuse the material together. In the latter case, the constituent materials are also heated in air, but a significant chemical reaction takes place. In particular, in the calcination process metals are used and transformed into oxides via oxidation. Sintered and calcinated film normally consist of either a single metal oxide or a combination of several metal oxides and other materials and are referred to as a ceramic.

Sintered and calcinated films have a physical structure that is a combination of the single crystal, polycrystalline, and powder forms. Examples of two different structures that may be associated with these films are shown in Figure 2.23.

The conductivity or resistance of these necks is strongly dependent on the neck cross-sectional area. In the two-dimensional illustration shown in Figure 2.23, the dimension, d, will dictate whether the conductance is crystalline type or intercrystallite type. Finally, considering all the possible "neck sizes," one can see that the conduction process in these kinds of films may

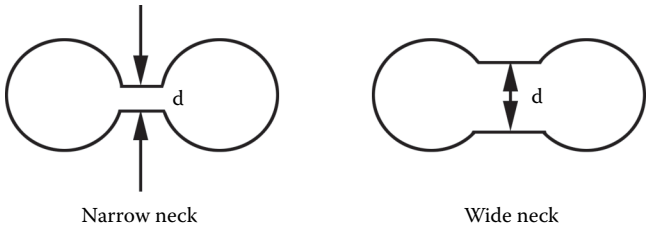


FIGURE 2.23
Examples of narrow and wide necks in sintered or calcinated films.

be quite complex. Unfortunately, there is no consistent technique that can be used in the preparation of the sintered and calcinated films that can accurately control neck sizes.

Finally, one must note that the previous discussion related to film structure and electron transport in one dimension (see Figure 2.22). This, of course, does not occur in the real world. The film structure in three dimensions is somewhat analogous to a boulder-strewn beach. Electrons must navigate through the maze of three-dimensional crystallites in order to go from one electrode to the other. To characterize this transport process exactly is a formidable task.

2.3.2.7 Gas-Semiconductor Film Interactions

In order to try to understand the interaction between a sensing film and a target gas, the processes occurring at the gas–solid interface need to be examined. These types of processes can include interactions with the ambient atmospheric gases such as oxygen, hydrogen, and water vapor and various target gases.

A gas can be adsorbed on a solid by either physisorption, chemisorption, or ionosorption. In physisorption, the bonding is considered to be weak, while in chemisorption the bonding is at least 2½ times greater. In ionosorption, no bonding occurs and the adsorbate acts as a surface state capturing an electron or a hole. It is therefore held to the surface by an electrostatic interaction. When the physisorption or chemisorption processes take place, the surface atoms in the film may move about and relocate themselves. Bonds of the surface atom to the bulk atoms will break, and new bonds will form with the adsorbing species. This will result in the formation of a new surface phase. Using experimental spectroscopic techniques, such as low-energy electron diffraction (LEED), one can track the formation of this new phase.

Instead of trying to discuss the adsorption processes microscopically, the present discussion will focus on the transferring of charge from the conduction or valence band in the adsorbent (film) to the adsorbate

(adsorbing species). This transfer of charge will ionize the adsorbate. The rate of physisorption, chemisorption, or ionosorption of an adsorbate is complex, particularly when one is considering the formation of one or more monolayers of adsorbate with adsorbent. For submonolayer coverages that normally occur in gas sensing, the rate is proportional to the adsorbate pressure.

In physisorption, the force between the adsorbate and the adsorbent is a dipole-dipole interaction. Basically, the molecule in the ambient can be approximated to be a dipole. This adsorbate dipole can then polarize or induce dipoles in the absorbent, resulting in a weak dipole-dipole interaction.

The interactions inherent in chemisorption are much stronger. Normally the adsorbate molecules will dissociate into atoms. These atoms can then form surface and chemical bonds with the absorbent. Ionosorption is very important in gas sensing, particularly as it relates to oxygen. The ionosorbed oxygen basically exists in two forms, O_2^- and O^-. These negatively charged surface states augment the surface states discussed earlier relative to the double-layer case encountered in metal oxide semiconductors (see Figure 2.11). The surface states also provide the majority of the surface charge, N_s, and increase the surface barrier potential. This in turn causes the high resistivity or low conductivity normally encountered in metal oxide semiconductors.

The ionosorbed oxygen can be removed from the surface by reducing agents. For example, if the surface is exposed to a hydrocarbon (reducing agent) such as methane, one would obtain water and carbon monoxide as byproducts. The surface potential barrier would decrease and the conductivity would increase. The conductivity increase would be directly related to the concentration of methane in the ambient. Many gas sensors operate on this principle.

2.3.2.7.1 Effect of Adsorbed Gases on the Metal Oxide Electrical Properties

When gases adsorb on a semiconductor surface, two major changes in the semiconductor film can occur. First, the electron or hole density can change, and second, the height and width of the barrier potential can either increase or decrease. Since the most common ambient in which metal oxide sensors operate consists of air, one must determine the effects that ambient air has on the film properties. Ambient air basically consists of about 80% nitrogen and 20% oxygen with trace amounts of other gases. It is known experimentally that the film properties do not change when exposed to a 100% nitrogen atmosphere. However, the same cannot be said for oxygen. Further, since many sensing films are either compressed powder or polycrystalline, they contain a large number of potential barriers that electrons must either tunnel through or hop over. Therefore, it is important to analyze in detail the effects adsorbed oxygen has on the metal oxide film properties.

The adsorption of oxygen can be described in terms of the following reaction steps:

$$e^- + O_2 \longrightarrow O_2^- \tag{2.23}$$

$$O_2 \longrightarrow 2O \tag{2.24}$$

$$e^- + O \longrightarrow O^- \tag{2.25}$$

The ionosorbed entities are O_2^- and O^-. Basically, either O_2 from the ambient directly picks up an electron to form O_2^-, or O_2 is broken into two oxygens on the metal oxide surface and then attracts an electron from the metal oxide bulk. The amount of oxygen adsorbed is directly proportional to the oxygen pressure in the ambient.

One can obtain the amount of oxygen ionosorbed on the surface by noting that gas sensors are operated in the ambient. This places limits on the value of the barrier potential, which is usually between 0.5 and 1 V. Referring to Equations 2.15 and 2.21 one can see that density of surface charged states is limited, and that the corresponding surface coverage for oxygen is about 10^{11} to 10^{12} ions/cm^2.

The intergranular or intercrystallite boundaries provide most of the electrical resistance in compressed powder and polycrystalline films. The resulting electrical current is controlled by the density of electrons, N_s, at the semiconductor surface, which is proportional to the barrier potential, qV_s (see Equations 2.15 and 2.21). In addition, one must also contend with the intercrystallite potential barrier.

In metal oxide sensors two different scenarios can occur when a target gas adsorbs on the sensor surface. The amount of ionsorbed oxygen can either decrease or increase. In the case of the former the individual barrier heights decrease. The target gas in this case is referred to as a reducing agent, and the resulting target gas-film interaction is called reduction. In the latter case the ionsorbed oxygen concentration increases, causing a corresponding increase in the barrier height. In this case the target gas is referred to as an oxidizing agent and the resulting target gas-film interaction is called oxidation.

The reaction with a reducing agent, R, can be expressed as follows:

$$R + O_2^- \xrightarrow{k_1} RO_2 + e^- \tag{2.26}$$

and

$$R + O^- \xrightarrow{k_2} RO + e^- \tag{2.27}$$

where k_1 and k_2 are the rate constants for the reactions.

The reactivity of O_2^- is much less than that of O^-; therefore, Equation 2.26 can be ignored. Examples of typical reducing agents include carbon, carbon monoxide, hydrogen, and methane. The associated reactions may be written as follows:

$$C + O^- \longrightarrow CO + e^- \tag{2.28}$$

$$CO + O^- \longrightarrow CO_2 + e^- \tag{2.29}$$

$$H_2 \longrightarrow 2H; 2H + O^- \longrightarrow H_2O + e^- \tag{2.30}$$

$$CH_4 \longrightarrow C + 4H; C + 4H + 3O^- \longrightarrow CO + 2H_2O + 3e^- \tag{2.31}$$

In the four cases presented it can be clearly seen that the amount of ionsorbed oxygen is decreased, causing the barrier potential to decrease and excess electrons to be produced. These two processes cause the metal oxide film resistance to go down or the conductivity to increase. This is the basis of operation for many gas sensors [46]. The results of a Taguchi sensor exposed to carbon monoxide (CO) are presented in Figure 2.24 [47].

The reaction with an oxidizing gas is typically a multistep process and occurs most frequently with oxygen-bearing gases. The most common target oxidizing gases are NO and NO_2, which result in the following reactions:

$$2NO \longrightarrow N_2 + O_2 \tag{2.32}$$

$$O_2 \longrightarrow 2O$$

$$O + \bar{e} \longrightarrow +O^-$$

and

$$2NO_2 \longrightarrow N_2 + 2O_2 \tag{2.33}$$

$$2O_2 \longrightarrow 4O$$

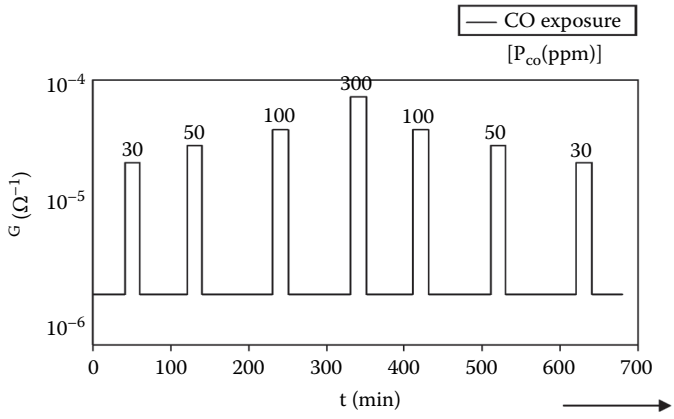


FIGURE 2.24
(Please see color insert following page 146) Response of a TGS 812 sensor (SnO_2) in air (50% humidity) to various CO exposures. (Reprinted from Schierbaum, K. D., Weimar, V., Kowalkowski, R., and Göpel, W., *Sensors Actuators* B3 [1991]: 205. With permission from Elsevier.)

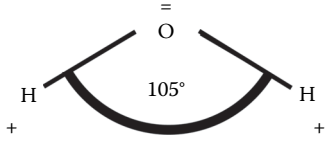


FIGURE 2.25
Oxygen molecule.

$$O + e^- \longrightarrow O$$

Examination of Equations 2.32 and 2.33 reveals that electrons in the metal oxide are used to combine with oxygen to form ionsorbed O^-. This results in an overall decrease in free electrons and a corresponding increase in the barrier potential height. As a result, resistance increases and conductivity goes down.

Another interesting interaction that takes place in metal oxide semiconductors is the interaction with hydrogen (H_2). Contrary to oxygen, which extracts electrons from the metal oxide, hydrogen donates electrons. It can interact with the ionosorbed oxygen as shown in Equation 2.30, or it can interact directly with the metal oxide, as shown below for ZnO:

$$H_2 \longrightarrow 2H; H + ZnOH^+ + e^- \tag{2.34}$$

When a large amount of hydrogen or hydrogen-bearing gases are present, the surface can actually become positively charged, resulting in the formation of an accumulation layer (see Figure 2.16).

In addition to interacting with reducing and oxidizing agents the metal oxide can also interact with water. It is important to discuss this situation since water vapor always occurs in the ambient atmosphere when a sensor is operating. Also, H_2O is produced in the interaction of hydrogen and reducing agents such as methane (see Equations 2.30 and 2.31). In order to completely understand the interaction of H_2O with a metal oxide surface we must realize that water is a polar molecule in which the hydrogen ions are separated by 105°, as shown in Figure 2.25.

One also notices that the H_2O molecule can be approximated as a dipole. There are basically two processes associated with water adsorption. In the first process a hydroxylated surface is formed. In this process the OH^- is attracted to the metal ion while the H^+ ion or proton is attracted to the oxide ions. This results in a net change of carrier density in the oxide since it eliminates the O^- ion from the surface state. This hydroxylated surface is shown in Figure 2.26. The second process involves the physisorption of water on the hydroxylated surface. This can appear at the interface between grains or polycrystallites in the metal oxide and will affect the resistance of the metal oxide.

Heat treatments and operating the sensor at elevated temperatures will minimize the effect of water. Normally, the physisorbed water can be removed at low temperatures (~400° K); however, hydroxylated water needs a temperature of about 700° K to be removed.

The final gas interaction with a semiconductor surface, which will be discussed, relates to surface ion exchange. The previous interactions that have been discussed relate to surface states, adsorption (physisorption, chemisorption, and ionosorption), and reduction-oxidation reactions. The surface ion exchange relates to the replacement of atoms at the semiconductor oxide surface, which will most likely affect metal oxide electrical properties. The surface ion exchange interaction is not often seen in metal oxide gas sensors. It has been suggested [48] that the hydrogen sulfide sensor that uses tungsten trioxide as the metal oxide film operates on this basis. In particular, it has been postulated that the lattice oxygen is replaced with sulfur to form a tungsten disulfide (WS_2) layer. This could result in higher conductivity since WS_2

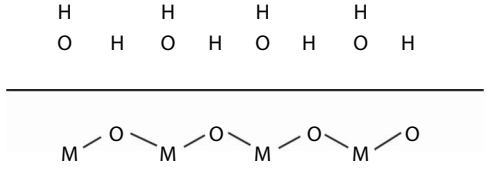


FIGURE 2.26
Hydroxylated surface on a typical metal oxide (MO).

is conductive. However, contributions due to hydrogen and water vapor also play a role in the interaction and should be considered. Also, the question of reversibility in responses would be difficult to explain for anion exchange interaction.

2.3.2.7.2 Sensitivity and Selectivity

As a result of the previous discussion, it is obvious that the interesting gas-semiconductor film interactions take place at the film surface and at the interfaces between the grains or crystallites. These interactions modify the barrier potential and carriers in the film, which in turn change the film impedance. The question of how one can impart sensitivity and selectivity to the film has not been explored fully.

It has been pointed out earlier that sputtered films that have been heat treated or cured offer advantages over films prepared by other techniques. Film parameters such as conductivity can be controlled to a reasonable degree by carefully adjusting the sputtering parameters. Curing the film was found to cause the film to become polycrystalline and stabilize the film's electrical properties (see Figure 2.20). Polycrystalline films are particularly attractive in gas sensing applications since they have a large interface area between the crystallites, therefore maximizing the gas-film interactions. In fact, one can make an interesting comparison: the surface of a cured film is more like the Rocky Mountains, whereas the surface of an as deposited film is more like the Great Plains. Also, it has been shown [44] in cured WO_3 films that the response times to a target gas (H_2S) are shorter and the sensitivity is improved compared to an as deposited WO_3 film [43]. This is clearly seen by referring to Figure 2.27.

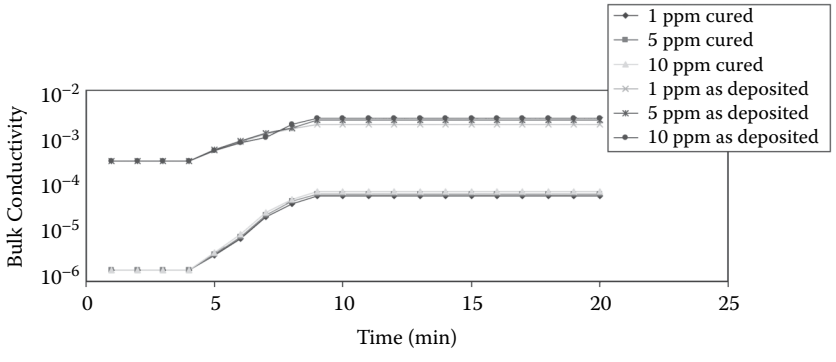


FIGURE 2.27
(Please see color insert following page 146) Bulk conductivity of an as deposited and cured WO_3 film sample vs. time for various H_2S gas concentrations (sample temperature 200°C and humidity 20%). (From Smith, D. J., Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications, MS thesis, Electrical Engineering, University of Maine, 1992.)

If the sensor is maintained at a constant temperature, this can have an effect on the sensitivity and selectivity. The rate of reaction with a particular target gas is temperature dependent. The case of the reaction with a reducing or oxidizing agent serves as an excellent example, since these reactions are often seen in sensors. If the temperature is too low, the reaction is too slow to yield a high sensitivity. If the temperature is too high, the reaction proceeds so quickly that the concentration at the surface becomes diffusion limited, and again the sensitivity is low. For example, in the case of a SnO_2-based sensor, the optimum temperature for methane is 450°C, while for CO and C_3H_8 the corresponding temperatures are 200 and 350°C. Therefore, one can use a single SnO_2 sensor to selectively and sensitively detect three different gases by simply operating the sensor at different temperatures. The influence of temperature on sensitivity and selectivity is also seen in the WO_3-based H_2S and H_2 sensor [43], as shown in Figure 2.28. In this case the optimum temperature for H_2S is 200°C, while for H_2 it is about 240°C.

Perhaps one of the most interesting techniques used to impart selectivity and decrease the response time is the use of a catalyst. Noble metal catalysts such as platinum (Pt), palladium (Pd), iridium (Ir), rhodium (Rh), silver (Ag), and gold (Au) have been frequently used in semiconductor metal oxide sensors. In order to completely understand how to use catalysts in gas sensors, one should have a complete understanding of catalysis. Even with this background it will become clear that much more research needs to be done in order to optimize the use of catalysts in gas sensing.

In order to understand the role of the catalyst in a gas sensor, a short discussion of catalysis will be given. Before a reaction can be initiated, a certain amount of energy called activation energy must be supplied. For example, in the adsorption of a gas such as O_2 in a gas sensor, the O_2 molecule must be dissociated as described in Equation 2.24 before it can react with another atom or molecule. In addition to the dissociation of a molecule, other processes that may have to take place before a reaction can appear on the sensor surface include ionization or some intermediate reaction. The energy necessary to initiate the reaction is usually in the form of thermal energy. This is most commonly supplied by heat, light, or photons. The actual amount of activation energy necessary to initiate the reaction depends upon the reaction path. If a particular path requires a high activation energy, the reaction rate will be low. However, if another path requires a lower activation energy, the reaction rate will be faster.

Typically, a catalyst is added to a metal oxide semiconductor to speed up a chemical reaction. Ideally, finely divided particles of the catalyst should be added so that a large surface area is covered in the metal oxide. The catalyst, although critical to the chemical reaction, does not undergo any changes during the reaction. The catalyst can play two important roles in the chemical reaction. First, it can concentrate the reactants by adsorbing the appropriate gas molecules, and second, it can provide a reaction path that requires a small amount of activation energy. The second role is best illustrated by the

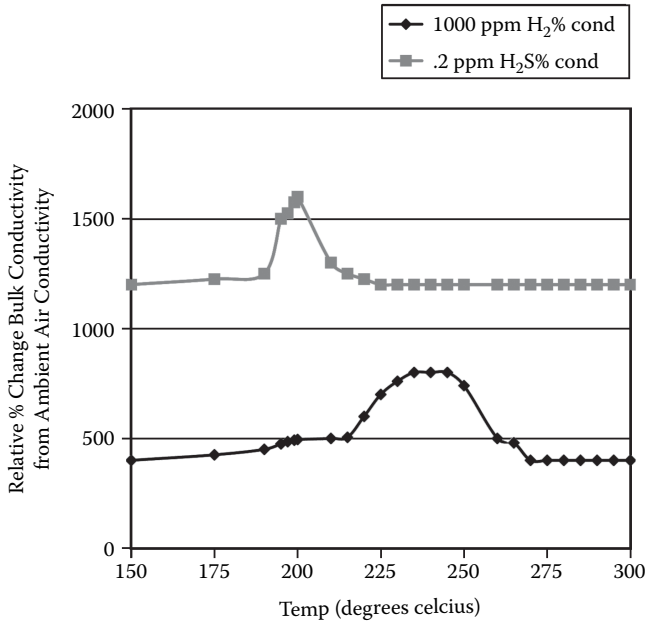


FIGURE 2.28
(Please see color insert following page 146) Average relative percent change in bulk conductivity of a WO_3 cured film from its ambient air conductivity vs. temperature for exposures of 0.2 ppm H_2S and 1,000 ppm H_2 (humidity 20%). (From Smith, D. J., Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications, MS thesis, Electrical Engineering, University of Maine, 1992.)

fact that a metal catalyst such as platinum can rapidly dissociate a molecule such as hydrogen into reactive H atoms. Once hydrogen is in this form, the necessary activation energy required for a chemical reaction is much less. Without a catalyst present the required activation energy is much higher since it first has to dissociate molecular hydrogen (H_2) before supplying the energy necessary for the chemical reaction to proceed.

In addition to increasing the rate of the chemical reaction, the catalyst also plays an important role in selectivity. Certain catalysts will greatly enhance the rate of some reactions and have little or no effect on other reactions. This is particularly true for gold. In the case of gases such as H_2S and H_2 a gold catalyst rapidly dissociates these gas molecules as follows:

$$H_2 \leftrightarrow 2H \tag{2.35}$$

and

$$H_2S \leftrightarrow 2H + S \tag{2.36}$$

If oxygen-bearing gases such as CO and CO_2 are exposed to gold, little or no dissociation occurs. Platinum and palladium, however, serve as excellent catalysts for many gases, a feature that is not desirable for selectivity in a gas sensor.

In order to be effective in gas sensing applications, a catalyst has to change the resistance of the semiconducting metal oxide. The catalyst is basically deposited on the surface of the grains or crystallites in the metal oxide semiconductor. Since the catalyst can concentrate the reactants by adsorbing the appropriate gas molecules, one can see that the rate of the chemical reactions on the catalyst will be high. This is shown for the case of a reducing agent reacting with surface oxygen in Figure 2.29. Since the products of the reactions will disappear into the atmosphere, it is not clear as to how this will affect the metal oxide semiconductor resistance. In fact, the catalyst may "burn" all of the reducing agent. This will have little or no effect on the film resistance, and therefore the sensitivity of the sensor. Fortunately, there are two techniques by which the catalyst can significantly affect the metal oxide resistance. These techniques are called *spillover* and *Fermi energy control*.

In spillover the metal catalyst dissociates the molecule and the atoms spill over onto the semiconducting metal oxide surface, as shown in Figure 2.30 for O_2.

Similar dissociations can occur for hydrogen and other atmospheric gases. Once the oxygen or hydrogen has been dissociated on the catalyst, the following sequence of reactions takes place:

$$O + e^- \longrightarrow O^- \tag{2.37}$$

$$2H + O^- \longrightarrow H_2O + e^- \tag{2.38}$$

$$RH_2 + 2O^- \longrightarrow RO + H_2O + 2e^- \tag{2.39}$$

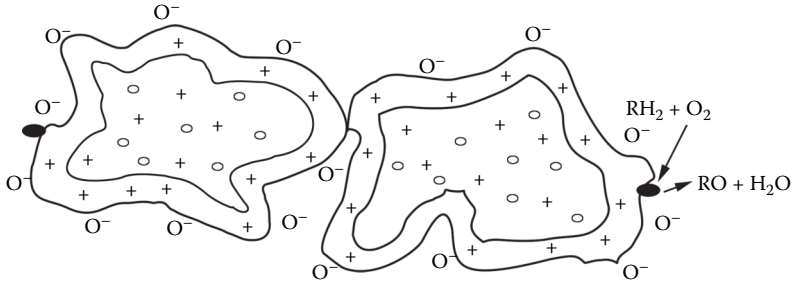


FIGURE 2.29
Catalysts (black dots) on the surface of a crystallite. Since they are isolated and not located at the crystallite interfaces, the catalysts will have no significant effect on the film resistance.

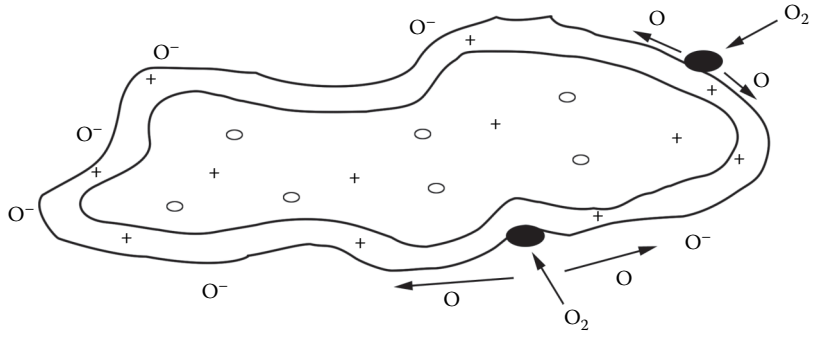


FIGURE 2.30
Spillover effect where the metal catalyst dissociates a molecule, in this case O_2.

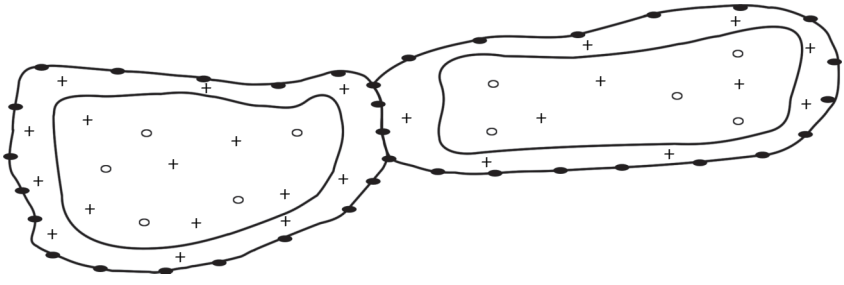


FIGURE 2.31
Finely distributed catalysis over two crystallites insuring that the intercrystallite boundary is populated by the catalyst.

Note that the concentration of O^- ions on the semiconductor surface is critical to the reactions. Since high concentrations of O^- means high resistance and the reducing agent removes O^-, one concludes that the resistance decreases quickly with increasing concentrations of the reducing agent. Obviously for the resistance to change, the spillover oxygen and hydrogen atoms must be able to migrate to the intercrystallite boundaries. This can be facilitated if the catalysis is distributed finely over the crystallite surface as shown in Figure 2.31.

The fact that this occurs in a real metal oxide is dramatically shown in Figures 2.32 and 2.33 for polycrystalline samples undoped and gold-doped WO_3 [44]. The small white dots in the gold-doped sample are the finely divided gold particles.

The other way in which the catalyst can affect the metal oxide resistance is by Fermi energy control. In this situation the oxygen that is concentrated on the metal catalyst obtains its electrons from the metal catalyst. The metal catalyst in turn removes electrons from the metal oxide. If the deposited catalyst is in high concentration and other surface adsorbates, such as charged oxygen on the metal oxide, can be neglected, the catalyst controls the barrier potential. It is then clear that the catalyst controls the flow of electrons from

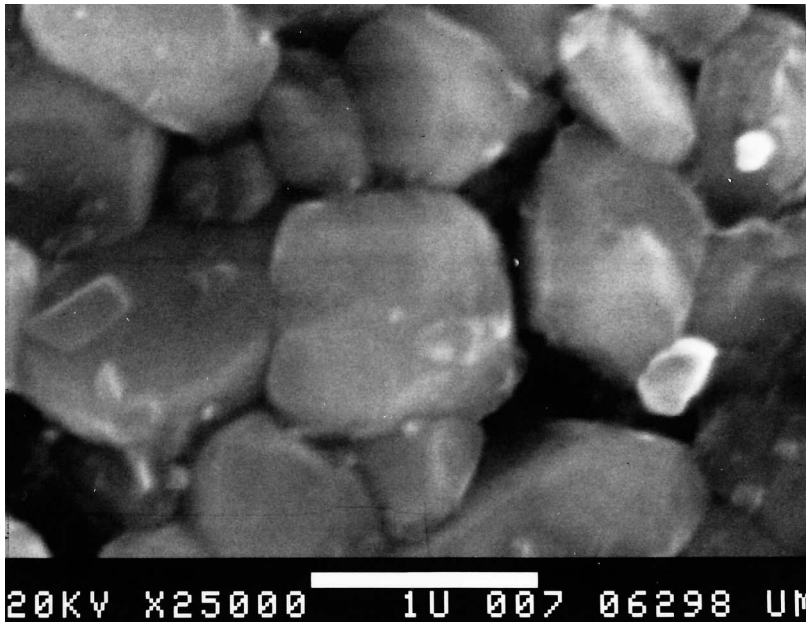


FIGURE 2.32
(Please see color insert following page 146) Scanning electron micrographs (SEMs) of an undoped polycrystalline WO_3 film on an alumina substrate. (From Xu, Z., Electrical Properties of Thin Tungsten Trioxide Films, MS thesis, Electrical Engineering, University of Maine, 1990.)

the metal oxide. Obviously for Fermi energy control to work, not only must the concentration of the metal catalyst be high, but catalyst particles should be finely dispersed, as shown in Figure 2.31. Only then will intercrystallite resistance be affected. Typically the catalyst separation should be less than the depletion width, or roughly less than 500 Å apart.

Often certain materials called promoters are added to metal oxide semiconductors to help improve the selectivity or reactivity of the catalyst. They typically influence the effect of the catalyst microscopically and have an influence on valence states or stoichiometry. Overall, little is understood on the exact role played by promoters. Usually promoters are stumbled upon by trial and error.

2.4 Chemiresistors

If a conductimetric sensor uses a film other than a metal oxide semiconductor, it is usually referred to as a chemiresistor. The films that are typically used in the chemiresistor include organic films and metallic films.

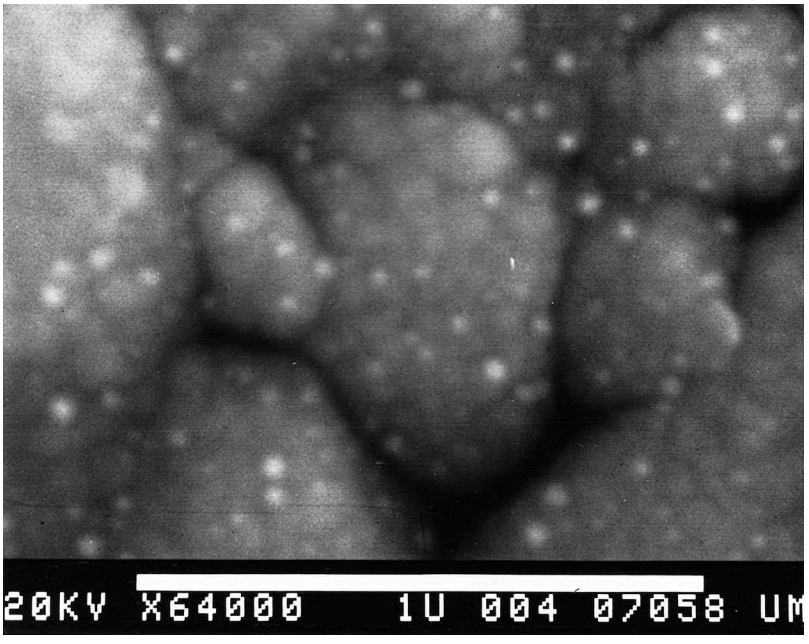


FIGURE 2.33
(Please see color insert following page 146) SEMs of a gold-doped polycrystalline WO_3 film on an alumina substrate. The little white dots are present on the finely divided gold particles. (From Xu, Z., Electrical Properties of Thin Tungsten Trioxide Films, MS thesis, Electrical Engineering, University of Maine, 1990.)

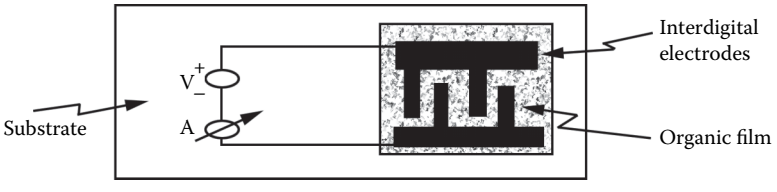


FIGURE 2.34
An example of an interdigitated chemiresistor with an organic film.

A common configuration for a chemiresistor with an organic semiconductor film is shown in Figure 2.34. The film is applied over interdigital electrodes that are vacuum deposited on an insulating substrate such as quartz or alumina. The film is usually deposited using solvent spin coating or the Langmuir-Blodgett (L-B) technique [49]. Spin coating is a familiar technique that is often used in photolithographic processes. The L-B technique is a method by which one can deposit an organic film layer by layer. This technique offers a high degree of precision in terms of controlling the film thickness. Unfortunately, it is limited primarily to organic materials. The more common organic materials used are based upon the metal phthalocyananines

(MPCs), shown in Figure 2.35. MPCs basically have a metal ion in the center that is attached through nitrogen atoms to various benzene rings. MPCs are *p*-type organic semiconductors that can interact strongly with gases such as nitrogen oxide (NO_2), ammonia (NH_3), and chlorine (Cl_2), resulting in large changes in electrical conductivity. By varying the metal ion (Cu, Pb, Mg, Fe, Co, Ni) in the MPCs, sensor properties such as sensitivity, selectivity, reversibility, and response times can be varied.

Although promising results have been achieved using organic semiconductors such as MPCs, the long-term stability of these organic semiconductor sensors is poor. The sensors may find limited use as disposable devices in various biosensor applications or in cheap badges used by workers exposed to various types of gases.

An example of a chemiresistor that utilizes a metallic film is shown in Figure 2.36. In this sensor, the metallic film selectively adsorbs the target gas molecule. This molecule then acts as a scattering center for electrons

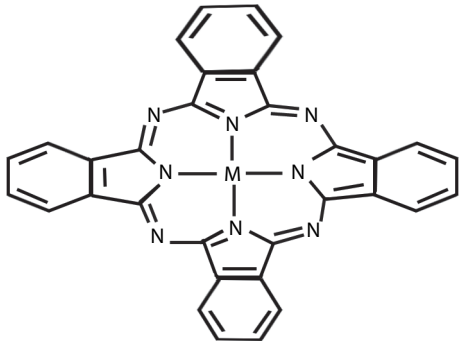


FIGURE 2.35
The chemical structure of metallophthalocyananine (MPC). The hexagons represent benzene type rings; M, the metal ion; and N, nitrogen atoms.

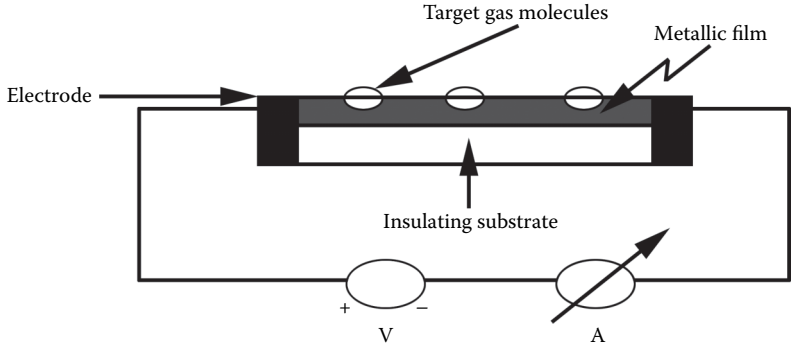


FIGURE 2.36
An example of a metal film chemiresistor.

propagating through the metallic film, resulting in an increase in film resistance. In preparation of this sensor, several precautions need to be taken. First, the electrodes should be of a different metal than the sensing film or shielded from the target gas. Second, the films should be relatively thin so that the contributions to conductivity changes can be accurately monitored. Although any metal can be used, gold is particularly attractive since it is reasonably selective to potential dangerous gases such as mercury, halogen gases (Cl_2, F_2, I_2, Br_2), and H_2S. The major weakness in this type of sensor is dynamic range. The transition of the metal films from a perfect conductor to an insulator takes place over a small range of gas concentrations. Therefore, this type of sensor, although very sensitive (ppb range), saturates very quickly. This type of sensor has been exploited commercially, as is evidenced by the gold film H_2S sensor available from Jerome Instrument Corporation [50]. This sensor has an operational range of 1 to 500 ppb of H_2S, and although the principle of operation is reasonably simple, it costs about $10,000!

