

SMART SENSOR SYSTEMS



UNIT-I

The term smart sensor was coined in the mid-1980s, and since then several devices have been called smart sensors.



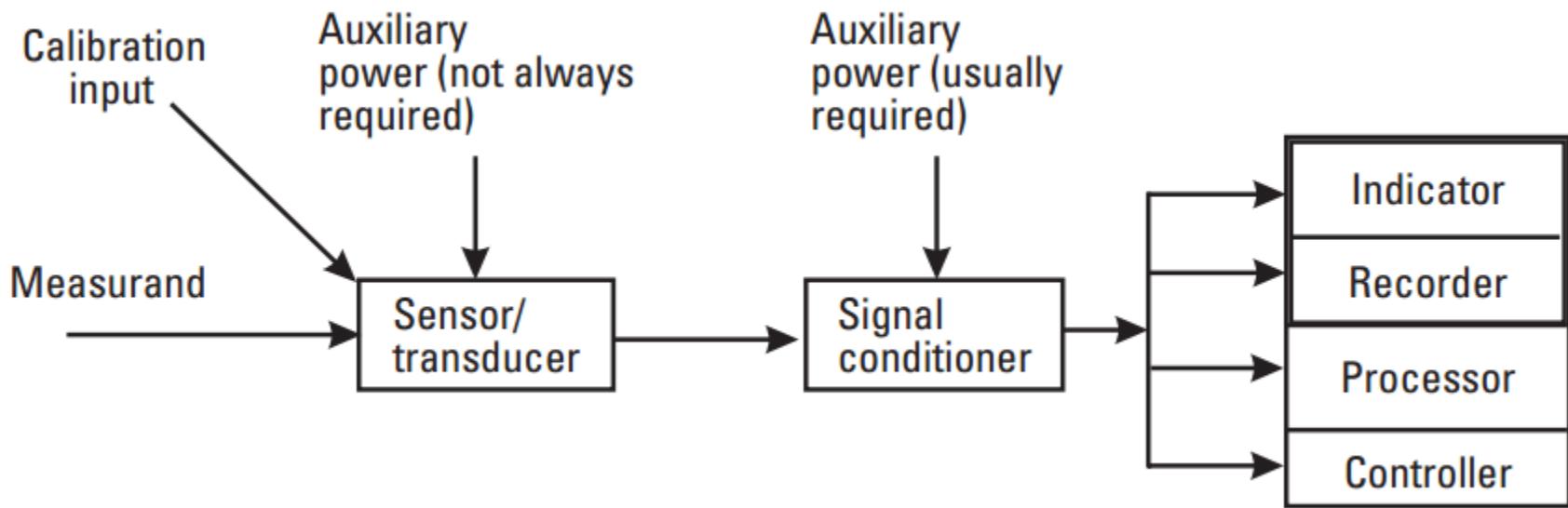
The intelligence required by such devices is available from microcontroller unit (MCU), digital signal processor (DSP), and application-specific integrated circuit (ASIC) technologies developed by several semiconductor manufacturers.

Before the availability of microelectronics, the sensors or transducers used to measure physical quantities, such as temperature, pressure, and flow, usually were coupled directly to a readout device, typically a meter that was read by an observer.

Many home thermostats, tire pressure gauges, and factory flow meters still operate in the same manner.

However, the advent of microprocessor technology initiated the requirement for sensors to have an electrical output that could be more readily interfaced to provide unattended measurement and control.

GENERAL SENSING SYSTEM



The recently approved IEEE 1451.2 specification defines a **smart sensor** as a **sensor that provides functions beyond those necessary for generating a correct representation of a sensed or controlled quantity.**

This function typically simplifies the integration of the transducer into applications in a networked environment”

Third Industrial Revolution

Automation has three phases:

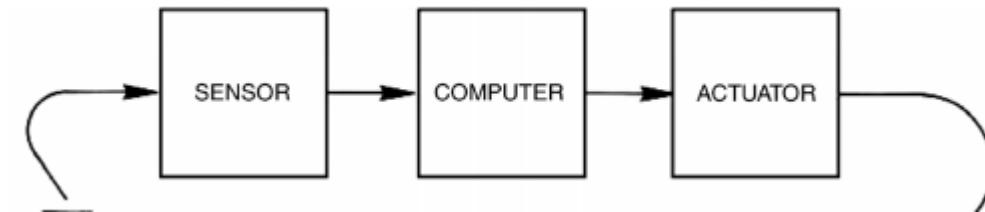
- (1) Mechanization**
 - (2) Informatization**
 - (3) Sensorization**
-

Humans have always tried to extend their capabilities.

Firstly, they extended their mechanical powers. They invented the steam engine, the combustion engine, the electric motor, and the jet engine. Mechanization thoroughly changed society. **The first industrial revolution was born.**

Secondly, they extended their brains, or their ratio. They invented means for artificial logic and communication: the computer and the internet. This **informatization phase is changing society again**, where we cannot yet fully predict the end result.

Sensorization together with mechanization and informatization will bring about the **third industrial revolution of full automation or robotization**



Pilotless Aircraft



Would you fly in one?

It includes many sensors to monitor the flight. The computers process the signals, compare them with the designed values, and provide control signals for the engines, rudders, and flaps that move the plane.

Definition of Sensors

Sensors transform signals from different energy domains to the electrical domain.

Definition of Smart Sensors

If we combine a sensor, an analog interface circuit, an analog to digital converter (ADC) and a bus interface in one housing, we get a smart sensor.

A smart sensor is a system-in-package in which a sensor and dedicated interface electronics are realized. **It may consist of a single chip**, as is the case with smart temperature sensors, image sensors and magnetic field sensors.

However, in cases when **the sensor cannot be implemented in the same technology as the interface electronics**, a **two-chip solution is required**. Since this also decouples the production yield of the circuit from that of the sensor, a two-chip solution is often more cost effective, even in cases where the sensor could be co-integrated with the electronics

A *sensor* is a device that detects or measures a physical quantity,

The opposite device is an *actuator*, which converts a signal (usually electrical) to some action, usually mechanical.

A *transducer* is a device that converts energy from one form into another.

Actuators and sensors are therefore forms of transducers.





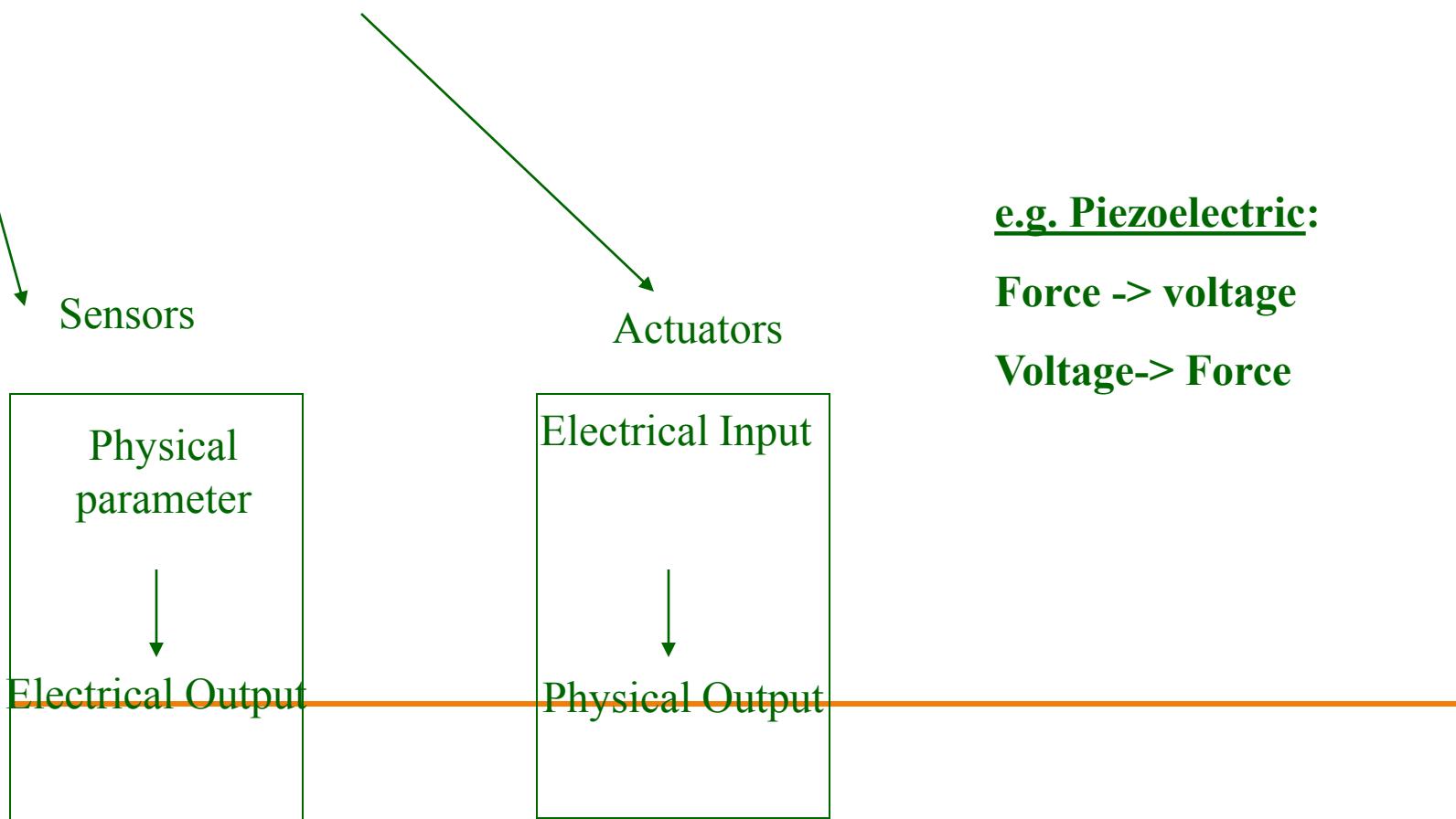
The differences between sensors and transducers are often very slight. A sensor performs a transducing action, and the transducer must necessarily sense some physical quantity. The difference lies in the efficiency of energy conversion. The purpose of a sensor is to detect and measure, and whether its efficiency is 5% or 0.1% is almost immaterial, provided the figure is known. A transducer, by contrast, is intended to convert energy, and its efficiency is important, though in some cases it may not be high. Linearity of response, defined by plotting the output against the input, is likely to be important for a sensor, but of much less significance for a transducer. By contrast, efficiency of conversion is important for a transducer but not for a sensor. The basic principles that apply to one, however, must apply to the other, so that the descriptions that appear in this book will apply equally to sensors and to transducers.

Transducer?

A device which converts one form of energy to another

When input is a physical quantity and output electrical → Sensor

When input is electrical and output a physical quantity → Actuator



Sensors classifications



PASSIVE AND ACTIVE

A passive sensor does not need any additional energy source and directly generates an electric signal in response to an external stimulus. That is, the input stimulus energy is converted by the sensor into the output signal.

The examples are a thermocouple, a photodiode, and a piezoelectric sensor.



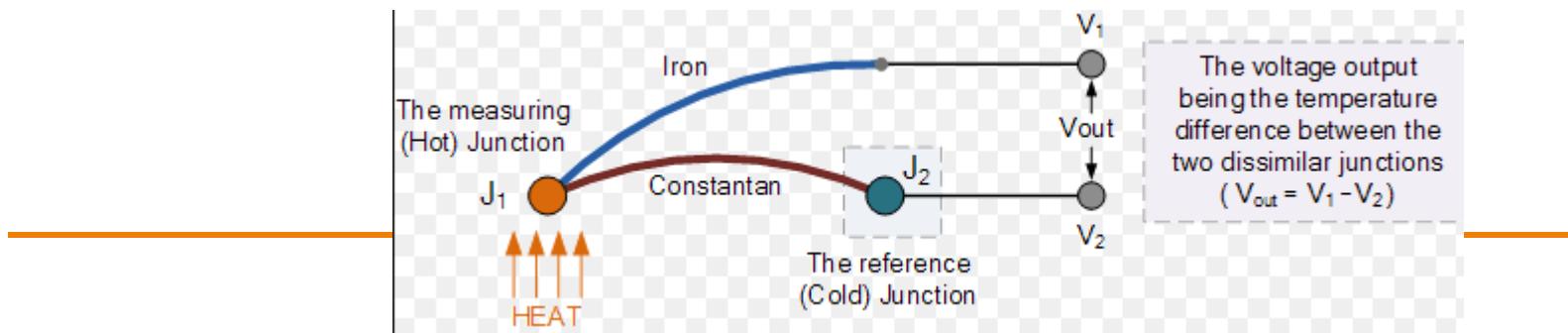
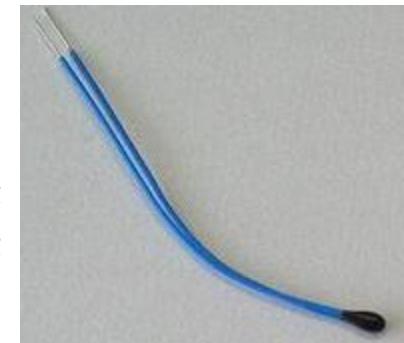
PASSIVE AND ACTIVE

The active sensors require external power for their operation, which is called an excitation signal. That signal is modified by the sensor to produce the output signal.

The active sensors sometimes are called parametric because their own properties change in response to an external effect and these properties can be subsequently converted into electric signals.

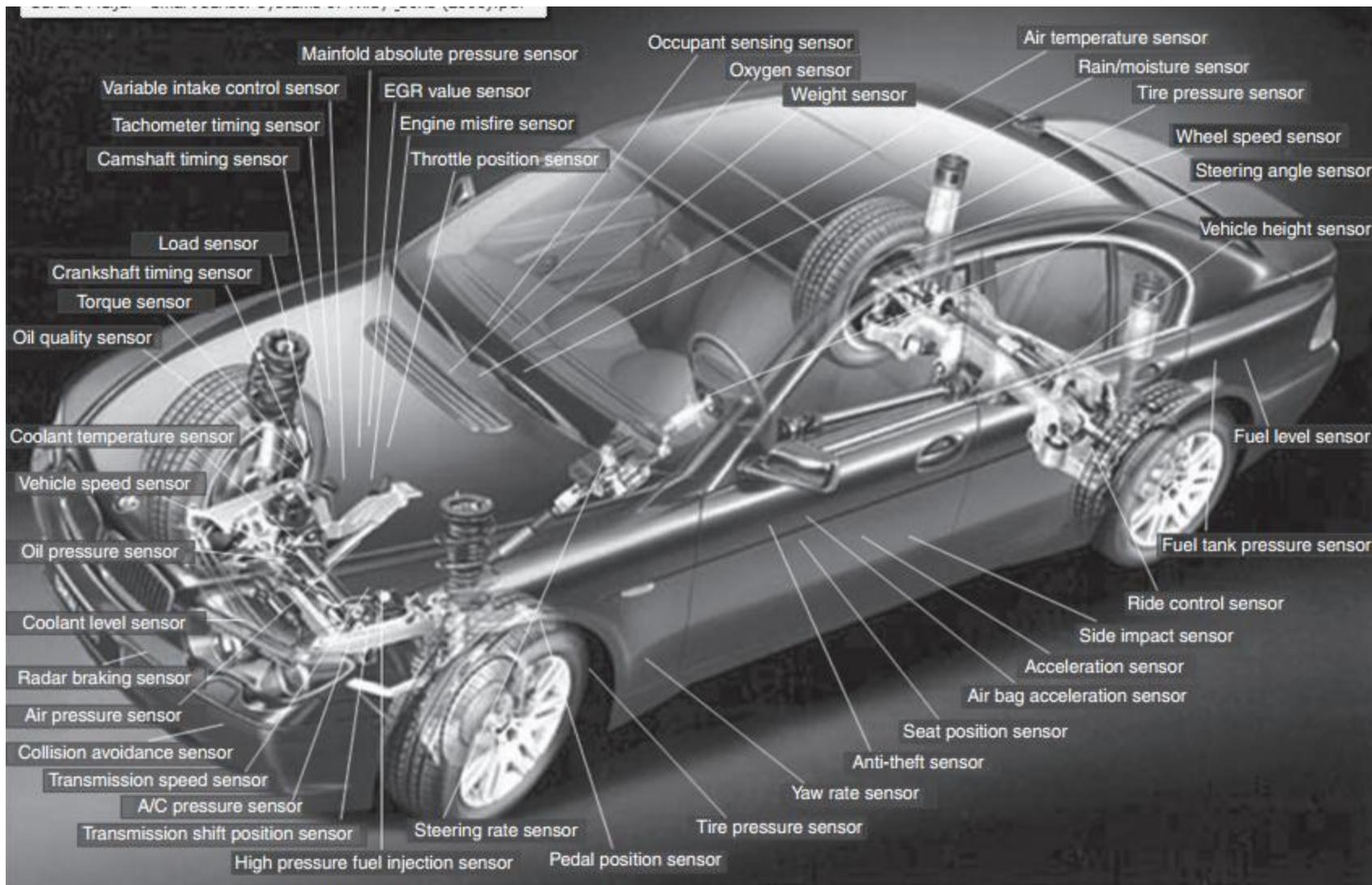
For example, a thermistor is a temperature sensitive resistor. It does not generate any electric signal, but by passing an electric current through it (excitation signal) its resistance can be measured by detecting variations in current and/or voltage across the thermistor. These variations (presented in ohms) directly relate to temperature through a known transfer function. Another example of an active sensor is a resistive strain gauge in which electrical resistance relates to a strain.

To measure the resistance of a sensor, electric current must be applied to it from an external power source.



There are a variety of ways by which a sensor can be categorized. Often, a sensor is located within the environment of interest. This type of sensor, which usually is mechanical or electrical, is considered an **invasive, or in situ, sensor**. Ideally , invasive sensors should not disturb the environment, which could alter the process under investigation.

A sensor also can be located outside the environment. For example, an optical pyrometer senses temperature remotely This is a **noninvasive sensor**.



Sensors can be categorized into domains, according to the type of physical variables that they sense

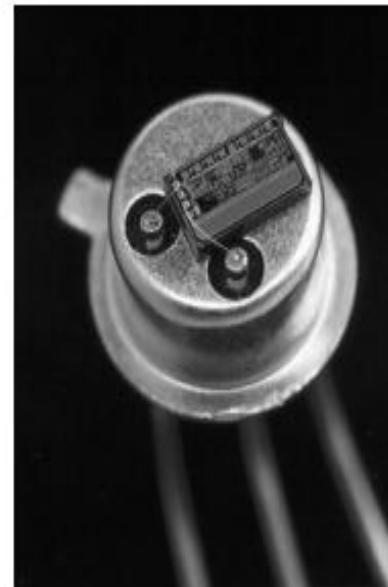


Figure 2.2 Photograph of a smart temperature sensor packaged in a TO-18 metal can. After sealing the encapsulation (not shown in the figure) the packaging is tested for gross and fine leakage. (Reproduced by permission of Smartec)

Sensors can be categorized into domains, according to the type of physical variables that they sense

Physical

Thermoelectric

Photoelectric

Photomagnetic

Magnetoelectric

Electromagnetic

Thermoelastic

Electroelastic

Thermomagnetic

Thermooptic

Photoelastic

Chemical

Chemical transformation

Physical transformation

Electrochemical process Spectroscopy

Biological

Biochemical transformation, Physical transformation

Spectroscopy

Sensors also can be organized with respect to the physical basis of how they sense.

These are

Electric, piezoresistive,

Fluid mechanic, optic

photoelastic, thermoelectric, and electrochemical.

Sensor material

Conductor

Insulator

Semiconductor

Liquid, gas or plasma

Biological substance

Other

Detection means used in sensors/how they sense

Biological

Chemical

Electric

magnetic electromagnetic wave

Heat, temperature

Mechanical displacement or wave

Radioactivity, radiation

Acoustic
Wave amplitude, phase, polarization
Spectrum
Wave velocity
Other

Biological
Biomass (types, concentration, states)
Other

Chemical
**Components (identities, concentration,
states)**
Other

Electric

Charge, current

Potential, voltage

Electric field (amplitude, phase, polarization, spectrum)

Conductivity

Permittivity

Magnetic

Magnetic field (amplitude, phase, polarization, spectrum)

Magnetic flux

Permeability

Radiation

Type

Energy

Intensity

Other

Thermal

Temperature

Flux

~~Specific heat~~

Thermal conductivity

Other

Optical

Wave amplitude, phase, polarization, spectrum

Wave velocity

Refractive index

Emissivity, reflectivity, absorption

Mechanical Position (linear, angular)

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

SENSORS APPLICATIONS

Civil engineering, construction

Domestic, appliances

Distribution, commerce, finance

**Environment, meteorology,
security**

Energy, power Information, telecommunication

Health, medicine Marine

Manufacturing

Recreation, toys

Military

Space

Scientific measurement Other

Transportation

What is Measurand ?

A physical quantity, property or condition which is measured.

Actually the following five terms define the nature of a sensor.

Measurand:

A physical quantity, property or condition which is measured.

Sensor:

A device which provides a usable output in response to a specified measurand.

Output:

The electrical quantity, produced by a sensor, which is a function of the applied measurand.

Sensing element:

That part of a sensor which responds directly to the measurand.

Transduction element:

The electrical portion of a sensor in which the output originates.

Sensor Characteristics/Properties Relating to the Measurand

An ideal or theoretical input–output (stimulus–response) relationship exists for every sensor.

If a sensor is ideally designed and fabricated with ideal materials by ideal workers working in an ideal environment using ideal tools, the output of such a sensor would always represent the true value of the stimulus. This ideal input–output relationship may be expressed in the form of a table of values, a graph, a mathematical formula, or as a solution of a mathematical equation.

If the input–output function is time invariant it is commonly called transfer function

SENSOR CHARACTERISTICS

Static

Dynamic

STATIC CHARACTERISTICS

Accuracy

Precision

Resolution

Minimum detectible Signal

Threshold

Sensitivity

Selectivity

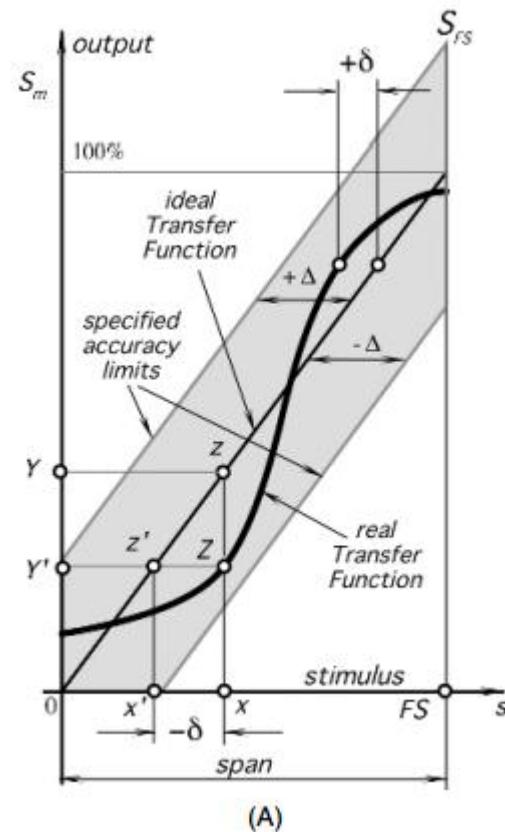
Nonlinearity

Hysteresis

Output impedance, Isolation, Grounding

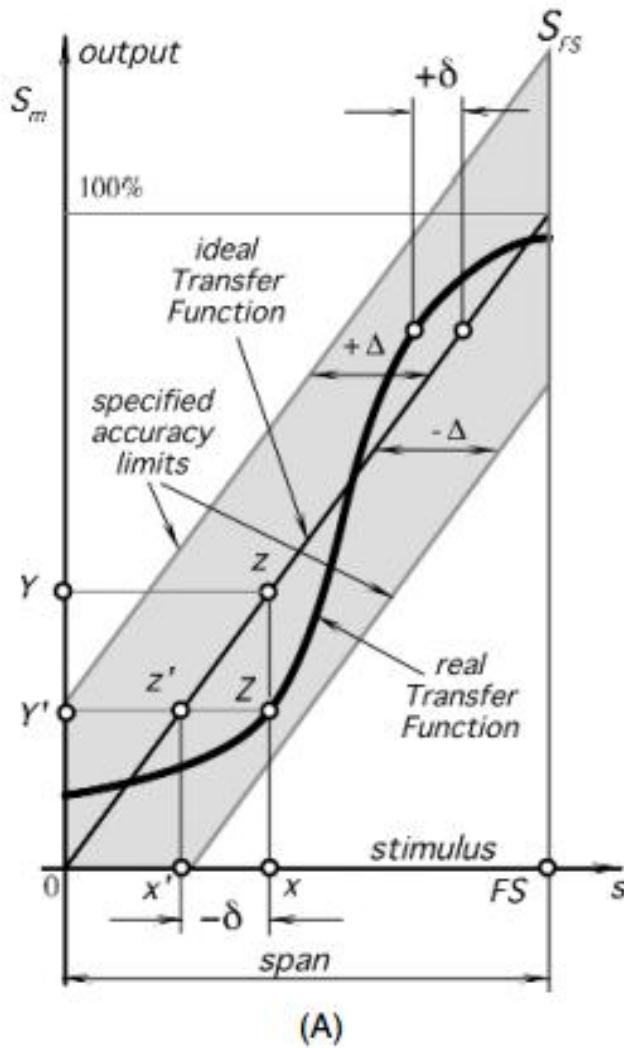
Span (Full-Scale Input)

A dynamic range of stimuli which may be converted by a sensor is called a span or an input full scale (FS). It represents the highest possible input value that can be applied to the sensor without causing an unacceptably large inaccuracy.



Full-Scale Output

Full-scale output (FSO) is the algebraic difference between the electrical output signals measured with maximum input stimulus and the lowest input stimulus applied. This must include all deviations from the ideal transfer function. For instance, the FSO output in Fig. is represented by S_{FS}



Accuracy

A very important characteristic of a sensor is accuracy which really means inaccuracy. Inaccuracy is measured as a highest deviation of a value represented by the sensor from the ideal or true value at its input. The true value is attributed to the object of measurement and accepted as having a specified uncertainty

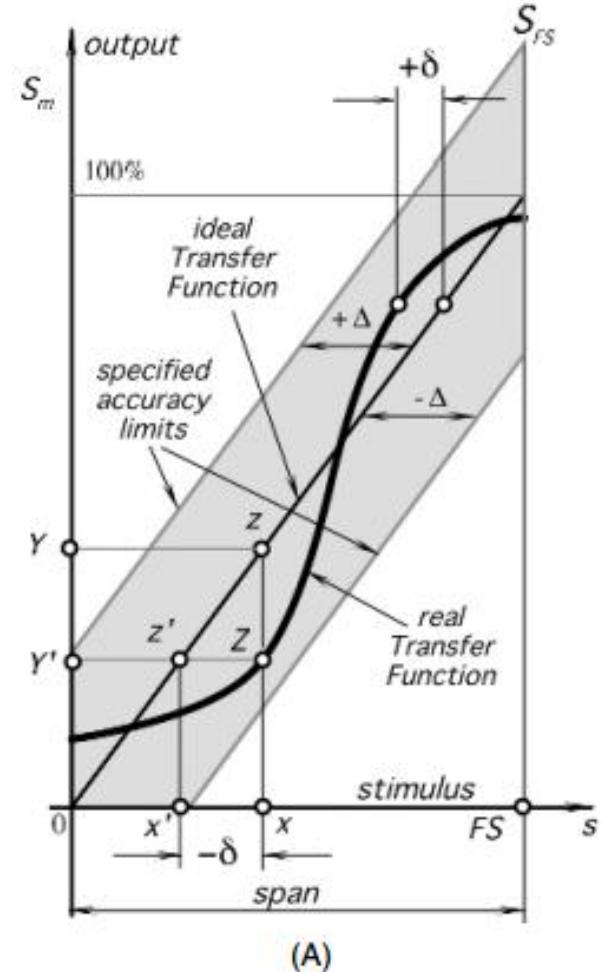


Figure shows an ideal or theoretical transfer function. In the real world, any sensor performs with some kind of imperfection. A possible real transfer function is represented by a thick line, which generally may be neither linear nor monotonic. A real function rarely coincides with the ideal. Because of material variations, workmanship, design errors, manufacturing tolerances, and other limitations

However, all runs of the real transfer functions must fall within the limits of a specified accuracy. These permissive limits differ from the ideal transfer function line by $\pm\Delta$

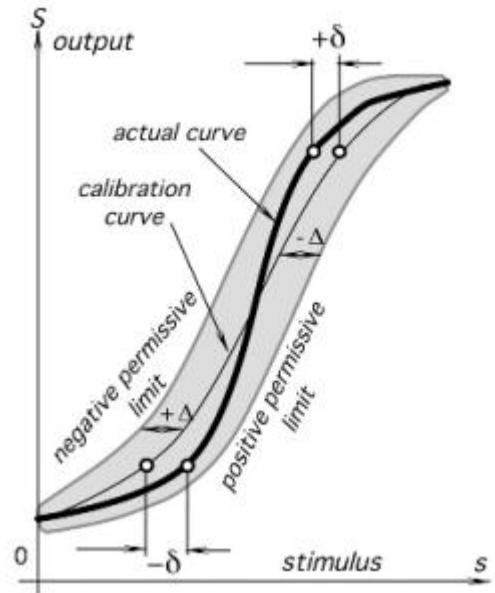
The real functions deviate from the ideal by $\pm\delta$, where $\delta \leq \Delta$

Thus, in this example, imperfection in the sensor's transfer function leads to a measurement error of $-\delta$.

The accuracy rating includes a combined effect of part-to-part variations, a hysteresis, a dead band, calibration, and repeatability errors (see later subsections). The specified accuracy limits generally are used in the worst-case analysis to determine the worst possible performance of the system.

The inaccuracy rating may be represented in a number of forms:

1. Directly in terms of measured value
2. In percent of input span (full scale)
3. ~~In terms of output signal~~



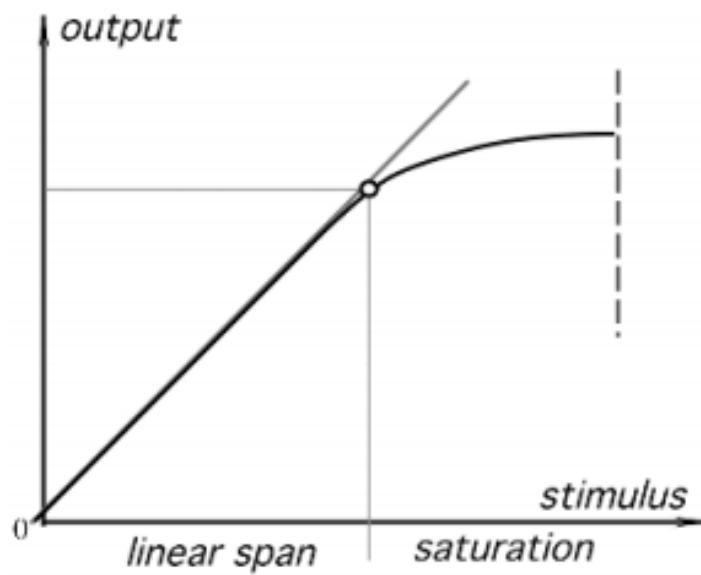
Error is specified in terms of input value

Hysteresis

A hysteresis error is a deviation of the sensor's output at a specified point of the input signal when it is approached from the opposite directions (Fig. 2.4). For example, a displacement sensor when the object moves from left to right at a certain point produces a voltage which differs by 20 mV from that when the object moves from right to left. If the sensitivity of the sensor is 10 mV/mm, the hysteresis error in terms of displacement units is 2 mm. Typical causes for hysteresis are friction and structural changes in the materials.

Saturation

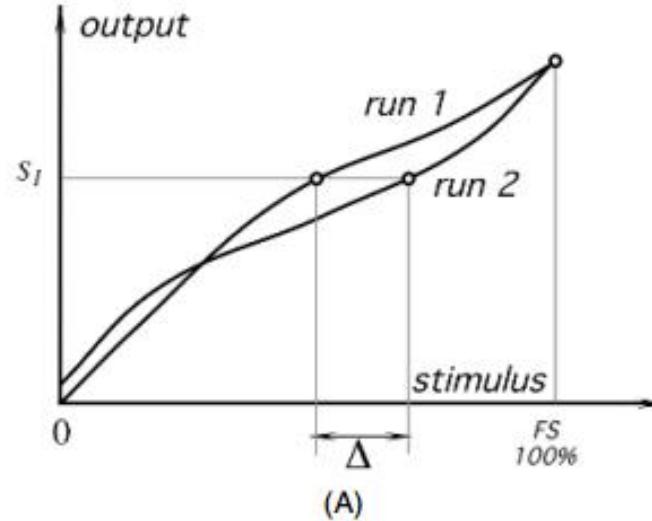
Every sensor has its operating limits. Even if it is considered linear, at some levels of the input stimuli, its output signal no longer will be responsive. A further increase in stimulus does not produce a desirable output. It is said that the sensor exhibits a span-end nonlinearity or saturation.



Repeatability

A repeatability (reproducibility) error is caused by the inability of a sensor to represent the same value under identical conditions. It is expressed as the maximum difference between output readings as determined by two calibrating cycles (Fig. , unless otherwise specified. It is usually represented as % of FS

$$\delta_r = \frac{\Delta}{FS} \times 100\%.$$



Possible sources of the repeatability error may be thermal noise, buildup charge, material plasticity, and so forth.

Resolution

Resolution describes the smallest increments of stimulus which can be sensed. When a stimulus continuously varies over the range, the output signals of some sensors will not be perfectly smooth, even under the no-noise conditions. The magnitude of the input variation which results in the output smallest step is specified as resolution under specified conditions (if any)

Resolution indicates the minimum change in input variable that is detectable, resolution is also defined in terms of percentage as:

$$\text{Resolution} = \frac{\Delta I}{I_{\max} - I_{\min}} \times 100$$

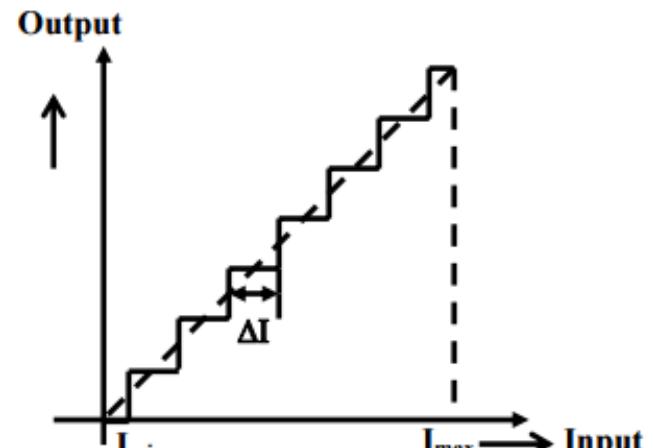


Fig. 3 Resolution

Dynamic Characteristics

Under static conditions, a sensor is fully described by its transfer function, span, calibration, and so forth. However, when an input stimulus varies, a sensor response generally does not follow with perfect fidelity. The reason is that both the sensor and its coupling with the source of stimulus cannot always respond instantly. In other words, a sensor may be characterized with a time-dependent characteristic, which is called a dynamic characteristic.

- . If a sensor does not respond instantly, it may represent the stimulus as somewhat different from the real, that is, the sensor responds with a dynamic error. A difference between a static and dynamic error is that the latter is always time-dependent. If a sensor is part of a control system, which has its own dynamic characteristics, the combination may cause at best a delay in representing a true value of a stimulus or at worst-cause spurious oscillations.
-

In a control system theory, it is common to describe the input-output relationship through a constant-coefficient linear differential equation. Then, sensor's dynamic (time-dependent) characteristics can be studied by evaluating such an equation. Depending on the sensor design, the differential equation can be of several orders. A zero-order sensor is characterized by a transfer function that is time independent. Such a sensor does not incorporate any energy storage devices, like a capacitor. A zero-order sensor responds instantaneously. In other words, such a sensor does not need any dynamic characteristics to be specified.

A first-order differential equation describes a sensor that incorporates one energy storage component. The relationship between the input $s(t)$ and output $S(t)$ is differential equation

$$b_1 \frac{dS(t)}{dt} + b_0 S(t) = s(t) \quad (2.33)$$

A typical example of a first-order sensor is a temperature sensor where energy storage is thermal capacity.

The first-order sensors may be specified by a manufacturer in various ways. A typical is a frequency response, which specifies how fast a first-order sensor can react to a change in the input stimulus. The frequency response is expressed in Hz or rad/s to specify the relative reduction in the output signal at certain frequency (Fig. 2.17a). A commonly used reduction number (frequency limit) is -3 dB. It shows at what frequency the output voltage (or current) drops by about 30%. Frequency response limit f_u is often called the upper cutoff frequency, as it is considered the highest frequency, which a sensor can process.

The frequency response directly relates to a speed response, which is defined in units of input stimulus per unit of time. How to specify, frequency or speed, in any particular case depends on the sensor type, its application, and the preference of a designer.

Another way to specify speed response is by time, which is required by the sensor to reach 90% of a steady-state or maximum level upon exposure to a step stimulus. For the first-order response, it is very convenient to use a so-called time constant. Time constant τ is a measure of the sensor's inertia. In electrical terms, it is equal to a product of electrical capacitance and resistance: $\tau = CR$. In thermal terms, thermal capacity and thermal resistances should be used instead. Practically, the time constant can be easily measured.

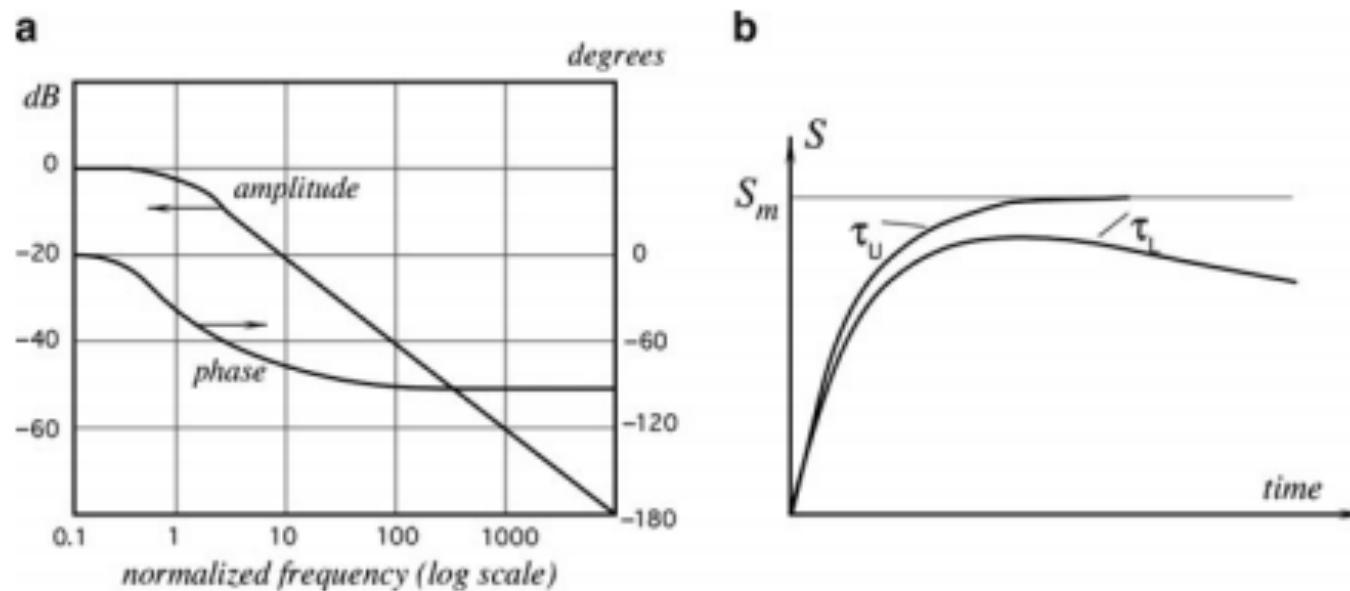


Fig. 2.17 Frequency characteristic (a) and response of a first-order sensor (b) with limited upper and lower cutoff frequencies τ_u and τ_l are the corresponding time constants

A first order system response is as follows:

$$S = S_m(1 - e^{-t/\tau}), \quad (2.34)$$

where S_m is steady-state output, t is time, and e is base of natural logarithm.

Substituting $t = \tau$, we get

$$\frac{S}{S_m} = 1 - \frac{1}{e} = 0.6321 \quad (2.35)$$

In other words, after an elapse of time equal to one time constant, the response reaches about 63% of its steady-state level. Similarly, it can be shown that after two time constants, the height will be 86.5% and after three time constants it will be 95% of the level that would be reached at infinite time.

Cutoff frequency shows what is the lowest or highest frequency of stimulus the sensor can process. The upper cutoff frequency shows how fast the sensor reacts, the lower cutoff frequency shows how slowly changing stimuli the sensor can process. Figure 2.17b depicts the sensor's response when both upper and lower cutoff frequencies are limited. As a rule of thumb, a simple formula can be used to establish a connection between the cutoff frequency f_c (either upper and lower) and time constant in a first-order sensor:

$$f_c \approx \frac{0.159}{\tau} \quad (2.36)$$

Phase shift at a specific frequency defines how the output signal lags behind in representing the stimulus change (Fig. 2.17a). The shift is measured in angular degrees or rads and is usually specified for a sensor that processes periodic signals. If a sensor is part of a feedback control system, it is very important to know its phase characteristic. The phase lag reduces the phase margin of the system and may result in overall instability.

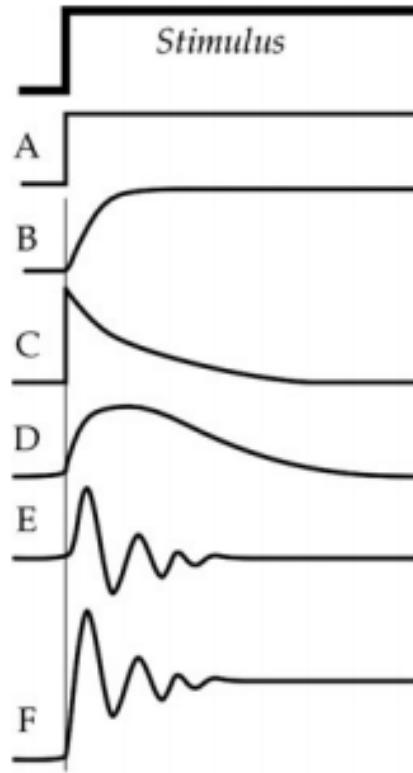
A second-order differential equation describes a sensor that incorporates two energy storage components. The relationship between the input $s(t)$ and output $S(t)$ is differential equation

$$b_2 \frac{d^2S(t)}{dt^2} + b_1 \frac{dS(t)}{dt} + b_0 S(t) = s(t) \quad (2.37)$$

An example of a second-order sensor is an accelerometer that incorporates an inertial mass and a spring (Fig. 2.18).

A second-order response is specific for a sensor that responds with a periodic signal. Such a periodic response may be very brief and we say that the sensor is damped, or it may be of a prolonged time and even may oscillate continuously. Naturally, for a sensor such a continuous oscillation is a malfunction and must be avoided. Any second-order sensor may be characterized by a resonant (natural) frequency, which is a number expressed in Hz or rad/s. The natural frequency

Fig. 2.18 Types of responses
unlimited upper and lower frequencies (A); first-order limited upper cutoff frequency (B); first-order limited lower cutoff frequency (C); first-order limited both upper and lower cutoff frequencies (D); narrow bandwidth response (resonant) (E); wide bandwidth with resonant (F)



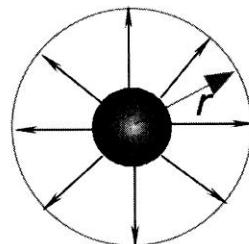
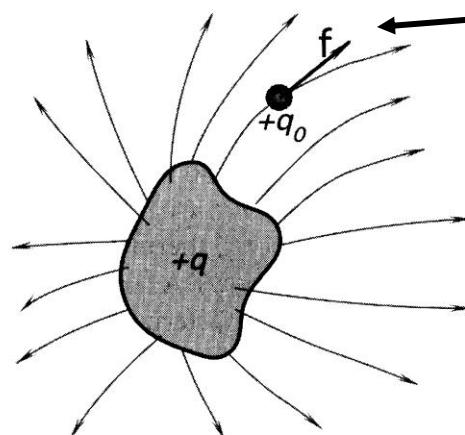
shows where the sensor's output signal increases considerably. When the sensor behaves as if the output conforms to the standard curve of a second-order response, the manufacturer will state the natural frequency and the damping ratio of the sensor. The resonant frequency may be related to mechanical, thermal, or electrical properties of the detector. Generally, the operating frequency range for the sensor should be selected well below (at least 60%) or above the resonant frequency. However, in some sensors, the resonant frequency is the operating point. For instance, in glass breakage detectors (used in security systems) the resonant makes the sensor selectively sensitive to a narrow bandwidth, which is specific for the acoustic spectrum, produced by shattered glass.

Physical Principles of Sensing

- Charges, fields & potentials
 - Capacitance
 - Magnetism
 - Induction
 - Resistance
 - Piezoelectric effect
 - Seebeck and Peltier effects
 - Thermal properties of materials
 - Heat transfer
 - Light
-

Electric Charges, Fields, & Potentials

- Any charged object is subject to a force when in the region of an electric field.
 - A field can be used to detect the presence of charge or the opposite can be true and the force on a charge determined to detect a field.



$$\mathbf{E} = \frac{\mathbf{f}}{q_0}$$

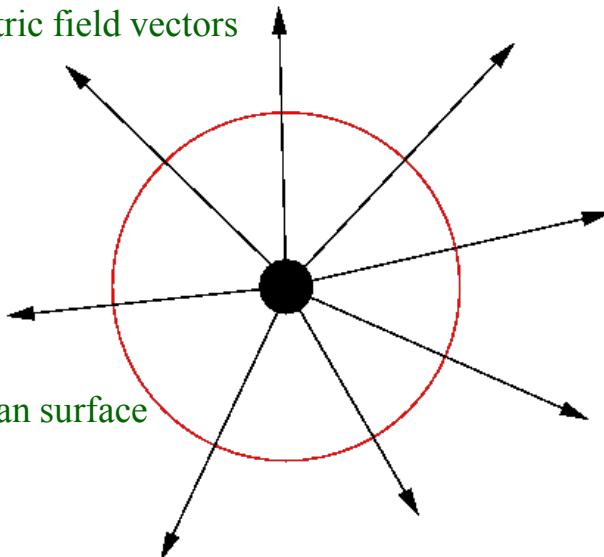
$$E = \frac{q}{4\pi\epsilon_0 r^2}$$

Point or Spherical charge

What is the field around a point charge (e.g. an electron)?

The electric field is everywhere perpendicular to a spherical surface centred on the charge.

Electric field vectors

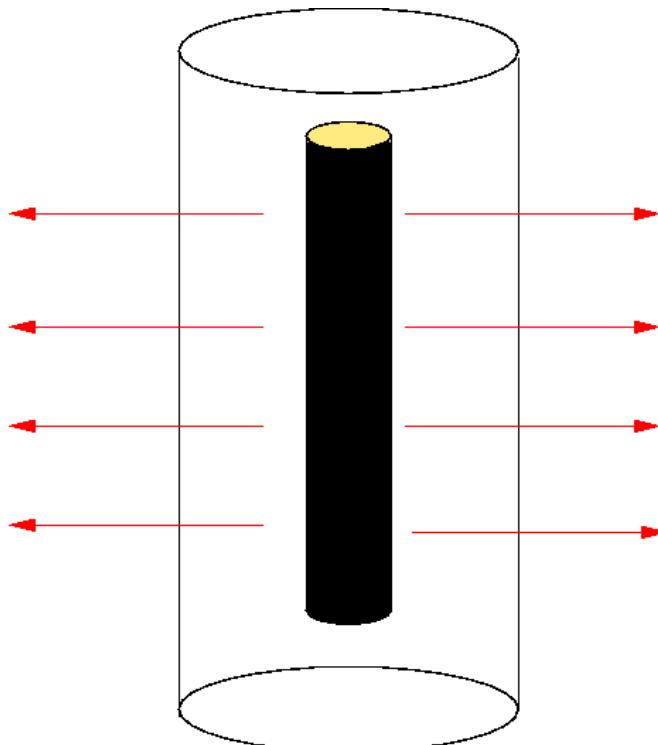


Gaussian surface

$$\text{So } \oint_S \vec{E} \cdot d\vec{s} = E \times 4\pi r^2 = \frac{Q}{\epsilon_0}$$

$$E = \frac{Q}{4\pi r^2}$$

Line of Charge



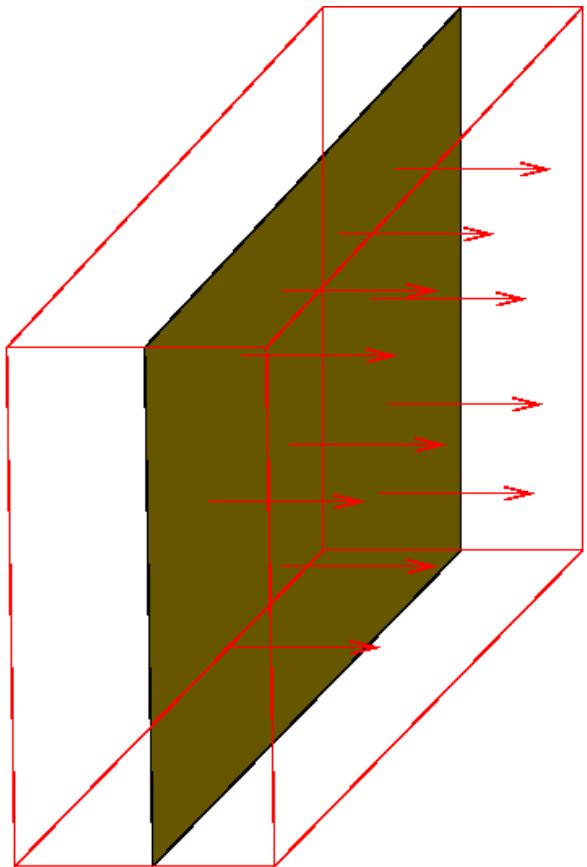
For a very long line of charge (eg a wire), the cylindrical surface has electric field perpendicular to a cylindrical surface.

$$\text{So } \oint_S \vec{E} \cdot d\vec{s} = E \times 2\pi r L = \frac{Q}{\epsilon_0}$$
$$\Rightarrow E = \frac{Q}{2\pi\epsilon_0 r L} = \frac{\lambda}{2\pi\epsilon_0 r}$$

Where λ = linear charge density

(Coulombs/meter)

Plane of Charge



For a very large flat plane of charge
the electric field is perpendicular to a
box enclosing a segment of the sheet

So

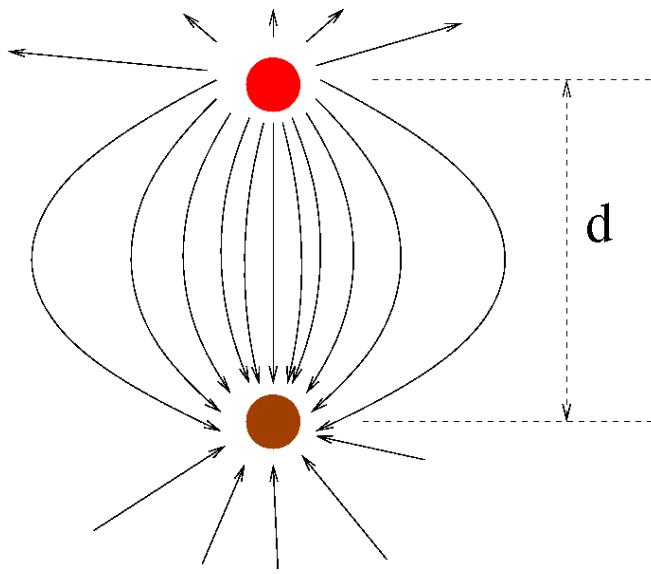
$$\oint_S \vec{E} \cdot d\vec{s} = E \times A = \frac{Q}{\epsilon_0}$$

$$E = \frac{Q}{A\epsilon_0} = \frac{\sigma}{\epsilon_0}$$

Where σ = Charge/Unit area on the surface

Electric Dipole

Dipoles are found in crystalline materials and form a foundation for piezoelectric and pyroelectric detectors



- An electric dipole is two equal and opposite charges Q separated by a distance d .
- The electric field a long way from the pair is

$$E = \frac{1}{4\pi\epsilon_0} \frac{Qd}{r^3} = \frac{1}{4\pi\epsilon_0} \frac{p}{r^3}$$

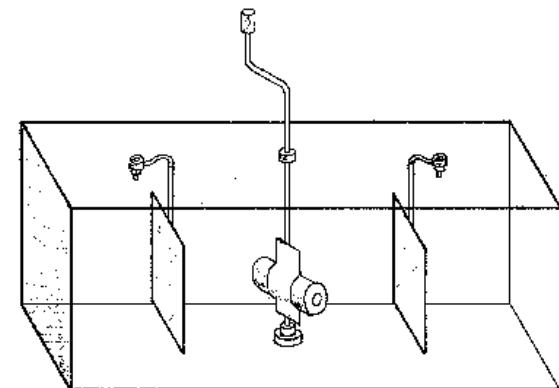
- $p = Qd$ is the Electric Dipole moment

-
- p is a measure of the strength of the field generated by the dipole.

Electrocardiogram

- Works by measuring changes in electric field as heart pumps
- Heart can be modeled as a rotating dipole

- Electrodes are placed at several positions on the body and the change in voltage measured with time



Electrocardiographs record small voltages of about one millivolt (mV) that appear on the skin as a result of cardiac activity. The voltage differences between electrodes are measured; these differences directly correspond to the heart's electrical activity.

Electrocardiogram

- Interior of Heart muscle cells negatively charged at rest
- Called “polarisation”
- K^+ ions leak out, leaving interior –ve
- Depolarisation occurs just prior to contraction:

Na^+ ions enter cells

Occurs in waves across the heart

Re-polarisation restores –ve charge in interior



Polarisation

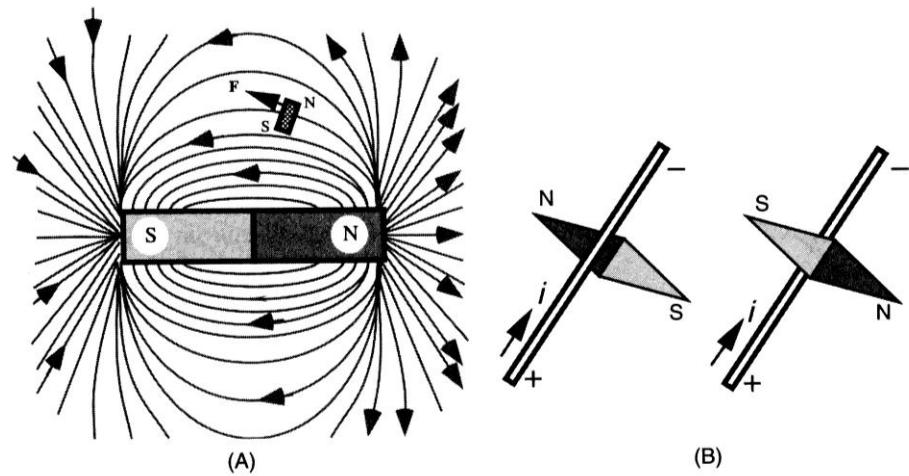


Depolarisation

Magnetism

- There are two methods of generating a magnetic field:
 - Permanent magnets (magnetic materials).
 - The magnetic field generated by a current.

Force is generated on a test magnet in the field of magnetic materials.

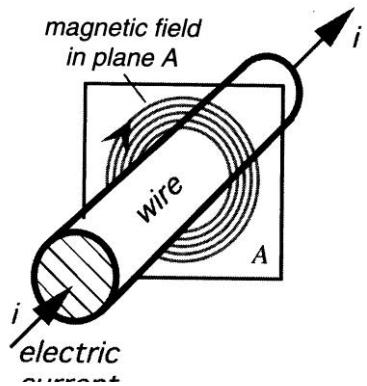


A compass needle will respond to the magnetic field generated by a current.

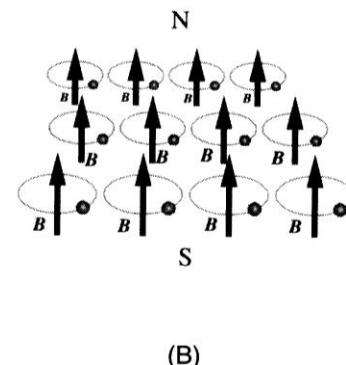
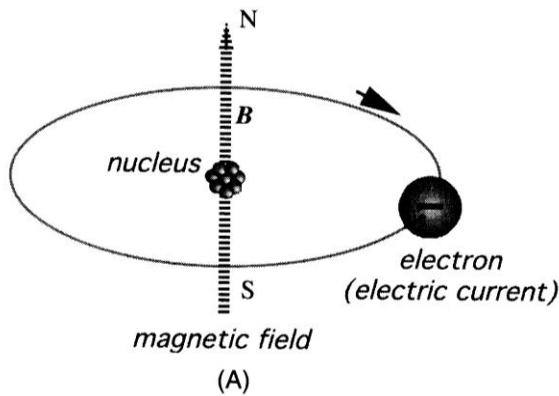
Magnetic field, \mathbf{B}

“flux” is the field density, Φ_B

Sources of Magnetic Field



Electric current sets a circular magnetic field around a conductor.



Moving electron sets a field, superposition of field vectors results in a combined magnetic field of a permanent magnet.

Magnets are useful for fabricating magnetic sensors for the detection of motion, displacement, and position.

Electric field due to a point charge

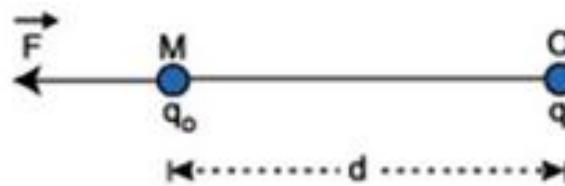
The test charge q_0 is placed at M. The force on it,

$$\vec{F} = \frac{1}{4\pi\epsilon_0} \frac{qq_0}{d^3} \vec{d}$$

Electric field strength at M,

$$\vec{E} = \frac{\vec{F}}{q_0} = \frac{1}{4\pi\epsilon_0} \frac{q}{d^3} \vec{d}$$

Or, $E = \frac{1}{4\pi\epsilon_0} \frac{q}{d^3}$



The direction of electric field is away from the positive point charge

3.5 Resistance

In any material, electrons move randomly like gas in a closed container. There is no preferred direction and an average concentration of electrons in any part of material is uniform (assuming that the material is homogeneous). Let us take a bar of an arbitrary material. The length of the bar is l . When the ends of the bar are connected to the battery having voltage V (Fig. 3.16), an electric field E will be setup within the material. It is easy to determine strength of the electric field

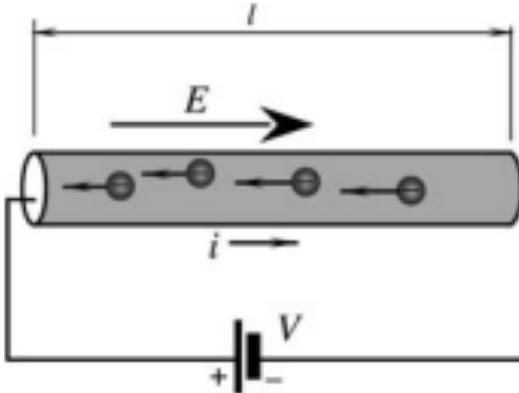
$$E = \frac{V}{l} \quad (3.49)$$

For instance, if the bar has a length of $l = 1$ m and the battery delivers 1.5 V, the electric field has the strength of 1.5 V/m. The field acts on free electrons and sets them in motion against the direction of the field. Thus, the electric current starts flowing through the material. We can imagine a cross-section of the material through which passes electric charge q . The rate of the electric charge flowing (unit of charge per unit of time) is called electric current

$$i = \frac{dq}{dt} \quad (3.50)$$

No matter what the cross-section of the material is, whether it is homogeneous or not, the electric current through any cross-section is always the same for a given electric field. It is similar to water flow through a combination of serially connected pipes of different diameters: the rate of flow is the same throughout of the pipe combination. The water flows faster in the narrow sections and slower in the wide

Fig. 3.16 Voltage across a material sets electric current



section, but amount of water passing through any cross-section per unit of time is constant. The reason for that is very simple: water in the pipes is neither drained out nor created. The same reason applies to electric current. One of the fundamental laws of physics is the law of conservation of charge. Under steady state conditions, charge in a material is neither created nor destroyed. *Whatever comes in must go out*. In this section, we do not consider any charge storages (capacitors), and all materials we discuss are said have pure resistive properties.

The mechanism of electrical conduction in a simplified form may be described as follows. A conducting material, say copper wire, can be modeled as a semirigid spring-like periodic lattice of positive copper ions. They are coupled together by strong electromagnetic forces. Each copper atom has one conduction electron, which is free to move about the lattice. When electric field \mathbf{E} is established within the conductor, force $-e\mathbf{E}$ acts on each electron (e is the electron charge). The electron accelerates under the force and moves. However, the movement is very short as the electron collides with the neighboring copper atoms, which constantly vibrate with intensity, which is determined by the material temperature. The electron transfers its kinetic energy to the lattice and is often captured by the positive ion. When captured, it frees another electron, which keeps moving in the electric field until, in turn, it collides with the next portion of the lattice. The average time between collisions is designated as τ . It depends on the material type, structure, and impurities. For instance, at room temperature, a conduction electron in pure copper moves between collisions for an average distance of $0.04 \mu\text{m}$ with $\tau=2.5 \times 10^{-14} \text{ s}$. In effect, electrons, which flow into the material near the negative side of the battery, are not the same that outflow to the positive terminal. However, the constant drift or flow of electrons is maintained throughout the material. Collisions of electrons with the material atoms further add to the atomic agitation and, subsequently, raise the material temperature. This is why passing of electric current through a resistive material results in the so called Joule heat liberation.

It was arbitrarily decided to define the direction of current flow along with the direction of the electric field, i.e., in the opposite direction of the electronic flow. Hence, the electric current flows from the positive to negative terminal of the battery while electrons actually move in the opposite direction. It is interesting to note that unlike water flowing through a pipe, electrons do not need to initially “fill up” the conductor before they start flowing out at a positive side. Electrons are

always present in a conductor. Since electric field in a conductor propagates with the speed of light in the conductor’s material, electric current appears at all parts of the conductor nearly instantaneously.

3.5.1 Specific Resistivity

If we fabricate two geometrically identical rods from different materials, say from copper and glass and apply to them the same voltage, the resulting currents will be quite different. A material may be characterized by its ability to pass electric current. It is called *resistivity* and material is said has electrical *resistance*, which is defined by Ohm’s law, which means that a ratio of voltage to current is a constant

$$R = \frac{V}{i} \quad (3.51)$$

For the pure resistance (no inductance or capacitance) voltage and current are in-phase with each other, meaning that they are changing simultaneously.

Any material has electric resistivity⁴ and therefore is called a *resistor*. The SI unit of resistance is 1 ohm (Ω) = 1 volt/1 ampere. Other multiples and submultiples of Ω are as follows:

1 milliohm ($m\Omega$)	$= 10^{-3} \Omega$
1 kilohm ($k\Omega$)	$= 10^3 \Omega$
1 megohm ($M\Omega$)	$= 10^6 \Omega$
1 gigohm ($G\Omega$)	$= 10^9 \Omega$
1 terohm ($T\Omega$)	$= 10^{12} \Omega$

If we compare electric current with water flow, pressure across the pipe line (Pascal) is analogous of voltage (V) across the resistor, electric current (C/s) is analogous of water flow (L/s) and electric resistance (Ω) corresponds to water flow resistance in the pipe (no special unit). It is clear that the resistance to water flow is smaller when the pipe is short, wide and has no obstructions. When the pipe has, for instance, a filter installed in it, resistance to water flow will be higher. Consider a human body where coronary blood flow may be restricted by cholesterol deposits on the inner lining of arteries. These deposits increase the flow resistance (called vascular resistance). The arterial blood pressure increases to compensate for rise in the vascular resistance but not always can keep up with it. If it is the case, the heart action that develops the arterial pressure is no longer sufficient to provide a necessary blood supply to vital organs, including a heart. This may result in a heart attack.

The basic laws that govern the electric circuit designs are called Kirchhoff's Laws, after the German physicist Gustav Robert Kirchhoff (1824–1887). These laws were originally devised for the plumbing networks, which, as we have seen, are analogous to the electric networks.

Resistance is a characteristic of a device. It depends on both the material and the geometry of the resistor. Material itself can be characterized by a *specific resistivity*, ρ , which is defined as

$$\rho = \frac{E}{j}, \quad (3.52)$$

where current density $j = i/a$ (a is the area of the material cross section). The SI unit of resistivity is $\Omega\text{-m}$. Resistivities of some materials are given in Appendix (Table A.7). Quite often, a reciprocal quantity is used which is called conductivity: $\sigma = 1/\rho$. The SI unit of conductivity is siemens having dimension [1/ Ω].

Resistivity of a material can be expressed through mean time between collisions, τ , the electronic charge, e , the mass of electron, m , and a number of conduction electrons per unit volume, n

$$\rho = \frac{m}{ne^2\tau}. \quad (3.53)$$

To find the resistance of a conductor the following formula may be used:

$$R = \rho \frac{l}{a}, \quad (3.54)$$

where a is the cross-sectional area and l is the length of the conductor. The ratio l/a is called a geometry factor.

Formula (3.54) establishes the fundamental relationship between resistance and its parameters. Thus, if one wants to design a resistive sensor, she should find ways of modulating either the specific resistivity or the geometry factor of the resistor.

Resistive

A sensor based upon the principle that a change in resistance can be produced by a change in a physical variable is, perhaps, the most common type of sensor.

**A resistance sensor can be used to measure
displacement, strain,
force, pressure,
acceleration, flow velocity ,
temperature, heat or light
flux, and gas concentration.**

Strain Gage

The resistance of a conductor also changes measurably with strain. When a fine wire of length L is stretched, its length increases by ΔL , yielding a longitudinal strain of $\epsilon_L = \Delta L/L$. This produces a change in resistance. Its width decreases by $\Delta d/d$, where d is the wire diameter. This defines the transverse strain $\epsilon_T = \Delta d/d$, Poisson's ratio, ν , is defined as the negative of the ratio of transverse to longitudinal local strains, ϵ_T / ϵ_L

For a wire, the resistance R can be written as $R = \rho \frac{L}{A}$,
in which ρ is the resistivity, L the length, and A the cross-sectional area.

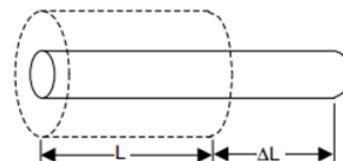
Taking the total derivative of Equation yields

$$dR = \frac{\rho}{A} dL + \frac{L}{A} d\rho - \frac{\rho L}{A^2} dA.$$

divided by the above equation to give the relative change in resistance,

$$\frac{dR}{R} = (1 + 2\nu)\epsilon_L + \frac{d\rho}{\rho}.$$

Above Equation shows that the relative resistance change in a wire depends on the strain of the wire and the resistivity change.



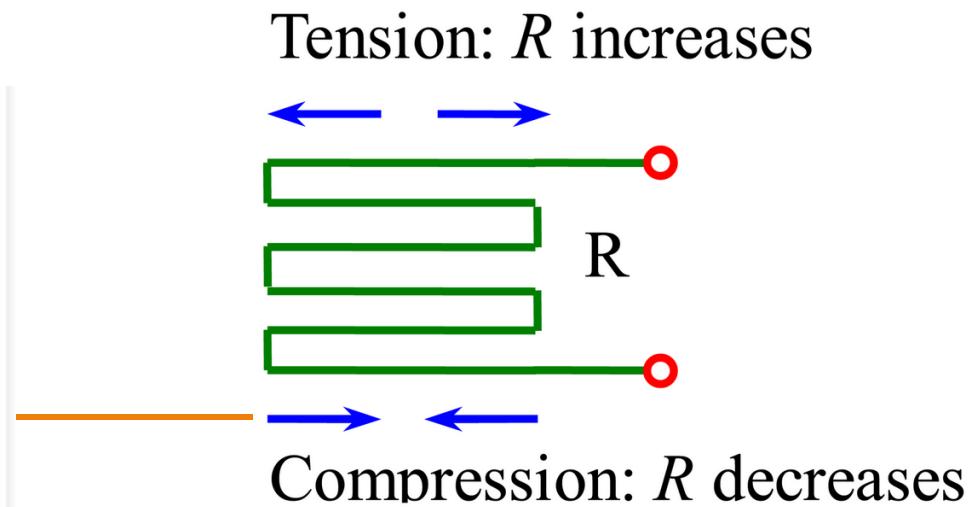
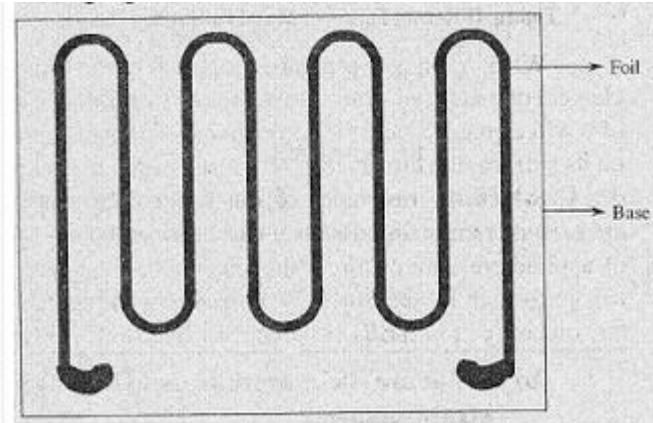
The gauge factor and characteristics of foil strain gauges are similar to the wire strain gauges. be easily etched on a the metal foil gauges can flexible insulating carrier film. The metal foil type strain gage is manufactured by photo-etching technique.

In the construction of etched foil strain gauge first a layer of strain sensitive material is bonded to a thin sheet of backelite or paper.

The part of some masking material and then to this unit an etching solution is applied.

Therefore, the unmasked part of the metal will be removed thereby leaving the required grid structure.

By this method of construction, the etched foil strain gauges are made in thinner sizes.