2.5 Other Solid-State Electrochemical Gas Sensors

2.5.1 Background

Even though resistance-based metal oxide semiconductor sensors are the most common gas sensors commercially available, there are several other types of solid-state electrochemical gas sensors. They include solid-state capacitors, metal oxide semiconductor field effect transistors (MOSFETs), metal semiconductor Schottky type diodes, metal interfacial layer semiconducting (MIS) heterostructure diodes, and metal-insulating metal (MIM) heterostructure diodes. All of the aforementioned sensors exhibit changes in electrical properties such as conductivity, impedance, voltage, current, or capacitance upon exposure to a target gas. Complete descriptions of each of these sensors can be found in the literature [51]. The two most common solid-state electrochemical gas sensors, the MOS capacitor and the Schottky type sensors, will be discussed.

2.5.2 Solid-State Capacitive Gas Sensors

Any capacitor sensor consists of an arrangement of conductors and dielectrics. The specific materials and the geometric arrangement of the sensor itself enable one to classify different types of capacitive sensors. The two capacitive type sensors that will be discussed are the metal oxide semiconductor (MOS) capacitor and the micromechanical polymer capacitor.

2.5.2.1 The MOS Capacitive Gas Sensor

The physical structure and the associated band structure for the MOS capacitor are presented in Figure 2.37. The function of the insulating layer is to block carrier flow. The potential, φ_I, is the drop in potential across the insulating layer, and φ_S is the barrier potential variation in the space charge region. The barrier potential and the corresponding space charge region are similar to the situation that occurs in a double-layer metal oxide semiconductor, as discussed earlier. In the double-layer semiconducting metal oxide surface, the barrier potential in the space charge region can be modified by an applied voltage. In addition, a sorbed target gas can also modify the barrier potential, height, and width. Typically, oxidizing gases increase the barrier potential, whereas reducing gas causes a potential barrier decrease. It is critical, however, that the insulating layer thickness, t, is wide enough in order to inhibit the tunneling of electrons or holes through the potential barrier. Typically the value for t is greater than 100 Å.

The total capacitance per unit area, C, of the insulator and the space charge region is given as

$$C = \frac{C_{SC}C_I}{C_{SC} + C_I} \tag{2.40}$$

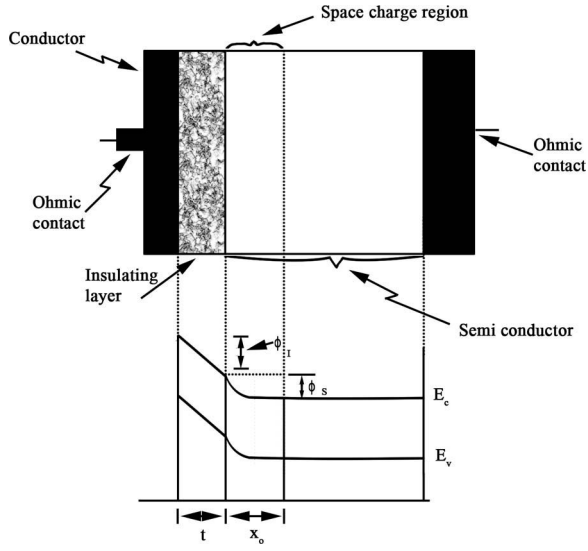


FIGURE 2.37
The physical structure and associated band structure for MOS capacitor. (From Schuetzle. D. and Hammerle. R. 1986. *Fundamentals and Applications of Chemical Sensors*. American Chemical Society. With permission.)

where C_{sc} = capacitance per unit area of the space charge region, and C_I = capacitance per unit area of the insulator.

Assuming that the structure shown in Figure 2.37 is approximated as a parallel plate capacitor, C_{sc} and C_I may be represented as follows:

$$C_{SC} = \varepsilon_{SC}/w \tag{2.41}$$

and

$$C_I = \varepsilon_I/t \tag{2.42}$$

where ε_{SC} and ε_I are the dielectric constants of the space charge region and the insulator, respectively, and w is the depletion width in the semiconductor.

As a target gas interacts with a space charge region, both the dielectric constant, ε_{SC}, and the thickness, w, of the space charge region change. Hence the space charge capacitance becomes a measure of the gas concentration. Typically the semiconductor that is used in the MOS capacitor is silicon and the insulating layer is SiO_2. Noting that the dielectric constant of Si is about four times greater than that of SiO_2, and since w is of the order of microns and t is typically of the order of a few 100 Å, the total capacitance as measured is approximately the same as the capacitance of the space charge region, and therefore a direct measure of the target gas concentration.

The materials used in the MOS capacitor sensor are transition materials such as Palladium (Pd) for the conductor, SiO_2 for the insulating layer, and silicon for the semiconductor. Since it is well known that Pd readily sorbs hydrogen and hydrogen-bearing gases, most of the work on an MOS capacitor relates to the detection of these gases. In Figure 2.38 the capacitive applied bias voltage (C-V) characteristics of a Pd/1,000 Å SiO_2/*n*-type silicon capacitor sensor are presented upon exposure to 100% O_2 and 600 ppm of the H_2 in O_2 at room temperature [54].

It can be seen from Figure 2.38 that the sensitivity, which is defined as the change in capacitance divided by the capacitive with no bias voltage, is maximum when the bias voltage is zero. It is interesting to note that in contrast to the semiconducting metal oxide sensor discussed earlier, which operates at elevated temperatures, the MOS H_2 sensor operates at room temperature. This sensor, however, can also be used at elevated temperatures to sense non-hydrogen-bearing gases such as CO. As in the case of the semiconducting metal oxide sensors discussed earlier, H_2O vapor affects the MOS capacitor sensitivity. In particular, although increased water vapor concentration decreases the response time, the sensitivity is reduced.

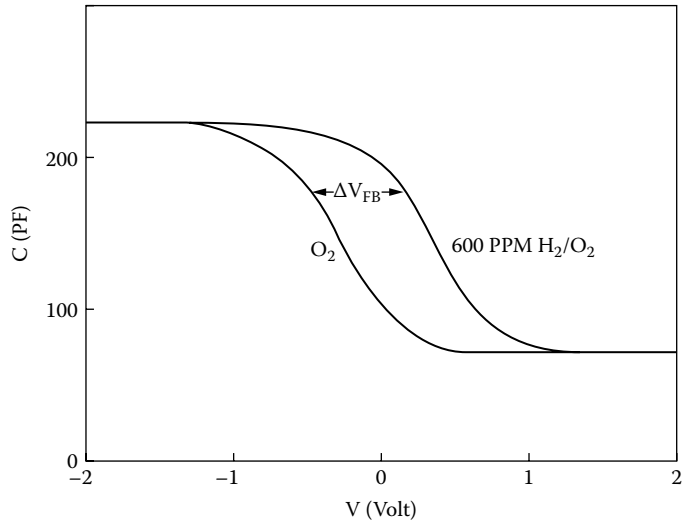


FIGURE 2.38
The physical structure and associated band structure for the MOS capacitor. (Reprinted from Schuetzle. D. and Hammerle. R. 1986. *Fundamentals and Applications of Chemical Sensors*. American Chemical Society. With permission.)

2.5.2.2 Micromachine Capacitive Polymer Gas Sensor

Recently there has been considerable interest in using polymer-based capacitive arrays for gas sensing [52]. In contrast to the single MOS capacitor sensors previously discussed, the novel use of the electrodes and the selective properties of polymer coatings have resulted in the fabrication of multiple capacitors in a single substrate. One of the major obstacles in a capacitor-based sensor is the exposure of the target gas to the chemi-selective layer. One such capacitor array configuration is shown in Figures 2.39 and 2.40. This type of arrangement can accommodate many capacitors, and the perforated top conductor plate allows the target gas to diffuse into the polymer layer.

Using different types of chemi-selective films operating at different temperatures and interfacing this structure with the appropriate electronics enables one to detect a variety of gases simultaneously [53]. In these individual sensing elements two basic changes occur: changes in the dielectric constant and changes in the volume of the polymer due to polymer swelling. This type of sensor array has been used to detect chemical warfare agents and other toxic industrial chemicals. The polymer coatings are typically deposited from an ink jet and then cured in an oven. Examples of polymers include propylaminopropyl polysiloxane (PAPPS) and siloxenefluoroalcohol (SXFA), which are selective to various target gases such as hydrazine, hydrogen cyanide, and sarin. The actual sensor system array, which includes the preconcentrator, heater

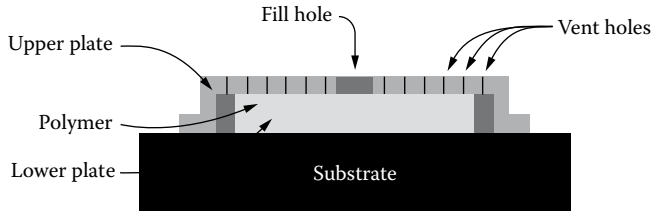


FIGURE 2.39
Side view of a single capacitor in micro machine capacitive polymer gas sensor. (Reprinted from Mlsna, T.E., Cemalouic, S., Warburton, M., Hobson, S.T., Mlsna, D.A., and Petel, S.V. 2006. *Sensors and Actuators* B 116 pp. 192–201. With permission from Elsevier.)

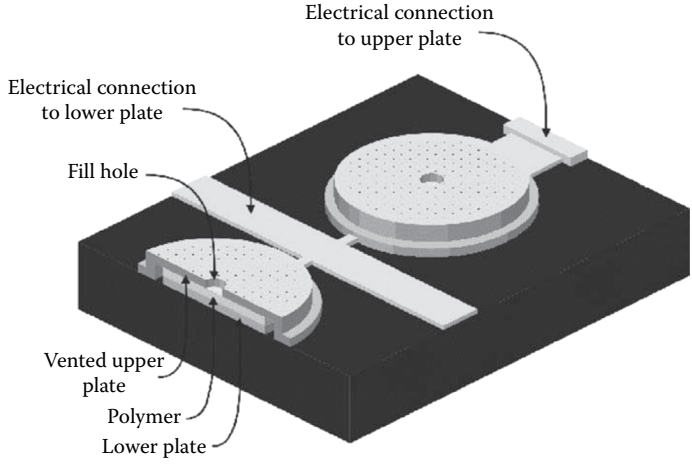


FIGURE 2.40
Top view of a single capacitor in micromachine capacitative polymer gas sensor. (Reprinted from Mlsna, T.E., Cemalouic, S., Warburton, M., Hobson, S.T., Mlsna, D.A., and Petel, S.V. 2006. *Sensors and Actuators* B 116 pp. 192–201. With permission from Elsevier.)

control board, sensor board, and battery, is relatively small and portable, as shown in Figure 2.41.

2.5.3 Schottky Diode Type Gas Sensor

The Schottky diode basically consists of a metal semiconductor (MS) configuration. However, if the metal is not a noble one such as gold, palladium, etc., then the diode would consist of a metal insulator semiconductor (MIS) configuration. When a non-noble metal is used, oxidation can occur at the metal semiconductor interface, causing an MIS type structure to result. The MIS structure and associated band structure are the same as the MIS capacitor shown in Figure 2.37, while the MS and associated band structure are shown in Figures 2.42. The critical element in these structures is the semiconductor,

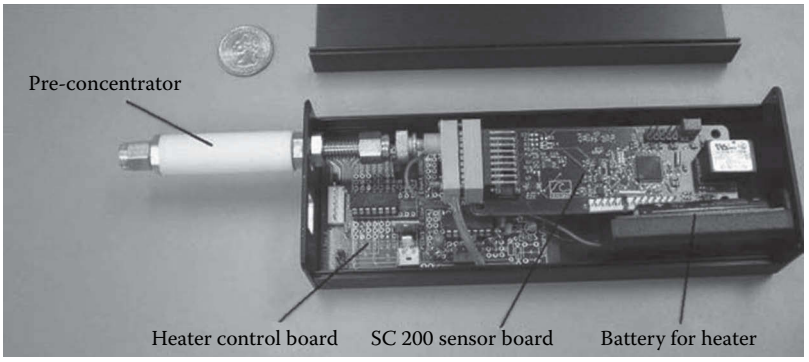


FIGURE 2.41
(Please see color insert following page 146) Photograph of an actual micromachine capacitative polymer gas sensor. (Reprinted from Mlsna, T.E., Cemalouic, S., Warburton, M., Hobson, S.T., Mlsna, D.A., and Petel, S.V. 2006. *Sensors and Actuators* B 116 pp. 192–201. With permission from Elsevier.)

since this is where the gas will interact, resulting in changes in the barrier potential and electron flow. The metal does serve a secondary purpose in that it can selectively absorb certain gases, dissociate them, and then make them available to the semiconductor. For example, it is well known that palladium absorbs hydrogen and certain hydrogen-bearing gases, whereas gold will absorb gases such as NO_x, SO_x, H_2S, and mercury.

The basic changes observed in the MS and MIS structures are in the current-voltage (I-V) relationship. One must note, however, that in the MIS structure the oxide layer must be less than 50 Å. This is due to the fact that electron transport occurs through the oxide layer via quantum mechanical tunneling. The MS configuration is really no different than the metal semiconductor interface that occurs in the metal oxide semiconductor gas sensor previously discussed.

As in the case of the MOS capacitive gas sensors, the most common target gases are hydrogen or hydrogen-bearing gases such as NH_3 and SiH_4. The most common metal used is palladium for both structures, whereas the common semiconductors and insulators used in the structure include TiO_2, *n*-type silicon, and SiO_2, respectively. Also note that in both cases a space charge region exists. An example of the I-V characteristics of a Pd/TiO_2 diode in air containing various concentrations of H_2 is shown in Figure 2.43.

2.6 Concluding Remarks

Electrochemical sensors form an important group of sensors whose response is the result of the interaction between electricity and chemistry. In this

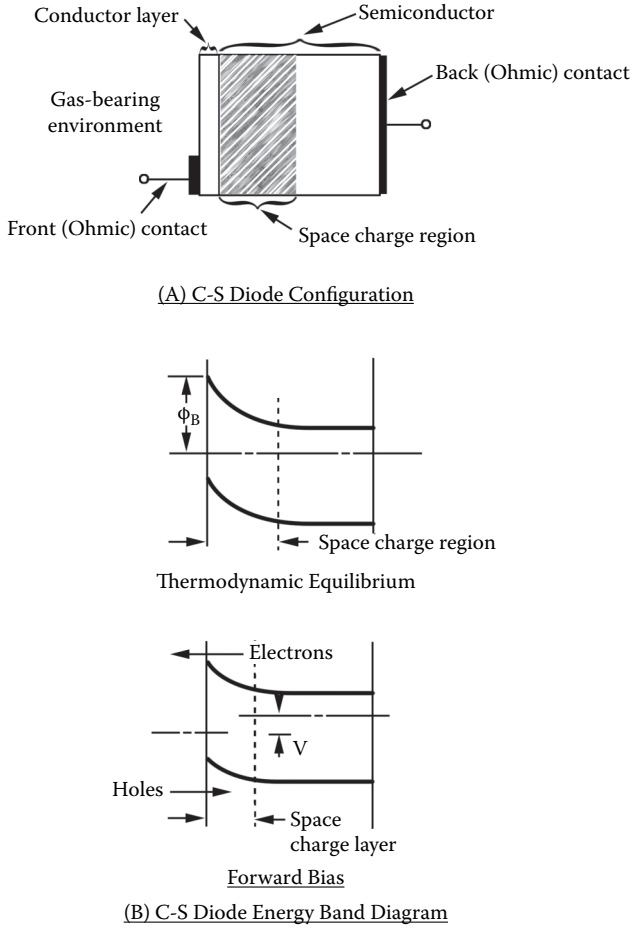
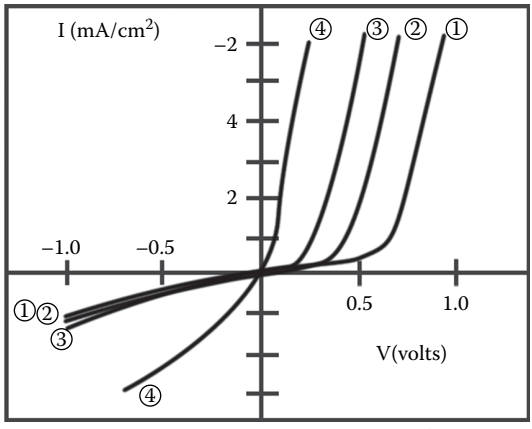


FIGURE 2.42
MS structure. (Reprinted from Schuetzle, D., and Hammerle, R., *Fundamentals and Applications of Chemical Sensors*, ACS Symposium Series [Washington, DC: American Chemical Society, 1986], 179. With permission.)

chapter we have discussed in detail the three types of electrochemical sensors: conductimetric, potentiometric, and amperometric. Moreover, the chapter also talks about semiconducting metal oxides and their properties in detail. Special emphasis is given on gas-film interactions. The chapter also discusses one of the most important semiconducting metal oxide sensors, the Taguchi sensor, as well as explains the working of chemiresistors and other solid-state electrochemical gas sensors, like the MOS capacitive gas sensor, capacitive polymer gas sensors, and the Schottky diode type gas sensor. A lot of cutting-edge research is being done in the field of electrochemical sensors,



① air, ②140 ppm H^2/air, ③ 1400 ppm H^2/air, ④ 1% H^2/air

FIGURE 2.43
I-V characteristics of a Pd/TiO_2 diode in air containing various concentrations of H_2. (From Schuetzle, D., and Hammerle, R., *Fundamentals and Applications of Chemical Sensors*, ACS Symposium Series, Washington, DC: American Chemical Society, 1986, 181. With permission.)

especially in the University of Maine, and providing all the details is beyond the scope of this book. This chapter provides all the necessary tools for the readers to have a firm understanding of electrochemical sensors, thus assisting them in further reading.

References

1. Janata, J. 1990. *Principles of chemical sensors*. New York: Plenum Press.
2. Bardeen, J. 1947. *Phys. Rev.* 71:717.
3. Brattain, W. A., and Bardeen, J. 1953. *Bell Syst. Tech. J.* 32:1.
4. Morrison, S. R. 1953. *J. Phys. Chem.* 57:860.
5. Heiland, G. 1954. *J. Phys.* 138:459. Heiland, G. 1955. *J. Phys.* 142:415.
6. Morrison, S. R. 1955. *Adv. Catal.* 7:259.
7. Weller, S. W., and Voltz, S. E. 1957. *Adv. Catal.* 9:215.
8. Bube, R. H. 1957. *J. Chem. Phys.* 24:496.
9. Heiland, G., Mollwo, E., and Stockman, F. 1959. In *Solid state physics*, ed. F. Seitz and D. Turnbell, 8. New York: Academic Press.
10. Wohlhenstein, Tb. 1960. *Adv. Catalysis* 12:180.
11. Seiyama, T., Kato, A., Fujiishi, K., and Negatani, M. 1962. *Anal. Chem.* 34:1502.
12. Seiyama, T., and Kagawa, S. 1966. *Anal. Chem.* 38:1069.
13. Setter, J. 1978. *J. Colloid Interface Sci.* 65:432 and references contained therein.

14. Taguchi, N. 1971. U.S. Patent 3,625,756, December 7. Taguchi, N. 1972. British Patent 1280809. Taguchi, N. 1972. U.S. Patent 3695848, October 3. Taguchi, N. 1972. Japanese Patent 547-38840, September. Taguchi, N. 1975. Japanese Patent Gazette 75-23317, August.
15. Jarzebski, Z. M., and Martin, J. P. 1976. *J. Electrochem. Soc.* 123:199c. Jarzebski, Z. M., and Martin, J. P. 1976. *J. Electrochem. Soc.* 123:299c.
16. Boyle, J. F., and Jones, K. A. 1977. *J. Electron. Mat.* 6:717.
17. Nitta, M., Kanefusa, S., and Harsdome, M. 1978. *J. Electrochem. Soc.* 125:1676.
18. Nitta, M., and Haradome, M. 1979. *IEEE Trans. Electron Devices* ED-26:219. Nitta, M., and Haradome, M. 1979. *IEEE Trans. Electron Devices* ED-26:247.
19. Chang, S. C. 1979. *IEEE Trans. Electron Devices* ED-26:1875.
20. Windischmann, H., and Mark, P. 1979. *J. Electrochem. Soc.* 126:627.
21. Yamazue, N., Fuchigami, T., Kishihawa, M., and Seiyama, T. 1979. *Surface Sci.* 86:335.
22. Advani, G. N., and Jordan, A. G. 1980. *J. Electron. Mat.* 9:29.
23. Schulz, M., Balm, E., and Heiland, G. 1979. *Technisches Messenten* J521–23:405.
24. Nitta, M., Ohtani, S., and Haradome, M. 1980. *J. Electron. Mat.* 9:727.
25. Pink, H., and Tischer, P. 1981. *Siemens-Forsch U. Emtwichl.-Ber. Ed.* 10:78.
26. Tischer, P., Pink, H., and Treitinger, L. 1980. *Jap. J. Appl. Phys.* 19-1:513.
27. Cooper, R. B., Advani, G. N., and Jordan, A. G. 1981. *J. Electron Mat.* 10:455.
28. Advani, G., and Nanis, L. 1981–1982. *Sensors Actuators* 2:201.
29. Clifford, P. K., and Tuma, D. T. 1982–1983. *Sensors Actuators* 3:235. Clifford, P. K., and Tuma, D. T. 1982–1983. *Sensors Actuators* 3:255.
30. Advani, G. N., Komen, Y., Hasenkopf, J., and Jordan, A. G. 1981–1982. *Sensors Actuators* 2:139.
31. Heiland, G. 1982. *Sensors Actuators* 2:343.
32. Lalauze, R., and Pijolat, C. 1984. *Sensors Actuators* 5:55.
33. Watson, J. 1984. *Sensors Actuators* 5:29.
34. Kanefusa, S., Nitta, M., and Haradome, M. 1985. *J. Electrochem. Soc.* 132:1770.
35. Bareesel, D., Gellert, W., Sarhole, W., and Scharner, P. 1984. *Sensors Actuators* 6:35.
36. Coles, G. V., and Gallagher, K. J. 1985. *Sensors Actuators* 7:89.
37. Xu, Z., Vetelino, J. F., Lec, R., and Parker, D. C. 1990. *J. Vac. Sci. Technol.* A8:3634.
38. Smith, D. J., Falconer, R. S., Wittman, E. L., and Vetelino, J. F. 1993. *Sensors and Actuators B Vol*. 13–14:264.
39. Morrison, S. R. 1982. *Sensors Actuators* 2:329.
40. Jones, A., Jones, T. A., Mann, B., and Firth, J. G. 1984. *Sensors Actuators* 5:75.
41. Bott, B., Jones, T. A., and Mann, B. 1984. *Sensors Actuators* 2:65.
42. Detector Electronics, Minneapolis, MN. Private communication.
43. Smith, D. J. 1992. Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications, MS Thesis, Electrical Engineering, University of Maine.
44. Xu, Z. 1990. Electrical Properties of Thin Tungsten Trioxide Films, MS Thesis, Electrical Engineering, University of Maine.
45. Nemoto, H. and Oda, I. 1981. *Adv. Ceram.* l, 167.
46. Madou, M. J. and Morrison, S. R. 1989. *Chemical Sensing With Solid State Devices,* Academic Press, NY.
47. Schierbaum, K.D., Weimar, V., Kowalkowski, R. and Göpel, W. 1991. *Sensors and Actuators,* B3, 205.

48. Barrett, E. P. S, Georgiades, G. C., and Sermon, P. A. 1990. *Sensors and Actuators* Bl, 116.
49. Biddle M. B. and Rickert, S. E. 1987. *Ferroe1ectrics* 76, 133.
50. Jerome Instrument Corporation, Jerome, AZ.
51. Gopal, W., Hesse, T., and Zemel, J.N. Sensors, A Comprehensive Survey. Vo. 3 Chemical and Biochemical Sensors, Part 1,VCH 1996.
52. Schuetzle. D. and Hammerle. R. 1986. *Fundamentals and Applications of Chemical Sensors.* American Chemical Society.
53. Mlsna, T.E., Cemalouic, S., Warburton, M., Hobson, S.T., Mlsna, D.A., and Petel, S.V. 2006. *Sensors and Actuators* B 116 pp. 192–201.
54. Seacoast Science, Inc. Carlsbad, CA.

Questions

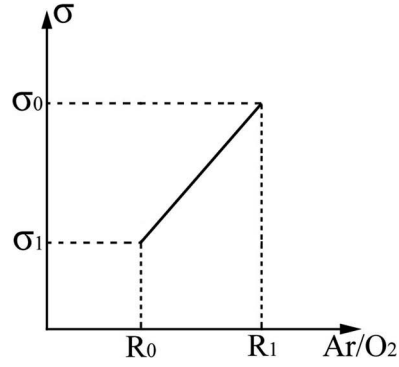
Electrochemical Sensors

1. What is an MIS contact and explain how an MIS contact can come about in an SMO sensor?
2. Why is the semiconducting metal oxide *n*-type?
3. If you were asked to deposit two metal oxide films of conductivity σ_0 and $10\sigma_0$, explain in detail how you would proceed.
4. Explain why changes in carrier mobility are not as significant as the changes in carrier concentration in a metal oxide when temperature is increased.
5. If the mobility of electrons in SnO_2 is μ at room temperature (24°C), what is the mobility at 200°C?
6. Draw the double-layer energy band structure defining each feature in detail.
7. What is more stable, a metal oxide film that is amorphous or one that is polycrystalline? Explain.
8. Derive the following relation for the surface capacitance, *C*, in a metal oxide sensor:

$$1/C^2 = \frac{2V_S}{\in A^2 q N_D}$$

where V_S = surface barrier potential, A = cross section area, ε = metal oxide dielectric constant, q = electronic charge, and N_D = oxygen vacancy concentration.

9. Experimentally several metal oxide films were deposited at various argon/oxygen (Ar/O_2) ratios and their conductivities were measured. Below a plot of σ versus argon/oxygen ratio is given.



Assuming that the relationship between σ and Ar/O_2 is linear and that electron mobility in the semiconductor is μ, develop an equation that relates the number of oxygen vacancies to the Ar/O_2 ratio. Assume each oxygen vacancy contributes one free electron.

10. What are the advantages of an impedance-based sensor over a resistance-based sensor?
11. Graphically show the variation of the conductivity as a function of temperature for an amorphous WO_3 film heated to 400°C.
12. A semiconducting metal oxide film is removed from the RF magnetic sputtering chamber and the conductivity decreases. Explain what is taking place.
13. A Taguchi sensor is exposed to hydrogen. Explain what could happen to conductivity.
14. Using chemical equations, show how oxygen ions are formed on a metal oxide surface.
15. A metal oxide film is exposed to methane. Using chemical equations, explain what happens to the resistivity of the film upon exposure to methane gas.
16. Repeat problem 15, except consider the case when the target gas is NO_x.
17. In detail describe why metal oxide films are often doped with noble metals.
18. Briefly describe the operation of the RF magnetron sputtering system.
19. Why is voltage usually applied to an interdigital electrode configuration instead of two single electrodes in a metal oxide sensor?

20. Explain why a thin gold film may not be used to detect Hg above 10 ppm.
21. In a capacitive polymer sensor, describe the two changes that can occur in the polymer upon exposure to a gas such as dimethylmethyl phosphonate (DMMP).
22. Explain in detail the operation of an array of capacitive sensors.

3

Piezoelectric Sensors

3.1 Introduction

Piezoelectric sensors use the **piezoelectric effect** as the principle of operation. The piezoelectric effect refers to the appearance of an electric field in a material when it is subjected to a stress force or a strain. Since it is a reversible effect, if an electric field or voltage is applied to a piezoelectric material, it stresses or strains. This first effect is called the **direct piezoelectric** effect, the second, the **converse piezoelectric effect**.

The following experiments will illustrate the two piezoelectric effects. A piezoelectric material with two metal electrodes, one on the top face and the other on the bottom face, is shown in Figure 3.1a. If one attempts to "stretch" the material or apply what is commonly called a tensile stress, an electric field will appear across the material as shown in Figure 3.1b. Note that for this material, the electric field polarity is positive on the top electrode. If one attempts to "squeeze" the material or apply what is commonly called a compressive stress to the same material as shown in Figure 3.1c, an electric field with a negative polarity on the top electrode will appear.

When an electric field of the polarity shown in Figure 3.1d is applied, the material will contract. Likewise, an electric field of the opposite polarity will cause the material to stretch or elongate as shown in Figure 3.1e. Finally, it is interesting to note that if the electric field or stress is doubled, the material deformation or electric field will also double. This is shown in Figure 3.1f and g. Obviously, if a very high electric field or stress is applied, the material would most likely fracture and not return to its original shape once the electric field or stress is removed. Therefore, under the application of a low to moderate electrical field or stress, a linear reversible relationship exists between electric field and stress.

As a result of these simple experiments, the following conclusions can be made:

1. Mechanical and electrical phenomena are directly coupled in piezoelectric materials. In particular, a stress produces an electric field, and hence a voltage, and an electric field produces a stress, and hence a deformation.

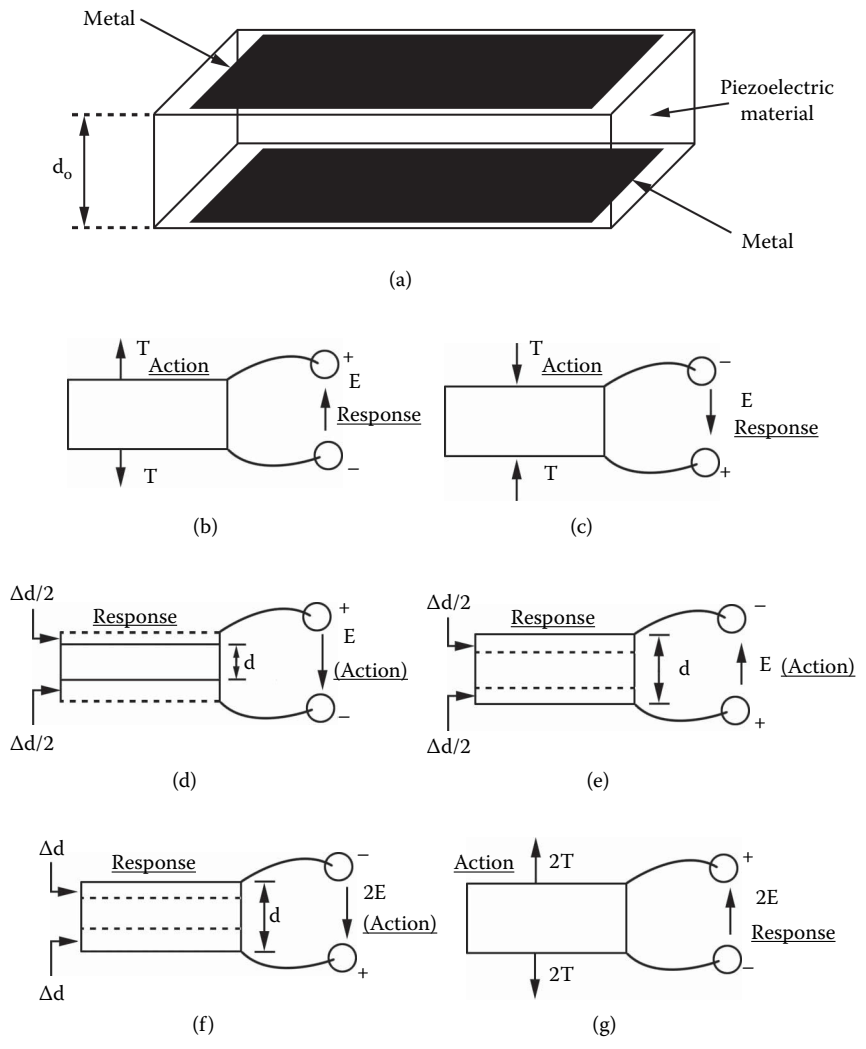


FIGURE 3.1
(a) Natural state of a sample of piezoelectric material. (b, c) Experiments showing the direct piezoelectric effect. (d, e) Experiments showing the converse piezoelectric effect. (f) Experiment showing the influence of the magnitude of the applied electric field on the magnitude of the deformation. (g) Experiment showing the influence of the magnitude of the stress on the magnitude of the electric field.

2. The direction of the electric field developed across the material due to stress depends on whether the stress is tensile or compressive.
3. A type of deformation, such as the elongation or compression, caused by the applied electric field, depends on the direction of the applied electric field.

4. The piezoelectric effect is linear; i.e., the stress and hence the deformation are proportional to the electric field, and an electric field is proportional to the applied stress.

3.2 Classification of Piezoelectric Sensors

Piezoelectric sensors can be classified as being either passive or active. In **passive piezoelectric sensors**, the measurands may be detected without the use of an external energy source and the measurand of interest is usually a force. This force may be pressure or the force resulting from acceleration or deceleration. An electric field may also serve as the measurand; however, this type of passive piezoelectric sensor is seldom used.

When the piezoelectric sensor is used as an **active sensor**, an **energy** source such as an AC electric field or voltage is used. The AC voltage sets up alternating stresses and therefore displacements in the material, which manifest themselves as acoustic waves. If these waves are then confined within the bulk or surface of finite volume in the material, they may become resonant acoustic waves. If the material with an excited resonant acoustic wave is then exposed to a measurand, acoustic wave properties such as velocity, attenuation, or resonant frequency change are then used to monitor the measurand of interest. Common active piezoelectric sensors include bulk and surface acoustic wave sensors. These sensors can be used to monitor mechanical measurands such as mass and viscosity and, if coated with a target analyte selective film, a wide range of chemical and biological target analytes that can be present in gaseous or liquid media.

3.3 Piezoelectric Materials as Intelligent or Smart Materials

Piezoelectric materials possess a very important feature. It is possible to transduct a mechanical measurand such as force, pressure, acceleration, or deceleration **directly** into an output electrical signal. This direct transduction may be considered an **inherent intelligence or smartness of the material**. In recent years the importance of piezoelectric materials or what is sometimes referred to as smart materials in the development of sensors has been growing very rapidly. The use of the smart materials allows one to simplify sensor design, and hence to eliminate usually complex hybrid signal transduction or conditioning components of the sensor.

In order to illustrate the smartness inherent in a piezoelectric material consider two pressure sensors, one a classical mechanical pressure sensor and another a piezoelectric pressure sensor. A standard approach in the classical

mechanical sensor is to use a diaphragm as a sensing element. An example of a classical mechanical sensor is shown in Figure 3.2a. This simple pressure sensor consists of a diaphragm, connecting rod, potentiometer, and DC voltage source. Under pressure, the diaphragm displaces the rod, which in turn moves the wiper of the potentiometer. The output voltage from the potentiometer is proportional to the pressure. The diaphragm is made of an elastic material like steel or aluminum.

In the piezoelectric pressure sensor the diaphragm is made of a piezoelectric material such as the polymer polyvinyl difluoride (PVDF), as shown in Figure 3.2b. In this case the pressure is directly converted into an electrical signal in PVDF through the piezoelectric effect. One can clearly see the advantage of the piezoelectric material. PVDF replaces the bulky mechanical

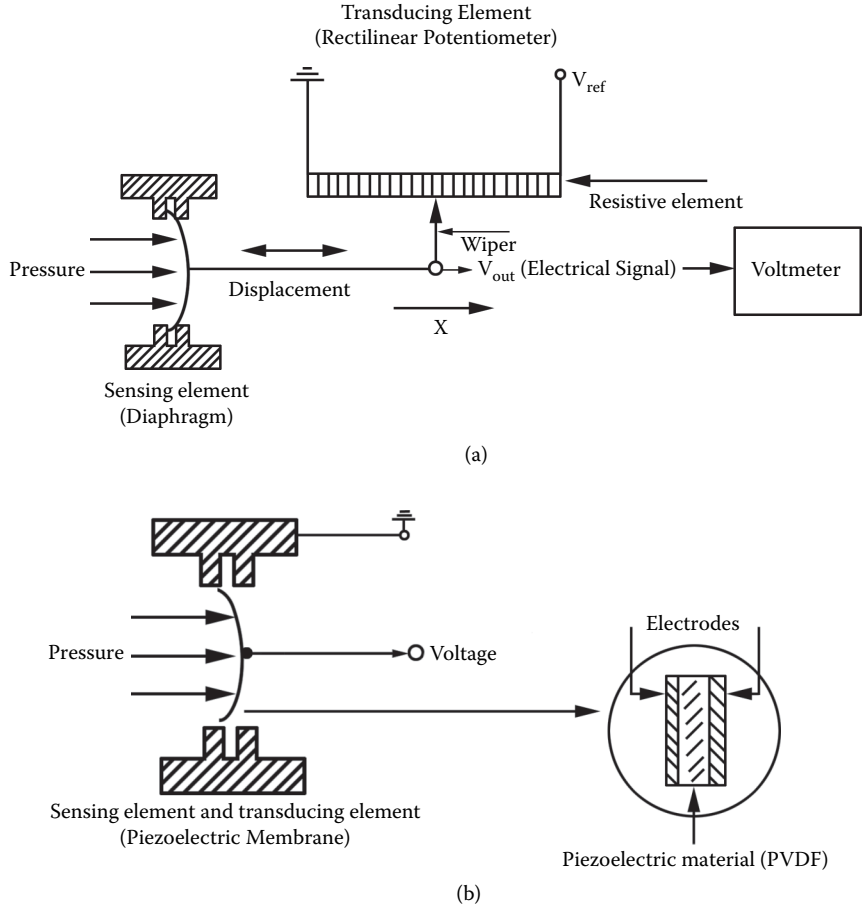


FIGURE 3.2
Two types of pressure sensors: (a) a mechanical pressure sensor and (b) a piezoelectric pressure sensor.

and electrical components of the classical mechanical sensor. As a result the piezoelectric pressure sensor is small, accurate, rugged, easy to manufacture, and inexpensive.

3.4 The Piezoelectric Effect

Only certain materials display the piezoelectric effect. The first materials known to exhibit the effect were naturally occurring, highly polar crystals such as quartz (SiO_2) and Rochelle salt. Barium titanate, lead zirconium titanate (PZT), lithium niobate ($LiNbO_3$), and PVDF include some of the synthetic materials found to possess piezoelectric properties.

In order to explain piezoelectricity, a two-dimensional hexagonal lattice consisting of oppositely charged atoms of identical charge is used as an example. The basic element of the hexagonal cell is the dipole shown in Figure 3.3a. The atoms are held together by interatomic forces, which to a first approximation are represented as springs in Figure 3.3.

Referring to Figure 3.3a, the electric field, E, from the dipole at a point W where $r >> \ell$, is given by

$$E = p/2\pi\varepsilon_o r^3 \tag{3.1}$$

where p is the electric dipole moment defined as

$$p = q\ell \tag{3.2}$$

and ε_o is the dielectric constant of vacuum.

Under the equilibrium conditions shown in Figure 3.3b and noting that $r >> \ell$, the electric field of the hexagonal lattice at point W is the sum of the contributions of the individual fields from the dipoles AB, FC, and ED, which may be expressed as follows:

$$E = \left[(q\ell)_{AB} - (2q\ell)_{FC} + (q\ell)_{ED}\right]/2\pi\varepsilon_O r^3 = 0 \tag{3.3}$$

Thus, the total electric field at W from the hexagonal lattice is zero.

Now, if a tensile stress, t, is applied to the lattice along the axis OW, the lattice will be deformed and each ion will be displaced, as shown in Figure 3.3c. The displacements, $\delta\ell_1$ and $\delta\ell_2$, will be different due to the fact that the two atoms have different mass. However, to a first approximation, one may assume that $\delta\ell_1 = \delta\ell_2 = \delta\ell$. Noting that the dipoles AB and ED are

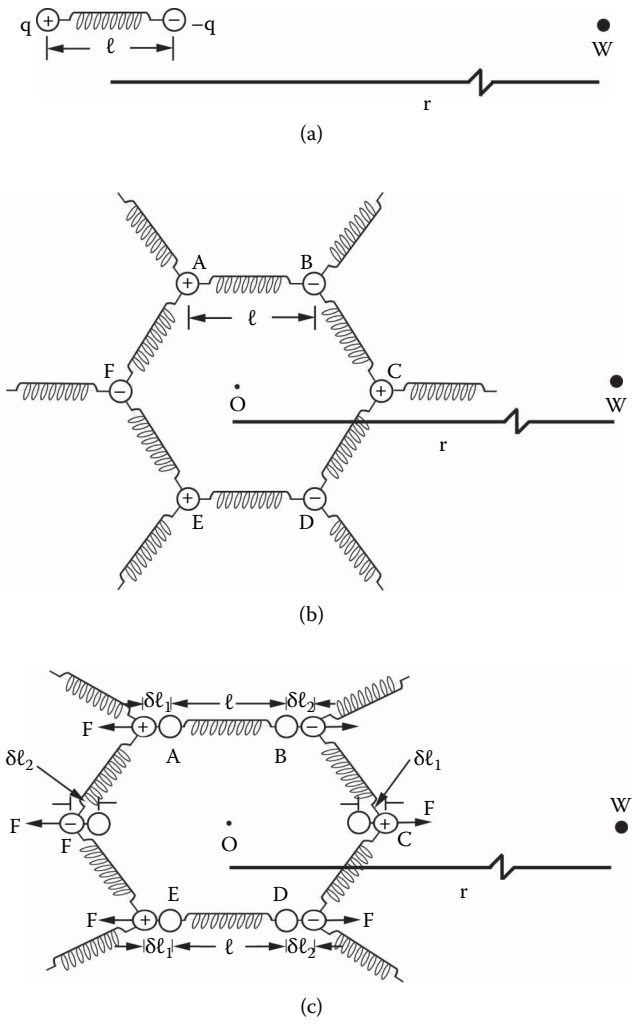


FIGURE 3.3
(a) An electric dipole. (b) A two-dimensional hexagonal lattice of oppositely charged atoms of identical charge. (c) The hexagonal lattice subjected to a symmetric tensile force, *F*.

of length $\ell + 2\delta\ell$ and that dipole FC is of length $2\ell + 2\delta\ell$, the electric field at point W becomes

$$E = p'/2\pi\varepsilon_O r^3 \tag{3.4}$$

where

$$p' = 2q\delta\ell \tag{3.5}$$

And p' may be thought of as the equivalent dipole moment of the stressed hexagonal lattice.

In contrast to no electrical field occurring for the hexagon in equilibrium, it is observed that the application of tensile stress has resulted in the appearance of an electrical field.

Noting that the strain, s, due to the tensile stress, t, is defined as

$$s = \delta\ell/\ell \tag{3.6}$$

Equation 3.5 may be written as follows:

$$p' = 2q\ell s \tag{3.7}$$

Since stress, t, is related to strain by Hooke's law, which is given as

$$s = s_c t \tag{3.8}$$

where s_c is the mechanical compliance, the equivalent dipole moment, p', can be written as follows:

$$p' = 2q\ell s_c t \tag{3.9}$$

Assuming that the two-dimensional hexagon contains N molecules per unit area, it follows that the electric polarization, P, which in this two-dimensional case is defined as the dipole moment per unit area, is given as

$$\mathrm{P} = dt \tag{3.10}$$

where

$$d = 2q\ell s_c N \tag{3.11}$$

The quantity, d, is by definition the **piezoelectric constant**, which is the proportionality factor between the mechanical stresses and induced electrical polarization. Note that in an actual three-dimensional lattice the polarization is defined as the dipole movement per unit volume.

In order to determine the units and magnitude of d, some representative values of the parameters in Equation 3.11 will be used. Noting in three dimensions that $N = 5 \times 10^{28}/\mathrm{m}^3$, $q = 1.6 \times 10^{-19}$ coulombs (C), $s_c = 12.7 \times 10^{-11}$ m^2/N, and $\ell = 1.5 \times 10^{-10}\mathrm{m}$, it is found that $d = 1.52 \times 10^{-12}$ C/N. It is interesting to point out that this result is comparable to an actual piezoelectric constant of quartz, $d_{11} = 2.41 \times 10^{-12}$ C/N.

In a similar fashion it may be shown that when an external electric field, E, is applied along OW, the lattice will undergo a deformation that is a linear function of the electric field. The resulting proportionality constant between the electric field and the resultant strain is also a piezoelectric constant. It is interesting to examine the structural properties of a hexagonal lattice to see if there is a correlation between piezoelectricity and structure. Assuming that the origin of the lattice is at an atomic site (for example, site A), one notes that this lattice does not possess a center of symmetry. In particular, for each ion at a position (xy) there is no similar ion located at ($-x,-y$). If there were a like atom at ($-x,-y$), the piezoelectric effect would not be present.

It is interesting to note that if stress or electric field is applied along a direction different than OW, the piezoelectric constant may be different. It is therefore obvious that in an actual crystal the piezoelectric constant will have different values depending on the direction of the applied stress or electric field.

3.5 The Properties of Nonpiezoelectric and Piezoelectric Materials

A knowledge of the electrical and mechanical properties of piezoelectric materials is important in the design and operation of a piezoelectric sensor. In general, these properties are described by the **constitutive equations**, which relate mechanical forces such as stress and strain to electrical properties such as electric field and electric displacement.

In order to deduce the constitutive equations for a piezoelectric material, both mechanical and electrical properties of nonpiezoelectric materials must be discussed.

3.5.1 Mechanical Properties of Nonpiezoelectric Materials

A nonpiezoelectric material subjected to a stress responds by changing its shape, and there is no electric field accompanying the stress, as shown in Figure 3.4. Similarly, if the material is subjected to a strain, a stress appears. The general relationship in a three-dimensional solid cube between stress and strain can be written as follows:

$$[\mathrm{T}] = [c] \cdot [s] \tag{3.12}$$

where

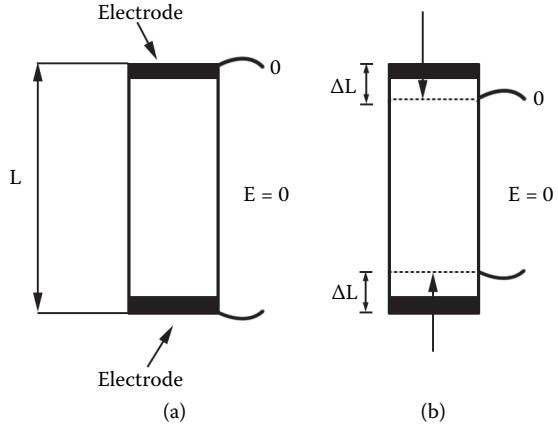


FIGURE 3.4
Elastic nonpiezoelectric material (a) under stress-free conditions and (b) under stress.

$$[T] = \begin{bmatrix} T_{xx} \\ T_{yy} \\ T_{zz} \\ T_{yz} \\ T_{zx} \\ T_{xy} \end{bmatrix} \tag{3.13}$$

$$[c] = \begin{bmatrix} c_{11} & - & - & - & c_{16} \\ | & & & & | \\ | & & & & | \\ | & & & & | \\ c_{61} & - & - & - & c_{66} \end{bmatrix} \tag{3.14}$$

$$[s] = \begin{bmatrix} s_{xx} \\ s_{yy} \\ s_{zz} \\ s_{yz} \\ s_{zx} \\ s_{xy} \end{bmatrix} \tag{3.15}$$

In order to better understand the elements in the [T] matrix, refer to Figure 3.5, which represents all types of forces that may be applied to a small cube of the material.

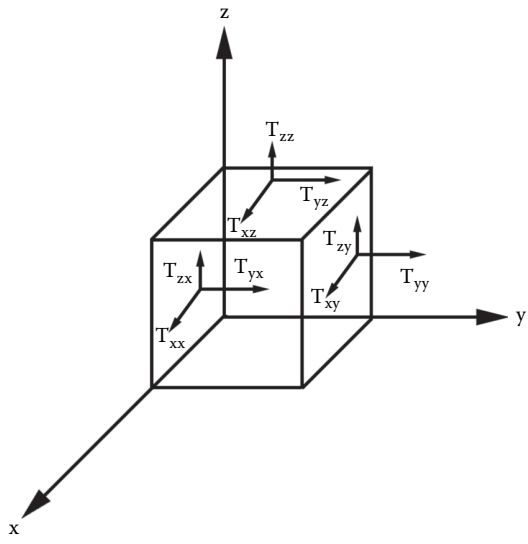


FIGURE 3.5
Various forces that may be applied to a small cube of the material.

The first subscript in the individual stresses indicates the direction of the applied force, whereas the second subscript denotes the plane upon which the force is acting. For example, the term T_{yz} corresponds to a force acting in the y direction on the z plane. The forces with the same subscripts, e.g., T_{zz}, correspond to compressional forces, whereas the forces with different subscripts, e.g., T_{yz}, correspond to shearing forces. Also note that in Figure 3.5 there are a total of nine forces; however, only six forces appear in Equation 3.13. This is due to the fact that the angular acceleration of the cube must be zero, resulting in the following relationship:

$$T_{ij} = T_{ji} \tag{3.16}$$

where

$$i \neq j$$

and

$$i, j = x, y, z$$

Similar arguments also apply for the individual strains; hence

$$s_{ij} = s_{ji} \tag{3.17}$$

where i and j are defined in Equation 3.16.

The matrix [c] is defined as the elastic constant matrix and in general may have 36 terms. Equation 3.12 may also be written in the following form:

$$[s] = [s_c] \cdot [\mathrm{T}] \tag{3.18}$$

where

$$[s_c] = [c]^{-1} \tag{3.19}$$

The matrix [s_c] is defined as the compliance matrix and is the same rank order as the elastic constant matrix.

In the previous discussion relating to the planar hexagonal lattice, the stresses, strains, mechanical compliance, and piezoelectric constant had only one value. This is due to the fact that the force was applied in only one direction. Because the distance between individual atoms in an actual crystalline material is dependent upon crystallographic direction, one can expect that a crystalline material will have several different constants, which is reflected in the multivalued elastic and compliance matrices. Fortunately, due to the symmetry inherent in an actual crystal, the number of independent elastic or compliance constants is quite small. In fact, in the case of an isotropic solid, there are only two independent constants, whereas for a cubic crystal, only three independent constants exist. The respective elastic constant matrices for an isotropic solid and a cubic crystal are given as follows:

$$[c] = \begin{bmatrix} (c_{12}+2c_{44}) & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & (c_{12}+2c_{44}) & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & (c_{12}+2c_{44}) & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{bmatrix} \tag{3.20}$$

and

$$[c] = \begin{bmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{bmatrix} \tag{3.21}$$

3.5.2 Electrical Properties of Nonpiezoelectric Materials

In order to describe the effect an electric field has on a nonpiezoelectric material, consider the parallel plate capacitor shown in Figure 3.6. The applied voltage or electric field causes a displacement of the electrons around the atomic nuclei, resulting in each atom becoming a small dipole. The material is then said to be polarized and has a dipole moment per unit volume, or what is defined as the polarization, [*P*]. The magnitude of the polarization relates to the individual atoms comprising the material and the strength of the applied electric field. As a consequence of the polarization, one capacitor plate acquires a positive charge and the other plate a negative charge, as shown in Figure 3.6. The amount of charge per unit area is characterized by the **electric displacement** or **electric flux density**, [*D*], and is given as follows:

$$[D] = \varepsilon_o [\mathrm{E}] + [P] \tag{3.22}$$

where [*P*] = polarization of the nonpiezoelectric material, and ε_o = permittivity of a vacuum.

Since [*P*] is proportional to the electric field, [*E*], and the ability of the atoms to deform into dipoles, it may be represented as follows:

$$[\mathrm{P}] = \varepsilon_o [\eta] \cdot [\mathrm{E}] \tag{3.23}$$

where [η] is a 3 × 3 matrix defined as the **dielectric susceptibility**.

The total electric displacement, [*D*], can now be expressed as

$$[D] = [\varepsilon] \cdot [E] \tag{3.24}$$

$$[\varepsilon] = \varepsilon_0 \{[I] + [\eta]\}$$

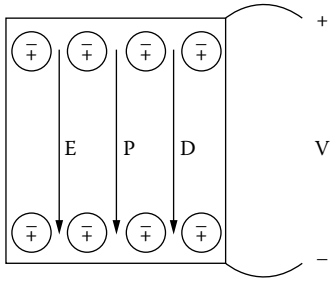


FIGURE 3.6
Parallel plate capacitor with a nonpiezoelectric material between the electrodes.

where [*I*] is the identity matrix, defined as

$$[I] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.25}$$

Equation 3.24 is called the **dielectric constitutive equation**, and [ε] is the **dielectric permittivity**.

It is important to note that [*E*], [*D*], and [*P*] in general are 1 × 3 matrices that may have *x*, *y*, or *z* components, depending upon the reference coordinate system defined in the material. It then follows that the dielectric permittivity, [ε], is a 3 × 3 matrix, which can be represented as follows:

$$[\varepsilon] = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{bmatrix} \tag{3.26}$$

Normally the dielectric permittivity matrix contains one or two independent terms and the off-diagonal terms are usually equal to zero.

In addition to polarization there is a very small change in the shape of a material due to the elongation of the atoms caused by an applied electric field. This change in shape is very small in comparison to changes due to an applied electric field in piezoelectric materials. This phenomenon is called electrostriction and is present in any nonpiezoelectric material. Also, it is not reversible, which means that a change in a material's shape due to a stress will not create an electric field.

3.5.3 Electrical and Mechanical Properties of Piezoelectric Materials

The electrical and mechanical properties of piezoelectric materials can be described by two piezoelectric constitutive equations. In order to obtain these equations, first consider the experiment shown in Figure 3.1c. A piezoelectric material with two metal electrodes, one on the top face and the second on the bottom face, is subjected to a compressive stress. As a result, the material will initially be deformed according to Hooke's law, and then, as a result of the deformation, an electric field, [*E*], appears, which then causes a polarization, [*P*], and an electric displacement, [*D*], to occur in the material. Mathematically, this may be represented by the following piezoelectric constitutive equation:

$$[\mathrm{T}] = [c] \cdot [s] - [e] \cdot [E] \tag{3.27}$$

where [*e*] is defined as the piezoelectric constant matrix. Examination of Equation 3.27 reveals that the [*e*] matrix is a 3 × 6 matrix defined as follows:

$$[e] = \begin{bmatrix} e_{11} & e_{21} & e_{31} \\ | & & | \\ | & & | \\ | & & | \\ | & & | \\ e_{16} & - & e_{36} \end{bmatrix} \tag{3.28}$$

Note that if the material is nonpiezoelectric, Equation 3.27 reduces to Equation 3.12, the common representation of Hooke's law. In order to obtain the second constitutive equation, consider the experiment shown in Figure 3.1d, where the piezoelectric material is subjected to an applied electric field. As a result, the material will acquire a polarization, [*P*], or electric displacement, [*D*], and a strain, [*S*], or a stress, [*T*]. The second constitutive equation is usually written in the following form:

$$[D] = [e]^{\mathrm{T}}[s] + [\varepsilon][E] \tag{3.29}$$

where the superscript T denotes the matrix transpose operation. Again, note that if the material is nonpiezoelectric, Equation 3.29 reduces to the relationship given in Equation 3.24.

3.5.4 The Piezoelectric Coupling Coefficient

The piezoelectric coupling coefficient, *k*, is one of the most important properties of a piezoelectric material. It characterizes the piezoelectric material from the energy conversion point of view, whereas the piezoelectric constants characterize a material in terms of the relation between the field quantities, such as stress, strain, electric field, and electric displacement.

In order to describe the piezoelectric coupling coefficient from an energy balance point of view, refer to Figure 3.7. In Figure 3.7 the mechanical energy, W_{m}, is supplied to a piezoelectric material through an applied stress. This mechanical energy, W_{m}, is stored as a mechanical energy via the deformation of the material, W_{mm}, and as electric energy, W_{me}, due to the creation of an electric field as shown in Figure 3.7. A similar energy balance consideration can be obtained in the case where an applied electric energy, W_{e}, is divided into polarization or electric displacement energy, W_{ee}, and mechanical energy, W_{em}.

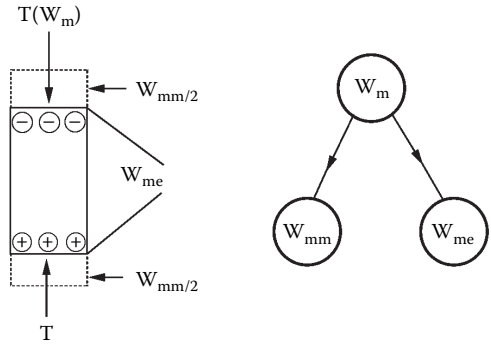


FIGURE 3.7
Energy balance in piezoelectric materials upon application of a compressive stress, *T*.

The piezoelectric coupling coefficient, k, is a measure of how much of the mechanical energy can be converted into electrical energy in the first experiment, and how much electric energy can be converted to mechanical energy in the second experiment. Mathematically, this may be expressed as

$$k_1 = \frac{(W_{me})^{1/2}}{W_m} \tag{3.30}$$

and

$$k_2 = \frac{(W_{em})^{1/2}}{W_e} \tag{3.31}$$

It can be shown [1] that $k_1 = k_2 = k$, and k is given by

$$k = e/(s\varepsilon)^{0.5} \tag{3.32}$$

where e, s, and ε represent single-valued equivalent piezoelectric, compliance, and dielectric constants in the crystallographic direction for which the stress or electric field is applied. These single-valued constants are linear combinations of the independent constants. The value of the piezoelectric coupling coefficients for selected piezoelectric materials is given in Table 3.1.

The piezoelectric ceramic, PZT, has the highest piezoelectric coupling coefficients and is the strongest piezoelectric material. It has a value of $k = 68\%$, whereas for quartz, $k \approx 1\%$.

TABLE 3.1

Material Constants for Piezoelectric Materials

Sensor Material	Dielectric Constant $\varepsilon/\varepsilon_o$ Unitless	Compliance s	Elastic Constant c	Piezoelectric Constant d	Piezoelectric Constant e	Piezoelectric Coupling Coefficient k
		$10^{-12}\frac{m^2}{N}$	$\left[10^{10}\frac{N}{m^2}\right]$	$\left[10^{-12}\frac{C}{N}\right]$	$\left[\frac{C}{m^2}\right]$	[%]
Quartz	4.5	12.77	8.67	2.31	0.17	1.05
$LiNbO_3$	85.2	50.21	24.2	6.0	1.33	9.7
PZT	1,320	15.5	11.5	289	15.1	68

Source: Auld, B. A., *Acoustic Fields and Waves in Solids*, Vol. II, Malabar, FL: Robert E. Krieger, 1990 With permission.

3.6 Piezoelectric Stress/Pressure Sensor

The most common passive piezoelectric sensor is one in which stress or pressure is the measurand of interest. The design steps necessary to develop this are as follows:

1. Determination of the magnitude of the sensing element response in the measurand range for different piezoelectric materials
2. Development of an equivalent electrical circuit for the sensing element
3. Determination of the time response of the sensing element
4. Development of a signal conditioning system to interface the sensing element response to the observer
5. Choice of the piezoelectric material for the sensing element
6. Sensor construction

The stress/pressure sensor used will be designed to respond to static stresses and pressures ranging from 10^2 to 10^5 N/m^2. The piezoelectric sensing element will be a square piezoelectric plate of dimensions 10 × 10 × 1 mm, with electrodes on the top and bottom faces (see Figure 3.8). The above stress range corresponds approximately to the force generated from a gram to a 1 kg weight on the surface plate.

3.6.1 Determination of the Magnitude of the Sensing Element Response in the Measurand Range for Different Piezoelectric Materials

In Section 3.5 it was shown that a stress will produce a strain or deformation, an electric field, and an electric displacement. In order to determine

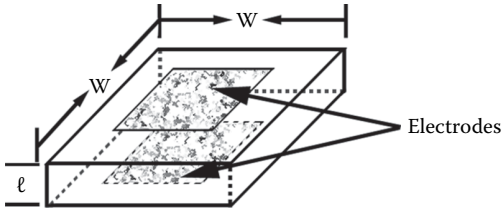


FIGURE 3.8
Square piezoelectric plate sensing elements.

the magnitude of each sensing response, refer to Equations 3.27 and 3.29. The strain associated with an applied stress, *T*, normal to the plate, can be obtained from Equation 3.18. Noting the definition of strain in terms of deformation given by Equation 3.6, one obtains the following relationship for the deformation:

$$\delta\ell = s_c \ell \mathrm{T} \tag{3.33}$$

where s_c = equivalent mechanical compliance normal to the plate, and ℓ = plate thickness.

Note that even though the mechanical compliance is in general a 6 × 6 matrix, in a given crystallographic direction (in this case, normal to the plate) it has only a single value. This so-called equivalent mechanical compliance is a linear combination of independent mechanical compliance constants.

The applied stress also produces an electric field or voltage across the piezoelectric plate. In this case Equation 3.27 reduces to the following:

$$T = -eE \tag{3.34}$$

Solving Equation 3.34 for *E* one obtains

$$E = -e^{-1}\, T \tag{3.35}$$

Since the applied stress is normal to the plate, and noting that the induced voltage may be expressed as

$$V = E\ell \tag{3.36}$$

one obtains

$$V = -\ell e^{-1}\mathrm{T} \tag{3.37}$$

TABLE 3.2

Operating Range Associated with Different Sensing Mechanisms for Static Stresses Ranging from 10^2 to 10^5 N/M^2 (Results Given for X-Cut Quartz, Z-Cut $LiNbO_3$, and Z-Cut PZT)

Sensor Material	Deformation (angstroms)		Voltage (volts)		Charge (coulombs)	
	Min	Max	Min	Max	Min	Max
SiO_2 (X-cut)	.1277	127.7	-5.8×10^{-3}	–5.8	2.3×10^{-14}	2.31×10^{-11}
$LiNbO_3$ (X-cut)	.502	502	-2.23×10^{-3}	–2.23	6×10^{-14}	6×10^{-11}
PZT (Z-cut)	.155	155	-26.1×10^{-3}	–26.1	2.89×10^{-12}	2.89×10^{-9}

As in the case of the mechanical compliance, the term e represents an equivalent piezoelectric constant.

Finally, the applied stress also produces an electric displacement that is equal to the surface charge density on the piezoelectric plates. Referring now to Equations 3.18 and 3.29, one obtains

$$[D] = [e]^T [S_c] [T] \tag{3.38}$$

Noting that the surface charge, q, may be represented as

$$q = DA \tag{3.39}$$

where A = surface area of the piezoelectric plate, one obtains

$$q = Ae'S_c T \tag{3.40}$$

The term e' represents another equivalent piezoelectric constant.

Assuming that the piezoelectric plates are X-cut quartz, Z-cut $LiNbO_3$, and Z-cut PZT, respectively, the corresponding minimum and maximum values for the deformation, voltage, and charge can be obtained. These values are summarized in Table 3.2. It may be seen that the only significant response that can easily be measured is the voltage.

3.6.2 Equivalent Circuit for the Sensing Element

In order to interface the sensing element to the signal conditioning and processing segment of the sensor, it is necessary to represent the sensing element in terms of an equivalent electrical circuit. The piezoelectric plate is essentially a capacitor in which the capacitance may be represented as follows:

$$C_o = \varepsilon A/\ell. \tag{3.41}$$

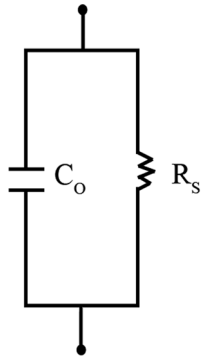


FIGURE 3.9
Equivalent circuit for the sensing element of a stress/pressure sensor.

TABLE 3.3

Values for the Circuit Elements for the Equivalent Circuit of the Piezoelectric Plate

Sensor Material	C_o **(picofarads)**	R_s **(ohms)**
SiO_2 (X-cut)	3.98	2×10^{12}
$LiNbO_3$ (Z-cut)	75.4	5×10^3
PZT (Z-cut)	1,250	5×10^9

Although piezoelectric materials are in general good dielectrics, they do have a finite conductivity, which introduces a resistance, R_s, that is parallel to C_o. This resistance may be expressed as

$$R_s = \ell/\sigma A\,, \tag{3.42}$$

where σ = piezoelectric plate conductivity. The resulting equivalent circuit for the piezoelectric plate is shown in Figure 3.9.

The values for C_o and R_s for various orientations in several piezoelectric materials are summarized in Table 3.3.

The resistance for $LiNbO_3$ was measured at 400°C, and for the other crystals, the resistance was measured at 20°C.

3.6.3 Time Response for the Stress/Pressure Sensor

The response of the sensing element will be examined for a static force, F. Using Equation 3.40 and noting that stress is force (F) per surface area of the piezoelectric plate, one obtains the following expression for the charge per unit area on the piezoelectric plate:

$$q = e'ScF. \tag{3.43}$$

The charge generated on the surface of the piezoelectric plate causes a voltage to appear across the plate. Noting the definition of capacitance, namely,

$$C_0 = q/V, \tag{3.44}$$

and using Equation 3.43, the voltage on the plate, V_0, may be expressed as follows:

$$V_0 = e'ScF/C_0. \tag{3.45}$$

Since the charge leaks through the resistor, R_S, it may be shown that the voltage on the piezoelectric plate decays exponentially with time (t) as follows:

$$V_0(t) = V_0 e^{-t/\tau_0}, \tag{3.46}$$

where

$$V_0 = e'ScF/C_0$$

and

$$\tau_0 = R_s C_0$$

The term V_0 may be written as follows:

$$V_0 = KF, \tag{3.47}$$

where

$$K = e'Sc/C_0. \tag{3.48}$$

The parameter K depends only upon properties of the piezoelectric plate and is called the sensor parameter. The term τ_0 is the time it takes for the voltage to decay to 1/e of its original value and is called the time constant of the circuit. The rate of decrease of the voltage, $V_0(t)$, depends directly on τ_0, which is also a function of the material properties of the piezoelectric plate.

The approximate time responses of sensing elements made from SiO_2, $LiNbO_3$, and PZT are shown in Figure 3.10. Depending on the piezoelectric material, there is a wide range of decay rates. For example, for $LiNbO_3$, $\tau = 38$ μs, while for PZT and SiO_2 the values for τ are 1.15 s and 7.96×10^3 s (2.2 h), respectively. Therefore, SiO_2 (quartz) appears to be a logical choice for a static force sensor due to its very slow decay rate.

In Table 3.4, values for the time constant, τ_0, and the sensor parameter, K, are summarized for orientations in quartz, $LiNbO_3$, and PZT.

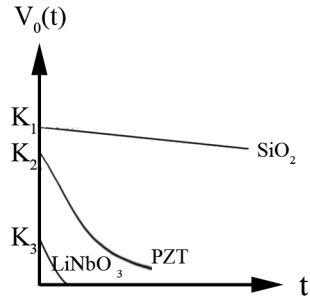


FIGURE 3.10
The approximate time response of different sensing elements to a static or step force. K_1, K_2, and K_3 are sensor parameters for SiO_2, PZT, and $LiNbO_3$.

TABLE 3.4

The Time Constant, τ_O and the Sensor Parameter, *K*, for Various Piezoelectric Materials

Sensor Material	Time Constant τ_0 (s)	Sensor Parameter *K* (V/N)
SiO_2 (X-cut)	7.962×10^3 (2.2 h)	5.82×10^{-1}
$LiNbO_3$ (Z-cut)	3.82×10^{-5}	7.962×10^{-2}
PZT (Z-cut)	1.15	5.142×10^{-1}

3.6.4 Signal Conditioning System to Interface the Sensing Element Response to the Observer

Examining Tables 3.2 and 3.3, one can clearly see that piezoelectric pressure sensing elements are inherently high-output resistive devices that generate small charges (on the order of picocoulombs [pC]) or currents (on the order of picoamperes [pA]), and small voltages (on the order of millivolts [mV]). Therefore, amplifiers may need to be used to condition the output signal. In the present analysis two cases will be considered. In the first case the sensing element output is directly connected to an output display. In the second case a signal conditioning circuit is placed between the sensing element output and the display device.

3.6.4.1 Sensing Element Output Interfaced Directly to the Display Device

The pressure sensing element is connected via a 3 ft long coaxial 50 Ω cable directly to a display system consisting of a Tektronix Model 7603 oscilloscope. The electric equivalent circuit, which includes the sensing elements, a connecting cable, and a display device (oscilloscope), is shown in Figure 3.11. The total capacitance, C_T, and resistance, R_T, of the circuit may be expressed as

$$C_T = C_O + C_{Cable} + C_{OSC} \tag{3.49}$$

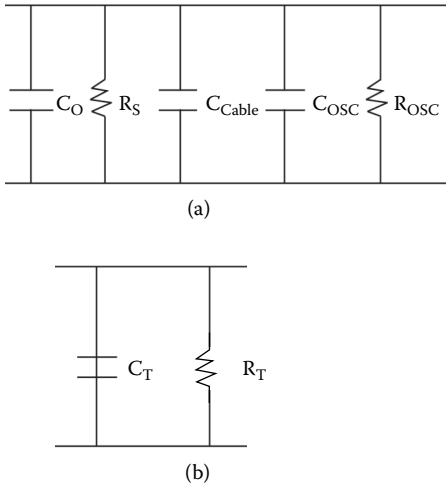


FIGURE 3.11
(a) Equivalent electrical circuit for a piezoelectric pressure sensing element connected to a display system consisting of an oscilloscope with a coaxial cable and (b) the resultant equivalent circuit.

and

$$R_T = \frac{R_S \cdot R_{OSC}}{R_S + R_{OSC}}, \tag{3.50}$$

where C_{Cable} = cable capacitance, C_{OSC} = oscilloscope capacitance, and R_{OSC} = oscilloscope resistance.

The time response is of the same form as the pressure sensing element response (Equation 3.46) and may be represented as follows:

$$V_0'(t) = V_0' e^{-t/\tau_0'}, \tag{3.51}$$

where

$$V_0' = e'ScF/c_T = K'F$$

and

$$\tau_0' = R_T C_T.$$

Typical values for the C_T, R_T, τ_0', and K' parameters for common piezoelectric materials are given in Table 3.5.