Capacitive sensors

A **capacitive sensor** consists of two small conducting plates, each of area, A , separated by a distance, h , with a dielectric material in between the two plates. The capacitance between the two plates is

$$C = \epsilon_0 \epsilon A / h, \quad (2.21)$$

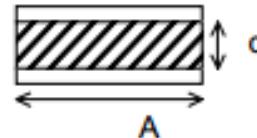
in which ϵ_0 is the permittivity of free space and ϵ the relative permittivity. Thus, a change in either ϵ , A , or h will produce a change in capacitance. Differentiating Equation 2.21 with respect to C gives

$$\frac{dC}{C} = \frac{d\epsilon}{\epsilon} + \frac{dA}{A} - \frac{dh}{h}. \quad (2.22)$$

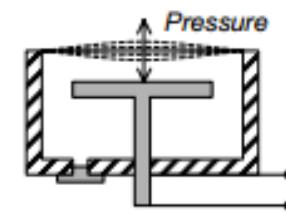
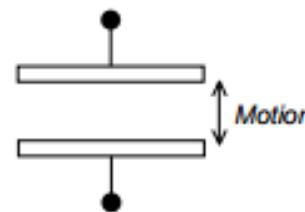
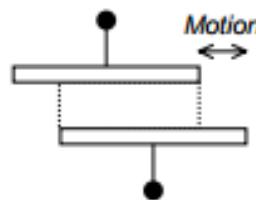
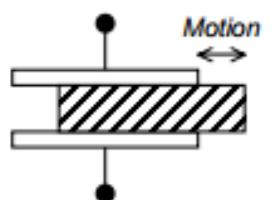
Each of these changes has been exploited in developing different capacitive-based sensors.

- The capacitance of a parallel plate capacitor is

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



- d is the separation between the plates, A is the area of the plates, ϵ_0 is the permittivity of air and ϵ_r is the relative permittivity of the dielectric
- A moving object is attached to the dielectric or the plates to generate capacitance changes



- Notes

- Variable distance (d) sensors operate over a range of a few millimeters
- Cross-sensitivity to temperature and humidity (specially if the dielectric is air)
- Capacitive sensors are also commonly used to measure pressure
 - "Condenser" microphones measure changes in air pressure of incoming sound waves

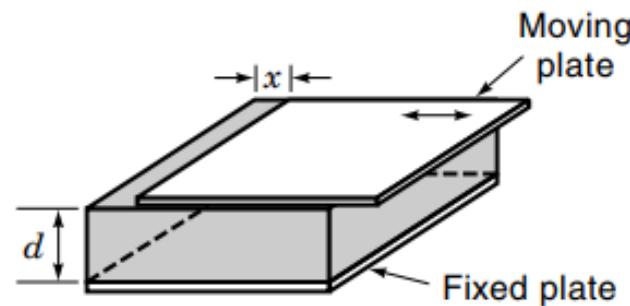
Variable Area Displacement Measurements

The displacements may be sensed by varying the surface area of the electrodes of a flat plate capacitor, as illustrated in Fig..

In this case, the capacitance is

$$C = \epsilon_r \epsilon_0 (A - wx)/d$$

and the transducer output is linear with displacement x . This type of sensor is normally implemented as a rotating capacitor for measuring angular displacement. The rotating capacitor structures are also used as output transducers for measuring electric voltages as capacitive voltmeters.



Variable Distance Displacement Measurements

If a capacitor is made from two flat coplanar plates separated by a variable distance x , ignoring fringe effects, the capacitance can be found by

$$C(x) = \epsilon A/x = \epsilon_r \epsilon_0 A/x$$

where ϵ is the dielectric constant or permittivity, ϵ_r is the relative dielectric constant (in air and vacuum, $\epsilon_r \approx 1$), $\epsilon_0 =$

8.854188×10^{-12} F/m ($10^{-9}/36 \pi$ F/m) is the dielectric constant of vacuum, x is the distance of the plates in m, and A is the effective area of the plates in m^2 .

The capacitance is a nonlinear function of the distance x , and has the characteristics of a hyperbolic transfer function. The sensitivity of capacitance to changes in plate separation may be found as

$$dC/dx = -\epsilon_r \epsilon_0 A/x^2$$

Equation indicates that the sensitivity increases as x decreases. Nevertheless, using Eqs. Above it can be proved that the percent changes in C are proportional to the percent changes in x which can be expressed as

$$dC/C = -dx/x$$

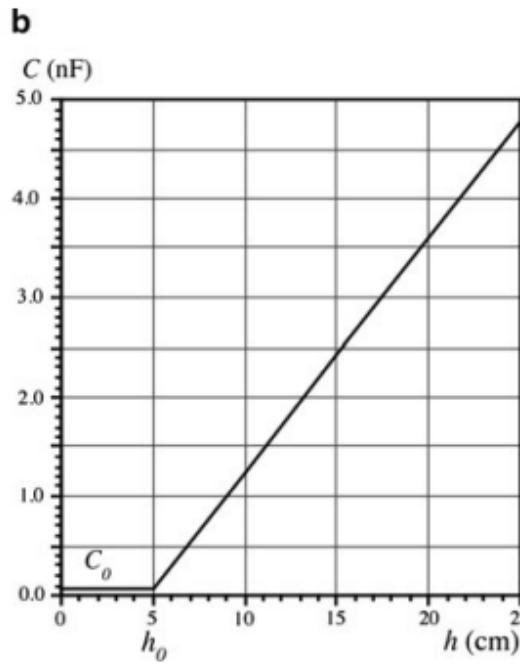
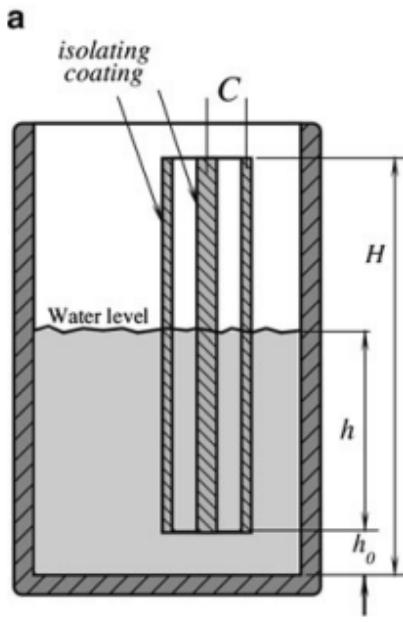
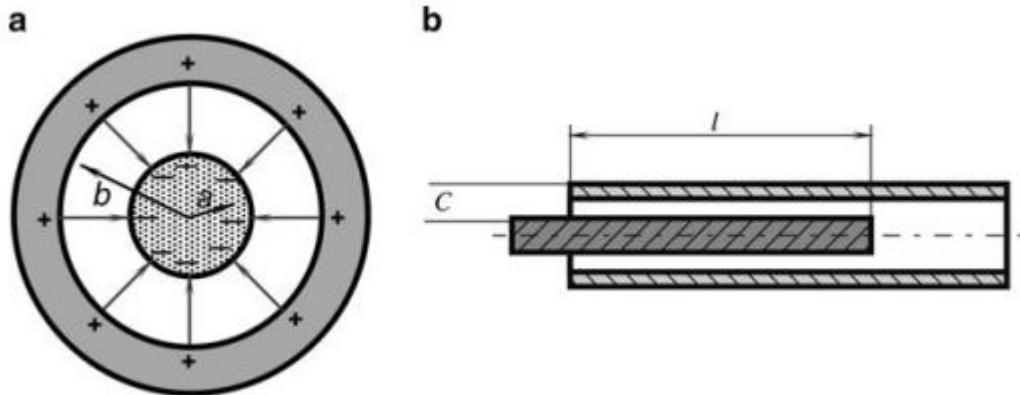


Fig. 3.8 Capacitive water level sensor (A); capacitance as function of the water level (B) (sensor's dimensions are as follows: $a=10$ mm, $b = 12$ mm, $H = 200$ mm, liquid—water)

Total capacitance of the coaxial sensor is

$$C_h = C_1 + C_2 = \epsilon_0 G_1 + \epsilon_0 \kappa G_2,$$

Where C_1 is the capacitance of the water-free portion of the sensor and C_2 is the capacitance of the water-filled portion. The corresponding geometry factors are designated G_1 and G_2 .



The Inductance Parameter

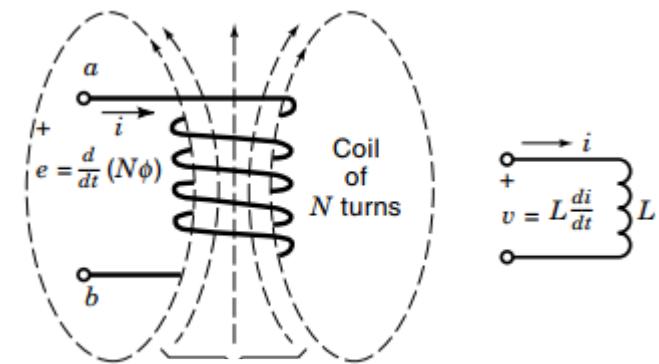
Inductance is the property whereby an electrical device sets up magnetic flux when current passes through it. In principle, any conductor of current has this property. However it is more pronounced in a coil of wire or any electrical device whose construction is akin to a coil.

Taking the solenoid of N turns shown in Fig. 1(a) as representing a general situation, let a current of i amperes through it produce a magnetic flux of ϕ webers. For simplicity, if we assume that the closed path of every flux line is interlinked with all N turns, then the coil has flux linkages of $N\phi$ weber-turns.

The inductance L of the coil, defined as the flux linkages produced in it per unit current, is measured in henrys. One henry corresponds to one weber-turn per ampere.

Any physical device, such as the coil considered, which has the property of inductance, is called an inductor.

$$L = N\phi / i$$

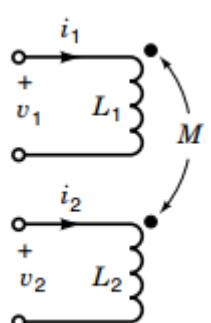
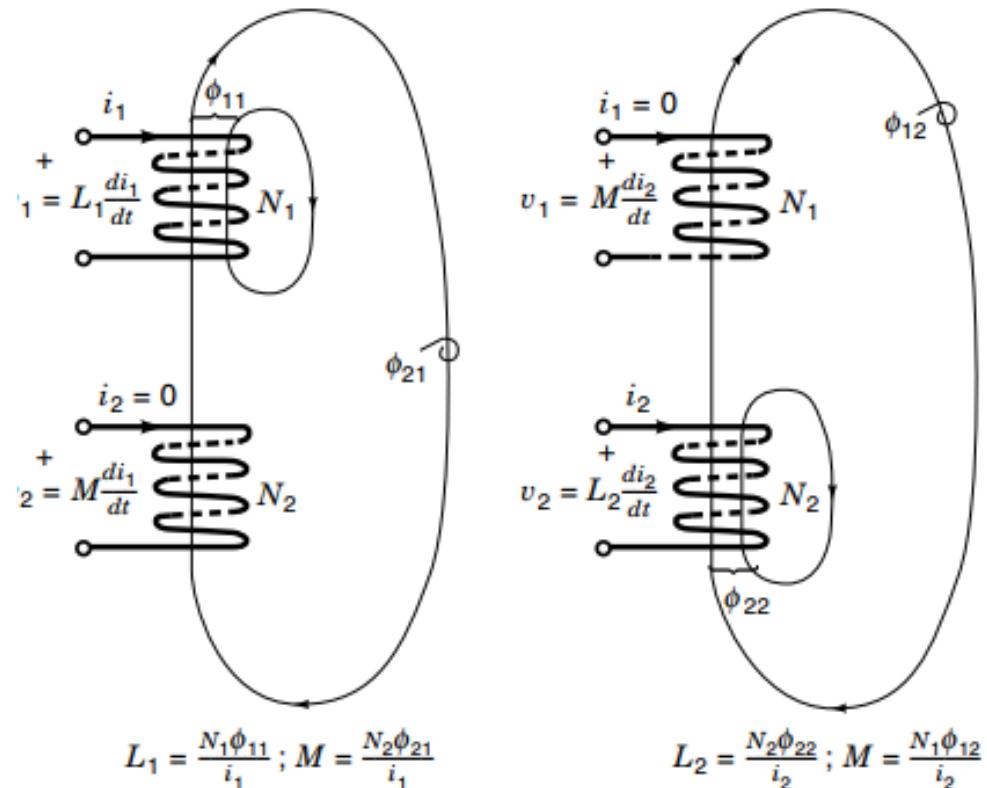


the induced voltage in the inductor

$$e = \frac{d(N\phi)}{di} \frac{di}{dt} = L \frac{di}{dt}$$

The Mutual Inductance Parameter

The inductance of a single isolated coil discussed so far is called its self-inductance. The property of mutual inductance arises when two inductors are in proximity so that some or all the flux lines set up by the current in one inductor also link with the other inductor. Because of this magnetic coupling, when the current in the first inductor varies, its self-flux linkages and also the flux linkages created by it in the other inductor (called mutual flux linkages) vary, thereby inducing voltage in the latter.



When they carry currents i₁ and i₂ simultaneously, their terminal voltages are given by

$$v_1 = L_1 \frac{di_1}{dt} + M \frac{di_2}{dt}$$

$$v_2 = M \frac{di_1}{dt} + L_2 \frac{di_2}{dt}$$

Magnetism

Magnetic properties were discovered in prehistoric times in certain specimens of an iron ore mineral known as magnetite (Fe_3O_4). It was also discovered that pieces of soft iron that rubbed against a magnetic material acquired the same property of acting as a magnet, i.e., attracting other magnets and pieces of iron. The first comprehensive study of magnetism was made by William Gilbert. His greatest contribution was his conclusion that the earth acts as a huge magnet. The word magnetism comes from the district of Magnesia in Asia Minor, which is one of the places at which the magnetic stones were found.

There is a strong similarity between electricity and magnetism. One manifestation of this is that two electrically charged rods have like and unlike ends, very much in the same way as two magnets have opposite ends. In magnets, these ends are called S (south) and N (north) poles. The like poles repeal and the unlike attract. Contrary to electric charges, the magnetic poles always come in pairs. This is proven by breaking magnets into any number of parts. Each part, no matter how small, will have a north pole and a south pole. This suggests that the cause of magnetism is associated with atoms or their arrangements or, more probably, with both.

A fundamental property of magnetism is that moving electric charges (electric current) essentially produce a magnetic field. Knowing this, Albert Einstein came up with explanation of the nature of a permanent magnet. A simplified model of a magnetic field origination process is shown in Fig.3.12a. An electron continuously spins in an eddy motion around the atom.

The electron movement constitutes a circular electric current around the atomic nucleus. That current is a cause for a small magnetic field. In other words, a spinning electron forms a permanent magnet of atomic dimensions. Now, let us imagine that many of such atomic magnets are aligned in an organized fashion so that their magnetic fields add up.

Fig. 3.11 Electric current sets a circular magnetic field around a conductor

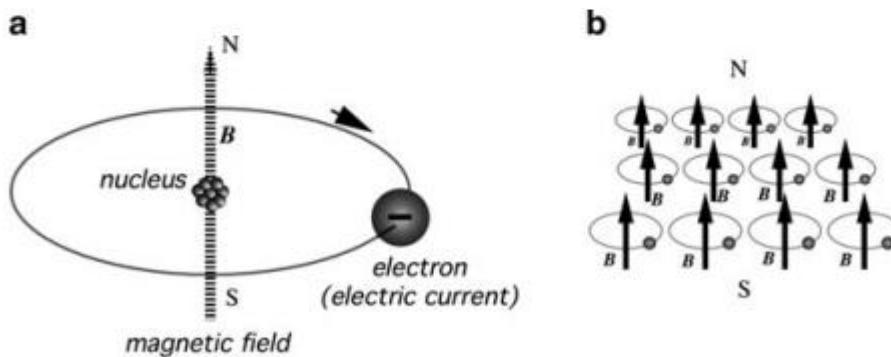
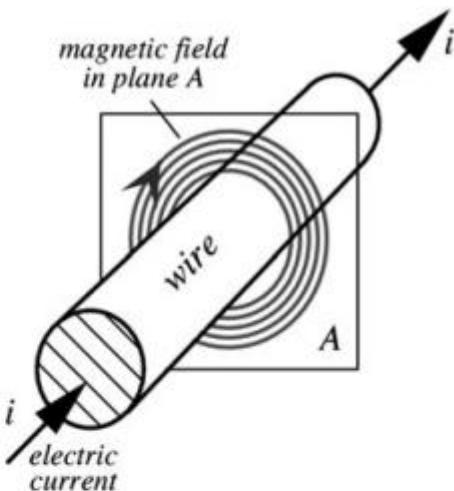


Fig. 3.12 Moving electron sets a magnetic field (a); superposition of field vectors results in a combined magnetic field of a magnet (b)

The process of magnetization then becomes quite obvious – nothing is added or removed from the material - only orientation of atoms is made. The atomic magnets may be kept in the aligned position in some materials, which have an appropriate chemical composition and a crystalline structure. Such materials are called ferro-magnetics.

Michael Faraday pondered the question, “If an electric current is capable of producing magnetism, is it possible that magnetism can be used to produce electricity?” It took him nine or ten years to discover how. If an electric charge is moved across a magnetic field, a deflecting force is acting on that charge. It must be emphasized that it is not important what actually moves either the charge or the source of the magnetic field. What matters is a relative displacement of those.

A discovery that a moving electric charge can be deflected as a result of its interaction with the magnetic field is a fundamental in electromagnetic theory. Deflected electric charges result in an electric field generation, which, in turn, leads to a voltage difference in a conducting material, thus producing an electric current.

To define the magnetic field vector \mathbf{B} we use a laboratory procedure where a positive electric charge q_0 is used as a test object. The charge is projected through the magnetic field with velocity \mathbf{V} . A sideways deflecting force \mathbf{F}_B acts on the charge (Fig.3.13a). By “sideways” we mean that \mathbf{F}_B is at a right angle to \mathbf{V} . It is interesting to note that vector \mathbf{V} changes its direction while moving through the magnetic field. This results in a spiral rather than parabolic motion of the charge (Fig.3.13b). The spiral movement is a cause for a magnetoresistive effect, which forms a foundation for the magnetoresistive sensors. Deflecting force \mathbf{F}_B is proportional to charge, velocity, and magnetic field

$$\mathbf{F}_B = q_0 \mathbf{V} \mathbf{B},$$

Vector \mathbf{F}_B is always at right angles to the plane formed by \mathbf{v} and \mathbf{B} and thus is always at right angles to \mathbf{v} and to \mathbf{B} , that is why it is called a sideways force. The magnitude of magnetic deflecting force according to the rules for vector products, is

$$F_B = q_0 v B \sin \phi, \quad (3.30)$$

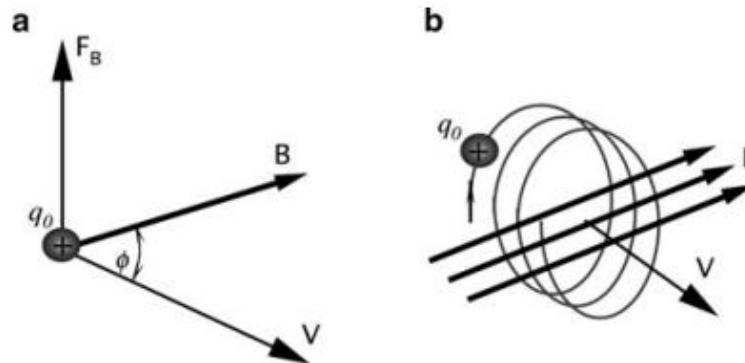


Fig. 3.13 Positive charge projected through a magnetic field is subjected to a sideways force (A); spiral movement of an electric charge in a magnetic field (B)

Where ϕ is the angle between vectors \mathbf{V} and \mathbf{B} . The magnetic force vanishes if \mathbf{V} is parallel to \mathbf{B} . Equation (3.30) is used for the definition of the magnetic field in terms of deflected charge, its velocity, and deflecting force. Therefore, the units of \mathbf{B} is (Newton/coulomb)/(meter/second). In the system SI it is given name tesla (abbreviated T). Since coulomb/second is an ampere, we have $1\text{T} = 1 \text{ newton}/(\text{ampere} \cdot \text{meter})$ or $1 \text{ weber per square meter}$. An older unit for \mathbf{B} still is sometimes in use. It is the gauss: $1 \text{ tesla} = 10^4 \text{ gauss}$.

Piezoelectric Effect

The piezoelectric effect is generation of electric charge by a crystalline material upon subjecting it to stress. The effect exists in natural crystals, such as quartz (chemical formula SiO₂), and poled (artificially polarized) human-made ceramics and some polymers, such as PVDF.

It is said that piezoelectric material possesses ferroelectric properties. The name was given by an analogy with ferromagnetic properties, though there is no iron in most piezoelectrics.

The word piezo comes from the Greek piesZ meaning “to press.” The Curie brothers discovered the piezoelectric effect in quartz in 1880, but very little practical use was made until 1917 when another Frenchman, professor P. Langevin used x-cut plates of quartz to generate and detect sound waves in water. His work led to the development of sonar.

A simplified, yet quite explanatory model of the piezoelectric effect was proposed in 1927 by A. Meissner [4]. A quartz crystal is modeled as a helix (Fig. 3.21a)

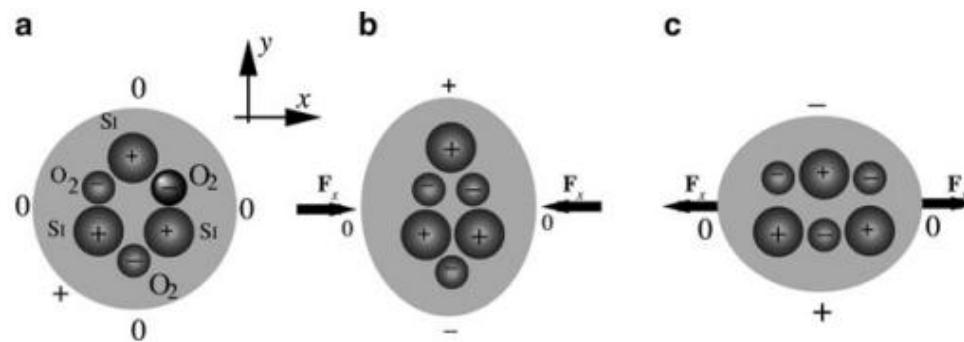


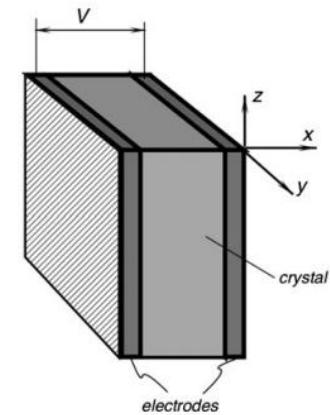
Fig. 3.21 Piezoelectric effect in a quartz crystal

with one silicon, Si, and two oxygen, O₂, atoms alternating around the helix. A quartz crystal is cut along its axes x, y, and z, thus Fig. 3.21a is a view along the z-axis. **In a single crystal-cell there are three atoms of silicon and six oxygen atoms. Oxygen is being lumped in pairs. Each silicon atom carries four positive charges and a pair of oxygen atoms carries four negative charges (two per atom). Therefore a quartz cell is electrically neutral under the no-stress conditions.** When external force, F_x , is applied along the x-axis, the hexagonal lattice becomes deformed. Figure 3.21b shows a compressing force, which shifts atoms in a crystal in such a manner as a positive charge is built up at the silicon atom side and the negative at the oxygen pair side. Thus, the crystal develops an electric charge along they-axis. If the crystal is stretched along the x-axis (Fig. 3.21c), a charge of opposite polarity is built along the y-axis, which is a result of a different deformation. This simple model illustrates that crystalline material can develop electric charge on its surface in response to a mechanical deformation.

To pickup an electric charge, conductive electrodes must be applied to the crystal at the opposite sides of the cut (Fig.3.22). As a result, a piezoelectric sensor becomes a capacitor with a dielectric material between the metal plates, where the dielectric is a piezoelectric crystalline material.

The dielectric acts as a generator of electric charge, resulting in voltage V across the capacitor. Although charge in a crystalline dielectric is formed at the location of an acting force, metal electrodes equalize charges along the surface making the capacitor not selectively sensitive.

**Piezoelectric
sensor is formed by applying
electrodes to a poled
crystalline material**

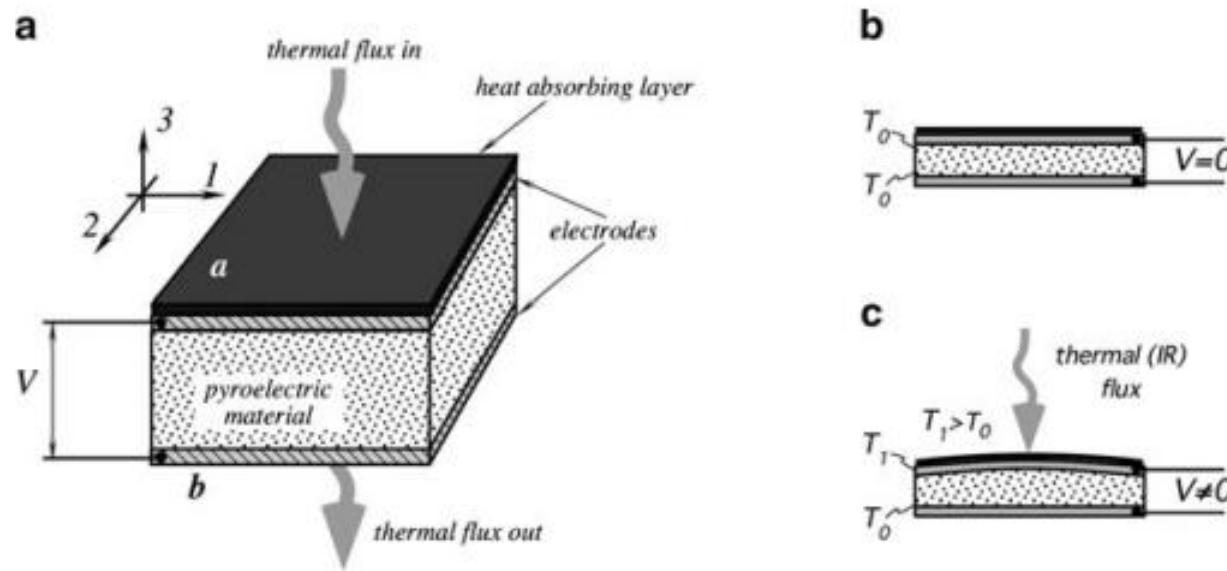


The piezoelectric effect is a reversible physical phenomenon. This means that applying voltage across the crystal produces mechanical strain. It is possible by placing several electrodes on the crystal to use one pair of electrodes to deliver voltage to the crystal and the other pair of electrodes to pick up charge resulting from developed strain.

PYROELECTRIC EFFECT

The pyroelectric materials are crystalline substances capable of generating an electrical charge in response to heat flow. The pyroelectric effect is very closely related to the piezoelectric effect.

Like piezoelectrics, the pyroelectric materials are used in form of thin slices or films with electrodes deposited on the opposite sides to collect the thermally induced charges .



The pyroelectric sensor is essentially a capacitor, which can be electrically charged by flux of heat. The detector does not require any external electrical bias (excitation signal), thus it is a direct converter of heat into electricity. It needs only an appropriate electronic interface circuit to measure the charge. Contrary to thermoelectrics (thermocouples), which produce a steady voltage when two dissimilar metal junctions are held at steady but different temperatures.

Pyroelectrics generate charge in response to a change in temperature. Since a change in temperature essentially requires propagation of heat, a pyroelectric device is a heat flow detector rather than heat detector. When a pyroelectric crystal is exposed to a heat flow (for instance, from an infrared radiation source or from touching a warm object), temperature of the exposed side is elevated and the side becomes a source of heat, which propagates through the pyroelectric material toward its opposite side. Hence, there is an outflow of heat from the crystal to the environment

A crystal is considered to be pyroelectric if it exhibits a spontaneous temperature-dependent polarization. Beside pyroelectric properties, all these materials exhibit piezoelectric properties as well; they generate an electrical charge in response to mechanical stress. Thus, when a pyroelectric sensor is designed, it is very important to minimize all potential mechanical disturbances.

A pyroelectric material can be considered as a composition of a large number of minute crystallities, where each behaves as a small electric dipole. All these dipoles are randomly oriented (Fig.3.23a). Above a certain temperature, known as the Curie point, the crystallities have no dipole moment. Manufacturing (poling) of pyroelectric materials is similar to that of the piezoelectric

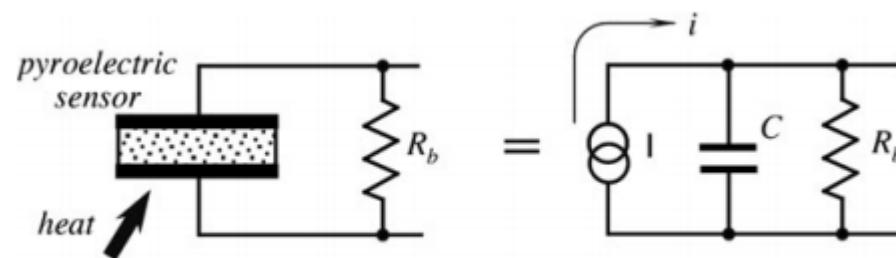


Fig. 3.27 Pyroelectric sensor and its equivalent circuit

There are several mechanisms by which changes in temperature will result in pyroelectricity. **Temperature changes may cause shortening or elongation of individual dipoles. It may also affect the randomness of the dipole orientations due to thermal agitation.** These phenomena are called primary pyroelectricity. There is also secondary pyroelectricity, which, in a simplified way, may be described as a result of the piezoelectric effect, that is, a development of strain in the material due to thermal expansion. Figure b shows a pyroelectric sensor whose temperature T_0 is homogeneous over its volume. That is, the sensor generates zero voltage across the electrodes. Now, let us assume that heat is applied to the top side of the sensor in form of thermal (infrared) radiation. The radiation is absorbed by the heat absorbing layer (e.g., goldblack or organic paint) and warms up the upper side of the pyroelectric material. As a result of the heat absorption, the upper side becomes warmer (the new warmer temperature is T_1), which causes the top side of the material to expand. The expansion leads to flexing of the sensor which, in turn, produces stress and change in a dipole orientation. Being a piezoelectric, the stressed material generates electric charges of the opposite polarities on the electrodes and thus a voltage is observed across the electrodes. Hence, we may regard a secondary pyroelectricity as a sequence of events: thermal radiation-heat absorption-thermally induced stress-electric charge

Let us analyze properties of a pyroelectric material. The dipole moment, M , of the bulk pyroelectric sensor is

$$M = \mu Ah,$$

where μ is the dipole moment per unit volume, A is the sensor's area, and h is the thickness. The charge, Q_a , which can be picked up by the electrodes, develops the dipole moment across the material

$$M_0 = Q_a \cdot h$$

M must be equal to M_0 , so that

$$Q_a = \mu \cdot A \quad (3.74)$$

As the temperature varies, the dipole moment also changes, resulting in an induced charge. Thermal absorption may be related to a dipole change, so that μ must be considered as function of both temperature, T_a , and an incremental thermal energy, ΔW , absorbed by the material

$$\Delta Q_a = A \cdot \mu(T_a, \Delta W) \quad (3.75)$$

Figure 3.27 depicts a pyroelectric detector (pyroelectric sensor) connected to a resistor R_b that represents either the internal leakage resistance or a combined input resistance of the interface circuit, which is connected to the sensor. The equivalent electrical circuit of the sensor is shown at right. It consists of three components: (1) the current source generating a heat induced current, i (remember that a current is a movement of electric charges), (2) the sensor's capacitance, C , and (3) the leakage resistance, R_b .

The output signal from the pyroelectric sensor can be taken in form of either charge (current) or voltage, depending on the application. Being a capacitor, the pyroelectric device is discharged when connected to a resistor, R_b . Electric current through the resistor and voltage across the resistor represent the heat flow induced charge. It can be characterized by two pyroelectric coefficients [22]

$$P_Q = \frac{dP_s}{dT} \quad \text{Pyroelectric charge coefficient}$$
$$P_V = \frac{dE}{dT} \quad \text{Pyroelectric voltage coefficient} \quad (3.76)$$

where P_s is the spontaneous polarization (which is the other way to say “electric charge”), E is the electric field strength, and T is the temperature in K. Both coefficients are related by way of the electric permittivity, ϵ_r , and dielectric constant, ϵ_0

HALL EFFECT

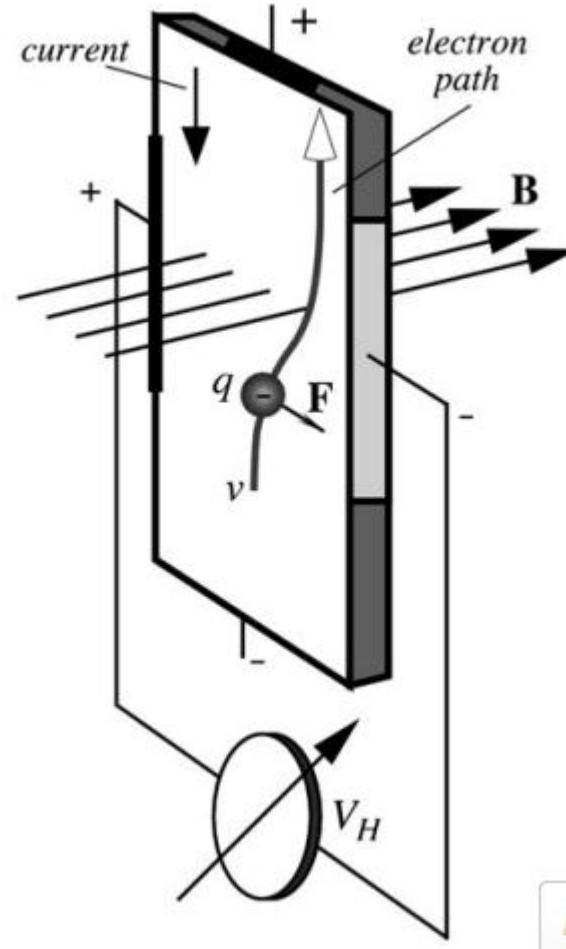
This physical effect was discovered in 1879 in Johns Hopkins University by E. H. Hall. Initially, the effect had a limited, however, a very valuable application as a tool for studying electrical conduction in metals, semiconductors, and other conductive materials. Nowadays, the Hall sensors are used to detect magnetic fields, position, and displacement of objects

The effect is based on the interaction between moving electric carriers and an external magnetic field. In metals, these carriers are electrons. When an electron moves through a magnetic field, upon it acts a sideways force

$$\mathbf{F} = q\mathbf{v}\mathbf{B},$$

Where q is an electronic charge, v is the speed of an electron, and B is the magnetic field. Vector notations (bold face) are an indication that the force direction and its magnitude depend on the spatial relationship between the magnetic field and the direction of the electron movement. The unit of B is 1 tesla=1 newton/(ampere-meter).

Many Hall effect sensors are fabricated from silicon and fall into two general categories: the basic sensors and the integrated sensors. Other materials used for the element fabrication include InSb, InAs, Ge, and GaAs. In the silicon element, an interface electronic circuit can be incorporated into the same wafer. This integration is especially important since the Hall effect voltage is quite small



Thermoelectric Effects

SEEBECK EFFECT

In 1821, Thomas Johann Seebeck (1770–1831), an Estonian born and Berlin and Gōttingen educated physician, accidentally joined semicircular pieces of bismuth and copper while studying thermal effects on galvanic arrangements A nearby compass indicated a magnetic disturbance (Fig. 3.34a). Seebeck experimented repeatedly with different metal combinations at various temperatures, noting related magnetic field strengths. Curiously, he did not believe that an electric current was flowing, and preferred to describe that effect as “thermomagnetism”

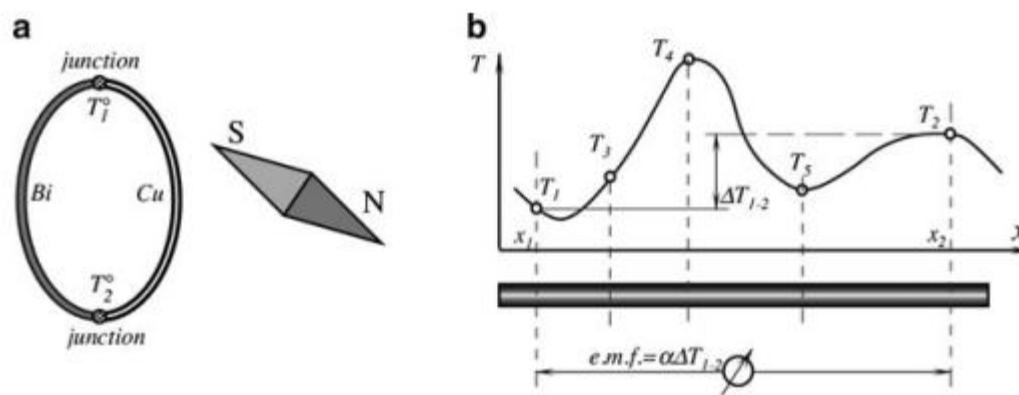
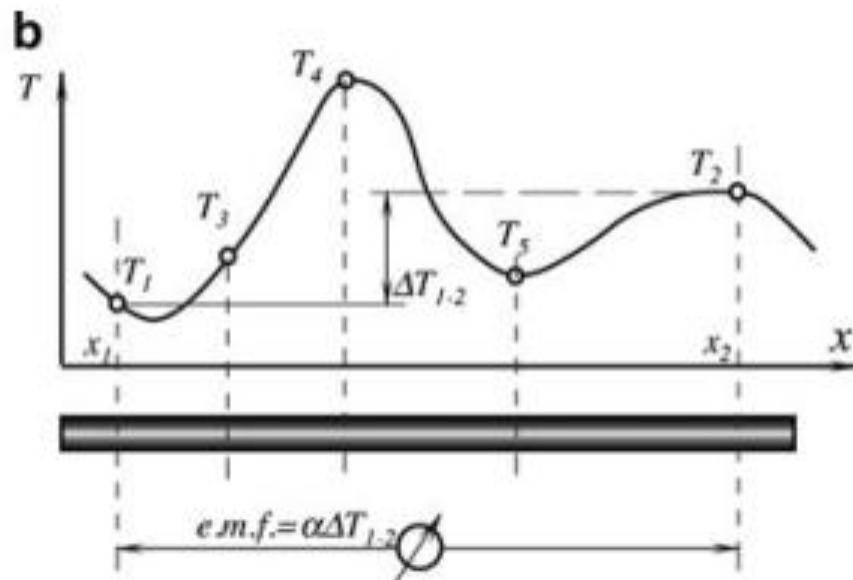


Fig. 3.34 Seebeck experiment (a), varying temperature along a conductor is a source of a thermoelectric e.m.f. (b)

If we take a conductor and place one end of it into a cold place and the other end into a warm place, energy will flow from the warm to cold part. The energy takes the form of heat. The intensity of the heat flow is proportional to the thermal conductivity of the conductor. Besides, the thermal gradient sets an electric field inside the conductor.

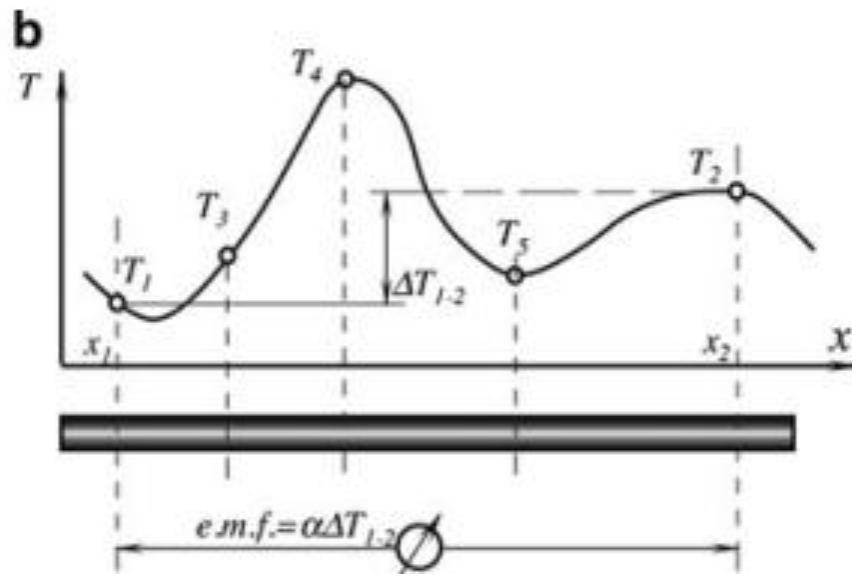
The field results in incremental voltage



$$dV_a = \alpha_a \frac{dT}{dx} dx,$$

where dT is the temperature gradient across small length dx and α_a is the *absolute* Seebeck coefficient of the material [29]. If the material is homogeneous, α_a is not a function of length and (3.87) reduces to

$$dV_a = \alpha_a dT. \quad (3.88)$$



$$dV_a = \alpha_a dT. \quad (3.88)$$

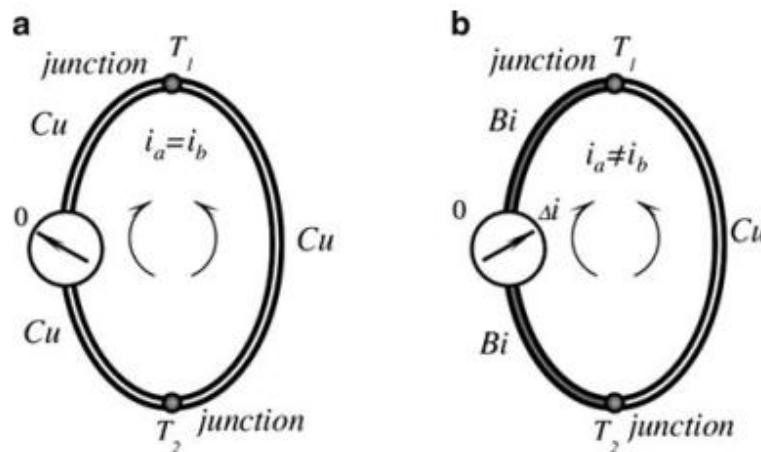


Fig. 3.35 Thermoelectric loop

Joints of identical metals produce zero net current at any temperature difference (a); joints of dissimilar metals produce net current Δi (b)

Equation (3.88) is a principle mathematical expression of a thermoelectric effect. Figure 3.34b shows a conductor having nonuniform temperature T along its length x . A temperature gradient between any arbitrary points defines an electromotive force (e.m.f.) between these points. Other possible temperatures between the selected points (temperatures T_3 , T_4 , and T_5 , for example) make no effect whatsoever on the value of e.m.f. between points 1 and 2.

What happens when two conductors are joined together? Free electrons in metal may behave as an ideal gas. Kinetic energy of electrons is a function of the material temperature. However, in different materials, energies and densities of free electrons are not the same. When two dissimilar materials at the same temperature are brought into a contact, free electrons diffuse through the junction.

The electric potential of the material accepting electrons becomes more negative at the interface, while the material emitting electrons becomes more positive. Different electronic concentrations across the junction set up an electric field, which balances the diffusion process and the equilibrium is established. If the loop is formed and both junctions are at the same temperature, the electric fields at both junctions cancel each other, which is not the case when the junctions are at different temperatures.

It can be stated that the thermoelectric properties of a conductor are in general just as much bulk properties as are the electrical and thermal conductivities. Coefficient α_a is a unique property of a material. When a combination of two dissimilar materials (A and B) is used, the Seebeck potential is determined from a differential Seebeck coefficient:

$$\alpha_{AB} = \alpha_A - \alpha_B \quad (3.89)$$

and the net voltage of the junction is

$$dV_{AB} = \alpha_{AB}dT. \quad (3.90)$$

The above equation can be used to determine a differential coefficient

$$\alpha_{AB} = \frac{dV_{AB}}{dT} \quad (3.91)$$

Note that the Seebeck coefficient is not really a constant. It is temperature dependent and thus the Seebeck potential will be different at different temperatures. **The Seebeck effect is a direct conversion of thermal energy into electric energy.**

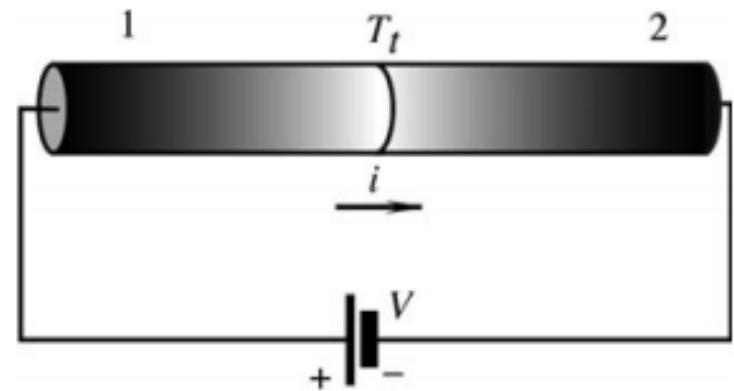
PELTIER EFFECT

In the early nineteenth century, a French watchmaker-turned-physicist, Jean Charles Athanase Peltier (1785–1845) discovered that if electric current passes from one substance to another and then heat may be given or absorbed at the junction.

Heat absorption or production is a function of the current direction

$$dQ_P = \pm pi dt, \quad (3.95)$$

Where i is the current and t is time. The coefficient p has a dimension of voltage and represents thermoelectric properties of the material. It should be noted that heat does not depend on temperature at the other junction.



The Peltier effect concerns the reversible absorption of heat, which usually takes place when an electric current crosses a junction between two dissimilar metals. The effect takes place whether the current is introduced externally or is induced by the thermocouple junction itself (due to Seebeck effect).

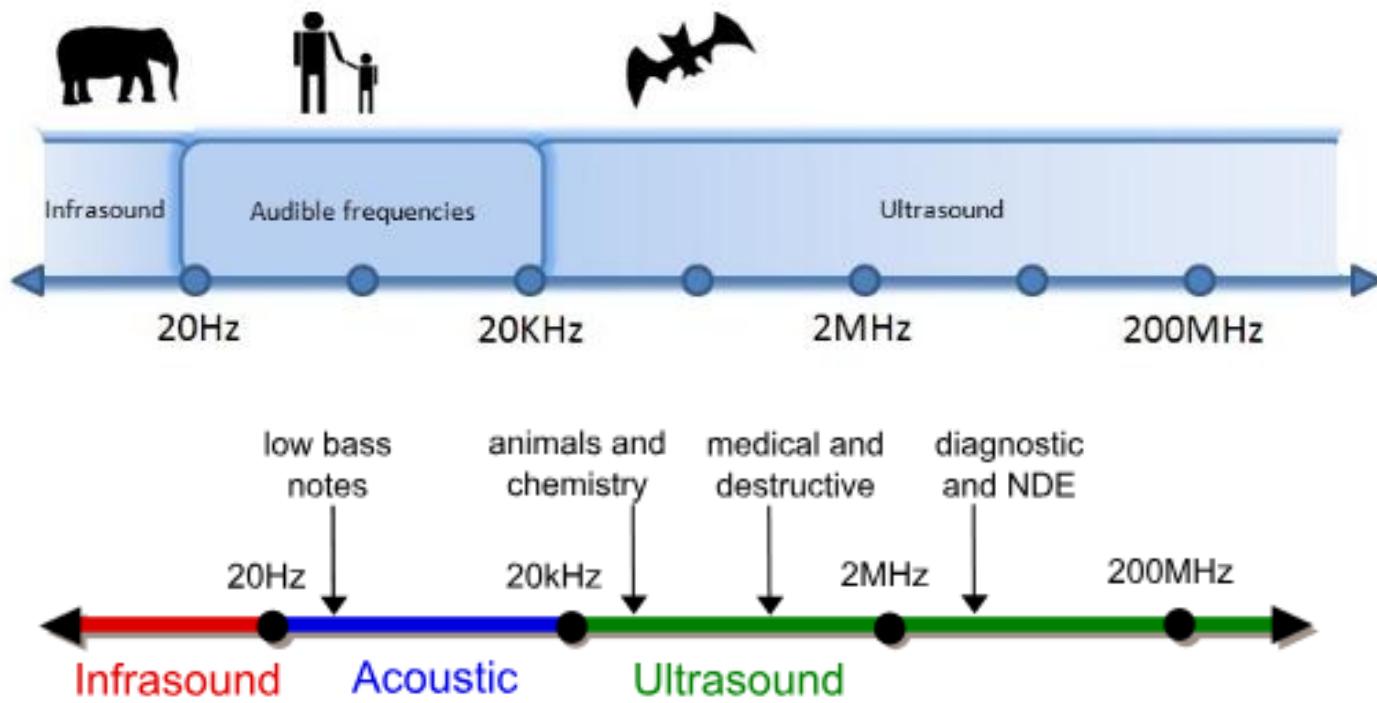
The Peltier effect is used for two purposes: it can produce heat or “produce” cold (remove heat), depending on the direction of electric current through the junction. This makes it quite useful for the devices where precision thermal control is required. Apparently, the Peltier effect is of the same nature as the Seebeck effect.

The magnitude and direction of Peltier heat do not depend in any way on the actual nature of the contact. It is purely a function of two different bulk materials, which have been brought together to form the junction and each material makes its own contribution depending on its thermoelectric properties. The Peltier effect is a basis for operation of thermoelectric coolers, which are used for cooling of the photon detectors operating in the far infrared spectral range

UNIT-II

Acoustic waves

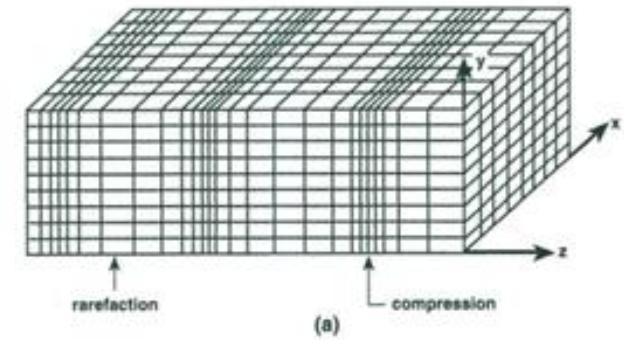
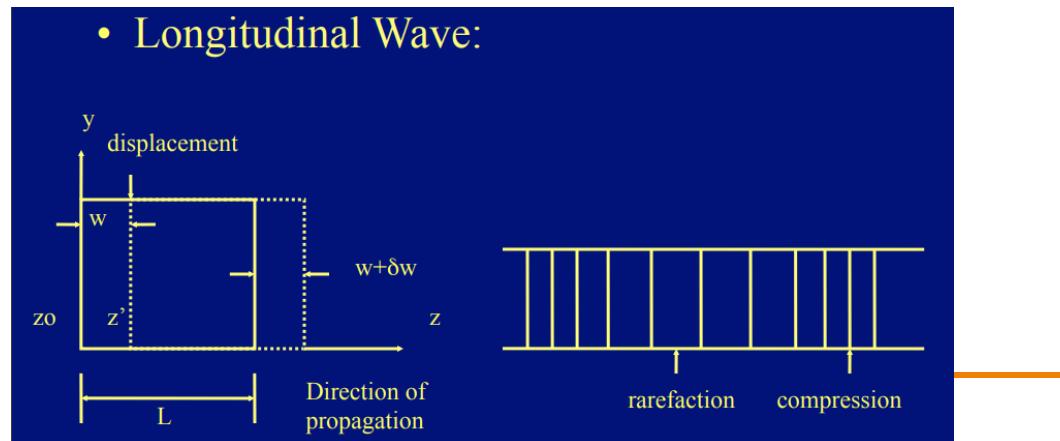


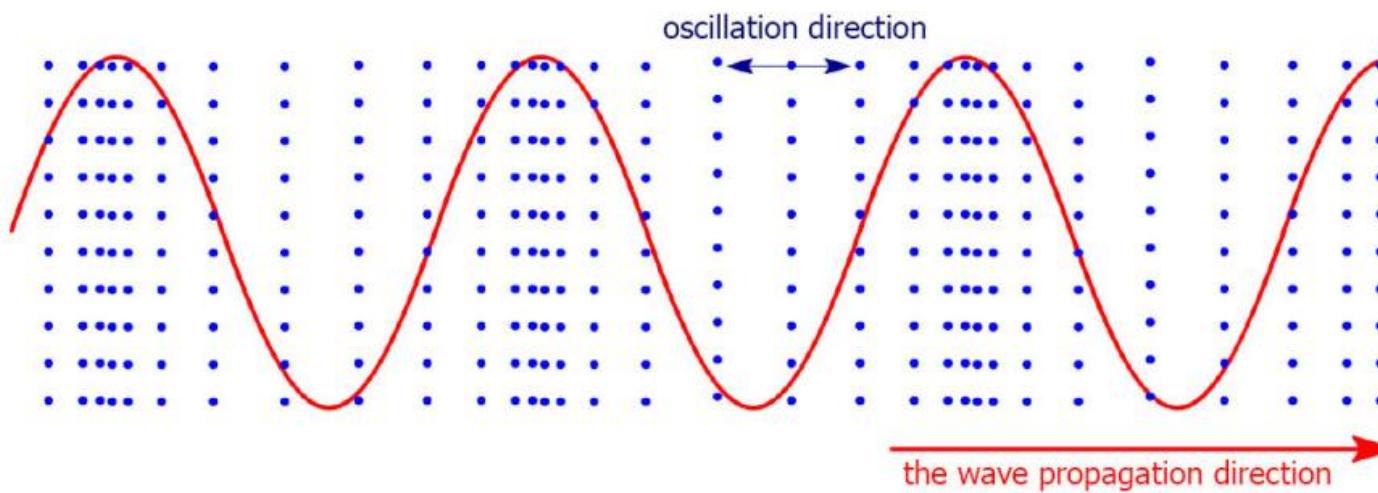


What are acoustic waves?

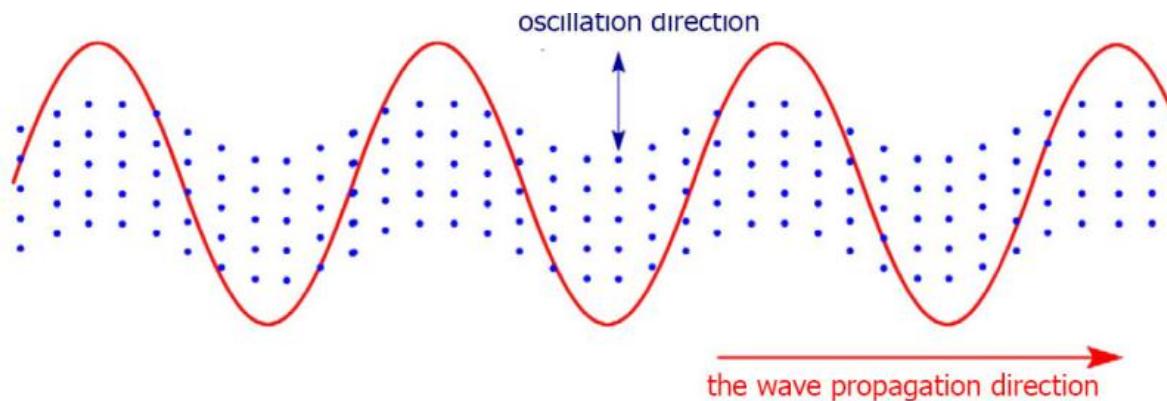
- The acoustic waves are mechanical waves that need a medium to carry their energy from any position to another.
- The acoustic waves, which are mechanical waves, are incapable to travel over a vacuum.

The **longitudinal wave**, in which the motion of a particle in the acoustic medium is only in the direction of propagation. Thus when a force is applied to the acoustic medium, the medium extends or contracts in the z direction, as shown in Fig

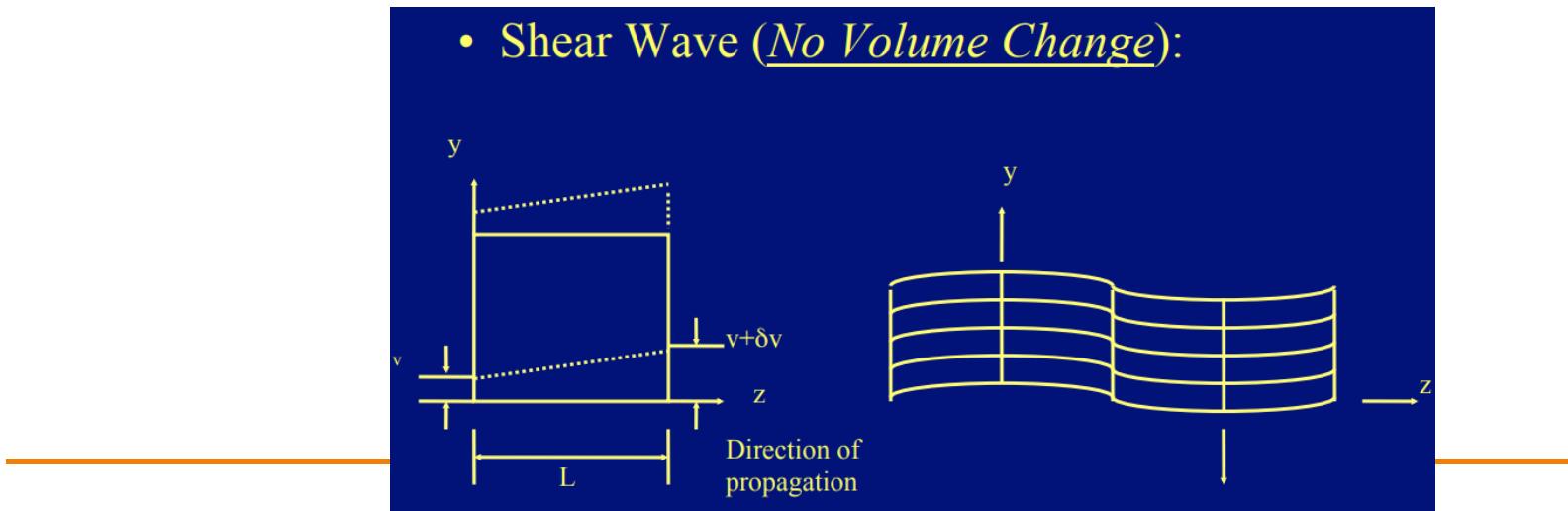




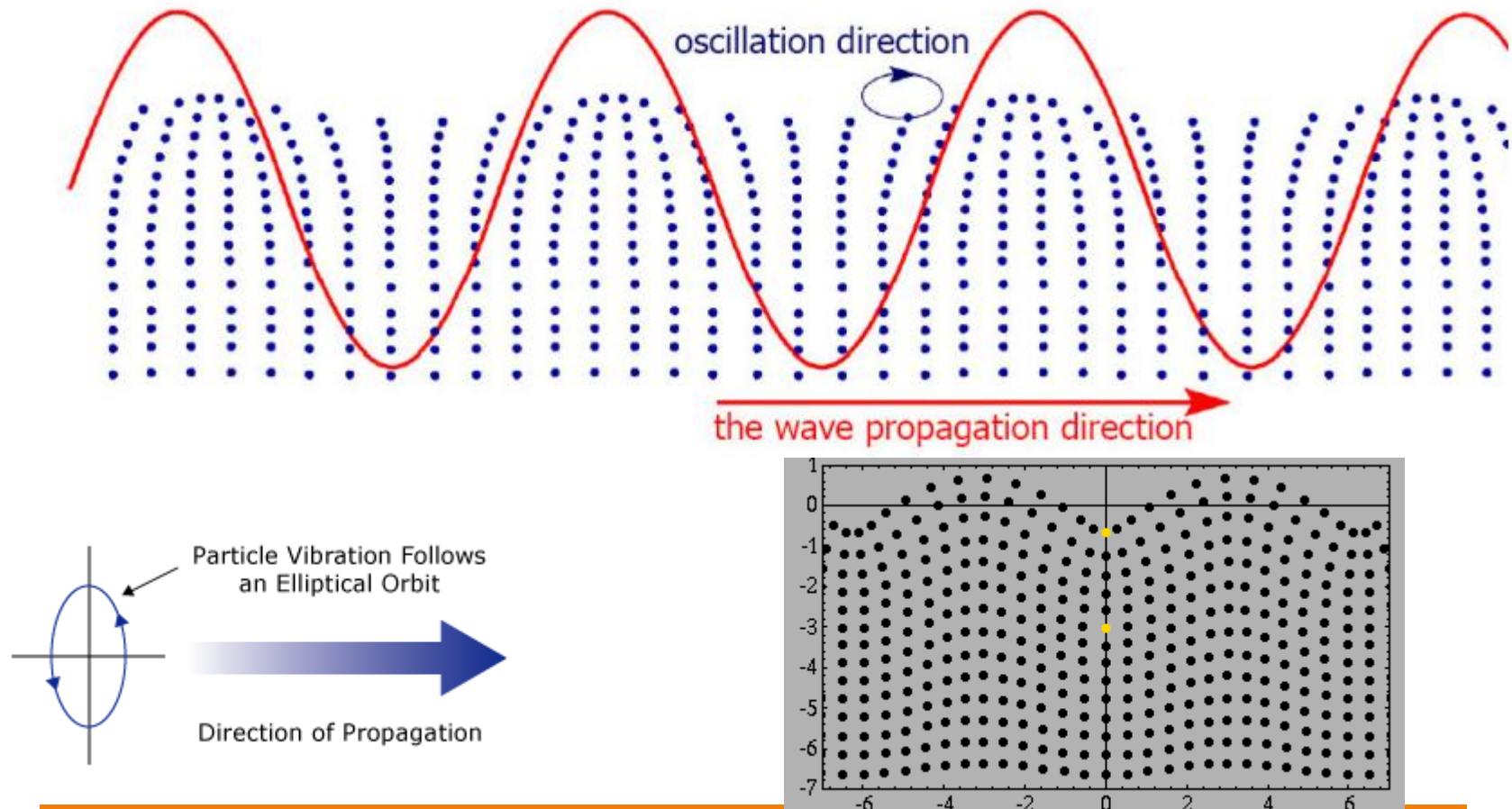
Transverse waves – are acoustic waves, in which the particles of the medium vibrate perpendicular to the direction of wave propagation. Under the influence of waves the material undergoes shear deformation.



- Shear Wave (No Volume Change):



Surface waves (Rayleigh waves) – are the acoustic waves that propagate along the boundary between two media, in layer with a typical thickness of 1.5 to 2 wavelength $(1.5\text{--}2) \cdot \lambda$. Fluctuations medium particles of the wave combine the longitudinal and transverse vibrations, the particles move along locked elliptical trajectories



There are various kinds of acoustic waves.

Bulk acoustic waves are acoustic waves propagated through the bulk substrate material.

Surface acoustic waves

propagate along the surface of a substrate.

There are some types of the surface acoustic waves. In the case of the Rayleigh waves particles in the surface layer move both up and down and back and forth tracing out elliptical paths.

Typical velocities are 6000 m/s for longitudinal waves and 3000 m/s for transverse waves.

Piezoelectric materials



On the surface of the piezoelectric material, electric charges are produced due to applying any mechanical stress, such as the one applied from the sound waves. In the direct piezoelectric effect, the convinced charges are proportional to the mechanical stress.

This piezoelectricity effect has a wide application in detecting the acoustic sound and electronic frequency/high voltage generation.

The negative and positive electrical charges are divided in the piezoelectric crystal leading to electrically neutral overall crystal.

This symmetry is disturbed with applying a stress to the piezoelectric materials, and the asymmetry of the charge produces a voltage.

The piezoelectric material can be categorized according its cutting procedure, namely, shear, longitudinal, and transverse, which are defined as follows

Shear effect: due to this effect, the generated charges are independent of the element's shape/size and proportional to the applied forces

Longitudinal effect: due to this effect, the displaced amount of charge is independent of the piezoelectric element's shape/size and proportional to the applied forces.

Transverse effect: due to this effect, the applied force along the y-axis transfers charges along the x-direction, which is perpendicular to the force line.

Piezoelectric Materials

Quartz remains still one of the most important piezoelectric materials and is in great demand.

A large part of this demand is increasingly being met by mass production of artificial quartz based on the technique of hydrothermal growth.

During World War II., about 75 million quartz plates were produced for the armed forces of the United States. As the supply of good quality raw material started to decline, interest in artificial growth was renewed.

The size of synthetic crystals increased tenfold thanks to the concentrated effort and modern technology. As a result, we now have a virtually unlimited supply.

Synthetic quartz has been commercially available since 1958. Today, the manufacture of synthetic quartz has become an important industry.

Rochelle salt was often used in various transducers because it has a great piezoelectric effect. In spite of being a piezoelectric crystal with the highest electromechanical coupling coefficient, Rochelle salt is of limited use because it is soluble in water, and some parameters do not have a suitable temperature characteristic

Since 1935 attempts were made to produce piezoelectric crystals, which could replace quartz. Piezoelectric crystals such as ammonium and potassium salts ($\text{NH}_4\text{H}_2\text{PO}_4$ – ADP, KH_2PO_4 – KDP), ethylene diamine tartrate (EDT), dipotassium tartrate (DKT) and lithium sulphate monohydrate (LH) were developed.

Many of these materials are no longer in use as a result of development and production of artificial quartz, ferroelectric crystals or piezoelectric ceramics. The discovery of the strong piezoelectric properties of ferroelectric ceramics is a major milestone in applications of piezoelectricity.

With the exception to quartz few single crystals are used in piezoelectric devices. Popular choices are LiNbO_3 , LiTaO_3 . In recent times considerable attention is given to the synthesis of berlinitite (aluminium phosphate) and langasite. This material could combine some useful characteristics of quartz with a high coupling factor. The single crystals are anisotropic, exhibiting different material properties depending on the cut of the materials and the direction of bulk or surface wave propagation.

In the last 20 years intensive research was done on the synthesis of new piezoelectric crystal compositions.

The main objectives consisted in the search for new single crystals with improved piezoelectric properties like high sensitivity, high electromechanical coupling and high stability at elevated temperatures.

More than 40 single crystal compounds used for sensor application like force sensors, strain sensors and high temperature pressure sensors were synthesized.

Pyroelectric and piezoelectric materials, which generate an electric field with the input of heat and stress, respectively, are called “smart” materials.

These offdiagonal couplings have corresponding converse effects, the electrocaloric and converse piezoelectric effects, and both “sensing” and “actuating” functions can be realized in the same materials. “Intelligent” materials must possess a function, which is adaptive to changes in environmental conditions.

An elastic medium behaves as a distributed mass-spring system in which displacement of a single element results in the propagation of a disturbance through out the medium. A particle at a free surface is different from one interior to the solid, in that it is constrained by adjacent particles from only one side. Thus, disturbances at a surface can behave somewhat differently from those in the interior of a solid.

Like the one-dimensional vibrating string, particle displacement in the solid is a function both of time and position.

The waves that can propagate in a solid depend upon both the properties of the solid and its boundaries

Acoustic sensing



There are several methods to measure the sound waves, namely, the

- (i) frequency;**
- (ii) wavelength, where the distance that the disturbance travels via the medium represents a complete wave cycle;**
- (iii) amplitude which is related to the sound volume, loudness, and intensity;**
- (iv) phase; and (v) speed of sound that depends on the medium state/type, which is affected by the elasticity and the inertia.**

Acoustic (sound) sensors can detect and transmit vibrations and sound waves from **infrasound** (very low frequencies) up to **ultrasound** (very high frequencies).

Acoustic wave sensors detect acoustic or mechanical waves produced by the human body.

During the propagation of the acoustic wave through the body, the propagation path characteristics change, which affect the amplitude/velocity of the acoustic wave.

Measuring the phase/frequency characteristics of sensed signals reflects the occurred changes in the velocity, which is correlated to the consistent physical measured quantity

Acoustic wave sensors

Acoustic wave sensors are so named because they utilize a mechanical, or acoustic, wave as the sensing mechanism.

As the acoustic wave propagates through or on the surface of the material, any changes to the characteristics of the propagation path affect the velocity and/or amplitude of the wave.

Changes in velocity can be monitored by measuring the frequency or phase characteristics of the sensor and can then be correlated to the corresponding physical quantity that is being measured.

Virtually all acoustic wave devices and sensors use a piezoelectric material to generate the acoustic wave.

PIEZOELECTRIC SUBSTRATE MATERIALS FOR ACOUSTIC WAVE SENSORS

There are several piezoelectric substrate materials that may be used for acoustic wave sensors and devices. The most common are quartz (SiO_2) and lithium tantalate (LiTaO_3), and to a lesser degree, lithium niobate (LiNbO_3). Each material has specific advantages and disadvantages, which include cost, Temperature dependence, attenuation, and propagation velocity.

Other materials that have commercial potential include gallium arsenide (GaAs), silicon carbide (SiC), langasite (LGS), zinc oxide (ZnO), aluminum nitride (AlN), lead zirconium titanate (PZT), and polyvinylidene fluoride (PVDF).

THE DIRECT AND THE CONVERSE PIEZOELECTRIC EFFECT

The direct piezoelectric effect may be defined as the change of electric polarization proportional to the strain. A material is said to be piezoelectric if the application of an external mechanical stress gives rise to dielectric displacement in this material. This displacement manifests itself as internal electric polarization.

It should be noted that the piezoelectric effect strongly depends on the symmetry of the crystal. A crystal having sufficiently low symmetry produces electric polarization under the influence of external mechanical force.

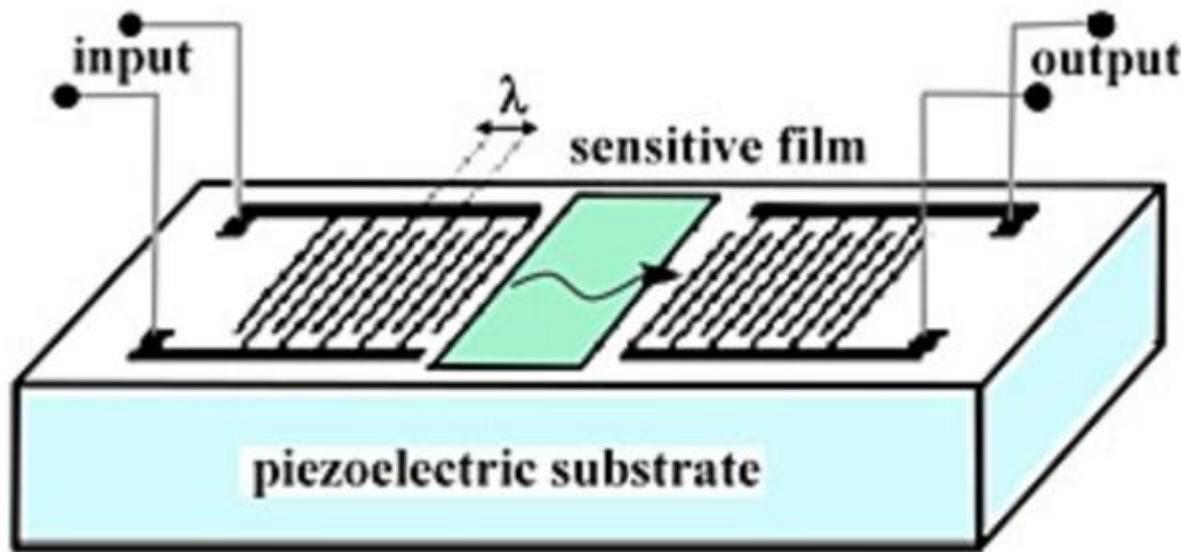
Closely related is the converse effect, whereby a piezoelectric crystal becomes strained if an external electric field is applied. Both effects are the manifestation of the same fundamental property of the non-centric crystal.

saw sensors



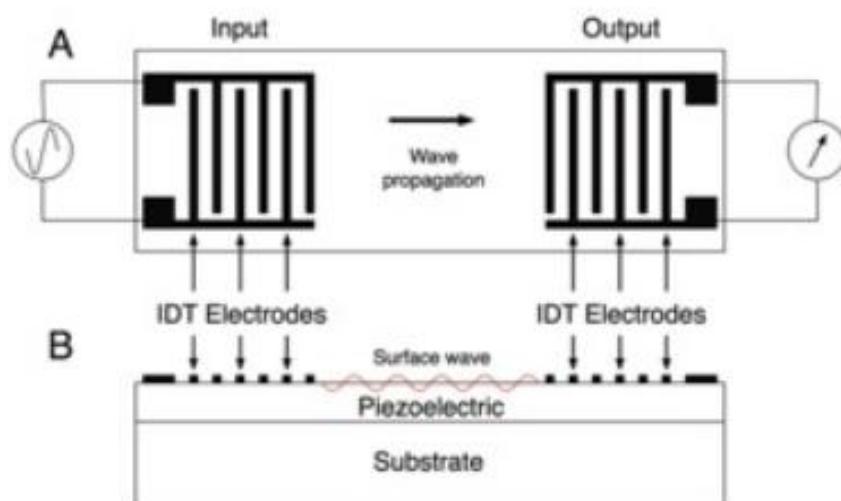
Surface Acoustic Wave Sensors

The largest consumer for SAW devices is the telecommunications industry, where SAWs are used as band pass filters in the radio frequency (RF) range in mobile phones and base stations. Some of the emerging applications for acoustic wave devices include sensors for the automotive (e.g., torque and pressure sensors), medical (e.g., chemical sensors) and industrial markets (e.g., humidity, temperature and mass sensors). SAWs are also used for advanced touch screens..