Comparing Tables 3.4 and 3.5, it can be seen that the display circuit elements significantly affect the relevant sensor parameters. In particular, for quartz, which initially appeared to have promise as the piezoelectric material for the

TABLE 3.5

The Sensor Parameters for the Piezoelectric Pressure Sensor That Includes the Sensing Element, the Connecting Cable, and a Display Device

Sensor Material	Total Capacitance C_T (pF)	Total Resistance R_T (Ω)	Time Constant τ_O' (s)	Sensor Parameter K' (V/N)
SiO_2 (X-cut)	174	10^6	1.742×10^{-4}	1.332×10^{-2}
$LiNbO_3$ (Z-cut)	254	9.92×10^5	2.452×10^{-4}	2.442×10^{-2}
PZT (Z-cut)	1420	9.12×10^5	1.22×10^{-3}	2.192×10^{-2}

sensing element, the time constant has decreased by about six orders of magnitude and the sensor parameter has decreased by a factor of 44. Degradation of sensor parameters was also observed for $LiNbO_3$ and PZT. In conclusion, one can easily see that the performance of the sensor appears to be more dependent on the coaxial cable and display device than on the sensing element output.

3.6.4.2 Sensing Element Interfaced to a Charge Amplifier and a Display Device

In this case a charge amplifier is placed between the coaxial line and the display device. The equivalent circuit for this sensor is shown in Figure 3.12. A charge amplifier is a circuit whose input impedance is a capacitance that provides a very high value of impedance at low frequencies. The function of the charge amplifier is to produce a voltage that is proportional to the charge on the piezoelectric plate while at the same time providing a low-output impedance. Hence, it is a charge-to-voltage converter, or what may be called an impedance converter. The charge amplifier, with its low-output impedance, essentially separates the display device from the sensing element.

The amplifier has a significant effect on the sensor performance. In particular, for an operational amplifier (op-amp) of gain, A, the output voltage from the sensing element, V_O, is equal to AV_O. Inherently, op-amps have high-input impedance, low-output impedance, and high-voltage gain. Note that the positive input of the op amp is grounded.

It may be shown using circuit theory that the output voltage, $V_0''(t)$, is of the same form as Equations 3.46 and 3.51; namely,

$$V_0''(t) = V_0'' \, e^{-t/\tau_0''}, \tag{3.52}$$

where

$$V_0'' = e'ScF/C_F = K''F$$

and

$$\tau_0'' = R_T C_F.$$

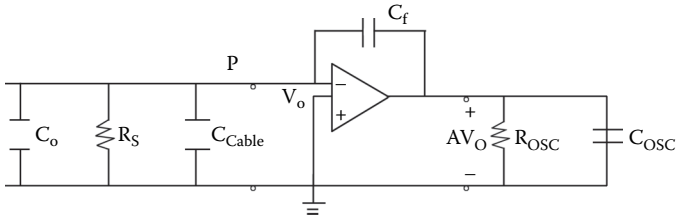


FIGURE 3.12
A piezoelectric pressure sensing element interfaced to a charge amplifier and a display device. C_f is the feedback capacitance, and the other circuit elements have been defined before.

TABLE 3.6

The Sensor Parameters for the Piezoelectric Pressure Sensor That Includes the Connecting Cable, a Charge Amplifier, and a Display Device

Sensor Material	Time Constant τ_0'' (s)	Sensor Parameter K'' (V/N)
SiO_2 (X-cut)	10^8	0.23
$LiNbO_3$ (Z-cut)	$5{\cdot}10^{-2}$	0.6
PZT (Z-cut)	0.5	28.9

In the derivation of Equation 3.52, the gain is assumed to be very large (typically 10^5 to 10^9).

The values of K'' and $\tau_{0''}$ obtained for $A = 10^6$ and $C_f = 10$ pF are given in Table 3.6.

It is interesting to compare the sensor parameters with the corresponding parameters given in Tables 3.4 and 3.5. It is seen that the most dramatic change occurs for the sensor parameter. In particular, for PZT it has increased three orders of magnitude. The charge amplifier has essentially eliminated the effect of the connecting cable and display device on the sensor performance.

3.6.5 The Choice of the Piezoelectric Material for the Sensing Element

The choice of the piezoelectric material to be used for the sensing element is dependent upon the specific application of the stress/pressure sensor. The applications are diverse and cover the range of measuring very weak stress, or low pressures, to measuring stress or pressure under adverse environmental conditions. Some typical examples of applications will be given, and appropriate piezoelectric materials to be used for the sensing element will be suggested.

In applications requiring high sensitivity, and therefore the need to measure very small stresses or pressures, the sensor parameter is the key factor. Examining Tables 3.5 and 3.6, it is obvious that the sensor should have

an amplifier stage and that the most desirable piezoelectric material for the sensing element is PZT followed by $LiNbO_3$ and SiO_2.

In the measurement of static forces the key factor is the time constant. In this case, it can be seen from Table 3.6 that the sensor must have an amplifier stage and the piezoelectric material of choice is quartz. PZT may also be used, particularly if one is concerned with sensitivity. However, appropriate signal processing needs to be performed when PZT is used since it has a rather small time constant.

In measurement of large stresses or pressure, the piezoelectric material must be able to withstand significant forces without fracturing or exceeding its elastic limit. Since those materials with high elastic constants are considered to be the strongest, the best choice for the piezoelectric material is SiO_2 followed by $LiNbO_3$ and PZT.

Often applications require that a sensor be used in an environment where the temperature varies over a wide range. The appropriate choice of material is then strongly dependent upon the temperature range of interest. For high-temperature applications, one needs to be concerned as to whether the material of interest maintains its piezoelectric properties. Materials such as quartz and PZT undergo structural phase transitions above 500 and 150°C, respectively. Therefore, these materials should not be used near or above those temperatures. On the other hand, $LiNbO_3$ can be used up to 1,000°C, since it still retains its piezoelectric properties. At low temperatures, the critical feature is the brittleness of the material. Since this case is similar to applications relating to large stress or pressure, the material of choice is SiO_2 followed by $LiNbO_3$ and PZT. Finally, in applications relating to high sensitivity, temperature can play a critical role. For example, in measuring very low stresses, a change of a few degrees in temperature can cause piezoelectric material parameters to change, and hence affect the sensing element response. This will result in a significant error in determining low stress values. In order to eliminate this problem, one should try to use a material that is both sensitive to stress and insensitive to temperature variations. In this case, quartz may be the material of choice over PZT and $LiNbO_3$. This is due to the fact that quartz, although it is not as stress or pressure sensitive as PZT or $LiNbO_3$, is relatively insensitive to temperature variations. PZT and $LiNbO_3$, on the other hand, are very sensitive to temperature changes.

In sensors that are expected to function over a long period of time, it is important that the material and hence sensor properties have minimal variations with time. Materials and sensors exhibiting this behavior are said to have excellent aging characteristics. For long-life sensors the material of choice is SiO_2 followed by $LiNbO_3$ and PZT.

Two other factors that can affect the choice of sensor material are the technological process used to make the sensing element and the cost of the sensing element. In terms of brittleness, PZT is the most brittle, followed by $LiNbO_3$ and SiO_2. Relative to size and shape, SiO_2 and $LiNbO_3$ have limits

due to the fact that they are single crystals, whereas PZT is a ceramic and hence can be made into a variety of sizes and shapes. Finally, in terms of cost, the least expensive is PZT followed by SiO_2 and $LiNbO_3$.

In many applications, the final choice of the piezoelectric material is obvious. For example, if the ultimate objective is to reduce cost, PZT is the choice. In other applications, the final choice is not so obvious and trade-offs need to be considered. An example would be a sensor for low pressure that needs to operate over a specified temperature range. PZT gives the necessary sensitivity, but SiO_2 is very temperature stable, therefore resulting in a choice that is not so obvious.

3.6.6 Sensor Construction

There are many possible physical configurations for the stress/pressure sensor. These configurations depend upon the source of the stress or pressure. The two types of sources that will be examined include mechanical force and gas.

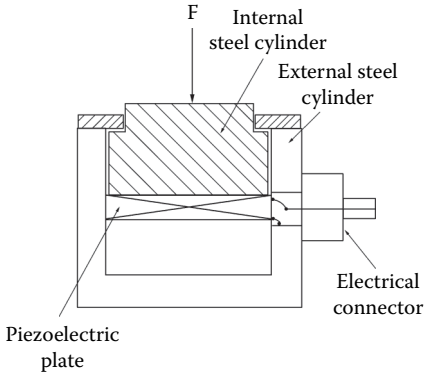
3.6.6.1 Piezoelectric Mechanical Force Sensor

In a force sensor a piezoelectric sensor (plate) is placed in an enclosure (cell) in order to form a mechanically rigid cell. A basic piezoelectric force sensor cell is illustrated in Figure 3.13a. The cell contains the piezoelectric plate and two hardened steel cylinders. The piezoelectric plate is mechanically connected to the internal cylinder, which is opened to a force and free to move. A change in force causes a mechanical change in the piezoelectric plate, which results in a generation of electric charge or voltage. The charge or voltage is transferred out of the cell using a shielded connector. A commercial model of a quartz force sensor produced by Kistler [2] is shown in Figure 3.13b.

3.6.6.2 Piezoelectric Pressure Sensor

A basic piezoelectric pressure sensor cell is illustrated in Figure 3.14a. The piezoelectric plate is mechanically connected to a pressure sensing diaphragm that acts as a force summing device and creates a stress in the plate. The diaphragm, in turn, is exposed to a gas pressure that acts through a perforated cap. The cap protects the sensor against dust, mechanical damage, and other hostile environmental conditions. A change in pressure causes a mechanical change in the piezoelectric plate, which results in a generation of electric charge or voltage, which is then transferred out of the cell using a shielded connector. A commercial model of a pressure sensor is shown in Figure 3.14b [3].

Once a sensor has been designed and constructed, various sensor parameters are usually supplied by the sensor manufacturer. An example of a sensor parameter specification sheet is given in Table 3.7 [3].



(a)

(b)

FIGURE 3.13
(Please see color insert following page 146) A force sensor cell: (a) cross section of a basic piezoelectric sensor and (b) a commercial quartz force sensor. (From Kistler Instrumente AG, www.kistler.com. With permission.)

Many applications exist for the stress/pressure sensor. Examples of stress or force sensor applications include the measurement of compression, tension, impact, vibrating, balancing, striking, rolling, cutting, forming, pressing, machining, and punching operations. Examples fo pressure sensor applications include gas pressure, combustion, explosion, pulsation, actuation, cavitation, fluidic, pneumatic, blast, turbulence, and sound pressure measurements.

Various advantages and disadvantages exist for the piezoelectric stress/pressure sensor. The advantages may be summarized as follows:

1. Simple sensing mechanism, direct conversion of mechanical quantities into electrical quantities
2. Passive or self-generating sensors
3. High sensitivity

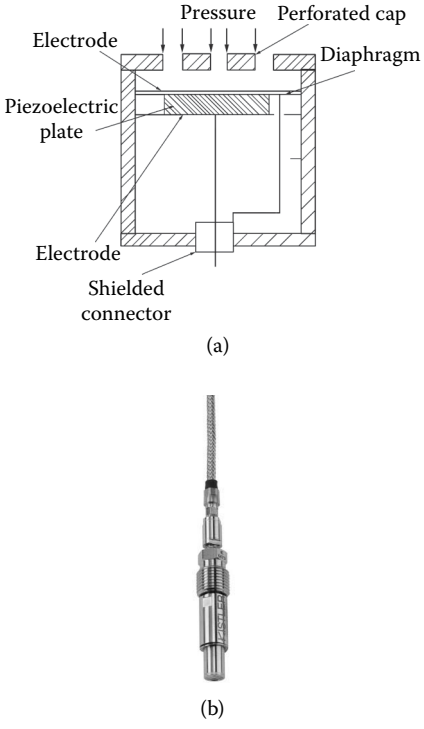


FIGURE 3.14
(Please see color insert following page 146) Piezoelectric pressure sensor cell: (a) cross section of a basic sensor and (b) a commercial pressure sensor. (Kistler Instrumente AG, www.kistler.com. With permission.)

4. High-frequency response, i.e., for measuring of small pressure fluctuations; high mechanical stiffness, i.e., they undergo small deformations (~micron or angstrom) and hence do not load the measurand
5. Light weight
6. Small size (~1 mm)
7. Rugged
8. Nominally low cost

Some of the disadvantages of the piezoelectric force and pressure sensors over other devices are as follows:

1. High-output impedance
2. Sensitive to temperature
3. Brittle

TABLE 3.7

Sensor Parameters for the Kistler Type 61258 Pressure Sensor

Measuring range	bar	0 … 300
Calibrated partial ranges	bar	0 … 100
		0 … 200,0 … 300
Overload	bar	300
Sensitivity at RT	pC/bar	≈ –37
Natural frequency	kHz	>75
Linearity, all ranges (at 23°C)	%FSO	≤±0,4
Acceleration Sensitivity		
axial	bar/g	<0,002
radial	bar/g	<0,003
Operating temperature range	°C	–20 … 350
Temperature, min./max.	°C	–50/400
Sensitivity Change		
250°C ± 100 °C	%	<±1
RT … 350°C	%	<±2
Thermal shock error		
at 1 500 1/min, p_{mi} = 9 bar		
Δp (short-term drift)	bar	≤±0,3
Δp_{mi}	%	<±1,5
Δp_{max}	%	<±1
Insulation resistance at 30°C	Ω	$\geq 10^{13}$
Ground isolation 30 °C	Ω	$\geq 10^{6}$
Torque wrench setting	N·m	10
Weight without cable	g	30
Connector, ceramic insulator	–	KIAG 10-32

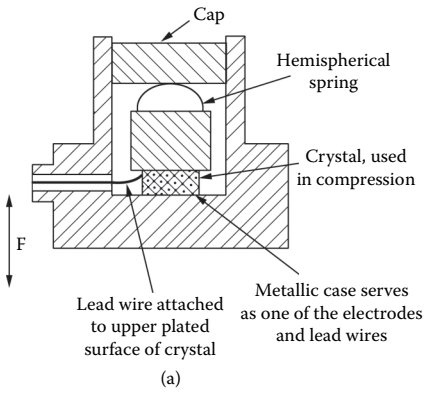
Source: Kistler Instrumente AG, www.kistler.com. With permission.

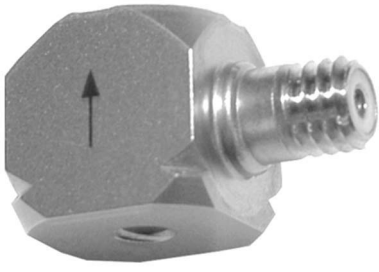
3.7 Piezoelectric Accelerometer

A piezoelectric sensor can also be used as a pressure sensor to determine acceleration. The principle of operation of an accelerometer is based on Newton's law, which states that an acceleration, a, that a body acquires under an influence of a force, F, is the ratio of the force, F, divided by its mass, M; therefore,

$$a = F/M \tag{3.53}$$

Basically, an accelerometer measures a force for a known mass to determine the acceleration.





(b)

FIGURE 3.15
(Please see color insert following page 146) (a) Cross section of a piezoelectric accelerometer and (b) a commercial model. (Kistler Instrumente AG, www.kistler.com. With permission.)

As an example of an acceleration force, consider driving an automobile and then pressing suddenly on the gas. One notices that the driver will experience a force pushing backwards. Similarly, if the brake is applied, the driver will be forced forward and experience a deceleration force.

A cross section of a typical accelerometer and an actual commerical model are shown in Figure 3.15[4]. This accelerometer basically consists of a piezoelectric crystal and a mass, M, which is restrained by a hemispherical spring of very high stiffness. When the accelerometer is set into a motion, the mass, M, acquires an acceleration, a, which manifests itself as the inertia force, $F = Ma$. This force acts on the crystal producing an output voltage proportional to the force, F. Since the mass, M, is known, the output voltage V_0 is proportional to the force and hence the acceleration (see Equation 3.53) as follows:

$$V_0 = K_0 a \tag{3.54}$$

where K_0 is a proportionality constant called sensitivity.

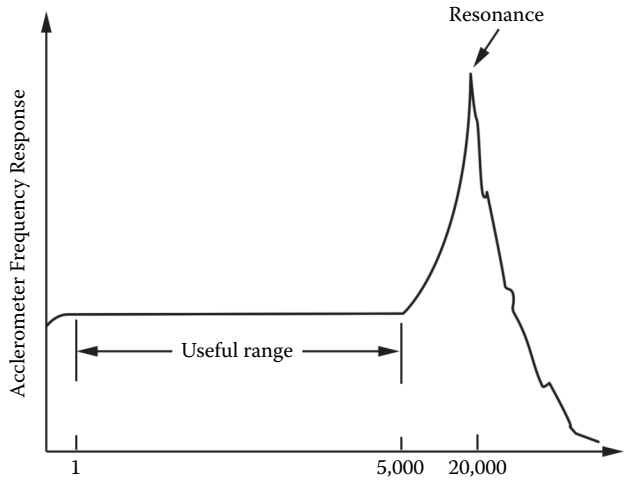


FIGURE 3.16
A typical frequency response curve for a piezoelectric accelerometer.

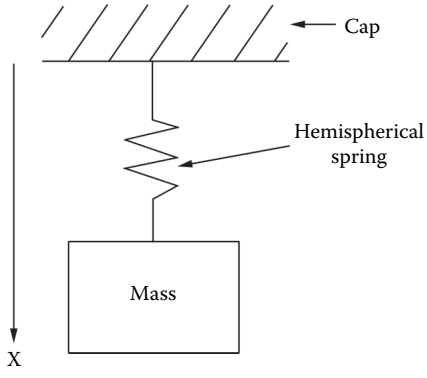


FIGURE 3.17
The equivalent mechanical circuit in a piezoelectric accelerometer.

A typical experimentally measured frequency response curve for the piezoelectric accelerometer is shown in Figure 3.16. The lower limit of the useful range (~1 Hz) is a function of the charge leakage in the piezoelectric element, the mounting methods, the limitations of the signal conditioning systems, and the signal-to-noise ratio. The upper frequency limit results from the existence of the natural resonant frequency of the restraining spring of stiffness, K_S, and the inertia mass, M. The simple equivalent mechanical circuit for the mechanical part of the accelerometer excluding the piezoelectric element is shown in Figure 3.17, which consists of mass, M, and the spring with spring constant, K_S. The resonant frequency for the mechanical circuit may be shown to be the following:

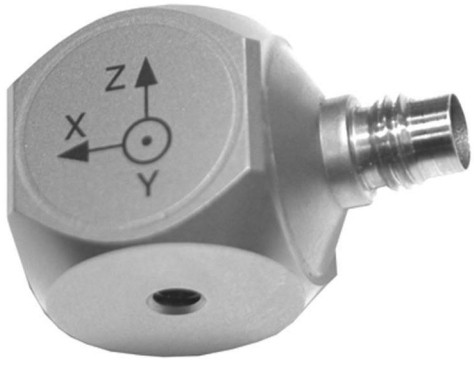


FIGURE 3.18
(Please see color insert following page 146) PCB Piezotronics Model 356A66 triaxial accelerometer for measuring acceleration in three dimensions. (Kistler Instrumente AG, www.kistler.com. With permission.)

$$f_r = (K_S/M)^{.5} \tag{3.55}$$

Typical values of M are in the range of several tens of grams, and K_S is of the order of 10^{10} N/M^2, and hence f_r is of the order of several tens of kilohertz. Therefore, the flat-band frequency response covers a wide range of frequencies up to 5 KHz.

The acceleration is a vector quantity, so it has both magnitude and direction. The piezoelectric element in Figure 3.15 operates in the compressional mode and will produce an output signal only when the force or acceleration is normal to the element. Therefore, in order to determine acceleration in space, one needs the three-dimensional accelerometer, which is shown in Figure 3.18[5].

A typical specification sheet supplied by a manufacturer with a piezoelectric accelerometer is given in Table 3.8[4].

The accelerometer sensitivity, K_O, is commonly expressed in terms of electrical voltage per unit acceleration. Noting that $F = Ma$, one can deduce from Equation 3.48 that K_O is defined as follows:

$$K_O = KM = (e'Sc/C_O)M \tag{3.56}$$

The sensitivity, K_O, depends on the properties of the piezoelectric material and the mass of the inertia element. If a large mass is used, the force on the crystal will be large for a given acceleration, thus generating a relatively large output signal. However, an accelerometer with a large mass results in several disadvantages, which may be summarized as follows:

1. The measurand (acceleration) is distorted if the accelerometer mass is comparable to the mass of the moving body.
2. A heavy accelerometer has a lower resonant frequency, and hence a smaller useful frequency range.

TABLE 3.8

Sensor Parameters for Kistler Single Axis Type 8640A Accelerometer

Specification	Unit	Type 8640A5 Type 8640A5T	Type 8640A10 Type 8640A1OT	Type 8640A50 Type 8640A50T
Acceleration range	g	±5	±10	±50
Acceleration limit	gpk	±8	±16	±86
Threshold (1 … 10 kHz)	grms	0,00014	0,00016	0,00036
Sensitivity (±10%)	mV/g	1000	500	100
Resonant frequency mounted nom	kHz	17	17	25
Frequency response (±5%)	Hz	0,5 … 3000	0,5 … 3000	0,5 … 5000
Phase shift <5°	Hz	2 … 3000	2 … 3000	2 … 5000
Amplitude non-linearity	%FSO	±1	±1	±1
Time constant nom.	sec	1,1	1,1	1,1
Transverse sensitivity typo (max. 3%)	%	1,5	1,5	1,5
Environmental				
Base strain sensitivity @ 250με	g/με	0,004	0,004	0,004
Random vibration max.	grms	50	50	100
Shock limit (1 ms pulse)	gpk	7000	7000	10000
Temperature coefficient of sensitivity	%/°C	0,17	0,23	0,23
Temperature range operating	°C	–40 … 55	–40 … 65	–40 … 65
Output				
Bias nom.	VDC	13	13	13
Impedance	Ω	100	100	100
Voltage Full Scale	V	±5	±5	±5
Power Supply				
Voltage	VDC	22 … 3	0,22 … 30	22 … 30
Constant current	mA	2 … 20	2 … 20	2 … 20
Construction				
Sensing element	Type	PiezoBeam	PiezoBeam	PiezoBeam
Housing/base	Material	Titanium	Titanium	Titanium
Sealing-housing/ connector	Type	Hermetic	Hermetic	Hermetic
Connector	Type	0-32 coaxial	10-32 coaxial	10-32 coaxial
Ground isolated		With accessory	With accessory	With accessory
Mass	grams	3,5	3,5	3,5

(*Continued*)

TABLE 3.8

Sensor Parameters for Kistler Single Axis Type 8640A Accelerometer (Continued)

Specification	Unit	Type 8640A5 Type 8640A5T	Type 8640A10 Type 8640A1OT	Type 8640A50 Type 8640A50T
Mounting	Type	Wax, adhesive, clip, magnet, stud (5-40 UNF-2B)	Wax, adhesive, clip, magnet, stud (5-40 UNF-2B)	Wax, adhesive, clip, magnet, stud (5-40 UNF-2B)
Mounting torque, stud	Nm	0,7	0,7	0,7

Source: Kistler Instrumente AG, www.kistler.com. With permission.
1 g = 9,80665 m/s^2, 1 Inch = 25,4 mm, 1 Gram = 0,03527 oz, 1 lbf-in = 0,113 Nm

There are many applications for an accelerometer that relate to measuring vibratory motion on heavy structures of machinery, machine tools, engines, pumps, compressors, pipes, and presses. The accelerometer can also be used for testing in aviation, space, and automotive vehicles.

3.8 Active Piezoelectric Sensors

In contrast to the passive piezoelectric sensors such as the stress/pressure sensor and the accelerometer, the active piezoelectric sensors utilize an energy source such as a time-varying or AC voltage to excite an acoustic wave. This wave then interacts actively with the measurand of interest, and acoustic wave properties such as velocity, attenuation, or frequency are used to monitor the measurand.

Before embarking on a discussion of various types of active piezoelectric sensors, several of the more popular acoustic waves used in sensing applications will be discussed. Perhaps the most popular wave used in acoustic wave sensors is the bulk acoustic wave. This wave is typically excited by an AC voltage and is accompanied by both mechanical displacements and an electric field or potential. The displacements associated with a bulk acoustic wave may either be predominantly in the direction of propagation or transverse to the direction of propagation. In the former case, the wave is called a longitudinal acoustic wave. The displacements associated with this wave are shown in Figure 3.19. Basically, the medium alternately compresses and expands in the direction of wave propagation. If the acoustic wave has displacements transverse to the propagation direction it is called a shear wave. The displacements associated with this wave are shown in Figure 3.20. In an actual piezoelectric crystal, there is one longitudinal acoustic wave and two shear waves in a particular direction. This is shown in Figure 3.21. The

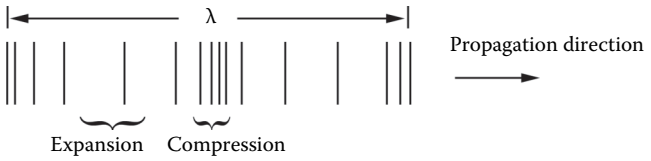


FIGURE 3.19
Schematic representations of the displacements associated with a longitudinal acoustic wave where the wavelength, λ, is shown.

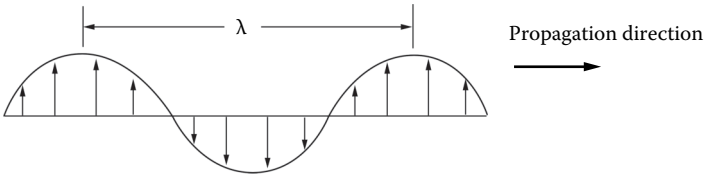


FIGURE 3.20
Schematic representation of the displacements associated with a transverse acoustic wave where the wavelength, λ, is shown.

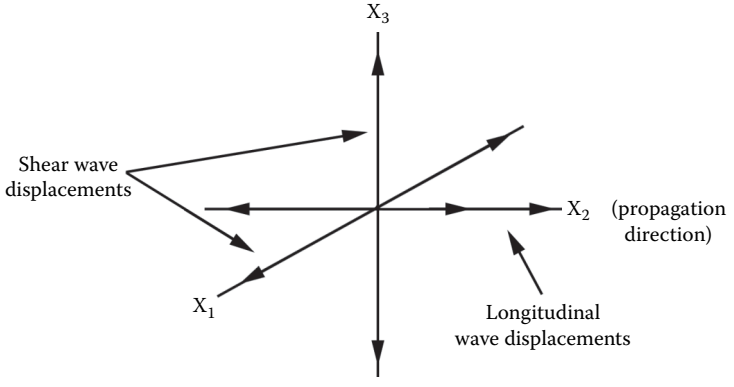


FIGURE 3.21
If the propagation direction is along the X_2 direction, the longitudinal wave would have displacements coincident with the X_2 axis, while the two shear wave displacements would be along X_1 and X_3, respectively.

longitudinal acoustic wave velocity is roughly twice the lowest shear wave velocity. The shear wave velocities are quite close together, and in some high-symmetry directions in crystals they are equal. Typical bulk wave velocities for some common piezoelectric materials are summarized in Table 3.9.

Another type of acoustic wave that has been utilized recently in sensor applications is the surface acoustic wave (SAW). This electrically excited acoustic wave propagates along the surface of the piezoelectric material

TABLE 3.9

Typical Bulk Wave Velocities for Selected Orientation in Quartz, $LiNbO_3$, and $LiTaO_3$

Piezoelectric Material	Orientation	Longitudinal Velocity (M/s)	Fast Shear Velocity (M/s)	Slow Shear Velocity (M/s)
Quartz	X-cut	5,996	4,314	3,913
	Y-cut, Z-cut	5,744	5,099	3,298
$LiNbO_3$	X-cut	6,838	4,410	3,938
	Y-cut, Z-cut	6,555	4,760	4,010
$LiTaO_3$	X-cut	5,688	3,882	3,520
	Y-cut, Z-cut	5,552	4,211	3,367

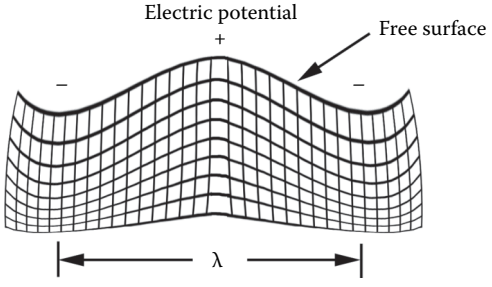


FIGURE 3.22
Schematic representation of the displacements and potential associated with a surface acoustic wave where the wavelength, λ, is shown.

and has both displacements and an electric potential or an electric field associated with it. Figure 3.22 gives a schematic representation of the displacements associated with the SAW. The SAW velocity is typically slightly less than the lowest bulk shear wave velocity. This type of wave has some distinct advantages over bulk waves. Perhaps the most important is related to the fact that this wave travels along the surface, thereby allowing it to have maximum interaction with a measurand. In contrast, a bulk wave can only interact with the measurand at the ends of a piezoelectric crystal.

Acoustic plate modes, which have also been suggested for sensor applications, are similar to electromagnetic waveguide modes and travel in the interior of the piezoelectric plate interacting with the surface at discrete locations. Figure 3.23 describes a propagating plate mode. These modes can have both longitudinally or transversely polarized displacements with respect to their propagation direction. Plate modes are uniquely different from bulk and surface acoustic waves. In particular, one may excite a large number of plate modes, while there is only one SAW or three bulk waves for a particular direction in a crystal. The number of plate modes is dependent upon the plate thickness and frequency of the excitation source.

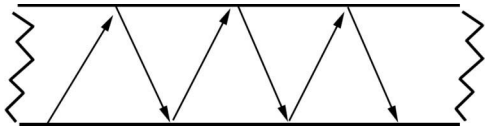


FIGURE 3.23
Description of the propagation of a plate mode in a piezoelectric plate.

In terms of sensor applications, both bulk and plate modes can be used in liquid and gas environments, whereas SAWs are usually used only in gas environments. If a SAW is used in a fluid environment, it will usually radiate much of its energy to the nearby fluid, and hence the SAW will attenuate very quickly along the propagation direction. An exception to this is a SAW, which has displacement components in the surface plane of the piezoelectric crystal. This type of horizontally polarized acoustic wave has been used in fluid sensing applications. Unfortunately, this wave only exists in a few crystals in isolated orientations.

3.9 Bulk Acoustic Wave Sensor

In a bulk acoustic wave sensor a resonant or standing wave is excited by an AC voltage applied to a piezoelectric crystal. This wave must be shear and therefore have displacement components that are purely transverse to the propagation direction. The allowed wavelengths associated with this wave are directly related to the crystal thickness, t, as follows:

$$\lambda_n = 2t/n \tag{3.57}$$

where $n = 1, 3, 5\ldots$.

The displacements associated with $n = 1$ and 3 resonant shear waves are shown in Figure 3.24. The $n = 1$ configuration is commonly called the fundamental mode, and the $n = 3, 5, \ldots$, configurations are referred to as higher-order harmonics. Note that even though acoustic waves with displacement components in the direction of propagation may be excited, they will not be resonant in this structure. This is due to the fact that energy will be leaked out at the crystal surfaces in the form of longitudinal acoustic waves.

Before describing the details of using a piezoelectric crystal as an acoustic wave sensing platform, methods of exciting the resonant acoustic wave will be discussed. This wave may be excited with a thickness field excitation (TFE) or a lateral field excitation (LFE). In the case of the TFE, metal electrodes are deposited on the top and bottom faces of the piezoelectric

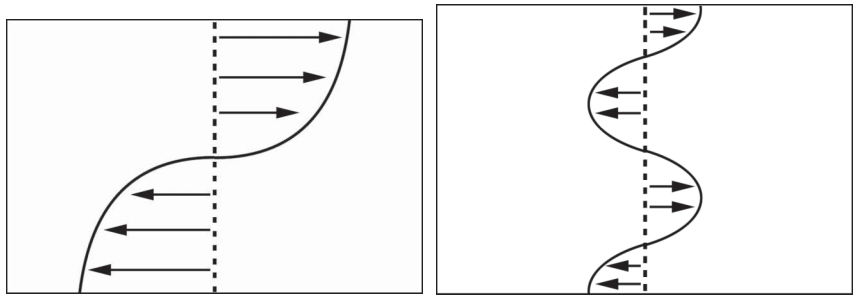
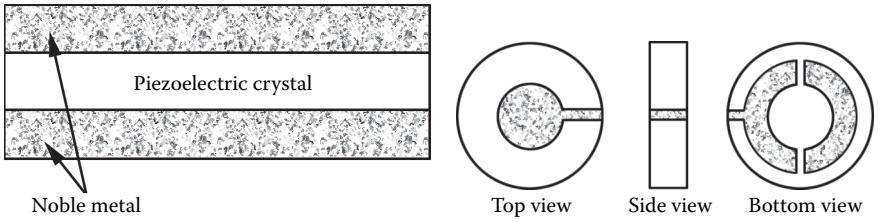


FIGURE 3.24
Displacement associated with the fundamental (n = 1) (a) and first-order harmonic (n = 3) (b) resonant shear mode in a piezoelectric crystal.

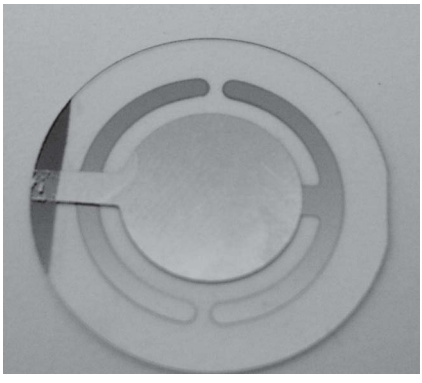
crystal. These electrodes are vacuum deposited noble metals, usually gold. The exact configuration associated with the electrodes could be a gold film on each crystal surface or a more complicated structure, each of which is shown Figure 3.25. In both configurations an AC voltage is applied to the electrodes, resulting in an electric field through the piezoelectric crystal—hence the nomenclature *TFE*. The structure given in Figure 3.25b requires a little elaboration. In this structure the electrical contact is made only on the bottom crystal face since the top electrode goes around the side of the crystal. Both of the structures shown in Figure 3.25 are commonly referred to as quartz crystal monitors (QCMs). The QCM structure shown in Figure 3.24a is used to control processes such as thin-film growth (integrated circuits [IC] industry), liquid evaporation (paper industry), or bacteria growth (biotechnology industry). In these applications the sensing platform is typically located in a holder. The exact configuration shown in Figure 3.25b came about as result of many years of research relating to obtaining a reproducible clean resonant signal as films were being simultaneously deposited on the sensor platform and the substrate of interest.

Lateral field excitation is another approach used to excite the transverse shear mode. A significant amount of work relating to LFE transducer configurations and their subsequent use as sensing platforms has been done at the University of Maine (UM) [6–12]. In this excitation, an AC voltage is applied to electrodes deposited on the bottom crystal face, resulting in an electric field along the lateral direction of the crystal. Two common types of LFE electrode patterns are shown in Figure 3.26.

Very recently researchers at UM have examined a novel approach using a photolithographically deposited spiral coil antenna to excite resonant acoustic waves. This antenna configuration is shown in Figure 3.27. An AC flows through the antenna radiating a time- and frequency-dependent electric field in the piezoelectric crystal. This electric field then excites the resonant acoustic wave.



(a) (b)



(c)

(d)

FIGURE 3.25
(Please see color insert following page 146) (a) Simple TFE electrode pattern, (b) more complex TFE electrode pattern, (c) photo of the top of the more complex TFE device, and (d) photo of the bottom of the more complex TFE device.

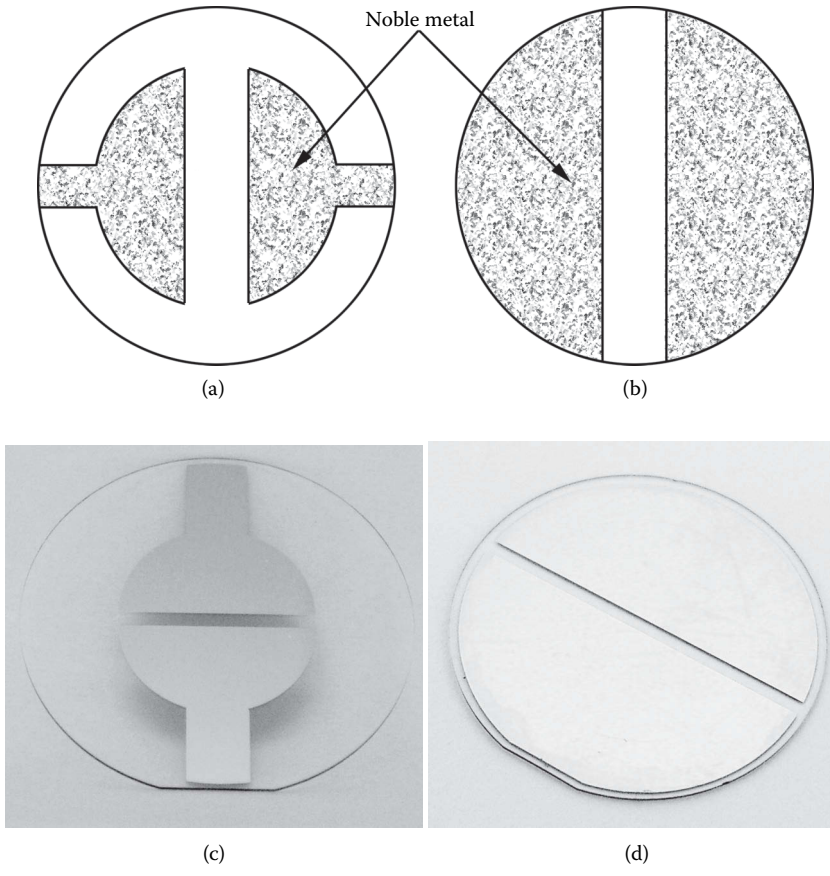


FIGURE 3.26
(Please see color insert following page 146) Electrode patterns for the bottom face of an LFE sensing platform: (a) bite-wing pattern, (b) half-moon pattern, (c) photo of LFE bite-wing pattern, and (d) photo of LFE half-moon pattern.

At first glance one might be led to the conclusion that any orientation of a piezoelectric crystal can be used to excite the purely transverse shear mode. This is not the case, and in fact, only certain orientations in piezoelectric crystals can excite this wave. The most common piezoelectric crystal and orientation is the AT-cut in quartz.

In this orientation only the pure transverse shear is excited using TFE and LFE. The coupling to the other shear mode and the longitudinal mode is zero. This orientation has an additional advantage in that it is temperature compensated, which means the acoustic properties such as velocity, frequency, and displacement are independent of temperature over a significant temperature range about room temperature. Orientations do exist in other piezoelectric crystals for selective excitation of a purely transverse shear mode, but they have not been exploited for sensor applications.

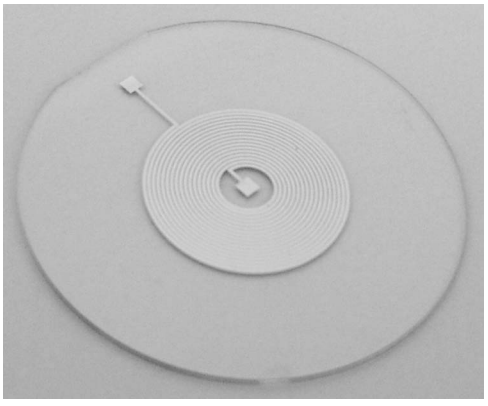


FIGURE 3.27
(Please see color insert following page 146) Monolithic spiral coil antenna configuration for the excitation of resonant acoustic waves in a piezoelectric crystal.

The LFE and spiral coil antenna sensing platforms have distinctive advantages over the TFE sensing platform in that the top sensing surface is bare. A bare top surface allows both mechanical displacements and electric fields associated with the transverse shear mode to penetrate into the adjacent medium or analyte selective film to sense both mechanical and electrical properties. In addition, the spiral coil antenna allows efficient excitation of very high-frequency (~GHz) acoustic waves, which cannot be achieved using TFE or LFE transducers on a single piezoelectric crystal. High-frequency acoustic waves have wavelengths that may potentially probe measurands down to the molecular level.

Once the resonant acoustic wave has been excited, the measurable quantities are the resonant frequency and the electrical impedance. The velocity of the resonant acoustic wave, v, is a function only of the piezoelectric crystal properties and may be expressed as follows:

$$\nu = f_n \lambda_n \tag{3.58}$$

where f_n = frequency of the resonant acoustic wave.

Substituting Equation 3.57 into Equation 3.58 allows the resonant acoustic wave frequency to be written as follows:

$$f_n = \frac{n\nu}{2t} \tag{3.59}$$

where n = 1, 3, 5….

The QCM sensing platform can be represented by an electrical circuit called the Butterworth–Van Dyke equivalent circuit [7], shown in Figure 3.28.

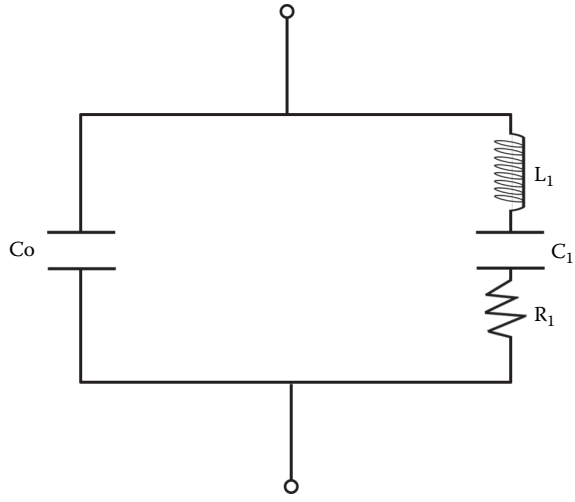


FIGURE 3.28
The Butterworth–Van Dyke equivalent circuit for the QCM.

C_0 is the electrostatic capacitance between the electrodes on the top and bottom faces of the piezoelectric crystal. The L_1, C_1, and R_1 arm of the circuit is referred to as the motional arm. L_1 represents the motional inductance, which is a measure of the inertia of the oscillating crystal. C_1 represents the motional capacitance, which is related to the elasticity or energy stored in the crystal. R_1 represents the motional resistance and is a measure of the viscosity or energy dissipation in the crystal. The circuit elements, L_1, C_1, and R_1, can be shown to be a function of the material properties and thickness of the piezoelectric crystal [13].

Although the LFE and spiral coil platforms have been used in several sensing applications, no definitive equivalent circuit currently exists for these platforms.

3.10 Bulk Acoustic Wave Sensor Response Measurement

The response of resonant bulk acoustic wave sensors to perturbations in the environment is commonly measured in two ways. The acoustic wave sensor can be utilized as the feedback element in a crystal-controlled oscillator, where the frequency and voltage of the output can be measured or the impedance of the acoustic wave sensor can be measured using an impedance analyzer or network analyzer. The advantage of the oscillator configuration is that the measurement system is compact and well suited to remote or portable

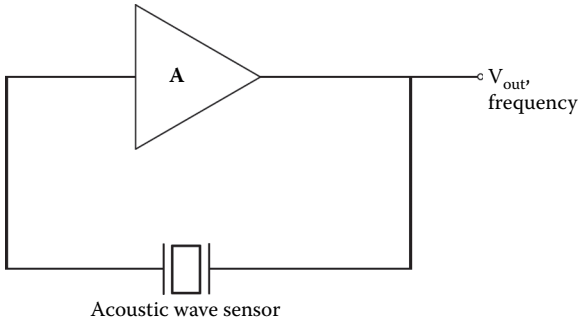


FIGURE 3.29
Crystal-controlled oscillator.

sensing systems. However, it is possible to measure only two parameters of the oscillator, voltage and frequency. Thus, it is not possible to determine the origin of the response if the sensor is responding to more than two analytes in the environment, or responding to more than two property changes of a single analyte, i.e., viscosity, permittivity, and conductivity changes in a liquid. The measurement of the impedance of the sensor will enable one to determine more exactly the nature of the response. Unfortunately, commercially available impedance analyzers are bulky and not well suited to be remote or portable sensors.

A crystal-controlled oscillator consists of two fundamental parts: an amplifier or gain circuit and a feedback network (Figure 3.29). The acoustic wave sensor serves as the feedback network, providing positive feedback at the frequency of oscillation [13]. The gain of the amplifier is A, and the gain of the feedback element is β. The two criteria that must be met for the oscillator are that the loop gain, $A \cdot \beta$, must be greater than unity, and that the total phase $\Theta_A + \Theta_\beta = 2n\pi$, where $n = 0, 1, 2, \ldots$. The loop gain must be greater than unity so that the signal level can increase to a usable level, where it is typically adjusted via automatic gain control. The phase criterion is crucial in sensor applications. This is due to the fact that since perturbations to the measurement environment affect the sensor response, the phase will change, resulting in a change to the frequency of oscillation. This is measured at the output. Additionally, losses in the feedback element due to perturbations in the measurement environment can be measured as an attenuation to the output signal, V_{out}.

Another method of measuring the sensor response of resonant acoustic wave sensors is to directly measure the impedance using an impedance analyzer. Changes to the impedance of the acoustic wave sensor due to perturbations of the measurement environment can then be stored and analyzed. In the case of the QCM whose equivalent circuit is given in Figure 3.28, the change in impedance can affect the series and parallel

resonant frequencies, the amplitudes and phases of both resonant frequencies, and the Q of the sensor. Due to this, the use of an impedance analyzer is preferred when the measurement environment is complex, such as in liquids. In the case of the LFE sensor, the exact equivalent circuit is yet to be determined. However, the Butterworth–Van Dyke circuit can be used as an approximation.

The TFE, spiral coil, and LFE platforms can be used as stand-alone sensors or coated with a film that may be selective to a particular chemical or biological measurand. The most common utilization of the TFE platform as a stand-alone sensor is in thin-film deposition systems. In this application a TFE platform is usually referred to as a QCM. As indicated in the name, quartz is the piezoelectric crystal that is used. Also, a very specific crystallographic orientation in quartz is used, namely, the AT-cut. In this orientation a purely transverse shear mode is excited and the acoustic wave properties do not change appreciably near room temperature. Typically in a deposition process the QCM is placed close to the substrate upon which a deposition is taking place, and the film is deposited on both the QCM and the substrate simultaneously. The frequency of the fundamental resonant acoustic wave in the QCM changes in the deposition process and is used to monitor the thickness of the film deposited on the substrate. In particular, it may be shown [14] that the relative change in frequency of the QCM, $\Delta f/f_q$, is directly proportional to the mass of the film being deposited as follows:

$$\Delta f \Big/ f_q = -\frac{M_f}{M_q} \tag{3.60}$$

where M_f = film mass, M_q = quartz crystal mass, and f_q = resonant frequency of the bare QCM.

M_f is a function of the QCM surface area, A, the film density, ρ_f, and thickness, h_f, as follows:

$$M_f = \rho_f A h_f \tag{3.61}$$

Substituting Equation 3.61 into Equation 3.60 and solving for h_f, one obtains

$$h_f = \frac{\Delta f M_q}{\rho_f A f_q} \tag{3.62}$$

where Δf = decrease in the QCM frequency.

Since Δf is measured and the other parameters in Equation 3.62 are known, the thickness of the deposited film is directly proportional to the decrease in the QCM frequency as follows:

$$h_f = K\Delta f \tag{3.63}$$

where

$$K = M_q / \rho_f A f_q$$

The TFE, LFE, and spiral coil sensing platforms can also be used as stand-alone sensors in a fluid environment. In this case the measurand of interest is a fluid property, which includes density, elasticity, viscosity, electrical conductivity, or permittivity. It is important to point out that the resonant acoustic wave does not propagate into the fluid, but rather the amplitudes of the mechanical displacement and electric field decay exponentially into the fluid. The decay rate and depth of penetration of the acoustic fields into the adjacent fluid are functions of the fluid properties and the acoustic wavelength. The resulting probing of the acoustic wave into the fluid changes the resonant frequency of the acoustic wave.

Figures 3.30 to 3.32 present the variations of the frequency shifts for selected mechanical and electrical fluid properties for different LFE sensors and a QCM. The electrode configuration for LFE sensors is the half-moon structure shown in Figure 3.26b with electrode gaps varying from 1.5 to 3.5 mm. In Figure 3.30 the frequency is shown to decrease as a function of NaCl solution in water. It is interesting to point out that the QCM frequency does not change

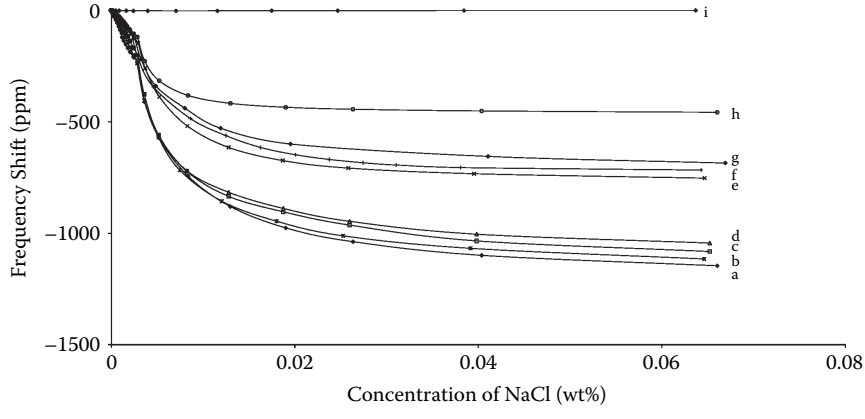


FIGURE 3.30
(Please see color insert following page 146) Frequency changes of the sensors as a function of NaCl concentration in water: (a) LFE sensor with electrode gap = 1.5 mm, (b) gap = 2.0 mm, (c) gap = 2.5 mm, (d) gap = 3.5 mm, (e) open-ring electrode, (f) small electrode, (g) closed-ring electrode, (h) gap = 12.7 mm, (i) standard QCM. (From Hu, Y., et al., *IEEE Trans. Ultrasonics Ferroelectrics Frequency Control* 51 [2004]: 1376–78. With permission.)

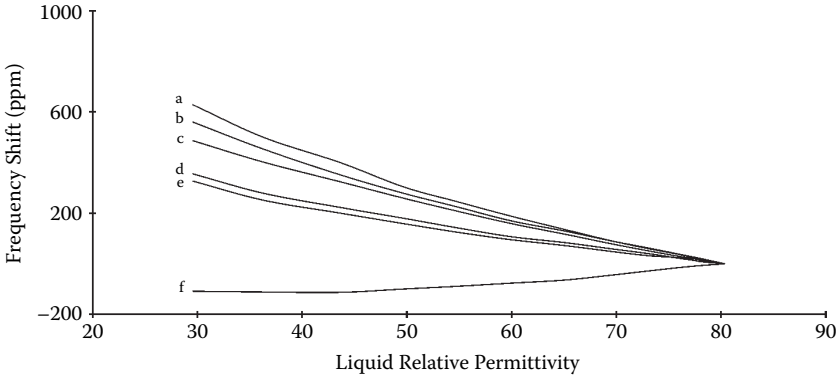


FIGURE 3.31
(Please see color insert following page 146) Frequency changes of the sensors as a function of the liquid relative permittivity of the 2-propanol solution: (a) LFE sensor with gap = 1.5 mm, (b) LFE sensor with gap = 2.0 mm, (c) LFE sensor with gap = 3.5 mm, (d) open-ring electrode geometry, (e) small electrode geometry, (f) standard QCM. (From Hu, Y., et al., *IEEE Trans. Ultrasonics Ferroelectrics Frequency Control* 51 [2004]: 1376–78. With permission.)

appreciably with increasing NaCl concentration due to the fact that the electrode on the top surface shields most of the electric field from penetrating into the adjacent fluid. In Figure 3.31 the frequency is shown to increase as the relative dielectric constant of a 2-propanol solution decreases for the LFE sensor, while for the QCM a slight decrease in frequency is observed. The QCM smaller frequency shift in a direction opposite to that of the LFE sensor is due to the fact that in a 2-propanol solution the viscosity increases with decreasing relative dielectric constant. Finally, in Figure 3.32, the frequency is seen to decrease as the viscosity of the water glycerol solution increases. Again, notice that the changes observed for the LFE sensor are larger than those observed for the QCM. This is most likely due to the bare sensing surface in the LFE sensor, which allows the displacements associated with the resonant acoustic wave to penetrate more deeply into the adjacent fluid.

The bulk acoustic wave sensor can also be used to sense chemical and biological targets if the sensing platform is coated with a film selective to the target of interest. As an example of chemical sensing, an LFE sensor and a QCM are coated with a polymer, polyepichlorohydrin (PECH), which is selective to phosmet, an organophosphate pesticide [7]. In Figure 3.33 the frequency shift from baseline due to successive additions of 10 µl of phosmet solution in 200 ml of water is shown. The phosmet solution exposed to the LFE sensor had a phosmet concentration of 1.3425 ppm, while for the QCM, the concentration was 1.4917 ppm. It is interesting to point out that the shifts observed for the LFE sensor were higher than those observed for the QCM, even though the 10 µl injections for the QCM contained a higher concentration level of phosmet. For biological sensing an LFE device is coated with a multilayer biochemical film (see Figure 3.34) selective to *E. coli* 0157:H7, the most dangerous

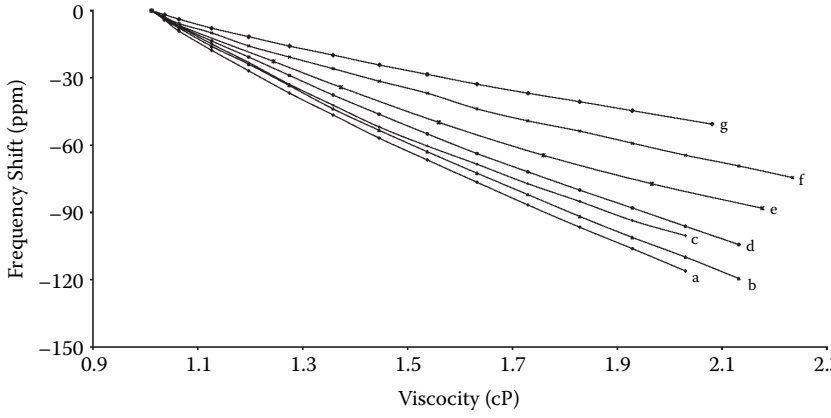


FIGURE 3.32
(Please see color insert following page 146) Frequency shifts of the sensors as a function of a water-glycerol solution: (a) LFE sensor with electrode gap = 1.5 mm, (b) gap = 2.0 mm, (c) gap = 3.5 mm, (d) gap = 12.7 mm, (e) open-ring electrode, (f) small electrode, (g) standard QCM. (From Hu, Y., et al., *IEEE Trans. Ultrasonics Ferroelectrics Frequency Control* 51 [2004]: 1376–78. With permission.)

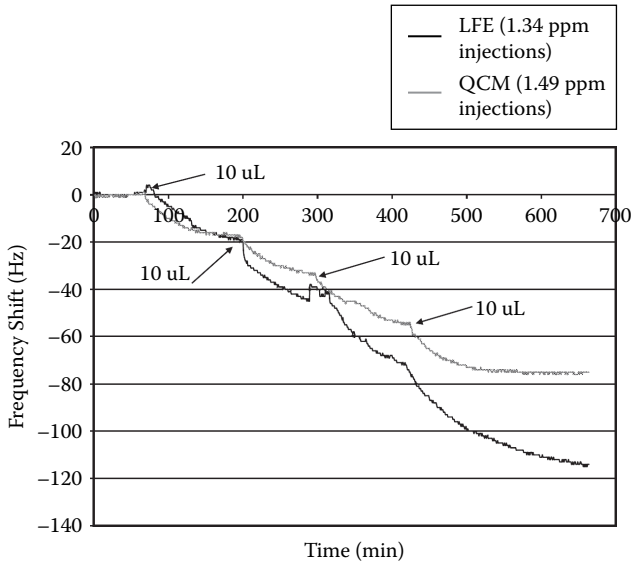


FIGURE 3.33
(Please see color insert following page 146) Frequency shift from baseline due to phosmet additions. (Reprinted from Hu, Y., et al., *Sensors and Actuators B* 108 [2005]: 910–16. With permission from Elsevier.)

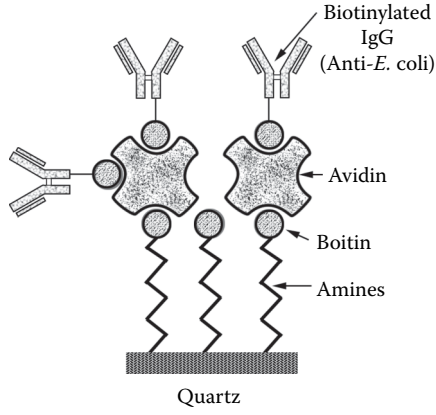


FIGURE 3.34
Schematic of antibody/antigen immobilization for *E. coli* O157:H7 selective film. (From McCann, D.F., Wark, M.S., French, L.A., and Vetelino, J.F. 2008. Novel transducer configurations for bulk acoustic wave sensors. 2008 *IEEE Conference*, Lecce, Italy, pp. 1448–1451, Oct. 26–29. © 2008 IEEE. With permission.)

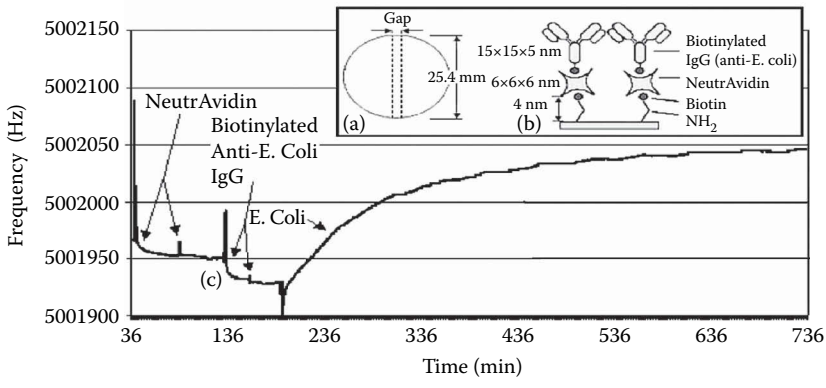


FIGURE 3.35
(Please see color insert following page 146) (a) LFE sensor: Top surface is bare and the dotted lines indicate the bottom surface electrode configuration. (b) Schematic of the antibody film on the sensing surface. (c) Frequency response of the LFE to neutrAvidin, anti-*E. coli* O157:H7 and *E. coli* O157:H7 bacteria. (From McCann, D.F., Wark, M.S., French, L.A., and Vetelino, J.F. 2008. Novel transducer configurations for bulk acoustic wave sensors. 2008 *IEEE Conference*, Lecce, Italy, pp. 1448–1451, Oct. 26–29. © 2008 IEEE. With permission.)

strain of *E. coli* bacteria. In Figure 3.35 the variation of the resonant frequency of an LFE sensor is shown while it is being coated with the biochemical film and exposed to *E. coli* 0157:H7. One notices that, in contrast to the sensor responses shown in Figure 3.32, the response time to the biological target is much longer. This is due in part to the fact that it takes more time for the *E. coli* 0157:H7 bacteria to diffuse to the LFE sensor surface.

3.11 Surface Acoustic Wave Sensors

In contrast to **bulk** acoustic wave sensors the initial work in SAW sensors was done in the late 1970s. Since that time a significant amount of research was performed at the University of Maine on this sensor for sensing inorganic and organic gases and biological molecules. Whereas in the BAW sensor one makes use of bulk acoustic waves, in the SAW sensor the wave of interest is the SAW.

SAWs are excited on the sensing surface of a piezoelectric crystal by means of an interdigital (IDT) electrode configuration. Figure 3.36 shows a simple SAW device.

An AC voltage is applied to the IDT on the left side of the device, and each adjacent IDT finger has opposite polarity at a particular instant of time. The alternating electric field between the metal electrodes, which are usually aluminum, excites an alternating stress on the crystal surface. If the electrode spacings are a half wavelength apart, the alternating stress will cause the wavelets excited between adjacent fingers to add in phase, resulting in the propagation of a SAW. The appropriate electrode spacing can be determined by the excitation frequency, f, and the known SAW velocity, V_{SAW}, which is a function of the piezoelectric material constants and the crystallographic orientation. In particular, the electrode spacing, d, is as follows:

$$d = \frac{V_{SAW}}{2f} \tag{3.64}$$

Given the size of single crystals, the limits of photolithography, and the SAW velocity, the frequency of SAW devices varies from about 50 MHz to 1 GHz.

If one places an amplifier in the feedback loop of the SAW device to compensate for losses in propagation, excitation, and reception, one has a simple SAW oscillator or resonator. Any material or foreign substance that comes in contact with the surface between the IDTs will modify the SAW velocity. Since d is fixed, one can see from Equation 3.64 that f changes. The change in SAW frequency then becomes a function of the amount of substance on the surface.

The common SAW sensor configuration [15] consists of two identical SAW devices fabricated on a single piezoelectric substrate, as shown in Figure 3.37.

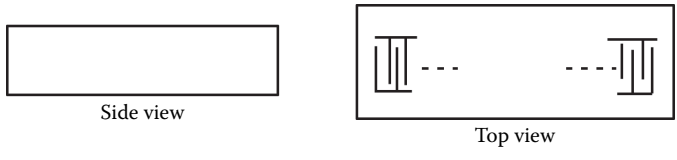


FIGURE 3.36
Simple SAW device.

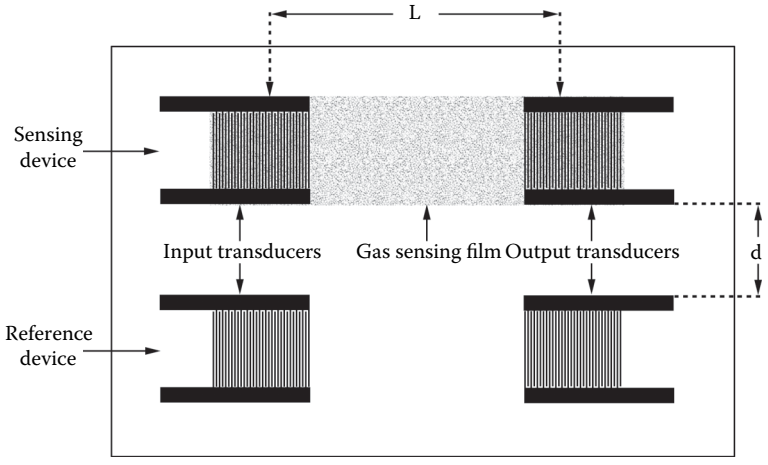


FIGURE 3.37
Schematic representation of the dual-delay line SAW sensor.

The SAW device, which has a sensing film deposited between the electrodes, is called the sensing device, whereas the other SAW device is called the reference device. In this so-called dual-delay line configuration, changes in the material properties of the piezoelectric substrate due to environmental effects such as temperature are eliminated. If one used only a single device and tried to operate the device over a specified range, the change in frequency would be a function of not only the sensing film changes due to the sorption of a target analyte, but also the change in SAW velocity over the temperature range of interest.

The complete measuring system for monitoring the difference frequency of the SAW sensing element is shown in Figure 3.38. The feedback amplifiers create the oscillating loops with the SAW delay lines, while the directional couplers in the loop allow the signals to be measured without interfering with the loops. The outputs from the directional couplers are connected to the inputs of a double balanced mixer, which multiplies the sinusoidal waveforms and outputs a signal containing the sum frequency of $(f_r + f_s)$ and the difference frequency of $(f_r - f_s)$. The low-pass filter allows the difference frequency to pass while blocking the sum frequency. The frequency counter then measures only the difference frequency and transmits it to a computer where it is displayed and monitored.

In order to illustrate the performance of the dual-delay line sensor configuration, the sensing path of a Y-cut Z-propagating $LiNbO_3$ crystal was coated with a WO_3 film and exposed to hydrogen sulfide. The response of the sensor is shown in Figure 3.39. Notice that the change in frequency is almost 600 Hz, which is very large since the measurement accuracy for frequency changes is typically in the single hertz range. One can clearly see that even though the SAW sensor is more complex than a bulk wave sensor, it does offer high sensitivity.

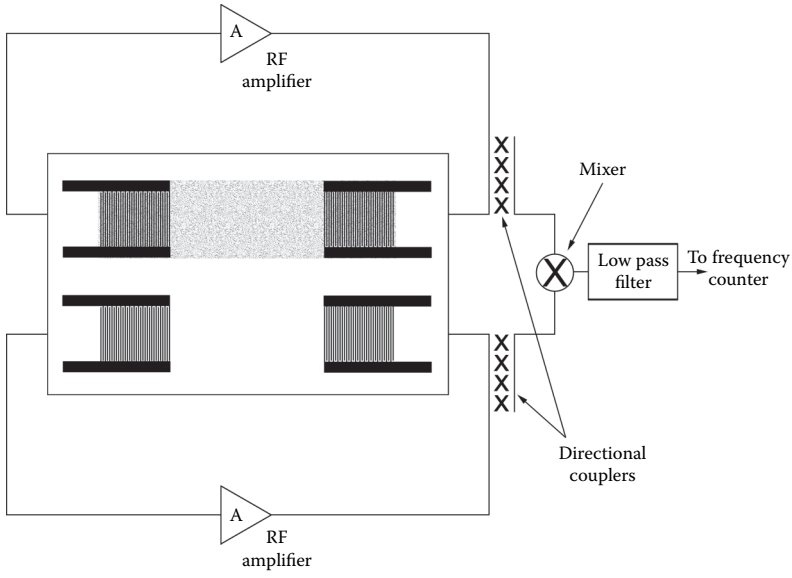


FIGURE 3.38
The dual-delay time SAW sensing element and its associated electronics.

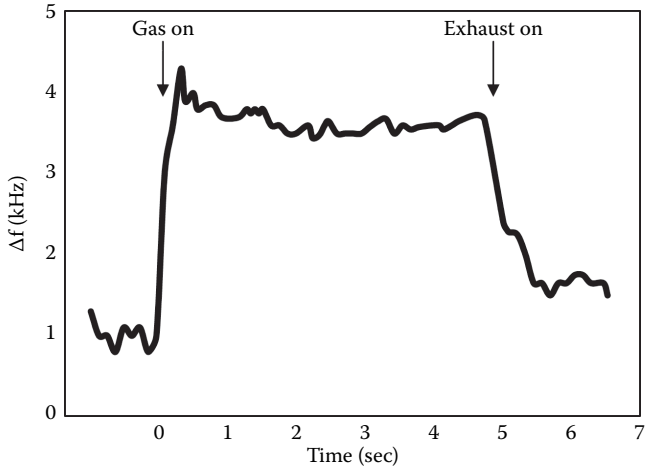


FIGURE 3.39
(Please see color insert following page 146) A typical SAW sensor response to H_2S. The gas concentration for this exposure was 0.25 ppm, and the device operating temperature was 98°C.

References

1. Auld, B.A. 1990. *Acoustic Fields and Waves in Solids*, Volume II. Malabar, FL: Robert E. Krieger.
2. Kistler Press Force Sensor Type 9001A.
3. Kistler Pressure Sensor Type 6125C.
4. Kistler single axis accelerometer Type 8640A.
5. Kistler triaxial accelerometer Type 8766A.
6. Hu Y., French L.A., Radecsky K., daCunha M.P., Millard P., and Vetelino, J.F. 2004. A Lateral Field Excited Liquid Acoustic Wave Sensor, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 51:11, 1373–1380.
7. Pinkham W., Frankel D., French L., Hu Y., and Vetelino, J. F. 2005. Pesticide detection using lateral field excited acoustic wave sensor, *Sensors and Actuators B Chemical* 108, 910–916.
8. Pollard T.B., Kenny T.D., daCunha M.P., and Vetelino, J.F. 2006. Pure SH-SAW propagation, transduction and measurements on KNb03, *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 53: 1, 199–208.
9. Hempel U., Lucklum R., Hauptmann P., and Vetelino, J.F. 2008. Advanced application of the impedance spectrum of a lateral field excited sensor, *Sensors and Actuators*: A. 142, 97–103.
10. McCann D. F., McGann J.M., Parks J.M., Frankel D.J., daCunha M.P., and Vetelino, J.F. 2009. Lateral field excited LiTa03 high frequency bulk acoustic wave sensor, *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 56, 4, 779–787.
11. McCann D., French L., Wark M. and Vetelino, J.F. 2009. Recent advances in lateral field excited and monolithic spiral coil acoustic transduction bulk acoustic wave sensor platforms, *Journal of Measurement Science and Technology*, 20, 12, pp. 124001–124012.
12. McCann D.F., Wark M.S., French L.A., and Vetelino J.F. 2008. Novel transducer configurations for bulk acoustic wave sensors, 2008 *IEEE Conference*, Lecce, Italy, pp 1448–1451, Oct. 26–29.
13. Bottom V.E. 1982. *Introduction to Quartz Crystal Unit Design*. New York: Van Nostrand Reinhold.
14. Sauerbrey G., 1959. *Zeitschrift für Physik*, 155, 206–222.
15. Vetelino J.F., Bryant A., and Lee D.L. 1981. A surface acoustic wave gas detector, *1981 IEEE Ultrasonics Symposium*, Chicago, IL, pp. 171–174, Oct. 14–16.

Questions

1. Which of the following are piezoelectric crystals?
   a. NaBr
   b. ZnTe

c. CdTe
d. Si

2. Name at least one piezoelectric material used in the following applications:
   a. Sonar
   b. Watches
   c. Cell phones

3. A stress is applied to a piezoelectric material. Explain in words and equation(s) the resulting polarization that occurs in the material.
4. A nonpiezoelectric material is known to have the following elastic constant matrix:

$$c = \begin{bmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{bmatrix}$$

A uniform compressive strain is applied to the crystal. Obtain expressions for the resulting stress.

5. An isotropic (nonpiezoelectric) solid is characterized by two independent elastic constants, λ and μ. If the elastic constant matrix is given as follows,

$$c = \begin{bmatrix} (\lambda+2\mu) & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & (\lambda+2\mu) & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & (\lambda+2\mu) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{bmatrix}$$

obtain expressions for the stresses induced by a strain in the x direction on the x face of the cube.

6. What is electrostriction?
7. Two crystals have the same piezoelectric constant. If the coupling coefficient of one crystal is twice as large as the other, what is the ratio of the effective dielectric constants of these crystals?
8. What is the equivalent circuit and time response for a piezoelectric sensing element used to measure a step function force? Define all terms.
9. When subjected to static force, current, charge, and voltage occur in a sensing element. What is the approximate magnitude of each?
10. For the following applications, indicate the best piezoelectric material to use and why.
    a. Static force
    b. High temperature
    c. Low stresses
    d. Long-term use
    e. Low temperature
11. What is the disadvantage of using a heavy accelerometer?
12. In a resonant piezoelectric sensor, derive an equation for the change in frequency when the sensor is loaded with a small mass, Δm.
13. Derive an equation for the resonant frequencies possible in an AT-cut quartz crystal monitor. Define all terms.
14. For the fundamental and third harmonics in an AT-cut quartz monitor, plot the displacements.
15. A bulk acoustic wave sensor may be operated at higher harmonics. Explain what this means.
16. What is a
    a. TFE sensor?
    b. LFE sensor?