Surface Acoustic Wave devices (SAW) are acoustic devices commonly used for high frequency filtering applications such as in mobile telephony. Acoustic waves are generated by applying an alternating voltage across electrodes on the piezoelectric. As the piezoelectric starts to oscillate it generates an acoustic wave, **the resonant frequency being a function of the spacing between the electrodes and the acoustic wave velocity of the material**. This wave is received at a second set of electrodes and converted back to **an a.c. voltage**.

The resonant frequency of such devices can be increased by either reducing the electrode spacing in the IDT or using a material of higher sound wave velocity. Diamond has the highest acoustic wave velocity of any material and thus offers the highest possible resonant frequencies of any SAW device.



Schematic of a Surface Acoustic Wave (SAW) device, plan view (A) and side view (B). The Inter - Digitated - Transducer (IDT) Electrodes are used to apply an alternating potential across the piezoelectric at the input. This generates a surface acoustic wave which is converted back to an electric potential at the output.

Diamond is not a piezoelectric and thus has to be combined with a piezoelectric material such as AlN in order to fabricate a discrete device. The integration of AlN onto diamond films requires very smooth surfaces, typically below 2 nm over several micron.

SAW parameters

The most important parameter for SAW device design is the center frequency, which is determined by the period of the IDT fingers and the acoustic velocity.

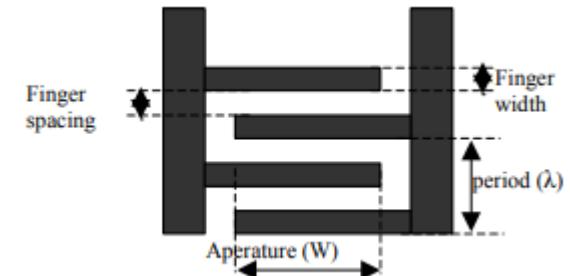
The governing equation that determines the operation frequency is:

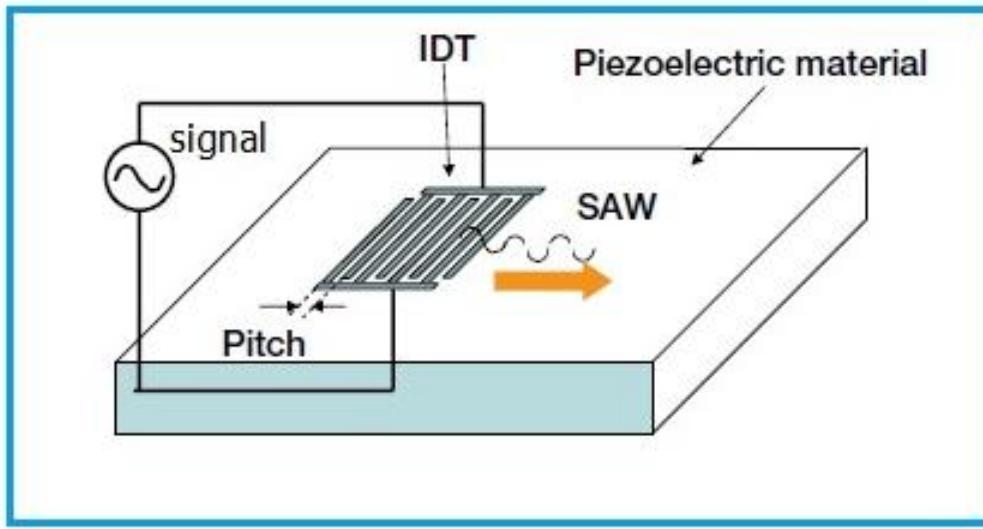
$$f_0 = v_{SAW}/f$$

Where f is the wavelength, determined by the periodicity of the IDT and v_{SAW} is the acoustic wave velocity . For the technology being used in this research:

$$f = p = \text{finger width} \times 4$$

with the finger is determined by the design rule of the technology which sets the minimum metal to metal distance. v_{SAW} is surface acoustic wave velocity.





Electrical signals are transformed into surface acoustic waves with the help of piezoelectric materials. When electric field is applied, piezoelectric material changes shape i.e. gets deformed.

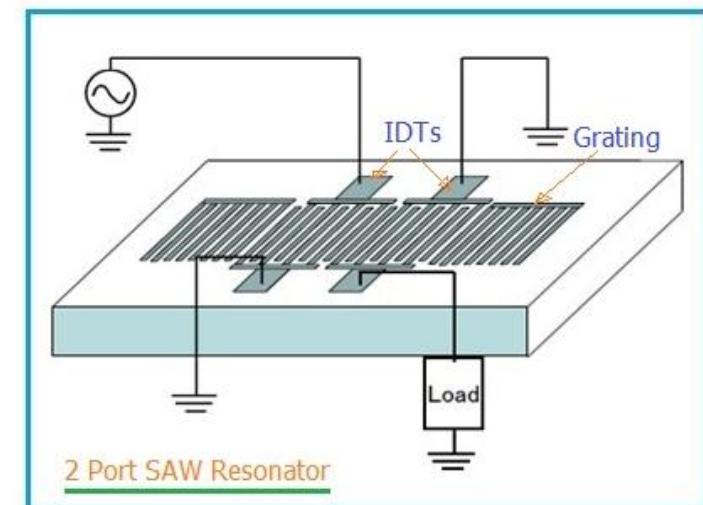
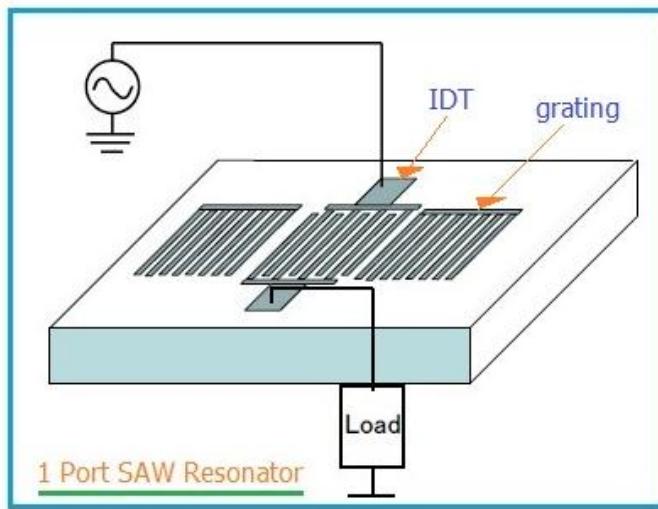
Hence by forming comb shaped electrodes known as IDT (Inter Digital Transducers) and with signal application, SAW waves are generated. This is shown in the figure-1. The velocity of acoustic wave usually depends on substrate type or wave type (rayleigh wave,) used.

The variation in pitch of IDT electrodes will change frequency of generated waves.

What is Resonator?

The device which naturally oscillates (or resonates) at some frequency with higher magnitude is known as resonator. The frequency at which maximum amplitude is obtained is known as resonant frequency or resonance frequency. The oscillations can be either EM (electro-magnetic) or acoustic in nature.

The resonator is used to generate wave at specific frequency or to select particular frequencies from signal. The resonator which uses structure as shown in the figures below are known as SAW resonator.



2 port SAW resonator. It uses two IDTs, one for input and the other for output between the gratings. ~~The 2-port SAW resonator is used as filter due to its propagation characteristics between IDTs.~~

Acoustic Sensing Remote Sensing with Sound

Acoustics is a key component of ocean observation at all scales. This is a result of the rapid attenuation of optical and electromagnetic waves in water that renders useless in the oceans most of the primary remote sensing tools we use in the atmosphere and in space.

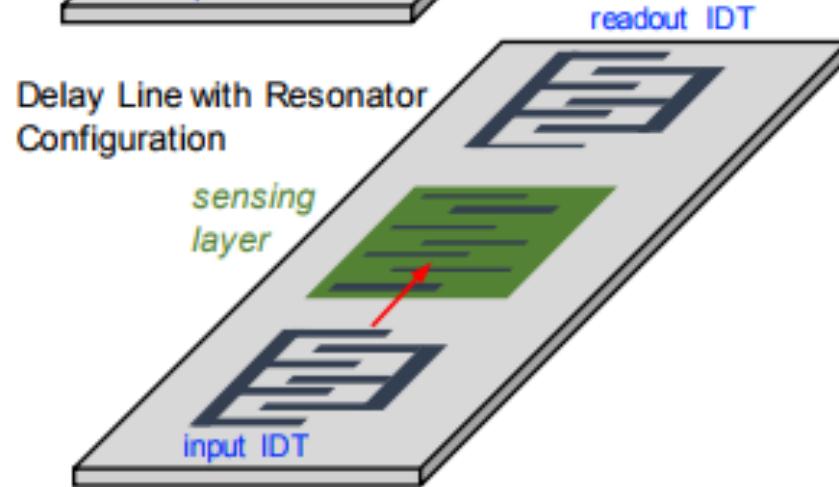
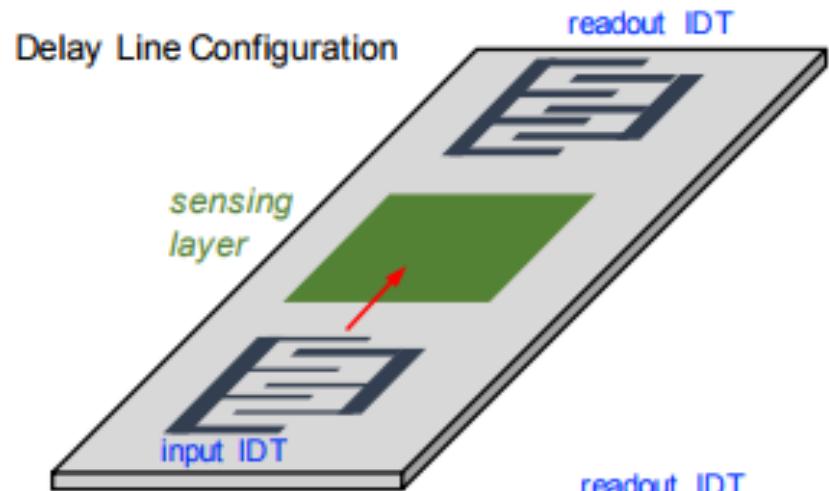
Acoustic waves, on the other hand, can travel great distances underwater. They provide the primary mechanism for remote sensing in the ocean by mankind and many marine creatures.

Approximately 4-5 billion SAW devices are produced each year

SAW-Based Chemical Sensing Fundamentals

The principle behind SAW-based sensing devices is that chemical targets that adsorb on a functionalized surface on the SAW substrate fundamentally change the frequency of the SAW, and measurement of the frequency shift (Δf) can be correlated to the concentration of the target species.

The main advantage associated with SAW sensors is their high sensitivity and capability for the trace detection of a wide variety of chemical materials including bio/chemicals, organic and inorganic vapors, and explosives.

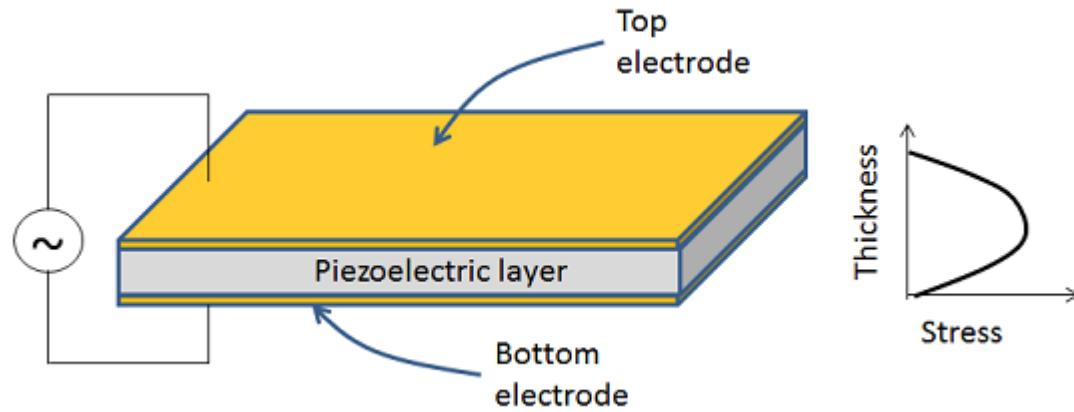


The distance between the input IDT and readout IDT is the delay line, and the sensing layer is deposited in this region of the substrate.

On the bottom is a similar configuration but with additional IDT electrodes beneath the sensing layer. This resonator IDT acts as a reflector to set up a resonant cavity and increase sensitivity.

What is a BAW Resonator?

A BAW resonator is an electromechanical device in which a standing acoustic wave is generated by an electrical signal in the bulk of a piezoelectric material. In the simplest configuration, a device will consist of a piezoelectric material (typically quartz, AlN, or ZnO) sandwiched between two metallic electrodes. The natural frequency of the material and the thickness are used as design parameters to obtain a desired operating frequency.



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

$$\lambda_n = 2t/n$$

where $n = 1, 3, 5, \dots$

The displacements associated with $n = 1$ and 3 resonant shear waves are shown in Figure. The $n = 1$ configuration is commonly called the fundamental mode, and the $n = 3, 5, \dots$, 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

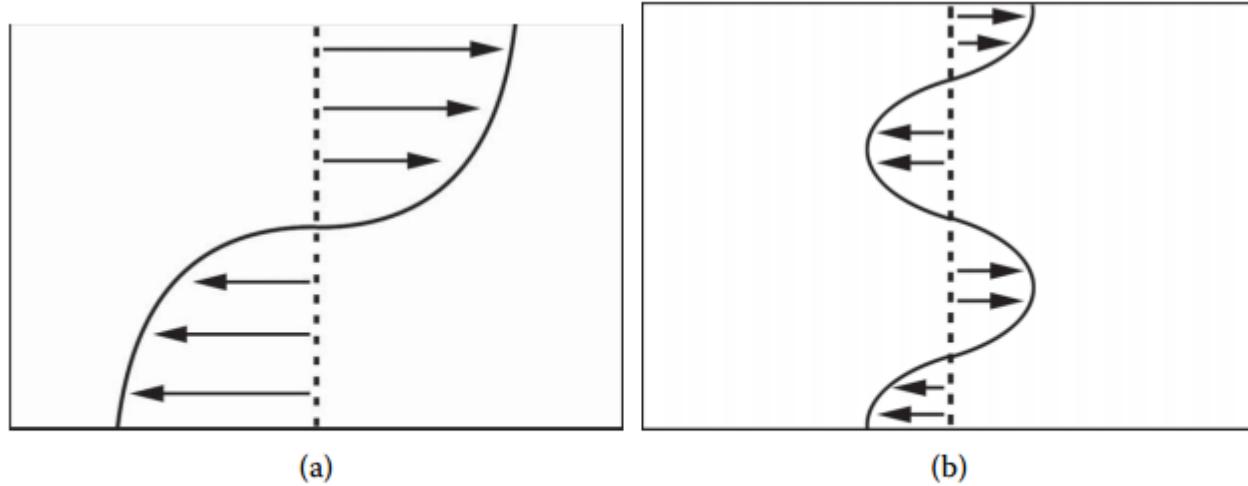
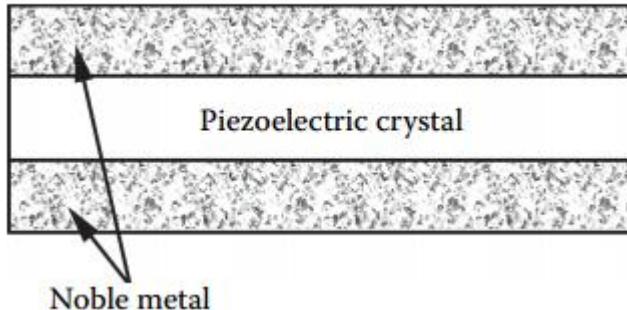


FIGURE 3.24

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

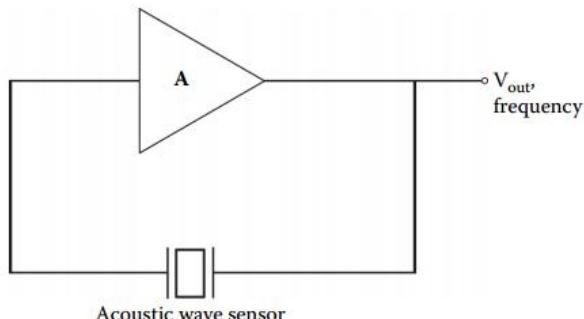
crystal. These electrodes are vacuum deposited noble metals, usually gold. The exact configuration associated with the electrodes could be a gold film on each crystal surface or a more complicated structure, each of which is shown Figure 3.25.



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 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.

Crystal-controlled oscillator.



Among the piezoelectric substrate materials that can be used for acoustic wave sensors and devices, the most common are quartz (SiO_2), lithium tantalate (LiTaO_3), and, to a lesser degree, lithium niobate (LiNbO_3). An interesting property of quartz is that it is possible to select the temperature dependence of the material by the cut angle and the wave propagation direction.

The advantage of using acoustic waves (vs electromagnetic waves) is the slow speed of propagation (5 orders of magnitude slower). For the same frequency, therefore, the wavelength of the elastic wave is 100,000 times shorter than the corresponding electromagnetic shortwave.

Magnetic sensors: effects and materials



Magnetic sensors account for a significant portion of the sensing market.

Manufactures such as **Honeywell, Phillips, Optek, Cherry, and Infineon** primarily make commercial and automotive sensors while Fujitsu, IBM, Maxtor and Seagate control the information sector additionally, Asahi Chemicals has a significant position in fan speed sensing.

There are **two types of magnetic sensors**.

The first type of magnetic sensor commonly used is the Hall-effect sensor.

The second type of magnetic sensor is the magnetoresistor

On the basis of the materials and structures, MR effects can be classified as

1. *Ordinary magnetoresistance* (OMR) effect in nonmagnetic metals;
2. *Anisotropic magnetoresistance* (AMR) effect in ferromagnetic alloys;
3. *Giant magnetoresistance* (GMR) effect in multiple alternating ferromagnetic-alloy and metallic layer structures;
4. *Tunneling magnetoresistance* (TMR) effect in multiple alternating ferromagnetic-alloy and thin-insulating layer structures;
5. *Ballistic magnetoresistance* (BMR) effect in multiple alternating ferromagnetic-alloy-layer and nonferromagnetic-point structures; and
6. *Colossal magnetoresistance* (CMR) effect in $\text{La}_{1-x}\text{M}_x\text{MnO}_{3+8}$ ($\text{M} = \text{Ca}$ or Sr) perovskite structures.

These commercial devices, manufactured by the above companies, have diverse applications such as **proximity sensors, gear-tooth sensors, and read head sensors**

A search of the United States Patent Data Base shows over **three thousand patents using hall-elements**. Additionally there are over **four hundred patents using magnetoresistors**.

There is an art and a science to building commercial sensors.
Often it takes a diverse group of scientists and engineers to characterize and model these sensors prior to committing a design to production.

When a magnetic field is applied to the plate so that it is at right angles to the current flow, as shown in Figure 1-1, a small voltage appears across the plate, which can be measured by the probes. If you reverse the direction (polarity) of the magnetic field, the polarity of this induced voltage will also reverse. This phenomenon is called the Hall effect, named after Edwin Hall.

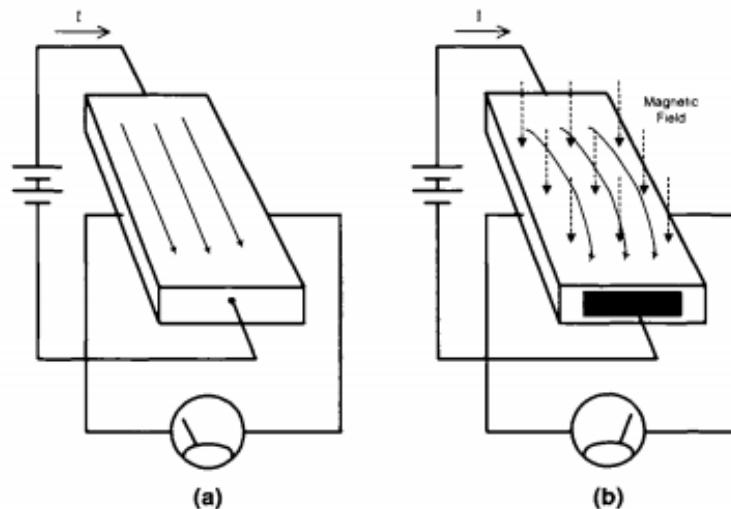


Figure 1-1: The Hall effect in a conductive sheet.

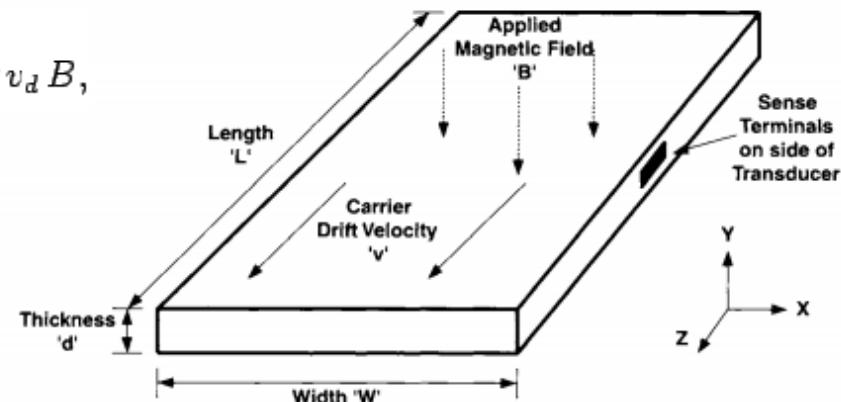
Although the magnetic field forces the charge carriers to one side of the Hall transducer, this process is self-limiting, because the excess concentration of charges to one side and consequent depletion on the other gives rise to an electric field across the transducer. This field causes the carriers to try to redistribute themselves more evenly.

It also gives rise to a voltage that can be measured across the plate. An equilibrium develops where the magnetic force pushing the charge carriers aside is balanced out by the electric force trying to push them back toward the middle .

$$\therefore E = V_H/w.$$

$$qE = \frac{qV_H}{w} = qv_d B,$$

E_H is the Hall electric field across the transducer



which means that the Hall field is solely a function of the velocity of the charge carriers and the strength of the magnetic field. For a transducer with a given width w between sense electrodes, the Hall electric field can be integrated over w , assuming it is uniform, giving us the Hall voltage.

$$V_H = v_d w B.$$

$$I = q n w d v_d,$$

$$V_H = \frac{IB}{qn d}.$$

The Hall voltage is therefore a linear function of:

- a) the charge carrier velocity in the body of the transducer,
- b) the applied magnetic field in the "sensitive" axis,
- c) the spatial separation of the sense contacts, at right angles to carrier motion.

Hall Effect in Metals

Even for the case of a magnetic field as strong as 10,000 gauss, the voltage resulting from the Hall effect is extremely small. For this reason, it is not usually practical to make Hall-effect transducers with most metals.

$$V_H = \frac{IB}{q_0 N d}$$

It can be seen that one means of improvement might be to find materials that do not have as many carriers per unit volume as metals do. A material with a lower carrier density will exhibit the Hall effect more strongly for a given current and depth. Fortunately, semiconductor materials such as silicon, germanium, and gallium-arsenide provide the low carrier densities needed to realize practical transducer elements. In the case of semiconductors, carrier density is usually referred to as carrier concentration.

Material	Carrier Concentration (cm^{-3})
Copper (est.)	8.4×10^{22}
Silicon	1.4×10^{10}
Germanium	2.1×10^{12}
Gallium-Arsenide	1.1×10^7

Consider the case of a transducer consisting of a piece of copper foil, similar to that shown back in Figure 1-1. Assume the current to be 1 ampere and the thickness to be 25 μm (0.001"). For a magnetic field of 1 tesla (10,000 gauss) the resulting Hall voltage will be:

$$V_H = \frac{1\text{A} \cdot 1\text{T}}{1.6 \times 10^{-19} \text{C} \cdot 8.42 \times 10^{28} \text{m}^{-3} \cdot 25 \times 10^{-6} \text{m}} = 3.0 \times 10^{-6} \text{V} \quad (\text{Equation 1-11})$$

Note the conversion of all quantities to SI (meter-kilogram-second) units for consistency in the calculation.

Even for the case of a magnetic field as strong as 10,000 gauss, the voltage resulting from the Hall effect is extremely small. For this reason, it is not usually practical to make Hall-effect transducers with most metals.

Consider a Hall transducer constructed from N-type silicon that has been doped to a level of $3 \times 10^{15} \text{ cm}^{-3}$. The thickness is 25 μm and the current is 1 mA. By substituting the relevant numbers into Equation 1-10, we can calculate the voltage output for a 1-tesla field:

$$V_H = \frac{0.001\text{A} \cdot 1\text{T}}{1.6 \times 10^{-19} \text{C} \cdot 3 \times 10^{21} \text{m}^{-3} \cdot 25 \times 10^{-6} \text{m}} = 0.083 \text{V} \quad (\text{Equation 1-12})$$

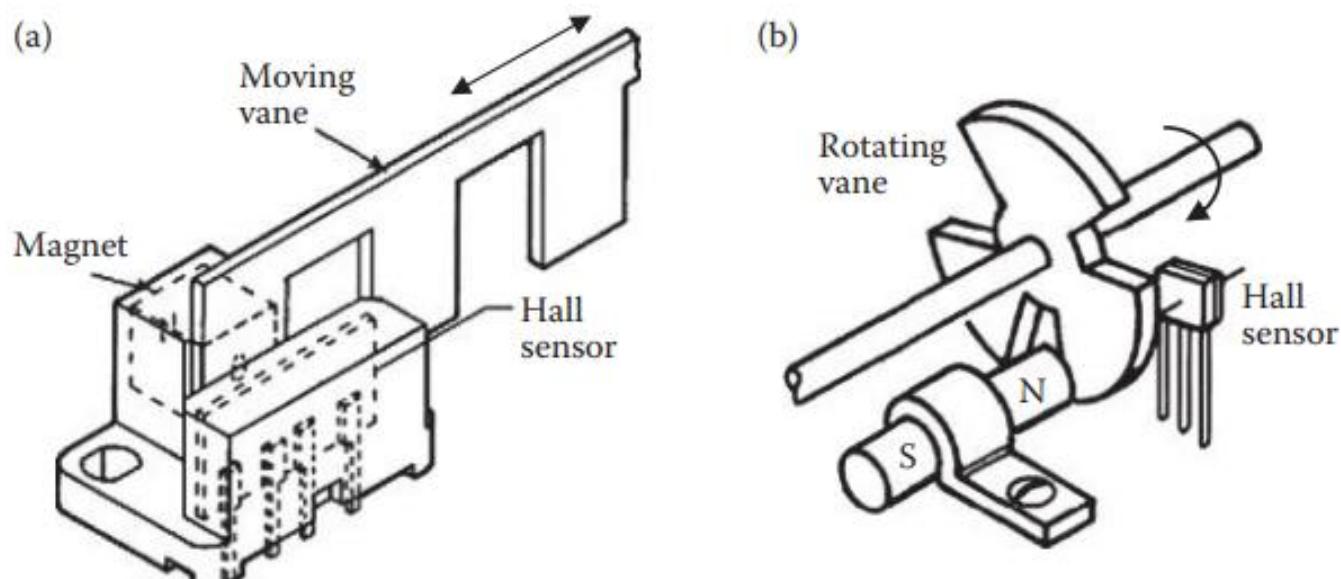
The resultant voltage in this case is 83 mV, which is more than 20,000 times the signal of the copper transducer described previously. Equally significant is that the necessary bias current is 1/1000 that used to bias the copper transducer. Millivolt-level output signals and milliamp-level bias currents make for practical sensors.

Applications of Hall Sensors

Hall sensors are widely used in automobile, security, brushless DC motors, damper control, various instrumentation, or any applications that involve electric current or magnetic field measurements.

High-quality Hall sensors can be constructed inexpensively with the standard IC processes used in the microelectronics industry, and integrate ancillary signal processing circuitry on the same silicon die.

It consists of a magnet and a Hall sensor—both are mounted rigidly in a fixture made of a nonmagnetic material. A ferrous vane can move in and out through the gap, which alters the magnetic flux lines in the gap. The Hall sensor then detects the presence, absence, or position of the vane,



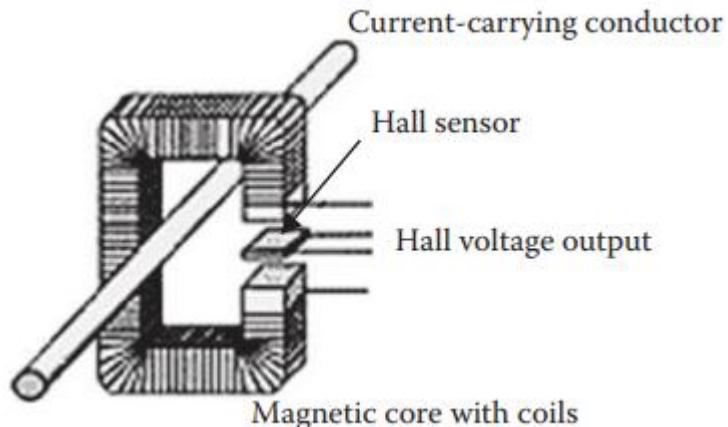
A Hall angular speed sensor

Allegro Microsystems,
LLC,, Massachusetts,
USA

**A Hall position sensor,
Honeywell Inc.**

Hall Current Sensor

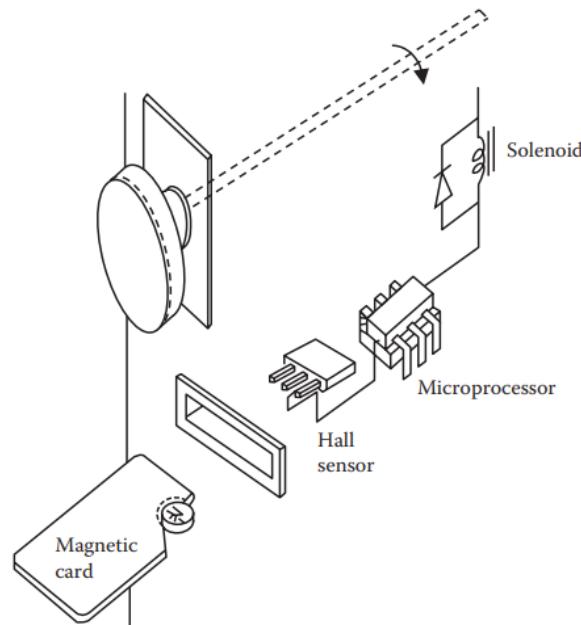
Hall sensors measure current via the intensity of the magnetic field generated by current flow. A larger current produces a stronger magnetic field. A Hall sensor's output voltage is therefore directly proportional to the current. Hall sensors can measure both AC and DC currents, and pulsed waveforms.



A C-core of soft magnetic material is placed around a conductor to concentrate the field. The Hall sensor, placed in the small air gap, delivers a voltage that is proportional to the current in the conductor. Hall current sensors are usually surface-mount types that can be mounted on a PCB to measure the current in the traces. Hall current sensors have advantages. of maintaining galvanic isolation between the sensor and the measuring circuits, and measuring the current without interrupting the circuit.

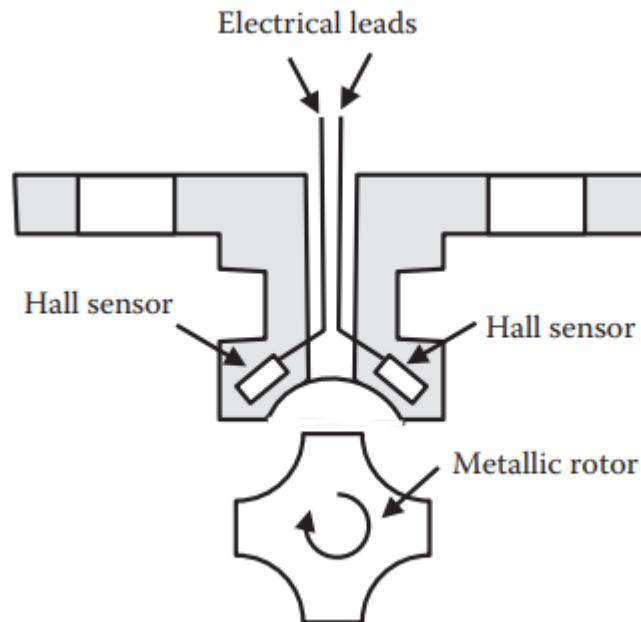
Door Security System

Figure shows a door interlock security system. It consists of a Hall sensor, a magnetic card, an actuator, and a circuit. When the keycard slides by, the Hall sensor detects the magnetic embedded in the card, and then sends an output signal to the microprocessor where the analog signal is converted into digital pulse to pull in the relay to open the door.



Flow Rate Meter

Hall sensors can also be used to measure flow rate . Two Hall sensors measure the rotational velocity of the metal rotor that relates to the flow rate of a liquid transmitted through a pipe. The rotor is modified to have four recessed areas that have magnetic properties on their surfaces. The two Hall sensors pickup the magnetic field changes due to the rotor's rotation, and the signals are converted into the flow rate.

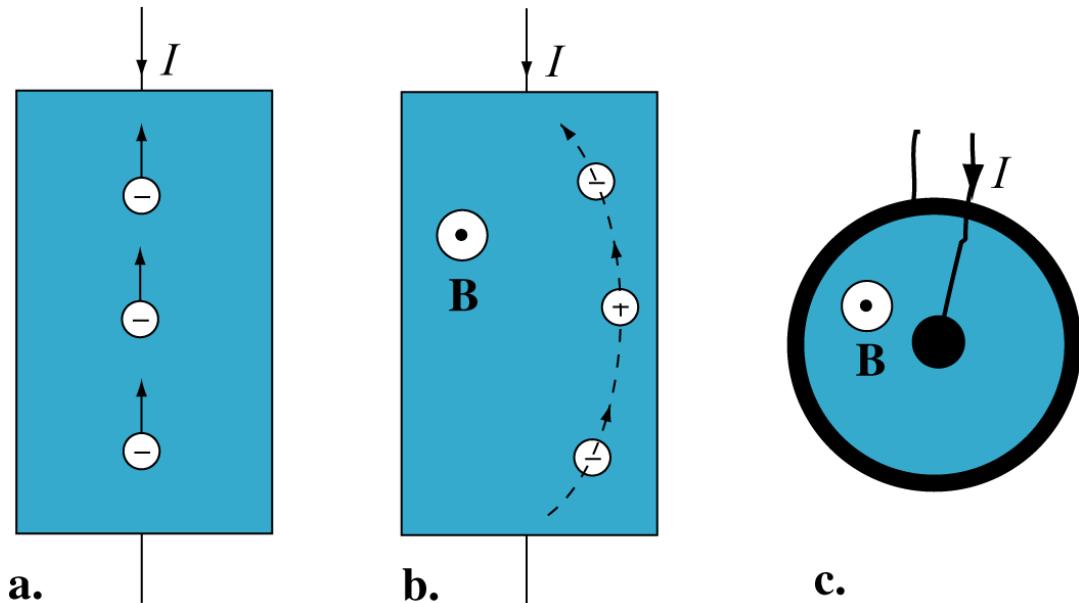


Flow-rate meter using two Hall sensors. (Based on US Patent)

Magnetoresistive sensors

- Two basic principles:
 - Similar to Hall elements
 - The same basic structure is used but
 - No Hall voltage electrodes.
 - The electrons are affected by the magnetic field as in the hall element
 - Because of the magnetic force on them, they will flow in an arc.
-

The magnetoresistive sensor



Magnetoresistive sensors

- The larger the magnetic field, the larger the arc radius
 - Forces electrons to take a longer path
 - The resistance to their flow increases (exactly the same as if the effective length of the plate were larger).
 - A relationship between magnetic field and current is established.
 - The resistance of the device becomes a measure of field.
-

Magnetoresistive sensors

- The relation between field and current is proportional to B for most configurations
- It is dependent on carrier mobility in the material used (usually a semiconductor).
- The exact relationship is rather complicated and depends on the geometry of the device.

Magnetoresistive sensors

- Magnetoresistors are also used where hall elements cannot be used.
- One important application is in magnetoresistive read heads where the magnetic field corresponding to recorded data is sensed.
- Much more sensitive than hall elements

The magnitude of the effect, MR (unit less), depends on the material of the conductor as

$$M_R = \frac{\rho_B - \rho_0}{\rho_0} \times 100\%$$

where ρ_B (in $\Omega \cdot m$) is the resistivity of the conductor under an applied magnetic field B, and ρ_0 (in $\Omega \cdot m$) is the resistivity of the conductor without an applied magnetic field B.

The value of the effect MR is in the order of few percentage points.

On the basis of the materials and structures, MR effects can be classified as

1. *Ordinary magnetoresistance* (OMR) effect in nonmagnetic metals;
2. *Anisotropic magnetoresistance* (AMR) effect in ferromagnetic alloys;
3. *Giant magnetoresistance* (GMR) effect in multiple alternating ferromagnetic-alloy and metallic layer structures;
4. *Tunneling magnetoresistance* (TMR) effect in multiple alternating ferromagnetic-alloy and thin-insulating layer structures;
5. *Ballistic magnetoresistance* (BMR) effect in multiple alternating ferromagnetic-alloy-layer and nonferromagnetic-point structures; and
6. *Colossal magnetoresistance* (CMR) effect in $\text{La}_{1-x}\text{M}_x\text{MnO}_{3+8}$ ($\text{M} = \text{Ca}$ or Sr) perovskite structures.

Ordinary Magnetoresistance (OMR) Effect

OMR effect is present in normal (nonmagnetic) metals. It arises from the effect of Lorentz force acting on an electron in a magnetic field, causing a circular or helical motion of the electron. The resistivity ρ (in $\Omega \cdot m$) to determine the MR value for the OMR effect is

$$\rho = \frac{B}{nqc\omega_c\tau_r}$$

where B is the applied magnetic field strength (in T); n is the electron density of the metal (in m^{-3}); q is the electron charge ($1.602 \times 10^{-19} C$); c is the speed of light ($3 \times 10^8 m \cdot s^{-1}$); ω_c is the cyclotron frequency (in $rad \cdot s^{-1}$); and τ_r is the electron relaxation time (the mean time between collisions, in s). ω_c can be found by



$$\omega_c = \frac{qB}{m^*c} \quad (5.10)$$

where m^* is the *effective mass* of an electron (in kg). Thus, the resistivity becomes

$$\rho = \frac{m^*}{nq^2\tau_r} \quad (5.11)$$

In Equation 5.9, $\omega_c\tau_r$ is the dominant factor for the OMR effect. In metals, such as Cu, Ag, and Au, the dependence of $\omega_c\tau_r$ value on magnetic field can be approximately described by $\omega_c\tau_r \approx 0.005B$. Thus, their OMR value is smaller than 1% under 1 T. To have a substantial M_R , $\omega_c\tau_r$ should be at least of order 1. Research has found that for some metals (e.g., Bi) the $\omega_c\tau_r$ value can be increased up to 100 times [7].

Anisotropic Magnetoresistance (AMR) Effect

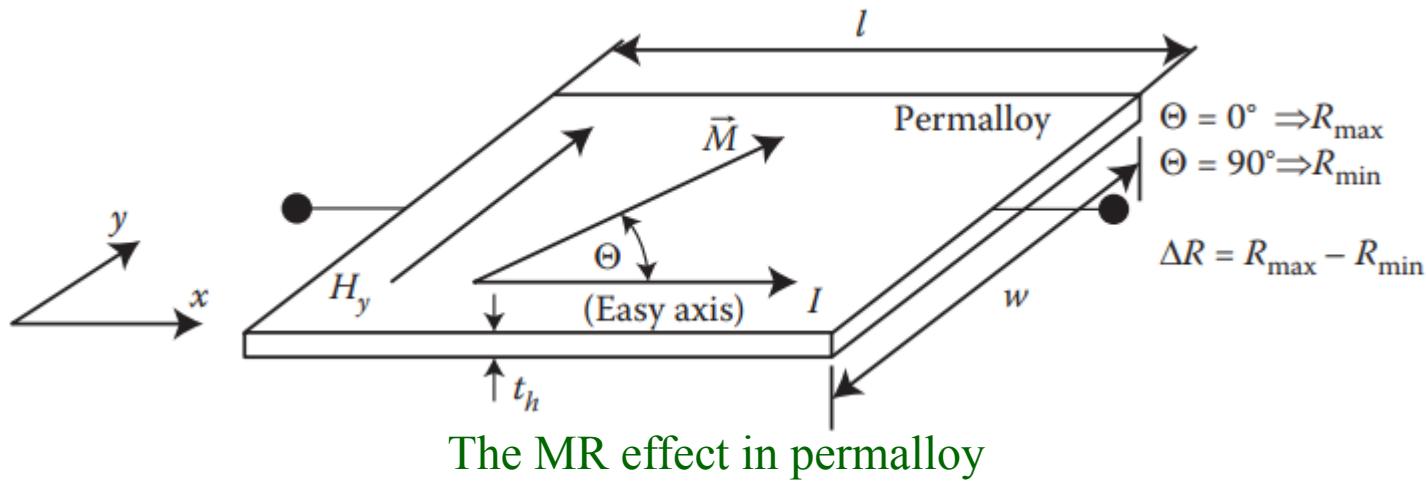
AMR, discovered in 1857 by William Thomson, is a typical effect in ferromagnetic (FM) materials. The term anisotropic is from the dependence of the resistivity on the orientation of the magnetic field relative to the current direction Θ .

Mathematically it is expressed by:

$$\rho(\Theta) = \rho_0 + \Delta\rho \cos^2 \Theta$$

where ρ_0 is the resistivity of the material without an applied magnetic field, and $\Delta\rho (= \rho(0^\circ) - \rho(90^\circ))$ is the resistivity difference between the parallel ($\Theta = 0^\circ$) and perpendicular ($\Theta = 90^\circ$) relationship between the current direction and the applied magnetic field direction. Thus, the resistance is at maximum when the current and the magnetic field directions are parallel and is at minimum when their directions are perpendicular. The AMR effect reflects the change of electron scattering in the atomic orbitals due to a magnetic field.

The dependence of the sensor resistance R on the angle Θ is described by



$$R(\Theta) = \underbrace{\rho_{\perp} \frac{l}{wt_h}}_{R_0} + \underbrace{(\rho_{||} - \rho_{\perp}) \frac{l}{wt_h}}_{\Delta R} \cos^2 \Theta = R_0 + \underbrace{(R_{\max} - R_{\min})}_{\Delta R} \cos^2 \Theta$$

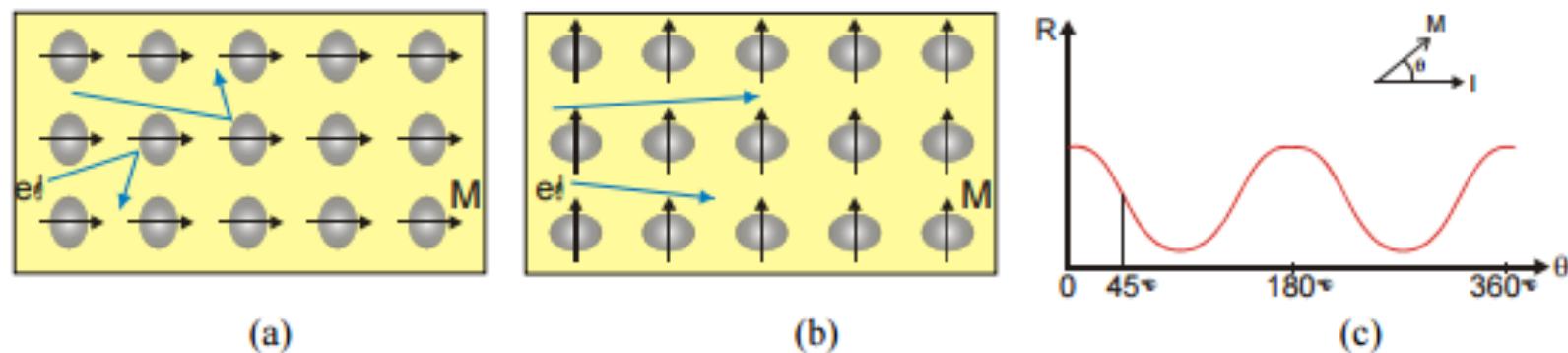
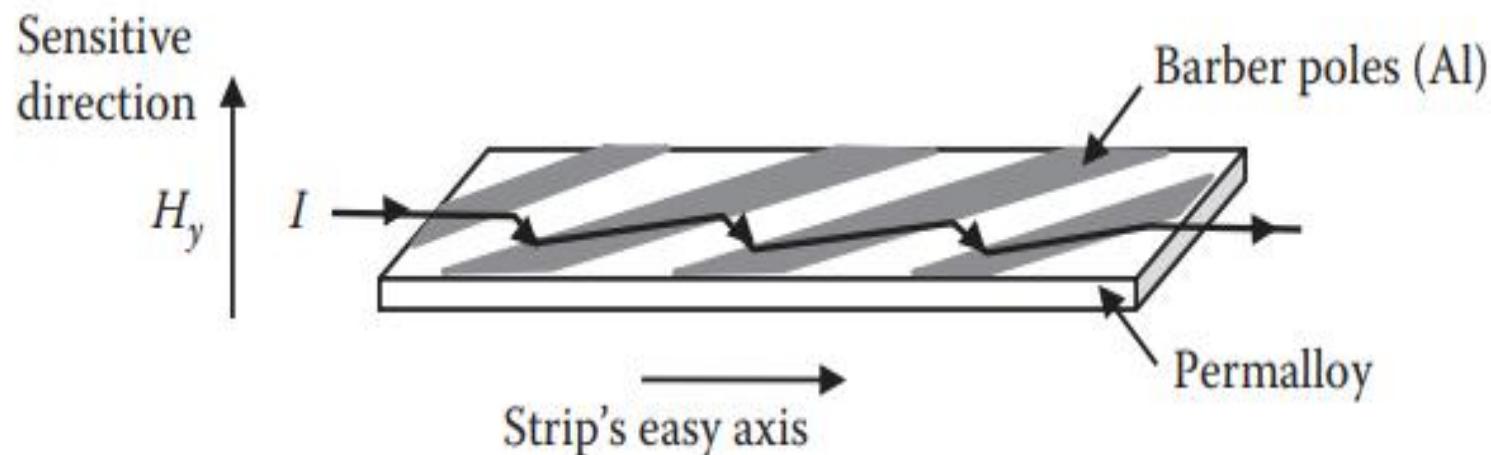


Fig. 1. Illustration of AMR effect showing distortion of electron orbitals and resulting difference in scattering when the magnetization is (a) parallel to the current or (b) perpendicular to the current direction. (c) Variation of resistance as a function of angle between the current and magnetization. The optimum operating point is at 45°.



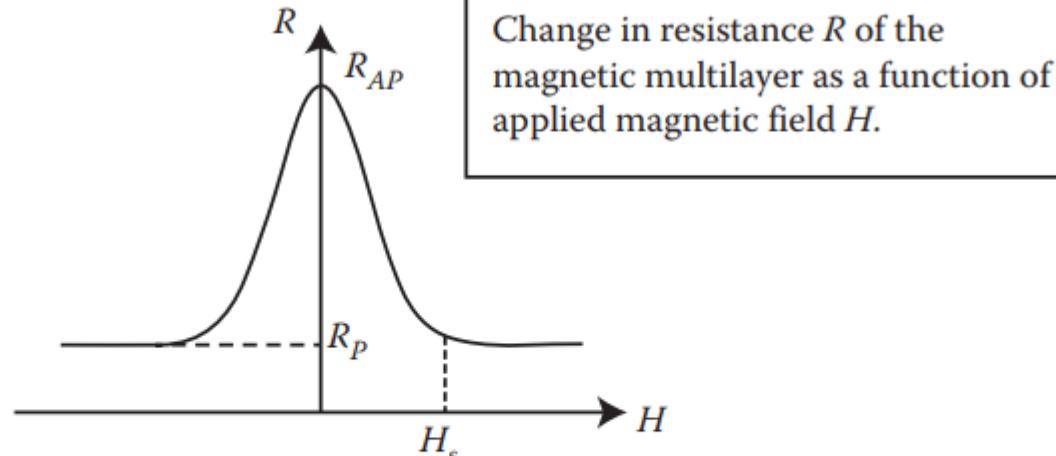
Giant Magnetoresistance (GMR) Effect

Baibich et al. and Binasch et al. are the first who reported “Giant” magnetoresistance measured on Fe/Cr/Fe thin multilayers.

They demonstrated that the electric current was strongly influenced by the relative orientation of the magnetizations of the magnetic layers. The cause of this giant change in resistance is attributed to the scattering of the electrons at the layers’ interfaces. Thus, any structure with metal–FM interfaces is a candidate to display the GMR effect. Since then, a huge effort has been carried out on improving structures to maximize the effect.

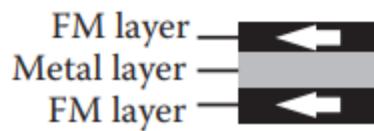
Today above 200% of MR value can be achieved at room temperature, which is much higher than the MR value of either OMR or AMR effect.

(a)



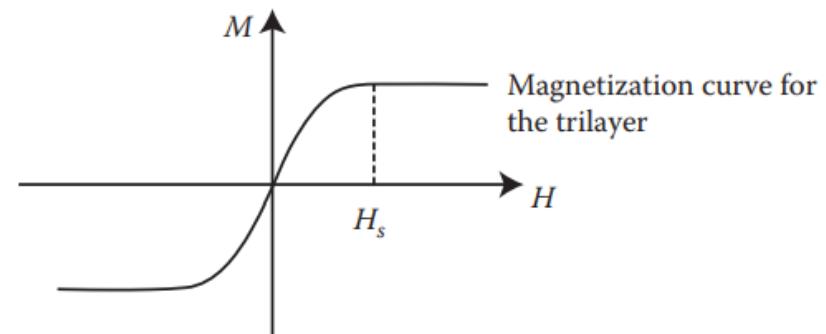
Change in resistance R of the magnetic multilayer as a function of applied magnetic field H .

(b)



Three magnetization configurations:
parallel, antiparallel, and parallel

$$M_R \text{ (for GMR)} = \frac{R_{AP} - R_P}{R_P} \times 100\%$$



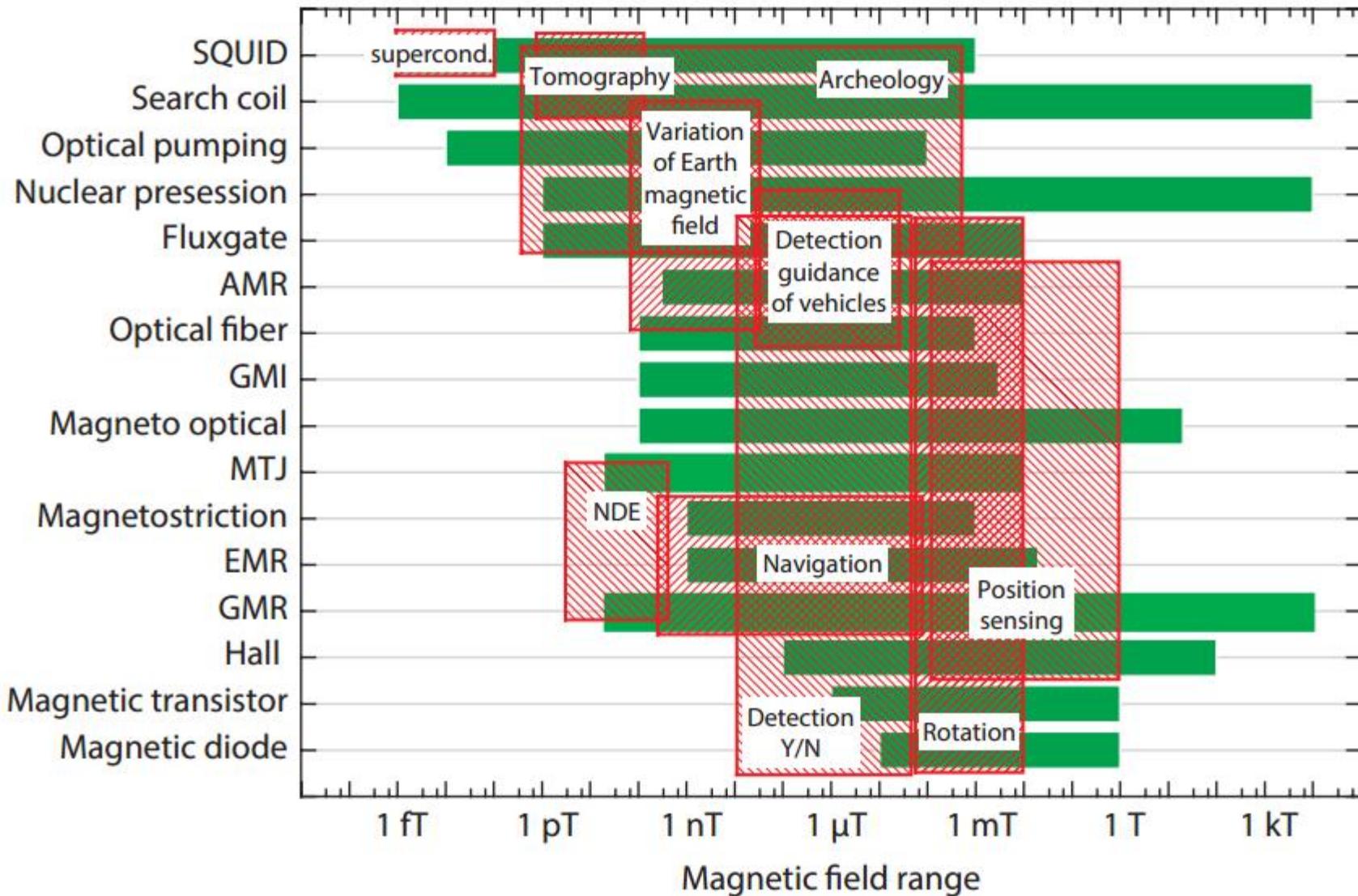
The change in electrical resistance can be made to over 200% of MR, depending on the number and thickness of layers, materials, and manufacturing processes.

Comparison of MR Effects

MR Effect	$\Delta R/R$ (%)	B (T)	Mechanism	Comments
OMR	<1	1 T (metal)	Lorentz force	There is no saturation at large magnetic field.
AMR	1–5	0.5 mT–1 T (depend on bulk or wire permalloy)	Electron spin–orbit interaction (leads to scattering of conducting electrons).	R is directly related to the orientation of magnetization M relative to current I .
GMR	10–200	1–10 T	Spin-dependent electron transport. In FM-metal-FM alternating layer structure.	Interface quality is crucial. They are broadly applied in sensors and read-heads of magnetic hard disks.

The discovery of giant magnetoresistance (GMR) by Albert Fert and Peter Grünberg in 1988 raised MR sensors' maximum MR value from 2–5% to 10% or more , a feat honored by the 2007 Nobel Prize in Physics.

Range of applications of different magnetic field sensors



Integrated Hall sensors



The first Hall effect magnetic sensors became commercially available in the mid-1950s, a few years after the discovery of high-mobility compound semiconductors. Since then, the development of Hall effect devices has taken advantage of using high-quality materials and sophisticated, highly productive fabrication methods available in the microelectronics industry.

Today, Hall effect magnetic sensors form the basis of a mature and important industrial activity. They are mostly used as key elements in contactless sensors for linear position, angular position, velocity, rotation, electrical current, end so on Most of currently produced Hall magnetic sensors are discrete elements; but the sales of discrete Hall elements stagnates, whereas the sales of integrated Hall sensors grows at more than 10 % per year. Integrated Hall magnetic sensors are ‘smart’: they incorporate electronic circuits for biasing, offset reduction, compensation of temperature effects, signal amplification, and more. The integration helps improve sensor sys-tem performance at moderate costs, which allows a continuous penetration of Hall magnetic sensors into new application areasA great majority of integrated circuits (IC) have been made of silicon. Silicon IC technology is very mature, easily accessible and low cost. This makes this technology very attractive for the realization of Hall plates, in spite of moderate mobility of electrons in silicon. We shall now describe a few conventional structures of Hall devices fabricated with the aid of silicon integrated circuit technology. The conventional Hall plates are parallel with the chip surface; so considering a chip as an ocean in which floats a Hall plate, we say that such a Hall plate is ‘horizontal’. ~~Horizontal Hall plates are sensitive to the magnetic field perpendicular to the chip plane~~

MOS Hall effect devices

The idea of implementing a Hall device in the form of an MOS device looks especially attractive for sensor applications: such a device is readily integrable with MOS bias and signal-processing circuits.

Unfortunately, an MOS Hall effect device also has a few serious drawbacks: the carrier mobility in the channel amounts to only half its value in the bulk.

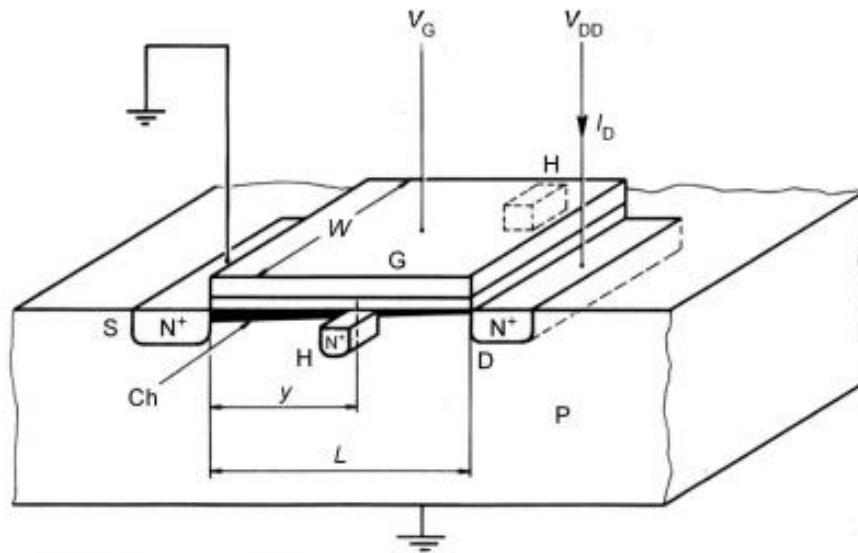


Figure 7.1. MOS Hall plate: the channel (Ch) of a MOSFET is exploited as an extremely thin conductive plate. Two Hall voltage probes H are added to the usual MOSFET structure. (Adapted from [1]. © 1986 IEEE.)

thin Hall plate, the source (S) and drain (D) as the biasing contacts, and the two additional heavily doped regions (H) are provided for sensing the Hall voltage (compare with figure 5.7). The two sense regions are fabricated simultaneously with the source and drain regions. The channel length is L, its width is W, and the sense probes are positioned at a distance y from the source.

Let us now consider the operation of an MOS Hall device in the Hall voltage mode. An MOS Hall plate working in the linear region is equivalent to a conventional Hall plate. The only difference is that, in a conventional Hall plate, the charge carriers in the plate are provided by the material itself; and in an MOS Hall plate, the charge carriers are due to the surface field effect.

$$V_H \simeq \mu_{ch} \frac{W}{L} G_H V_D B_{\perp}.$$

At a higher drain voltage, the channel charge in the MOS Hall plate decreases with an increase in the distance from the source 'y'. The lower the channel charge, the higher the Hall electric field. Therefore, if we could neglect the short-circuiting effect by the supply electrodes, the Hall voltage would steadily increase by moving the sense contacts towards the drain.

At very low drain voltages, when

$$V_D \ll V_G - V_T \quad (7.1)$$

the area density of carriers in the channel is approximately constant over the channel. This charge density is given by

$$Q_{ch} \simeq C_{ox}(V_G - V_T) \quad (7.2)$$

where C_{ox} denotes the gate oxide capacitance per unit area. Then the drain current is given by

$$I \simeq \frac{W}{L} \mu_{ch} C_{ox} (V_G - V_T) V_D \quad (7.3)$$

where μ_{ch} denotes the drift mobility of carriers in the channel. This is the linear region of operation of a MOSFET.

At higher drain voltages, the carrier charge density in the channel continuously decreases with increasing distance from the source. The drain current is generally given by

$$I \simeq \frac{W}{L} \mu_{ch} C_{ox} [(V_G - V_T) V_D - \frac{1}{2} V_D^2] \quad \text{for } V_D \leq V_G - V_T. \quad (7.4)$$

When the drain voltage reaches the value

$$V_{Dsat} \simeq V_G - V_T \quad (7.5)$$

the carrier density at the drain end of the channel reduces virtually to zero. This is the pinch-off point. Beyond the pinch-off point, the drain current stays essentially constant. The saturated drain current is given by I_{Dsat} (V_{Dsat}) according to (7.4) and (7.5).

Magnetotransistors (BJT)

A magnetotransistor is a bipolar junction transistor (BJT) whose structure and operating conditions are optimized with respect to the magnetic sensitivity of its collector current.

The magnetic sensitivity of a magnetotransistor is usually defined

$$S_I = \left| \frac{1}{I_c} \frac{\Delta I_c}{B} \right|. \quad (7.15)$$

Here, I_c denotes the collector current, and ΔI_c is the change in the collector current due to a magnetic induction B :

$$\Delta I_c = I_c(B) - I_c(0). \quad (7.16)$$

The sensitivities of magnetotransistors reported hitherto cover a surprisingly wide range: from 10^{-2} to over 10 T^{-1} . This large spread of sensitivities indicates that the operation of various magnetotransistors is based on different effects. Indeed, the Hall effect may interfere with the action of a bipolar transistor in many ways and give rise to different end effects. The following three major end effects may be distinguished:

- The current deflection effect. This is essentially the same effect that we studied in large-contact Hall plates.
- The injection modulation. The Hall voltage generated in the base region of a magnetotransistor modulates the emitter-base voltage, and thus also the carrier injection.
- The magnetodiode effect. The emitter-base diode of a transistor may function as a magnetodiode. This leads to a magnetic sensitivity of collector current. In principle, these three effects coexist and cooperate in any magnetotransistor.

The magnetodiode effect comes about as a result of the cooperation of the Hall effect and a few other effects pertinent to a p–n junction diode. The basic effects are: the conductivity modulation due to a high injection level ; the current deflection ; and the magnetoconcentration effect

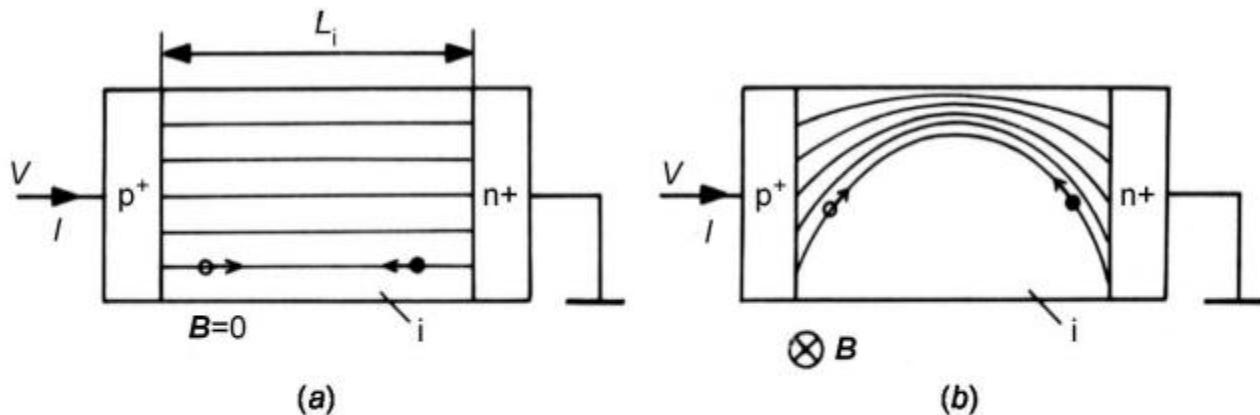


Figure 7.10. Current density lines in a magnetodiode with volume recombination. (a) No magnetic field; (b) a magnetic field perpendicular to the drawing plane is present. Then both electrons (●) and holes (○) are deflected towards the same boundary of the slab, and the current lines get longer.

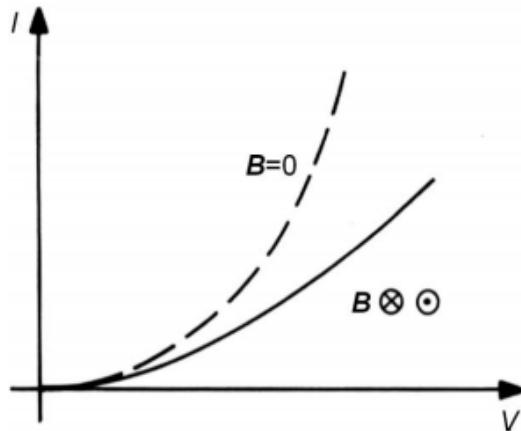


Figure 7.11. Current–voltage characteristics of a magnetodiode with volume recombination. At a constant bias voltage, a magnetic field causes a decrease in diode current, irrespective of the sign of the magnetic field.

The current in the magnetodiode decreases in the presence of a magnetic field. The lower the current, the lower the injection level and the higher the resistance of the i-region. Then the larger portion of the diode voltage drops across the i-region, the voltage across the injecting junctions decreases, and the diode current decreases. Thus at a high injection level, a magnetic field triggers a cumulative process of current reduction, which greatly boosts the magnetic sensitivity of the magnetodiode.

Magnetodiodes have been made of various semiconductors, including germanium, silicon and GaAs. Early devices were discrete

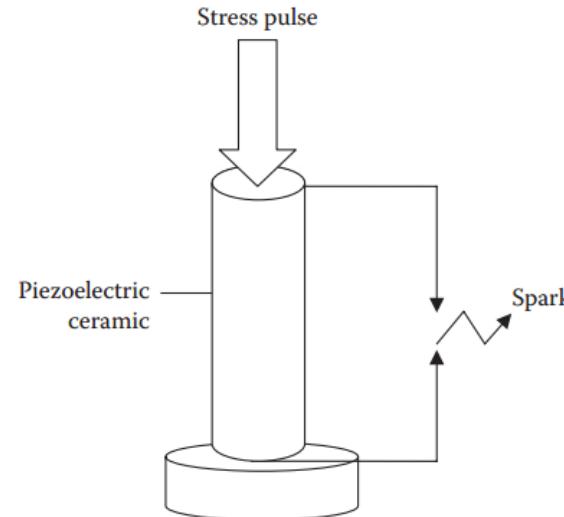
REQUIREMENTS OF PIEZOELECTRIC MATERIALS FOR SENSORS

- high piezoelectric sensitivity,**
- high mechanical strength,**
- high rigidity (high modulus of elasticity),**
- high electric insulation resistance (also at high temperatures),**
- minimal hygroscopicity,**
- linear relationship between mechanical stress and electric polarization,**
- absence of hysteresis,**
- high stability of all properties,**
- low temperature dependence of all properties within a wide temperature range,**
- low anisotropy of mechanical properties, such as thermal expansion coefficients**
- and elastic constants,**
- good machinability,**
- low production cost.**

Gas Lighter

A gas lighter is a common piezoelectric household device which makes use of the direct piezoelectric effect to generate electric sparks.

In a gas lighter, a high voltage pulse is required to be generated across a narrow electrode gap. A piezoelectric gas lighter consists of a PZT cylinder which is subjected to a stress pulse using a spring mechanism. When a button is pressed, a stress pulse is applied on the piezoelectric cylinder. The stress causes a high voltage to be generated which is made to appear across a small air gap between two closely spaced electrodes. The arrangement is shown in Figure 4.1. The voltage developed is high enough to cause breakdown of the air gap between the two electrodes, resulting in a spark.



Piezoresistivity

Piezoresistivity derives its name from the Greek word *piezin*, meaning “to press.” It is an effect exhibited by various materials that exhibit a change in resistivity due to an applied pressure. The effect was first discovered by Lord Kelvin in 1856, who noted that the resistance of copper and iron wires increased when in tension. He also observed that iron wires showed a larger change in resistance than those made of copper. The first application of the piezoresistive effect did not appear until the 1930s, some 75 years after Lord Kelvin’s discovery. Rather than using metal wires, these so-called strain gauges are generally made from a thin metal foil mounted on a backing film, which can be glued onto a surface. A typical metal foil strain gauge is depicted in Figure 5.1.

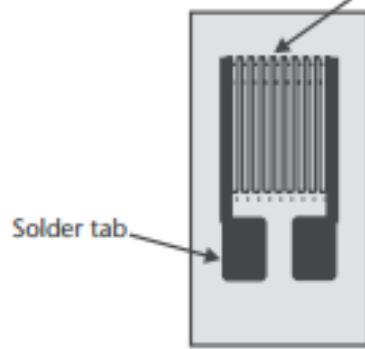


Illustration of a metal foil strain gauge.

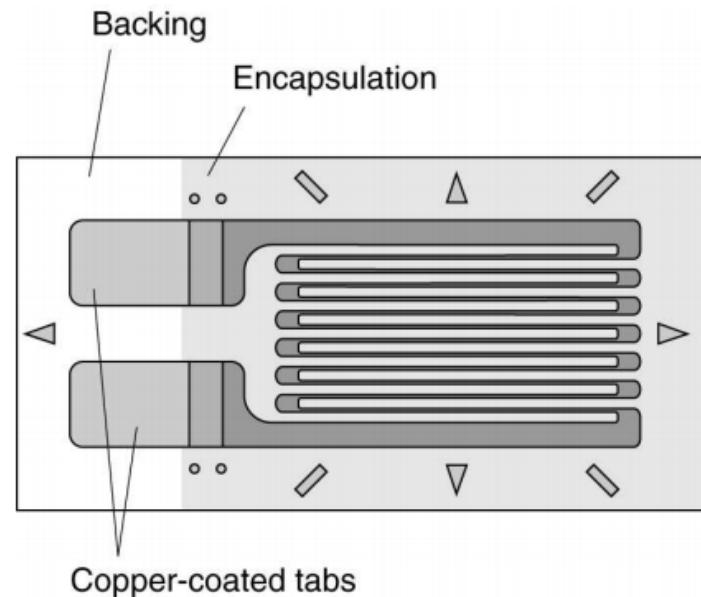
The strain gage is the most frequently used resistive sensor. A typical strain gage is shown in figure The gage consists of a very fine, etched wire of length L that winds back and forth over a flat, insulating sensing area. For the strain gage shown, there are 12 wire segments, yielding a total wire length of 12 cm.

A local gage factor, G_l , can be defined as the ratio of the relative resistance change to the relative length change,

$$G_l = \frac{dR/R}{dL/L}.$$

An engineering gage factor, G_e , can be defined as

$$G_e = \frac{\Delta R/R}{\Delta L/L}.$$

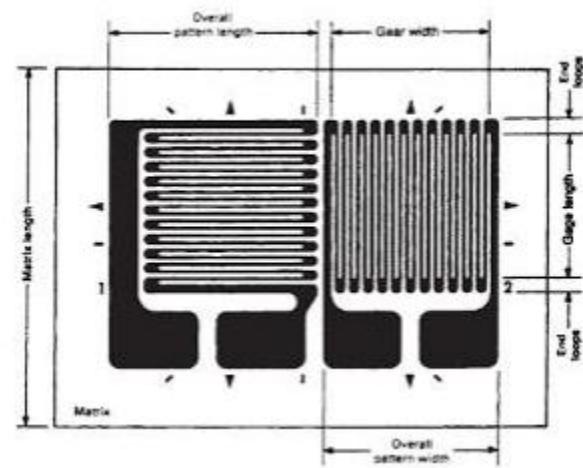


Different forms of metal foil strain gauges

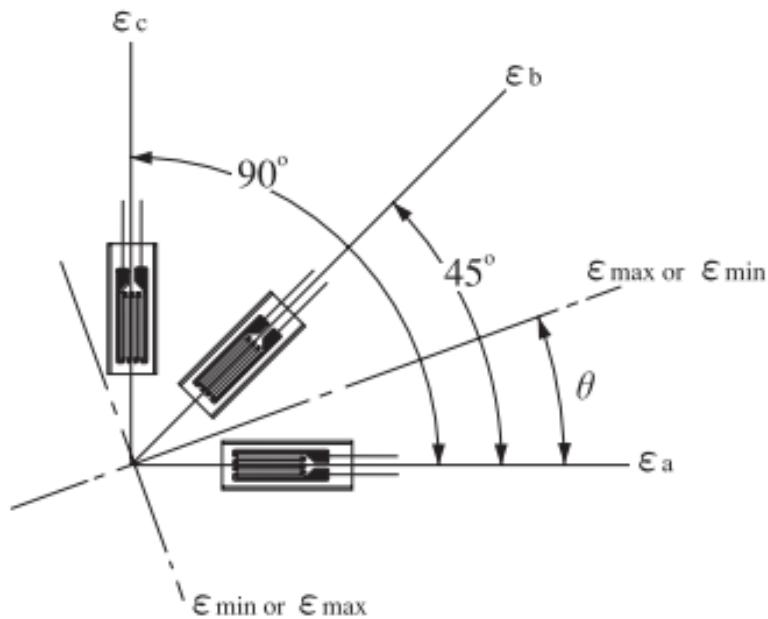
A two-element 90° rosette-type foil gage.

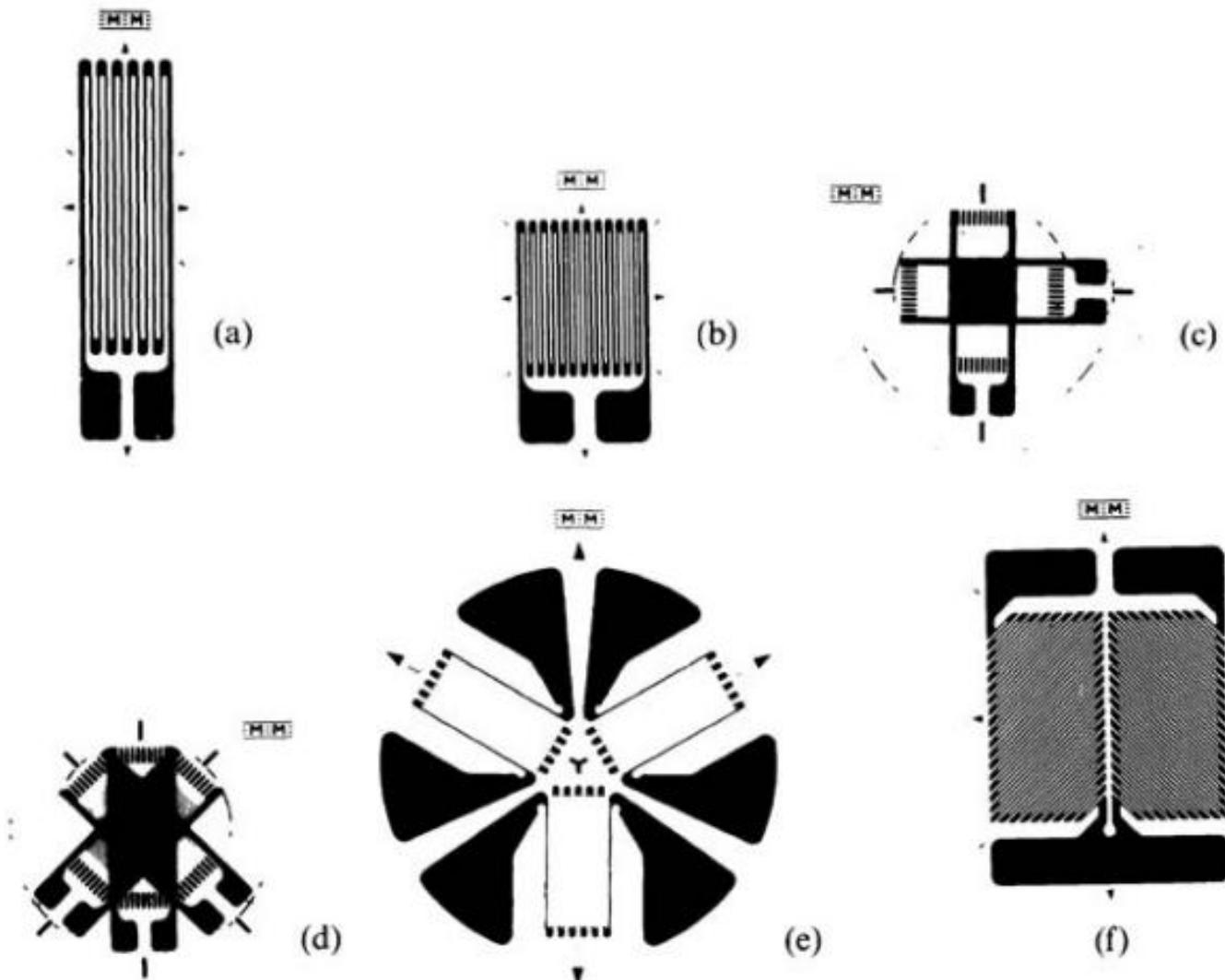
When a force or pressure is applied to the sensing element of metal foil strain gauge the physical dimensions of it will change.

Since, the strain gauge element is pasted on its surface, the dimensions of the strain gauge changes due to which the resistance of the gauge changes.



If we want to measure the strain in two or more directions at the same point, strain gage rosette, which is manufactured by stacking multiple strain gages in different directions, is used. Fig shows a three-element strain gage rosette stacked at 45^0 .





Foil strain gages, (a, b) Single-element gages, (c) Stacked two-element rectangular rosette, (d) Stacked three-element rectangular rosette, (e) Three-element delta rosette, (f) Two-element rectangular torque gage

The sensitivity of a strain gauge is generally termed the gauge factor. This is a dimensionless quantity and is given by

$$GF = \frac{\text{relative change in resistance}}{\text{applied strain}} = \frac{\Delta R / R}{\Delta L / L} = \frac{\Delta R / R}{\epsilon} \quad (5.1)$$

where R is the initial resistance of the strain gauge and ΔR is the change in resistance. The term $\Delta L/L$ is, by definition, the applied strain and is denoted as ϵ (dimensionless). For all elastic materials, there is a relationship between the stress $\sigma(\text{N/m}^2)$ and the strain ϵ ; that is, they obey Hooke's law and thus deform linearly with applied force. The constant of proportionality is the elastic modulus or Young's modulus of the material and is given by

$$\text{Young's modulus, } E = \frac{\text{Stress}}{\text{Strain}} = \frac{\sigma}{\epsilon} (\text{N/m}^2) \quad (5.2)$$

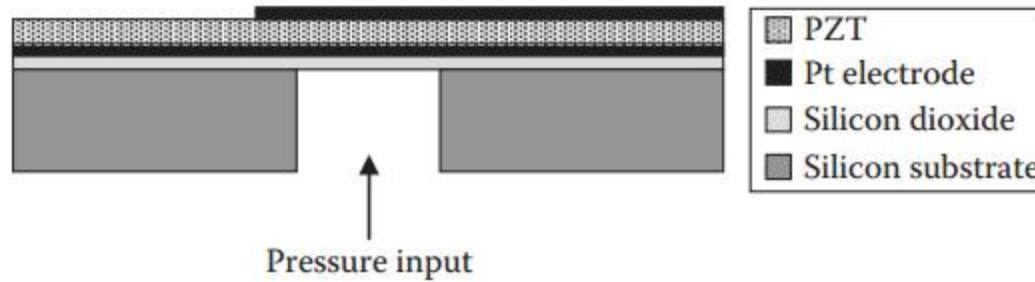
The Young's modulus of silicon is 190 GPa (1 Pa = 1 N/m²), which is close to that of typical stainless steel (around 200 GPa). For a given material, the higher the value of Young's modulus, the less it deforms for a given applied stress (i.e., it is stiffer).

Thick-film resistors, often used in hybrid circuits, have also been shown to be piezoresistive. Their gauge factor is around 10, and therefore, they offer a sensitivity between that of a semiconductor and foil strain gauge. The TCR is around 100 parts per million (ppm) per degree Celsius and matching between adjacent resistors is often less than 10 ppm/°C, making them well suited for use as active elements in Wheatstone bridge circuits, which reduce the overall temperature sensitivity.

An associated effect that has been observed in semiconductors is the so-called piezojunction effect, whereby a shift in the I-V characteristic of a *p-n* junction is observed as a result of an applied stress. Although this is an interesting physical effect, it has found little use in commercial micromachined devices.

Basic design of piezoresistive silicon pressure sensors

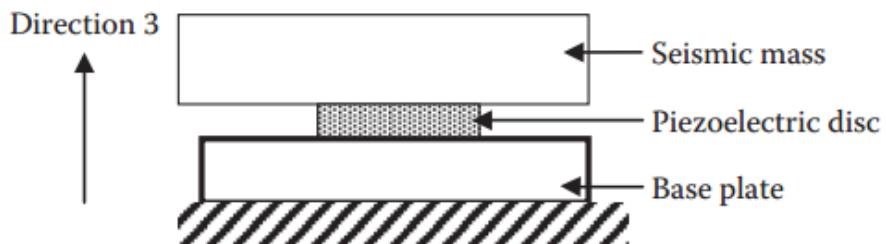
A quartz crystal piezoelectric pressure sensor. Courtesy: Intertechnology Inc.—Testing and Measurement Solutions, Ontario, Canada.



MEMS.PZT.pressure.sensor.

Accelerometer

Accelerometers are used for measurement of vibrations in many applications which include impact acceleration levels experienced by vehicles during crash, shock experienced by space vehicles and cargo during stage separation, testing of shock resistance of packaged products, vibrations in mining activities, seismic vibrations during earthquakes, etc.



When the system is subjected to acceleration, the seismic mass exerts a force F on the piezoelectric disc given by

$$F = Ma \quad (4.3)$$

where a is the acceleration experienced by the disc. The mechanical stress in direction 3 on the piezoelectric disc is given by

$$X_3 = \frac{F}{A} = \frac{Ma}{A} \quad (4.4)$$

where A is the area of the disc. The mechanical stress causes an electric field E to be generated across the thickness of the disc given by (Equation 2.4)

$$E_3 = g_{33} X_3 \quad (4.5)$$

The open circuit voltage across the piezoelectric disc of thickness t will be

$$V = E_3 t = g_{33} X_3 t \quad (4.6)$$

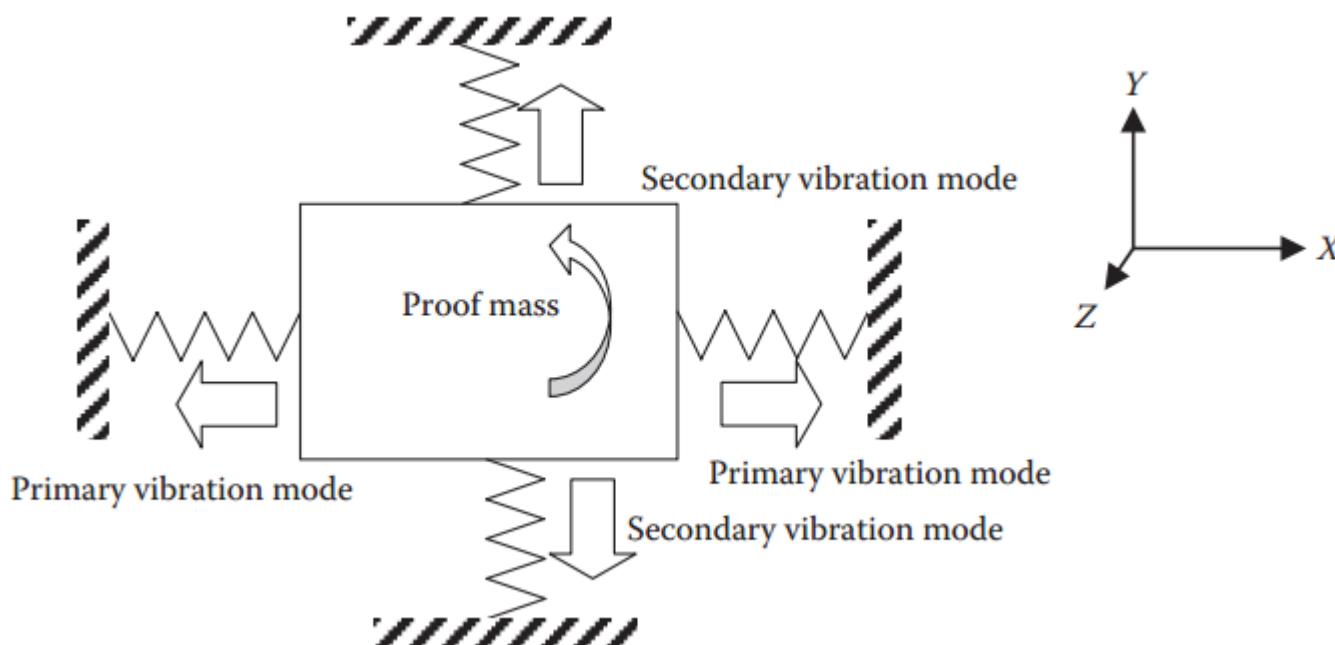
Substituting for X_3 from Equation 4.4,

$$V = g_{33} \frac{t}{A} Ma \quad (4.7)$$

The output voltage is proportional to the acceleration. The proportionality constant is determined by the seismic mass M , the piezoelectric coefficient g_{33} , and the dimensions of the piezoelectric disc.

Piezoelectric Gyroscope

Gyroscopes are devices used for measuring the angular rate of rotating objects. They have several engineering applications: in automobiles they are used for stability control, navigation assistance, and rollover detection; in marine engineering, they are used for stabilization and navigation of ships and in military applications for missile stabilization and guidance. There are



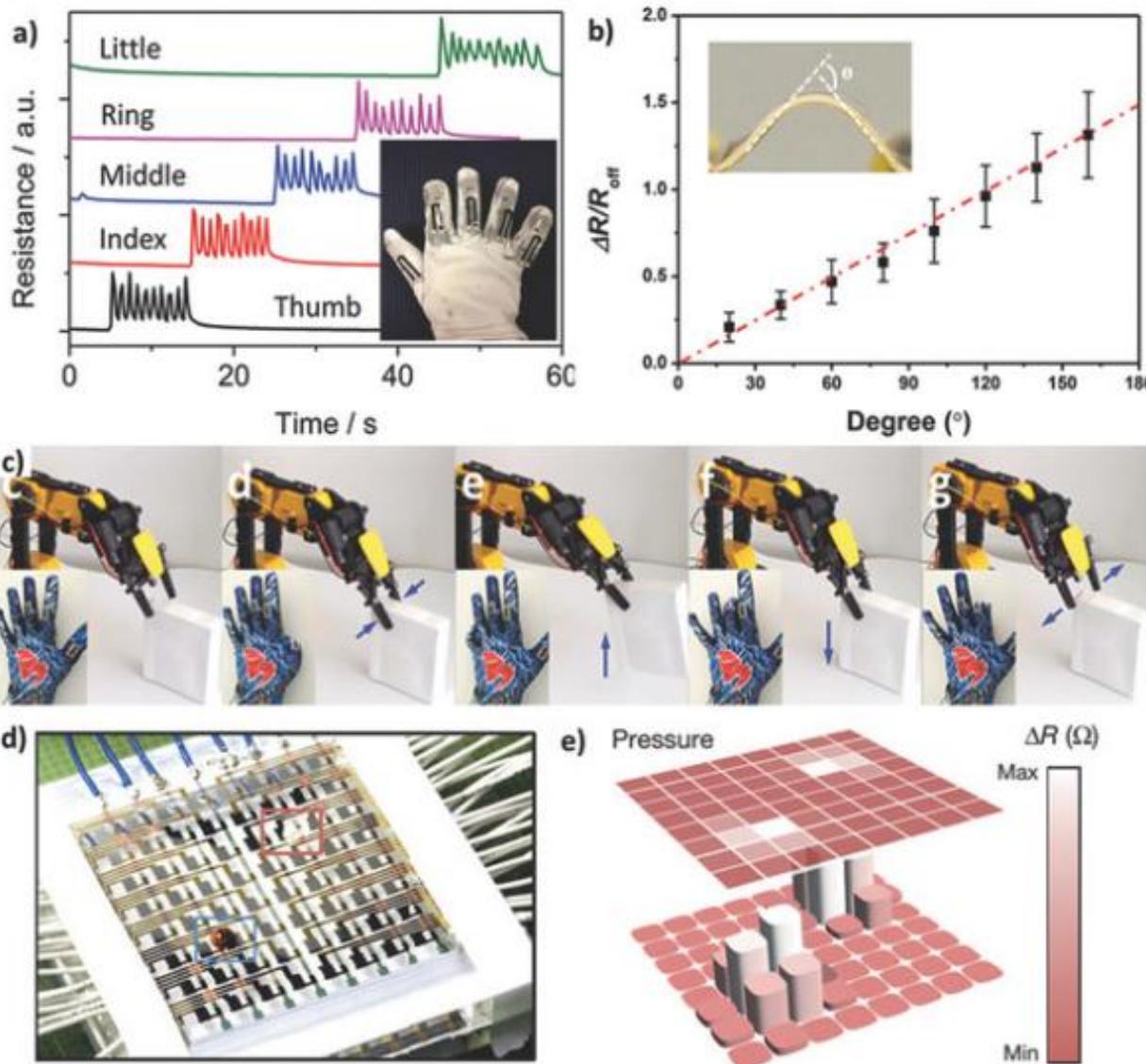
The primary mode of vibration is along the X-direction. The rotation of the object is anticlockwise with the axis of rotation in the Z-direction. The secondary mode of vibration acquired due to the Coriolis effect is in the Y-direction as shown. In a piezoelectric gyroscope, the primary mode is induced using a piezoelectric actuator, and the secondary mode is detected using a piezoelectric sensor.

The principle used in all types of gyroscopes is the Coriolis effect, which arises in a rotating frame of reference. The Coriolis effect may be stated as follows: "When a moving object is subjected to rotation about an axis perpendicular to the direction of motion, the object experiences an acceleration in a direction mutually perpendicular to the original direction of motion and the axis of rotation." The equation that describes the Coriolis effect is

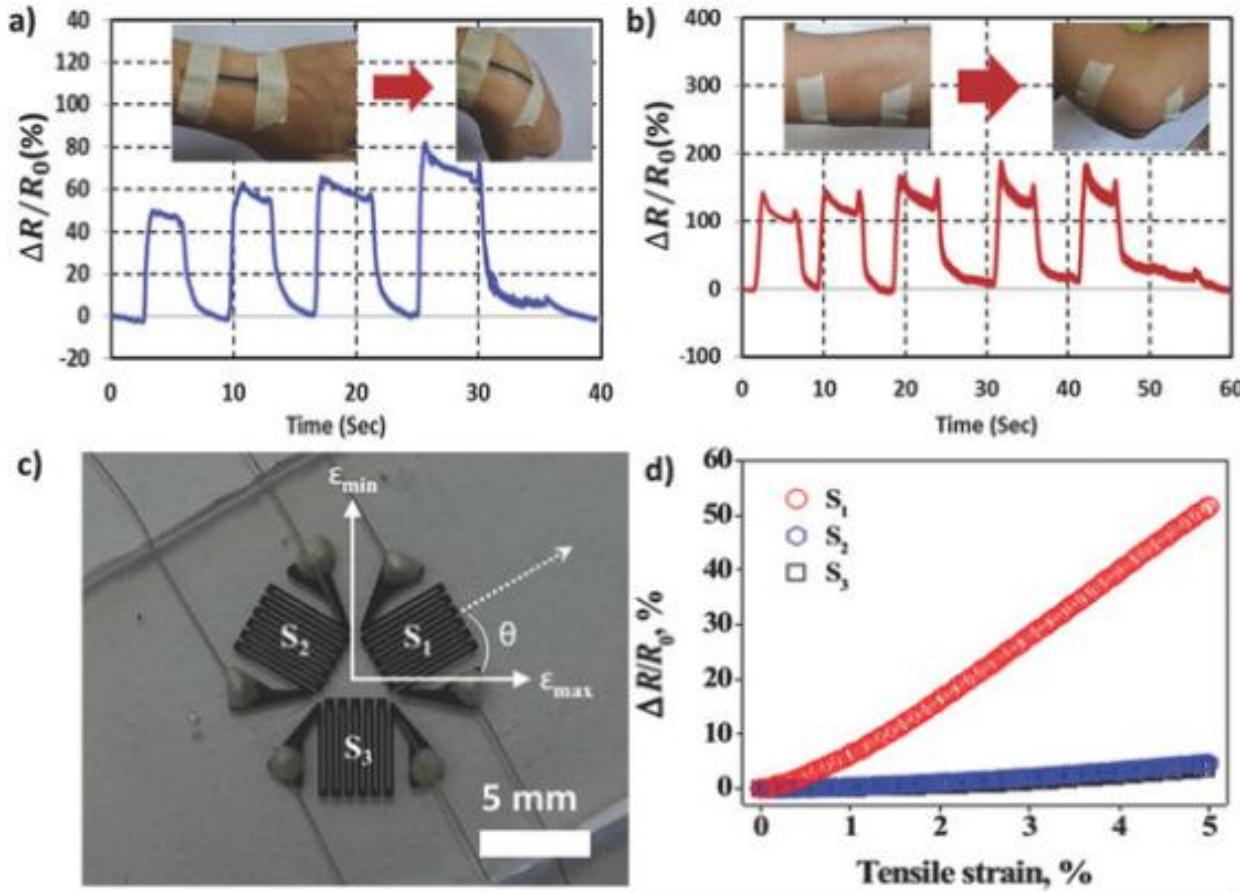
$$\vec{a} = -2\vec{\Omega} \times \vec{v} \quad (4.8)$$

where \vec{v} is the initial velocity of the object, $\vec{\Omega}$ is the angular velocity of rotation, and \vec{a} is the acceleration acquired.

Robotic applications of flexible strain sensors

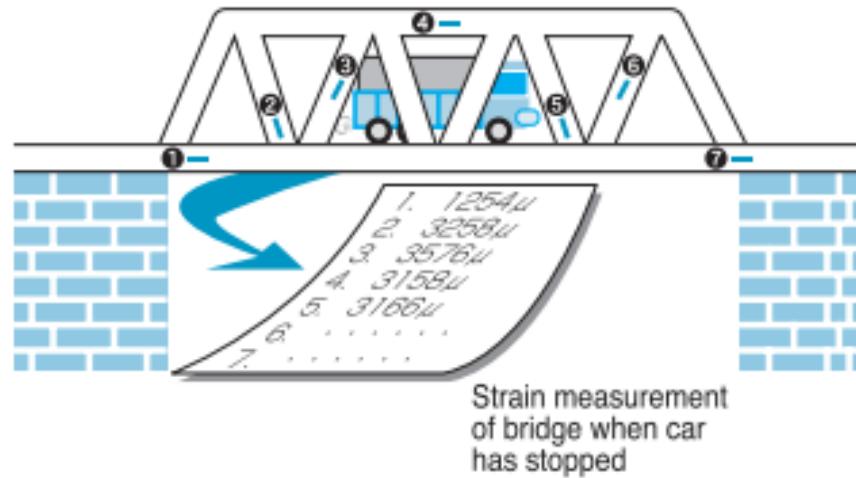


Wearable and skin-mountable strain sensors for sport performance monitoring



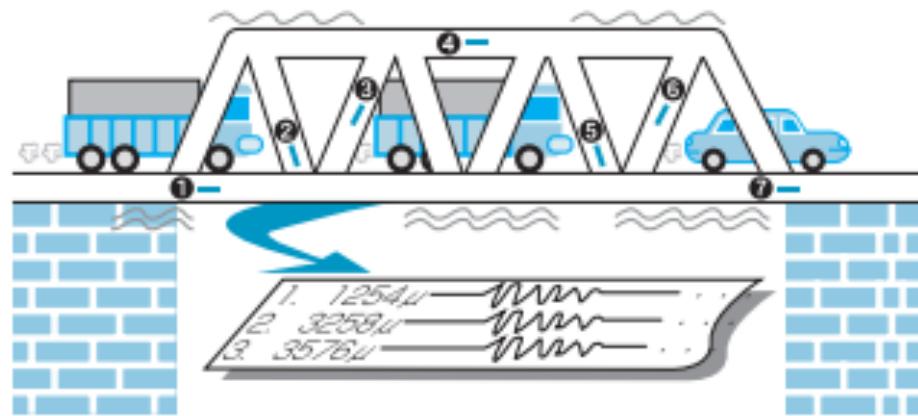
Static Strain Measurement

Measurement of strain initiated on a bridge by a stopped vehicle



Dynamic Strain Measurement

Measurement of strain initiated on a bridge by running vehicles

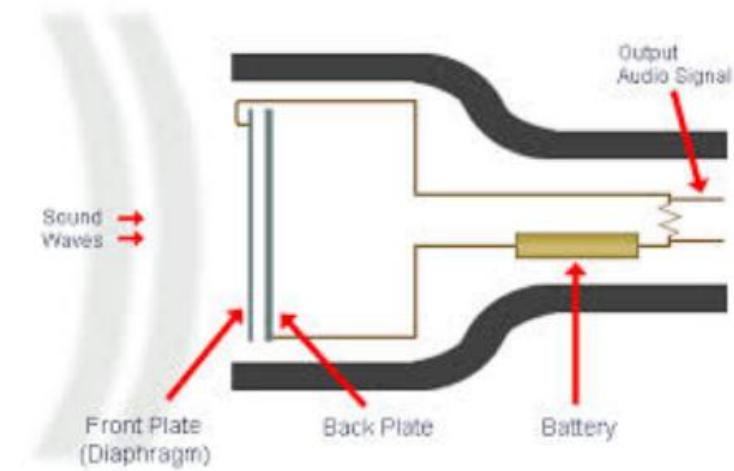


Condenser Microphones

If a parallel-plate capacitor is given an electric charge, q , voltage across its plates is governed by the equation below. On the other hand, according to equation the capacitance depends on distance d between the plates. Thus solving these two equations for voltage we arrive at

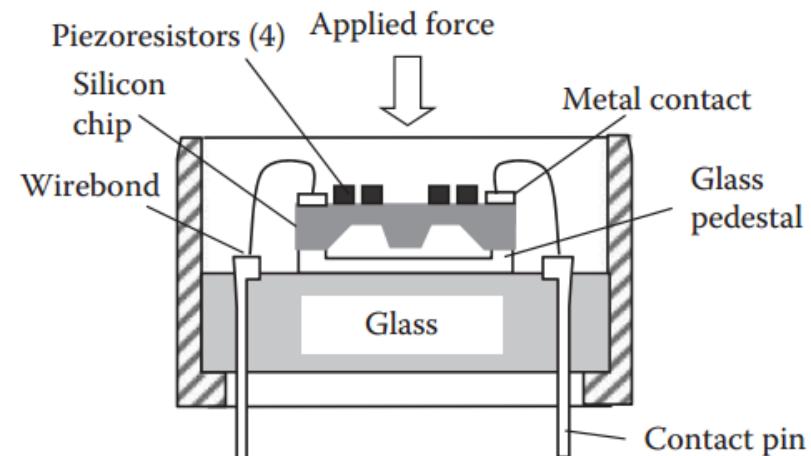
$$V = q \frac{d}{A\epsilon_0},$$

The above equation is the basis for operation of the condenser microphones, which is the other way to say “capacitive” microphones. Thus, a capacitive microphone linearly converts a distance between the plates into electrical voltage, which can be further amplified. The device essentially requires a source of an electric charge q whose magnitude directly determines the microphone sensitivity. The charge can be provided either from an external power supply having a voltage in the range from 20 to 200 V, or from an internal source capable of producing such a charge. This is accomplished by a built-in electret layer, which is a polarized dielectric crystal.



Piezoresistive Pressure Sensor

Piezoresistive pressure sensors are critical devices in a variety of control and automobile applications. Figure shows an internal combustion engine sensor designed by Kulite Semiconductor Products Inc., Leonia, New Jersey. It uses four piezoresistors to measure the stress in a silicon diaphragm caused by the force or pressure of the media. These four piezoresistors are connected electrically to form a Wheatstone bridge. At the corners of the diaphragm, five 0.024-mm-diameter gold bond wires (ultrasonically ball bonded to the sensor) allow electrical connections to the bridge. The sensor has a resonant frequency above 150 kHz, which also meets the stringent combustion requirements . This sensor can withstand the engine's harsh environment—extreme operating temperature of ~~500°C~~ and high vibration.



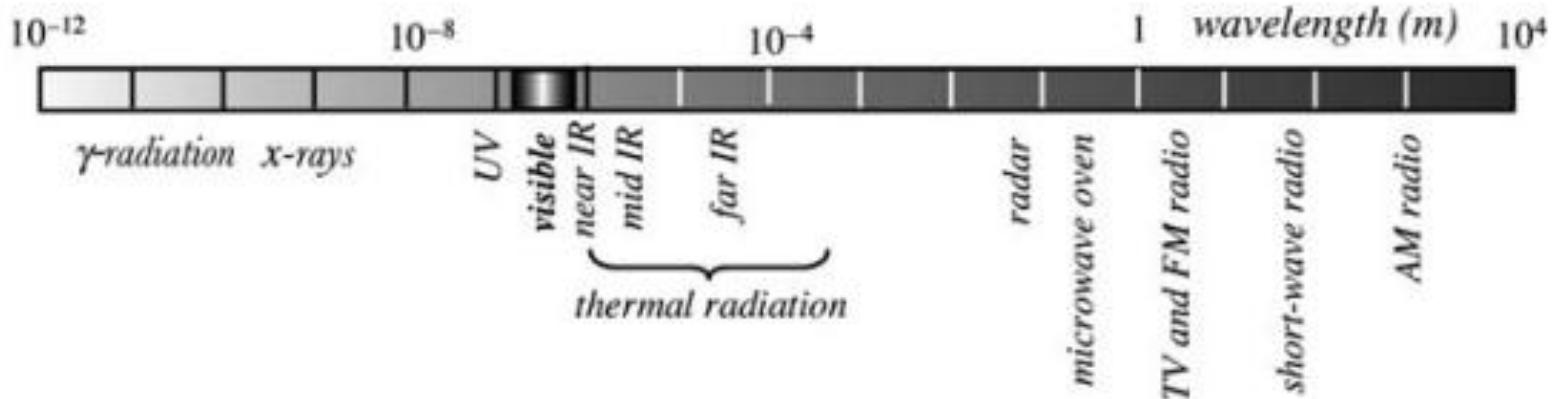
A piezoresistive internal combustion engine sensor.

UNIT -3

Radiation basics



Electromagnetic radiation spectrum spreading From g rays to radio waves.



The electromagnetic waves can be reflected, filtered focused

On its left-hand side, there is a region of the g-radiation. Then, there are the X-rays that depending on the wavelengths are divided into hard, soft, and ultrasoft rays.

However, a spontaneous radiation from the matter not necessarily should be electromagnetic: There is the so-called nuclear radiation, which is emission of particles from the atomic nuclei.

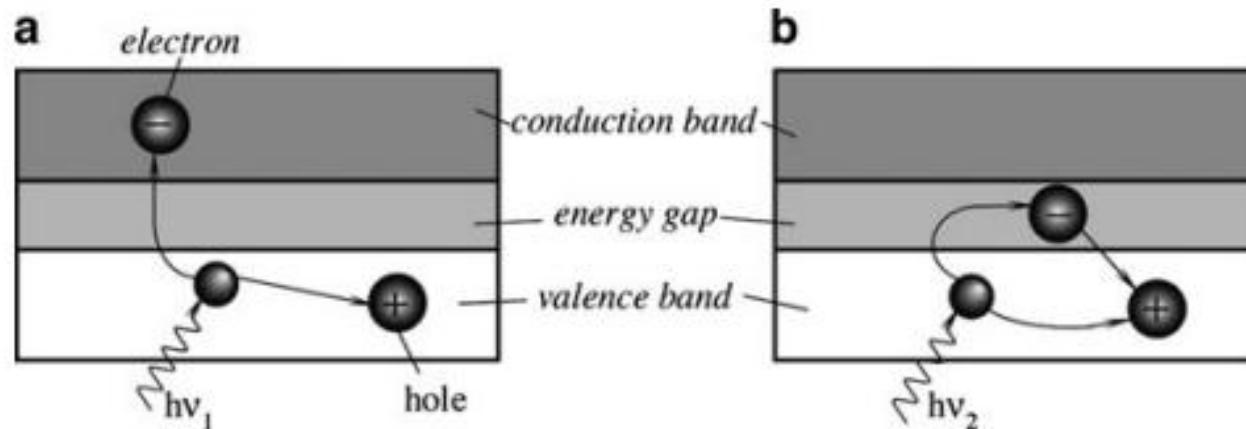
A spontaneous decay can be of two types: The charged particles (alpha and beta particles, and protons) and uncharged particles that are the neutrons. Some particles are complex like the alpha particles, which are the nuclei of helium atoms consisting of two neutrons, while other particles are generally simpler, like the beta particles that are either electrons or positrons. Ionizing radiations are given that name because as they pass through various media that absorb their energy, additional ions, photons, or free radicals are created.

Detectors of electromagnetic radiation in the spectral range from ultraviolet to far infrared are called light detectors.

Therefore, all light detectors are divided into two major groups that are called quantum and thermal.

The quantum detectors operate from the ultraviolet to mid-infrared spectral ranges, while thermal detectors are most useful in the mid- and far-infrared spectral ranges where their efficiency at room temperatures exceeds that of the quantum detectors.

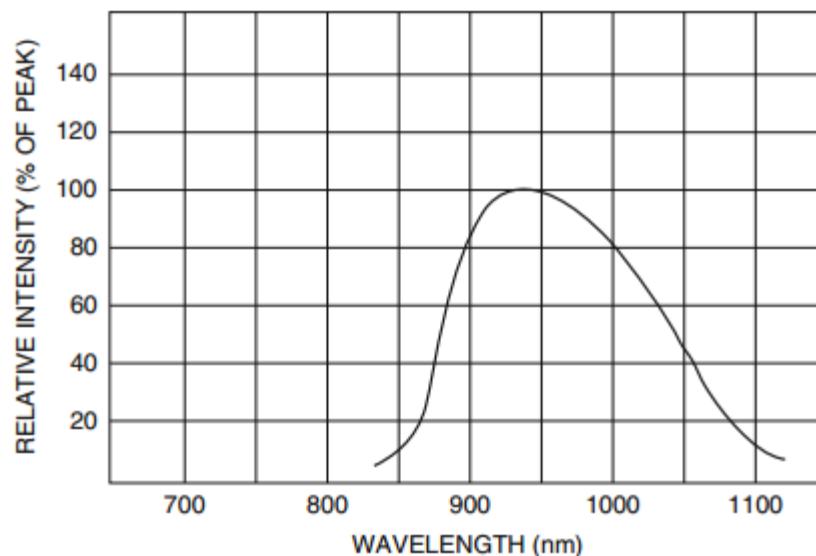
Solid-state quantum detectors (**photovoltaic and photoconductive devices**) rely on the interaction of individual photons with a crystalline lattice of semiconductor materials. Their operations are based on the photoeffect that was discovered by Albert Einstein, which won him the Nobel Prize.



14.1 Photoeffect in a semiconductor for high (a) and low (b) energy photons

A typical spectral response of a semiconductive material is shown in Fig. The light response of a bulk material can be altered by adding various impurities in the material. They can be used to reshape and shift a spectral response of the material.

All devices that directly convert photons of electromagnetic radiation into charge carriers are called quantum detectors, which are generally produced in the form of photodiodes, phototransistors, and photoresistors.



! Spectral response of an infrared photodiode

Photodetector principle

Photodetectors are of many types but they can be divided into two main classes

1. Thermal - thermal detectors detect light by a rise in temperature when the light is absorbed. They work mostly in the far IR region.
2. Photon - photon detectors work by creating electron-hole pairs on absorption of the incident radiation. The carrier concentration is proportional to the intensity of the incident radiation.



If λ is the wavelength of the incident radiation, then it can be ‘detected’ if

$$\lambda \leq \frac{hc}{\Delta E}$$

where ΔE is the energy of the *relevant transition* within the active region

Photodetectors are usually wavelength specific i.e. the material and device are chosen to work over a specific wavelength region

Photodetectors are usually wavelength specific i.e. the material and device are chosen to work over a specific wavelength region. Solar cells, on the other hand, are designed to work with the solar spectrum, which extends from the IR region to the visible and then UV.

An important factor in choosing the photodetector is the absorption coefficient of the semiconductor material. Optical absorption coefficients for different semiconductor materials are shown in figure 1. The absorption coefficient decides the *penetration depth* of the radiation into the device. This is given by Beer-Lambert law, and the penetration depth is the inverse of the absorption coefficient. If α is very large then most of the absorption will be close to the surface. On the other hand, if α is very small, then most of the light will pass through without absorption. The absorption coefficient, and through it the penetration depth, determines the working wavelength range of the photodetector, especially the lower limit. The upper wavelength limit

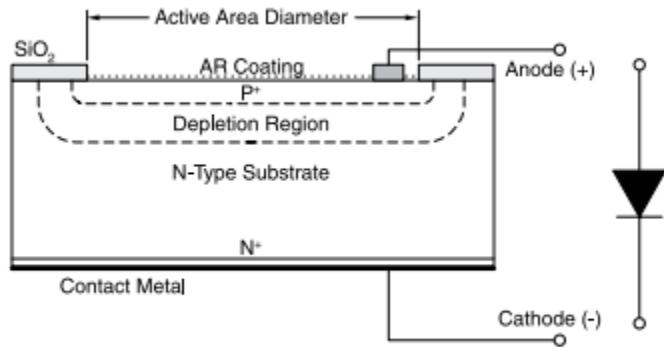


Figure 1. Planar diffused silicon photodiode

Photodetector I-V Curves

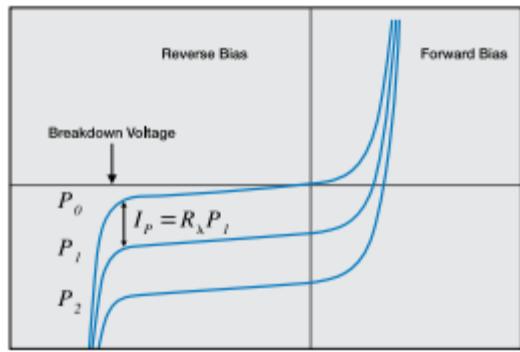


Figure 7. Characteristic I-V Curves of an OSi Optoelectronics photodiode for Photoconductive and Photovoltaic modes of operation. P_0-P_2 represent different light levels.

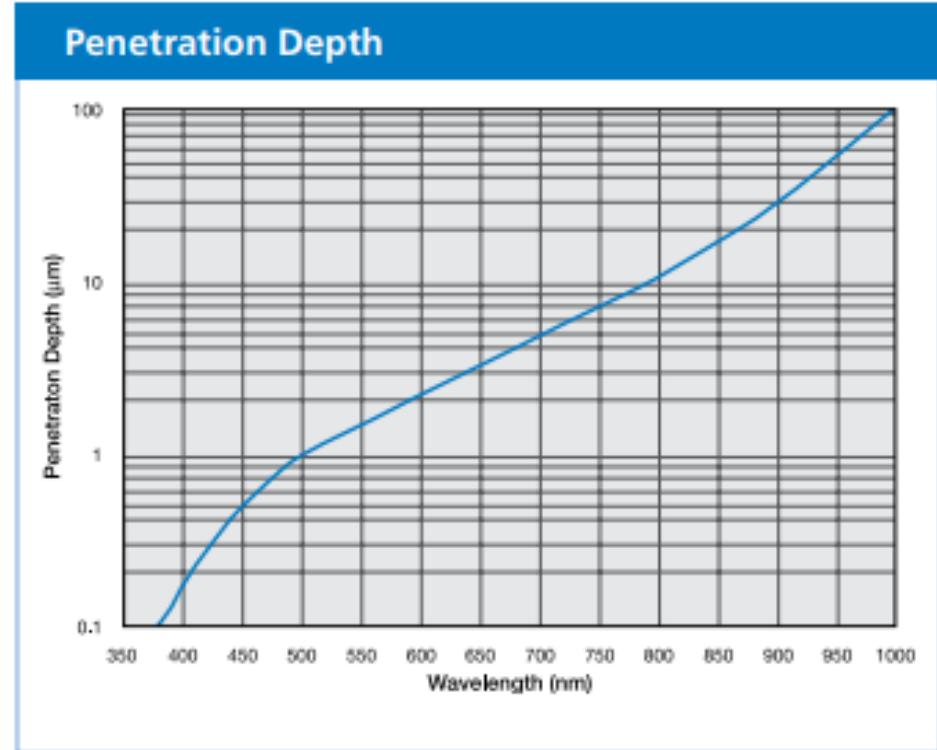
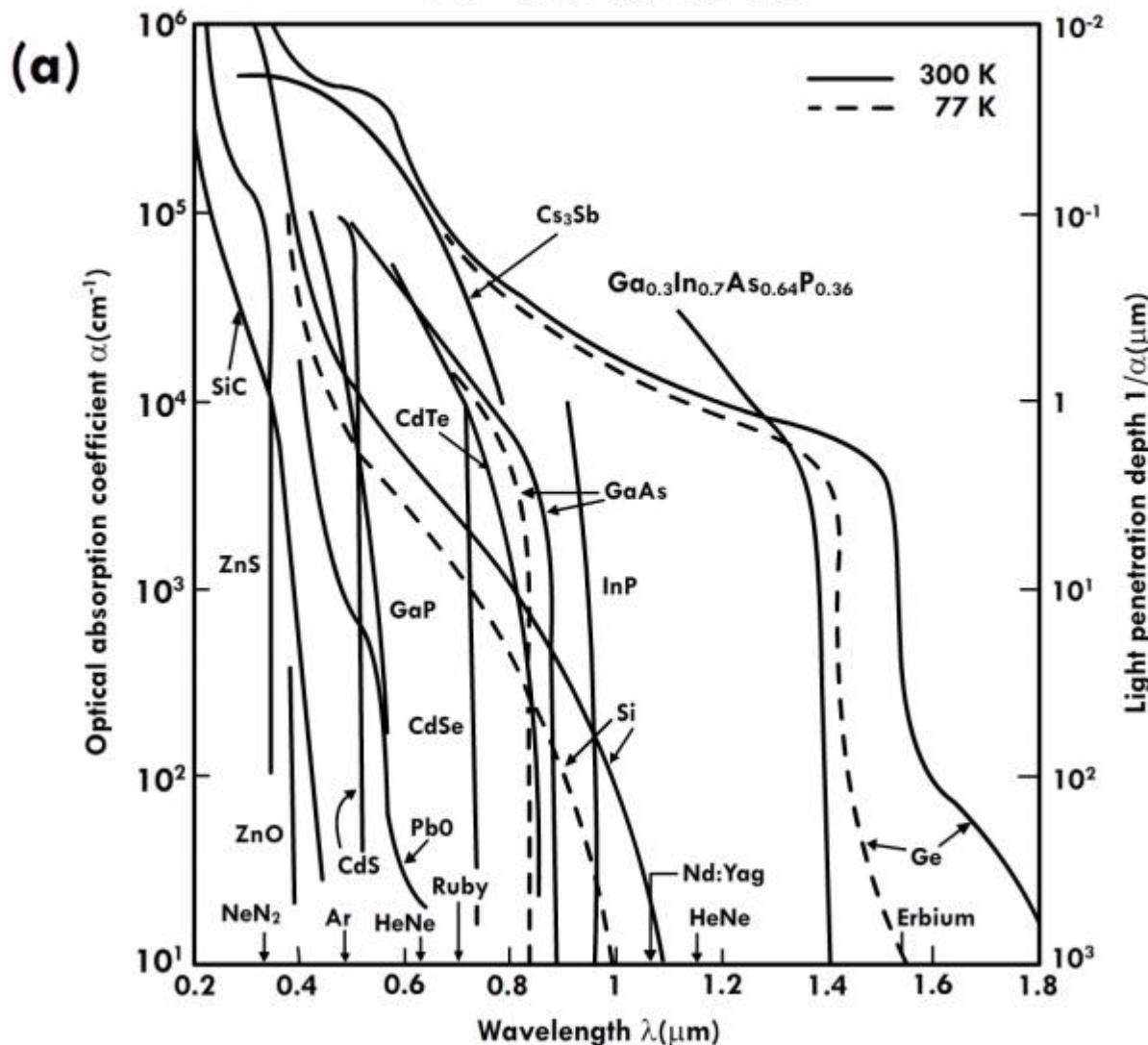
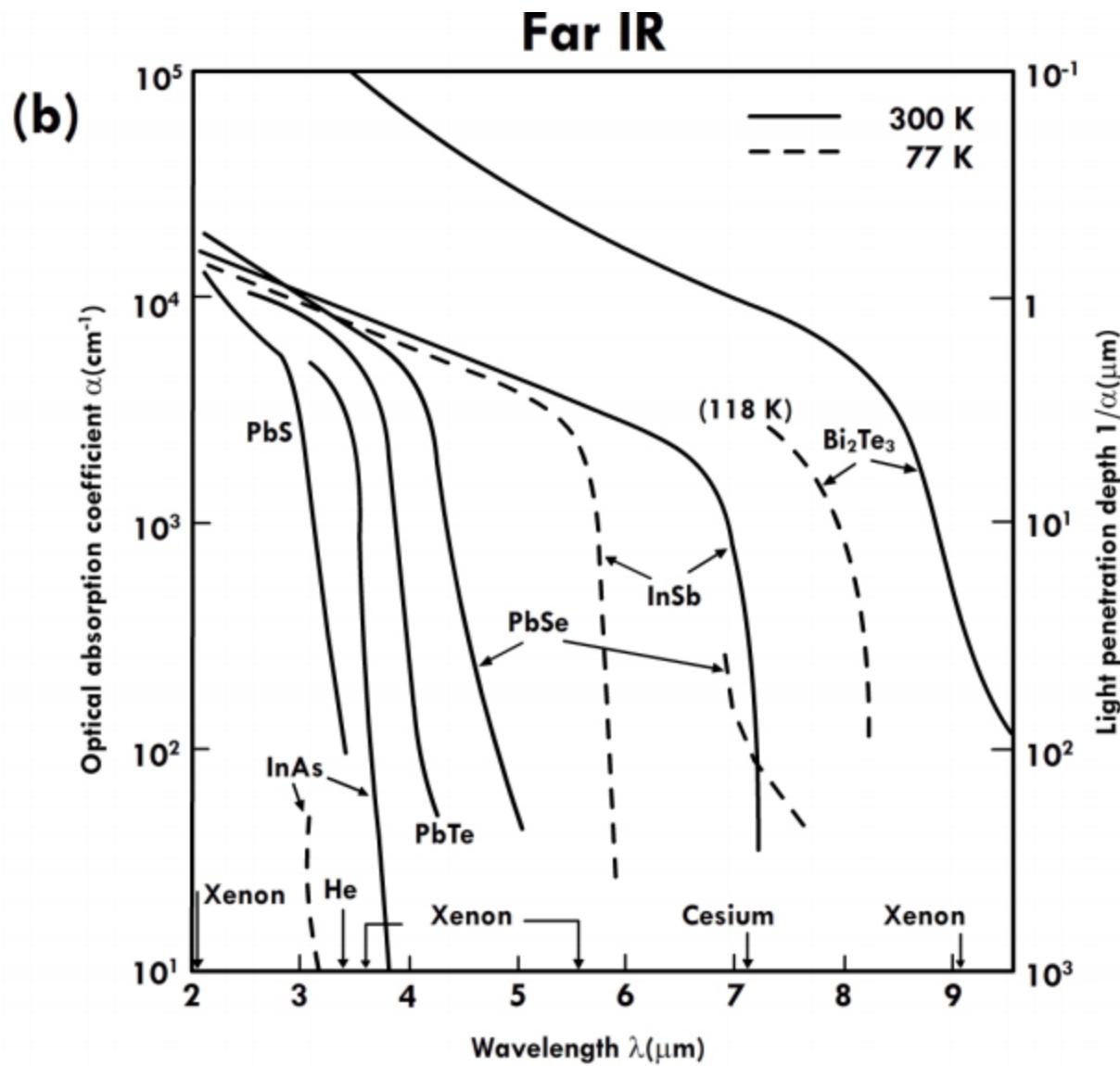


Figure 2. Penetration depth ($1/e$) of light into silicon substrate for various wavelengths.

Optical absorption coefficients for different semiconductors as a function of wavelength

UV-Vis-near IR





Another factor in photodetectors is the response time, especially when the radiation arrives in the form of pulses. The device speed is determined by the carrier generation rate. Also, the carrier generation and detection should be faster than the arrival rate of the next pulse. Carrier detection is related to the lifetime and also the distance the carriers have to travel before reaching the electrodes. One way to reduce the transit time is to reduce the size of the device, esp. in the active region. For a *pn* junction based photodetector, the active region is the depletion region, and this can be made smaller by increasing the doping concentration. But a smaller depletion region will also lead to a lower sensitivity, since the amount of light absorbed will be smaller.

The basic metric of the photodetector is the *quantum efficiency* (η). This is defined as the *number of carriers generated per photon*.

$$\eta = \frac{I_{ph}}{e\phi} = \frac{I_{ph}}{e} \left(\frac{h\nu}{P_{ot}} \right) \quad (2)$$

where I_{ph} is the photocurrent generated from the photon flux, ϕ , and this is related to the optical power P_{ot} .

Photosensors made of semiconductor materials gained much attention in recent years because they can be specially tailored for the needs of any optical application and they can be produced at low cost in big numbers within a rather short time of a few months. The size, the geometry in general, the spectral response and other parameters can be chosen almost freely. There are various structures available like PIN diodes, drift diodes, avalanche photodiodes, CCDs, phototransistors, etc. and the sensors can be made from all sorts of semiconductor materials like Si, Ge, GaAs, InGaAs.

HgCdTe(MCT) infrared sensors





HgCdTe infrared sensors

At present, HgCdTe is the most widely used variable gap semiconductor for infrared (IR) photodetectors. Over the last forty years it has successfully fought off major challenges from extrinsic silicon and lead-tin telluride devices, but despite that it has more competitors today than ever before.

Progress in IR detector technology is connected with semiconductor IR detectors, which are included in the class of photon detectors. In this class of detectors the radiation is absorbed within the material by interaction with electrons either bound to lattice atoms or to impurity atoms or with free electrons. The observed electrical output signal results from the changed electronic energy distribution. The photon detectors show a selective wavelength dependence of response per unit incident radiation power. They exhibit both perfect signal-to-noise performance and a very fast response.

But to achieve this, the photon detectors require cryogenic cooling. Photon detectors having long-wavelength limits above about 3 pm are generally cooled.

This is necessary to prevent the thermal generation of charge carriers. The thermal transitions compete with the optical ones, making non-cooled devices very noisy.

Cooling requirements are the main obstacle to the more widespread use of IR systems based on semiconductor photodetectors, making them bulky, heavy, expensive and inconvenient to use.

Current cooled IR detector systems use material such as HgCdTe, InSb, PtSi, and doped Si.

HgCdTe has inspired the development of three “generations” of detector devices.

The first generation, linear arrays of photoconductive detectors, has been produced in large quantities and is in widespread use today.

The second generation, two-dimensional arrays of photovoltaic detectors, is now in high-rate production – thousands of arrays annually.

Third generation devices, defined here to encompass the more exotic device structures embodied in two-colour detectors, avalanche photodiodes, and hyperspectral arrays, are now fielded in demonstration programs.

HgCdTe infrared sensors

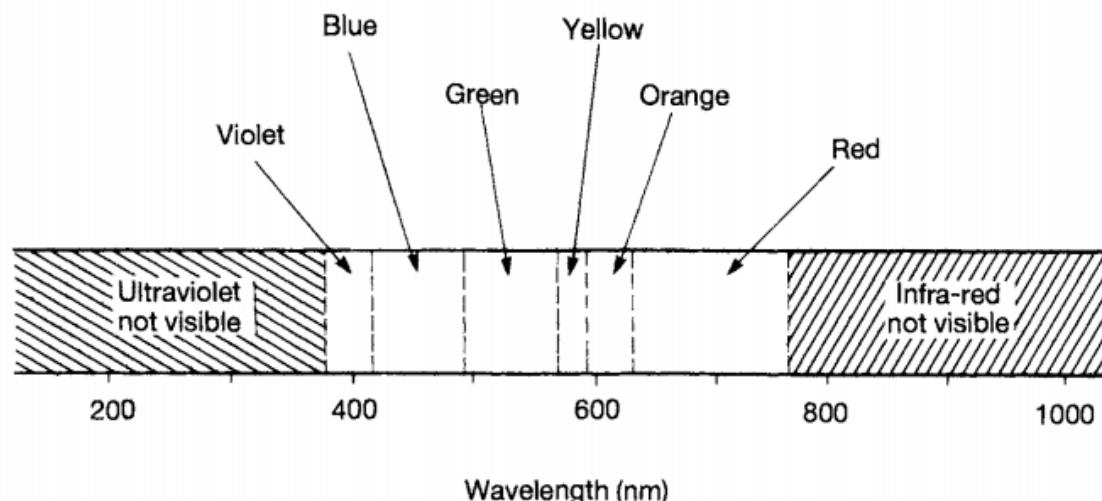


Fig. 1.2 The spectrum, with a rough indication of the colours seen.

$\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ is the most widely used material for high performance infrared detectors at present. By changing the composition x , the detector spectral response can be made to cover the range from $1 \mu\text{m}$ to beyond $17 \mu\text{m}$. The advantages of this system arise from a number of features, notably: close lattice matching, high optical absorption coefficient, low trap density, high electron mobility and readily available doping techniques. These advantages mean that background limited performance can be achieved at relatively high operating temperatures. HgCdTe continues to be developed as the material of choice for high performance long wavelength ($8\text{--}12 \mu\text{m}$) arrays and has an established market at the medium ($3\text{--}5 \mu\text{m}$) and short wavelength ($1\text{--}3 \mu\text{m}$) ranges.

In MEDIUM WAVE arrays the current–voltage ($I - V$) characteristics are usually close to ideal but at longer wavelength a number of leakage currents can impact on the $I - V$ characteristics and degrade the performance of the detector. Figure illustrates a typical $I - V$ characteristic for a LONG WAVE diode, along with two types of MCT photodiode, a planar diode and a via-hole diode.

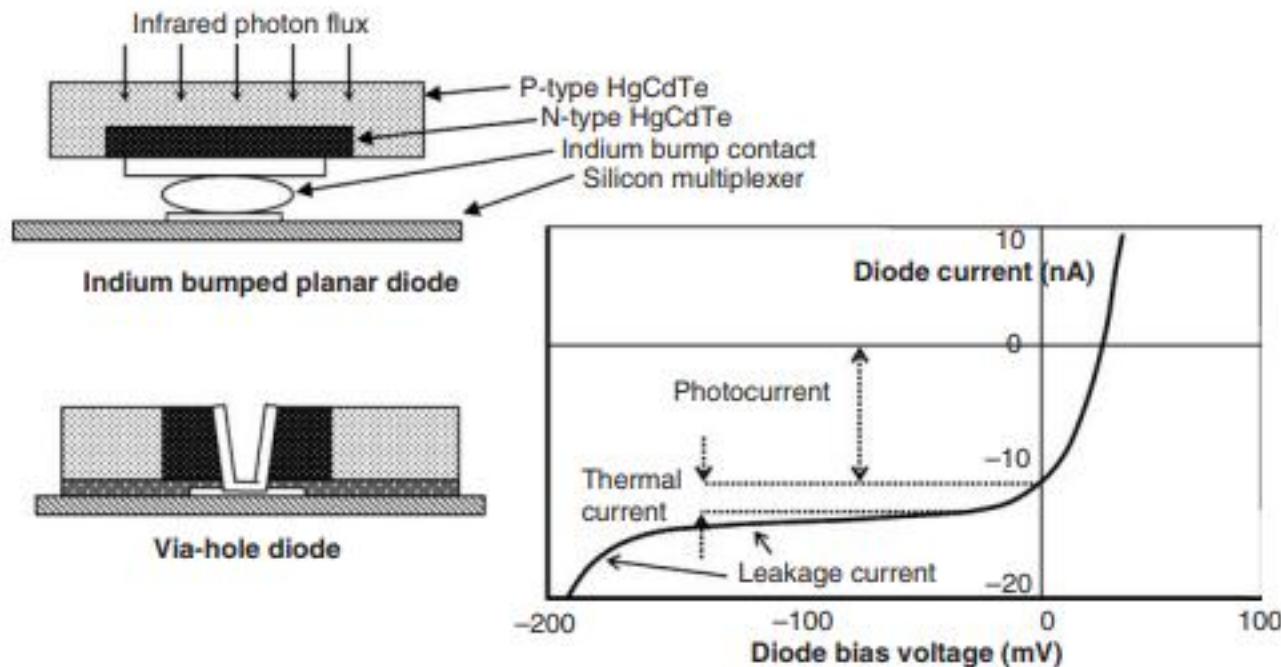
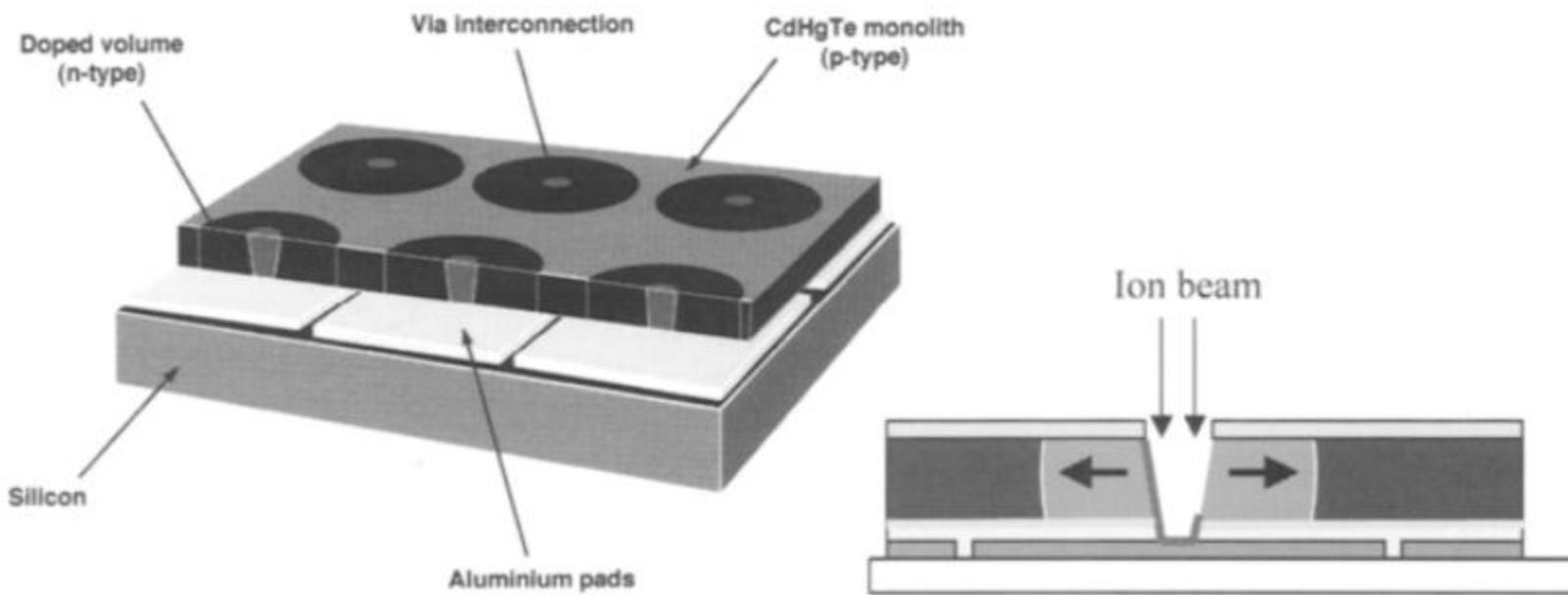


Figure 19.2 Types of HgCdTe diode and a typical current–voltage characteristic for a LW diode showing main non-ideal features.



Photocurrent and quantum efficiency

MCT has a strong optical absorption coefficient and only thin layers are needed to produce high quantum efficiency. Typically in MW detectors the absorber need only be 4–5 μm thick and about twice this in LW detectors. Ideally, the absorption should occur well within a diffusion length of the p–n junction to avoid signal loss due to recombination.

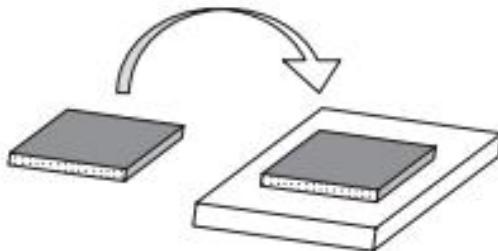
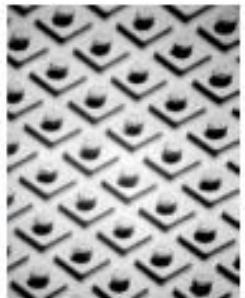
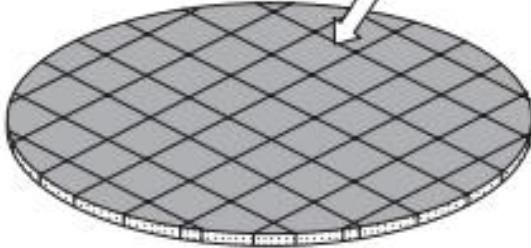
A long carrier lifetime is nearly always observed in n-type material with low carrier concentration. Device engineers tend to favor using n-type absorbers for the best quantum efficiency and try to minimize the volume of the p-region for lower thermal leakage currents.

MANUFACTURING TECHNOLOGY FOR MCT ARRAYS

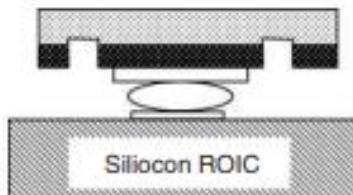
Considerable progress has been made in the past two decades on the epitaxial growth of MCT. Manufacturers select a technique that suits their device technology and the type of detectors they require. It is the aim of most manufacturers to produce high-quality layers in large areas at low cost, but this ideal has been elusive. At the present time the best structural quality material is grown using liquid phase epitaxy (LPE) onto lattice-matched crystals of CdZnTe, and this has been used successfully in homojunction technologies. CdZnTe is expensive, however, and both LPE and VPE now use a variety of alternative substrates. Many groups favor VPE because the composition and doping profile can be easily controlled to produce complex devices, such as two-color detectors.

Two VPE methods are commonly used: metal organic vapor phase epitaxy (MOVPE) and molecular beam epitaxy (MBE).

Photomicrograph of mesa diode array with indium bumps

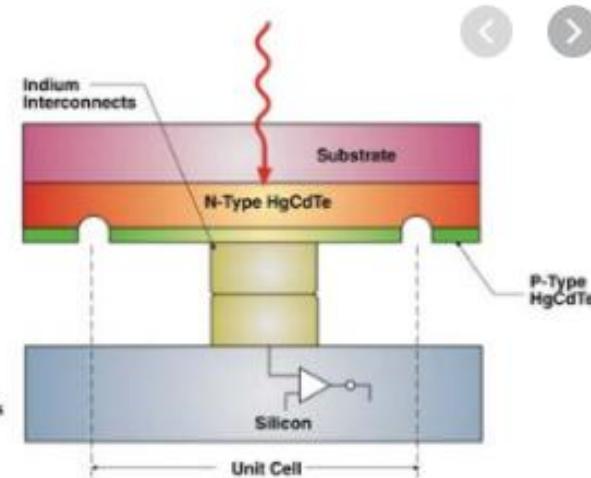
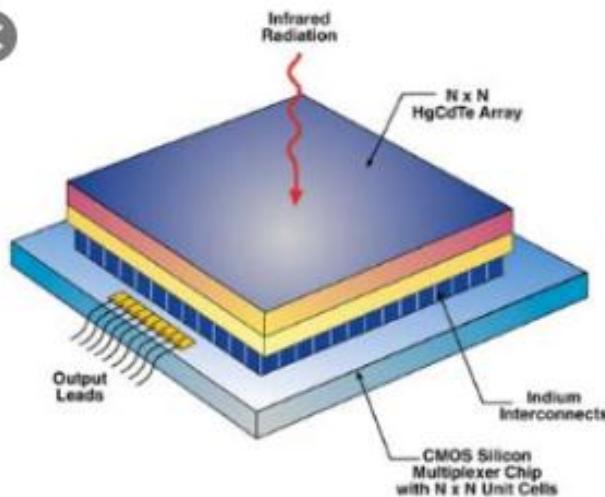


Flip-chipped onto silicon ROIC with substrate left on or thinned



Indium bumped mesa diode

HgCdTe layer grown on large area substrate (Si, GaAs or sapphire) using VPE growth method. Processed into arrays with indium bumps and sawn up into array blocks.



The main commercial drive though is to cost reduce the manufacture of large-area arrays and avoid expensive CdZnTe substrates. Silicon, gallium arsenide and sapphire are the commonest substrates used because they are available commercially in large area wafers at relatively low cost

Wafer-scale processing using VPE MCT material, mesa diodes, and indium bump interconnect technology

Visible-light color sensors



Visible-light color sensors

In the food & beverages industry, color is a very important measurand which can be used either for quality estimation, or for automatic selection of products in different categories.

Reflectivity measurements can be easily employed for solid foods, whereas for liquid samples, it is much more appropriate to measure the sample in transmission.

dependence of blood color on the concentration of deoxygenated and oxygenated hemoglobin can also be applied to estimate quantitatively the freshness of fresh meat. Thus, by analyzing the reflectance spectrum of beef meat and its variation in time during storage, it was deduced that the typical spectrum changes in time due to exposure to oxygen, which causes the oxymyoglobin to be oxidized into methemyoglobin, resulting in a quality degradation-induced color change that can be automatically sensed so that packaged meat that exceeded its shelf-life could be promptly removed and replaced with fresh one.

Similar color-based quality analysis has also been successfully performed on other foods (for diagnosis of storage conservation or for estimation of their alimentary properties), e.g., milk, orange or strawberry juice, or peach nectar.

Another example is the reported realization of an accurate real-time color classification microsystem that integrated the photosensors, together with pre-processing circuitry and a subsequent neural network processor, onto a single IC. The usage of neural networks offers distinct advantages, such as extreme ease of usage by the user of the final product, flexibility by **self-adaptation to new circumstances, and reduced cost and extreme suitability of monolithic integration with standard CMOS fabrication processes.**

This one-chip smart sensing microsystem had low cost, was robust, was mass-producible using standard commercial CMOS processes, and exhibited a significantly higher performance. The chip was successfully applied practically in freshness tests for several fruits (apples, tangerines, and lemons)

Phenolic compounds are responsible for the characteristic color, flavor, and aroma in wines and also act as antioxidants, with alleged beneficial effects on human health (reduced incidence of coronary heart disease and certain forms of cancer).

Color sensing has been employed not only for fresh produce but also to characterize the cooking of foods. This is important in order to avoid under- or over-cooking, and also to ensure objectivity of assessment since human evaluation is highly subjective, may also depend on other variables, and does not guarantee accurate reproducibility.

One reported realization for this purpose comprised an optical fiber sensor in conjunction with a small portable Ocean Optics spectrometer in order to monitor the color of the food while it was being cooked by examining the light reflected from both the sample's surface and core.

Many modern manufacturing processes require the detection of different colors and hues of visible light, i.e., wavelengths in the range 400–700 nm.

Color detection can be used to sort objects, verify position of objects, recognize color sequence, control color in dyeing and coating applications, and detect changes in color of liquid during titration.

Hence, color sensors are becoming an integral part in many industries, e.g., cosmetics, textile, food, publishing, optoelectronics, and image processing, including digital cameras.

Photodiodes

Photodiodes are semiconducting optical sensors, which if broadly defined may even include solar batteries. However, here we consider only the information aspect of these devices rather than the power conversion.