    Also, what advantages does an LFE sensor have over the common TFE quartz sensor?
17. Plot the impedance (magnitude and phase) of an AT-cut quartz crystal monitor as a function of frequency, clearly pointing out the location of the resonant frequency.
18. Qualitatively describe a SAW sensor.

4

Fiber Optic Sensors

4.1 Introduction

Fiber optic technology is based upon the transmission of light through fibers that can be as small as a human hair. The light, which is electromagnetic waves, can have a wide range of frequencies ranging from megahertz (10^6 cps) to the high gigahertz (10^9 to 10^{13} cps). The fact that different frequency light waves can be transmitted through these fibers allows one to transmit a tremendous amount of information. It is not uncommon to have bundles of fibers (100 or more) in a single fiber optic link. It is not surprising that fiber optic links have immediate application in transmission of data from one point to another. Fiber optic links may transmit energy over both short distances (~1 m) and long distances (5×10^3 km or greater). Relative to short distances, common applications are networking computers within an office or industry, while for long distances, transatlantic fiber optic links are used for telephone communications. Another very important application for optical fibers is the area of sensors. These sensors can be used for applications ranging from probing the human body to determining structural flaws in bridges and buildings. More details relating to sensor applications are discussed later.

4.2 Background

The first work relating to the transmission of light from one point to another in a controlled fashion can be traced back to the late 1800s. William Wheeler used light from an electric arc lamp placed in the basement of his house to direct light around the house using a system of pipes [1]. The pipes had reflective linings and diffuse optics to facilitate the transmission of the light waves. Wheeler's invention relating to piping light is documented in U.S. Patent 247,229 (1880). Figure 4.1 describes the piping light system developed by Wheeler.

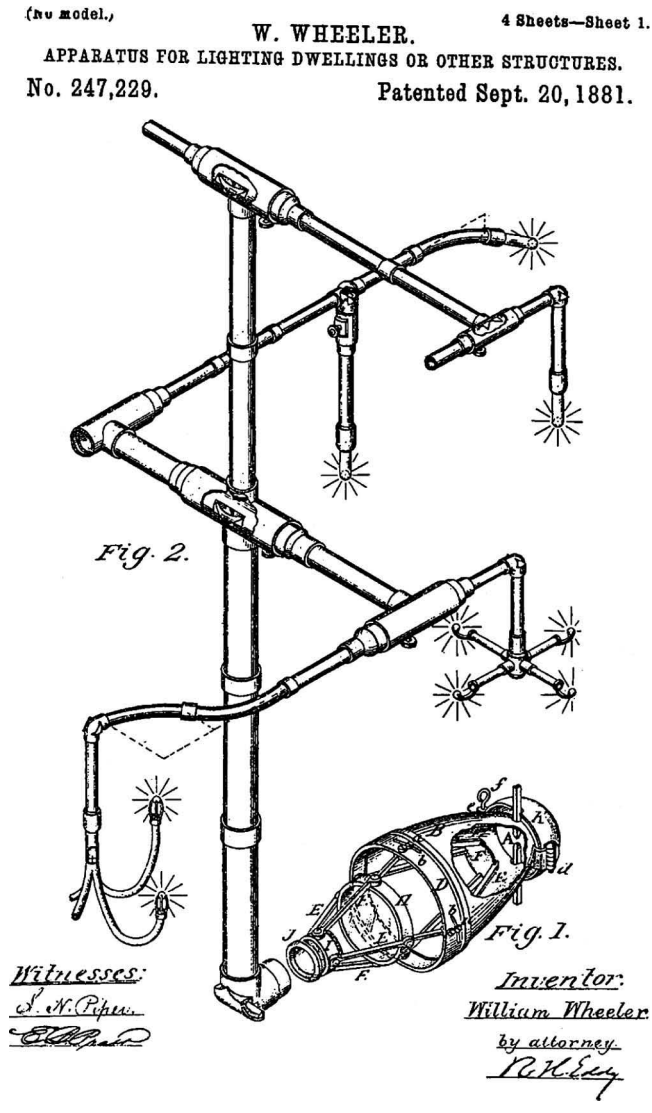


FIGURE 4.1
Piping light system.

The piping light system, although functional, was not completely successful due to nonideal reflecting surfaces in the pipe.

The next serious attempt at using light piping did not appear until the 1950s. This work was motivated by the fact that if a piping light system could be realized, a tremendous amount of data could be transmitted from one point to another through pipes. One of the major obstacles to light piping was the development of a transmission system. Research began to focus on

the use of glass fibers as the transmission material. However, many problems had to be resolved relative to the glass fiber. First, one had to be certain that the light energy did not leak out of the fiber. Second, the fiber had to be mechanically strong so that environmental factors would not alter the fiber properties. This is illustrated in Figure 4.2. When there is no coating around the fiber other than air, surface impurities and surface roughness can cause light energy to leak. If, however, a coating is applied to the fiber, it may be possible to minimize the leakage.

Researchers tried a wide range of coatings, such as beeswax, plastic, and margarine, for the fibers. Finally, in 1956, a University of Michigan undergraduate student by the name of Larry Curtis proposed a glass-clad fiber that consisted of melting a low-index glass onto a high-index glass [2]. This combination for the fiber quickly became the standard for fiber optic links.

The development of the glass-clad fibers clearly opened the door for fiber optic technology; however, several other important questions needed to be resolved before fiber optics became accepted as a standard. It was recognized that light could not be transmitted over long distances even though it did not leak out of the optical fiber. The light was being scattered in the fibers and also absorbed by impurities in the glass fiber. The scattering was attributed to structural abnormalities in the fiber, such as polycrystalline boundaries and voids. The absorption was due to impurities that were embedded in the fiber. The light energy was being absorbed by the impurities and excited vibrational modes in the impurity molecules, which resulted in the heating of the fiber. Some of the work in this area was done by Charles Kao and George Hockman, who were research scientists at Corning Glass [3]. Finally

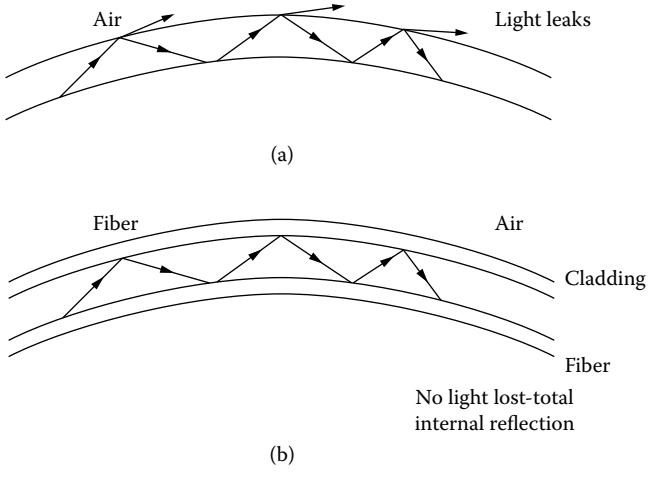


FIGURE 4.2
(a) Leakage occurring when fiber is exposed to air. (b) A cladding minimizes light leakage.

in 1970, Corning Glass was able to produce low-loss optical fibers that lost only 10% of the light over a distance of 30 miles. The question of efficient light generation and reception still needed to be addressed in order to realize a viable fiber optic communication system. The invention of the laser in 1960 solved the question of light sources. Light could now be efficiently generated over a wide range of frequencies. It was determined that electrons in various materials could be excited to levels above their ground state. When these electrons dropped from their excited states they emitted quanta of light energy at characteristic frequencies. The emitted energy can be described by Planck's relation:

$$E = hf \tag{4.1}$$

where h = Planck's constant and f = frequency.

By using different varieties of material it was then determined that the lasers could emit light covering a wide frequency range. The question of light reception was solved when silicon technology in the 1960s resulted in the realization of photodetectors. In these photodetectors the arriving light excites electrons in the silicon into the conduction band. They in turn cause a current to appear in the semiconductor. This current is often referred to as photo-excited current. Lasers, photodetectors, and improved fibers resulted in the realization of fiber optics as a viable technology for data transmission.

4.3 Theory

To completely describe the theory of excitation, transmission, and reception of light in fiber optic links requires a background in solid-state physics and electromagnetics. Light must be interfaced into the individual fiber. This is often done using lasers that can be tuned to transmit light at very specific frequencies. The light then propagates in the fiber optic links as modes. In order to determine whether the light will propagate in the fiber, one must realize that the fiber is a waveguiding structure. Therefore, one must use Maxwell's equations and solve the equations subject to the appropriate boundary conditions. This problem is more difficult than the traditional rectangular waveguide problem that is normally treated in undergraduate electromagnetic courses. In the fiber optic link the cladding on the glass fiber is not a perfect conductor. Therefore, the electromagnetic boundary conditions require the continuity of electric and magnetic fields at the glass-cladding interface. One also realizes that if the fiber is bent, it is possible for some of the light energy to be transmitted into the cladding. This obviously contributes to loss and corresponds to the light energy being leaked

into the cladding. This point is discussed in more detail later. The geometry of the glass fiber further complicates the mathematics in that the problem cannot be readily treated in rectangular coordinates but must be treated in cylindrical coordinates. Although involved, this problem can be solved and the propagating modes and their appropriate frequencies identified. Finally, the light energy must be received and appropriately processed. The light energy is received using an array of photodetectors. In this case, the light energy impinges on a semiconducting material and electrons are excited into the conduction band, giving rise to a current that is then measured. In order to fully understand the solid-state and electromagnetic theory associated with fiber optics, one should have at least one course on solid-state physics and two courses in electromagnetics, one relating to basic electromagnetic theory and the second relating to microwave theory. This is normally beyond the scope of most undergraduate students.

Fiber optics can fortunately be described in a qualitative sense. Normally light travels in a straight line, as it does in free space. Confining light within a prescribed structure will allow more direct point-to-point communications. Also, light need not propagate in a straight line. However, one has to be careful that the bending in the fiber is not too great; otherwise, light energy will leak out.

4.4 Light Leaking and Absorption in the Fiber Optic Link

Two very important topics that need to be addressed relative to the materials used in a fiber optic link are leaking and absorption. Leaking can be described using basic electromagnetics. It is necessary that total internal reflection occurs when the light is incident from the glass fiber onto the cladding. This may be illustrated by referring to Figure 4.3.

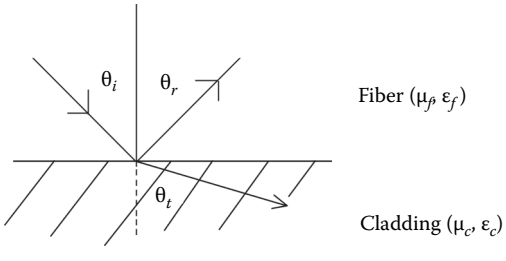


FIGURE 4.3
Incident electromagnetic wave in the glass fiber impinging on the cladding. (Note that the angle of incidence and reflection are the same, and therefore $\theta_i = \theta_r$.)

In order for no leakage of light to occur, θ_t must be equal to 90°. This may be explained using Snell's law, which is given as follows:

$$k_f \sin\theta_i = k_c \sin\theta_t \tag{4.2}$$

where

$$k_f = \omega\sqrt{\mu_f \varepsilon_f}$$

and

$$k_c = \omega\sqrt{\mu_c \varepsilon_c}$$

The terms μ_c and μ_f are the permeabilities of the two media, and ε_c and ε_f are permittivities of the two media. Total internal reflection starts when $\theta_t = 90°$. Therefore, the transition angle, θ_{ic}, for total internal reflection from Equation 4.2 becomes

$$\theta_{ic} = \sin^{-1}\left[\frac{\mu_c \varepsilon_c}{\mu_f \varepsilon_f}\right]^{1/2} \tag{4.3}$$

In this case the light wave travels along the interface between the fiber and the cladding. When $\theta_i > \theta_{ic}$ there is no light wave propagating into the cladding. An exponential decay occurs for the electromagnetic fields associated with the light wave in the cladding. In this case the wave is said to be evanescent into the cladding. In nonmagnetic materials $\mu_c \approx \mu_f$; therefore, from Equation 4.3 one obtains for internal reflection to be possible the following condition:

$$\varepsilon_c < \varepsilon_F \tag{4.4}$$

Absorption of light energy occurs when light interacts with impurity molecules in the glass exciting a vibrational mode in the fiber. This may be described using the water molecule as an example. In Figure 4.4 an electromagnetic wave (light wave) is incident on a water molecule.

The electric and magnetic fields have a harmonic variation with time. Therefore, at one instant of time the field may be oriented as shown in the Figure 4.4, and at the next incident of time, oriented in the opposite direction. Since electric fields go from plus to minus, this may cause a motion of the

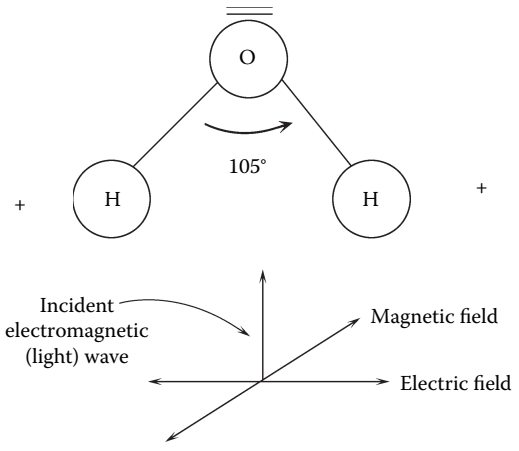


FIGURE 4.4
Interaction of a light wave with a water molecule.

H atoms in the water molecule. It is well known that a resonant "wagging" mode exists for the hydrogen atom in which the hydrogen atoms move back and forth at a characteristic resonant frequency. If the light energy is at that frequency, the light energy is then absorbed by the water molecule, causing the wagging mode to be excited. Molecular absorption typically occurs in the 10^{11} to 10^{13} cps frequency range. In addition to energy loss, this will lead to localized heating in the glass. Similar arguments may be made for other impurity atoms in the glass that occur at different frequencies. Therefore, in order to minimize absorption effects, the number of impurities in the fiber must be minimized.

In addition to leaking and absorption, losses also occur due to structural effects and voids. Ideally the silica fibers should be amorphous and have no void regions. However in the growth of the fiber heating and cooling occur resulting in the formation of polycrystallites. Polycrystallite boundaries and interfaces with the amorphous glass (silica) along with voids give rise to additional loss.

In order to overcome material losses repeaters are used at regular spacing intervals. The repeaters amplify the light waves, recovering some of the material losses. The principle of the repeaters is similar to that used in coaxial transmission lines, except that they operate at much higher frequencies.

4.5 Fiber Link and Materials

The cross section of an optical fiber is shown in Figure 4.5. The thickness and dielectric properties of the fiber and cladding often determine the

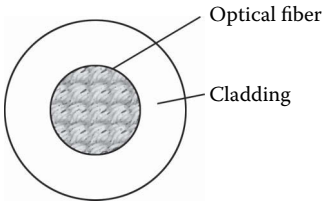


FIGURE 4.5
Cross section of an optical fiber.

specific application. Large fiber diameters relative to cladding thickness are good for illumination applications. On the other hand, for communication purposes the reverse is true: the optical fiber thickness is much smaller than the cladding thickness. The fiber may also have a graded dielectric constant, which is defined as a dielectric constant that changes linearly into cladding rather than a constant dielectric constant. This is often used to eliminate an abrupt boundary, which can result in aging problems. The actual diameter of the fiber may range from less than 100 microns to much larger sizes, depending upon the application. The normal material used in the cladding is glass (SiO_2), which has been doped with impurities so as to achieve a graded dielectric constant. In some cases plastics are used in order to obtain more flexibility in the fiber. Fiber optic links are often described as being similar to monofilament fishing line.

4.6 Communication Applications

Fiber optic links can be used to communicate both images and data. Each fiber can be used to carry one point of an image from one end of the fiber to the other. A bundle of fibers can then transmit a color image of a particular object over a long distance. In particular, the intensity and frequency of the received optical signals yield information concerning the color, size, and shape of the object. This technology is currently being exploited in an attempt to realize "picture" phones. The communication of data from one point to another can be represented as a block diagram (see Figure 4.6). When one uses many light sources operating at different frequencies and interfaces them to fibers and a receiver, a fiber optic communication system results.

Note that if the receiver and transmitter were both at the same end of the system, we would have an imaging system. In this case light is reflected from the probe area and returned to the transmitting location. This is illustrated in Figure 4.7. By analyzing the returned light one can obtain very

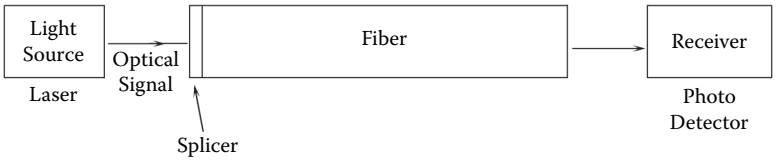


FIGURE 4.6
Block diagram of a fiber optic communication system.

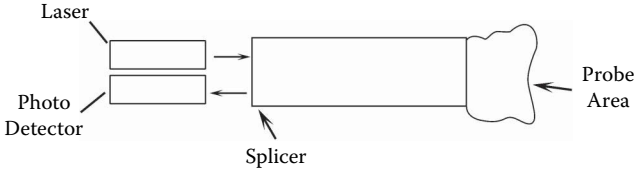


FIGURE 4.7
Example of a fiber optic imaging system.

useful information regarding the probe area. Fiber optic imaging systems are often used to image abnormalities in the human body. In the fiber optic communication system shown in Figure 4.6, large amounts of data can be transmitted from one point to another using bundles of individual fibers. In the system, however, alignment is critical with both the transmitter and receiver. Therefore, connectors and splicers must be carefully designed.

At this point it is appropriate to summarize the advantages of fiber optic systems:

- Carry a large amount of information over both short and long distances at high speeds
- Transmit both digital and analog signals
- Are noise-free and cheap if fiber optic links properly designed

Specific applications include:

- Land and sea long-haul communications (e.g., submarine cables, telephone networks, Internet transmission systems)
- Short-haul communications (e.g., local area networks [LANs])
- Picture phones
- Imaging of objects and abnormalities
- Use of fiber optics in treating diseases and performing surgery

In regard to the latter, treating diseases and performing microsurgery internally in the human body is possible by increasing the intensity of the light.

4.7 Fiber Optic Sensors

Previous sections have discussed how fiber optic technology has affected many different areas. However, the most logical question that one may ask is: "Is it possible to use fiber optic technology to realize real-world sensors?" The following sections will demonstrate that indeed fiber optic sensors can be realized. Previously, we have seen how modifications in current or resistance can be used in metal oxide semiconducting sensors, and modifications in acoustic waves can be used in acoustic wave sensors. In a likewise fashion, modifications in light can be used in fiber optic sensors. This becomes clear when specific applications are examined.

4.7.1 Fiber Optic Probes

Light impinging upon a candidate area or volume can yield very useful information. Fiber optic probes are often used in assembly line sensing. This is illustrated in Figure 4.8.

In a very simple sensing application, blockage or transmission of the light would indicate whether the assembly part is present or absent. The imaging of an object may also be obtained using a bundle of fibers. In this case

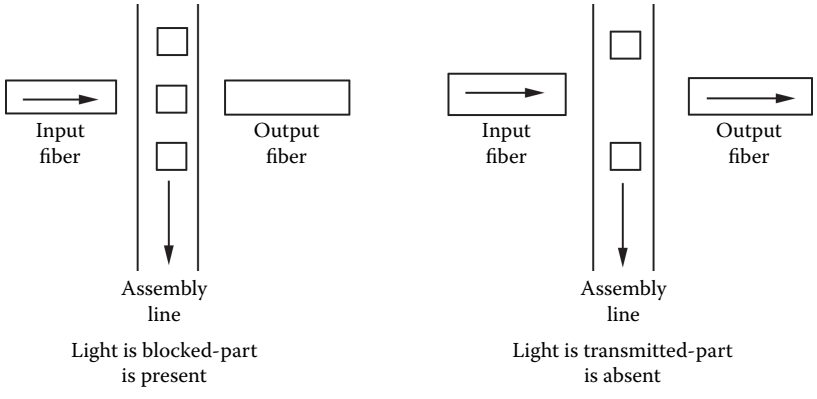


FIGURE 4.8
Assembly line application of a fiber optic sensor.

particular parts that do not meet specifications can be identified and removed from the production line.

Relative to sensors, the most important application is imaging within the human body. In this case the light reflects from the target region in the human body and is sent back to a receiver. The received light is then analyzed to determine features of the target region. If specific coatings are placed at the end of a fiber optic link, numerous other applications can be envisioned. In particular, internal physical parameters such as temperature and pressure may be determined. If the coatings were made to be specific to certain ions, molecules, or gases, internal sensing diagnostics could be performed. The target measurand would sorb the light, causing variations in reflected light signals, which can then be analyzed to determine the type of measurand within the target area in the human body. Since the individual fibers are so small, they may be used nondestructively in areas such as the stomach, heart, lung, and brain.

Fiber optic sensors are also very commonly used to obtain information about target regions in structures that are very difficult to access. For example, internal cracks and voids can be discovered by probing the area of interest using fiber optic probes. In this case, nondestructive evaluation of weaknesses in various types of structures can be uncovered. Another interesting application relates to a liquid-level sensor to determine levels of chemicals or any other type of fluid in a tank. A liquid-level sensor is shown in Figure 4.9. The actual sensor consists of an input and an output fiber interfaced with a prism. In Figure 4.9a, the light is totally reflected into the output fiber, whereas in the second case, shown in Figure 4.9b, the light is lost to the liquid. This is due to the fact that the liquid has different reflection characteristics than air. In the air case total internal reflection occurs, whereas the different value of the liquid dielectric constant causes the light to propagate into the liquid, consequently never reaching the output fiber.

Fiber optic sensors to determine accurate temperatures are yet another example. In this case ultraviolet (UV) light is transmitted in a fiber that has a glass blob at the end. The glass blob contains phosphor, and the light at different wavelengths stimulates fluorescence in the glass blob. The amount of fluorescence at different wavelengths changes with temperature. The fiber collects the light and transmits it to an optical analyzer that compares intensities at different wavelengths, and therefore measures temperature.

Another fiber optic application utilizes the fact that outside influences in the fiber can directly affect light transmission and reflection. In contrast to fiber links, which are used in communications in which the fibers are designed to be isolated from the environment, fibers can be designed to respond strongly to their environment. In this case one has a fiber optic sensor. The coatings on the fiber can be doped with materials that change their index of refraction with temperature, pressure, or other types of

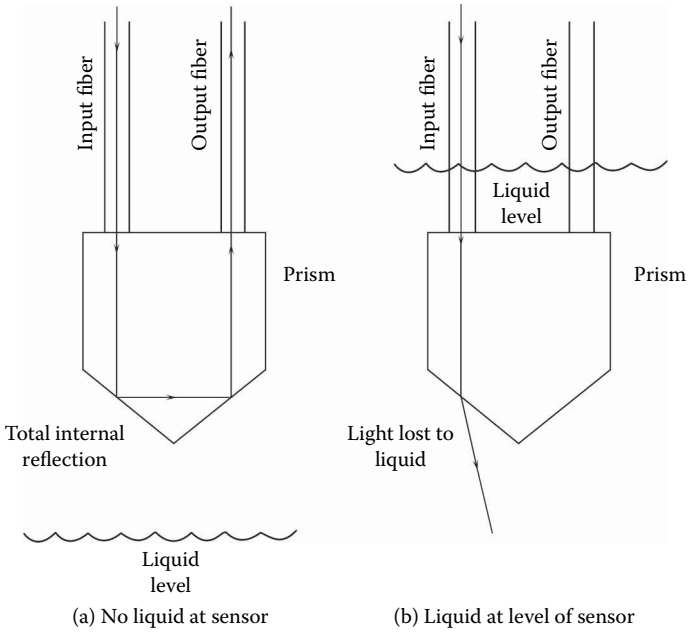


FIGURE 4.9
Illustration of a liquid-level sensor.

environments, like gases, etc. In this case fiber optic links can measure changes in light intensity, polarization, and phase.

Fiber optic links are also very cheap and can be installed within a structure when it is being built. For example, fibers imbedded in a bridge structure can be used to detect cracks. If the material remains intact, the fiber continues to transmit light. However, when a crack occurs the light is no longer transmitted and indicates that close examination of structural integrity is perhaps needed. Another example is a microbending sensor, which can detect degrees of bending. This is illustrated in Figure 4.10.

Under no bending the light output is high. However, when bending occurs the light output gets lower and, eventually, under heavy bending reaches a very low level.

Another example of a fiber optic sensor relates to a magnetic field sensor. The appropriate geometry is presented in Figure 4.11.

The unknown magnetic field may be in or out of the plane of the paper. Although it is possible to solve the out-of-plane problem, the present treatment relates to an in-plane unknown magnetic field. If the original light intensity is I_i and the output intensity is I_O, the ratio I_O/I_i for two different input light intensities will enable one to measure the unknown magnetic

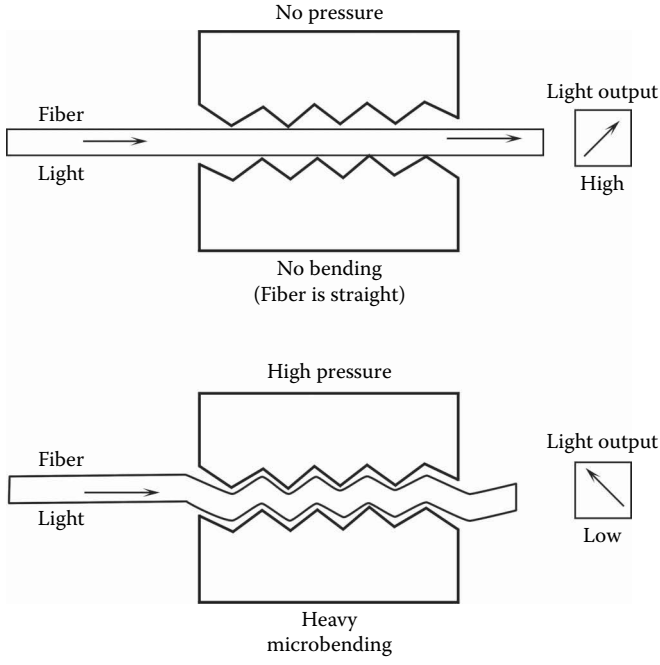


FIGURE 4.10
Microbending fiber optic sensor.

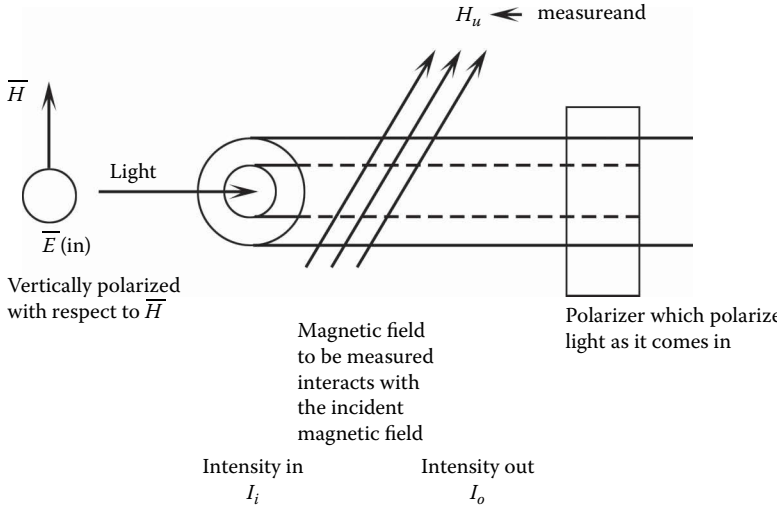


FIGURE 4.11
A magnetic field fiber optic sensor.

field magnitude and orientation. This may be explained as follows. The magnetic field, H, entering the polarizer is given as follows:

$$\bar{H} = \bar{H}_u + \bar{H}_i \tag{4.5}$$

where $\bar{H}_i$ = magnetic field associated with the incident light, and $\bar{H}_u$ = unknown magnetic field.

The polarizer allows only the vertical component of $\bar{H}$ (H_n) to pass through, which is given as follows:

$$\bar{H}_n = \left[\bar{H}_u + \bar{H}_i\right]\cos\theta \tag{4.6}$$

where θ = angle between the resultant $\bar{H}$ field in the fiber $\left(\vec{H}_u + \vec{H}_i\right)$ and $\vec{H}_i$.

The intensity of the light that passes out of the polarizer is related to the Poynting vector, $\bar{P}_O$, which is given as follows:

$$\bar{P}_O = \bar{E}_i x \bar{H}_n \tag{4.7}$$

The ratio of P_O/P_i is measured. Note that $\bar{P}_O$ and $\bar{P}_i$ are in the same direction. Using Equation 4.6, this yields the following relation:

$$R_i = \frac{P_O}{P_i} = \frac{E_i\left(H_u + H_i\right)\cos\theta}{E_i H_i} \tag{4.8}$$

For two different light intensities one obtains the following equations:

$$R_1 = \frac{E_1\left(H_u + H_1\right)}{E_1 H_1}\cos\theta = \left(\frac{H_u + H_1}{H_1}\right)\cos\theta \tag{4.9}$$

and

$$R_2 = \frac{E_2\left(H_u + H_2\right)}{E_2 H_2}\cos\theta = \left(\frac{H_u + H_2}{H_2}\right)\cos\theta \tag{4.10}$$

where E_1, H_1, E_2, H_2, R_1, and R_2 are known quantities.

The simultaneous solution of Equations 4.9 and 4.10 will yield values for H_u and θ, therefore determining the magnitude and direction of the unknown magnetic field.

Obviously one could describe an almost limitless number of applications for fiber optic sensors. One can refer to the literature to obtain information

on the most common sensors. One also could design a specific fiber optic sensor to detect a particular measurand.

Finally, one use of fiber optic sensors has led to the identification of structures with sensors called smart structures or smart skins. The former refers to the case when the sensor is embedded in a structure, while the latter refers to the sensor being on the surface or skin of a structure. These sensors can monitor the general health of the structure in a number of different ways. These include the following:

1. Monitor fabrication and curing of a composite
2. Monitor temperature to ascertain that curing conditions meet requirements
3. Monitor strain
4. Monitor cracks to identify serious stress problems
5. Use fiber optic sensors for monitoring structural integrity in aircraft wings, dams, buildings, bridges, ships, etc.

In conclusion, an obvious question one may ask is: "What type of sensor is better—metal oxide semiconductor sensor, acoustic sensor, or fiber optic sensor?" In order to answer that question intelligently, one must carefully examine sensor specifications such as sensitivity, selectivity, etc., peculiar to the particular measurand of interest. Only after an exhaustive study of the needs is made may one say one type of sensor is more suitable than another.

References

1. Wheeler, W. 1880. U.S. Patent 247,229.
2. Curtiss, L. 1993. *Understanding fiber optics*, 6. Upper Saddle River, NJ: Prentice-Hall.
3. Kao, K. C., and Hockham, G. A. 1966. *Proc. IEE* (London) 113:1151–58.

Questions

1. What is the primary loss mechanism in an optical fiber that has no bends?
2. What are the preferred materials for a fiber optic cable?
3. Using equations and figures, explain why the dielectric constant of the fiber optic cladding must be less than the dielectric constant of the fiber.

4. List four different applications for a fiber optic sensor.
5. Explain how a fiber optic fluid sensor works.
6. Fiber optic sensors are often used to determine inherent weaknesses in structure. Explain how these sensors can determine these weaknesses.
7. How is a polarizer used in a magnetic fiber optic sensor? Explain its operation.

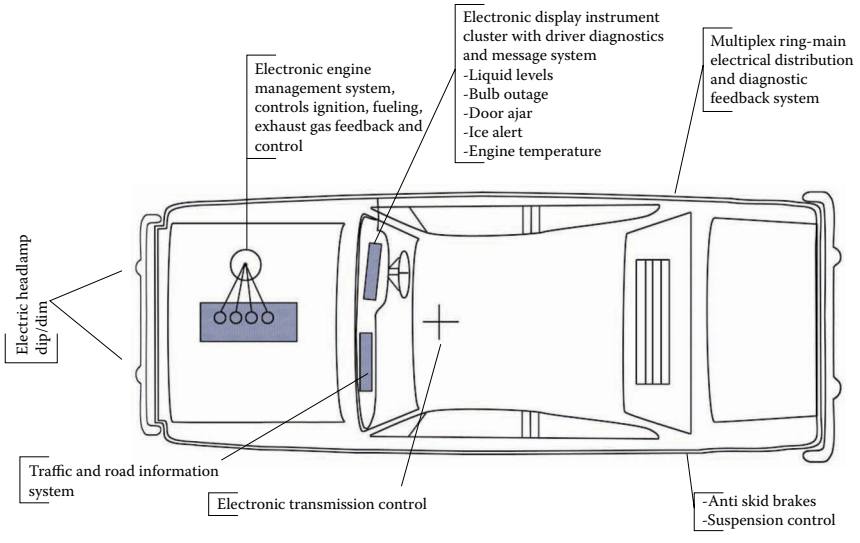


FIGURE 1.1
Areas where sensors can be utilized in the automobile. (From Westbrook, M. H., and Turner, J. D., *Automotive Sensors,* London: Institute of Physics Publishing, 1994, 9. With permission.)

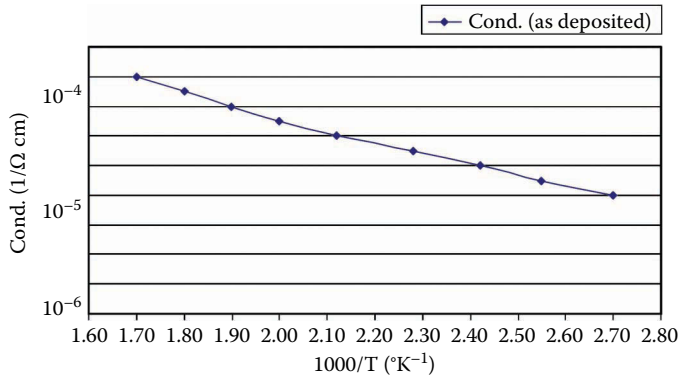


FIGURE 2.7
Typical conductivity results vs. 1,000/temperature (K^{-1}) for (as deposited)[2] 500 Å thick WO_3 film on $LiNbO_3$ in ambient air (humidity 50%). (Smith, D. J. 1992. Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications. MS thesis, Electrical Engineering, University of Maine.)