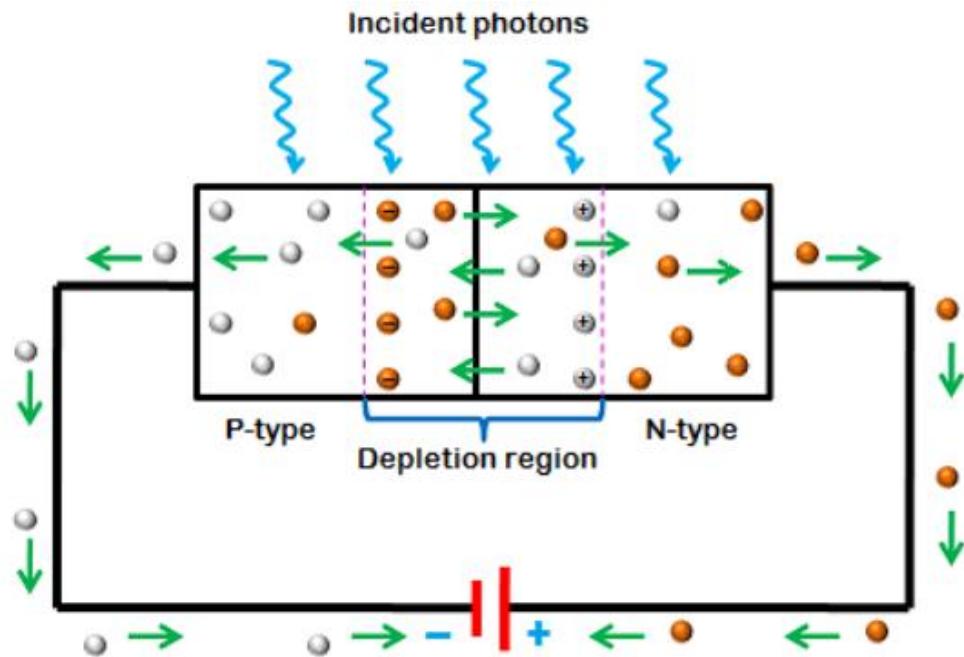
In a simple way, the operation of a photodiode can be described as follows.

If a pn-junction is forward biased (positive side of a battery is connected to the p side) and is exposed to light of proper frequency, the current increase will be very small with respect to a dark current. In other words, the bias current is much greater than the current generated by light, and the diode is just a diode, not really useful for sensing light.

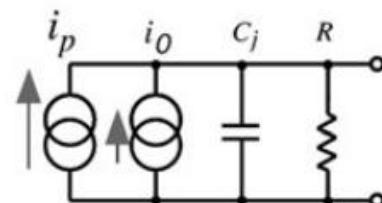
If the junction is reverse biased , when light strikes the semiconductor, the current will increase quite noticeably. Impinging photons create electron– hole pairs on both sides of the junction. When electrons enter the conduction band, they start flowing toward the positive side of the battery. Correspondingly, the

created holes flow to the negative terminal, meaning that photocurrent i_p flows in the network. Under dark conditions, dark current i_0 is independent of applied voltage and mainly is the result of thermal generation of charge carriers. Thus, a reverse-biased photodiode electrical equivalent circuit contains two current sources and a RCnetwork

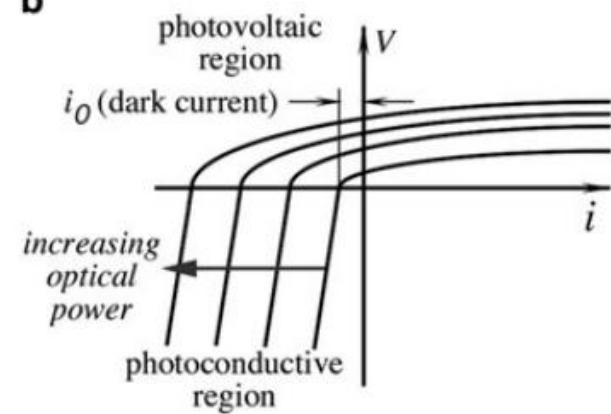
Photodiodes



a



b



4 An equivalent circuit of a photodiode (a) and its volt-ampere characteristic (b)

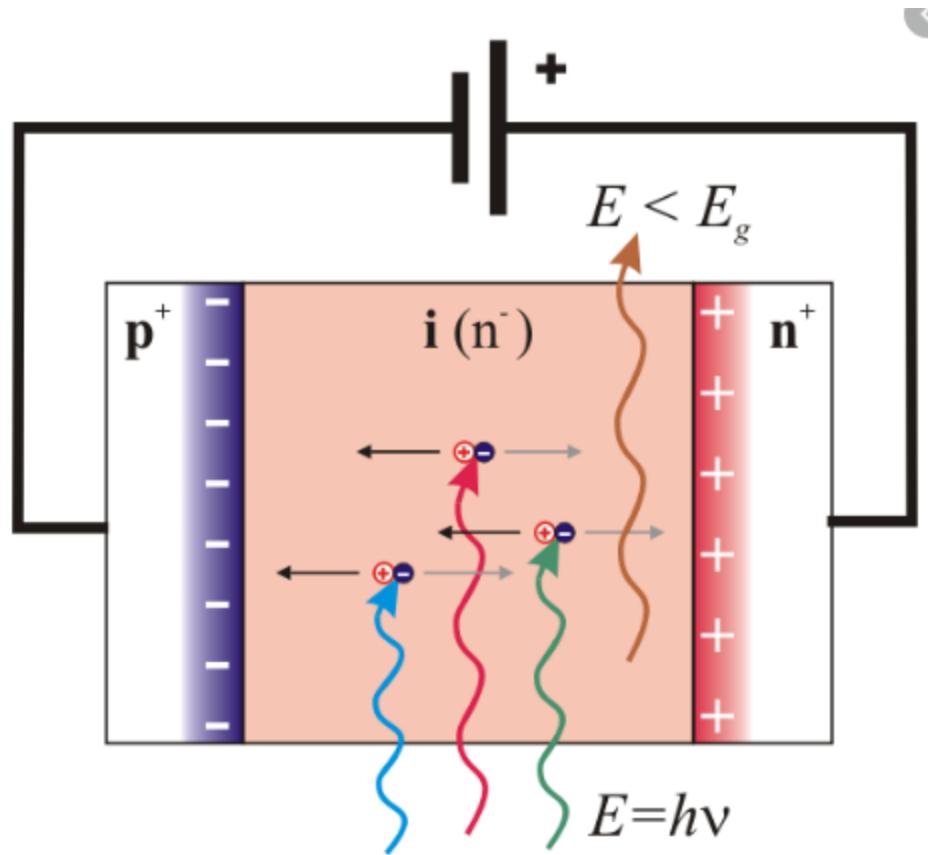
The current-to-voltage response of the photodiode is given by

$$i = i_0 \left(e^{eV/k_b T} - 1 \right) - i_s,$$

$$i = i_0 \left(e^{eV/k_b T} - 1 \right) - \frac{\eta e P}{h v},$$

$$i = \dots \frac{\eta e P}{h v},$$

If the probability that a photon of energy $h v$ will produce an electron in a detection is η then the average rate of the production of electrons for an incident beam of optical power P is given by



PN photodiode:

- A PN photodiode does not require a reverse bias and as a result is more suitable for low light applications as a result of the improved noise performance.

PIN photodiode:

- Reverse bias required by the PIN photodiode introduces a noise current which reduces signal to noise ratio
- The reverse bias offers better performance for high dynamic range applications
- The reverse bias required offers better performance for high bandwidth applications as the capacitance between the P and N regions as well as charge storage is small.

Avalanche photodiode advantages

- High level of sensitivity as a result of avalanche gain

Avalanche photodiode disadvantages:

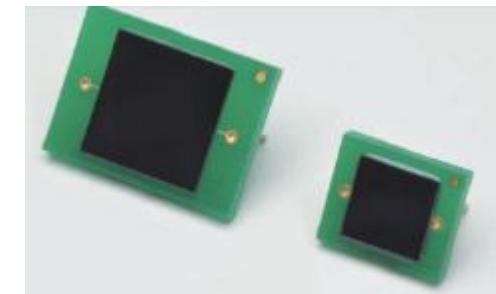
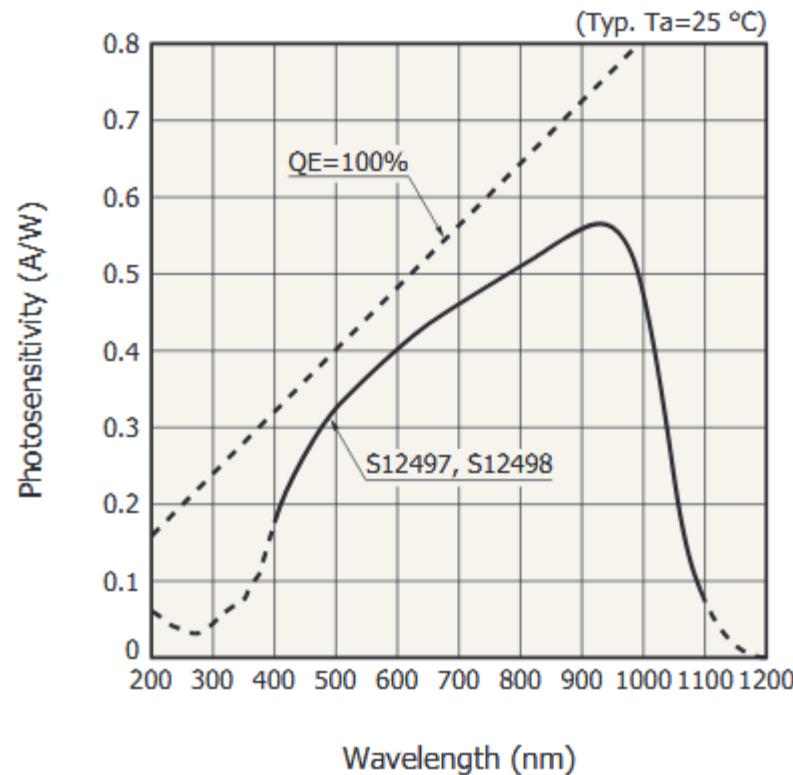
- Much higher operating voltage may be required.
- Avalanche photodiode produces a much higher level of noise than a PN photodiode
- Avalanche process means that the output is not linear

Si photodiodes

A commercial sensor example

Spectral response range :
400 to 1100 nm

Photosensitive area
S12497: 9.5×9.5 mm,
S12498: 6.0×6.0 mm



► Electrical and optical characteristics (Ta=25 °C)

Parameter	Symbol	Condition	S12497			S12498			Unit
			Min.	Typ.	Max.	Min.	Typ.	Max.	
Spectral response	λ		400	-	1100	400	-	1100	nm
Peak sensitivity wavelength	λ_p		-	920	-	-	920	-	nm
Photosensitivity	S	$\lambda=540$ nm	0.32	0.36	-	0.32	0.36	-	A/W
		$\lambda=920$ nm	0.52	0.57	-	0.52	0.57	-	
Short circuit current	Isc	100 lx, 2856 K	60	75	-	15	30	-	µA
Dark current	Id	VR=10 mV	-	50	200	-	10	150	pA
Rise time	tr	VR=0 V, RL=1 kΩ 10 to 90%, $\lambda=658$ nm	-	15	-	-	15	-	µs
Terminal capacitance	Ct	VR=0 V, f=10 kHz	750	950	1150	330	380	430	pF



Si Photodiode



PHOTORESISTORS

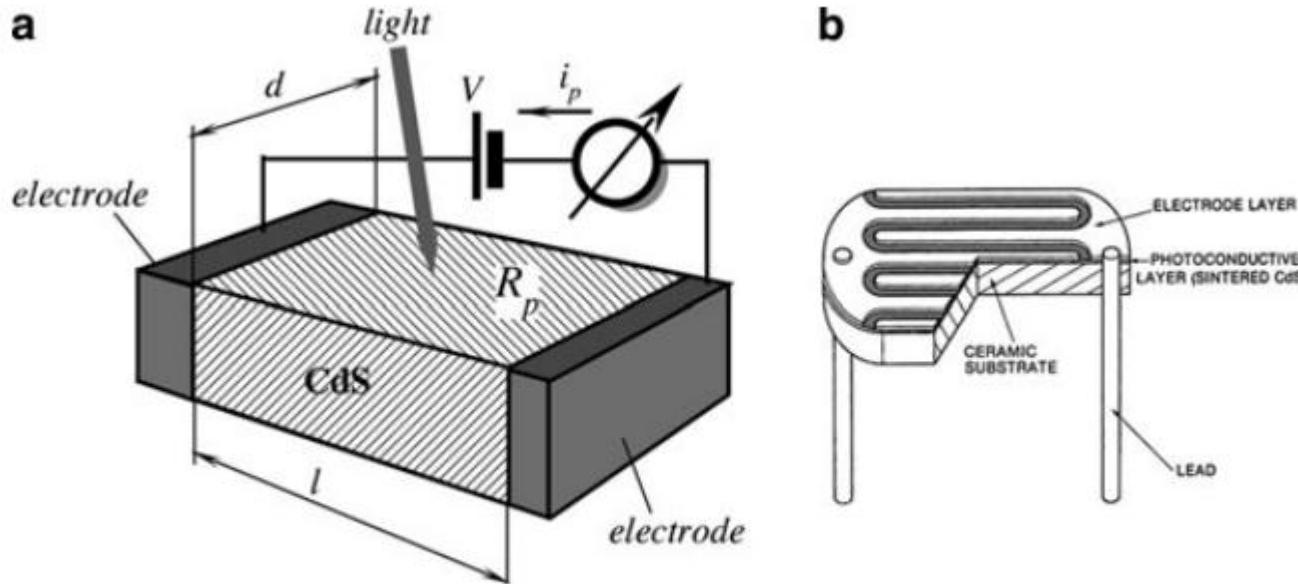


Fig. 14.11 Structure of a photoresistor (a) and a plastic-coated photoresistor having a serpentine shape (b)

As a photodiode, a photoresistor is a photoelectric device. It is a resistor whose resistance is called photoresistance R_p that changes as the function of incident light. The most common materials for its fabrication are cadmium sulfide (CdS) and cadmium selenide (CdSe), which are semiconductors whose resistances change upon light entering the surface. For its operation, a photoresistor requires a power source (excitation signal) because unlike a photodiode or phototransistor, it does not generate photocurrent – a photoeffect is manifested in change in the material's electrical resistance.

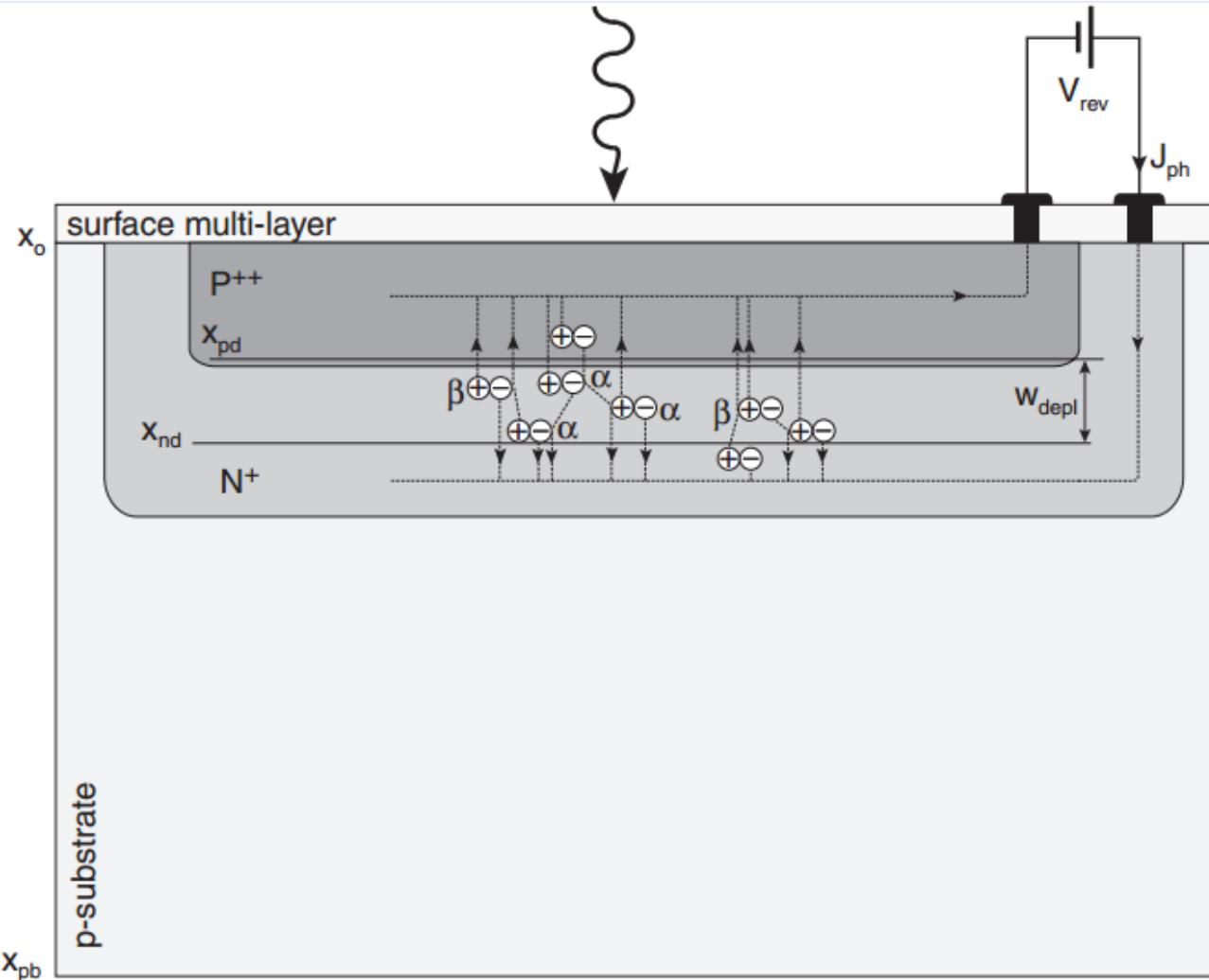
In darkness, the resistance of the material is high. Hence, the applied voltage V results in small dark current, which is attributed to temperature effect. When light is incident on the surface, current i_p flows.

When light illuminates the photoconductive crystal, photons are absorbed, which results in the added-up energy in the valence band electrons. This moves them into the conduction band, creating free holes in the valence band, increasing the conductivity of the material. Since near the valence band is a separate acceptor level that can capture free electrons not as easily as free holes, the recombination probability of the electrons and holes is reduced and the number of free electrons in the conduction band is high.

CdS has a band gap of 2.41 eV, the absorption edge wave length is 515 nm, which is in the visible spectral range. Hence, the CdS detects light shorter than 515 nm wavelengths. Other photoconductors have different absorption edge wavelengths. For instance, CdS is the most sensitive at shorter wavelengths range, while Si and Ge are the most efficient in the near infrared.

Avalanche photodiode

The avalanche photodiode is similar to the conventional pn junction, apart from the fact that the doping levels are relatively high. The operation of the avalanche photodiode is based on impact ionization within the depletion layer. The basic device structure is shown in Figure



The doping concentration of the p⁺⁺ layer exceeds 10^{19} cm^{-3} and n⁺ is in the order of $2 \times 10^{18} \text{ cm}^{-3}$ (this is generally referred to as the single-sided abrupt junction). Under these conditions and $V_{\text{rev}} \gg 10 \text{ V}$,

$$w_{\text{depl}} \approx \sqrt{\frac{2\epsilon}{qN_D} V_{\text{rev}}} \approx 0.1 \mu\text{m}$$

$$E_{\text{max}} = E(x_j) \approx \sqrt{\frac{2qN_D}{\epsilon} V_{\text{rev}}} \approx 10^7 \text{ V/m}$$

The avalanche photodiodes are named so because if a reverse bias is applied to the p-n junction and a high-intensity field is formed with the depletion layer, photon carriers will be accelerated by the field and collide with the atoms, producing the secondary carriers. In turn, the new carriers are accelerated again, resulting in the extremely fast avalanche-type increase in current. There-fore, these diodes work as amplifiers, making them useful for detecting extremely low levels of light.

The device is operated fully depleted and has two field regions: i) a light absorption region followed by the π region which is about (30–150) μm wide with a relatively low field ($\sim 2 \times 10^4 \text{ V cm}^{-1}$) wherein charge carriers are drifting; ii) a narrow (a few micrometers) high field region ($\sim 10^5 \text{ V cm}^{-1}$) wherein the multiplication takes place.

The principle of operation of an APD is based on the conversion of the energy of incoming photons into electron–holes pairs in silicon as in PIN and their further multiplication through an avalanche process in the depleted region by application of a very high reverse bias voltage to a p – n junction. The charge carriers created in the depleted region of the junction by the incident radiation, will drift in the π region towards the corresponding electrodes under the electric field applied across the junction: the holes will drift toward the p + side in a low electric field, while the electrons will drift toward the narrow n + side in a high electric field where an avalanche process will eventually take place, i.e., the electrons will be accelerated and acquire a kinetic energy large enough so that they can create additional electron-hole pairs via ionization through their collisions with atoms in the crystal lattice. At sufficiently high voltage, i.e., high electric field values, holes can also produce electron-hole pairs and therefore produce additional electrons. This electron multiplication generates avalanche breakdown that occurs at the so-called breakdown voltage which is the maximum reverse bias voltage which can be applied to an APD (when an APD operates close to avalanche breakdown, its gain is a very rapidly increasing function of the applied reverse bias voltage). The value of the breakdown voltage varies with the type of APD. Then, the operating voltage of an APD has to be kept lower than this breakdown voltage. In practice, this avalanche multiplication process - referred to as avalanche multiplication of photocurrent - occurs beyond a value of the applied voltage corresponding to an internal electric field

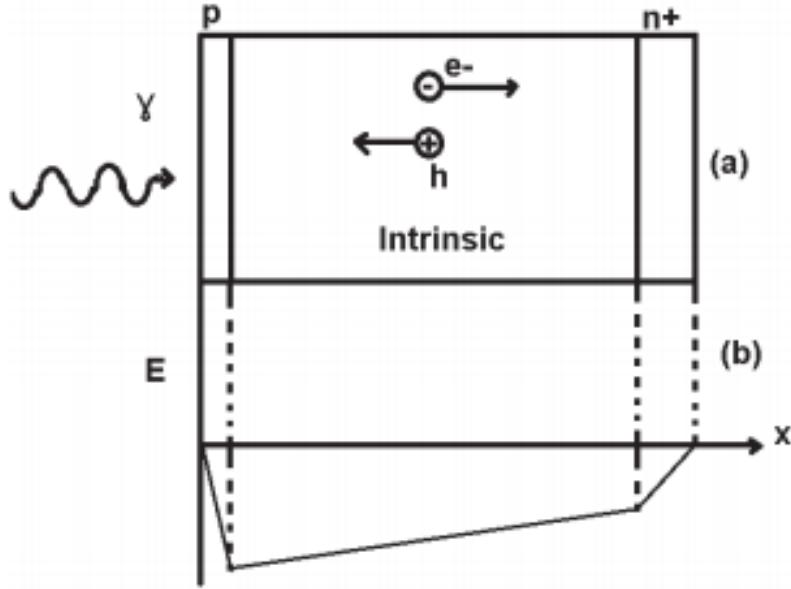


Fig. 6.1 a) Standard schematic representation of a photodiode; b) the electric field E generated by the application of a bias voltage.

depends on the voltage applied to the junction and gain of about 102 can be normally achieved for a bias voltage of (100–200) volts in silicon. The shape of the avalanche gain is an exponential (plus a constant) function of the applied bias voltage.

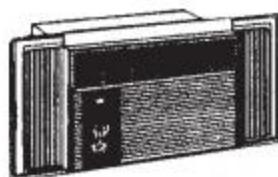
Heat transfer, thermal structures



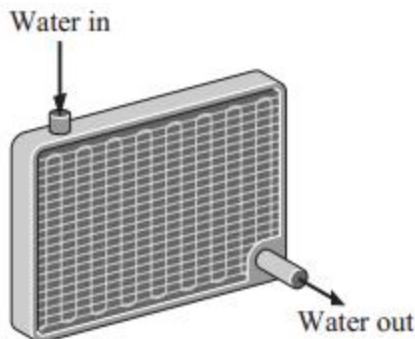
Application Areas of Heat Transfer



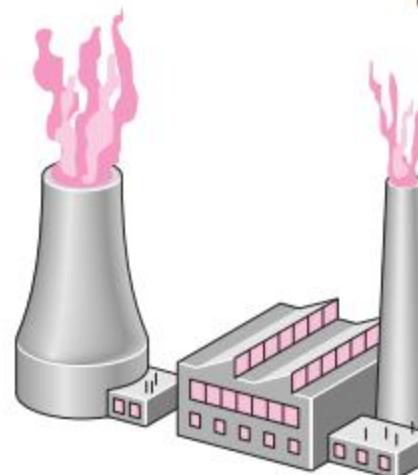
The human body



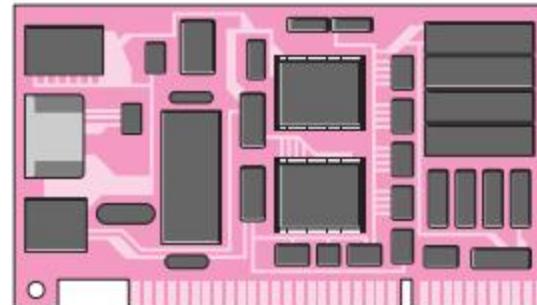
Air-conditioning
systems



Car radiators



Power plants



Circuit boards



Refrigeration systems

Heat transfer equipment such as heat exchangers, boilers, condensers, radiators, heaters, furnaces, refrigerators, and solar collectors are designed primarily on the basis of heat transfer analysis.

The heat transfer problems encountered in practice can be considered in two groups: (1) rating and (2)sizing problems.

The rating problems deal with the determination of the heat transfer rate for an existing system at a specified temperature difference.

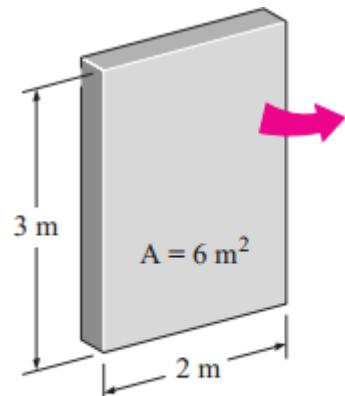
The sizing problems deal with the determination of the size of a system in order to transfer heat at a specified rate for a specified temperature difference.

Keeping in line with current practice, we will refer to the thermal energy as *heat* and the transfer of thermal energy as *heat transfer*. The amount of heat transferred during the process is denoted by Q . The amount of heat transferred per unit time is called **heat transfer rate**, and is denoted by \dot{Q} . The overdot stands for the time derivative, or “per unit time.” The heat transfer rate \dot{Q} has the unit J/s, which is equivalent to W.

The rate of heat transfer per unit area normal to the direction of heat transfer is called **heat flux**, and the average heat flux is expressed as (Fig. 1–11)

$$\dot{q} = \frac{\dot{Q}}{A} \quad (\text{W/m}^2) \quad (1-8)$$

where A is the heat transfer area. The unit of heat flux in English units is Btu/h · ft². Note that heat flux may vary with time as well as position on a surface.

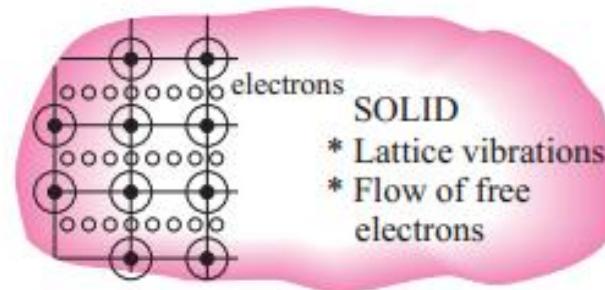
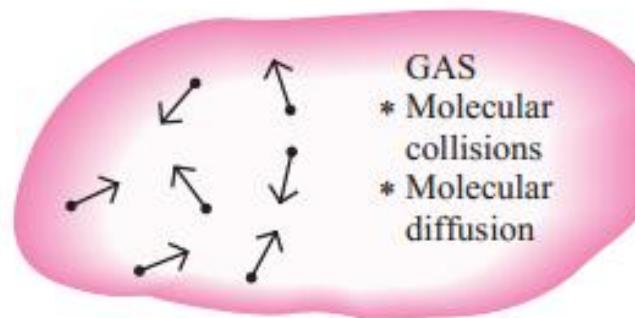


Heat flux is heat transfer *per unit time* and *per unit area*, and is equal to $\dot{q} = \dot{Q}/A$ when \dot{Q} is uniform over the area A .

$$\dot{q} = \frac{\dot{Q}}{A} = \frac{24 \text{ W}}{6 \text{ m}^2} = 4 \text{ W/m}^2$$

The transfer of energy as heat is always from the higher-temperature medium to the lower-temperature one, and heat transfer stops when the two mediums reach the same temperature.

Thermodynamics describes the fundamental behavior of heat and temperature and includes the three laws of thermodynamics. Heat transfer goes further and describes the mechanisms of heat exchange and the rate at which heat flows, giving us a way to calculate heat flow within, to and from objects or the environment. **There are three modes of heat transfer: conduction, convection and radiation.**



The mechanisms of heat conduction in different phases of a substance.

Conduction is the transfer of energy from the more energetic particles of a substance to the adjacent less energetic ones as a result of interactions between the particles. Conduction can take place in solids, liquids, or gases.

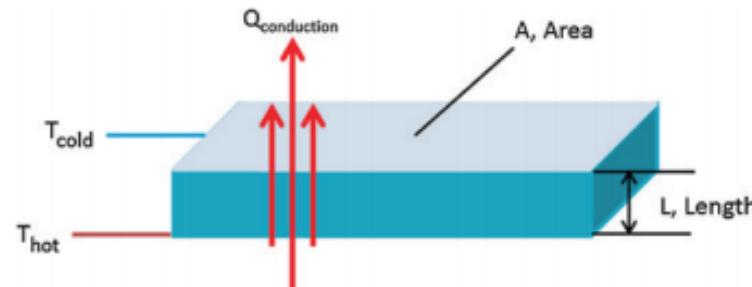
In gases and liquids, conduction is due to the collisions and diffusion of the molecules during their random motion. In solids, it is due to the combination of vibrations of the molecules in a lattice and the energy transport by free electrons.

Conduction is the heat transfer through solids. It can also occur with stagnant fluids. The one-dimensional rate of conductive heat transfer is determined by Eq. below:

Consider steady heat conduction through a large plane wall of thickness $\Delta x = L$ and area A , as shown in Fig. 1–21. The temperature difference across the wall is $\Delta T = T_2 - T_1$. Experiments have shown that the rate of heat transfer \dot{Q} through the wall is *doubled* when the temperature difference ΔT across the wall or the area A normal to the direction of heat transfer is doubled, but is *halved* when the wall thickness L is doubled. Thus we conclude that *the rate of heat conduction through a plane layer is proportional to the temperature difference across the layer and the heat transfer area, but is inversely proportional to the thickness of the layer*. That is,

$$\text{Rate of heat conduction} \propto \frac{(\text{Area})(\text{Temperature difference})}{\text{Thickness}}$$

$$Q_{\text{conduction}} = \frac{kA(T_{\text{hot}} - T_{\text{cold}})}{L}$$



1.3 Heat transfer through an object by conduction

where Q conduction is heat flow, k is the thermal conductivity of the material, A is the cross-sectional area of heat flow, T_{hot} is the temperature of the hot surface, T_{cold} is the temperature of the cold surface and L is the length of the material through which heat is conducting. Figure depicts the heat transfer through a solid material by conduction. The different variables of the conduction heat transfer are shown. The conduction resistance is defined by Eq.:

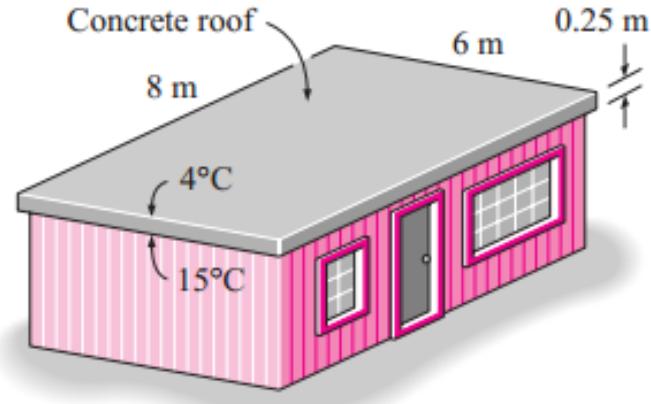
$$R_{\text{conduction}} = \frac{L}{kA}$$

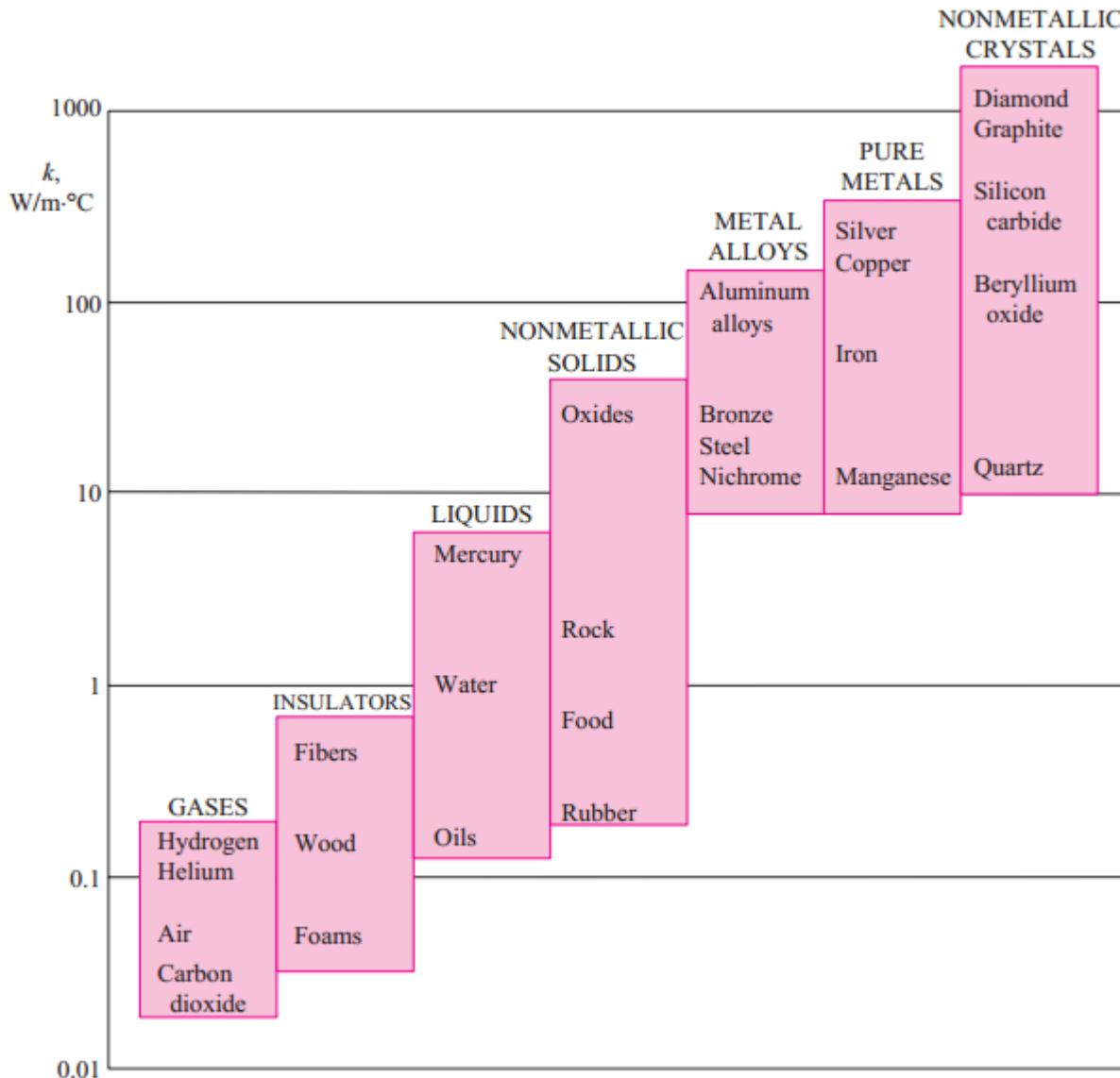
As shown in, in order to minimize the conduction resistance, conductivity of the material and cross-sectional area of material is maximized while the through-path (length) of the material is minimized.

The roof of an electrically heated home is 6 m long, 8 m wide, and 0.25 m thick, and is made of a flat layer of concrete whose thermal conductivity is $k = 0.8 \text{ W/m} \cdot ^\circ\text{C}$. The temperatures of the inner and the outer surfaces of the roof one night are measured to be 15°C and 4°C , respectively, for a period of 10 hours. Determine the rate of heat loss through the roof that night

Assume Steady operating conditions exist during the entire night since the surface temperatures of the roof remain constant at the specified values.

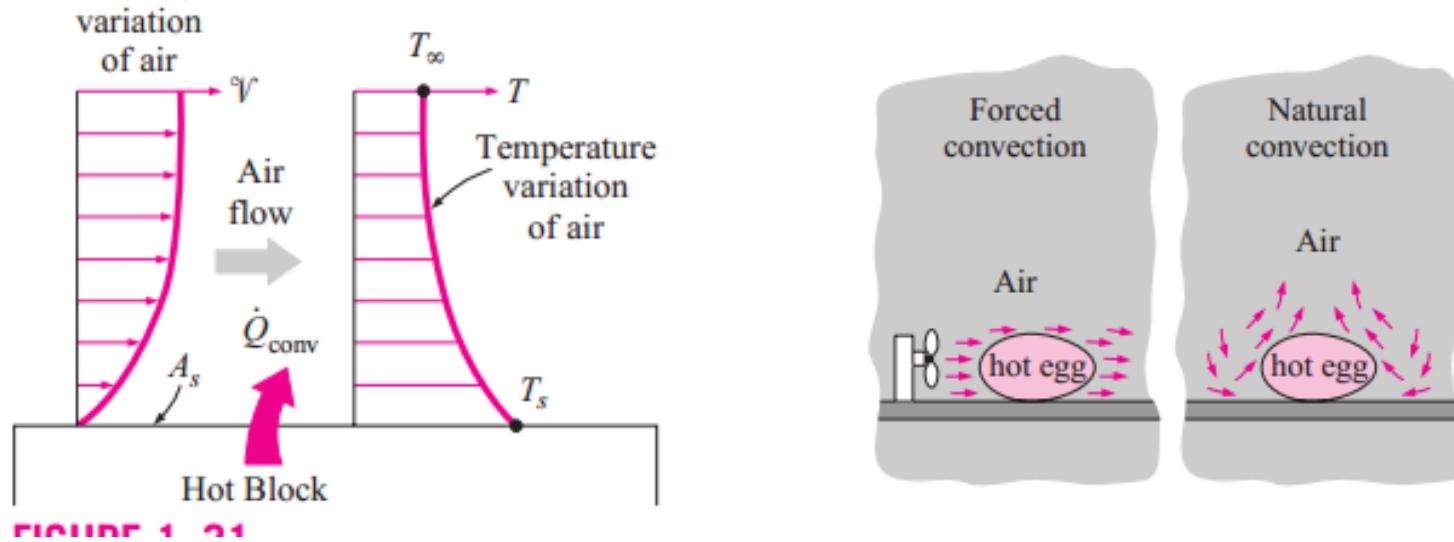
Constant properties can be used for the roof





The range of thermal conductivity of various materials at room temperature.

Convection is the mode of energy transfer between a solid surface and the adjacent liquid or gas that is in motion, and it involves the combined effects of *conduction* and *fluid motion*. The faster the fluid motion, the greater the convection heat transfer. In the absence of any bulk fluid motion, heat transfer between a solid surface and the adjacent fluid is by pure conduction. The presence of bulk motion of the fluid enhances the heat transfer between the solid surface and the fluid, but it also complicates the determination of heat transfer rates.

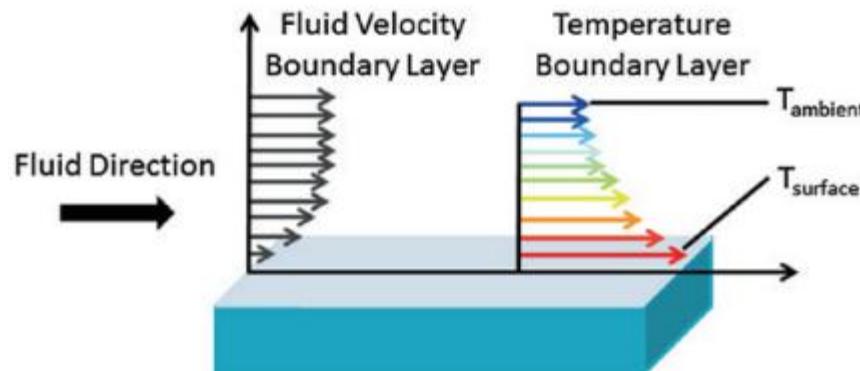


Convection: Convection is the heat transfer from a surface to a fluid. Some common fluids include air and water. Other fluids such as alcohol and oil can also be mentioned in cooling electronics. The rate of convection heat transfer is determined by Eq. 2.3 below

$$Q_{convection} = h_c A (T_{ambient} - T_{surface}) \quad (2.3)$$

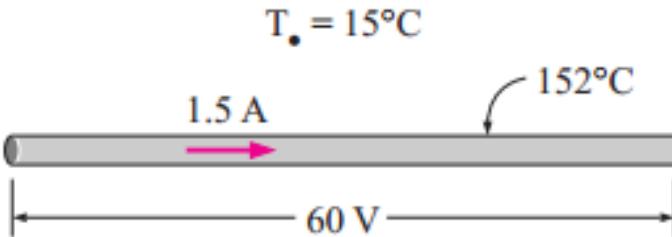
where Q convection is heat flow, h_c is the convection heat transfer coefficient, A is the surface area, T ambient is the ambient temperature of the fluid, and $T_{surface}$ is the temperature of the surface of the material. Figure 2.5 depicts the airflow profile and temperature profile you would expect through convection from the surface of the rectangular object. T surface is warmer than T ambient, which corresponds to the lower airflow velocity at the surface. The convection resistance is described by Eq. 2.4:

$$R_{convection} = \frac{1}{h_c A} \quad (2.4)$$



A 2-m-long, 0.3-cm-diameter electrical wire extends across a room at 15°C. Heat is generated in the wire as a result of resistance heating, and the surface temperature of the wire is measured to be 152°C in steady operation. Also, the voltage drop and electric current through the wire are measured to be 60 V and 1.5 A, respectively.

Disregarding any heat transfer by radiation, determine the convection heat transfer coefficient for heat transfer between the outer surface of the wire and the air in the room.



Radiation is the energy emitted by matter in the form of electromagnetic waves (or photons) as a result of the changes in the electronic configurations of the atoms or molecules. Unlike conduction and convection, the transfer of energy by radiation does not require the presence of an intervening medium.

In fact, energy transfer by radiation is fastest (at the speed of light) and it suffers no attenuation in a vacuum. This is how the energy of the sun reaches the earth.

In heat transfer studies we are interested in thermal radiation, which is the form of radiation emitted by bodies because of their temperature. It differs from other forms of electromagnetic radiation such as x-rays, gamma rays, microwaves, radio waves, and television waves that are not related to temperature.

All bodies at a temperature above absolute zero emit thermal radiation. Radiation is a volumetric phenomenon, and all solids, liquids, and gases emit, absorb, or transmit radiation to varying degrees.

However, radiation is usually considered to be a surface phenomenon for solids that are opaque to thermal radiation such as metals, wood, and rocks since the radiation emitted by the interior regions of such material can never reach the surface, and the radiation incident on such bodies is usually absorbed within a few microns from the surface.

The maximum rate of radiation that can be emitted from a surface at an absolute temperature T_s (in K or R) is given by the **Stefan–Boltzmann law** as

$$\dot{Q}_{\text{emit, max}} = \sigma A_s T_s^4 \quad (\text{W}) \quad (1-25)$$

where $\sigma = 5.67 \times 10^{-8} \text{ W/m}^2 \cdot \text{K}^4$ or $0.1714 \times 10^{-8} \text{ Btu/h} \cdot \text{ft}^2 \cdot \text{R}^4$ is the *Stefan–Boltzmann constant*. The idealized surface that emits radiation at this maximum rate is called a **blackbody**, and the radiation emitted by a blackbody is called **blackbody radiation** (Fig. 1–34). The radiation emitted by all real surfaces is less than the radiation emitted by a blackbody at the same temperature, and is expressed as

$$\dot{Q}_{\text{emit}} = \varepsilon \sigma A_s T_s^4 \quad (\text{W}) \quad (1-26)$$

where ε is the **emissivity** of the surface. The property emissivity, whose value is in the range $0 \leq \varepsilon \leq 1$, is a measure of how closely a surface approximates a blackbody for which $\varepsilon = 1$. The emissivities of some surfaces are given in Table 1–6.

Emissivities of some materials at 300 K

Material	Emissivity
Aluminum foil	0.07
Anodized aluminum	0.82
Polished copper	0.03
Polished gold	0.03
Polished silver	0.02
Polished stainless steel	0.17
Black paint	0.98
White paint	0.90
White paper	0.92–0.97
Asphalt pavement	0.85–0.93
Red brick	0.93–0.96
Human skin	0.95
Wood	0.82–0.92
Soil	0.93–0.96
Water	0.96
Vegetation	0.92–0.96

When a surface of emissivity ε and surface area A_s at an *absolute temperature* T_s is *completely enclosed* by a much larger (or black) surface at absolute temperature T_{surr} separated by a gas (such as air) that does not intervene with radiation, the net rate of radiation heat transfer between these two surfaces is given by (Fig. 1–36)

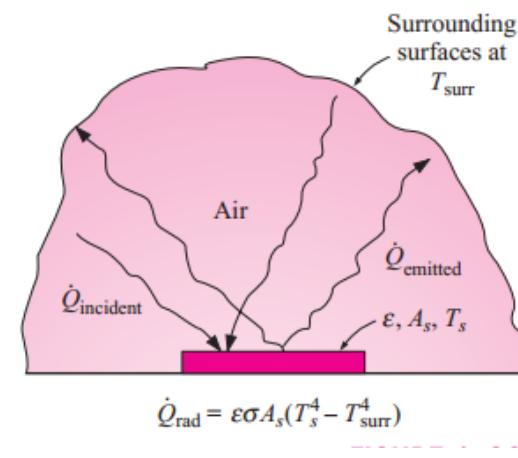
$$\dot{Q}_{\text{rad}} = \varepsilon \sigma A_s (T_s^4 - T_{\text{surr}}^4) \quad (\text{W}) \quad (1-28)$$

In this special case, the emissivity and the surface area of the surrounding surface do not have any effect on the net radiation heat transfer.

Radiation heat transfer to or from a surface surrounded by a gas such as air occurs *parallel* to conduction (or convection, if there is bulk gas motion) between the surface and the gas. Thus the total heat transfer is determined by *adding* the contributions of both heat transfer mechanisms. For simplicity and convenience, this is often done by defining a **combined heat transfer coefficient** h_{combined} that includes the effects of both convection and radiation. Then the *total* heat transfer rate to or from a surface by convection and radiation is expressed as

$$\dot{Q}_{\text{total}} = h_{\text{combined}} A_s (T_s - T_{\infty}) \quad (\text{W}) \quad (1-29)$$

Note that the combined heat transfer coefficient is essentially a convection heat transfer coefficient modified to include the effects of radiation.



1. Consider a person standing in a room maintained at 22°C at all times. The inner surfaces of the walls, floors, and the ceiling of the house are observed to be at an average temperature of 10°C in winter and 25°C in summer.

Determine the rate of radiation heat transfer between this person and the surrounding surfaces if the exposed surface area and the average outer surface temperature of the person are 1.4 m² and 30°C, respectively

Assumptions 1 Steady operating conditions exist.

2 Heat transfer by convection
is not considered.

3 The person is completely surrounded by the interior surfaces of the room.
4 The surrounding surfaces are at a uniform temperature.

The emissivity of a person is 0.95, σ (5.67×10^{-8} W m⁻² K⁻⁴) is the Stefan-Boltzmann constant

2. A rectangular block has an area of 400 mm²

. An engineer would like to use it to cool his heat source that is producing a total of 50 W with a specification of 90 °C. The heat source is placed in a chamber with an air temperature of 65 °C. If the through-length is 10 mm, what is the minimum conductivity of the material in order to cool the heat source to specification using the block alone?

Thermal Structures



In practical sensor structures, thermal effects are induced in the sensor by physical effects that interact with it. The sensitivity and accuracy of the sensors have to be as large as possible, while the influence of other physical effects, such as heat ‘leakage’ along connections and suspensions, and self-heating effects have to be minimized.

Usually it is possible to design structures in such a way that only a few well-known parameters dominate their behavior. When certain parameters are not well known, their influence ought to be negligible. Therefore, a good thermal structure design is simple and can be described by simple models.

In thermal sensors, the thermal signal is the temperature difference induced in the sensor by a physical effect. Optimizing the structure is aimed at optimizing the conversion of the power P to the temperature difference ΔT , the thermal resistance : $R_{th} = \Delta T/P$. This has led to the widespread use of very thin membranes in which the thermal resistance between the physical interaction area and the ambient is maximized..

When designing and optimizing a thermal structure, the choice of the thermal sensing element is an important factor, because its presence influences the thermal resistances and capacitances. For example, with transistors, diodes, or resistors in micromachined structures, connection leads can be made as thin and long as desired so that they have little influence on the thermal resistance between ambient and sensitive areas.

On the other hand, a thermopile often forms both the sensing element and the thermal connection between the hot and cold areas of the sensor, and its design directly influences the thermal resistance between the ambient and the sensitive area.

An important design aspect of a sensor structure is the physical transduction process on which the sensor is based. Some sensors require large interaction areas, e.g., infrared sensors, while others hardly need any interaction area at all.

Electrical-thermal analogies

The behavior of the heat flow and temperature in thermal systems is mathematically described by the same differential equations as electrical currents and voltages in electrical systems. This allows describing thermal systems by means of electrical equivalents, which is convenient because of the many excellent tools available for electrical circuit analysis and the familiarity of solving electrical-network problems

Some examples of thermal resistances

Axial heat flow in bodies having an arbitrary but uniform cross-section

Let us consider the stationary one-dimensional heat flow along the axis of a body of length L and uniform cross-sectional area A (see Figure 6.1), which is exposed to a temperature difference $\Delta T = T_1 - T_2$ between its ends.

In the steady state, the temperature distribution is given by a constant temperature gradient equal to $\Delta T/L$, and the total heat flow through the rod is $\kappa A \Delta T/L$. Hence, for this

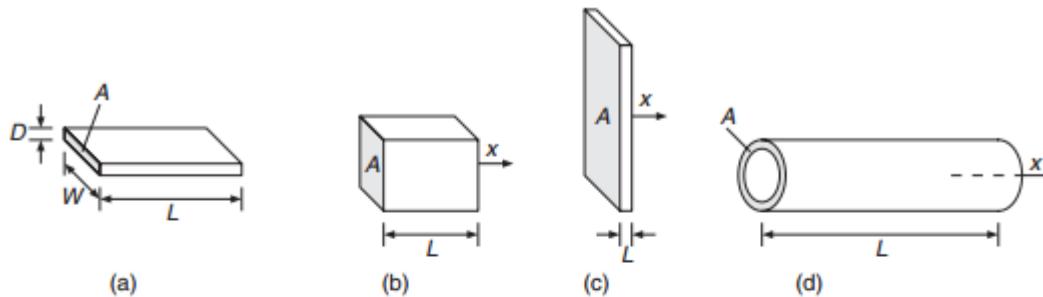


Figure 6.1 Structures with one-dimensional heat flow

configuration, we find a thermal resistance R_{th} :

$$R_{\text{th}} = \frac{L}{\kappa A} \quad (\text{structure with length } L \text{ and uniform cross section } A) \quad (6.13)$$

This expression can be used for the one-dimensional heat flow in all kinds of structures with a uniform cross-section, such as the rods, plates or wires (Figure 6.1).

For example, the thermal resistance of a flat plate with a rectangular cross-section of width W and thickness D (Figure 6.1a) is equal to

$$R_{\text{th}} = \frac{L}{W \kappa D} \quad (\text{sheet with length } L, \text{ width } W \text{ and thickness } D) \quad (6.14)$$

For a square plate, where the length L is equal to the width W , the thermal resistance is equal to $(\kappa D)^{-1}$, which is called the thermal sheet resistance R_{st} of the plate.

Radial heat flow in circular bodies

Multiple thermal resistances can be combined in series or in parallel, and the resulting resistance is calculated just as in the electrical case. This method can be used to calculate complex geometries. If the geometry is such that the heat flow is one-dimensional or if a cylindrical or spherical symmetry makes it possible to assign surfaces of equal temperatures, the effective thermal resistance follows from the integration between the beginning and the end of the structure.

For instance, the thermal resistance for radial heat flow in circular bodies (see Figure 6.2) with height D can easily be found by calculating the thermal resistance of a tube-shaped cylindrical element with height D , thickness dr and radius r , which is $dR = dr/(\kappa 2\pi r D)$. The thermal resistance of a structure with inner radius r_{inn} and outer radius r_{out} can then be obtained by means of integration:

$$R_{\text{th,rad}} = \frac{1}{2\pi\kappa D} \ln \frac{r_{\text{out}}}{r_{\text{inn}}} = \frac{1}{2\pi\kappa D} \ln \frac{d_{\text{out}}}{d_{\text{inn}}} \quad (\text{radial heat flow in tubes}) \quad (6.15)$$

where d_{out} and d_{inn} are the corresponding diameters of the tube and $1/\kappa D$ is the thermal sheet resistance of the body (tube or membrane).

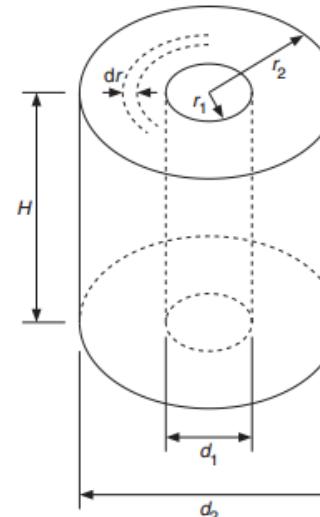


Figure 6.2 Radial heat flow

Radial heat flow in spheres

The thermal resistance for a radial heat flow (see Figure 6.3) of a spherical shell element with thickness dr and radius r is $dR = dr/(4\pi\kappa r^2)$. Integration between the limits r_1 and r_2 yields the thermal resistance between concentric spherical surfaces:

$$R_{\text{th,sphere}} = \frac{1}{4\pi\kappa} \left(\frac{1}{r_1} - \frac{1}{r_2} \right) \text{(radial heat flow in spheres)} \quad (6.16)$$

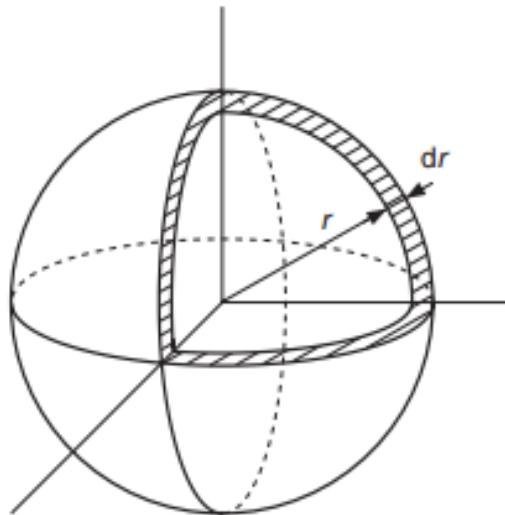


Figure 6.3 Spherical heat flow

Floating Membranes

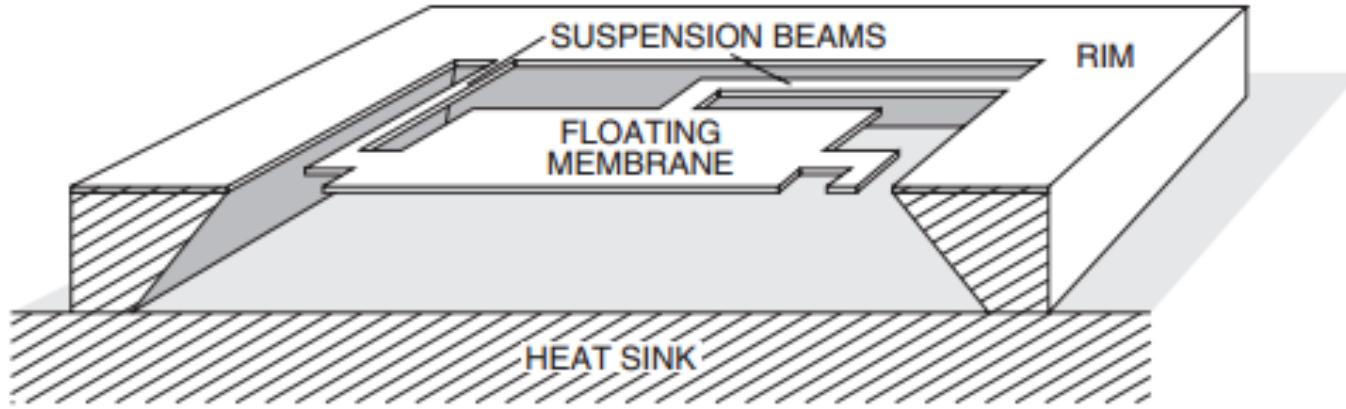
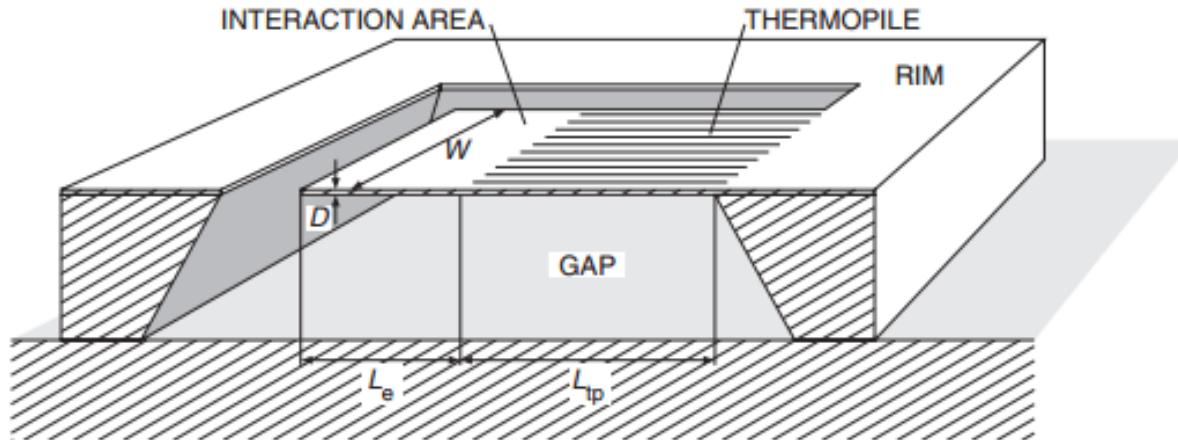


Figure 6.4 Floating-membrane structure

The simplest membrane structure in terms of a thermal model is the floating membrane. In silicon, a floating membrane is made by etching a (closed) membrane in the silicon wafer.

Cantilever beam structure



The cantilever beam is a rectangular beam etched out of a thin membrane, attached on one side to the rim of a silicon chip and free on the other planes. This structure is characterized by a high thermal resistance between the tip of the beam and the base where it is attached to the rim. Heat dissipated at the tip of the beam will flow through the silicon to the rim, creating a temperature difference in the beam. In addition, heat may be lost to the ambient through emission of infrared radiation and conduction and convection if gases are present.

In vacuum, there is no heat loss from the cantilever beam through gas conduction. Supposing that also infrared radiation is negligible, then the thermal resistance R_{th} of the cantilever beam is given by the thermal sheet resistance R_{st} ($= 1/(\kappa D)$) times the length-to-width ratio L/W of that part of the cantilever beam across which the temperature difference is being measured, for instance, for a thermopile of length L_{tp} :

$$R_{th} = R_{st} \frac{L_{tp}}{W} \quad (6.19)$$

Thermal-sensing elements



What Is a Thermal Sensor?

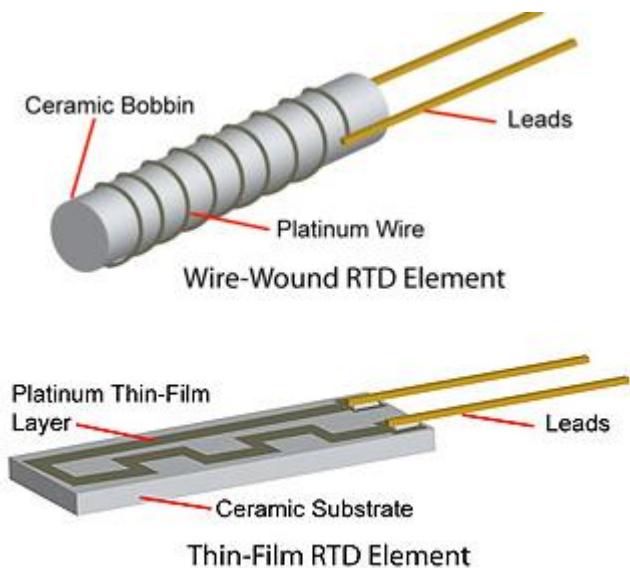
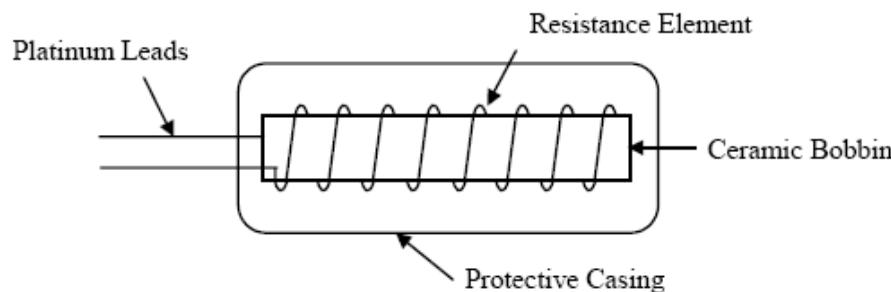
- Temperature is the measure of the average kinetic energy of the molecules of a gas, liquid, or solid.
- A thermal sensor is a device that is specifically used to measure temperature.

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.

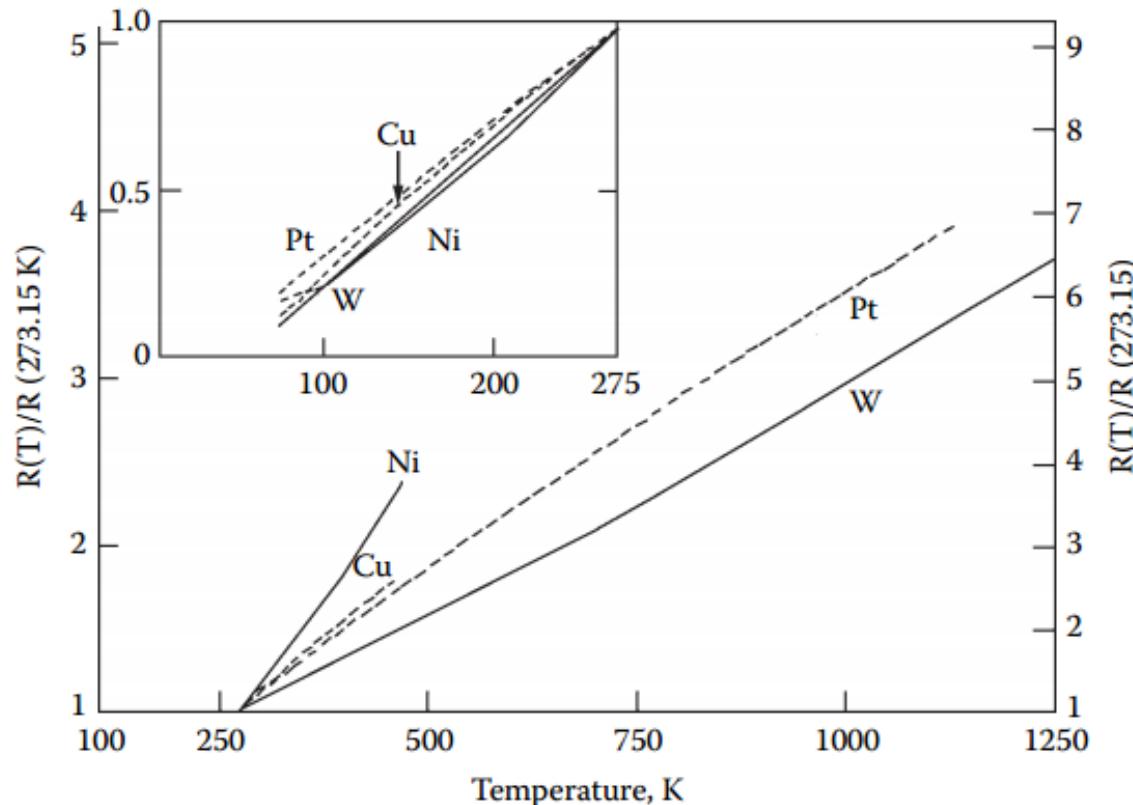
Resistance Thermometers

Resistance thermometers are also known as **resistance temperature detectors, or RTDs**. They are typically made of a single pure metal . Each metal has a material property of electrical resistance that is a function of temperature. The most accurate resistance thermometers are ones that use metals that have a very linear relationship with temperature, **such as platinum**. By using the relationship curves between electrical resistance and temperature, when the resistance of the metal is measured, a temperature can be calculated . Figure depicts the construction of one type of resistance thermometer



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

Experimental measurements clearly show that the resistance thermometers that use metals display an increase in resistance as temperature increases.



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.

Electrical conductivity of a metal sample may be expressed as follows:

$$\sigma = n e \mu$$

where n = electron density, e = electronic charge, and μ = electron mobility.

The electron mobility is defined as follows:

$$\mu = \frac{e\tau}{m} \quad (5.3)$$

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:

$$\rho = \frac{1}{\sigma} = \frac{m}{ne^2} \left(\frac{1}{\tau} \right) \quad (5.4)$$

The most common material used in metal thermometers is platinum. PRTs typically have a resistance range between .1 Ω and 1 KΩ.

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.

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

The conductivity is given by the following equation

$$\sigma = e(n\mu_n + p\mu_p)$$

where μ_n and μ_p are the mobilities of the electrons and holes, respectively.

$$\sigma = e^2 \left(\frac{n\tau_n}{m_n} + \frac{p\tau_p}{m_p} \right)$$

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

$$n = p = n_i \quad (5.10)$$

where n_i = intrinsic carrier density.

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.

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.

•

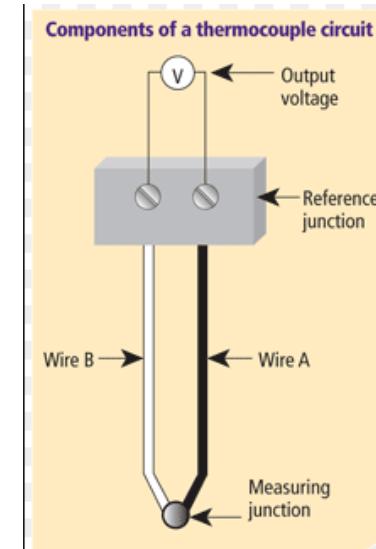
Thermistors typically have resistance between $10\ \Omega$ and $100\ M\Omega$. Due to the large resistance range, thermistors provide excellent sensitivity and are also very stable, particularly between 100 and 300°C . Thermistors can also be manufactured very cheaply

For applications greater than $1,000^{\circ}\text{C}$ thermistors made with zirconium oxide are the most popular. Zirconium is unique in its ability to withstand high temperatures.

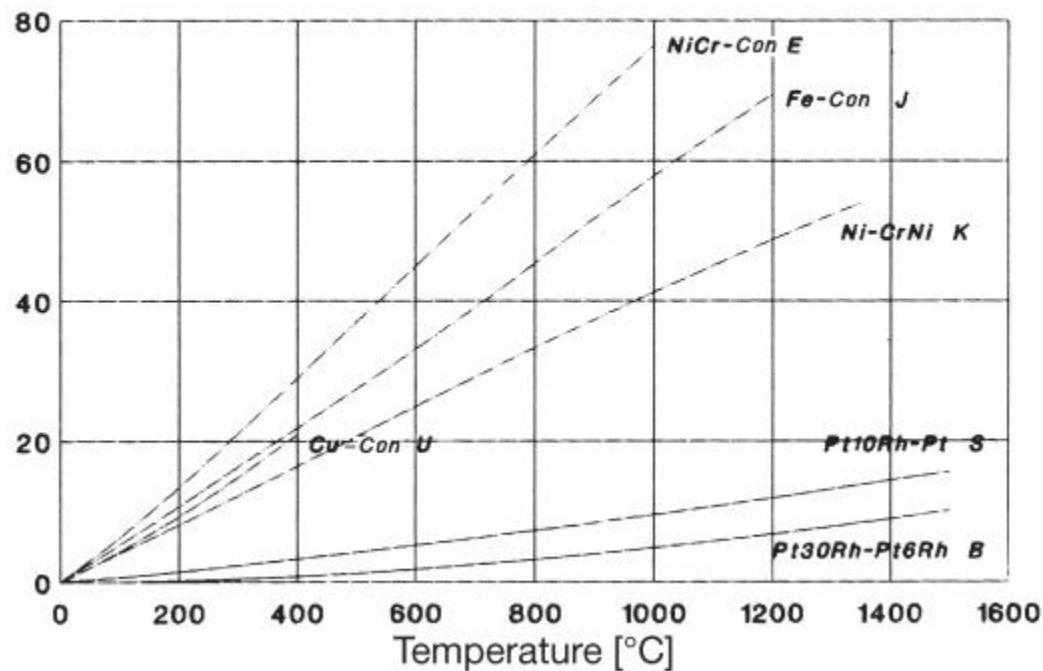
Temperature-Difference Sensing Elements

Thermocouples

Thermocouples are sensors composed of two different metals at their sensing end. A voltage is created when there is a temperature gradient between the hot sensor element and the cold reference junction. The change in voltage can be reported as a temperature through the Seebeck effect . **The Seebeck effect says that the change in voltage is linearly proportional to the change in temperature and the two variables are related to each other through a coefficient that is determined by the materials used in the thermocouple .** Figure depicts the construction of a thermocouple.



Voltage/mV

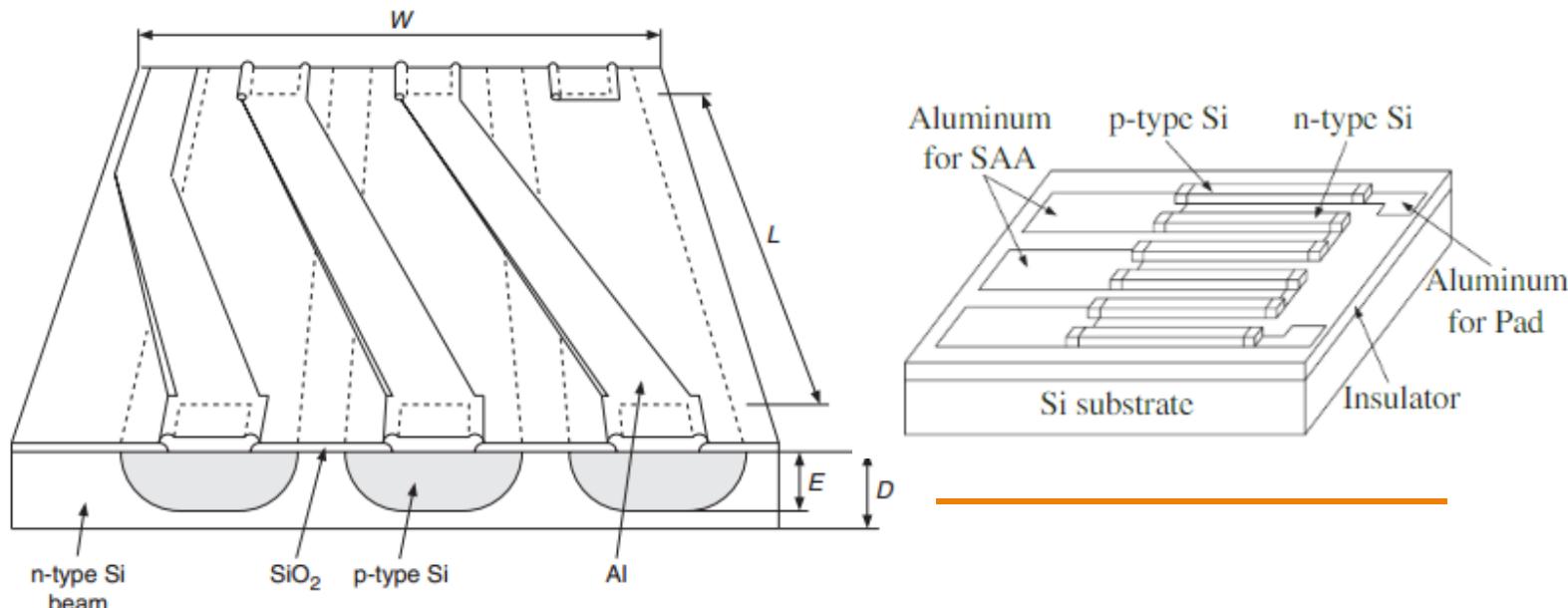


Characteristics of thermocouples

Thermocouples are two-lead elements that measure the temperature difference between the ends of the wires. The operating principle is based on the thermoelectric Seebeck effect, which says that a temperature difference ΔT in a (semi)conductor also creates an electrical voltage ΔV :

$$\Delta V = \alpha_s \Delta T \quad (6.29)$$

where α_s is the Seebeck coefficient expressed in V/K. The Seebeck coefficient α_s is a material constant. By taking two wires of materials with different α_s , we get different electrical voltages across the wires, even when the wires experience the same temperature gradients. With a junction of the wires at the hot point, the voltages are subtracted, and an effective Seebeck coefficient will remain. Thermocouples or thermopiles (several thermocouples in series) in thermal sensors are made of thin-film metals or polysilicon, or monocrystalline silicon



In practice, the Seebeck coefficient α_{mono} for monocrystalline silicon is related to the electrical resistivity ρ . At room temperature, this relation can be expressed as:

$$\alpha_{\text{mono}} = \frac{mk}{q} \ln \frac{\rho}{\rho_0} \quad (6.30)$$

with $\rho_0 \approx 5 \times 10^{-6} \Omega \text{ m}$ and $m \approx 2^{1/2}$ as constants [5] and k the Boltzmann constant, $k/q \approx 86.3 \mu\text{V/K}$. For practical doping concentrations, the Seebeck coefficients are of the order of 0.3 mV/K to 0.6 mV/K, where the sheet resistance depends on the layer depth.

For polycrystalline silicon, a similar expression is given by Von Arx [6] as a function of electrical resistivity:

$$\alpha_{\text{poly}} = \frac{m_{\text{poly}}k}{q} \ln \frac{\rho}{\rho_0} \quad (6.31)$$

with $\rho_0 \approx 1.4 \times 10^{-6} \Omega \text{ m}$ and $m_{\text{poly}} \approx 0.7$ as constants, and k the Boltzmann constant. In practice, the Seebeck coefficients are of the order of 0.1 mV/K to 0.2 mV/K, for instance at sheet resistances of $50 \Omega/\square$ to $100 \Omega/\square$ and a poly thickness of 300 nm.

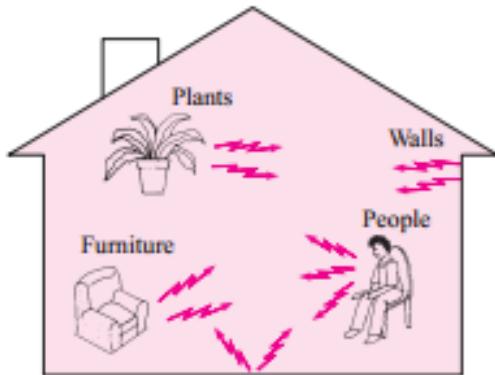
Seebeck coefficients for metals are much smaller than those for silicon and that the influence of aluminum interconnections on chips is negligible compared to the Seebeck coefficient for silicon. The electrical resistance and also the thermal conductivity play a part in determining how efficiently a thermopile functions in a thermal sensor. These parameters are much more favorable for bismuth telluride compounds or silicon–germanium compounds than for mono- or polysilicon [8]. However, the advantage of these compounds largely lies in their low thermal conductivity, compared with that of silicon, and in many microsensors the thermal resistance of the sensors is determined more by conduction through air or membranes than through the thermopile. In these, a silicon thermopile will lead to almost the same performance as thermopiles made of other compounds but has the big advantage that it can be produced in standard IC technology.

Table 6.3 Seebeck coefficients ($\mu\text{V/K}$) of some selected materials and standard thermocouples

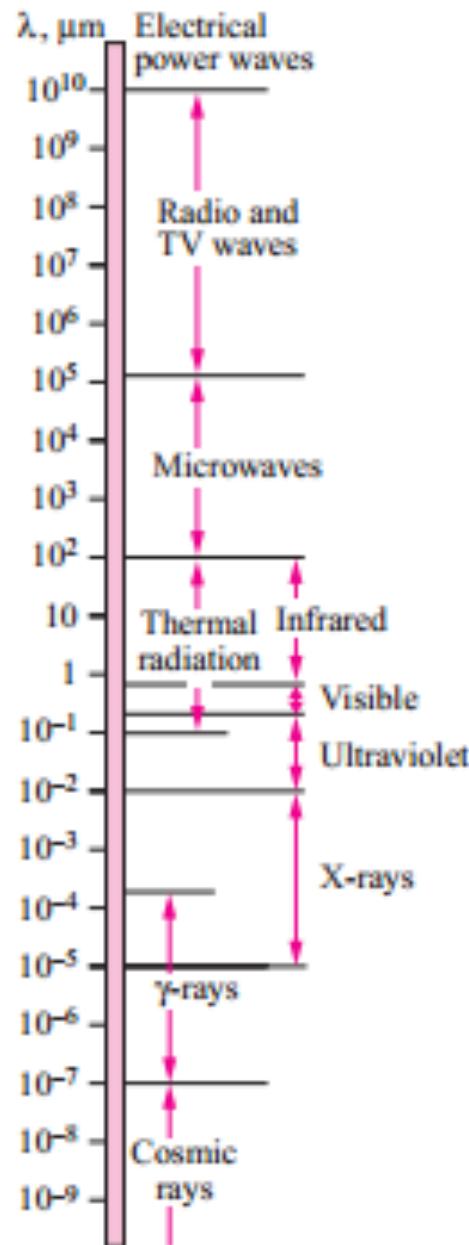
Material	273 K ($\mu\text{V/K}$)	300 K ($\mu\text{V/K}$)
Antimony (Sb)		43 ^a
Chromium (Cr)	18.8	17
Gold (Au)	1.79	1.94
Copper (Cu)	1.70	1.83
Silver (Ag)	1.38	1.51
Rhodium (Rh)	0.48	0.40
Lead (Pb)	-0.995	-1.047
Aluminum (Al)		-1.7
Platinum (Pt)	-4.45	-5.28
Nickel (Ni)	-18.0	
Bismuth (Bi)		-79 ^a
J-type thermocouple Fe–CuNi	50	51
T-type thermocouple Cu–CuNi	39	41
E-type thermocouple NiCr–CuNi	59	
K-type thermocouple NiCr–NiAl	39	41
S-type thermocouple Pt–Pt/10%Rh	5	7
T-type thermocouple Pt–Pt/13%Rh	5	6
Monocrystalline silicon (typical)	500	500
Polycrystalline silicon (typical)	150	150

^aAveraged over 0 °C to 100 °C.

The electromagnetic wave spectrum



Everything around us constantly emits thermal radiation



Thermal radiation is also defined as the portion of the electro-magnetic spectrum that extends from about 0.1 to 100 μm , since the radiation emitted by bodies due to their temperature falls almost entirely into this wave-length range. Thus, **thermal radiation includes the entire visible and infrared (IR) radiation as well as a portion of the ultraviolet (UV) radiation**

What we call light is simply the visible portion of the electromagnetic spec-trum that lies between 0.40 and 0.76 μm . Light is characteristically no different than other electromagnetic radiation, except that it happens to trigger the sensation of seeing in the human eye.
Light, or the visible spectrum, consists of narrow bands of color from violet (0.40–0.44 μm) to red (0.63–0.76 μm)

Radiation Thermometers

All substances and objects emit thermal radiation when it is at a temperature higher than absolute zero (0 K or -273.15°C). There is a **relationship between temperature and radiation energy emitted that can be used to calculate the temperature of the object surface**. Unlike other sensors discussed above, radiation thermometers are primarily used at a distance from the object of interest and can be used for hard-to-reach objects. An **example of a radiation thermometer is an infrared camera**, which measures infrared wavelengths that emit from an object.



Infrared Sensor

From the transduction point of view, the infrared sensor is fairly simple. The transduction from radiation to heat is carried out by a black absorber, which can have an efficiency up to 99 %. The first transduction step from incident radiation density P''_{inc} (in W/m²) to thermal power P is

$$P = Q P_{\text{inc}}$$

$$Q = \alpha A_D \tau_{\text{filter}}$$

where A_D is the sensitive area of the sensor (usually the area that is coated black), α is the absorptivity of the black coating of the sensitive area, and τ_{filter} is the transmittance of an infrared filter, which can be applied to select specific radiation wavelengths or simply for mechanical protection. The absorptivity α is between 0 and 1 and denotes the fraction of infrared radiation power which is absorbed by the black coating. Various types of black coatings are used for silicon infrared sensors.

Infrared sensors have many applications. Burglary alarms use infrared sensors of the pyroelectric type, which respond to changes in the infrared image. This is particularly useful in security applications, where an image without any movements is the proper one and a sudden movement indicates an intruder. For gas analysis, such as CO and CO₂, infrared sensors

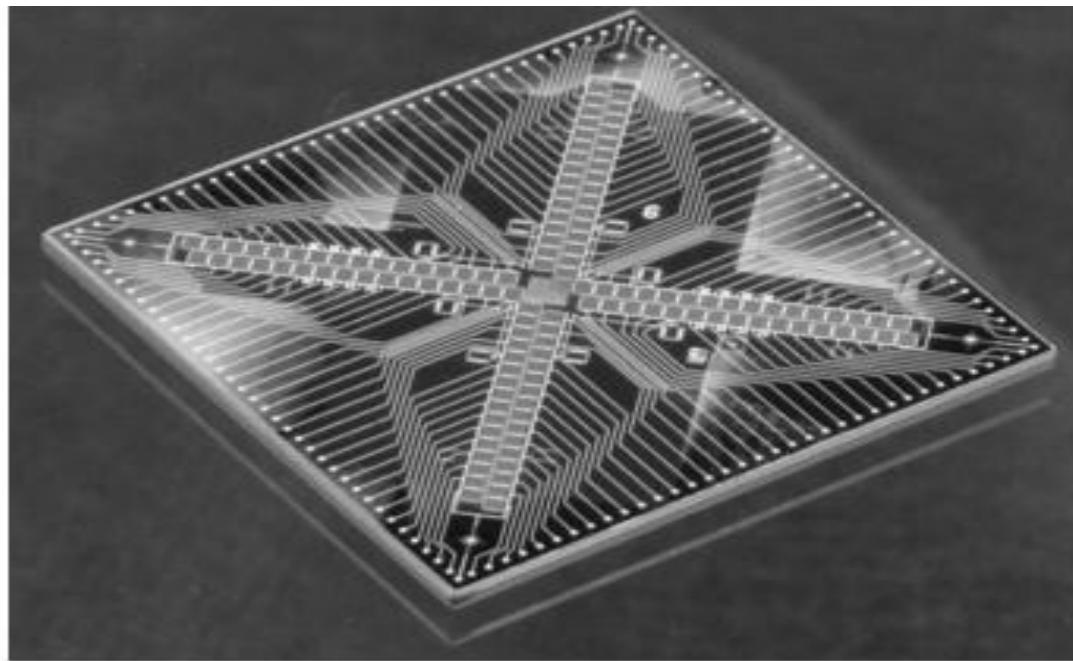


Figure 6.11 Focal plane array of 4×32 infrared detectors for satellite attitude control instrument, chip size $20.5 \text{ mm} \times 20.5 \text{ mm}$ (Xensor Integration)

based on thermopiles and thin membranes are used. Other applications include temperature measurement, where, for instance, the toasting of bread can be monitored using an infrared sensor. For space applications, EADS Sodern developed a focal plane array (FPA) of infrared detectors for a so-called Earth sensor, where the image of the Earth is projected onto the FPA chip using a germanium lens. The Earth sensor uses the image in the (14 to 16) μm band, emitted by the CO₂ of the atmosphere. This image of Earth is nicely round and is dependent neither on day or night nor on the seasons. In this way, the attitude of the satellite with respect to the Earth can be measured, and if necessary, corrected; as a result, for instance, a weather satellite will monitor the weather in the required region, instead of that of the North Pole or space. The FPA chip contains 132 infrared pixels (each with its own thermopile and black area) and measures $20.5 \text{ mm} \times 20.5 \text{ mm}$ (Figure 6.11) [17].

Interaction of gaseous species at semiconductor Surfaces-Catalysis



4.3.4.1 Adsorption/Desorption Mechanism

During the gas molecule–solid-state surface interaction, the gas species initially gets physisorbed and chemisorbed at the semiconductor surface. There is no charge transfer in case of physisorption. A molecule is considered to be chemisorbed when there is an electronic transfer between the gas and the solid, and the conductivity of the material gets affected.

As an illustrative example, the different oxygen species formed as the oxygen molecules interact with a solid-state surface are shown in Figure 4.12 (Kohl 1989, Liu et al. 2007). Oxygen can be adsorbed in the form of O_2^- , O^- , or O^{2-} . Due to energy considerations, near room temperature, the O^{2-} specie dominates the surface coverage, whereas at higher temperatures, O^- is the dominating specie. The adsorption of these species on the surface extracts electrons from the donor levels of the material. The extraction of electrons creates an electron-depleted or positive space charge region near the surface (Figure 4.12). This results in an increase in the resistance of an n-type material.

The adsorbed oxygen ions are present in large concentrations at any semiconductor surface. The gas-sensing mechanism depends on the decrease or increase in the concentration of the adsorbed oxygen ion at the semiconductor surface. It is clear that on interaction with a reducing gas, the number of adsorbed oxygen species will be reduced and a number of electrons will be released back to the semiconductor surface (conduction band of the semiconductor). The O_2^- and O^- species preferentially interact with the C–C bond of the incoming reducing gas. Increase in the concentration of electrons in the conduction band results in an increase in the conductivity of the n-type semiconductor. The large increase in the conductivity of the oxide semiconductor in the presence of a reducing gas is because of a decrease in the adsorbed oxygen species at the oxide surface. It is clear that the effect of an oxidizing gas will be reverse (increase in absorbed oxygen sites and decrease in the conductivity of the semiconductor).

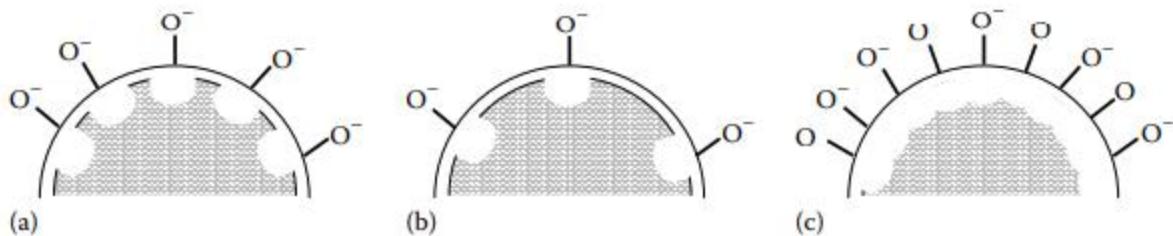
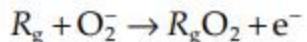
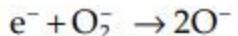
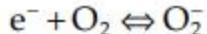


FIGURE 4.13

Effect of (a) air, (b) reducing gas, and (c) oxidizing gas on an n-type semiconductor material. Depletion region is shown as white region, and shadowed region shows the conducting regions of the oxide nanoparticles.

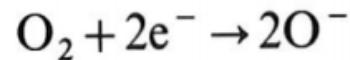
The reaction steps (Kohl 1989) are given below:



Here R_g represents the reducing gas.

If a p-type semiconductor is used in place of an n-type semiconductor, the reducing gas will decrease the adsorbed oxygen sites, take away electrons from the semiconductor surface, and create more holes in the semiconductor. This is also illustrated in Figure 4.13. The conductivity of the p-type semiconductor due to increase in the concentration of holes will thus increase.

- Typical operation principle “one explanation”
 - Resistive sensor, resistance decrease when a layer of powdered SnO_2 is exposed to combustible gas in present of ambient air.
 - Oxygen in air adsorbs at semiconductor surface, oxygen dissociates to form O^- . The electron is picked up from the semiconductor, which increases its resistance.



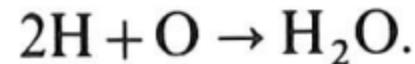
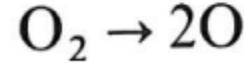
- O^- is highly reactive and easily react with combustible gases which contain hydrogen.



- Hydrogen react and form water, as a result an electron is released which lowers the resistance in the semiconductor.
- To increase the reaction rate a catalyst is often included.

Catalysis, the acceleration of chemical reactions

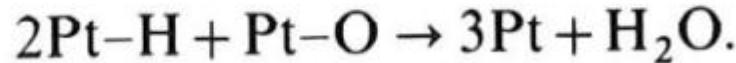
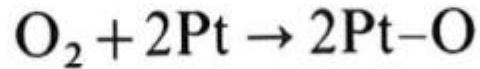
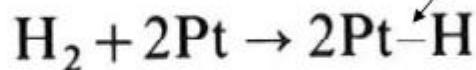
- Catalyst is needed to increase the sensitivity and accelerate the reaction in metal-oxide gas sensor
- A catalyst is not consumed and lower the activation energy for the reactions
- Without the catalyst, the reaction have a high activation energy:



chemical reactions

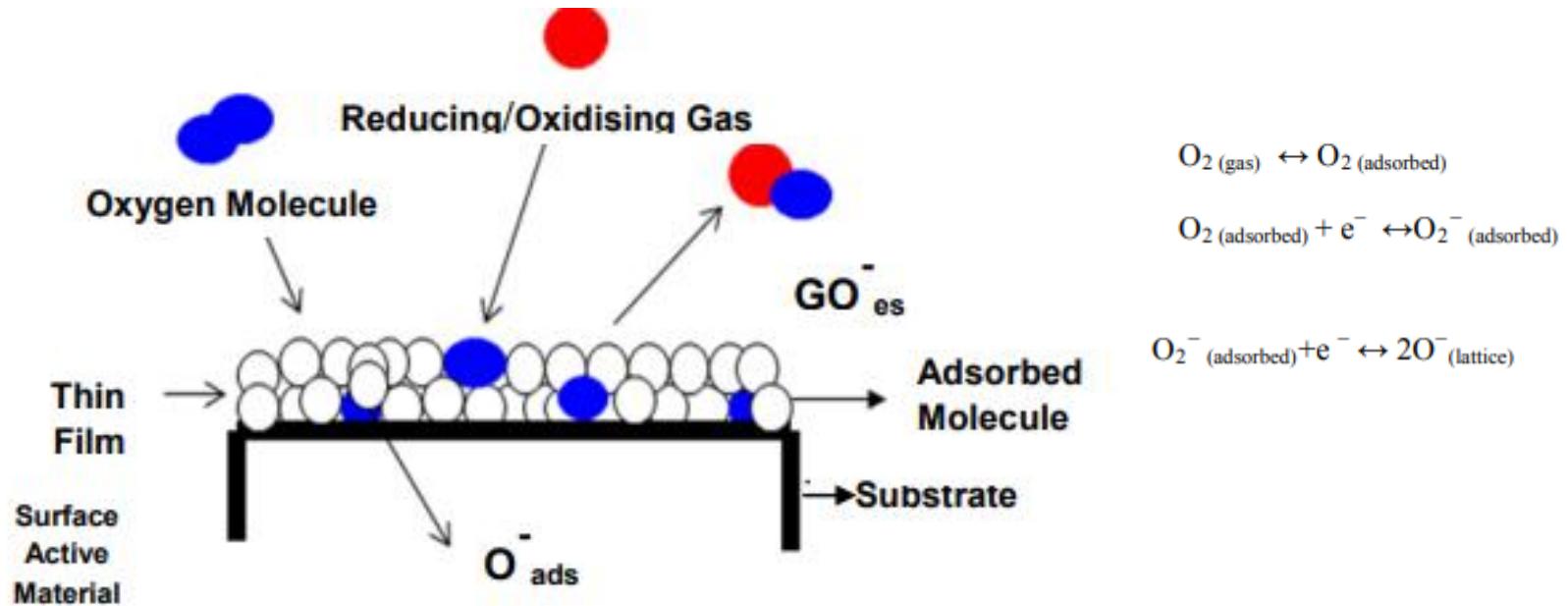
With Platinum as a catalyst

Hydrogen is adsorbed on a platinum surface ("group of")



The activation energy is drastic reduced

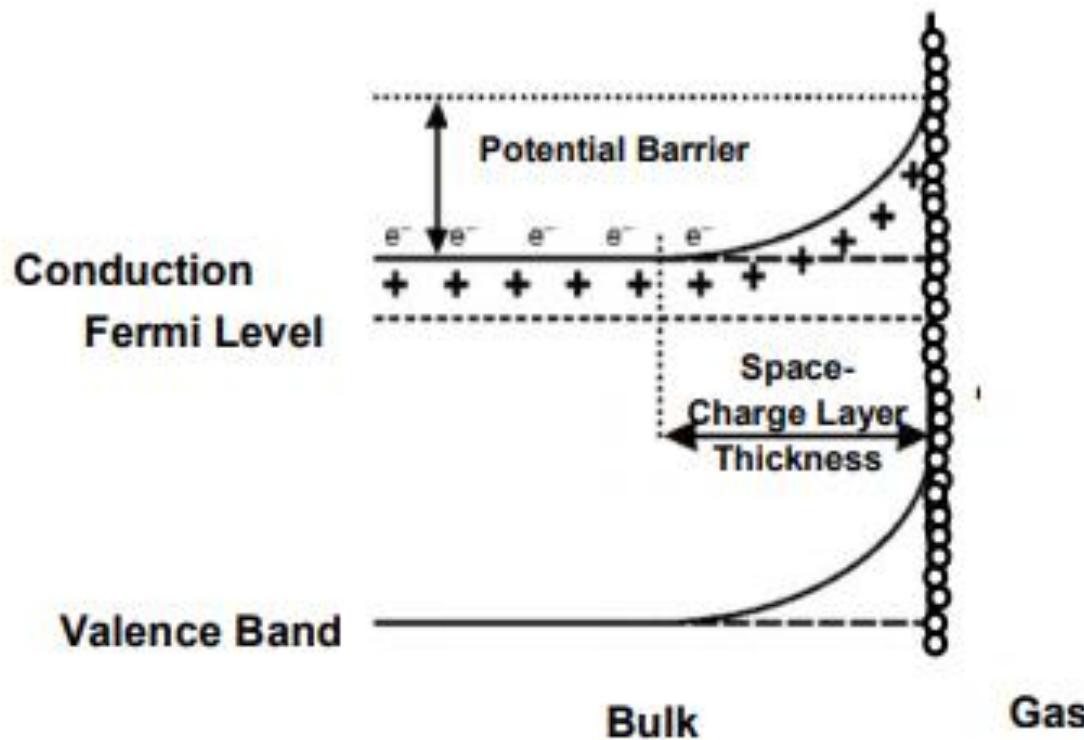
At elevated temperatures, reactive oxygen species such as O_2^- , O^{2-} and O^- are adsorbed on the surface of metal oxide semiconductor.



Schematic representation of metal oxide semiconductor as gas sensors

The sensing element of chemiresistive type sensors normally comprises of a semiconducting material with high surface-to-volume ratio on a ceramic (glass) substrate with ohmic contacts to measure the change in resistance/conductance. When, gas/volatile organic compounds (VOC) samples interact on the surface of metal oxide semiconductor, due to the combustion reaction that occurs with the oxygen species on the surface of metal oxide particles leads to change in resistance and forms the basic principle of detection.

As shown in Fig. when O₂ molecules are adsorbed on the surface of metal oxides, they would extract electrons from the conduction band E_c. This will lead to band bending and an electron depleted region (space charge region). The thickness of this space charge region is equivalent to the length of band bending region



Schematic representation of band bending after chemisorption of charged species

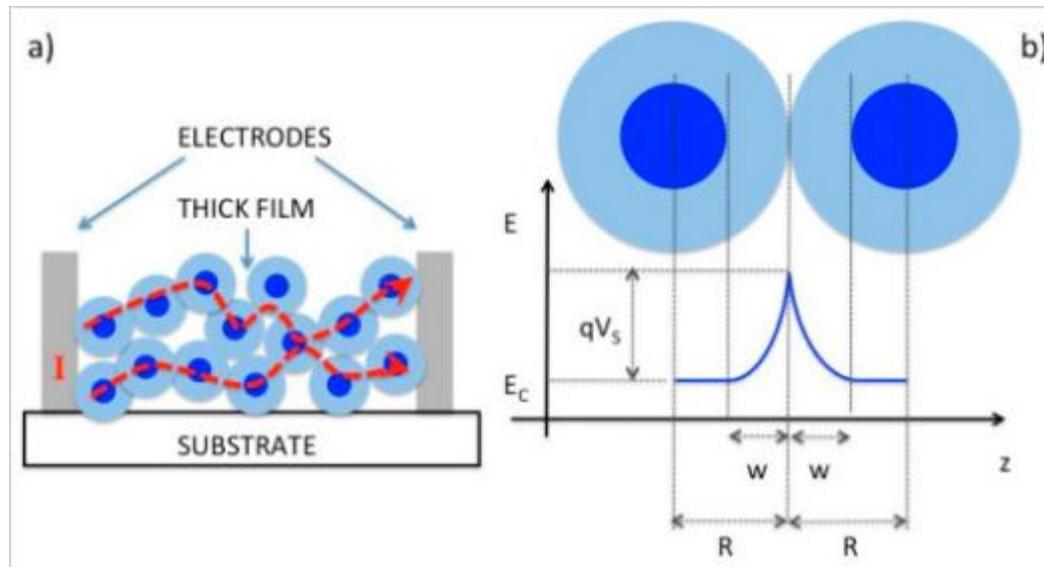
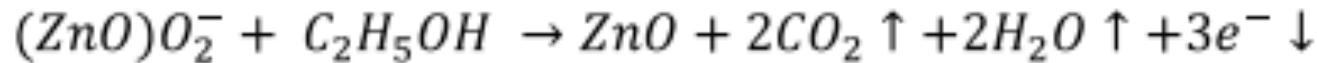
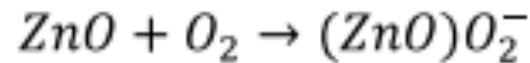
Table 1: Classification according to the changes in the response of sensing element

Classification	Oxidising Gases	Reducing Gases
n-type	Resistance increase	Resistance decrease
p-type	Resistance decrease	Resistance increase

Table 2: Classification of metal oxides based on the conductivity type

Type of Conductivity	Metal oxides
n – type	ZnO, MgO, CaO, TiO ₂ , WO ₃ , SnO ₂ , In ₂ O ₃ , Al ₂ O ₃ , Ga ₂ O ₃ , V ₂ O ₅ , Nb ₂ O ₅ , ZrO ₂
p – type	Y ₂ O ₃ , La ₂ O ₃ , CeO ₂ , Mn ₂ O ₃ , NiO, PdO, Ag ₂ O, Bi ₂ O ₃ , Sb ₂ O ₃ , TeO ₂

When the reducing gas like ethanol interacts with n-type semiconductor, then the surface conductivity increases based on the following reaction:



$$\therefore \frac{R_{\text{Gas}}}{R_{\text{Air}}} = \exp \frac{eV_{\text{Air}} - eV_{\text{Gas}}}{k_B T}$$

Thin-film sensors, FET devices for gas and ion sensing



Conducting polymer composites, intrinsically conducting polymers and metal oxides are three of the most commonly utilised classes of sensing materials in conductivity sensors. These materials work on the principle that a change in some property of the material resulting from interaction with a gas/odour leads to a change in resistance in the sensor.

The mechanisms that lead to these resistance changes are different for each material type; however, the structure and layout of conductivity sensors prepared using these materials are essentially the same.

A schematic of a typical conductivity sensor design is shown in Figure. 1. The sensing material is deposited over interdigitated or two parallel electrodes, which form the electrical connections through which the relative resistance change is measured. The heater is required when metal oxides are used as the sensing material because very high temperatures are required for effective operation of metal oxide sensors.

Typical structure of a conductivity sensor

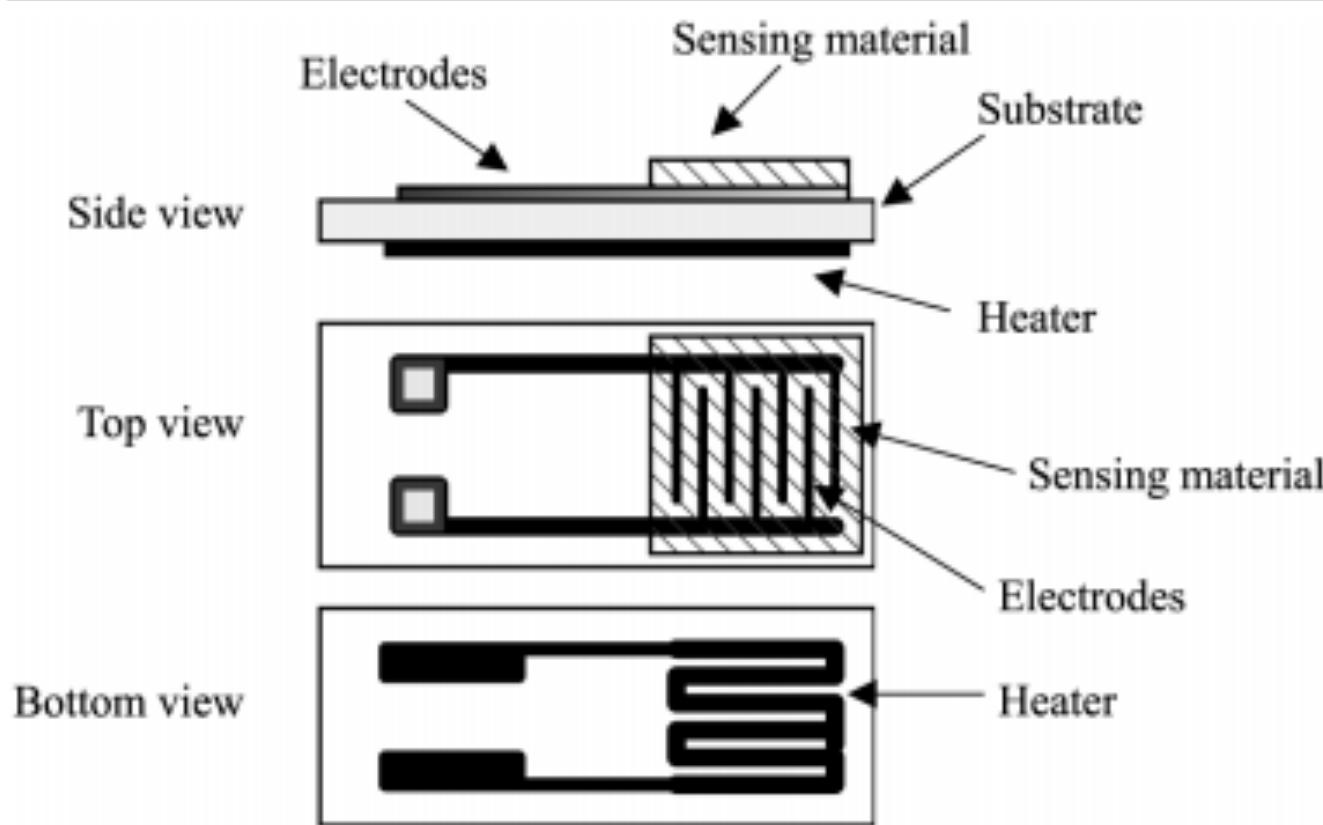
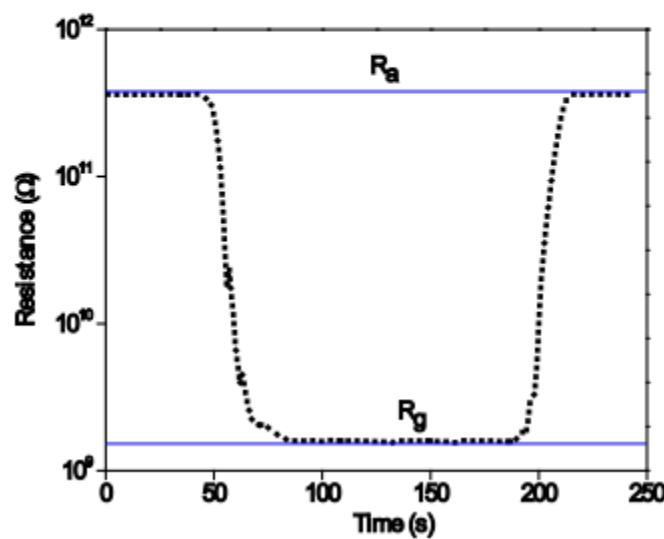


Figure. 1

Sensitivity It is the response of a gas sensor per unit change in the gas concentration. Since, metal oxide gas sensors are based on the principle of chemiresistivity; it is generally defined in terms of conductance or resistance. For n-type material in the presence of reducing gas and p-type material in the presence of oxidizing gas, sensitivity can be defined as,

$S = (R_a - R_g)/R_g$, where, R_a , R_g are the stable values of the resistance of the material before and after exposure to gas.

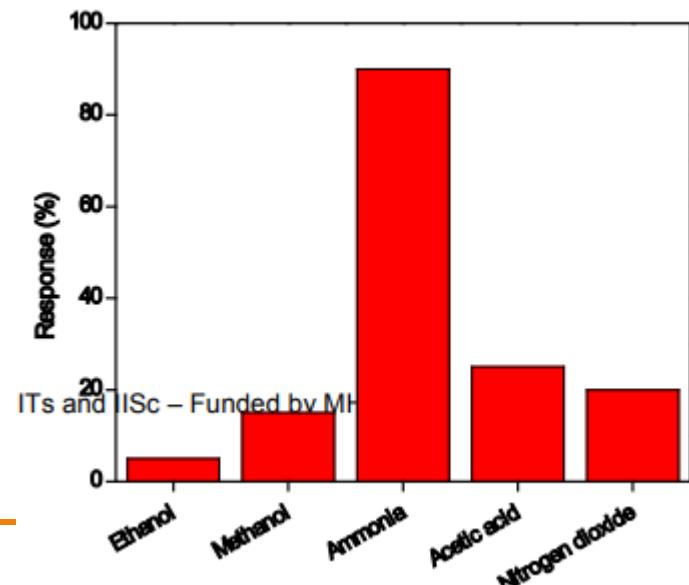


Selectivity

A sensor should respond to only a particular molecule in a mixture of environment. The selectivity of a gas sensor towards a particular molecule is the ratio of its response towards it and that of another dominant interfering molecule in the atmosphere.

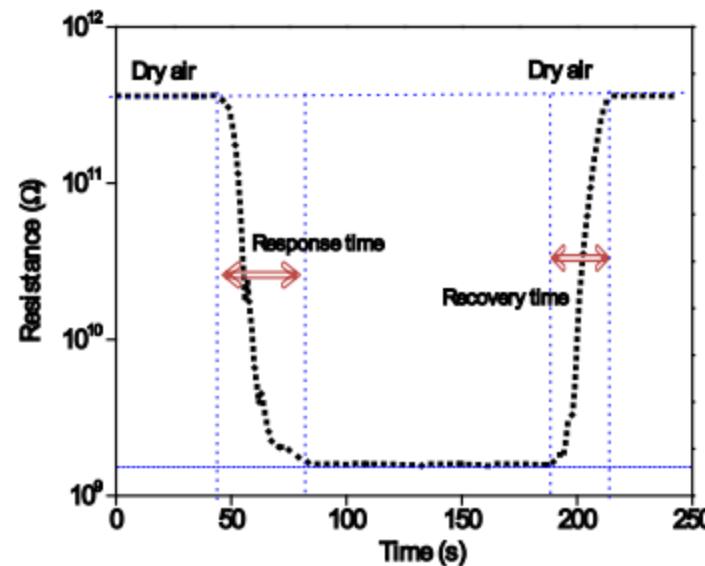
$$\text{Selectivity of an gas sensor} = \frac{\text{sensitivity of a particular molecule}}{\text{sensitivity towards an interferent}}$$

Selectivity of a gas sensor should be always greater than one.



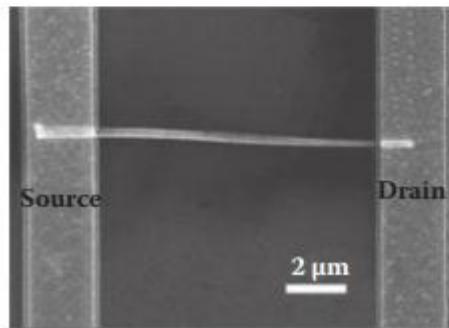
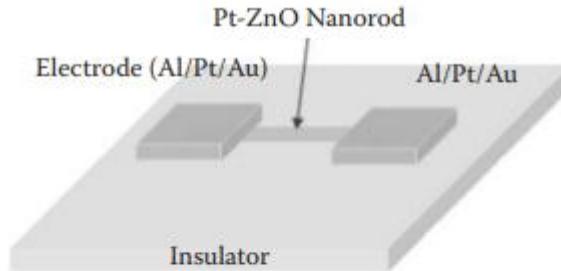
Speed of response

The time required for a sensor to reach 90% of total response of the signal such as resistance upon exposure to the target gas

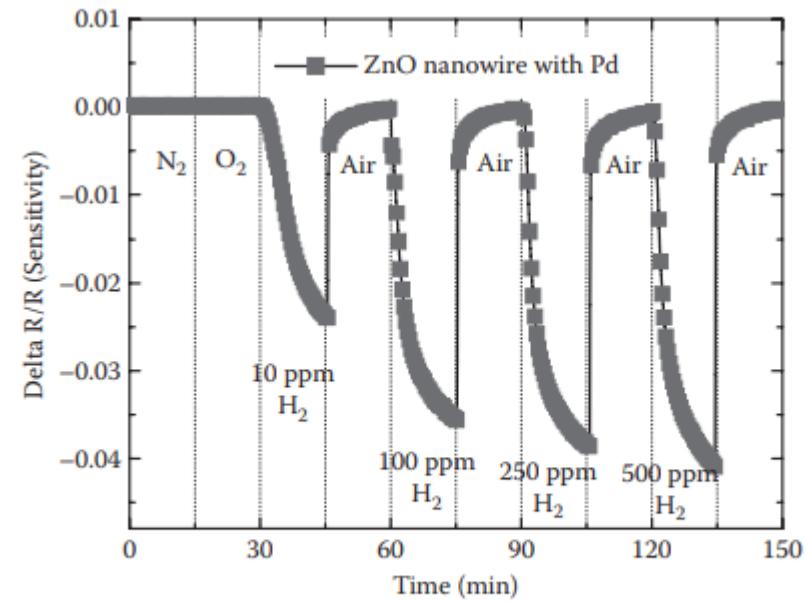


Recovery time

The time required for a sensor to return to 90% of original baseline signal upon removal of target gas



ZnO nanowire sensor (top), SEM of completed device



Change in resistance as a function of time when switching to H_2 -containing ambient.

Gas-sensitive FETs and field-effect devices combined with catalytic metal gates



Gas-sensitive FETs and field-effect devices combined with catalytic metal gates

Catalytic-gate FETs are one of types of gas-sensitive FETs. In 1975, Lundström et al. first reported a Pd-gate FET sensitive to hydrogen. Pioneering research on catalytic-gate FETs opened up the field of FET-based gas sensors and other gas-sensitive field-effect devices such as capacitor-based and Schottky diode-based sensors . Catalytic-gate filed-effect devices feature a nanoscale layer of catalytic metals, such as palladium and platinum, as a gate electrode on insulating layers in a metal-insulator-semiconductor (MIS) structure . Figure 1 shows reported schematic illustrations of this structure and the threshold voltage shift of a Pd-gate FET that is sensitive to hydrogen . In initial reports of catalyticgate FETs, Pd as a catalytic-gate electrode was deposited onto the insulating layer of the MIS structure of the FET .

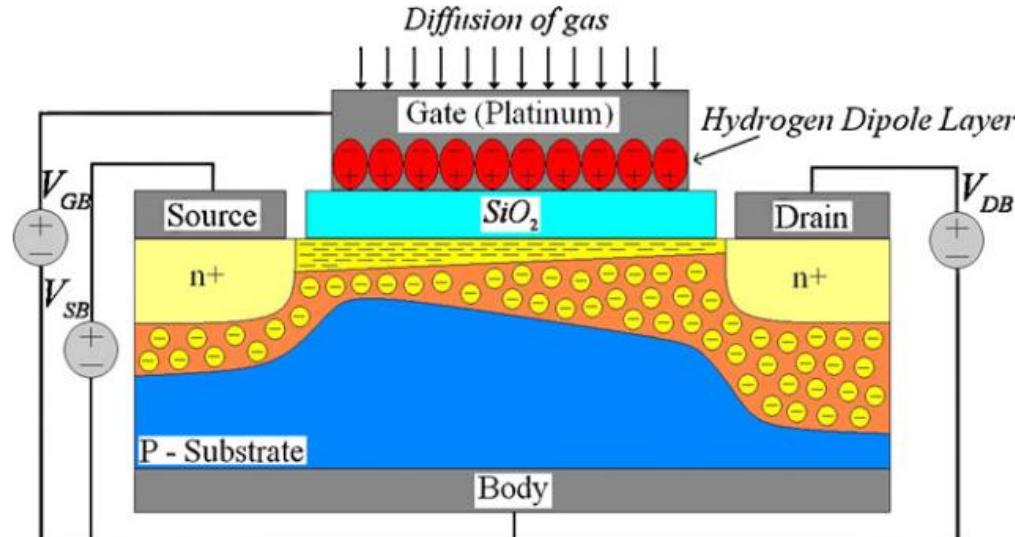
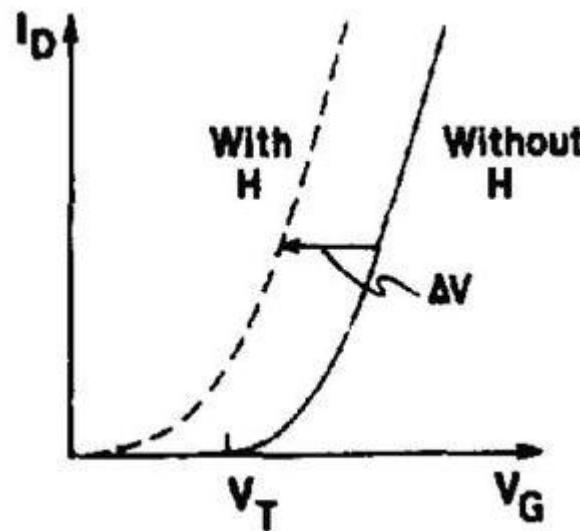


Fig. 1. Structure of a MOSFET gas sensor submitted to hydrogen gas.

Exposing the sensor to hydrogen gas causing formation of the dipole layer hydrogen molecules adsorb at the surface of the platinum and the attractive forces between the platinum and the hydrogen atoms weaken the H–H bond therefore the hydrogen molecules dissociate into the hydrogen atoms.

Some of the atoms landed on the platinum surface diffuse rapidly through the metal and adsorb at the Pt– SiO_2 junction where they become polarized under influence of the electric field and create the dipole layer shown Due to the dipole layer and causes shift in the metal work function:



MOSFET “Linear” Region

$$I_{DS} = \frac{W}{L} \mu_n C_{ox} (V_{GS} - V_{Th}) V_{DS} \quad V_{GS} > V_{Th} \quad V_{DS} \approx 100\text{mV}$$

$$V_{T0} = \phi_{GS} - 2\phi_F - \frac{Q_{d0}}{C_{ox}} - \frac{Q_i}{C_{ox}}$$

↓
 $\Delta W_M = W'_M - W_M$

$$\Delta V_T = \frac{\Delta W_M}{q}$$

Suspended-gate FETs

In 1983, Janata et al. reported an SGFET sensitive to dipolar molecules such as methanol and methylene chloride. In the SGFET shown in Fig.

Figure 12. Scheme of a suspended gate GasFET. The gate electrode is suspended and covered with a gas sensitive layer. The electrical potential generated by gas adsorption acts as an additional gate voltage and changes the source-drain current. Reprinted from [150] with permission from Elsevier.

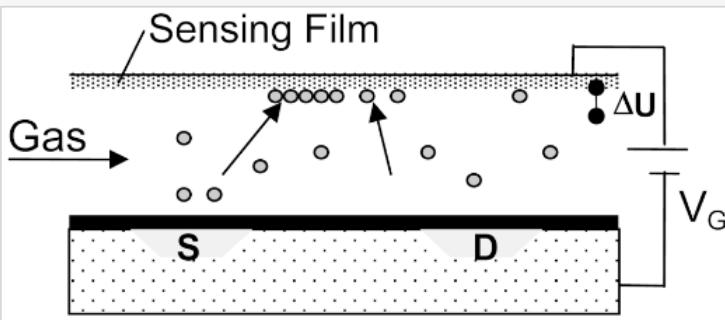
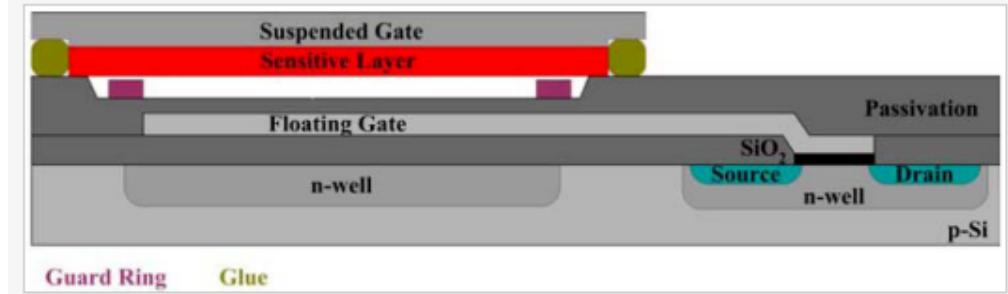


Figure 13. Schematic cross section of a floating gate (FGFET) type transducer that improves the coupling of the work function voltage to the FET. The capacitance well electrode can be additionally used to set the optimal working point in the transistor characteristics.

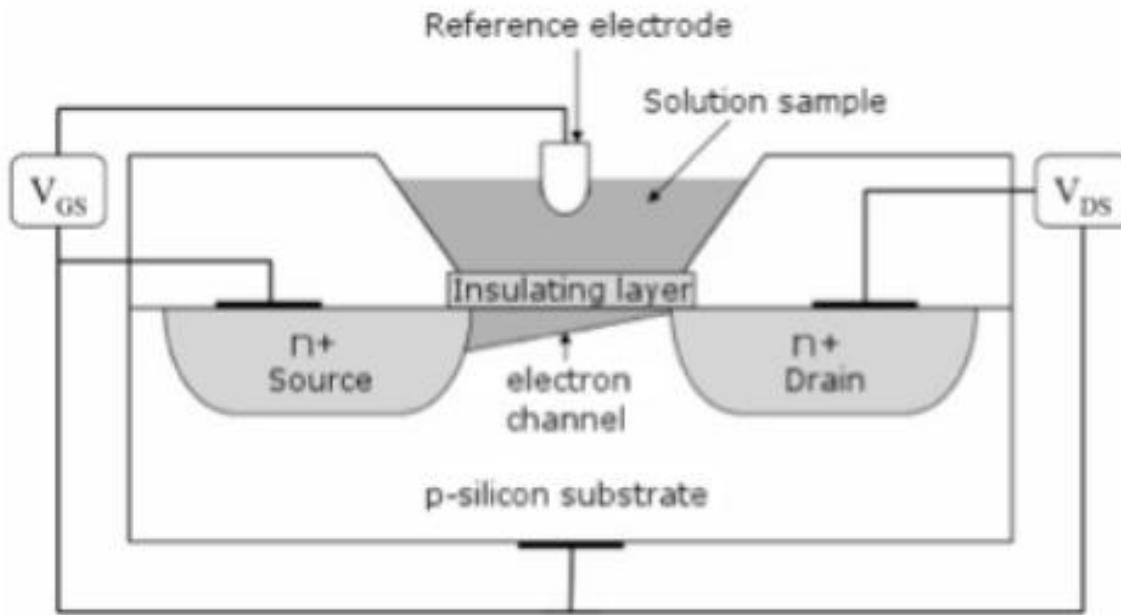


One significant improvement consists of the formation of a larger area capacitor build from the suspended gate and a floating gate. The floating gate then transmits the potential coming from the gas sensitive layer to a small FET-device with a short channel . This basically minimizes the loss of sensing signal due to weak coupling via the air gap

An appropriate sensing material is deposited on a flat carrier substrate forming what eventually becomes the gate-electrode. The preparation conditions are not limited by any Si-electronics related constraints.

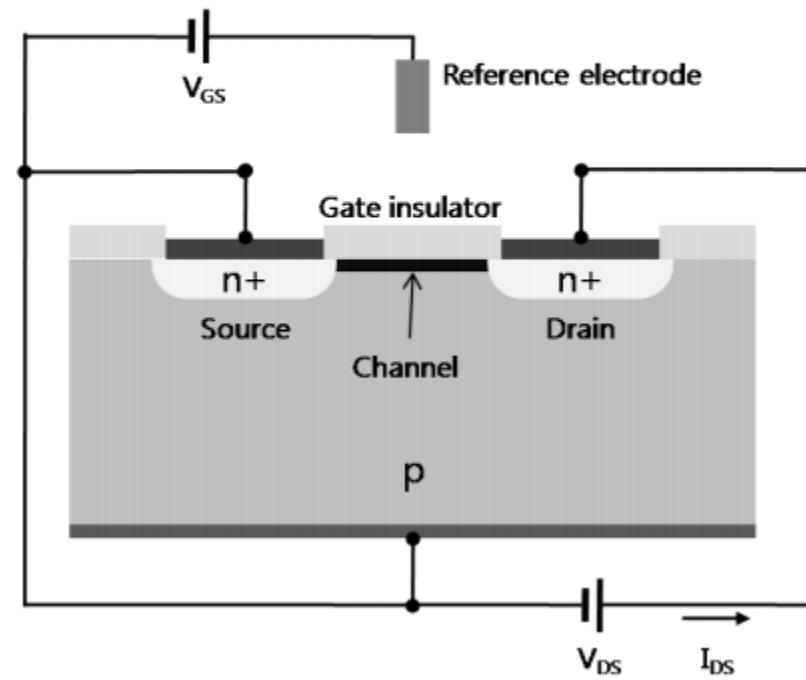
The Si FET-chip is separately prepared in standard CMOS. Electronics for driving the sensor may be integrated in the Si-Chip Finally both parts are bonded together so that a defined air gap is formed. These sensors are characterized by an unprecedented freedom in the choice of sensing materials. The gas receptor does not need to be an oxide. Organic molecules, polymers, metals, or salts may be used in the sensing platform. Since the sensors can be operated at room temperature, heating is not required. Multiple readout channels allowing for tiny sensing arrays can be realized on one single chip. The base potential of this sensing technology arises from the fact that the measurand comes from the direct measurement of surface effects. One consequence is that sensing materials in these devices do not need to be a semiconductor: metallic conductors or insulating materials can be used as well. One promising field for applications of such sensors is air monitoring in buildings. Due to their small dimensions and moderate costs, they can be used for distributed sensing networks allowing local sensing of air properties. Due to their low power consumption at ambient temperature operation, battery-operated sensing nodes that communicate wireless can be established. Since there are no high labor costs for wiring, an additional and decisive cost advantage for the sensor user occurs. An important application of such distributed sensors will be the control of the indoor air quality to allow for an on demand ventilation of different locations in a building. Good examples are meeting rooms. Usually, when the room is unused the ventilation is too high, and is much too low, when a meeting takes place. Similar situations occur at work places due to the changing occupancy and due to the varying loading of the air with contaminants

Ion Selective Field Effect Transistor



FET devices for ion sensing

Figure 1. Structure of ISFET. It consists of source, drain, gate insulator, and reference electrode.



In general, a field-effect transistor (FET) consists of three terminals; the source, drain, and gate. The voltage between the source and drain of the FET regulates the current flow in the gate voltage. Specifically, the current-control mechanism is based on an electric field generated by the voltage applied to the gate. The current is also conducted by only one type of carrier (electrons or holes) depending on the type of FET (n-channel or p-channel). A positive voltage applied to the gate causes positive charges (free holes) to be repelled from the region of the substrate under the gate. These positive charges are pushed downward into the substrate, leaving behind a carrier-depletion region. The depletion region is populated by the bound negative charge associated with the acceptor atoms. These charges are “uncovered” because the neutralizing holes have been pushed downward into the substrate [5]. The positive gate voltage also pulls negative charges (electrons) from the substrate regions into the channel region. When sufficient electrons are induced under the gate, an induced thin n-channel is in effect created, electrically bridging the source and drain regions. The channel is formed by inverting the substrate surface from p-type to n-type (inversion layer). When a voltage is applied between the drain and source with the created channel, a current flows through this n-channel via the mobile electrons (n-type FET). In the case of a p-type semiconductor, applying a positive gate voltage depletes carriers and reduces the conductance, whereas applying a negative gate voltage leads to an

accumulation of carriers and an increase in conductance (the opposite effect occurs in n-type semiconductors). The applied gate voltage generates an electric field which develops in the vertical direction. This field controls the amount of charge in the channel, and thus it determines the conductivity of the channel. The gate voltage applied to accumulate a sufficient number of electrons in the channel for a conducting channel is called the threshold voltage (V_{TH}). Note that V_{TH} for an n-channel (p-channel) FET is positive (negative).

With these properties, the FET can be configured as a biosensor by modifying the gate terminal with molecular receptors or ion-selective membranes for the analyte of interest. The binding of a charged biomolecule results in depletion or accumulation of carriers caused by change of electric charges on the gate terminal. The dependence of the channel conductance on gate voltage makes FETs good candidates for electrical biosensors because the electric field generating from the binding of a charged biomolecule to the gate is analogous to applying a voltage to a gate.

Generally, there are two types of planar FET-based biosensors, according to their structure; insulated-gate field-effect transistors (IGFET) and ISFET. In the case of IGFET, particularly MOSFET (metal-oxide-semiconductor field-effect transistor), the gate terminal is electrically isolated from the source and drain terminals. ISFET is similar to IGFET, but in the ISFET, the metal gate is replaced by an ion-selective membrane, electrolyte and a reference electrode (Figure 1). In the case of an ISFET biosensor, the amount of the current flow will be not only determined by the charges of biomolecules interacting on the gate dielectric, but also sensitive to pH, different ions, products of enzyme reactions, etc. An attractive feature of such FETs is that it is possible to detect biomolecular interactions in a label-free manner through a direct change in conductance or a related electrical property.

UNIT -4

BIOSENSORS



Biosensors

A biosensor is defined by IUPAC (1997) as:

“A device that uses specific biochemical reactions mediated by isolated enzymes, immunosystems, tissues, organelles or whole cells to detect chemical compounds usually by electrical, thermal or optical signals”.

Almost all biosensors are based on a three-component system: a **support**, a biological recognition element (**ligand**) that facilitates specific binding or biochemical reaction of a target, and a signal conversion unit (**transducer**).

- Biosensors thus have a wide range of applications including:
 - Clinical diagnostics
 - Drug development
 - Environmental monitoring (air, water, and soil)
 - Food quality control

A biosensor in general utilizes a biological recognition element that senses the presence of an analyte (the species to be detected) and creates a physical or chemical response that is converted by a transducer into a signal.

The important components of a biosensor are:

- (i) a sampling unit that introduces an analyte into the detector and can be as simple as a circulator;**
 - (ii) a recognition element that binds or reacts with a specific analyte, providing biodetection specificity (enzymes, antibodies, or even cells, such as yeast or bacteria, have been used as biorecognition elements**
 - (iii) stimulation that, in general, can be provided by optical, electric, magnetic, or other kinds of force fields that extract a response as a result of biorecognition; and**
 - (iv) a transduction process that transforms the physical or chemical response of the biorecognition in the presence of an external stimulation, into an optical, electrical, or magnetic signal that is detected by the detection unit. The detection unit may include pattern recognition for identification of the analyte**
-

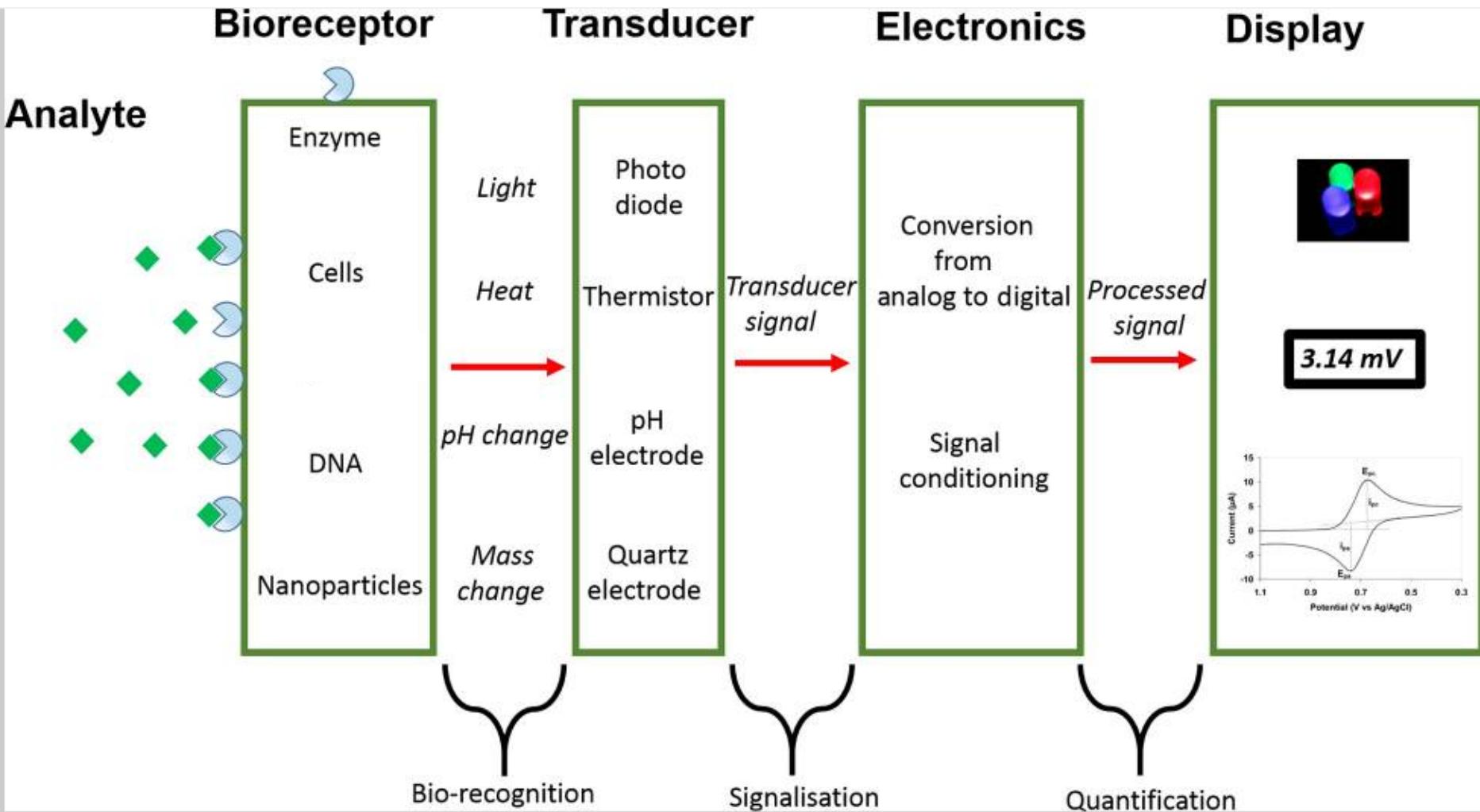
BIORECEPTORS

Bioreceptors can generally be classified into five different major categories

- 1) Antibody/ Antigen,**
- 2) Enzymes,**
- 3) Nucleic acids/DNA,**
- 4) Cellular structures/cells**
- 5) Biomimetic**

Bioreceptors are the key to specificity for biosensor technologies.

They are responsible for binding the analyte of interest to the sensor for the measurement.



Selectivity

Selectivity is perhaps the most important feature of a biosensor.

Selectivity is the ability of a bioreceptor to detect a specific analyte in a sample containing other admixtures and contaminants.

The best example of selectivity is depicted by the interaction of an antigen with the antibody. Classically, antibodies act as bioreceptors and are immobilised on the surface of the transducer. A solution (usually a buffer containing salts) containing the antigen is then exposed to the transducer where antibodies interact only with the antigens.

Biosensor = bioreceptor + transducer. A biosensor consists of two components: a bioreceptor and a transducer. The bioreceptor is a biomolecule that recognizes the target analyte, and the transducer converts the recognition event into a measurable signal. The uniqueness of a biosensor is that the two components are integrated into one single sensor (Figure 6.1.1). This combination enables one to measure the target analyte without using reagents (Davis et al, 1995). For example, the glucose concentration in a blood sample can be measured directly by a biosensor made specifically for glucose measurement, by simply dipping the sensor in the sample. This is in contrast to the commonly performed assays, in which many sample preparation steps are necessary and each step may require a reagent to treat the sample. The simplicity and the speed of measurements that require no specialized laboratory skills are the main advantages of a biosensor.

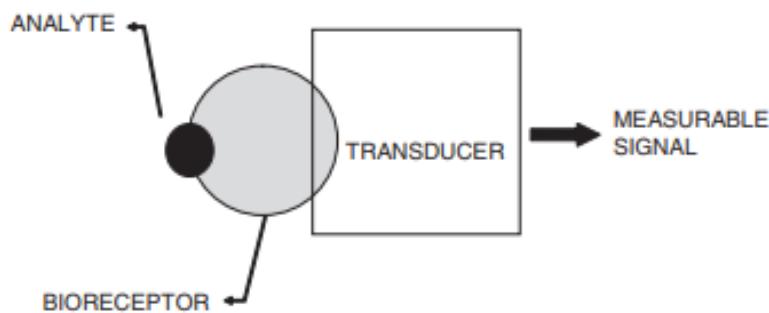
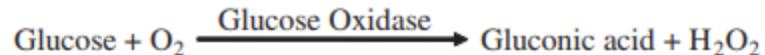


Figure 6.1.1: Biosensor configuration.

Enzyme is a Bioreceptor. When we eat food such as hamburgers and french fries, it is broken down into small molecules in our body via many reaction steps (these breakdown reactions are called *catabolism*). These small molecules are then used to make the building blocks of our body, such as proteins (these synthesis reactions are called *anabolism*). Each of these catabolism and anabolism reactions (the combination is called *metabolism*) are catalyzed by a specific enzyme. Therefore, **an enzyme is capable of recognizing a specific target molecule** (Figure 6.1.2). This biorecognition capability of the enzyme is used in biosensors. Other biorecognizing molecules (= bioreceptors) include antibodies, nucleic acids, and receptors.

Immobilization of Bioreceptor. One major requirement for a biosensor is that the bioreceptor be immobilized in the vicinity of the transducer. The immobilization is done either by physical entrapment or chemical attachment. Chemical attachment often involves covalent bonding to transducer surface by suitable reagents. A comprehensive treatment of immobilization is available in Hermanson (1996). It is to be noted that only minute quantities of bioreceptor molecules are needed, and they are used repeatedly for measurements.

Transducer. A transducer should be capable of converting the biorecognition event into a measurable signal (Figure 6.1.3). Typically, this is done by measuring the change that occurs in the bioreceptor reaction. For example, the enzyme glucose oxidase is used as a bioreceptor in a glucose biosensor that catalyzes the following reaction:



To measure the glucose in aqueous solutions, three different transducers can be used:

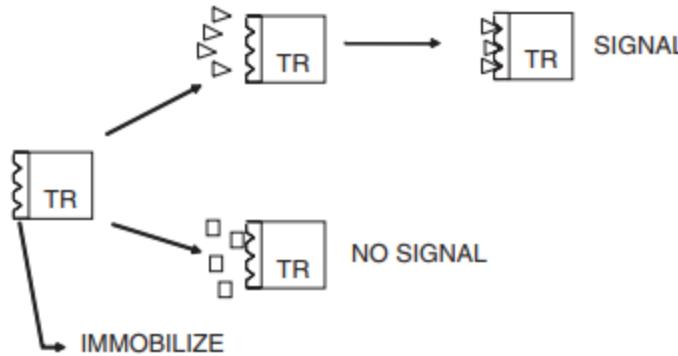
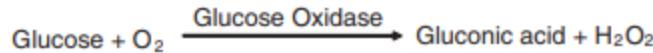
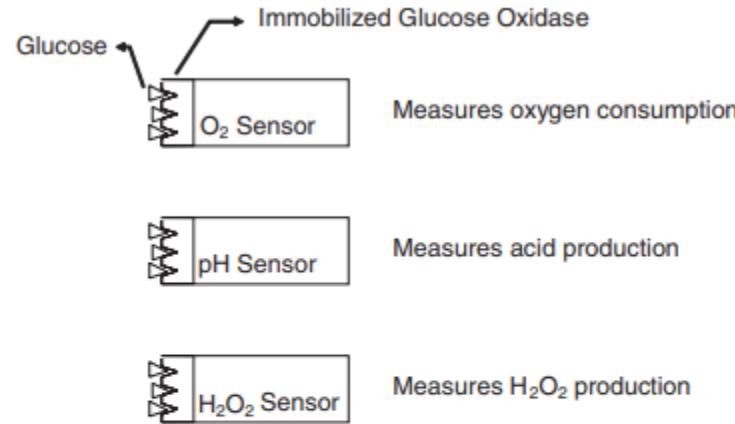


Figure 6.1.2: Specificity of biosensor (TR: transducer).



1. An oxygen sensor that measures oxygen concentration, a result of glucose reaction
2. A pH sensor that measures the acid (gluconic acid), a reaction product of glucose
3. A peroxidase sensor that measures H_2O_2 concentration, a result of glucose reaction



Note that an oxygen sensor is a transducer that converts oxygen concentration into electrical current. A pH sensor is a transducer that converts pH change into voltage change. Similarly, a peroxidase sensor is a transducer that converts peroxidase concentration into an electrical current. An excellent review of glucose sensing technologies was reported by Wilkins and Atansov (1996).

Figure 6.1.3: Three possible transducers for glucose measurement.

Biosensor Characteristics. Biosensors are characterized by eight parameters. These are: (1) *Sensitivity* is the response of the sensor to per unit change in analyte concentration. (2) *Selectivity* is the ability of the sensor to respond only to the target analyte. That is, lack of response to other interfering chemicals is the desired feature. (3) *Range* is the concentration range over which the sensitivity of the sensor is good. Sometimes this is called dynamic range or linearity. (4) *Response time* is the time required for the sensor to indicate 63% of its final response due to a step change in analyte concentration. (5) *Reproducibility* is the accuracy with which the sensor's output can be obtained. (6) *Detection limit* is the lowest concentration of the analyte to which there is a measurable response. (7) *Life time* is the time period over which the sensor can be used without significant deterioration in performance characteristics. (8) *Stability* characterizes the change in its baseline or sensitivity over a fixed period of time.

Considerations in Biosensor Development. Once a target analyte has been identified, the major tasks in developing a biosensor involve:

1. Selection of a suitable bioreceptor or a recognition molecule
2. Selection of a suitable immobilization method
3. Selection and design of a transducer that translates binding reaction into measurable signal
4. Design of biosensor considering measurement range, linearity, and minimization of interference, and enhancement of sensitivity
5. Packaging of the biosensor into a complete device

The first item above requires knowledge in biochemistry and biology, the second and third require knowledge in chemistry, electrochemistry and physics, and the fourth requires knowledge of kinetics and mass transfer. Once a biosensor has been designed, it must be packaged for convenient manufacturing and use. The current trend is miniaturization and mass production. Modern IC (integrated circuit) fabrication technology and micromachining technology are used increasingly in fabricating biosensors, as they reduce manufacturing costs. Therefore, an interdisciplinary research team, consisting of the various disciplines identified above, is essential for successful development of a biosensor.

Transduction principles



6.5 Transduction Mechanisms in Biosensors

Conventional Transducers. The majority of biosensors in use today use three types of transducers for converting the action of the bioreceptor molecule into a measurable signal. These are: amperometry based on H₂O₂ or O₂ measurement; potentiometry based on pH or pI on measurement; and photometry utilizing optical fibers. Biorecognition reactions often generate chemical species that can be measured by electrochemical methods. In these, typically the reaction product is H₂O₂ (or the reactant is O₂) which can be measured by a pair of electrodes. When a suitable voltage is impressed on one of the electrodes against a reference electrode (typically Ag/AgCl or Calomel), the target species (H₂O₂ or O₂) is reduced at the electrode and this generates electrical current (hence the name *amperometry*). In potentiometry, a glass membrane or a polymeric membrane electrode is used for measuring the membrane potential (hence the name *potentiometry*) resulting from the difference in the concentrations of H⁺ or other positive ions across the membrane. In photometry, the light

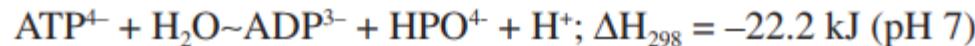
from an indicator molecule is the measured signal. In this method, one of the reactants or products of the biorecognition reaction results in colorimetric, fluorescent or luminescent changes that are measured using photodetectors. Usually, an optical fiber is used for guiding the light signals from the source to the detector. Adaptation and exploitation of these three routes, (**amperometric, potentiometric and photometric**), where user acceptability is already established, has been an obvious approach to the development of reagentless biosensor devices.