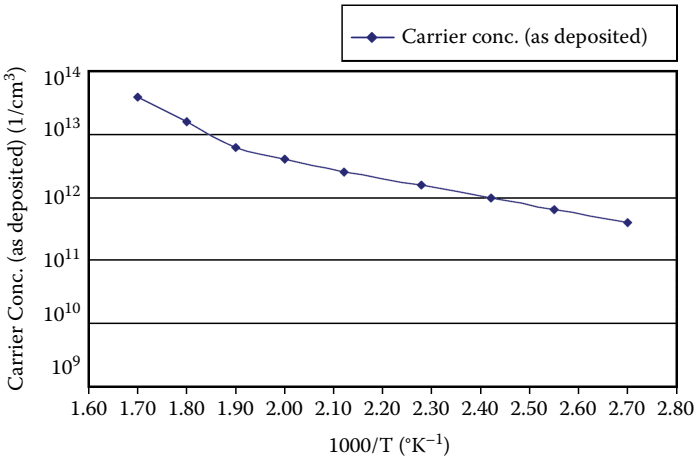


FIGURE 2.8
Typical electron carrier concentration results vs. 1,000/temperature (K^{-1}) for as deposited 500 Å thick WO_3 film on $LiNbO_3$ in ambient air (humidity 50%). (Smith, D. J. 1992. Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications. MS thesis, Electrical Engineering, University of Maine.)

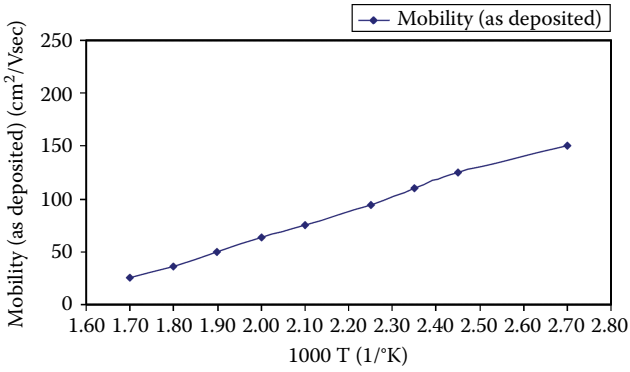


FIGURE 2.9
Typical carrier mobility results vs. 1.000/temperature (K^{-1}) for as deposited 500 Å thick WO_3 film on $LiNbO_3$ in ambient air (humidity 50%). (Smith, D. J. 1992. Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications. MS thesis, Electrical Engineering, University of Maine.)

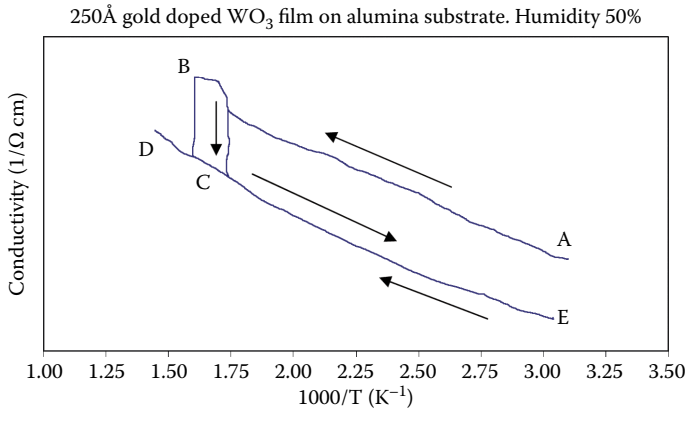


FIGURE 2.19
The electrical conductivity as a function of inverse temperature for a gold-doped 250 Å thick WO_3 film deposited on an alumina substrate. Arrows indicate the direction of increasing and decreasing temperature. The measurements were performed in air at 50% relative humidity. The reproducible measurement error was about 10%. (From Xu, Z., Electrical Properties of Thin Tungsten Trioxide Films, MS thesis, Electrical Engineering, University of Maine, 1990. With permission.)

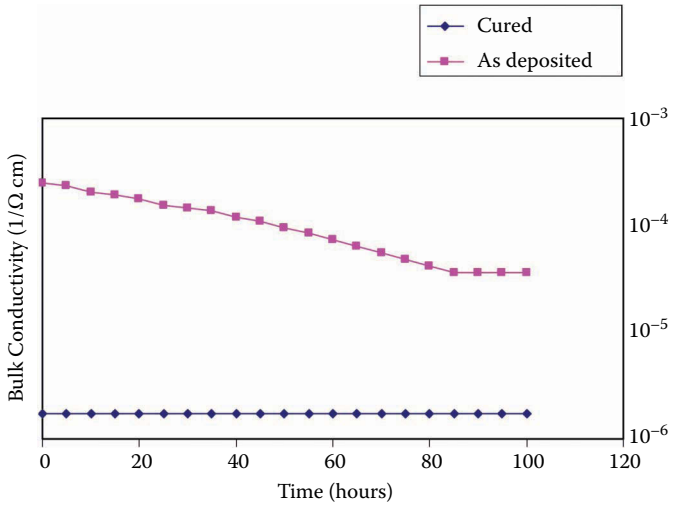


FIGURE 2.20
As deposited and cured WO_3 film conductivity variation vs. extended time at 200°C and relative humidity 20% in ambient air. (From Smith, D. J., Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications, MS thesis, Electrical Engineering, University of Maine, 1992. With permission.)

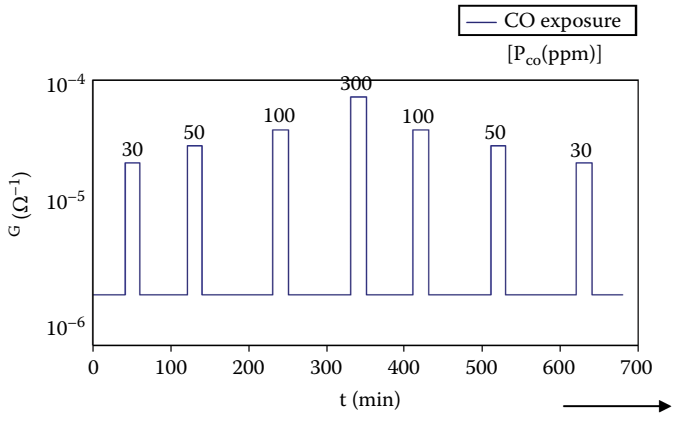


FIGURE 2.24
Response of a TGS 812 sensor (SnO_2) in air (50% humidity) to various CO exposures. (Reprinted from Schierbaum, K. D., Weimar, V., Kowalkowski, R., and Göpel, W., *Sensors Actuators* B3 [1991]: 205. With permission from Elsevier.)

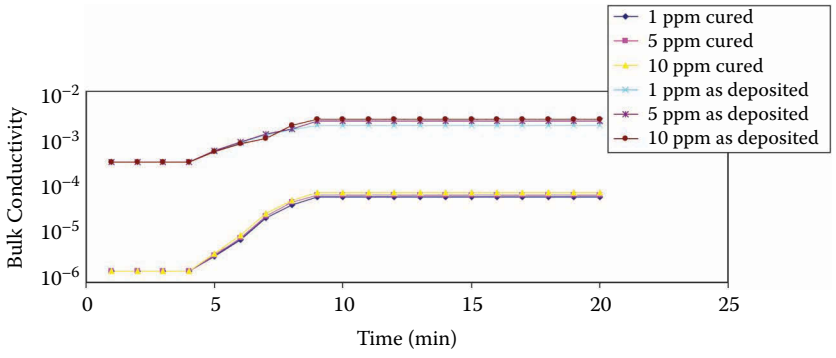


FIGURE 2.27
Bulk conductivity of an as deposited and cured WO_3 film sample vs. time for various H_2S gas concentrations (sample temperature 200°C and humidity 20%). (From Smith, D. J., Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications, MS thesis, Electrical Engineering, University of Maine, 1992.)

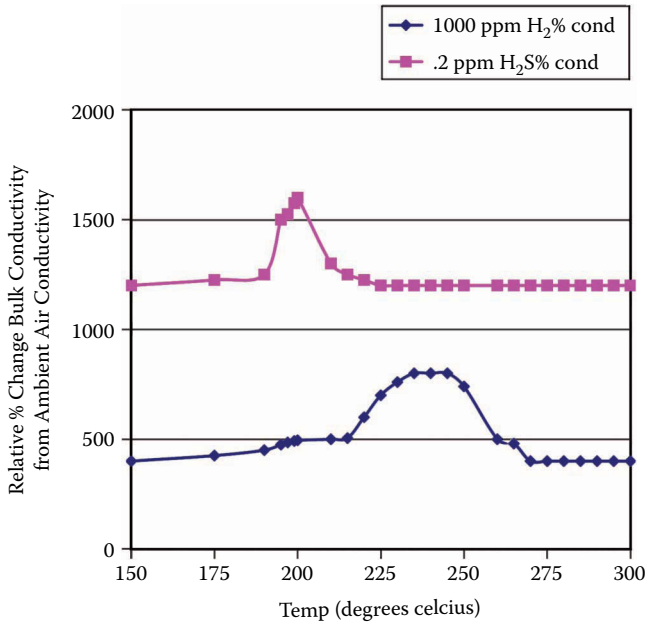


FIGURE 2.28
Average relative percent change in bulk conductivity of a WO_3 cured film from its ambient air conductivity vs. temperature for exposures of 0.2 ppm H_2S and 1,000 ppm H_2 (humidity 20%). (From Smith, D. J., Study of the Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications, MS thesis, Electrical Engineering, University of Maine, 1992.)

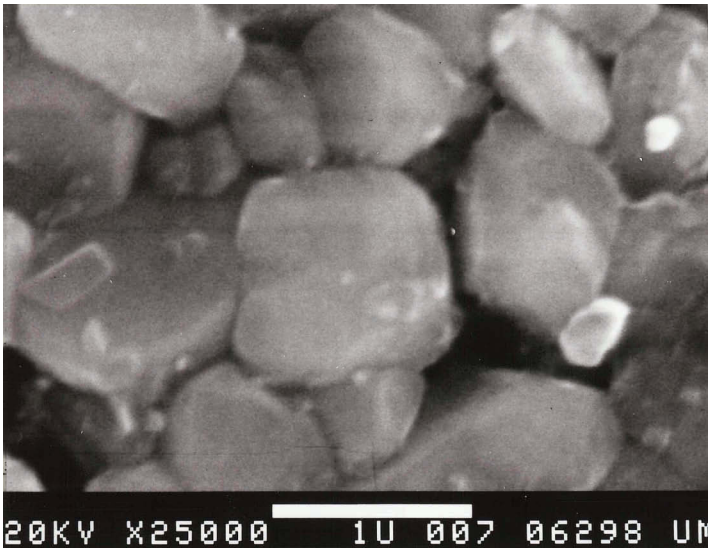


FIGURE 2.32
Scanning electron micrographs (SEMs) of an undoped polycrystalline WO_3 film on an alumina substrate. (From Xu, Z., Electrical Properties of Thin Tungsten Trioxide Films, MS thesis, Electrical Engineering, University of Maine, 1990.)

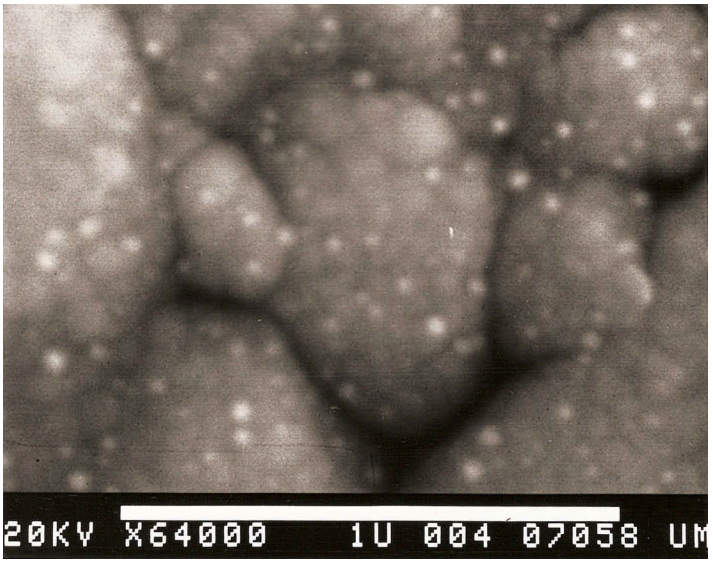


FIGURE 2.33
SEMs of a gold-doped polycrystalline WO_3 film on an alumina substrate. The little white dots are present on the finely divided gold particles. (From Xu, Z., Electrical Properties of Thin Tungsten Trioxide Films, MS thesis, Electrical Engineering, University of Maine, 1990.)

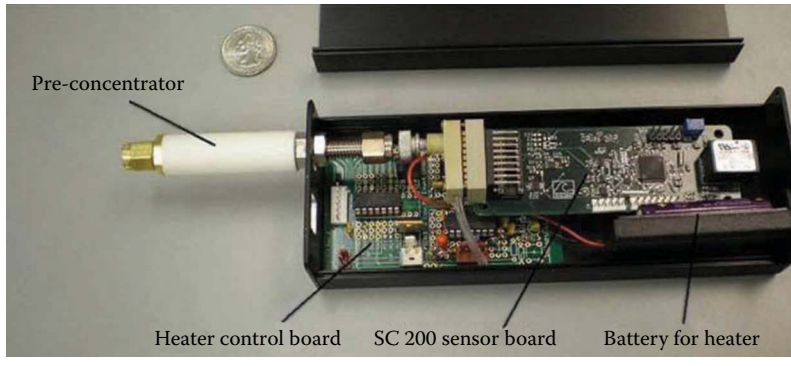


FIGURE 2.41
Photograph of an actual micromachine capacitative polymer gas sensor. (Reprinted from Mlsna, T.E., Cemalouic, S., Warburton, M., Hobson, S.T., Mlsna, D.A., and Petel, S.V. 2006. *Sensors and Actuators* B 116 pp. 192–201. With permission from Elsevier.)

(b)

FIGURE 3.13
(b) A force sensor cell: a commercial quartz force sensor. (From Kistler Instrumente AG, www.kistler.com. With permission.)

(b)

FIGURE 3.14
(b) Piezoelectric pressure sensor cell: commercial pressure sensor. (Kistler Instrumente AG, www.kistler.com. With permission.)

(b)

FIGURE 3.15
(b) A commercial model. (Kistler Instrumente AG, www.kistler.com. With permission.)

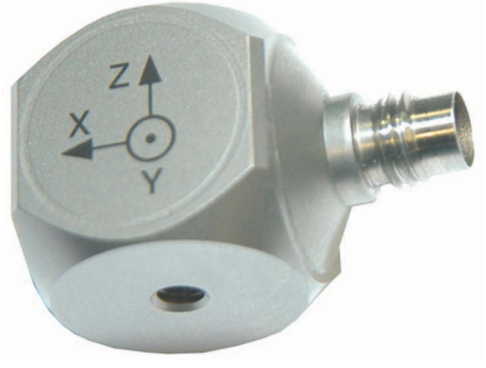
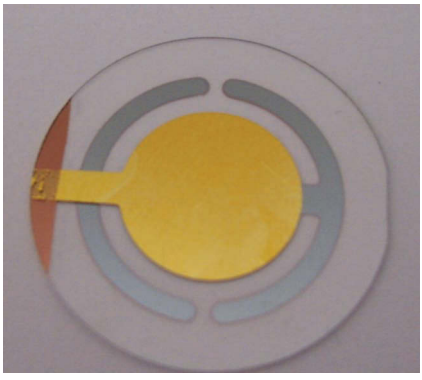
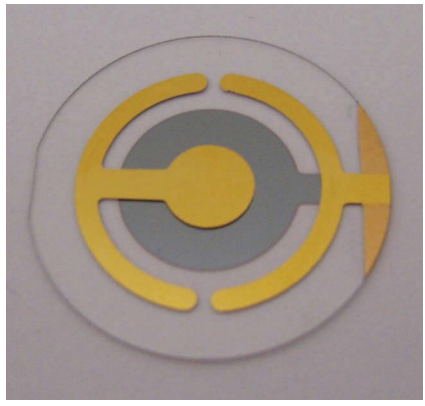


FIGURE 3.18
PCB Piezotronics Model 356A66 triaxial accelerometer for measuring acceleration in three dimensions. (Kistler Instrumente AG, www.kistler.com. With permission.)

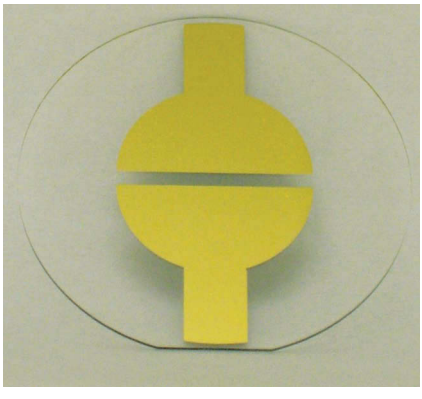


(c)

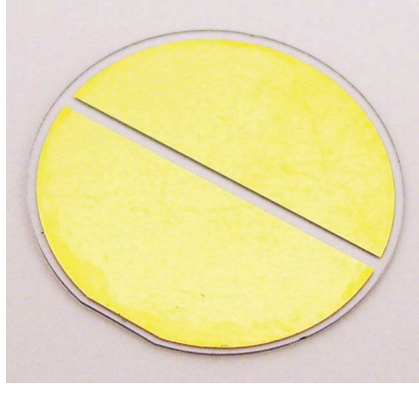


(d)

FIGURE 3.25
(c) photo of the top of the TFE device, and (d) photo of the bottom of the TFE device.



(c)



(d)

FIGURE 3.26
Electrode patterns for the bottom face of an LFE sensing platform: (c) photo of LFE bite-wing pattern, and (d) photo of LFE half-moon pattern.

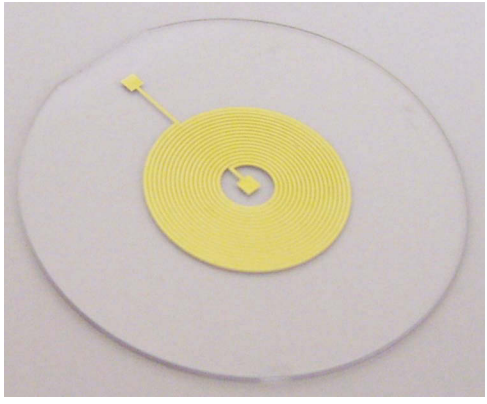


FIGURE 3.27
Monolithic spiral coil antenna configuration for the excitation of resonant acoustic waves in a piezoelectric crystal.

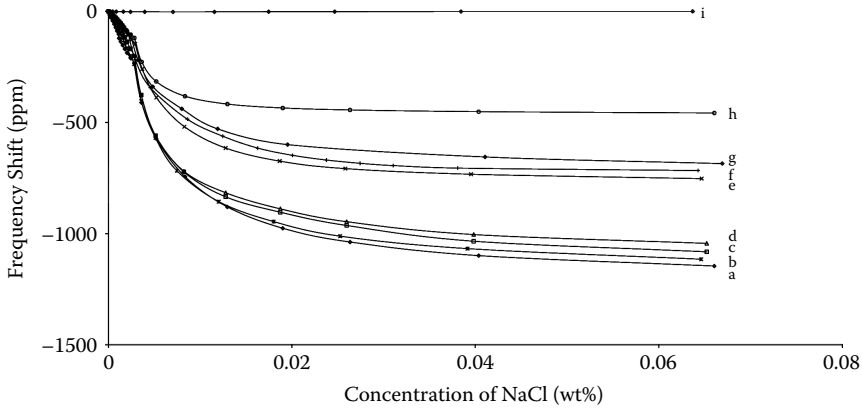


FIGURE 3.30
Frequency changes of the sensors as a function of NaCl concentration in water: (a) LFE sensor with electrode gap = 1.5 mm, (b) gap = 2.0 mm, (c) gap = 2.5 mm, (d) gap = 3.5 mm, (e) open-ring electrode, (f) small electrode, (g) closed-ring electrode, (h) gap = 12.7 mm, (i) standard QCM. (From Hu, Y., et al., *IEEE Trans. Ultrasonics Ferroelectrics Frequency Control* 51 [2004]: 1376–78. With permission.)

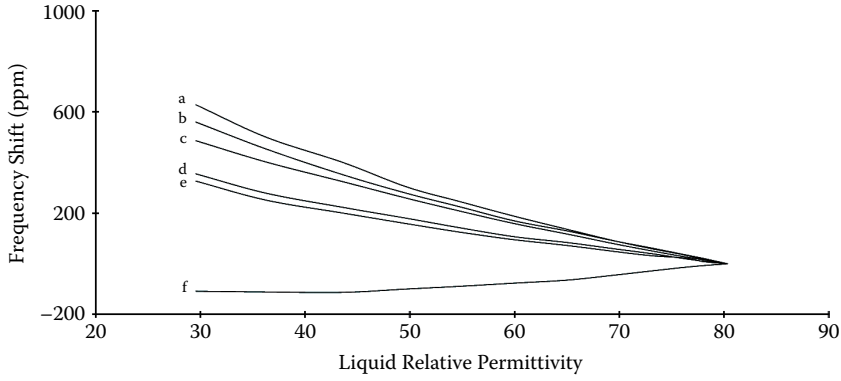


FIGURE 3.31
Frequency changes of the sensors as a function of the liquid relative permittivity of the 2-propanol solution: (a) LFE sensor with gap = 1.5 mm, (b) LFE sensor with gap = 2.0 mm, (c) LFE sensor with gap = 3.5 mm, (d) open-ring electrode geometry, (e) small electrode geometry, (f) standard QCM. (From Hu, Y., et al., *IEEE Trans. Ultrasonics Ferroelectrics Frequency Control* 51 [2004]: 1376–78. With permission.)

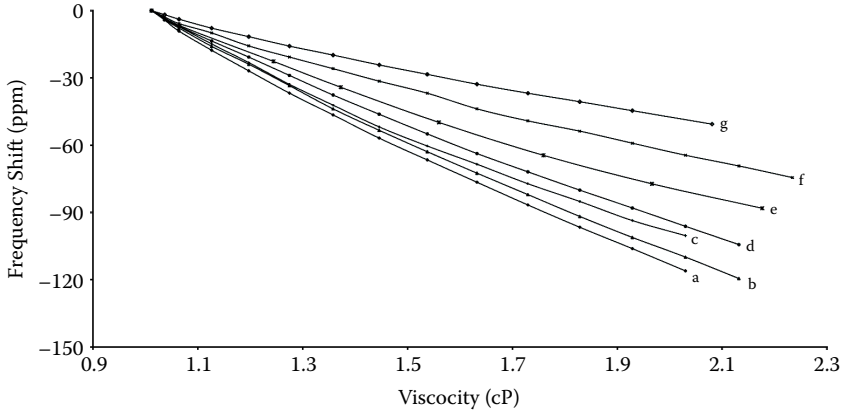


FIGURE 3.32
Frequency shifts of the sensors as a function of a water-glycerol solution: (a) LFE sensor with electrode gap = 1.5 mm, (b) gap = 2.0 mm, (c) gap = 3.5 mm, (d) gap = 12.7 mm, (e) open-ring electrode, (f) small electrode, (g) standard QCM. (From Hu, Y., et al., *IEEE Trans. Ultrasonics Ferroelectrics Frequency Control* 51 [2004]: 1376–78. With permission.)

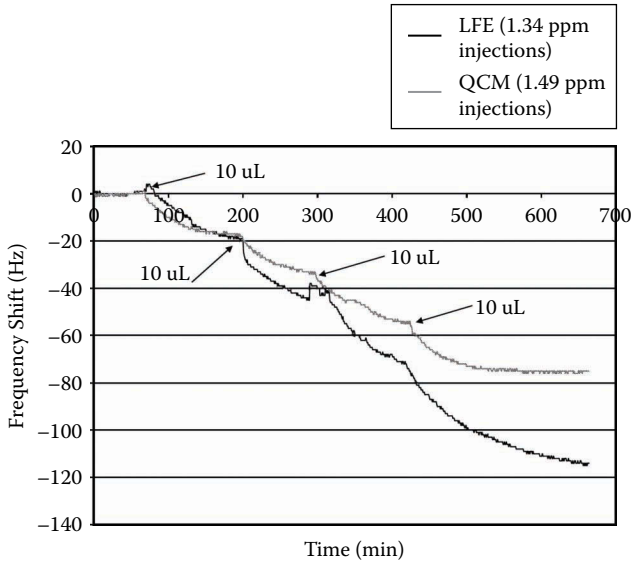


FIGURE 3.33
Frequency shift from baseline due to phosmet additions. (Reprinted from Hu, Y., et al., *Sensors and Actuators B* 108 [2005]: 910–16. With permission from Elsevier.)

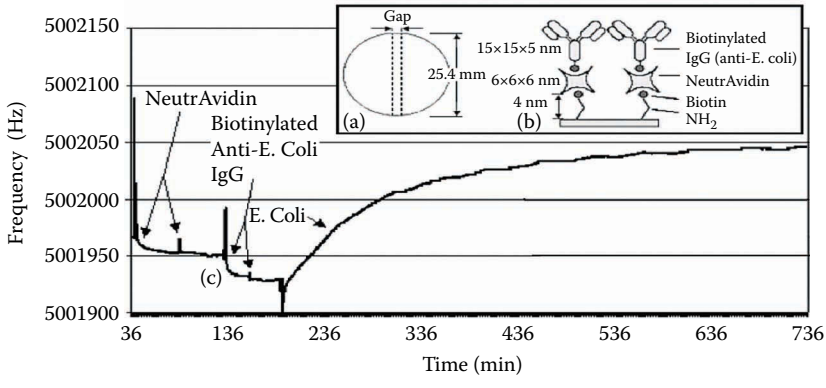


FIGURE 3.35
(a) LFE sensor: Top surface is bare and the dotted lines indicate the bottom surface electrode configuration. (b) Schematic of the antibody film on the sensing surface. (c) Frequency response of the LFE to neutrAvidin, anti-*E. coli* O157:H7 and *E. coli* O157:H7 bacteria. (From McCann, D.F., Wark, M.S., French, L.A., and Vetelino, J.F. 2008. Novel transducer configurations for bulk acoustic wave sensors. 2008 *IEEE Conference*, Lecce, Italy, pp. 1448–1451, Oct. 26–29. © 2008 IEEE. With permission.)

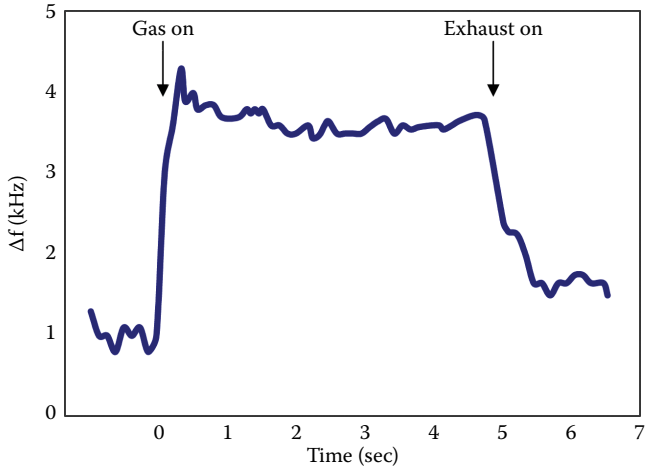


FIGURE 3.39
A typical SAW sensor response to H_2S. The gas concentration for this exposure was 0.25 ppm, and the device operating temperature was 98°C.

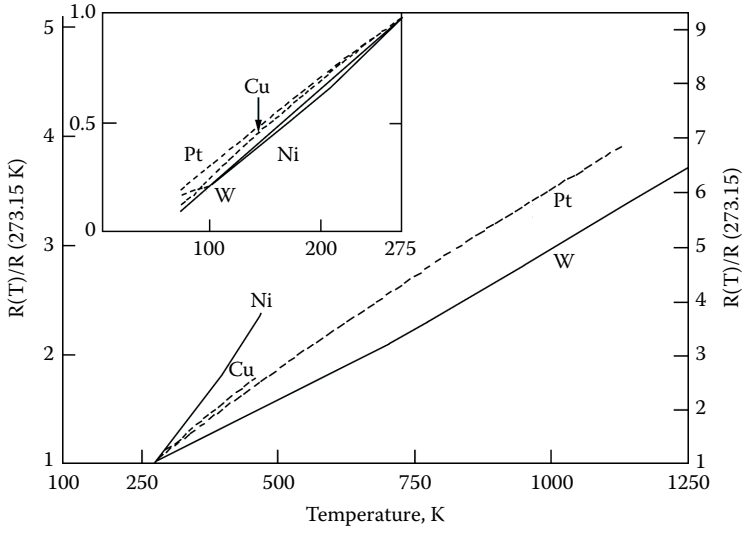


FIGURE 5.1
Resistance vs. temperature of Cu, Ni, Pt, and W. (From Gopal, W., Hess, J., and Zemel, J. N., Sensors, a comprehensive survey. *Thermal sensors*, Vol. 4. 1990. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

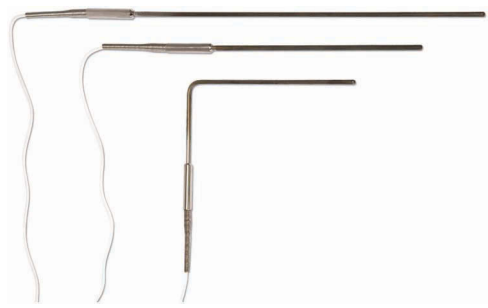


FIGURE 5.3
Long-stem standard platinum resistance thermometer. (From Fluke Secondary PRT with Calibration Options—5608 and 5609. With permission.)

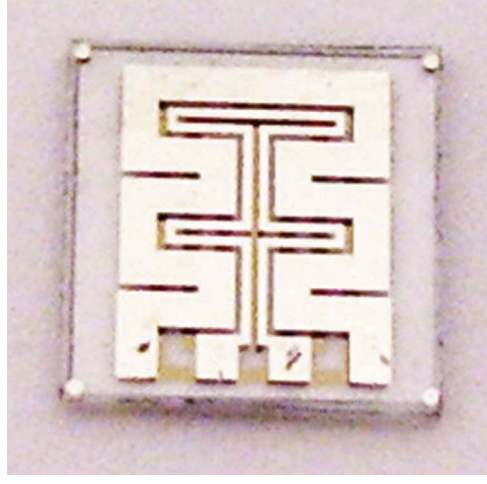


FIGURE 5.4
(b) Actual photo of the heater.

5

Thermal Sensors

5.1 Introduction

The most popular sensor in terms of market surveys and sales is the temperature sensor. This sensor is used in a wide range of applications critical to the medical, agricultural, industrial, and automotive communities, to name just a few. The particular application may range from measuring temperature within a furnace where accuracies of 10°C may be sufficient to the measurement of temperature associated with tumors, where accuracies of less than .01°C are necessary.

The different variety of temperature sensors cover a wide range, the most popular being the resistance thermometer, thermistors, thermocouples, and radiation thermometers. Other, less common varieties include acoustic thermometers, dielectric constant thermometers, refractive index thermometers, and quartz crystal thermometers. The obvious thermometer missing from this list is the mercury thermometer, which has been banned. This type of thermometer has generally not been used significantly for the accurate measurement of temperature. Each of these thermometers depends upon a particular technology for its operation. These technologies are typically based on such broad areas as solid state, electromagnetics, acoustics, and optics. Recent advances in microfabrication technology have enabled the realization of miniature temperature sensors that are extremely sensitive and stable.

5.2 Resistance Thermometers

This type of thermometer is the oldest and was first introduced by Ohm in 1827. He realized that the resistance of metals increased in a quasi-linear fashion with temperature. The first commercially available thermometer based on the variation of metal resistance with temperature appeared in 1871, when Von Siemens developed a pyrometer for furnaces by winding a .1 mm platinum wire on porcelain inside an iron tube. As a result of this work, it was

TABLE 5.1

Self-Heating Effect in Various Resistance Thermometers[a]

Type	Temperature Range (°K)	Resistance Ω	Medium
Platinum: Ceramic encased	250–350	100 (0°C)	Still air
Platinum (SPRT): Long stem	250–350	25.5 (0°C)	Still air
Thermistor: Glass enclosed	220–550	104 (25°C)	Still air
Thermistor: Glass enclosed	220–550	104 (25°C)	Still water
Germanium: Copper encased	10–30		Metal block
Germanium: Copper encased	<2		Metal block

Source: Gopal, W., Hesse, J., and Zemel, J. N., Sensors, a comprehensive survey. *Thermal sensors*, Vol. 4. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

[a] The given values are indicative of the behavior of different resistance thermometers. The manufacturers provide more accurate values for particular types.

realized that platinum was an excellent metal to use for a resistance thermometer. Callendar, in 1877, formalized the platinum resistance thermometer (PRT) as a stand-alone instrument to be used in a Wheatstone Bridge configuration. Platinum remained the most popular metal for the resistance thermometer in the 1900s and to this day remains the most popular. Other metals were suggested in the early 1930s. These include nickel, UO_2, $MgTiO_3$, germanium, rhodium/iron, copper, and iridium. Each had particular applications. For example, germanium and rhodium/iron were used for low-temperature (less than 20° K) applications; nickel, UO_2, and $MgTiO_3$ were used as cheap alternatives to platinum; copper was used in oceanographic studies; and iridium was used for surface temperature applications in industrial processes. Many of these resistance thermometers operated over a variety of temperature ranges, as shown in Table 5.1. The most versatile in terms of accuracy, linearity, and stability, however, was the PRT.

5.3 Theory of Metal-Based Thermometers

Experimental measurements clearly show that the resistance thermometers that use metals display an increase in resistance as temperature increases. This behavior is shown in Figure 5.1 for copper, nickel, platinum, and tungsten. It is convenient to define the temperature dependence of these metals by the temperature coefficient of resistance:

$$TCR = \frac{1}{R}\frac{dR}{dT} \tag{5.1}$$

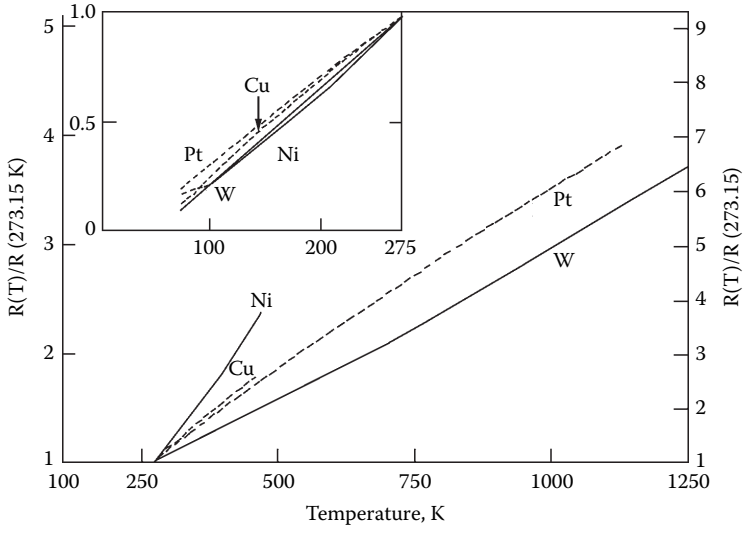


FIGURE 5.1
(Please see color insert following page 146) Resistance vs. temperature of Cu, Ni, Pt, and W. (From Gopal, W., Hess, J., and Zemel, J. N., Sensors, a comprehensive survey. *Thermal sensors*, Vol. 4. 1990. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

where R = resistance in ohms, and T = temperature in °K.

For the data shown in Figure 5.1, TCR ranges between 3.9×10^{-3} and $6.5 \times 10^{-10}\,K^{-1}$. Generally it is observed that the behavior of the resistance is quasi-linear in this range.