Piezoelectric Transducers. The transducer of a biosensor is not restricted to the three described above. In principle, any variable that is affected by the biorecognition reaction can be used to generate the transduced signal. The **piezoelectric materials** and surface acoustic wave devices offer a surface that is sensitive to **changes in mass**. These transducers have been used where the biorecognition reaction causes a change in mass. For example, piezoelectric silicon crystals—called quartz crystal microbalance (QCM)—have been used to measure very small mass changes in the order of picograms. For example see Bunde et al., (1998). QCM with immobilized antibody to pathogens have been successfully used to measure the presence of pathogens in aqueous samples. Piezoelectrically driven cantilevers have also been used to measure adsorption of very minute quantities of biochemicals (Raiteri et al., 2001).

Conductimetric Transducers. Monitoring **solution conductance** was originally applied as a method of determining reaction rates. The technique involves the measurement of changes in conductance due to the migration of ions. Many enzyme-linked reactions result in a change in total ion concentration and this would imply that they are suitable for conductimetric biosensors.

Electrical Capacitance as Transducer. When the biorecognition reaction causes a **change in the dielectric measurement constant** of the medium in the vicinity of the bioreceptor, the capacitance measurement method can be used as a transducer. Antigen-antibody reaction is a good example. Suppose antibody molecules are immobilized between two metal electrodes of known area. When antigen is added and binds with the antibody, the dielectric constant of the medium between the two electrodes is expected to change significantly. This change translates into a change in capacitance.

Thermometric Transducer. All chemical reactions are accompanied by the absorption (endothermic) or evolution (exothermic) of heat. Measurements of ΔH , the enthalpy of reaction at different temperatures, allows one to calculate ΔS (entropy) and ΔG (Gibbs free energy) for a reaction and therefore collect basic thermodynamic data. The hydrolysis of ATP for example is exothermic:



or the immunoreaction between anti-HSA and its antigen HSA yields -30.5 kJ/mol . For this latter reaction, the total increase in temperature for 1 mmol of antibody is of the order of 10^{-5} K , but many enzyme-catalyzed reactions have greater ΔH , and produce more easily measurable changes in temperature.

Enzyme Thermistor. For a biosensor device, the biorecognition compound must be immobilized on a temperature-sensing element capable of detecting very small temperature changes. The major initiative in this area has come from the Mosbach group at the University of Lund. Initially, they immobilized glucose oxidase or penicillinase in a small column, so that temperature changes in the column effluent were monitored by thermistors to give an *enzyme thermistor* sensitive to glucose and penicillin, respectively. They have also applied the technique to other substrates and to immunoassay using an enzyme-labeled antigen.

FET as a Transducer. As advances are made in biosensors, a need has developed for miniaturization and mass production. Field effect transistors (FET) used extensively in the semiconductor industry in memory chips and logic chips respond to changes in electric field (in front of the “gate” of the FET). An FET is thus capable of detecting changes in ion concentration when the gate is exposed to a solution that contains ions. Therefore, pH and ion concentration can be measured with an FET. The advantage of this transducer is that it can be incorporated directly into the electronic signal processing circuitry. In fact, a pen-size FET-based pH sensor is being marketed commercially.

Immobilization of biological elements



Choice of Supports

Supports can be classified as inorganic and organic according to their chemical composition.

The organic supports can be subdivided into natural and synthetic polymers.

Ideal support properties include physical resistance to compression, hydrophilicity, inertness toward enzymes, biocompatibility, resistance to microbial attack, and availability at low cost

The concept of irreversible immobilization means that once the biocatalyst is attached to the support it cannot be detached without destroying either the biological activity of the enzyme or the support. The most common procedures of irreversible enzyme immobilization are covalent coupling, entrapment or micro-encapsulation, and cross-linking.

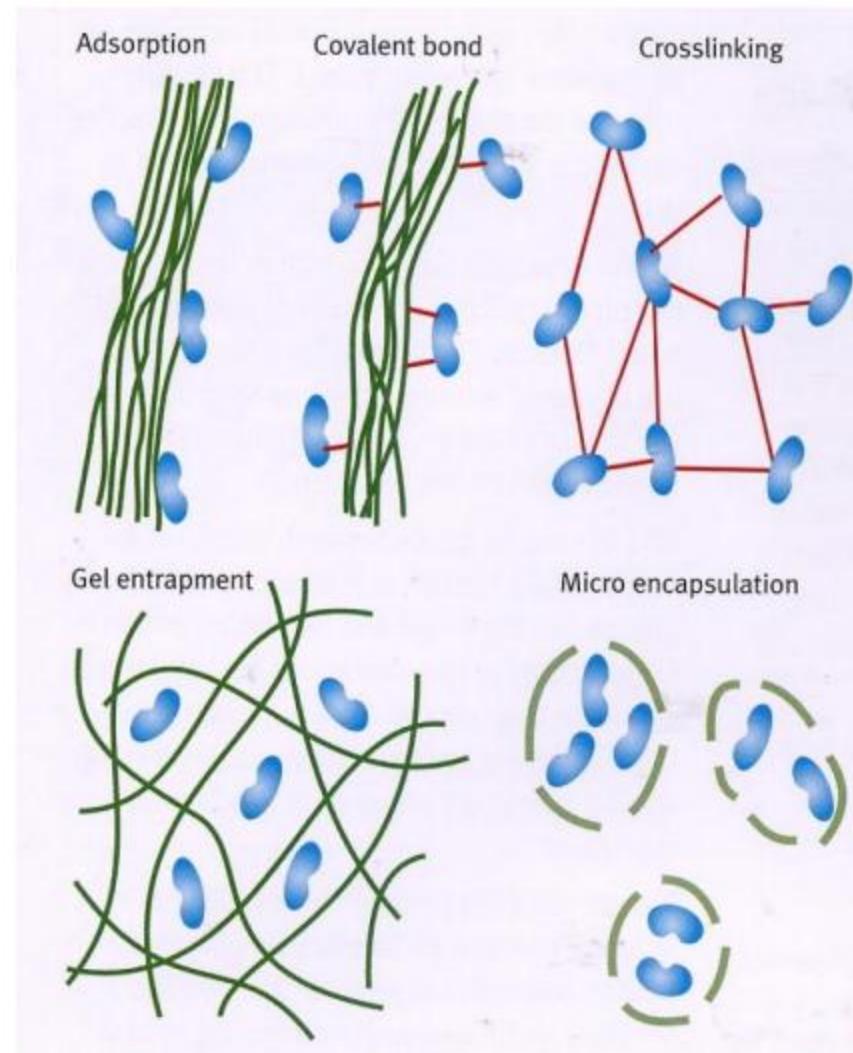


Why we need immobilization?



Immobilisation methods

- Adsorption
- Covalent bond
- Crosslinking
- Entrapment
- Encapsulation



The bioreceptor must be in intimate contact with the transducer and may be directly immobilised at the transducer surface.

However, the sensitive element is usually immobilised on or in some artificial substrate.

Immobilisation techniques were originally developed for enzymes,.

Immobilisation methods are usually broken down into five categories: **adsorption, inclusion, confinement, crosslinking, and covalent coupling on an active substrate**. Substrates of different kinds have been used Membrane substrates are ~~generally the best suited for making biosensors.~~

Receptors immobilisation

Optimal orientation of receptors (antibodies) results in:

- Greater affinity of the antigen
- Increased sensitivity of the affinity biosensors
- High reproducibility of the immobilisation



The biological elements are immobilized in/on a permeable membrane. Only specific analyte communicates with the immobilized biological elements residing within/on the membrane which grant products like heat, capacitance, electric potential, current, impedance, pH etc. The transducer reform the product signals into an electrical signal that is promoted to the electronic unit of the biosensor.

Adsorption

Adsorption is simply due to interactions between the bioreceptor and its support. If the latter is not charged, low energy bonds (hydrogen bonds, van der Waals forces, or hydrophobic effects) are established between bioreceptor and support. If it is charged, ionic bonds will also occur.

However, bioreceptors immobilised in this way can easily desorb through a change in pH or ionic strength, and this method is not widely used for bioreceptors.

Inclusion

The idea here is to incorporate the bioreceptor in a polymer which is usually in the form of a gel.

The bioreceptor is not therefore directly bound to the substrate. This technique, also easy to implement, allows a homogeneous distribution of the bioreceptor throughout the gel, but it may cause damage to biological structures, which may be affected by the reagents used in the polymerisation process.

Inclusion is mainly used to immobilise whole cells or subcellular fractions

Confinement

In the confinement technique, the bioreceptor remains in solution inside a compartment bounded by a semi-permeable membrane which only allows through small molecules. The bioreceptor thus remains in solution within a microcompartment of volume in the μL range or less, bounded by the membrane.

Crosslinking

The use of bifunctional agents such as glutaraldehyde $\text{CHO}-(\text{CH}_2)_3-\text{CHO}$ provides a way of crosslinking enzyme molecules one to another or co-crosslinking them in the presence of an inactive protein .

Crosslinking is also used to increase the stability of enzyme–substrate complexes obtained by adsorption or inclusion.



Covalent Bonding on an Activated Substrate

The covalent bonding of bioreceptors, mainly proteins, on activated substrates is achieved by setting up covalent bonds between functional groups of the substrate and functional groups of the bioreceptor that are not involved in the molecular recognition process.

The most widespread methods use --COOH , --NH_2 , --OH , and --SH groups on the substrate. These groups are chemically rather inactive and they must first be transformed into activated functions in order to be able to react under mild conditions with the functional groups of the bioreceptor.

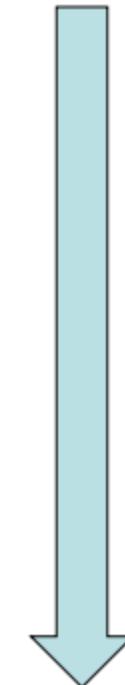
Immobilisation method	Substrate
Adsorption	Agarose, alumina, starch, carbon, cellulose, collagen, collodion, silica gel, ion exchange resins, porous glass, polyamide, polystyrene, polyurethane
Inclusion	Gels (alginate, starch, carrageenan, gelatine, polyacrylamide, polysiloxane, polystyrene, polyvinyl alcohol, silica)
Confinement	Dialysis and ultrafiltration membranes
Covalent coupling	Polyamino acids and proteins (polyglutamic acid, L-alanine and L-glutamic acid copolymer, collagen, fibroin) Polysaccharides (polygalacturonic acid, agarose, starch, cellulose, chitin, dextrose) Synthetic polymers (polyacrylamide, polyamide, polymethacrylate, polystyrene, polyvinyl alcohol) Inorganic substrates (silica gel, gold, quartz, porous glass)

Table 1. Immobilization procedures for enzymes

Method	Advantages	Disadvantages
<u>Adsorption</u> on insoluble matrices (e.g. by van der Waals forces, ionic binding or hydrophobic forces)	Simple, mild conditions, less disruptive to enzyme protein	Enzyme linkages are highly dependent on pH, solvent and temperature; insensitive
<u>Entrapment</u> in a gel (eventually behind a semipermeable membrane)	Universal for any enzyme, mild procedure	Large diffusional barriers, loss of enzyme activity by leakage, possible denaturation of the enzyme molecules as a result of free radicals
<u>Crosslinking</u> by a multifunctional reagent (such as glutaraldehyde bis-isocyanate derivatives or bis-diazobenzidine)	Simple procedure, strong chemical binding of the biomolecules; widely used in stabilizing physically adsorbed enzymes or proteins that are covalently bound onto a support	Difficult to control the reaction, requires a large amount of enzyme, the protein layer has a gelatinous nature (lack of rigidity), relatively low enzyme activity
<u>Covalent bonding</u> onto a membrane, insoluble supports	Stable enzyme-support complex, leakage of the biomolecule is very unlikely, ideal for mass production and commercialization	Complicated and time-consuming; possibility of activity losses due to the reaction involving groups essential for the biological activity (can be minimized by immobilization in the presence of the substrate or inhibitor of the enzyme)

Simple

Less stable



Complex

More stable



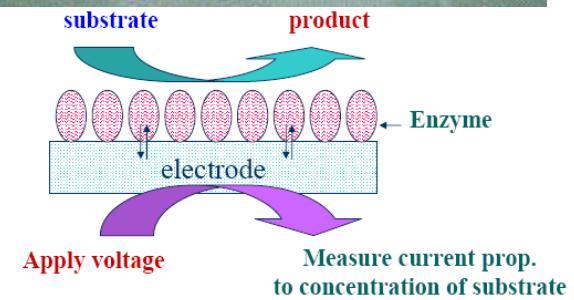
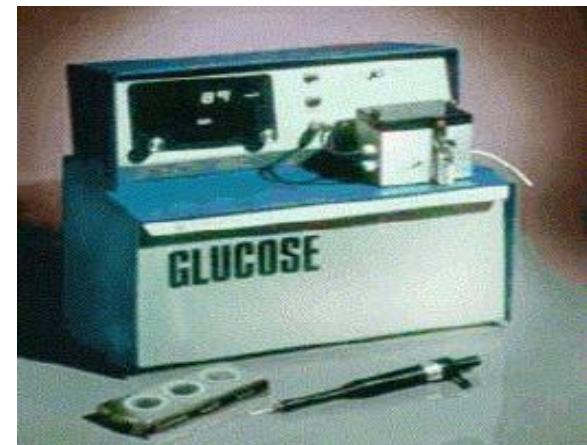
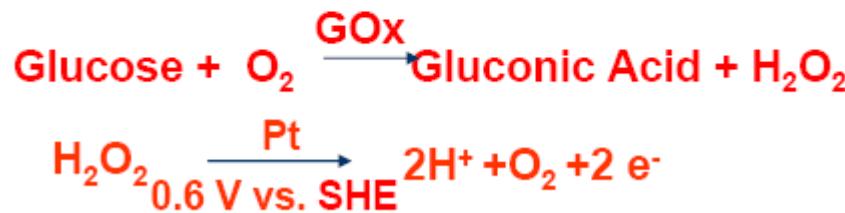
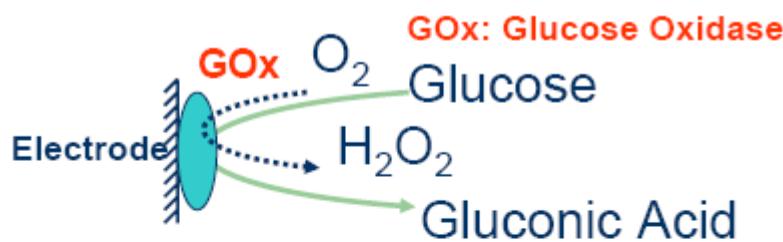
BLOOD GLUCOSE BIOSENSOR





**The first and the most commonly used commercial biosensor:
the blood glucose biosensor –
developed by *Leland C. Clark in 1962***

Basic Mechanism

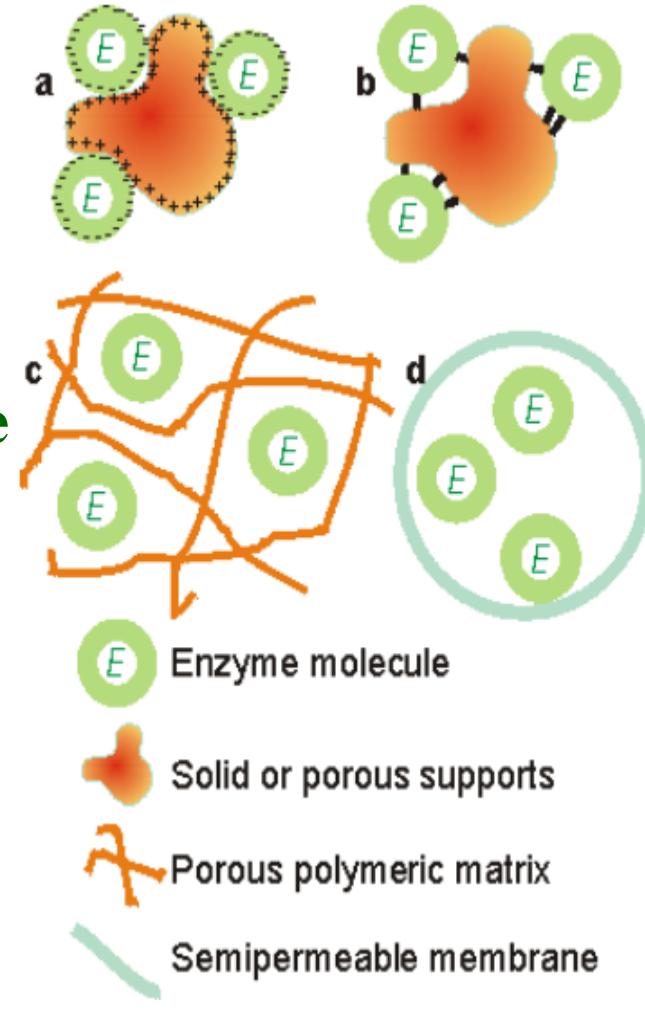


IMMOBILISING ENZYMES

Enzymes are biological catalysts that promote the transformation of chemical species in living systems.

These molecules, consisting of thousands of atoms in precise arrangements, are able to catalyze the multitude of different chemical reactions occurring in biological cells.

They have also been a key component in many ancient human activities, especially food processing, well before their nature or function was known.



Enzymes can catalyze reactions in different states: as individual molecules in solution, in aggregates with other entities, and as attached to surfaces. The attached—or “immobilized”—state has been of particular interest to those wishing to exploit enzymes for technical purposes.

The term “immobilized enzymes” refers to “enzymes physically confined or localized in a certain defined region of space with retention of their catalytic activities, and which can be used repeatedly and continuously.

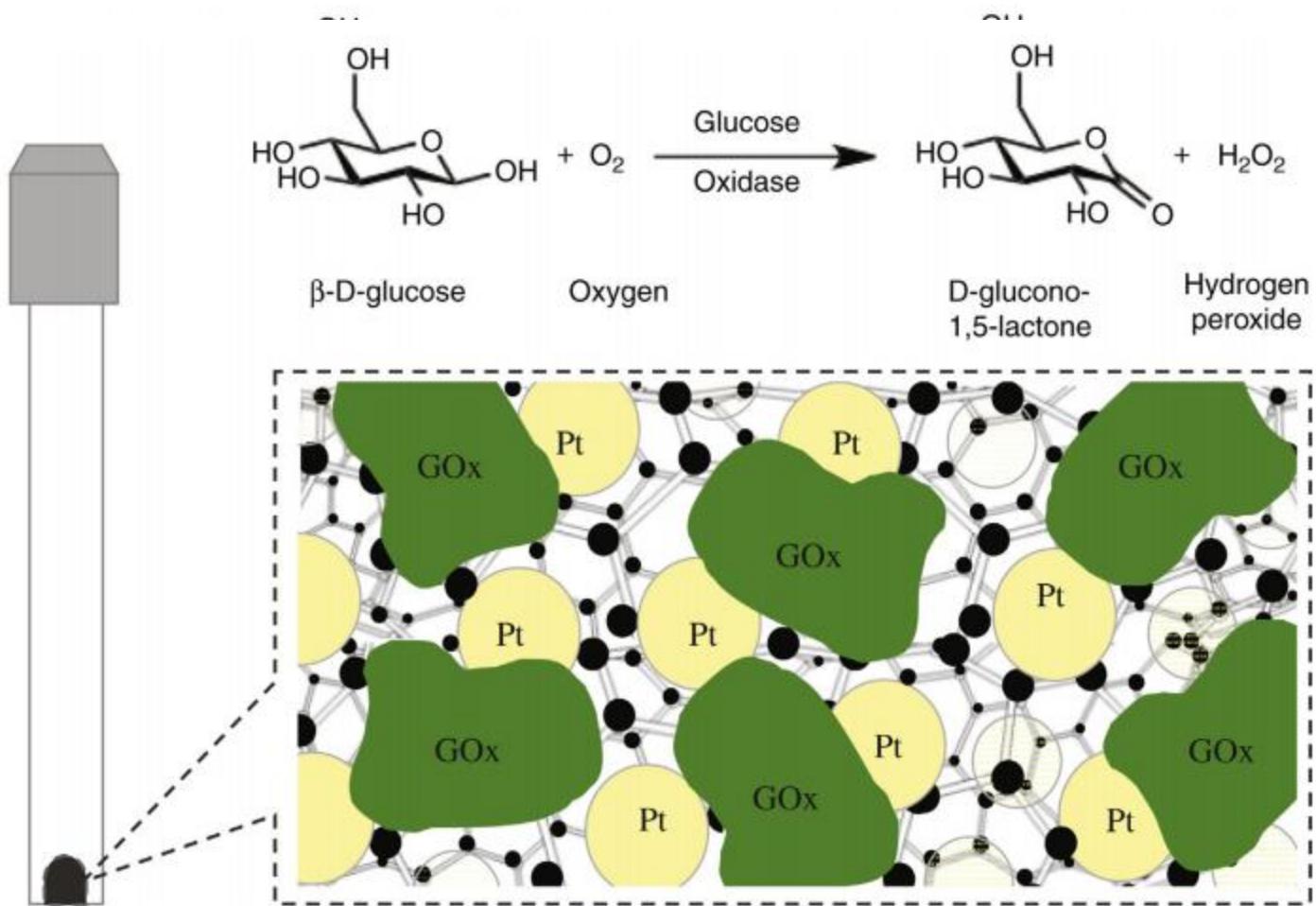


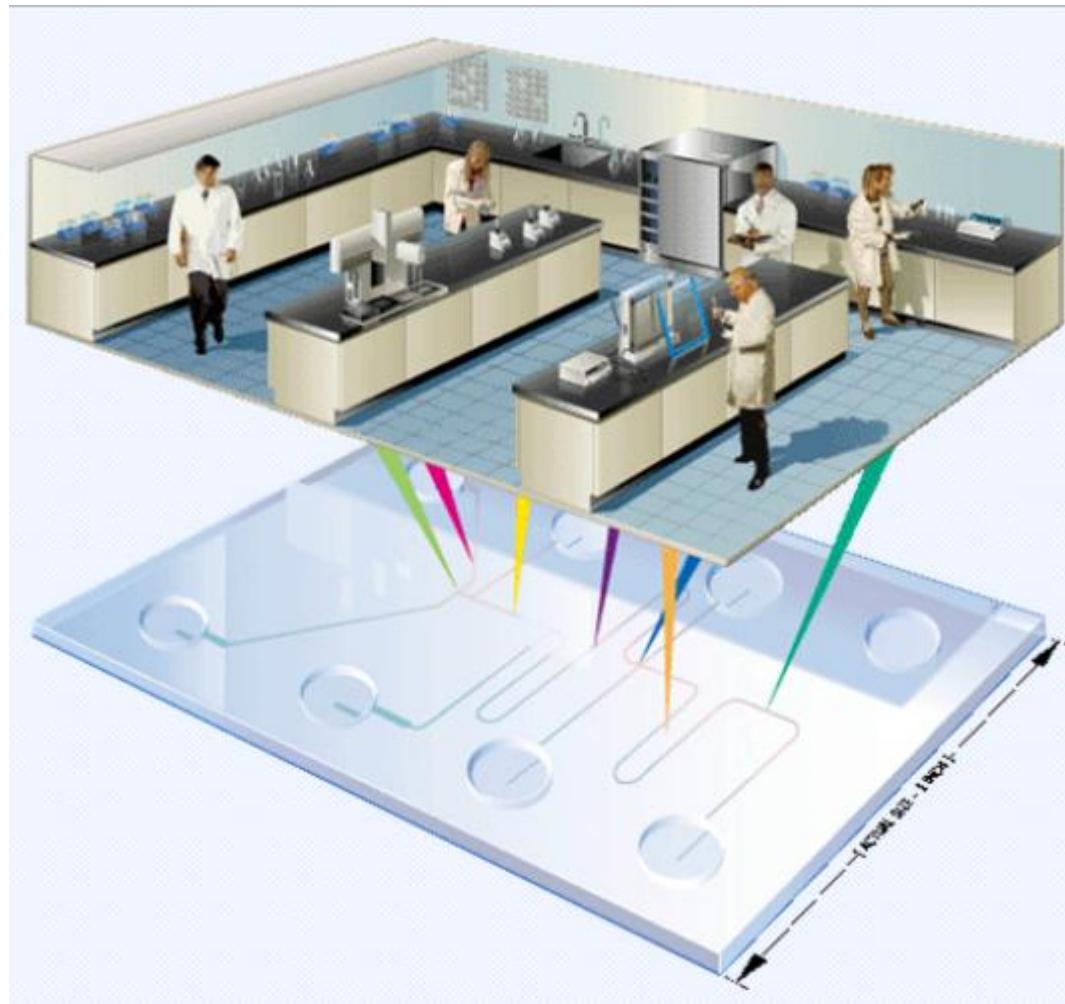
Figure 2.2. Schematic representation of glucose oxidase (GOx) immobilized on a platinum nanoparticle functionalized semi graphitic carbon support. The GOx/Pt/C electrocatalyst was packed into a glassy carbon electrode cavity.



Immobilized enzyme onto electrode surface (gel entrapment)

GOx can tolerate extreme changes in pH, temperature and ionic strength in comparison with other enzymes. Withstanding these conditions can be important during any manufacturing processes, making it a prime candidate for glucose monitoring devices

LAB-ON-CHIP SENSORS



A lab-on-a-chip (LOC) is an automated miniaturized laboratory system used for different clinical applications inside and outside the hospital.

**Examples of applications include
Measurements of blood gases,
blood glucose, and
cholesterol or
counting the number of HIV cells**

LOC-based applications are developing rapidly and that their number will increase in the near future.

Advantages compared to the current test methods.

Fast diagnostics at the location where diagnosis is needed (point of care) and

Small amounts of samples

and

Materials required to perform tests

Quality management aspects regarding calibration and maintenance of the device, and training and education of the user is necessary. This way, the benefits of LOC applications will not compromise quality of health care and patient safety

The technology used in LOC applications, microfluidics.

This technology provides the possibility to manipulate and handle fluids on microscale.

Devices incorporating the functionality of sample taking, sample preparation , sensing, and detection on a single microfluidic chip are commonly termed lab-on-a-chip or miniaturized total analysis systems (μ TAS),

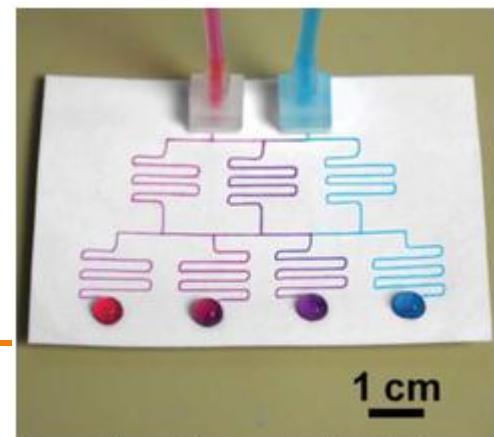
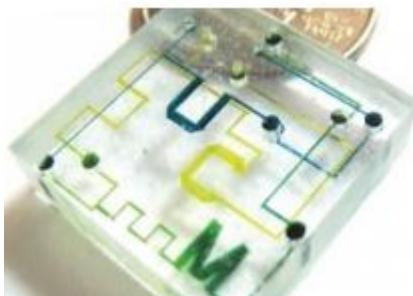
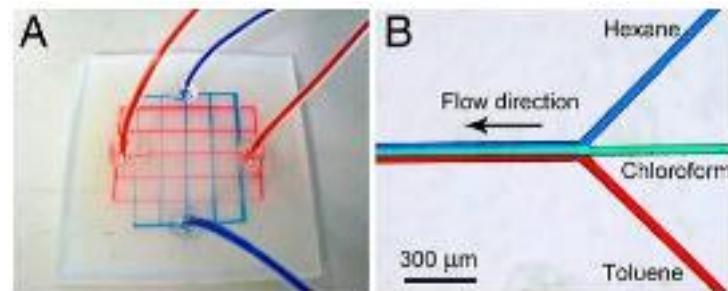
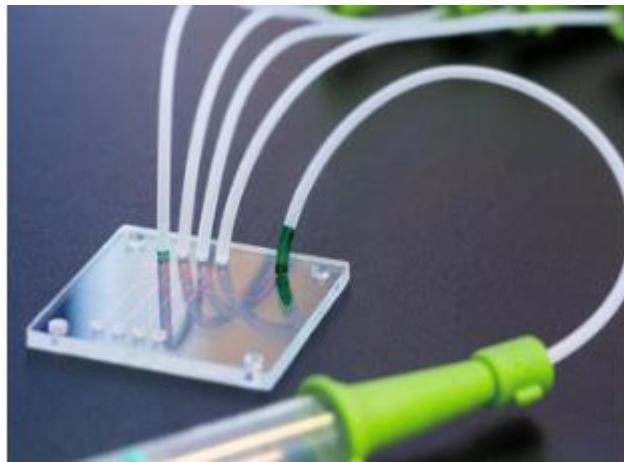
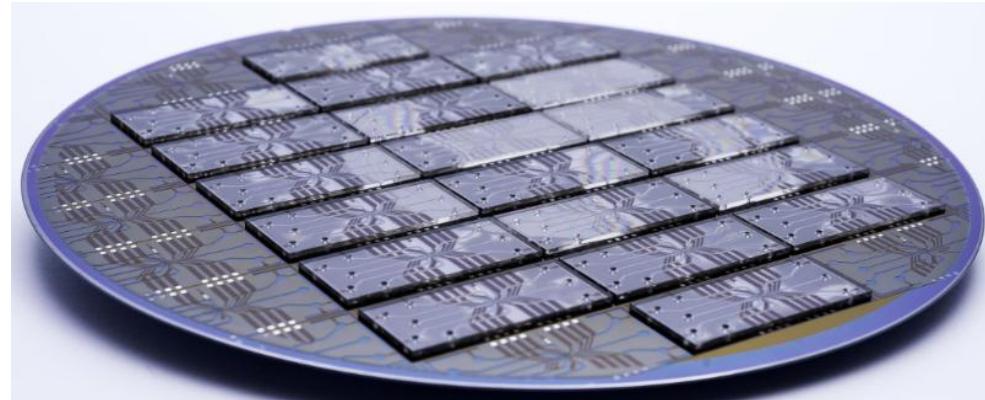
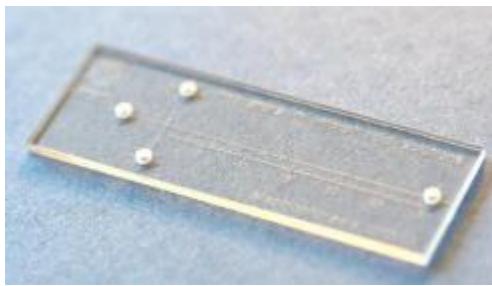
Microfluidics is a broader term that describes also mechanical flow control devices like pumps and valves or sensors like flowmeters and viscometers.

At present, there is no system on the market that completely integrates the full scope of these tasks. However, lab-on-a-chip systems have been developed that partially use macro periphery to perform crucial steps, for example, sample preparation and detection

For the manufacture of LOC devices, various materials can be used.

The most common device substrates are

**Silicon,
Glass,
Polymers such as Polydimethylsiloxane,
Paper.**



From Glavan et al. (2013)

Each dedicated microfluidic platform also contains sets of microfluidic elements, performing basic fluidic unit operations.

These **basic operations have to be able to transport fluids, storage of reagents and preparation of the sample.**

Microfluidic unit operations can be combined in different microfluidic platforms, such as lateral flow tests, linear actuated devices and surface acoustic waves. LOC can be used for the detection of proteins, nucleic acids, cells, pathogens, metabolites and other small molecules.

The largest LOC market segment clinical diagnostics, can be divided between point-of-care (POC) testing (i.e., a diagnostic test performed near the patients without needing a clinical laboratory) and central laboratory-based testing (i.e., diagnostic laboratory in a hospital).

Clinical diagnostics ranges from relatively simple immunochromatographic strips, similar to pregnancy tests, to highly complex systems requiring external machinery and expert training for their handling. Clinical diagnostic applications also include detecting nucleotides and peptides that are considered early indicators of disease. In many ways, the features of LOC devices fulfil the requirements for a POC diagnostic device: low consumption of reagents and sample, miniaturization of device and fast turn-around time for analysis.

It is a versatile technology that enables the miniaturization of complex fluid handling and integrated detection.

Microfluidic unit operations

Similar to the platforms in the application-specific integrated circuit industry in microelectronics, which provide elements and processes to make electronic circuitries, a **dedicated microfluidic platform comprises a set of microfluidic elements**. These elements have to be able to perform the basic fluidic unit operations

Pumping and valving

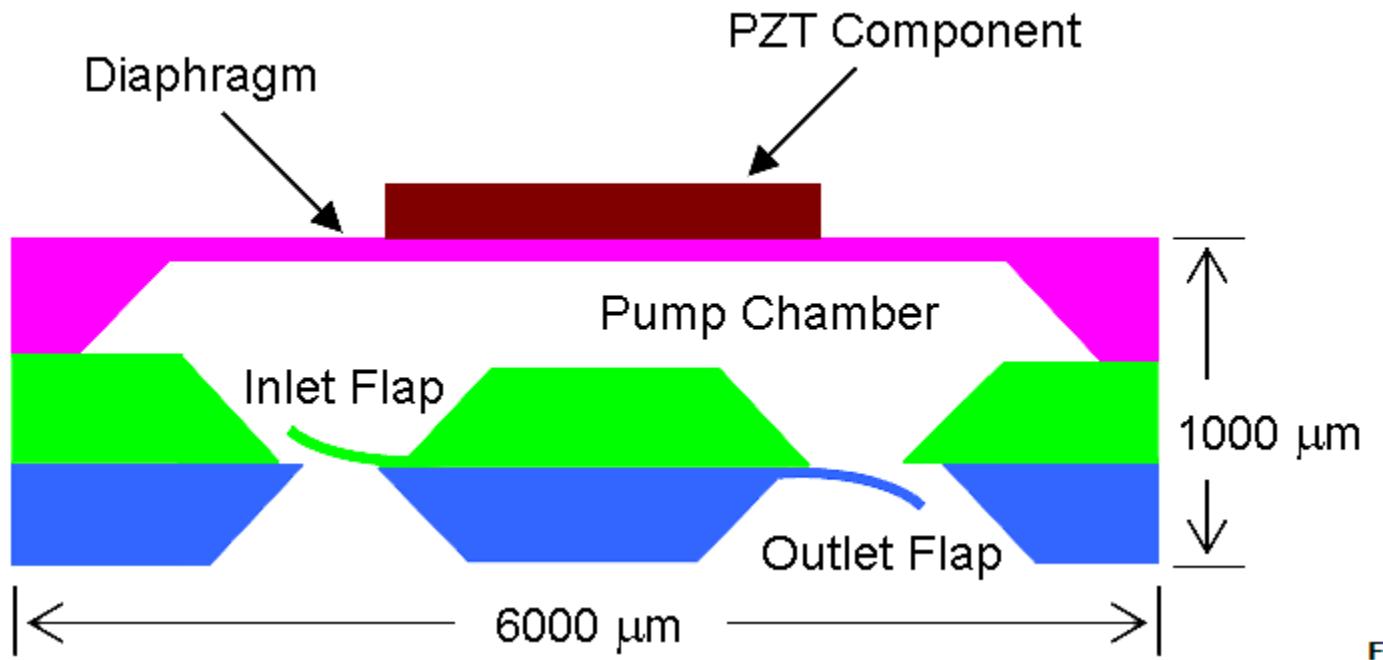
Microfluidic analytical systems require micropumps and microvalves **enabling precise control of sample, buffer, and reagent flow and delivery**

Pumping and valving

Several mechanisms have been used for transporting the fluids in microfluidic systems and they can be categorised in displacement and dynamic pumping.

Displacement pumps exert pressure forces on the fluid through one or more moving boundaries.

Micropumps can be based on reciprocating or rotary actuations or may have piezoelectric, (thermo)pneumatic, electrostatic and electromagnetic moving units to displace fluids



Sketch of micropump cross-section. Alternating voltage causes the PZT component to expand and contract along the horizontal direction. This induces a bending stress on the diaphragm, which in turn pumps the fluid through the chamber.

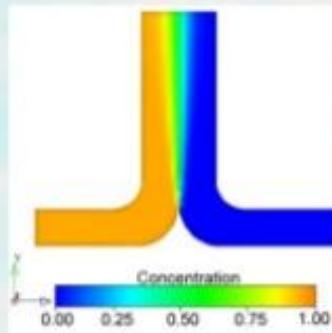
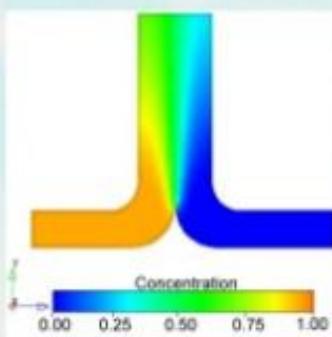
MIXING

Sample dilution, resuspension of dried reagents, and reaction of multiple reagents in LOC devices often require rapid and efficient mixing.

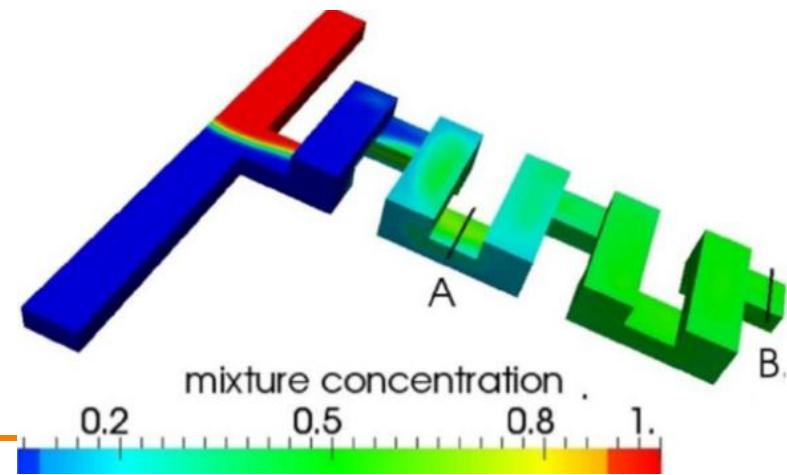
However, mixing in microfluidic platforms is difficult because flow is laminar and mixing is dominated by diffusion unless special measures are taken. Efficient micro-mixing can be achieved through a number of active and passive mixing mechanisms.

In active mixing, external driving forces such as acoustic waves, magnetic beads coupled with moving permanent magnets, or actuated air bubbles enhance the mixing of samples. In passive mixing, liquids are driven through microstructures designed to increase the contact area between different streams and to speed up diffusive or induce chaotic mixing.

Diffusion



2D solution for the diffusion of biotin ($D = 340 \mu\text{m}^2/\text{s}$, left image) and albumin ($D = 65 \mu\text{m}^2/\text{s}$, right image) through a T-sensor. A normalized concentration of 1.0 enters at the left inlet where the velocity flow rate is $125 \mu\text{m}/\text{s}$. Buffer enters at the right inlet at the same velocity.





When glacier water meets the river, Norway

SEPARATION

The beginning of modern microfluidic and LOC devices is closely linked to **separations of (bio)chemical substances**, in particular using electrophoresis.

Separation is important for LOC devices because it increases the target purity by removing interfering agents prior to detection.

Separation methods include capillary electrophoresis, dielectrophoresis, isoelectric focusing, liquid (electro)chromatography, size-based filtration, magnetic fields, acoustic waves, optical tweezers, and various combinations of flow, diffusion, and sedimentation based phenomena.

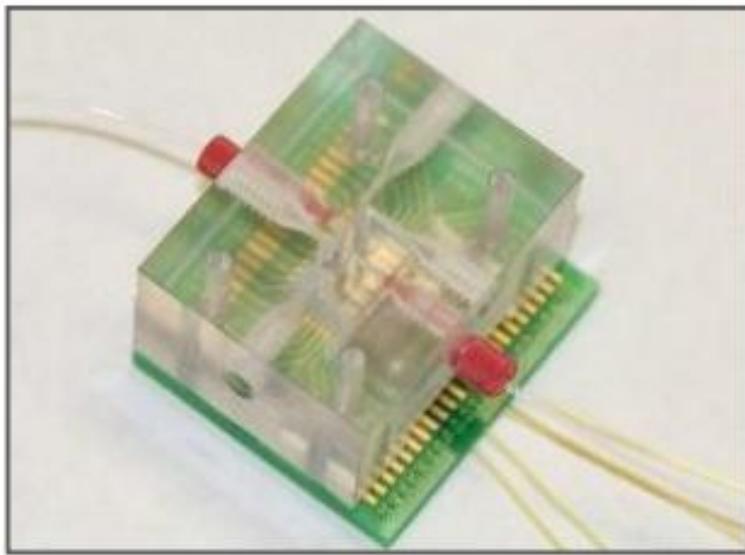
Reagent storage

For practical LOC devices, it is necessary to store reagents for extended periods on or in the device. Reagent, e.g. enzymes or antibodies, can be stored in a wet or dry state. The latter is often preferred in those cases where drying does not cause total and unrecoverable loss of activity, because reagents that are successfully dried typically exhibit improved stability relative to those stored wet

Sample preparation

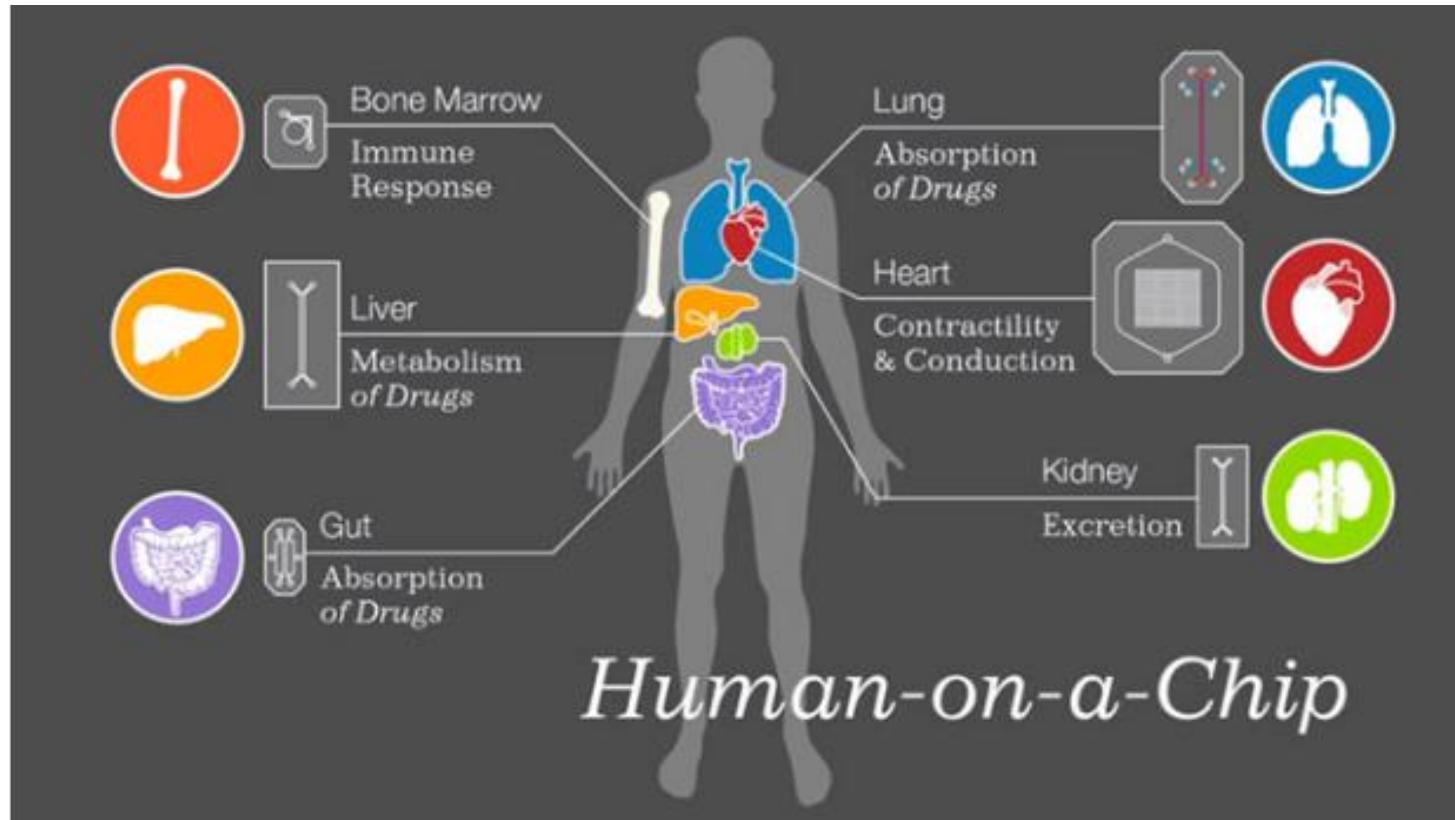
Sample preparation, a necessary analytical step, is important in achieving adequate sensitivity and specificity in any detection platform. This is especially important in the case of complex matrices, such as blood, saliva, and interstitial fluid. Sample preparation encompasses sample concentration, diffusion, filtration, purification and fractionation of analytes from analytically noisy background matrices. Although large numbers of LOC devices accommodate unprocessed blood samples, the range of assays that can be performed is limited by the lack of well-developed on-chip sample preparation methodologies

Below is a miniaturized, portable version of a blood-count machine that was tested by astronauts. On long missions, astronauts will need the ability to analyze blood samples in real-time to diagnose infection, allergies, anemia or deficiencies in the immune system. This device, about the size of a cell phone, is being designed to accomplish this task



Lab-on-a-chip – Blood Analysis [Photo courtesy of Y. Tai, California Institute of Technology]

Company	Country	Name of device/chip/system	M/D ¹	Application
Abaxis Inc	USA	Piccolo® Xpress	M	blood analysis
Abbott Diabetes Care Inc	USA	FreeStyle Lite®	M	blood glucose
Abbott Diabetes Care Inc	USA	FreeStyle Freedom Lite®	M	blood glucose
Abbott Diabetes Care Inc	USA	FreeStyle InsuLinx	M	blood glucose
Abbott Diabetes Care Inc	USA	Precision Xtra®	M	blood glucose and ketone
Abbott Diabetes Care Inc	USA	Precision Xceed Pro	M	blood glucose and beta-ketone (hospital setting)
Abbott Point of Care Inc	USA	i-STAT®	M	cardiac markers, blood gases, electrolyte analyses, lactate, coagulation, haematology
Abbott Point of Care Inc	USA	i-STAT® 1 Wireless	M	cardiac markers, blood gases, electrolyte analyses, lactate, coagulation, haematology
Achira Labs Pvt Ltd	India	ACHIRA 2000	M	thyroid disorders, infertility
Advanced Liquid Logic	USA	-	D	HIV diagnostics / CD4 count
Agilent Technologies Inc	USA	2100 Bioanalyzer	M	nucleic acids, proteins and cells
Akronni Biosystems Inc	USA	TruDiagnosis®	M	DNA, RNA, and antibody-based testing
Alere Inc	USA	Alere Pima™ CD4 Analyser	M	HIV diagnostics / CD4 count
Alere Inc	USA	NAT System	D	HIV diagnostics / CD4 count
Alere Inc	USA	Alere Cholestech LDX® System	M	cholesterol, blood glucose, liver enzymes
Alere Inc	USA	Alere™ Heart Check System	M	B-type natriuretic peptide
Alere Inc	USA	Alere Triage® MeterPro	M	BNP, CK-MB, D-dimer, myoglobin, NGAL, troponin I, PLGF
Alere Inc	USA	Alere™ INRatio® / INRatio® 2 PT / INR Monitor	M	Anticoagulation
Arkray Global Business Inc	Japan	GLUCOCARD 01	M	blood glucose
Arkray Global Business Inc	Japan	GLUCOCARD 01-mini	M	blood glucose
Arkray Global Business Inc	Japan	GLUCOCARD 01-mini plus	M	blood glucose
Arkray Global Business Inc	Japan	GLUCOCARD X-METER GT-1910	M	blood glucose
Arkray Global Business Inc	Japan	GLUCOCARD X-mini	M	blood glucose
Arkray Global Business Inc	Japan	GLUCOCARD X-mini plus	M	blood glucose
Arkray Global Business Inc	Japan	GLUCOCARD Σ	M	blood glucose
Arkray Global Business Inc	Japan	GLUCOCARD Σ-mini	M	blood glucose
Atonomics A/S	Denmark	Atolyzer	D	cardiovascular disease, prostate cancer
Axis-Shield plc	UK	Afineon	M	CRP, HbA1c, ACR, lipid
Axis-Shield plc	UK	NyoCard	M	CRP, HbA1c, D-dimer, U-albumine



The goal is really to do the whole human body, and then we can fluidically link multiple chips to capture interactions between different organs and eventually recreate a body on a chip

Integrated sensors: system organization and function



In a simple control system, the sensor is only one of three items required to implement a control strategy. The sensor provides an input to a controller with the desired strategy in its memory, and the controller drives an output stage to modify or maintain the status of a load, such as a light, a motor, a solenoid, or a display. As shown in Figure 1.4, a signal conditioning interface typically exists between the sensor(s) and the controller and between the controller and the output stage.

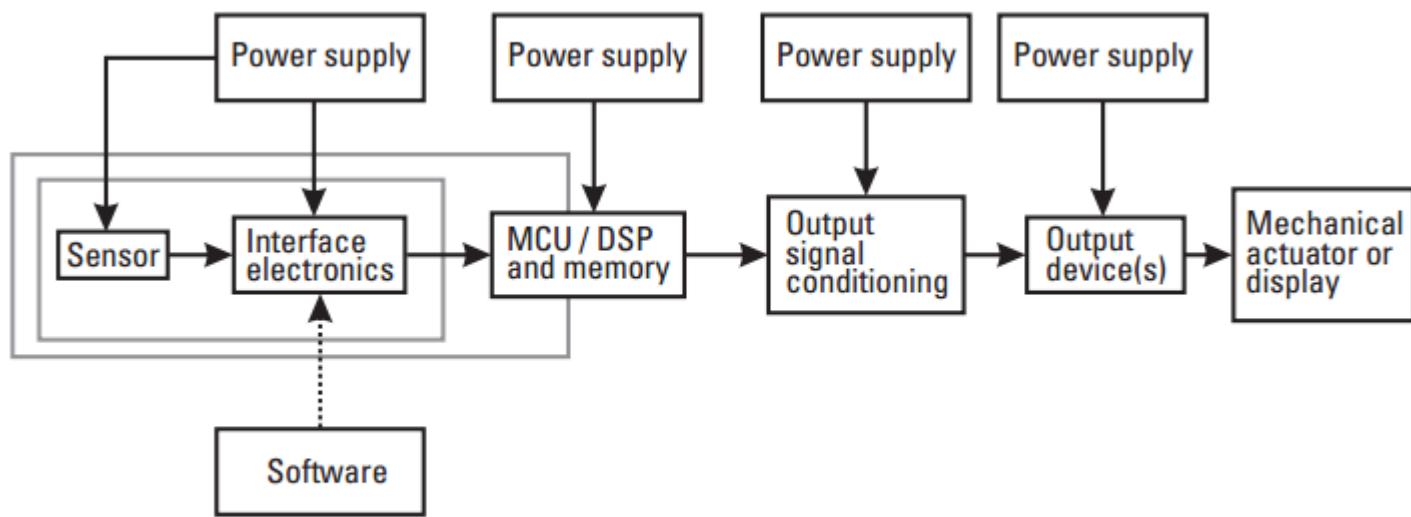


Figure 1.4 Generic control system.

and the output device. **Smart sensing includes a portion of the controller's functions in the sensor portion of the system. That means software will play an increasingly important role in smart sensors.** The power supply requirements for the electronics and the sensor represent an additional consideration that is becoming more important as MCU voltages are decreased and more sensors are used in battery power or portable applications. The number of supplies in Figure 1.4 may not be required for a particular application, but they serve as a reminder for considering the available voltage for the sensor and the interface versus the rest of the system.

As shown in Figure 1.5, in addition to the sensing element and its associated amplification and signal conditioning, an A/D converter, memory of some type, and logic (control) capability are

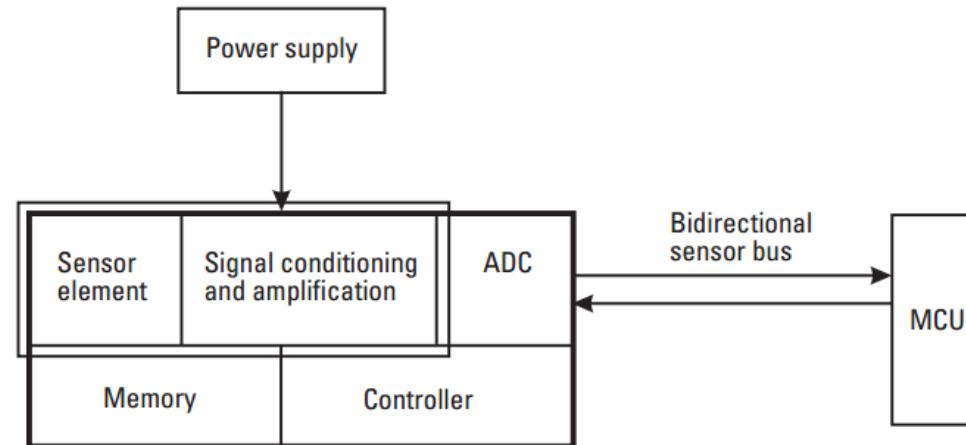


Figure 1.5 Smart sensor model.

included in the smart sensor. Once the signal is in digital format, it can be communicated by several communication protocols. The regulated power supply also required for the system and its effect on system accuracy must be taken into account. That is becoming more of an issue as power management issues are addressed in system design and different supply voltages proliferate.

Reducing the number of discrete elements in a smart sensor (or any system) is desirable to reduce the number of components, form factor, interconnections, assembly cost, and frequently component cost as well. The choices for how that integration occurs are often a function of the original expertise of the integrator. For example, as shown in Figure 1.6, a sensor manufacturer that already uses semiconductor, that is, bipolar or metal oxide semiconductor (MOS), technology for the sensing element may expand the capability and increase the value (and intelligence) of the sensing unit it produces by combining the signal conditioning in the same package or in a sensor module. Through integration, the signal conditioning can also be combined at the same time the sensor is fabricated or manufactured.

While the process of integration is more complex, the integrated sensor can be manufactured with the sensor and signal conditioning optimized for a particular application. Conversely, an MCU manufacturer using a complementary metal oxide semiconductor (CMOS) process typically integrates memory, A/D, and additional signal conditioning to reduce the number of components in the system. A variety of combinations are indicated in Figure 1.6. Processing technology is a key factor. However, manufacturers not only must be willing to integrate additional system components, they also must achieve a cost-effective solution. Combinations of hybrid (package level) and monolithic integration are discussed frequently in the remainder of this book. Different design philosophies and the necessity to partition the sensor/system at different points can determine whether a smart sensor is purchased or, alternatively, designed using a sensor signal processor or other components necessary to meet the desired performance of the end product.

The integration path can have a significant effect on the ultimate level of component reduction. As shown in Figure 1.7 [8], the input (demonstrated by a pressure sensor), computing (high-density CMOS [HCMOS] microcontroller), and output side (power MOS) are all increasing the level of monolithic integration. The choice of sensor technology, such as bipolar, can have a limiting effect on how far the integration can progress. For example, a bipolar sensor can increase integration level by adding signal conditioning and progress to a monolithic level III sensor. Through package-level integration, a two-chip sensor controller can be achieved by combining the sensor with an HCMOS microcontroller. However, the highest level of monolithic integration, level V, will be realized only by pursuing MOS-compatible sensing and power-control technologies. Realizing the full potential of those new sensors will require a new approach to identify sensor applications

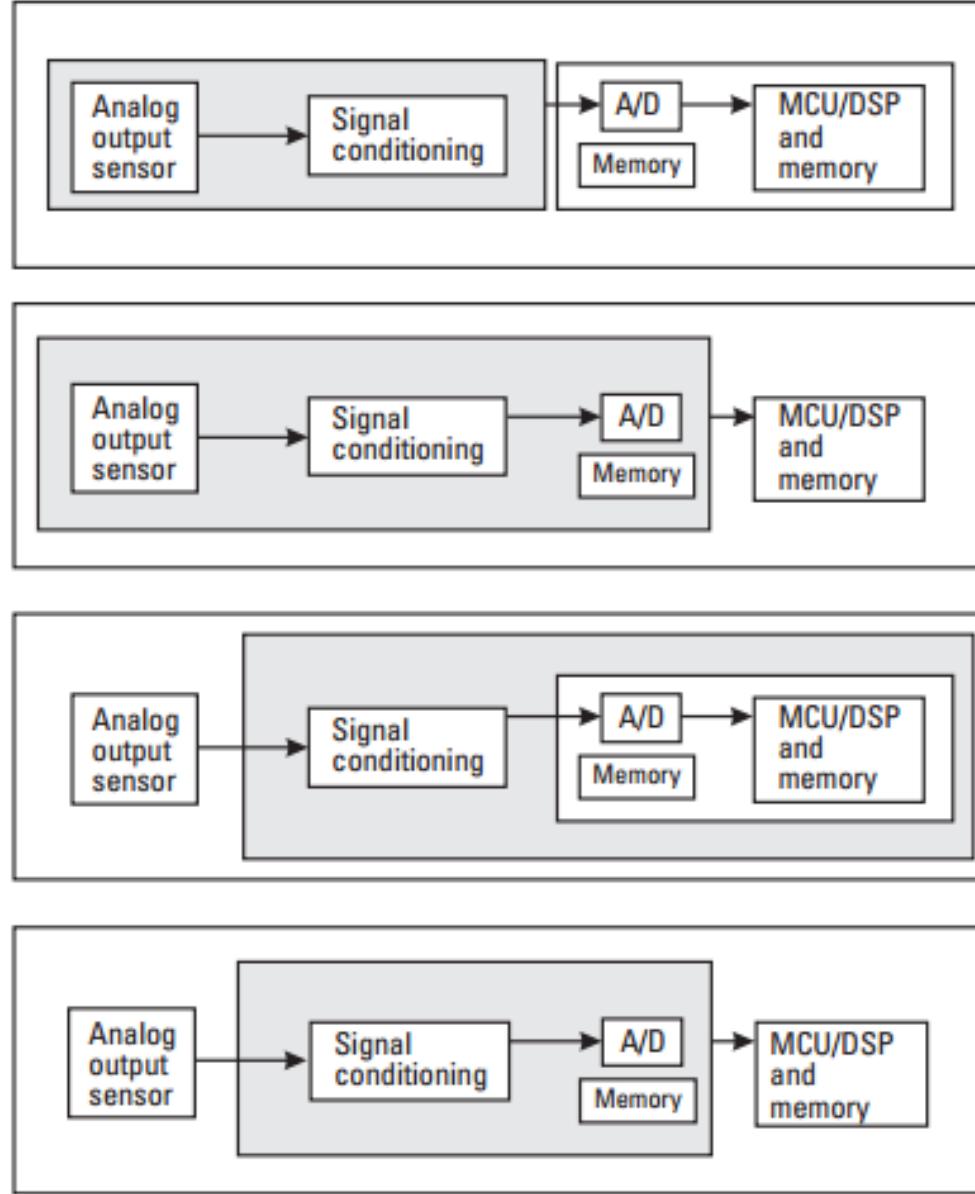


figure 1.6 Partitioning and integration possibilities.

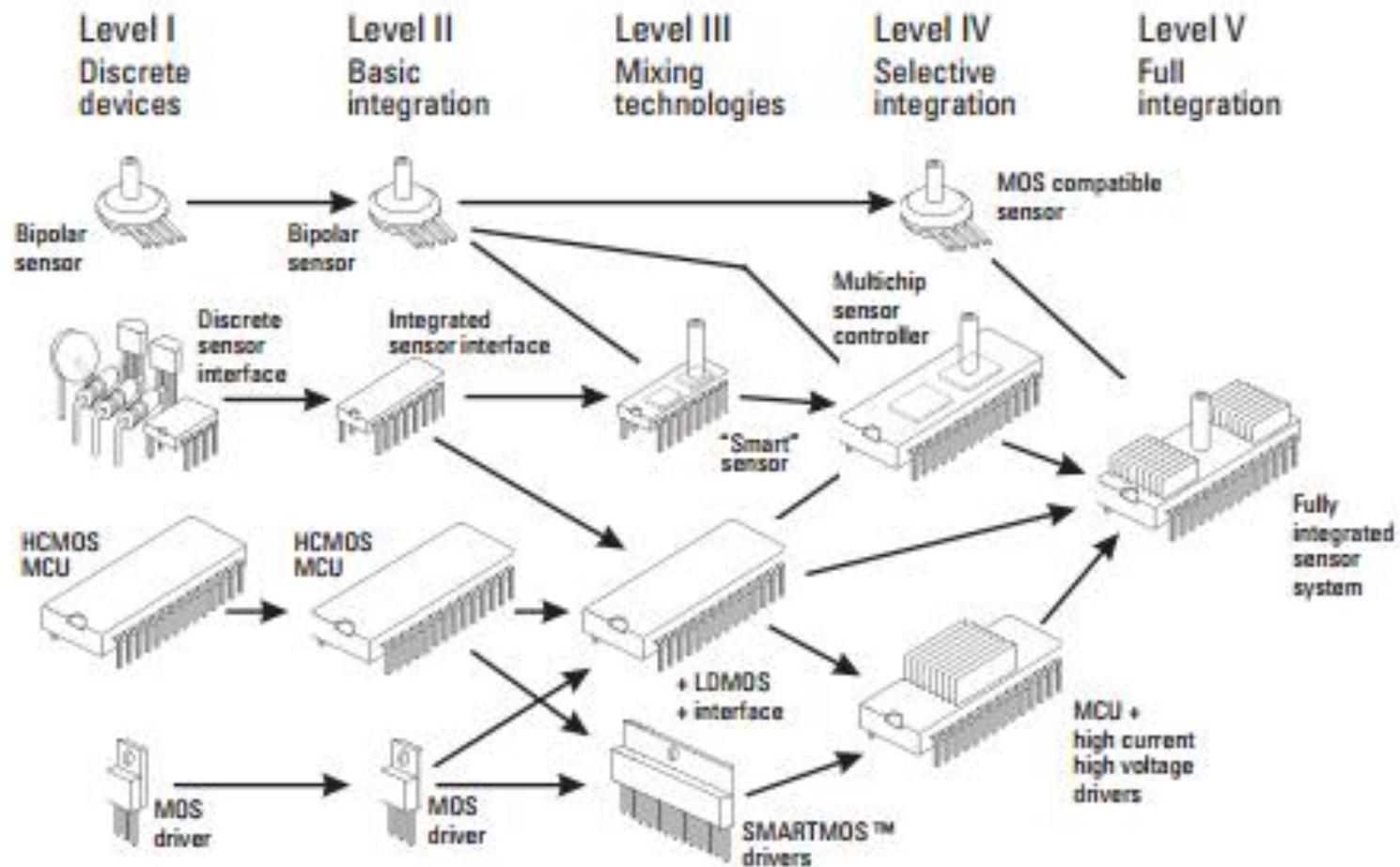
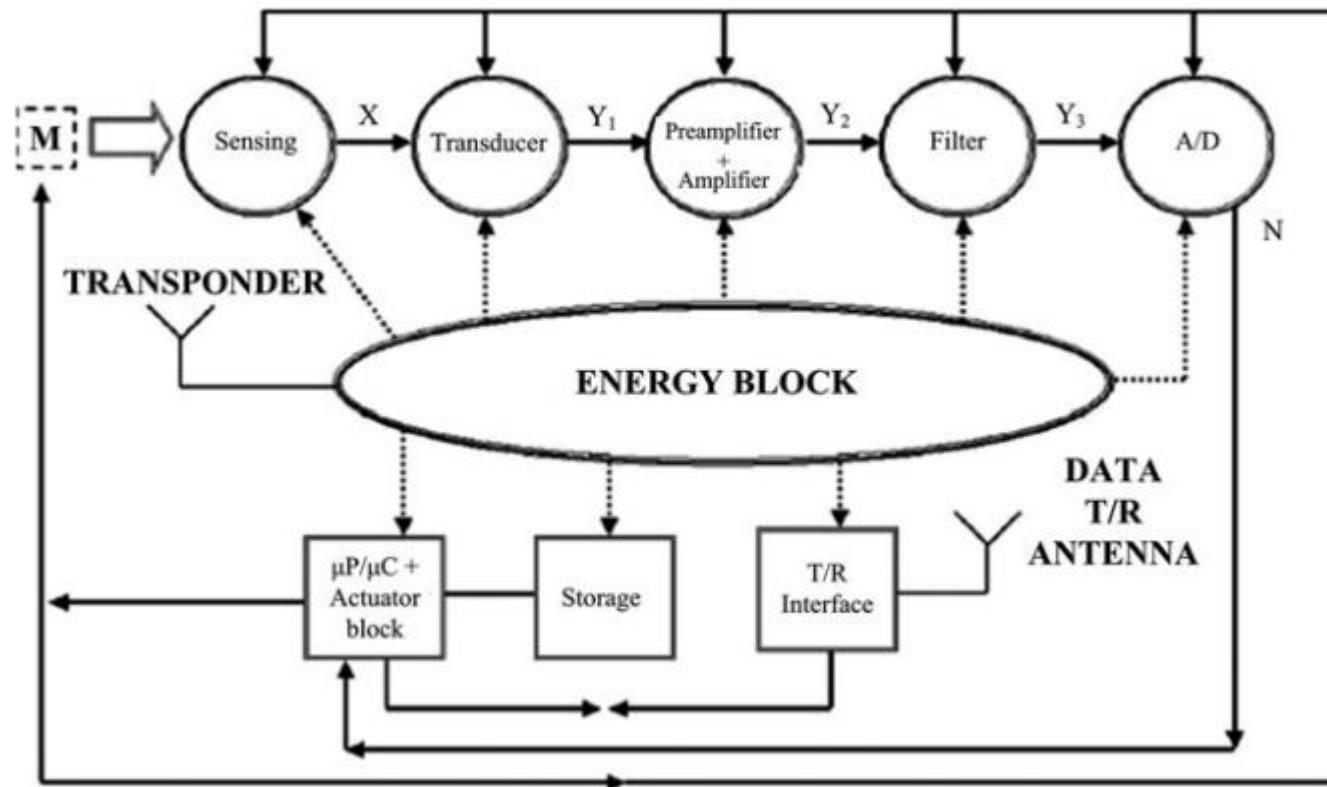


Figure 1.7 Sensor technology migration path.

SENSOR INTERFACING



For example, a gas-sensing microsystem typically consists of an array of gas-sensors, a temperature control circuit, an electronic readout block and a data processor. In order to develop a really portable device, the system has to be stand-alone, i.e. has to be able to operate without the aid of any laboratory instrument, while sensors, implemented with silicon based technologies, can detect different physical and chemical quantities with acceptable selectivity, sensitivity and resolution.



Block scheme of a smart sensor system

The sensor response (i.e., the output signal of the sensor) is typically analog and this is why it is said that “the real world is analog”. However, sometimes it can be also convenient to process the information in the digital electrical domain. In this case, a digital electronic system is required for converting the analog sensor response into a suitable digital electrical signal. **This is what electronic interfaces perform:**

they are circuits that convert the sensor responses into proper electric signals easy to be processed. If these interfaces are particularly “intelligent”, including special functions such as auto-calibration, sensor biasing, working temperature control, etc., they can be considered “smart”.

A smart system (if miniaturized, named microsystem) requires, all together, sensors (if miniaturized, named microsensors), actuators and suitable electronic interfaces.

More specifically, the sensor interface is an electronic circuit which allows to read-out the information coming from the signal generated by a sensor, providing a suitable output signal simple to display or to elaborate.

Signal Conditioning Circuits and Interface Devices

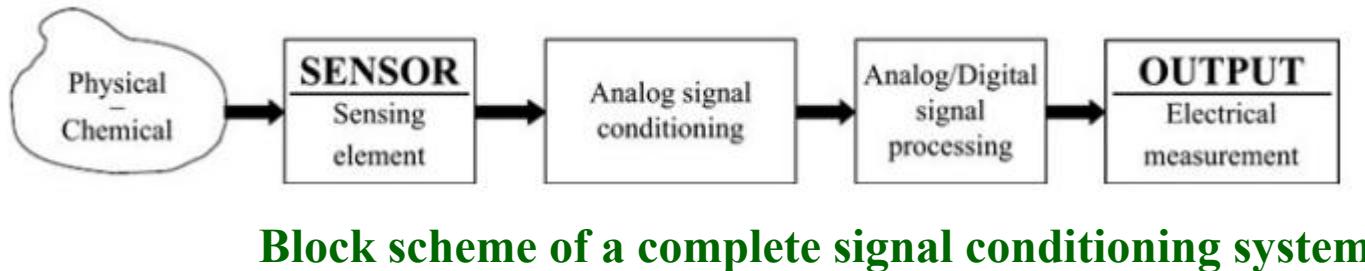
In order to get information from a sensor into a computer, the signal from the sensor must first be sent to an interface device of some form and from there to the computer. However, in order to be useful to the interface device, the signal from the sensor must often undergo some form of conditioning. Almost all interface devices designed to allow interfacing of sensors to computers are designed to accept a voltage signal in the range of 0 to 5 volts and to digitize this.

This means that the aim of the signal conditioning circuit is to take whatever output is available from the sensor, whether voltage or resistance, and convert it to a 0 to 5 volt signal. This process generally involves a combination of one or more simple processes: converting a resistance to a voltage, dividing a voltage, amplifying a voltage and shifting a voltage.