In order to explain why the resistance in metals increases with temperature, the metal must be modeled at the microscopic level. Basically the metal consists of ion cores immersed in a sea of electrons. The number of free electrons in the metal does not vary with temperature and is a constant. The only parameters that can vary with temperature are the motion of the electrons and the vibrations of the ion cores. As temperature increases, the electron motion increases and the electrons collide with themselves and the vibrating ion cores. Since the electrons are much smaller than the ion cores, one can neglect electron-electron collisions. The major contribution is therefore due to the electrons colliding with the vibrating ion cores. This is illustrated in Figure 5.2, where a simplified two-dimensional lattice is presented that shows the areas occupied by static ion cores (solid) and vibrating ion cores (dotted). One can clearly see that the increase in area occupied by the vibrating cores with increasing temperatures increases the probability of collisions. Mathematically the electrical conductivity of a metal sample may be expressed as follows:

$$\sigma = ne\mu \tag{5.2}$$

where n = electron density, e = electronic charge, and μ = electron mobility.

FIGURE 5.2
Two-dimensional lattice: solid circles, lattice at rest; dashed circles, vibrating lattice.

The electron mobility is defined as follows:

$$\mu = \frac{e\tau}{m} \tag{5.3}$$

where τ = average time between collisions, and m = electron mass.

Noting that the resistivity, ρ, of a metal is the inverse of the conductivity, one obtains, using Equations 5.2 and 5.3, the following relation:

$$\rho = \frac{1}{\sigma} = \frac{m}{ne^2}\left(\frac{1}{\tau}\right) \tag{5.4}$$

In a pure intrinsic material with no defects, the only collisions (excluding electron-electron collisions) occur between electrons and vibrating ion cores. However, defects such as vacancies, impurity atoms, dislocations, and grain boundaries also occur in a real material. These defects add to the resistance of the metal and can be incorporated into τ as follows:

$$\frac{1}{\tau} = \sum_{i=1}^{n} \frac{1}{\tau_i} \tag{5.5}$$

where n indicates the number of different types of scattering processes. Theoretically, the evaluation of each of the τ_i terms is a complicated process requiring knowledge of the allowed vibrational spectrum of the metal along with the contribution introduced by the defects.

5.4 Properties of Metal-Based Thermometers

The most common material used in metal thermometers is platinum. PRTs typically have a resistance range between .1 Ω and 1 KΩ. Although platinum is expensive, new processing techniques have made it very attractive. In particular, platinum can be drawn into very thin wires of high purity. The wires can be less than .05 mm in diameter and the purity greater than 99.99%. Also, platinum can be easily deposited using high-vacuum deposition or radio frequency (RF) magnetron sputtering. This results in much smaller volume and mass for a given resistance, which is critical in forming thick and thin film sensors. The thermal expansion of platinum is also well matched to insulating substrates such as silica, alumina, and sapphire. This results in minimum strain at the platinum-substrate interface, hence minimizing the possibility of the film disadhering with increasing temperatures.

Depending upon the particular application, platinum thermometers can exist in a variety of geometries, such as long wires, coiled wires, serpentine films, etc. An example of one such geometry is the long-stem thermometer shown in Figure 5.3 [2]. Researchers in the Laboratory for Surface Science and Technology at the University of Maine have developed a serpentine platinum heater that is used in semiconducting gas sensors. In this sensor, a gas-sensitive oxide film is deposited on a sapphire substrate and a serpentine platinum film is deposited on the back side of a sapphire substrate. When a voltage is applied to the serpentine platinum film it heats the substrate to

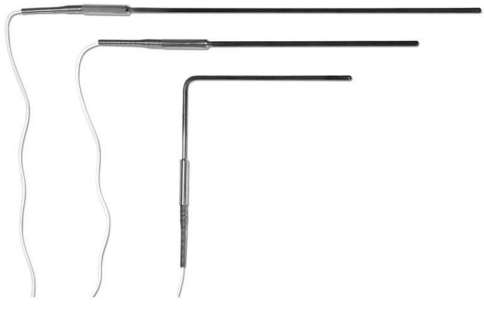


FIGURE 5.3
(Please see color insert following page 146) Long-stem standard platinum resistance thermometer. (From Fluke Secondary PRT with Calibration Options—5608 and 5609. With permission.)

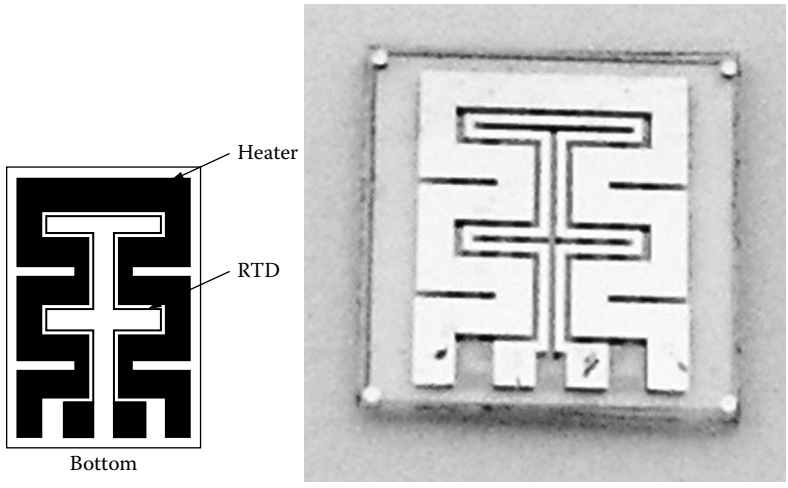


FIGURE 5.4
(Please see color insert following page 146) (a) Serpentine platinum heater developed in the Laboratory for Surface Science and Technology at the University of Maine, Orono. Shaded area, platinum; unshaded area, sapphire. (b) Actual photo of the heater.

a temperature that is appropriate for the target gas-film interactions to take place. A diagram and photo of this heater are shown in Figure 5.4.

Other, less popular metal resistance thermometers use copper (Cu), iridium (Ir), or combinations of rhodium and cobalt (Rh-Co) or platinum and cobalt (Pt-Co). The Cu thermometer is the most linear, but it has a limited temperature range (0 to 100°C). The Ir thermometer provides an excellent match to aluminum substrates and is normally used as a thin film in surface temperature measurements. Rh-Co and Pt-Co are used primarily for low-temperature (.5 to 30 K) applications.

5.5 Theory of Semiconductor-Based Thermometers

Thermometers that use semiconductors as the sensing element differ from metal-based thermometers in that the resistance decreases with increasing temperature. Also, in contrast to metals, the free electron density in semiconductors is not constant but increases with temperature. Also, semiconductors have vacancies called holes, whose concentration also increases with temperature. In an intrinsic (pure) semiconductor the electron and hole densities (n and p) can be represented as follows:

$$n = 2\left(\frac{2\pi m_e kT}{h^2}\right)^{3/2} e^{(E_f - E_c)/kT} \tag{5.6}$$

and

$$\mathrm{p} = 2\left(\frac{2\pi m_p kT}{h^2}\right)^{3/2} e^{-(E_f - E_v)/kT} \tag{5.7}$$

where k = Boltzman's constant, h = Planck's constant, E_f = Fermi level, E_c = energy level at the bottom of the conduction band, and E_v = energy level at the top of the valence band.

The resulting conductivity is then given by the following equation:

$$\sigma = e\left(n\mu_n + p\mu_p\right) \tag{5.8}$$

where μ_n and μ_p are the mobilities of the electrons and holes, respectively. Using Equation 5.3, the conductivity may be represented as follows:

$$\sigma = e^2\left(\frac{n\tau_n}{m_n} + \frac{p\tau_p}{m_p}\right) \tag{5.9}$$

where τ_n and τ_p are the average time between collisions for electrons and holes, respectively. In an intrinsic semiconductor the number of electrons is equal to the number of holes, and hence

$$n = p = n_i \tag{5.10}$$

where n_i = intrinsic carrier density.

Noting from Equation 5.10 that $n_i = \sqrt{np}$, substituting Equations 5.6 and 5.7 into Equation 5.9 one obtains

$$\sigma = 2e\left(\frac{2\pi kT}{h^2}\right)^{3/2}\left(m_n m_p\right)^{3/4}\left(\frac{e\tau_n}{m_n} + \frac{e\tau_p}{m_p}\right)e^{-E_g/2kT} \tag{5.11}$$

where

$$E_g = E_C - E_V$$

The major temperature-dependent term is the exponential, which is implicit in the equations for n and p (see Equations 5.6 and 5.7). τ_n and τ_p behave

roughly as $T^{-3/2}$, which cancels the $T^{3/2}$ term in Equation 5.11. As a result one can approximate Equation 5.11 as follows:

$$\sigma = \left(1/A\right)e^{-B/T}$$

or (5.12)

$$\rho = Ae^{B/T}$$

Examination of Equation 5.12 reveals that as temperature increases the resistance decreases.

Semiconductors can also be doped with either donors or acceptors. In the case of the former, the electron density increases significantly, resulting in an *n*-type semiconductor, whereas in the latter, the hole density increases, resulting in a *p*-type semiconductor. In the doped semiconductor the carrier concentrations increase with temperature, resulting in a behavior similar to that observed for the intrinsic semiconductor.

Thermometers that use semiconductors are thermally sensitive resistors, or what is commonly called thermistors. The semiconductors used in thermistors are not the common semiconductors such as silicon or gallium arsenide. Typically thermistors use metal oxides or combinations of metal oxides. These materials typically have oxygen vacancies resulting in compounds such as MO_{n-x}. When $x = 0$ the oxide is said to be stoichiometric; therefore, for each metal ion, *M*, there are *n* oxygen ions. However, normally x is not exactly zero but a very small number, resulting in $n - x$ oxygen atoms on an average associated with each metal ion. For example, in the case of tungsten trioxide, *n* would be 3 and x roughly about .001 or less. In this case, the tungsten ions without three oxygen atoms would have very loosely bound electrons, which upon application of temperature become free, resulting in the metal oxide being *n*-type.

Examples of metal oxides that are used in thermistors are the oxides of tungsten manganese, nickel, cobalt, iron, copper, lithium, magnesium, and chromium. The most stable mixed metal oxides are Mn-Ni and Mn-Ni-Co.

5.6 Thermistor Properties

Thermistors typically have resistance between 10 Ω and 100 MΩ. Due to the large resistance range, thermistors provide excellent sensitivity and are also very stable, particularly between 100 and 300°C. Thermistors can also be manufactured very cheaply. The two commonly used thermistors are the bead type and the disc type. In the bead type one starts with a platinum

wire that is drawn or stretched. Small blobs of metal oxide are then attached with a binder at intervals along the wire. The wire is then sintered at about 1,300°C to produce a series of thermistors with embedded leads. The platinum wire is then cut on each side of the metal oxide thermistor, and the resulting individual thermistors are coated with a protective coating (glass). This results in a large number of thermistors with embedded leads that can be used as temperature sensors. In the disc type thermistor the metal oxide in powder form is compressed and heated to about 1,100°C to form a disc. Silver is then deposited on the opposite sides of the disc by spraying, screen printing, vacuum deposition, or RF magnetron sputtering. The disc can then be diced to form several individual thermistors. Comparing the two techniques, the bead type thermistors are the most durable and the cheapest to produce. Finally, for applications greater than 1,000°C thermistors made with zirconium oxide are the most popular. Zirconium is unique in its ability to withstand high temperatures.

Another application for thermistors is in heat switches. Often at a particular temperature it is necessary to switch or truncate some ongoing process. Thermistors that exhibit large resistance changes over a very small temperature range can be used in these applications. Barium/strontium/titanate is an excellent oxide for switches since it can change its resistance by three orders of magnitude over a few tenths of a degree centigrade in the temperature range from 15° to 115°C. This resistance normally is related to ferroelectricity, and the temperature at which it occurs is related to the barium/strontium ratio. Hence, one can configure a heat switch anywhere in the range of 15 to 115°C by a judicious choice of the Ba/Sr ratio.

5.7 Concluding Remarks

In conclusion, a significant amount of research has been done on thermal sensors. Microfabrication techniques such as RF magnetron sputtering and photolithography have resulted in the realization of very small, accurate, and stable temperature sensors. The temperature range for these sensors is from about 1° K to over 1,200°C. Also, the sensors can be used either as temperature sensors or heat switches. Many of the temperature sensors can be purchased off the shelf or can be readily designed by the user for a particular application.

References

1. Gopal, W., Hess, J., and Zemel, J. N. 1990. Sensors, A Comprehensive Survey. *Thermal sensors*. Vol. 4.
2. Fluke secondary PRT with calibration options—5608 and 5609.

Questions

1. What is an RTD? Explain why platinum is often the material of choice in an RTD.
2. In an RTD, plot the variation of resistance with temperature and theoretically justify this variation.
3. List three different materials that are used in an RTD and explain the positive features of each one.
4. a. How does a thermistor work?
   b. Two different thermistors have different variations with temperature. Explain theoretically how this happens.
5. a. What material is used in a heat switch, and why?
   b. Show how heat switches can be used in a conventional oven. Plot a graph of resistance vs. temperature for the heat switch when you need to operate the oven at 24°C.
6. For an intrinsic semiconductor the electron and hole densities (n and p) can be represented as

$$n = 2\left(\frac{2\pi m_e kT}{h^2}\right)^{3/2} e^{(E_f - E_c)/kT}$$

$$p = 2\left(\frac{2\pi m_p kT}{h^2}\right)^{3/2} e^{-(E_f - E_v)/kT}$$

   where k = Boltzman's constant, h = Planck's constant, E_f = Fermi level, E_c = energy level at the bottom of the conduction band, and E_v = energy level at the top of the valence band.

Show that resulting conductivity may be expressed as

$$\sigma = \left[2e\left(\frac{2\pi kT}{h^2} \right)^{3/2} \left(m_n m_p \right)^{3/4} \left(\frac{e\tau_n}{m_n} + \frac{e\tau_p}{m_p} \right) \right] e^{\left(-E_g / 2kT \right)}$$

where

$$E_g = E_c - E_v$$

6

Magnetic Sensors

6.1 Introduction

Perhaps one of the first sensors used by man was a compass, which is essentially a magnetic sensor. The compass needle, which is steel, aligns itself with the earth's magnetic field lines and points toward the north. The compass became very important from about 1200 A.D. onward and was critical in guiding many explorers on both land and sea. At that time the production of the magnetic material (steel) suitable for compass needles was kept a close secret.

Christopher Columbus in his many transatlantic voyages was known to have a compass to guide his ships. Even prior to the compass, wildlife such as birds and fish have used the earth's magnetic field to guide them in their migration patterns. Although little is known about the exact internal sensor inherent in birds and fish such as geese, salmon, and eels, their migration patterns have been known to fluctuate due to subtle changes in the earth's magnetic field. In regard to Atlantic salmon, it has been postulated that they can sense not only the direction but also the magnitude of the earth's magnetic field, which allows them to return to their place of birth.

In 1820 Oersted discovered that a current-carrying wire deflected a compass needle, essentially initiating the age of electromagnetism. Oersted's work was followed by several fundamental laws of electromagnetism, such as the Biot-Savart law, Faraday's law of induction, Gauss's law, and Ampere's law. In 1862 these observed physical laws were put together by Maxwell and resulted in what is commonly called Maxwell's equation. These equations basically form the backbone of electrical engineering in that all the fundamental circuit and power relations along with the theory behind light waves and microwaves can be obtained and explained by Maxwell's equations. The formalization of the area of electromagnetics soon resulted in the development of a wide range of sensors associated with the laws and effects of magnetic or electromagnetic fields.

6.2 Natural and Man-Made Magnetic Fields

The most common natural magnetic field is the earth's, which varies depending upon where you are located. For example, the magnitude of the earth's magnetic field is about 0.5 A/cm at the poles and 0.25 A/cm at the equator. With the aid of sensors placed in satellites, we can very accurately map the exact direction and magnitude of the earth's magnetic field. On a much smaller scale humans are also known to produce small magnetic fields. These fields are caused by microcurrents in the heart, brain, and muscle tissue. The fields can be determined by attaching electrodes to the skin and measuring the voltage drop. This technique forms the basis for the electrocardiogram (ECG).

Magnetic fields may also be produced by man. For example, sources such as overhead electrical transmission lines and transformers can produce magnetic fields ranging from 0.1 to 10^4 A/cm. These fields no doubt have an effect on animals and human beings living near power lines since electrical currents exist in both. As to whether these effects are dangerous to health is a matter that has been debated by scientists for several years. Higher magnetic fields in the range of 10^2 to 10^5 A/cm can also be created using conventional or superconducting coils. The range of possible natural and man-made magnetic fields is presented in Figure 6.1 [1].

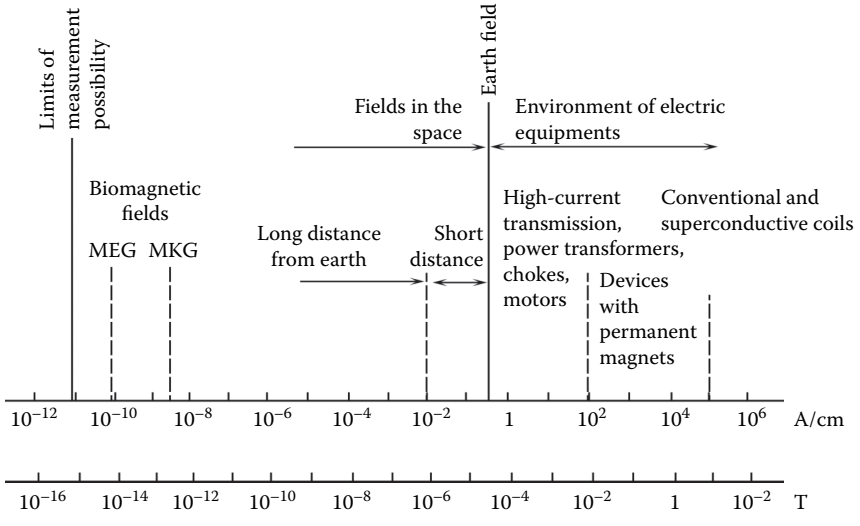


FIGURE 6.1
Scale of magnetic field strength and flux density (From Gopal, W., Hesse, J., and Zemel, J.N. 1996. Sensors, a comprehensive survey, *Magnetic Sensors*, Vol. 5. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

6.3 Materials Used in Magnetic Sensors

Magnetic sensors use both nonmagnetic and magnetic materials. The most common nonmagnetic materials in use are semiconductors, which include silicon, GaAs, and other lesser known group II-VI and III-V semiconductors. The magnetic materials used most commonly are iron based and are commonly called ferromagnetic materials.

In order to understand the nature of ferromagnetic materials, a brief background on materials stressing their magnetic properties will be given. All atoms consist of electrons revolving around the nucleus of the atom. Since electrons in motion are essentially current, electrons revolving around the nucleus can be approximated as a current loop. It may be shown that this current loop is equivalent to the magnetic dipole, as shown in Figure 6.2.

Obviously due to the number of electrons revolving around the nucleus the resulting magnetic dipoles may add constructively or destructively dividing materials into two groups: those in which permanent dipoles are absent and those in which permanent dipoles are present. The first group of materials is often referred to as being nonmagnetic or diamagnetic. The second group is divided into several subgroups depending upon the interactions between the individual dipoles. If the interaction between magnetic dipoles is very small, the material is paramagnetic. If the dipoles add up in parallel, the material is said to be ferromagnetic. If they are antiparallel, the material is said to be ferrimagnetic or antiferromagnetic. A schematic illustration of paramagnetic, ferromagnetic, antiferromagnetic, and ferrimagnetic materials is given in Figure 6.3.

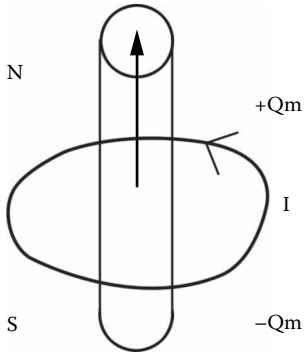


FIGURE 6.2
Current loop and the equivalent magnetic dipole.

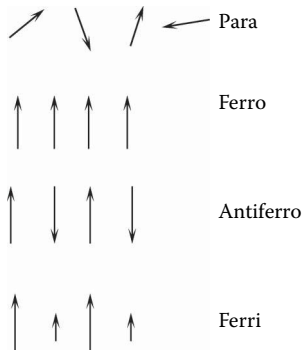


FIGURE 6.3
Schematic illustration of paramagnetic, ferromagnetic, antiferromagnetic, and ferrimagnetic materials. The arrows point from south to north.

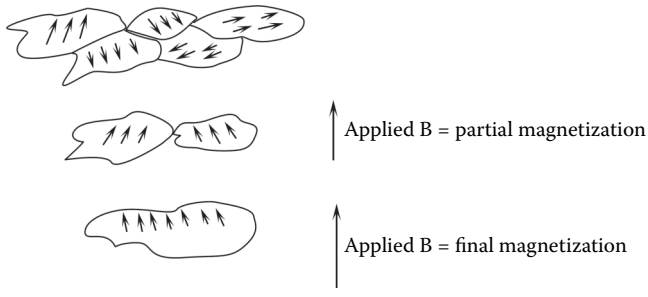


FIGURE 6.4
Magnetization of a ferromagnetic material where B = magnetic field.

The ferromagnetic materials respond most sensitively to a magnetic field, and the individual dipoles tend to align themselves with the applied magnetic field. This does not happen instantaneously. In particular, sections of the material called domains line up and finally the entire material is aligned. This is illustrated in Figure 6.4.

In the magnetization process stress occurs that leads to magnetostriction or the magnetoelastic effect. Many magnetic sensors can be realized using this effect.

6.4 Principles of Magnetic Sensors

In general a sensor can be represented as shown in Figure 6.5. If the sensor is passive then the input physical or electrical source is not present. This would be true in a simple piezoelectric force or pressure sensor in which

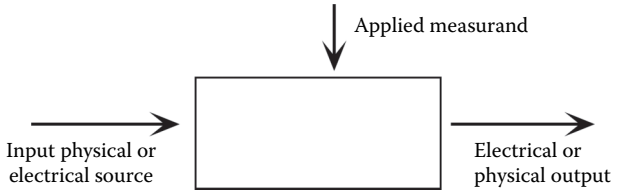


FIGURE 6.5
General sensor representation.

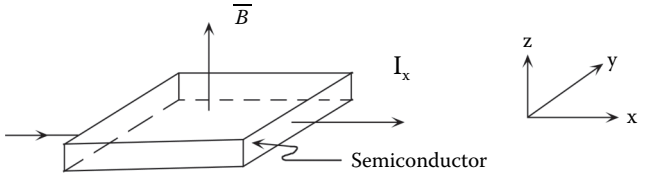


FIGURE 6.6
Hall effect geometry.

the output is a voltage or a stress. A Taguchi or metal oxide gas sensor is an active sensor and would have a voltage as an input, gas as the measurand, and a change in conductivity or resistivity as the output. In a magnetic sensor at least one of three components shown in Figure 6.5 would be a magnetic field. For example, we may have voltages as the input, magnetic field as the measurand, and change in current as the output. This configuration is the well-known magnetoresistive sensor in which a magnetic field is the measurand, which in turn changes the value of resistance in the material.

Several magnetic sensors can be realized as a result of the galvanomagnetic effect, which relates to interrelationships that occur between flowing electric current subjected to a magnetic field. This effect may be illustrated as shown in Figure 6.6, which represents the Hall effect.

A current passing through a semiconductor is subjected to an external magnetic field, $\overline{B}$, which gives rise to a Lorentz force as follows:

$$\overline{F} = e\overline{v} \times \overline{B} \tag{6.1}$$

where e = electronic charge, $\overline{B}$ = external magnetic field, and $\overline{v}$ = electron velocity.

Since the resulting force is in the y direction, the electrons will be deflected in that direction creating a voltage, which is commonly called the Hall voltage. Note that electron flow is opposite to flow of current. In addition to this voltage, the electron path from the left face of the semiconductor to the right face has been lengthened. This longer path for the

electrons results in an increase in the resistance of the semiconductor. This resistance is dependent on the magnetic field and is called magnetic resistance. This exactly describes the magnetoresistive sensor mentioned earlier. A more detailed discussion of the galvanomagnetic effect is given in the next section.

Another magnetic sensor can be realized from magnetostriction or the magnetoelastic effect. The application of a magnetic field to a cylindrical rod will cause a stress, and this stress may change the length of the rod. If this happens, the inducted stress or changes in rod length become a measure of the applied magnetic field. In order to understand magnetostriction, consider the application of a magnetic field to a demagnetized ferromagnetic material, as shown in Figure 6.7.

The applied magnetic field causes the individual magnetic dipoles to align themselves with the field, ultimately creating a magnetized material as shown in Figure 6.8.

In this magnetization process stresses and therefore strains occur in the magnetic material as the individual magnetic dipoles attempt to align themselves with the applied magnetic field. In general the stress, S, is proportional to the applied magnetic field, B, and the proportionality constant, K_m, referred to as the magnetoelastic constant, as follows:

$$S = K_m B \tag{6.2}$$

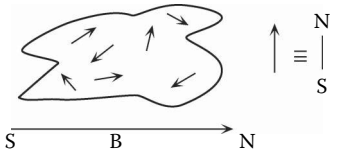


FIGURE 6.7
Application of a magnetic field to a demagnetized ferromagnetic material.

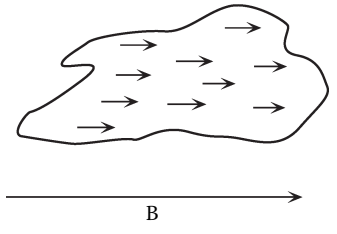


FIGURE 6.8
A magnetized ferromagnetic material.

This relationship is valid for isotropic homogeneous ferromagnetic materials. However, in crystalline anisotropic materials the relationship is not as simple as indicated in Equation 6.2 and one must account for the crystalline structure of the ferromagnetic material. Equation 6.2 in this case may be written as

$$[S] = [K_m]\,[B] \tag{6.3}$$

where $[S]$ is the 1×6 stress matrix, $[K_m]$ is the 3×6 magnetoelastic constant matrix, and $[B]$ is a 1×3 magnetic field matrix.

Equation 6.3 may also be represented in the following form:

$$[B] = [K_m]^{-1}\,[S] \tag{6.4}$$

In this case one sees that an applied stress, S, can cause a change in the magnetic field, $[B]$.

Obviously one can use the property of magnetoelasticity or magnetoconstriction to configure a large number of either stress or strain (force) and magnetic field sensors.

Additionally, one can also apply a magnetic field to a ferromagnetic material in which one has also applied an electric current. This is shown in Figure 6.9 for a longitudinal ferromagnetic rod.

In this case the applied magnetic field, B, and the magnetic field associated with the current, I, vectorially combine to create a resulting magnetic field on the rod surface, which has both longitudinal and transverse B field components. According to Equation 6.2, this resulting B field will then cause a stress resulting in the twisting of the rod. This type of configuration and the resulting effect are referred to as the Weidemann effect. This effect can also be used to configure a possible sensor for magnetic fields, current, or stress (twisting).

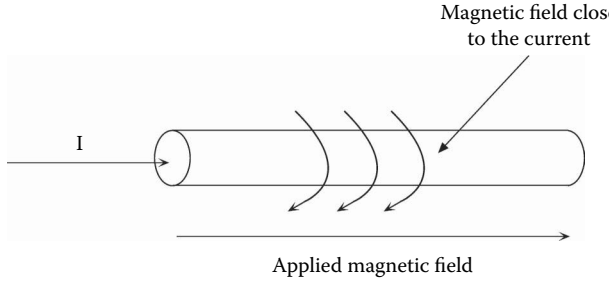


FIGURE 6.9
Application of longitudinal magnetic field to a current-carrying ferromagnetic rod.

6.5 Solid-State Magnetic Sensors

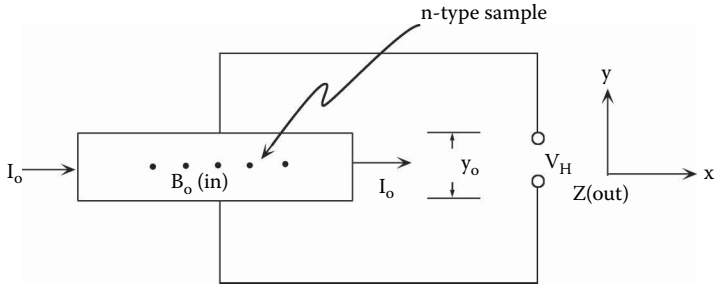
Due to the interrelationships between magnetic fields, electric fields (voltage), and current in semiconducting materials, a number of so-called solid-state magnetic sensors can be realized. In order to understand the operation of these sensors, it is necessary to explore galvanomagnetic efforts in more depth. In Figure 6.10 the geometry necessary to describe these effects is presented.

The Lorentz force equation may be written in centimeter gram second (CGS) units as follows:

$$\bar{F} = e\left(\bar{E} + \frac{1}{c}\bar{v} \times B_o\right) \tag{6.5}$$

where c = velocity of light.

In the absence of any applied voltage ($\bar{E} = 0$), a force is exerted on the electrons. This Lorentz force ($e/c(\bar{v}x\bar{B})$) in the negative y direction causes the electrons to build up on the lower edge of the sample (see Figure 6.11), resulting in an electric field that then balances the Lorentz force as follows:



I_o = Applied current in the x-direction
B_o = Applied magnetic field in the negative z direction (into the paper)

FIGURE 6.10
Basic geometry associated with galvanomagnetic effects.

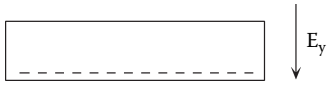


FIGURE 6.11
Electron buildup on lower face of the semiconductors causing an electric field.

$$0 = \frac{e}{c}\left(\bar{v}x\bar{B}_o\right)_y - eE_y \tag{6.6}$$

Solving Equation 6.6 for E_y, one obtains

$$E_y = \frac{v_x B_O}{c} \tag{6.7}$$

Noting that the current can be expressed as

$$I_O = -\text{nev}_x \tag{6.8}$$

one obtains

$$v_X = -\frac{I_O}{ne} \tag{6.9}$$

Substituting Equation 6.9 into Equation 6.7, one obtains

$$E_y = -\frac{I_O B_O}{nec} \tag{6.10}$$

Equation 6.10 is typically written in the following form:

$$E_y = RI_O B_O \tag{6.11}$$

where

$$R = -\frac{1}{nec} \tag{6.12}$$

R is referred to as the Hall coefficient and $V_H = E_y\, y_O$ is called the Hall voltage. R can be determined if V_H, I_O, and B_O are known. Note that I_O and B_O are applied and V_H is measured (see Figure 6.10). Once R is known we also can determine the carrier concentration, n, from Equation 6.12, obtaining

$$n = -\frac{1}{Rec} \tag{6.13}$$

In addition, if the semiconductor conductivity, σ, is known one can determine the carrier mobility as follows:

$$\sigma = ne\mu \tag{6.14}$$

and therefore,

$$\mu = \frac{\sigma}{ne} = -Rc\sigma \tag{6.15}$$

Equations 6.13 and 6.15 clearly indicate that it is possible to determine or sense the concentration and mobility of carriers in a semiconductor. Also, one may envision the semiconductor to be a metal oxide film wherein gas sorption would cause changes in n and μ, which can be accurately determined using the galvanomagnetic effect.

The effects just described are referred to as primary. Important secondary effects also exist. The electron buildup on the bottom face of the semiconductor also causes a temperature gradient, as shown in Figure 6.12.

Associated with this temperature gradient is an additional voltage or field,

$$E_T = V_T / y_O \tag{6.16}$$

A relationship then exists between the temperature gradient and E_T that may be expressed as

$$E_T = \tau \frac{\partial T}{\partial y} \tag{6.17}$$

This is commonly called the Ettinghausen effect, and τ is a proportionality constant. Notice also that since the electron path has been altered due to the Lorentz force, the path length for electrons from the left to the right side of the semiconductor has increased. This causes the sample resistance to increase, which is the magnetoresistance effect discussed earlier.

The galvanomagnetic effect can be utilized to construct a well-known magnetic field sensor called the Hall sensor. Figure 6.13 describes the basic geometry associated with this sensor.

This geometry consists of a thin rectangular semiconductor plate with four electrical contacts. A current, I, is applied via a voltage between contacts CS_1 and CS_2. The magnetic field, which is unknown in this case, is known from the previous galvanomagnetic discussion to produce a Hall voltage, which

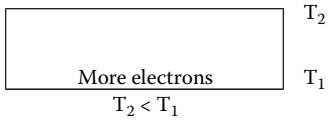


FIGURE 6.12
Electron buildup (causes a temperature gradient to appear in the semiconductor).

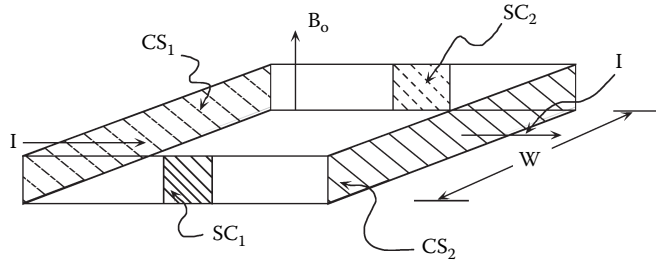


FIGURE 6.13
Geometry of the Hall effect sensor.

appears between the sensing contacts, SC_1 and SC_2. Referring to Equations 6.11 and 6.16, this voltage may be represented as

$$V_H = RIB_O W \tag{6.18}$$

For a long plate with very small contacts Equation 6.18 is valid. However, for a small plate with finite dimensions V_H is represented as follows:

$$V'_H = RIB_O WG \tag{6.19}$$

where G = geometric correction factor.

G may vary from 0 (large contacts) to 1 (point contacts). For sensing magnetic fields one commonly uses long plates with small contacts ($G \approx 1$). One then measures V_H and one knows R from Equation 6.12 and I, and therefore the magnetic field is given as follows from Equation 6.18:

$$B_O = \frac{V_H}{RIW} \tag{6.20}$$

By rotating the sample into different orientation one can now get all the components of the magnetic field.

In general in solid-state magnetic field sensors the following factors are critical:

- The magnetic field modified carrier transport
- Device design, which optimizes sensitivity

The present discussions have focused on semiconducting materials, but they may be extended to semiconducting devices such as diodes and transistors. As a result, one can realize magnetodiode sensors and bipolar

magnetotransistors. More recently, quantum theory and superconductivity have been utilized to realize superconductivity quantum interference devices (SQUIDs). These devices are among the most sensitive of all magnetic sensors and have been used to determine the low magnetic fields generated by the human brain and heart as well as unique instrumentation for the detection of gravitational waves. An in-depth discussion of the theory and applications associated with solid-state magnetic sensors may be found in Rosemerin [2].

References

1. Gopal, W., Hesse, J., and Zemel, J.N. 1996. Sensors, a comprehensive survey, *Magnetic Sensors*, Vol. 5.
2. Rosemerin, C. S. 1994. *Handbook of sensors and actuators: Solid state magnetic sensors*, ed. S. Middelhok. Vol. 1. Amsterdam: Elsevier.

Questions

1. Atlantic salmon return to the same river or stream in which they spawn. What might be the reason?
2. What is the Hall effect probe? Explain how the Hall effect probe works.
3. Explain:
   a. Wiedemann effect
   b. Ettinghausen effect
4. Explain, using equations, the property of magnetostriction.
5. What is the mathematical relationship between stress and the magnetic field in iron. Define all terms.
6. Theoretically, show how a magnetic sensor may be used to measure the carrier concentration and mobility of an *n*-type silicon wafer.
7. Explain in detail how you can determine the magnitude and direction of an unknown magnetic field.

Index

K

L

M