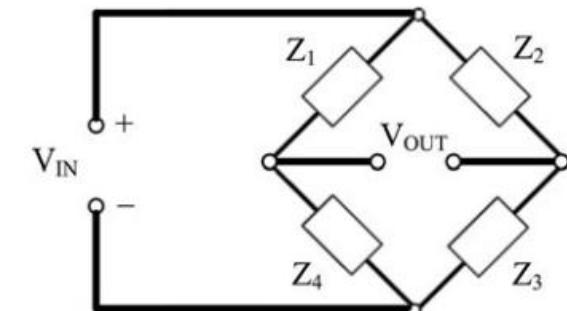
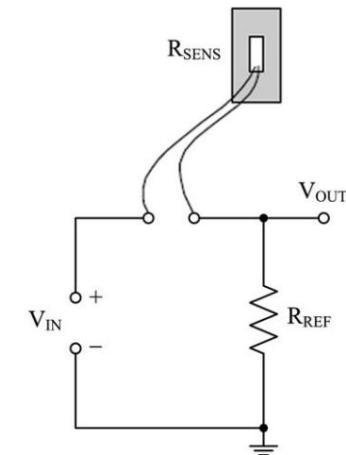
Basic Sensor Interfacing Techniques: Introduction to Signal Conditioning:

A signal conditioning system (or, in other words, electronic interface, read-out circuit, front-end, etc.) takes the output from a sensing element and converts it into a more suitable form for further processing (e.g., amplification, analog-digital conversion, frequency-voltage conversion, etc.), as described in Fig

Fig. 1



The simpler interface circuits, often utilized, for example, as basic signal conditioning stages in resistive sensors, are the voltage divider, shown in Fig. 1, and its differential version, the Wheatstone bridge where V_{IN} is the supply voltage and one (or more) of the bridge elements (impedances) are the sensors. These simple basic solutions are able to perform, more in general, a conversion from an impedance (e.g., a resistance) variation into a voltage one



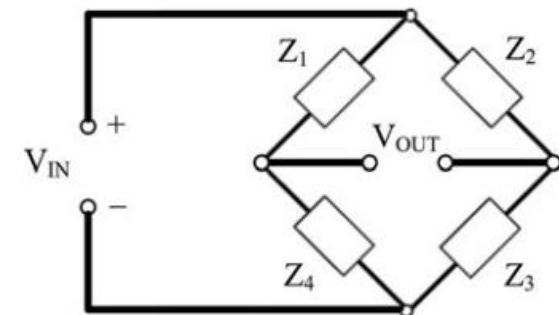
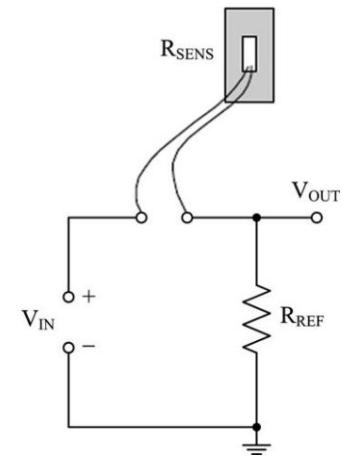
Resistive Sensors Basic Interfacing

When the sensor electrical parameter can be modelled by a resistance that, in particular, varies into a reduced range, not more than two to three decades, a resistive voltage divider circuit, operating a Resistance-to-Voltage (R-V) conversion (as yet shown in Fig.1), can be utilized as simple resistive sensor interface circuit. Typically, it applies a constant voltage so to measure the change of conductivity of the resistive sensing element

Another very simple interfacing circuit for resistive sensors, varying into a reduced range, can be implemented by the well-known Wheatstone bridge which operates also an R-V conversion.

This circuit configuration represents the “fully-differential” version of the basic voltage divider and shows its same sensitivity. In this case, one of the four resistances is the resistive sensor whose sensing element varies when an external physical or chemical phenomenon occurs. The main drawback of this kind of resistive sensor interface is in its unsettable and low sensitivity, only dependent on the total supply voltage.

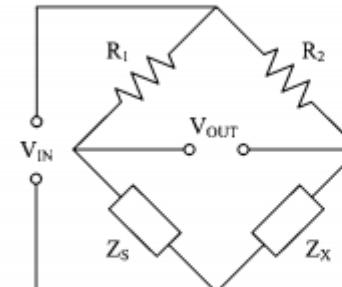
Fig.1



if larger variations of sensor resistive values happen, we can employ a Resistance-to-Time (R-T) conversion, which can be also considered as a Resistance-to-frequency (R-f) conversion when the “time” (period) is related to a periodic waveform. The R-T based interfaces exploit the easiness of measuring time intervals over a wide range of variation. As a consequence, no more scaling factor systems are needed. Typically, an R-T basic scheme is based on an oscillator architecture which exploits the sensor to be excited by a switched voltage (the AC excitation voltage). In this case, the simpler electronic interface which operates An R-T (or R-f) conversion can be implemented by an OA in an astable multivibrator configuration.

Capacitive Sensors Basic Interfacing

The typical simplest way to measure a capacitance is to convert it (or its variation) into a suitable voltage level, performing the so-called Capacitance-to-Voltage (C-V) conversion. This can be simply done by one of the bridge configurations shown in Figure, once that all the other passive components are known or can be accurately measured.



Comparison with series constants

Measures L or C

Balance equations:

$$R_x = R_s \frac{R_2}{R_1}$$

If inductive:

$$L_x = L_s \frac{R_2}{R_1}$$

If capacitive:

$$C_x = C_s \frac{R_1}{R_2}$$

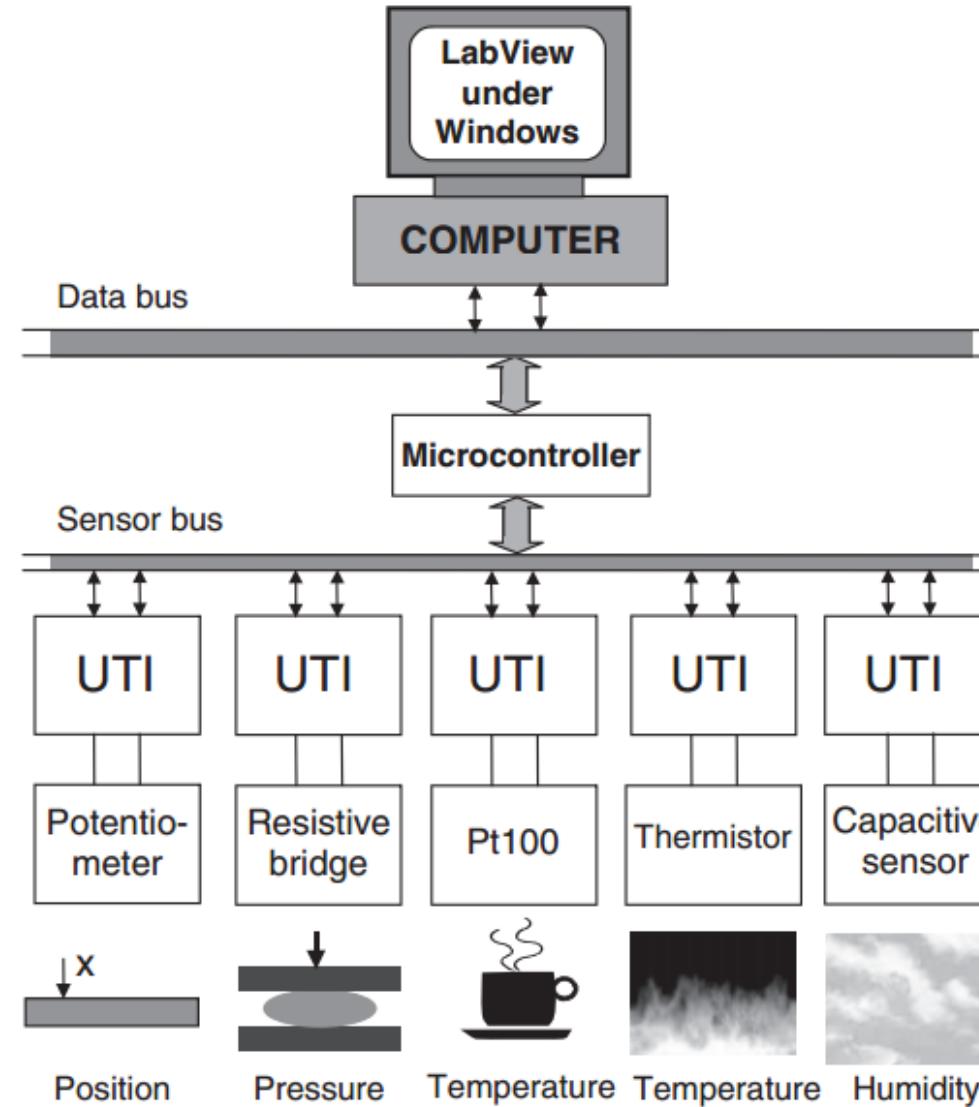
Nevertheless, the main problem related to all these interface solutions concerns the detection of either very low capacitance values or its small variations. In this sense, the proper design of a suitable read-out circuit, which has to be able to provide the smallest parasitic capacitances at its terminals, is another important task, while a special consideration for shielding to still reduce parasitic capacitances of the electronic front-end, which is essential to have suitable performances, has to be also done avoiding the need for large connectors. Therefore, the key aspect of the problem is related to the sensing system, where the sensitivity to parasitic elements, interconnection wires and noise has to be the lowest possible.



Universal transducer interface



Universal transducer interface



A common disadvantage of many digital interfaces is that many analog sensors cannot be interfaced in a low-cost way. In order to eliminate this disadvantage, the Universal Transducer Interface (UTI) circuit for different kinds of analog sensors was designed in the Electronics Research Laboratory, Delft University of Technology and is manufactured by Smartec.

Sensing elements can be directly connected to the UTI without extra electronics. Only a single reference element, of the same kind as the sensor, is required. The UTI is intended for the low-cost market, for systems where this interface provides an intermediate function between low-cost sensor elements on the one hand and the microcontroller on the other.

Some typical applications of the UTI on the measurement of some physical quantities, such as temperature, position, displacement, humidity, pressure, acceleration, etc..

The UTI is a complete front end for many types of passive sensors, such as resistive, resistive bridge and capacitive sensors.

The UTI converts low-level signals from a sensor to a period modulated microcontroller-compatible signal.

A common disadvantage of many digital interfaces is that many analog sensors cannot be interfaced in a low-cost way.

In order to eliminate this disadvantage, the Universal Transducer Interface (UTI) circuit for different kinds of analog sensors was designed in the Electronics Research Laboratory, Delft University of Technology

When the sensor signal is converted to the time domain, using a period modulator in the UTI, then the micro-controllers do not require a built-in ADC.

Sensing elements can be directly connected to the UTI without extra electronics.

The UTI is intended for the low-cost market, for systems where this interface provides an intermediate function between low-cost sensor elements on the one hand and the microcontroller on the other.

THREE-SIGNAL TECHNIQUE

The three-signal technique is a technique to eliminate the effects of unknown offset and unknown gain in a linear system.

In order to apply this technique, in addition to the measurement of the sensor signal, two reference signals are required to be measured in an identical way.

The signal conversion is carried out according to the linear law:

$$M_i = k \cdot S_i + M_{off},$$

where S_i is the analog output sensor's signal, k and M_{off} are measuring converter parameters that directly influence conversion error.

In order to achieve high accuracy the UTI operates in auto-calibration.

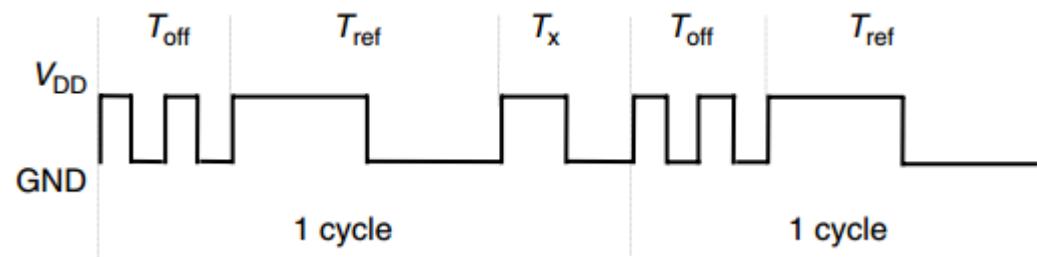
The given method allows removing the error caused by the above parameters.

The essence of this method of measurement consists of the measurement of three signals: $S_1 = 0$, $S_2 = S_{\text{ref}}$ and $S_3 = S_x$ (zero, reference and measurand) during one cycle:

$$M_1 = M_{\text{off}},$$

$$M_{\text{ref}} = k \cdot S_{\text{ref}} + M_{\text{off}},$$

$$M_x = k \cdot S_x + M_{\text{off}}.$$

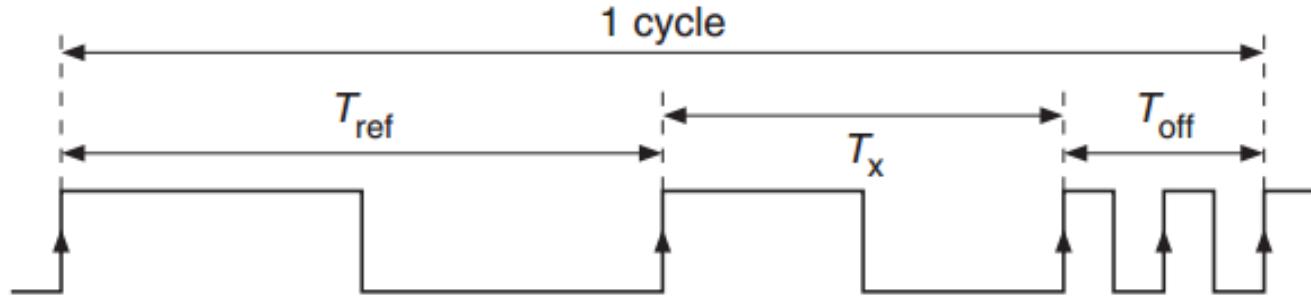


$$M = \frac{M_x - M_{\text{off}}}{M_{\text{ref}} - M_{\text{off}}} = \frac{Ex}{E_{\text{ref}}}$$

The output signal of the UTI has three informative component

When the system is linear, then in this ratio the influence of the unknown offset M_{off} and the unknown gain k of the measurement system is eliminated and the calculated value of **M** represents the ratio between the values of the (unknown) sensor element and the (known) reference element.

Period-modulated output signal of UTI for 3-phase mode



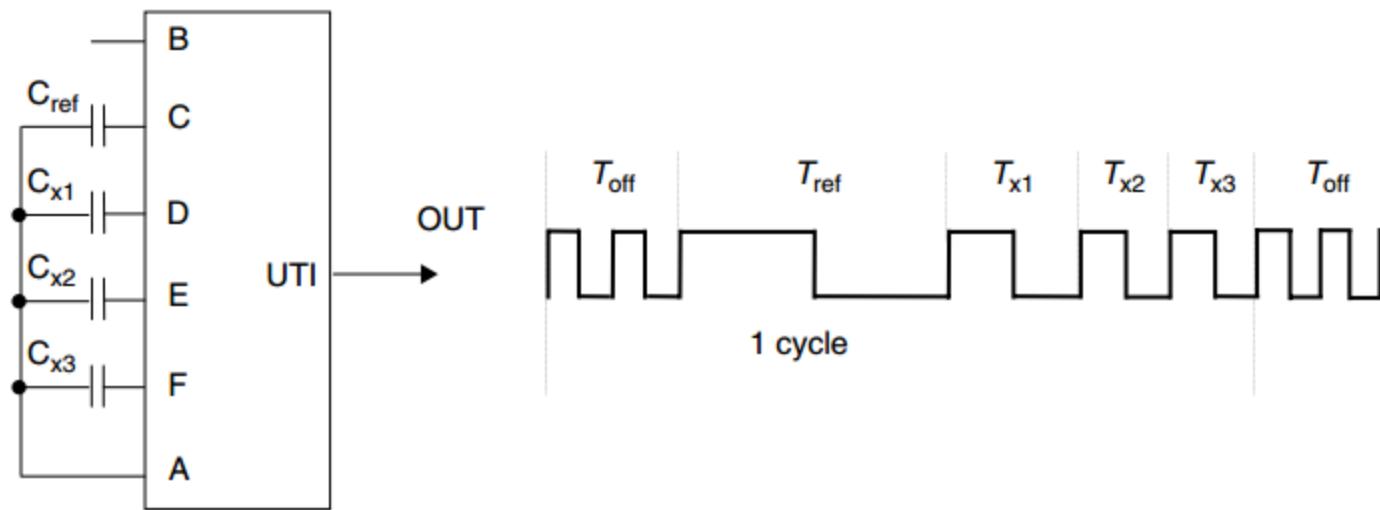
Output signal of the UTI interface for the case where a complete measurement cycle takes three phases. These phases correspond to three time intervals and four periods

During the first phase Toff, the offset of the complete system is measured. During the second phase Tref, the reference signal is measured and during the last phase Tx , the signal itself is measured. The duration of each phase is proportional to the signal that is measured during that phase. The result is the ratio:

$$M = \frac{T_x - T_{\text{off}}}{T_{\text{ref}} - T_{\text{off}}} = \frac{C_x}{C_{\text{ref}}} \text{ or } = \frac{R_x}{R_{\text{ref}}} \text{ or } = \frac{V_x}{V_{\text{ref}}}$$

It is possible to measure multiple capacitances as well as capacitances from 300 pF to 2 pF. Possible applications are liquid level sensors, humidity, position, rotation, movement, displacement sensors. The result can be calculated as follows:

$$M_i = \frac{T_{xi} - T_{\text{off}}}{T_{\text{ref}} - T_{\text{off}}} = \frac{C_{xi}}{C_{\text{ref}}}.$$



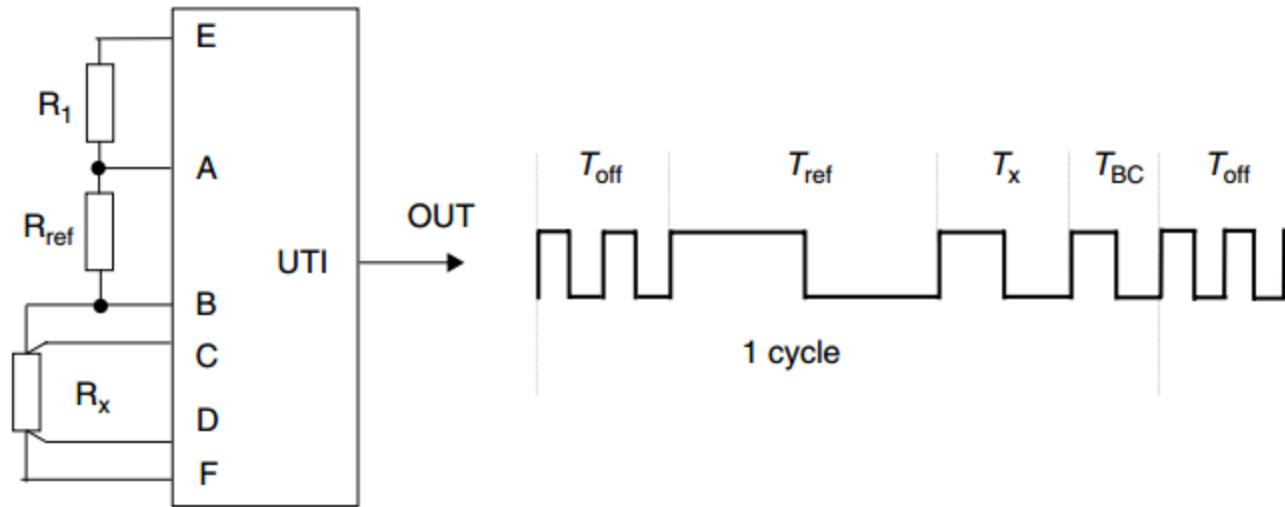
**Measurement circuit for small capacitance
and output signal of UTI**

The connection of the Pt100 (Rx) to the UTI and the time diagram is shown in Figure . The result can be calculated according to the formula:

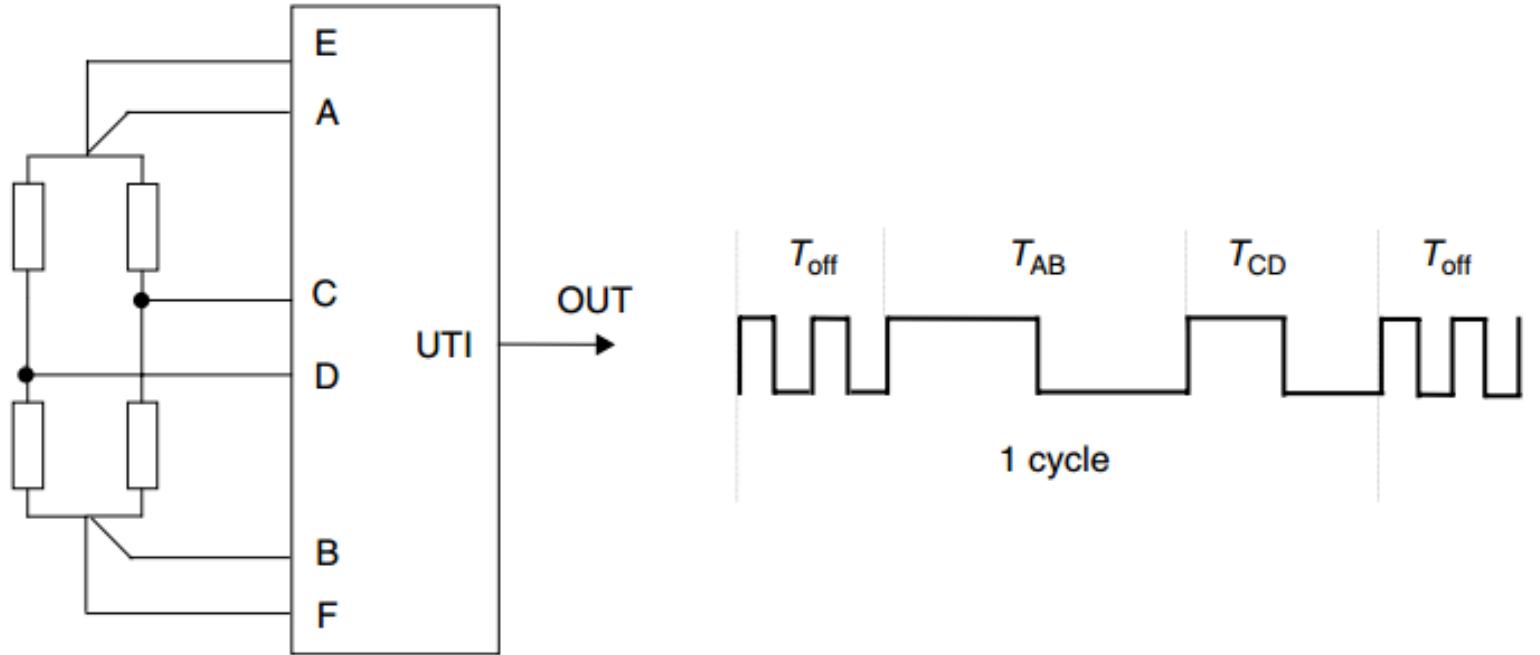
$$M = \frac{T_x - T_{\text{off}}}{T_{\text{ref}} - T_{\text{off}}} = \frac{R_x}{R_{\text{ref}}}$$

The connection of the resistive bridge to the UTI shown in Figure.





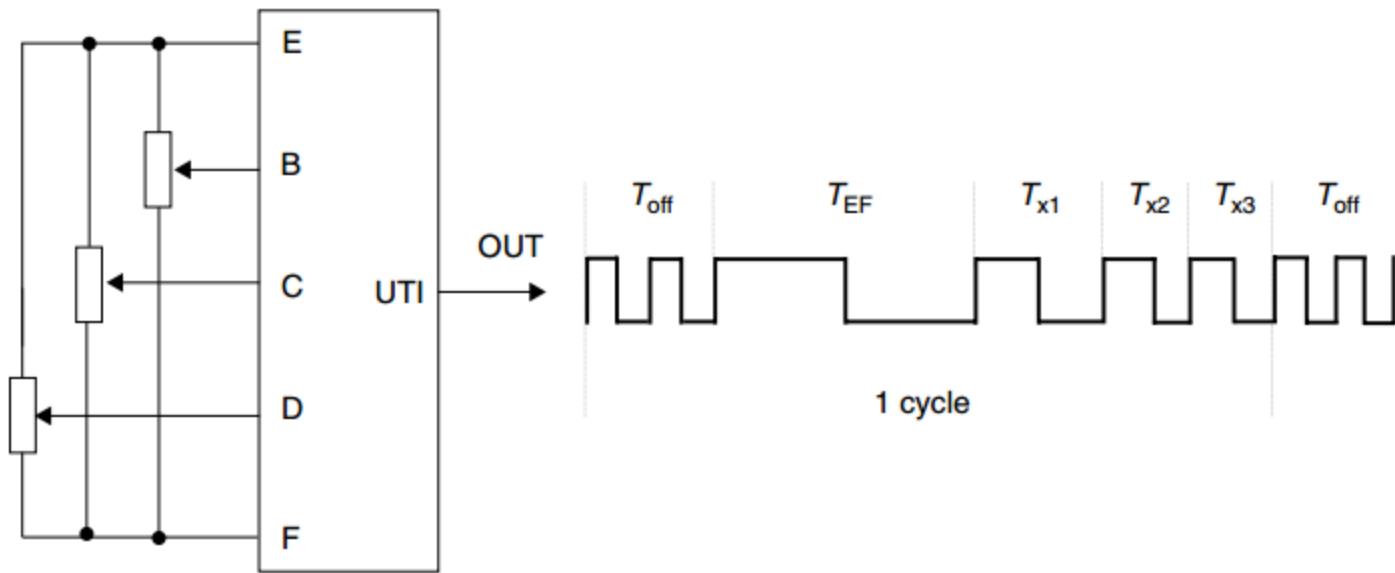
Measurement circuit for thermistor and output signal of UTI



Measurement circuit for resistive bridge and output signal of UTI

Possible applications are pressure sensors and accelerometers.
The measuring result can be calculated as follows:

$$M = \frac{1}{32} \cdot \frac{T_{CD} - T_{off}}{T_{AB} - T_{off}} = \frac{V_{CD}}{V_{AB}}$$

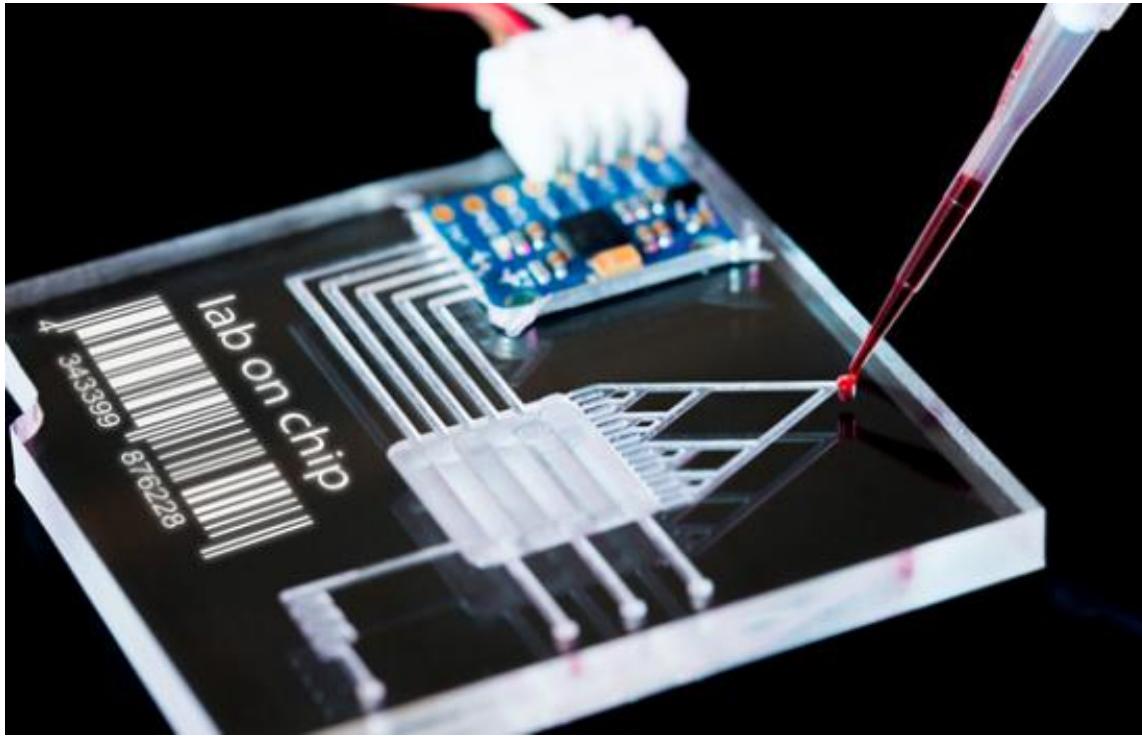


The UTI is ideal for use in smart microcontroller-based systems. The microcontroller is used to measure the period-modulated signal from the UTI, to process the measured data and to output digital data to a central computer via the communication interface.

MICROTECHNOLOGIES - MICROSYSTEMS ENGINEERING



MICROTECHNOLOGIES - MICROSYSTEMS ENGINEERING



Systems development: methods and tools,

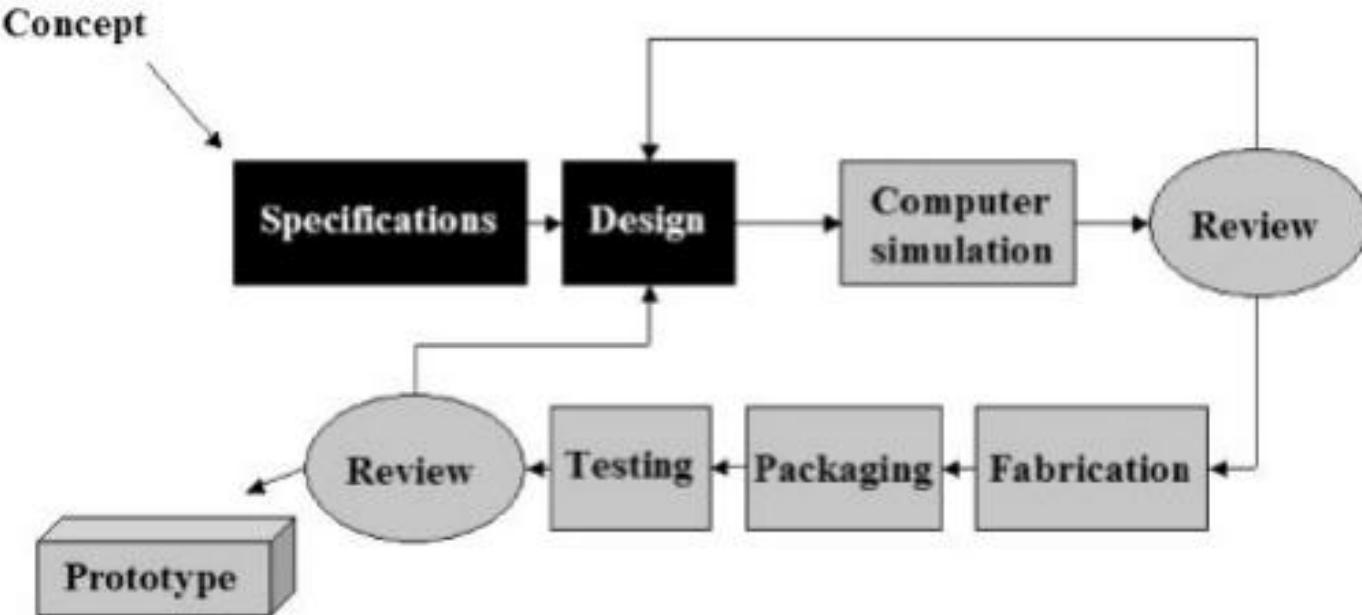
1.2.2 Systems-Development: Methods and Tools

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

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

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

constructive and connective techniques



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

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

Major ~~bottlenecks exist at present in the area for non-electronic components of microsystems engineering ;~~

The following basic developments are needed :

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

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

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

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

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

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

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

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

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

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

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

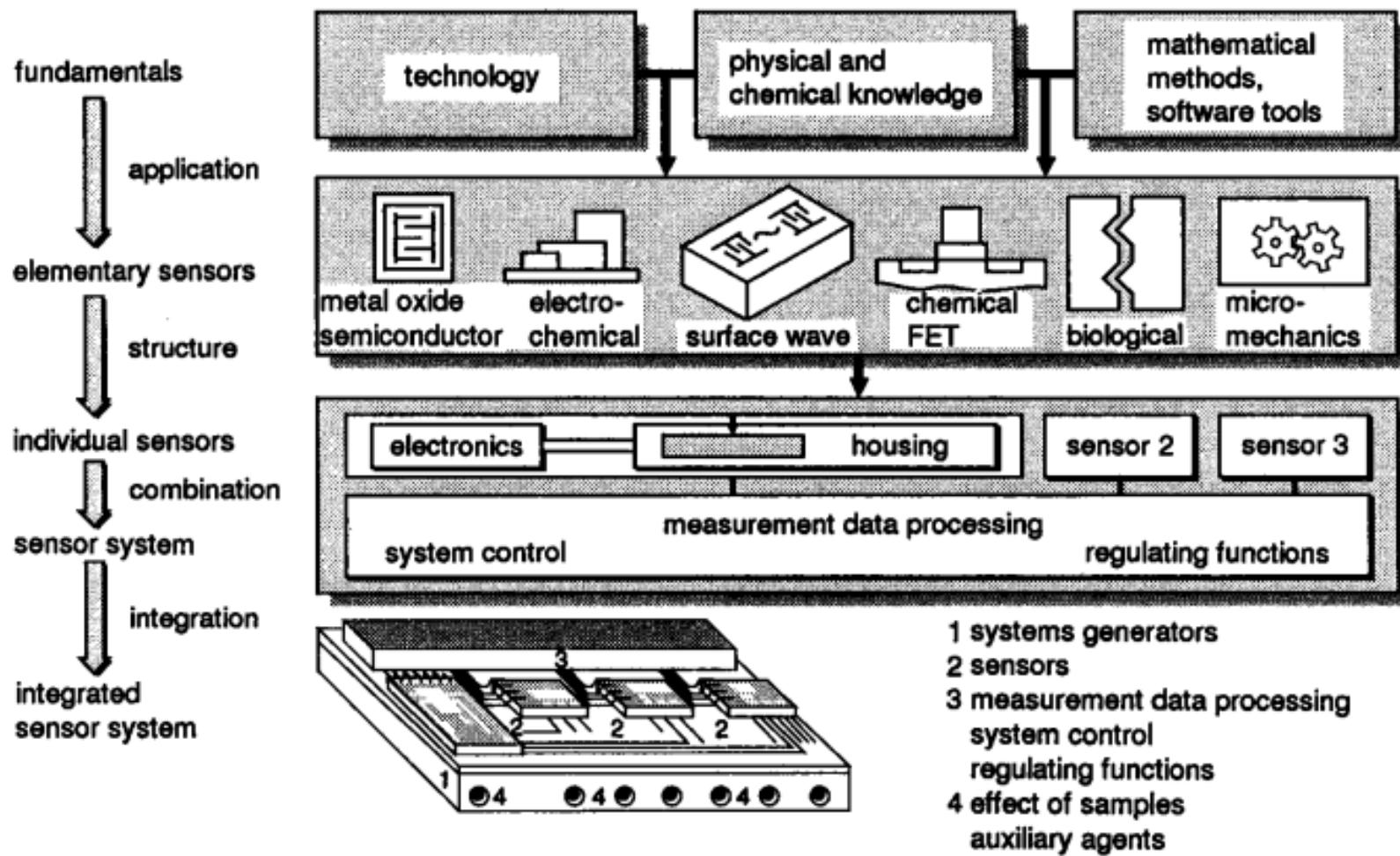


Figure 1-1. Sensors in microsystems engineering: from the elementary sensor to the bus-compatible sensor system.

UNIT -5

Fundamentals of MEMS fabrication: introduction and description of basic processes

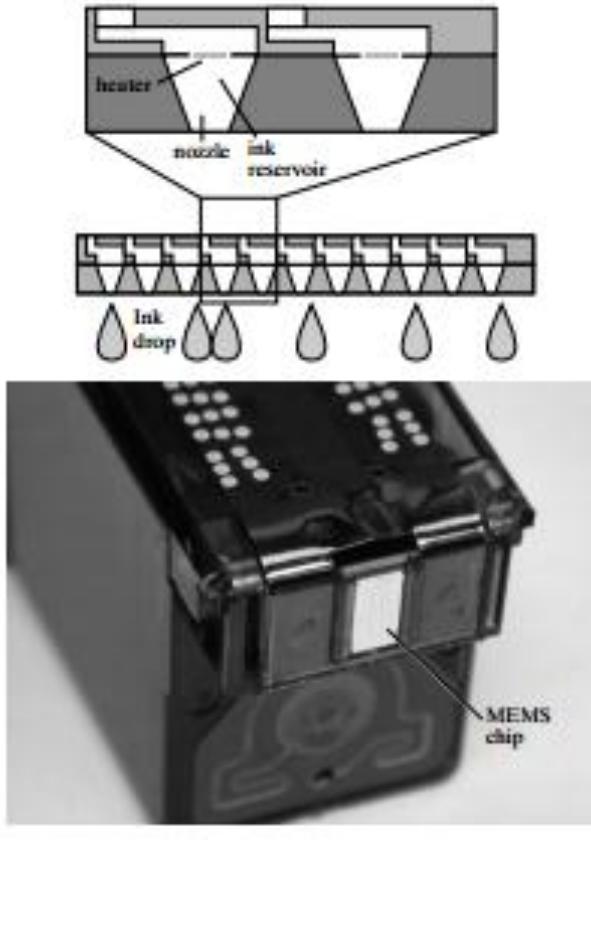


INTRODUCTION

Hewlett-Packard pioneered the technology of silicon micro-machined ink jet printer nozzles in 1978. Arrays of ink jet nozzles eject tiny ink droplets (“drop on demand”), upon expansion of liquid volume by thermally generated bubbles . The collapse of the bubble draws more ink into the ink cavity for the next firing. Color ink jet printing is achieved by dropping primary subtractive color dyes—cyan, magenta, and yellow (CMY).

Silicon micromachining technology played an enabling role for the ink jet printing technology. Using silicon micromachining, ink-ejection nozzles can be made extremely small and densely populated, an important trait for realizing high printing resolution and sharp contrast. Small-volume cavities with equally small heaters mean rapid temperature rise (during ink ejection) and fall, allowing ink jet printing to reach appreciable speed.

In 1995, the number of nozzles per cartridge has increased to 300 while the average weight of ink droplet is only 40 ng. In 2004, ink jet heads are based on a variety of principles, including thermal, piezoelectric, and electrostatic forces. The volume of each drop is on the order of 10 pl



Close-up view of a commercial inkjet printer head, and the silicon chip consisting of many nozzles. ~~Integrated circuits on chips control nozzle firing.~~

In the late 1980s, researchers in the nascent field called micromachining mainly focused on the use of silicon—either bulk silicon substrate (single crystalline silicon) or thin film silicon (polycrystalline silicon). These two forms of silicon were readily accessible as they were used heavily in the integrated circuit industry: bulk silicon is used as the substrate of circuitry, while polycrystalline silicon is used for making transistor gates. Three-dimensional mechanical structures, such as suspended cantilevers or membranes, can be made out of bulk silicon or thin film silicon.

Fundamentals of MEMS fabrication: introduction and description of basic processes

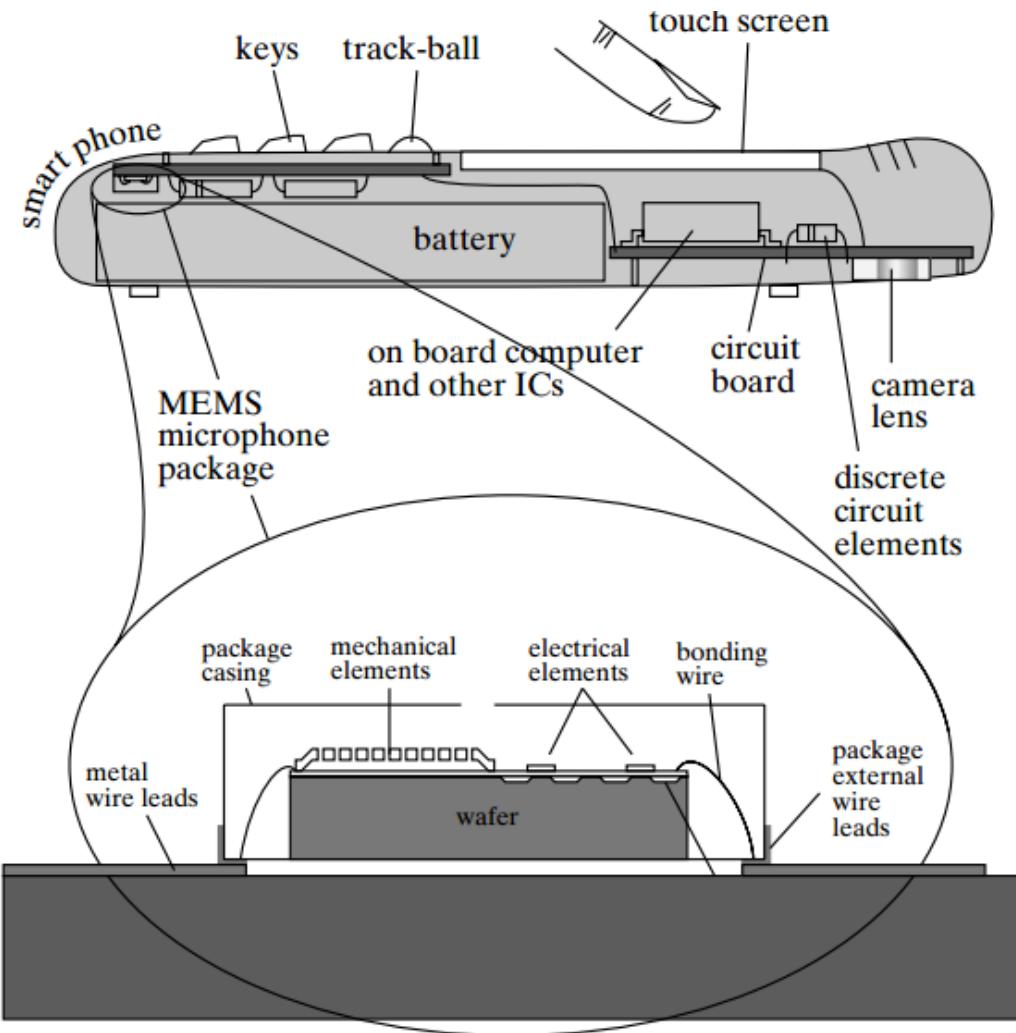
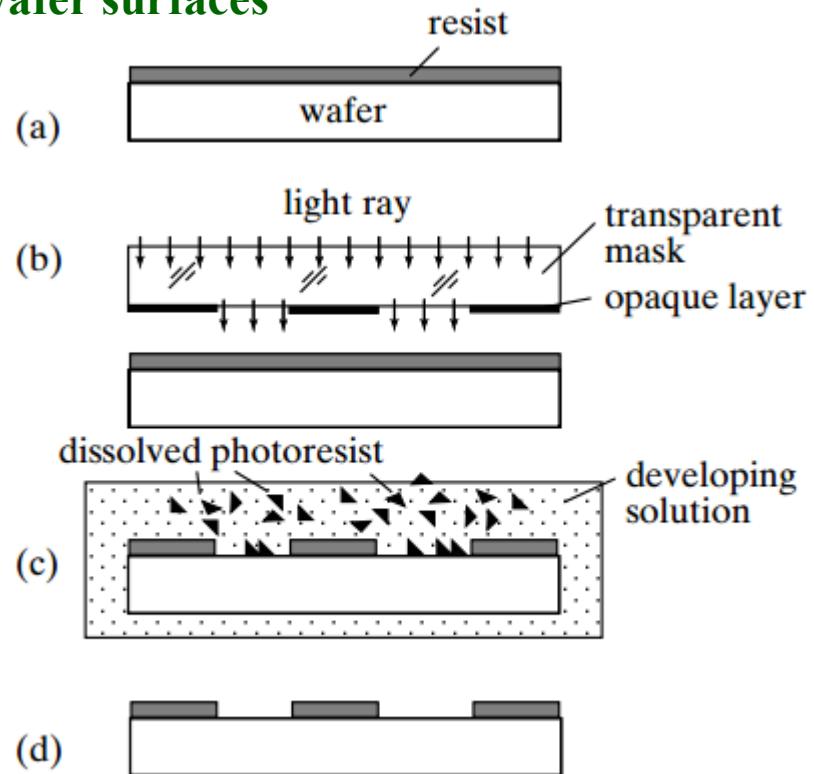
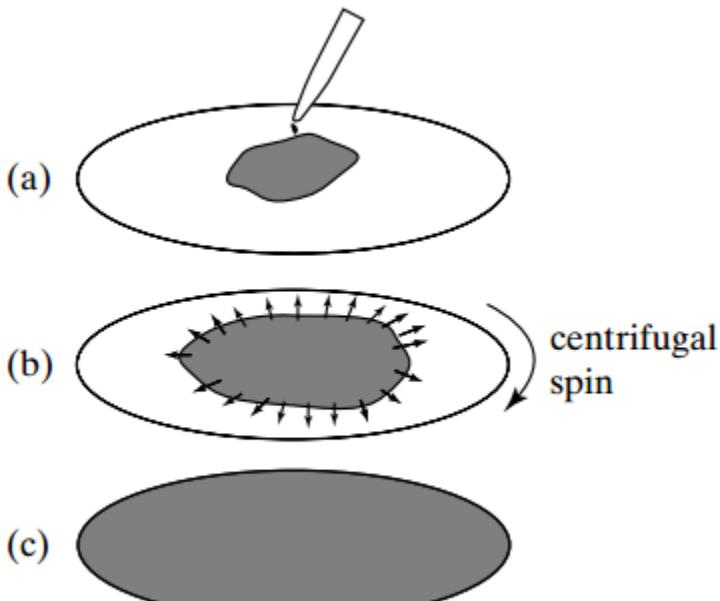


Diagram showing the interior elements of a smart phone, a pack-aged sensor chip, and interior components of the chip package.

FREQUENTLY USED MICROFABRICATION PROCESSES

Photolithography

to produce fine features on wafer surfaces

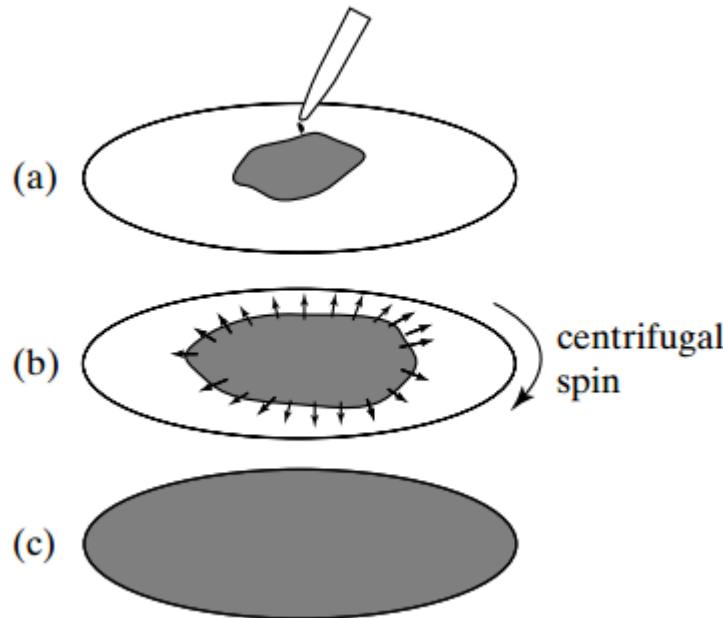


Process steps of
photoresist spin coating

Process flow for patterning
photoresist with a photomask

photoresist spin coating

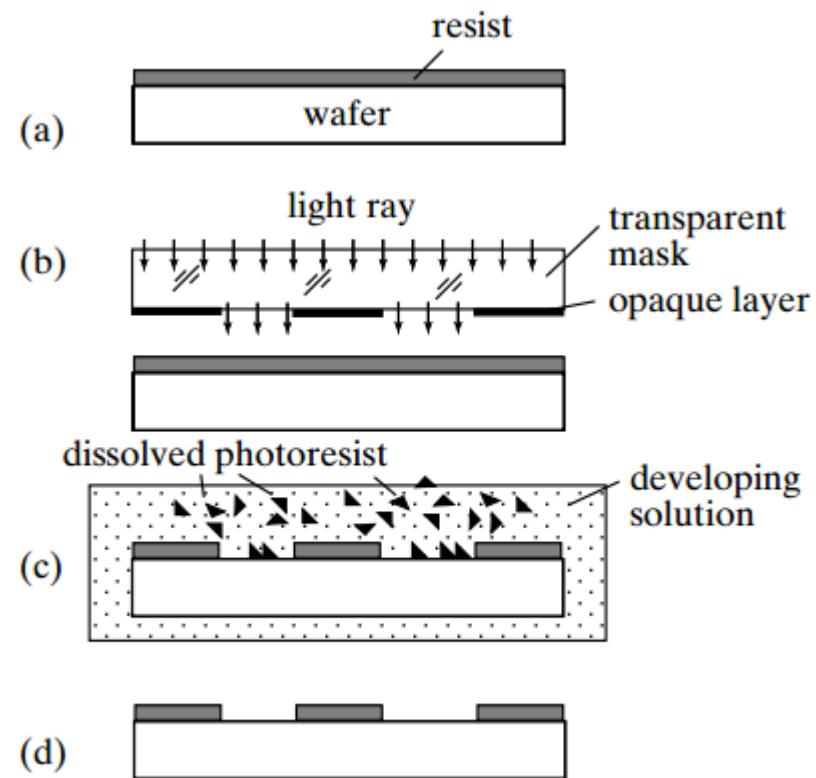
Photoresist is applied to the wafer surface by spin coating, or spinning. In this operation, the substrate is placed on a vacuum chuck which holds the wafer in place. Next, a quantity of resist is dispensed from a nozzle while the chuck is rotated at a slow speed. Finally, the speed is ramped up to several thousand rpm which causes all but a thin layer of resist (generally about 1 to 2 μm thick) to be thrown off. By carefully controlling the rotation rate and the resist viscosity, uniform layers with reproducible thickness can be obtained.



The wafers are then placed in an exposure tool and aligned so that the patterns projected onto the substrate during the exposure step are registered properly with previously defined patterns. If this is the first mask, the patterns are aligned to the wafer flat.

The resist is exposed to light from an intense source as it passes through the clear areas of the mask or reticle. The light and dark patterns on the mask are thus transferred to the resist. The effect of the light is to change the solubility of the resist in a solution called the developer.

Resists that increase in solubility upon exposure to light are called positive resists and are nearly universally used. (8) Because dark areas on the mask correspond to areas of resist which remain after development, the "polarity" of the mask pattern is duplicated in the positive resist.



The exposed wafer is then developed by either spraying or dipping the substrate in an appropriate solvent. Most positive resists are developed in proprietary alkaline solutions (e.g., dilute NaOH or KOH). The exposure process increases the solubility of the resist by a factor of about one thousand, so unexposed areas are largely untouched while the exposed portions are quickly washed away. However, since the solubility of unexposed resist is finite, careful control of the development process is essential in controlling the ~~pattern dimensions~~.

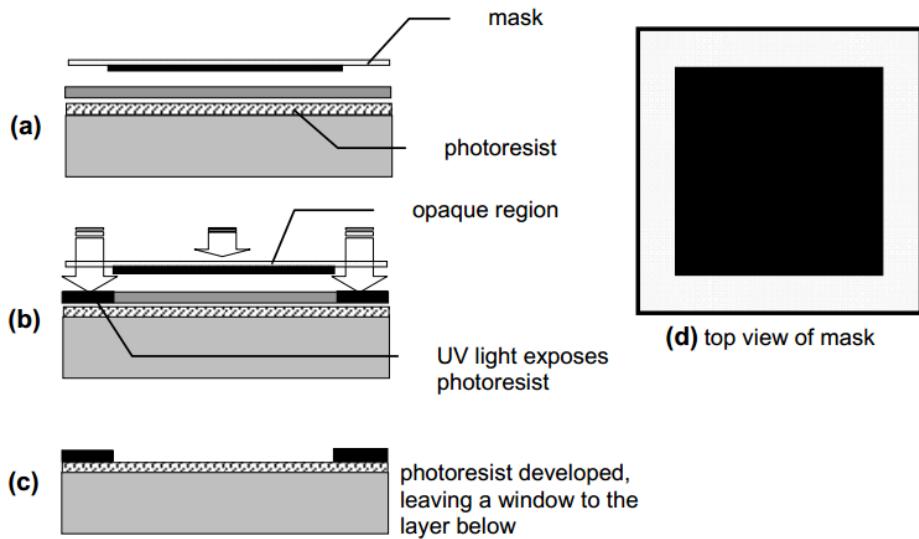
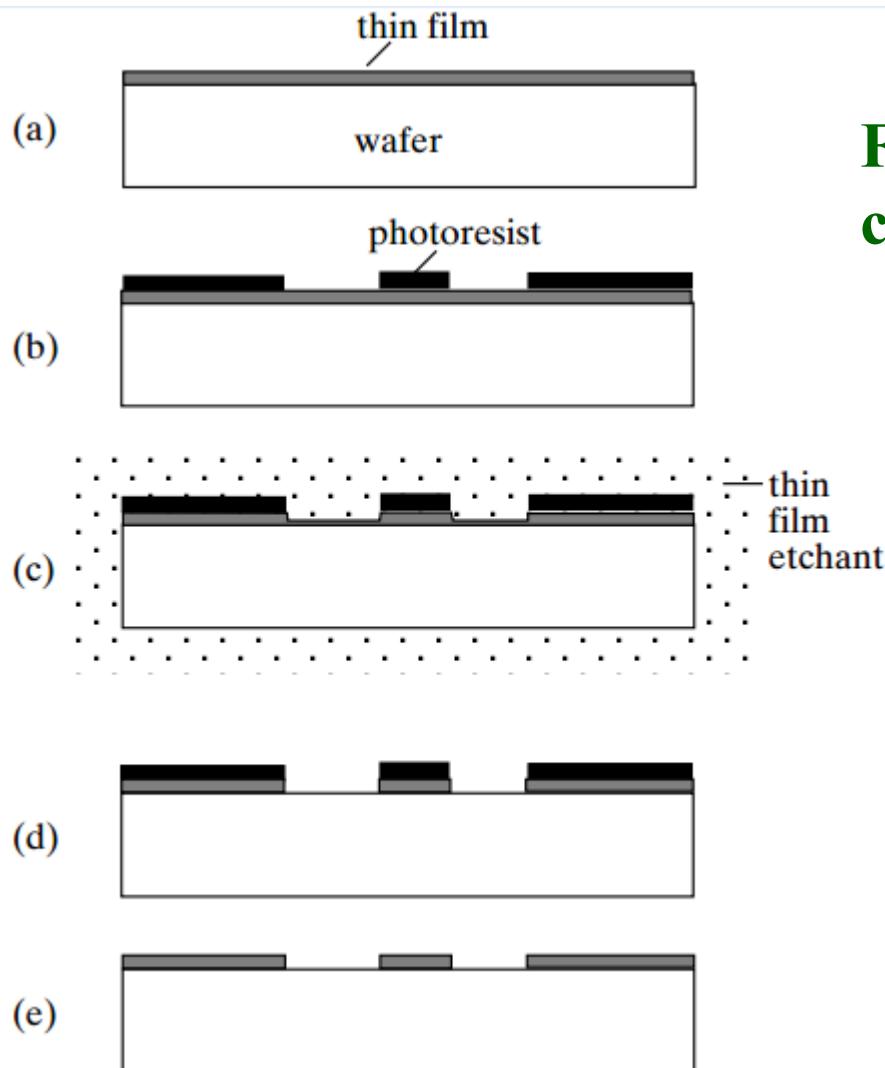


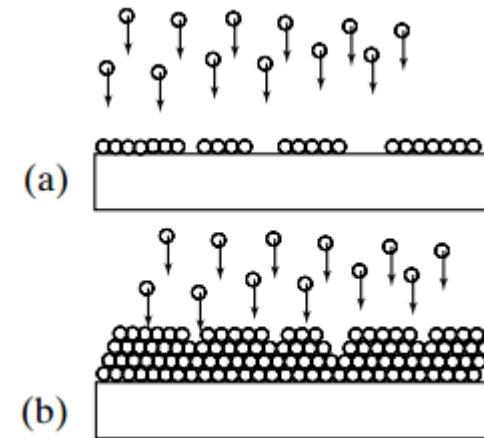
Fig. 3.1. Steps in a simple photolithographic process

Creating and transferring patterns—Photolithography

Thin Film Deposition



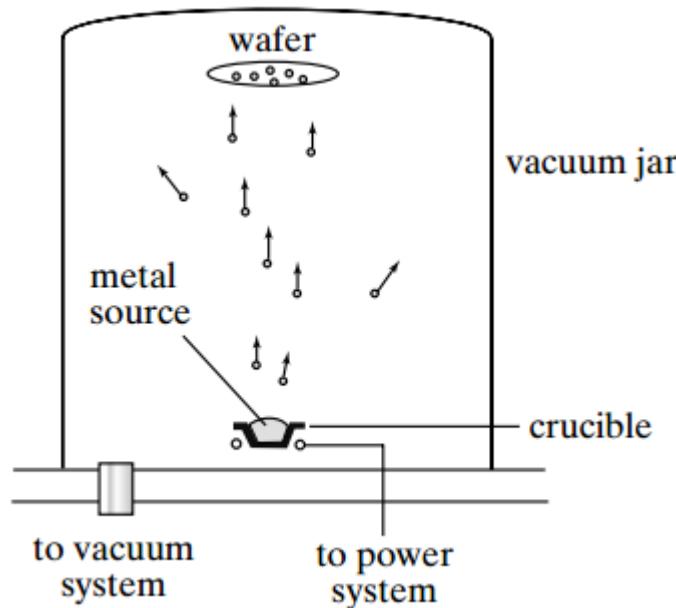
Functional materials, conductors and insulators



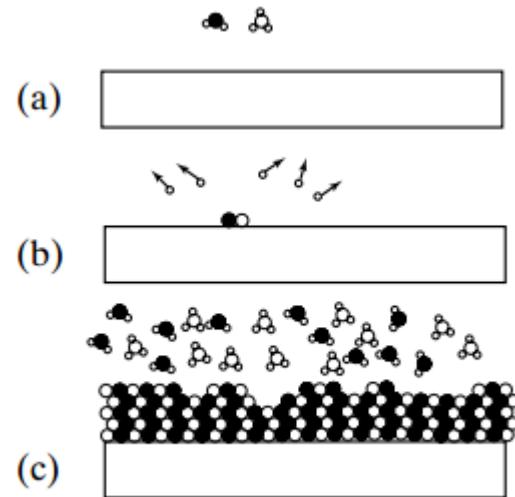
A process of evaporating a
thin film

Process of lithographically patterning a thin
film using
the photoresist as a mask

Thin Film Deposition



A schematic diagram of the metal evaporation equipment.



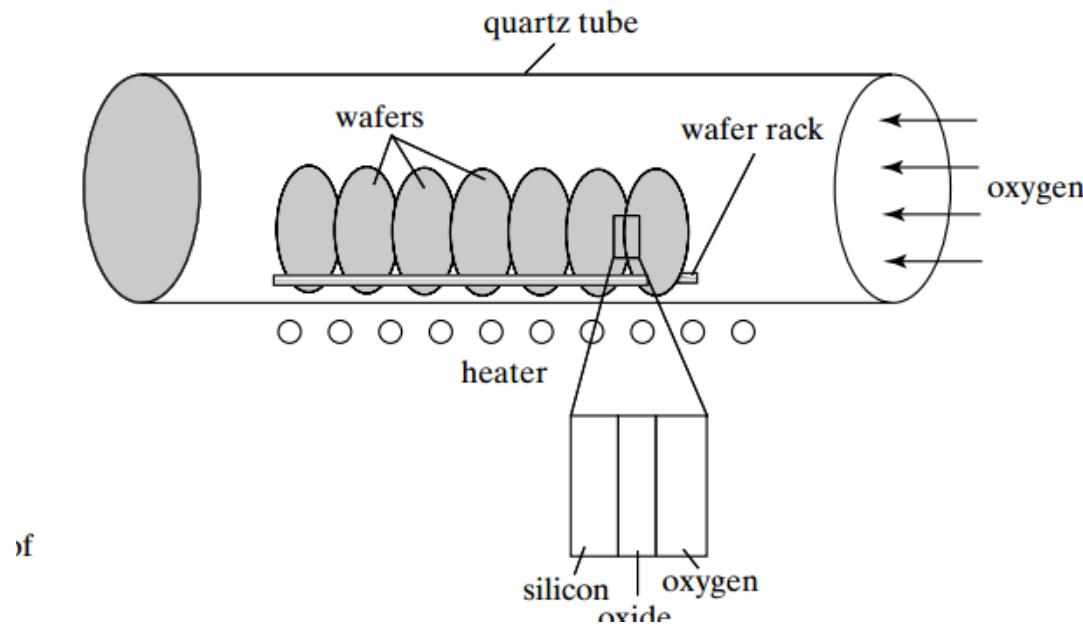
**A process of chemical reactive deposition
(e.g., chemical vapor deposition)**

Evaporation and sputtering are largely physical deposition processes. In both types of processes the material to be deposited starts out as a solid and is transported to the substrate surface where a film is slowly built up. In evaporation, the transport takes place by thermally converting the solid into a vapor. In sputtering, atoms or molecules of the desired material are removed (from the "target") by energetic ions created in a glow discharge.

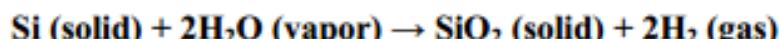
The configuration of an evaporation system is shown in Figure. The system consists of a large bell jar evacuated to a low base pressure, generally less than 10^{-7} torr. (The lower the base pressure, the fewer the impurities incorporated into the growing film.) The substrates are placed near the top of the vacuum chamber, with the side receiving the film face down. The material to be deposited is converted into a vapor, which condenses onto the substrates. Because the base pressure in the system is low, the mean free path of the evaporated atoms is high. This means that collisions between source atoms and background gas molecules are rare; thus there is little scattering of the vapor as it travels from the crucible to the wafers.

The result is that evaporated films are not particularly conformal. The cloud of atoms striking the substrate can be shadowed by features on the surface of the substrate. To prevent this, the substrates are mounted on a planetary system which rotates them about their own axis and rotates the hemispherical planetary assembly about its own axis. Also, since the source emits vapor over a well-defined solid angle, the substrates must be placed at a considerable distance from the source to obtain reasonable wafer-to-wafer uniformity.

Thermal oxidation



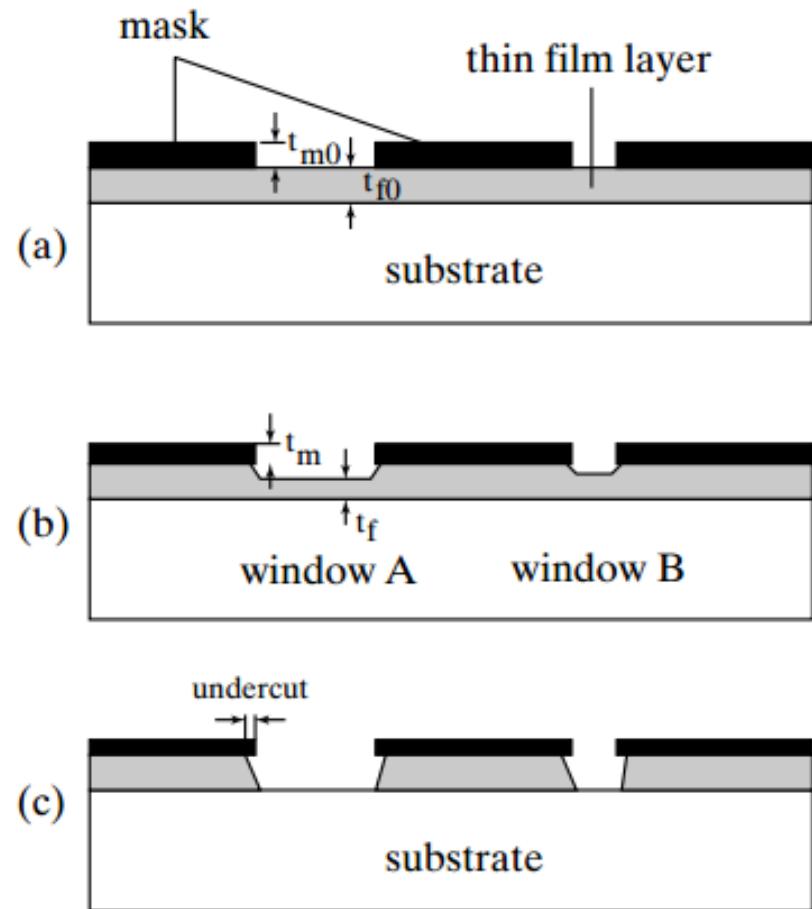
Wet oxidation uses steam or water vapor to form SiO_2 .

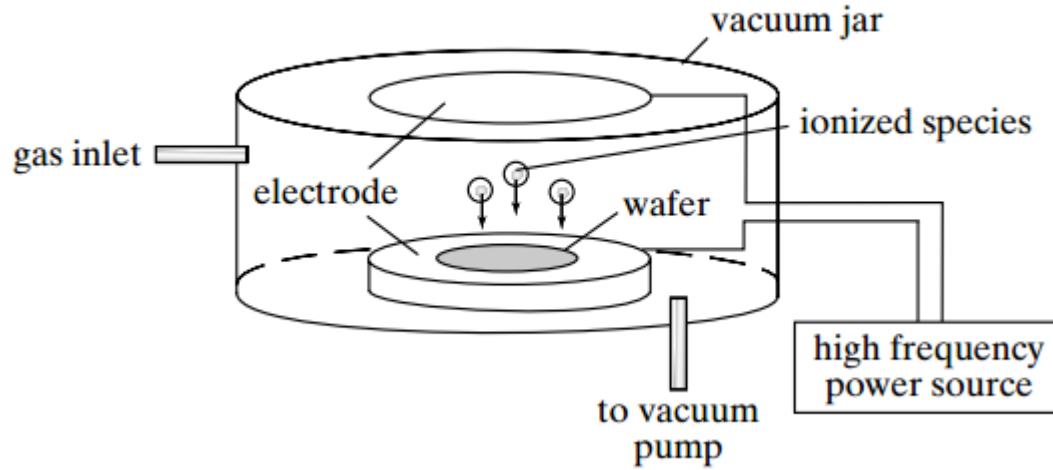


Two oxidation methods are used in thermal oxidation: dry and wet oxidation.
Dry oxidation uses In both processes, dry and wet, the process temperature affects the rate of oxidation (the rate at which the SiO_2 layer grows). The higher the temperature, the greater the oxidation rate (amount of oxide growth / time).
Also, wet oxidation has a higher oxidation rate than dry oxidation at any given temperature.

Wet Etching

Removing materials by wet chemical reaction is common. It is used for removing metal, di-electrics, semiconductors, polymers, and functional materials. The selectivity of etching against masking materials, substrates and the target material is a crucial issue for MEMS process design. Important performance characteristics include etch rate, temperature, and uniformity.



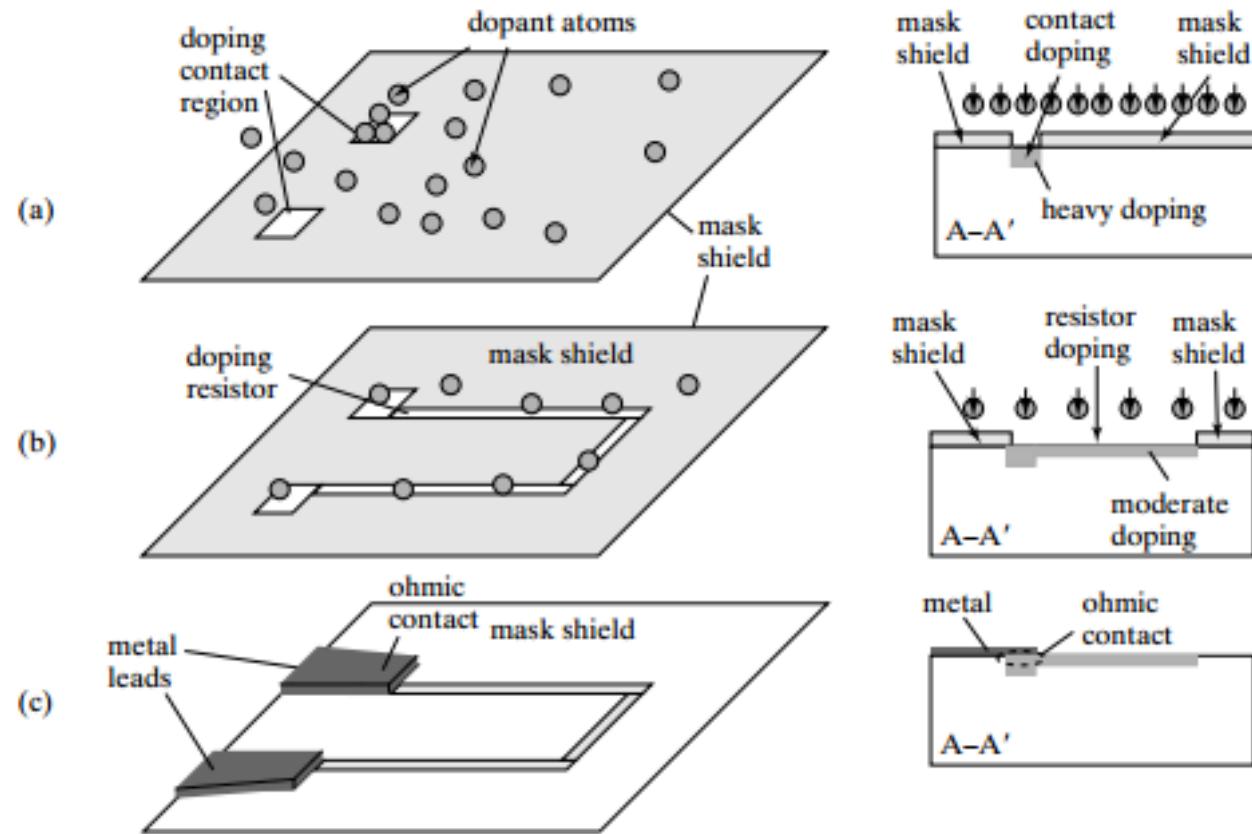
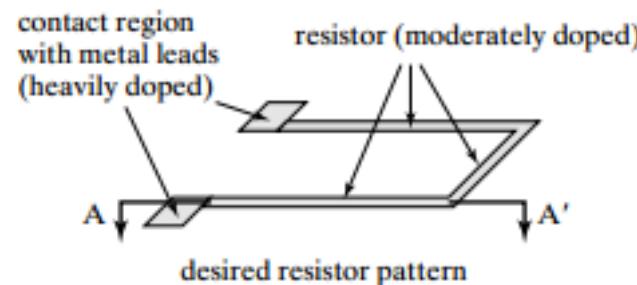


Plasma Etching and Reactive Ion Etching

Plasma etching is a very prominent method of removing materials from a wafer surface. Since the process does not involve wet chemicals, it is often referred to as dry etching. The etching is carried out in a specialized process equipment called the plasma etcher. A chamber with two opposing electrodes is filled with a chemically active gas species. The process pressure is typically rather low. Inside the plasma etcher, gas species are broken up by the electric field, creating active gaseous radicals that are electrically charged. The radicals may react with the wafer chemically. Meanwhile, the charged radicals may be accelerated in the electrical field to high speed and interact with materials on the wafer physically (bombardment, sputtering). Both physical and chemical processes may be present at the same time. In general, physical etching is more directional and anisotropic, whereas chemical etching is isotropic and material selective.

Doping

Doping is the process of planting dopant atoms into the host semiconductor lattice in order to change the electrical characteristics of the host material. The initial source dopants can be placed on the surface of the wafer or precisely injected into the silicon lattice using the ion implantation method

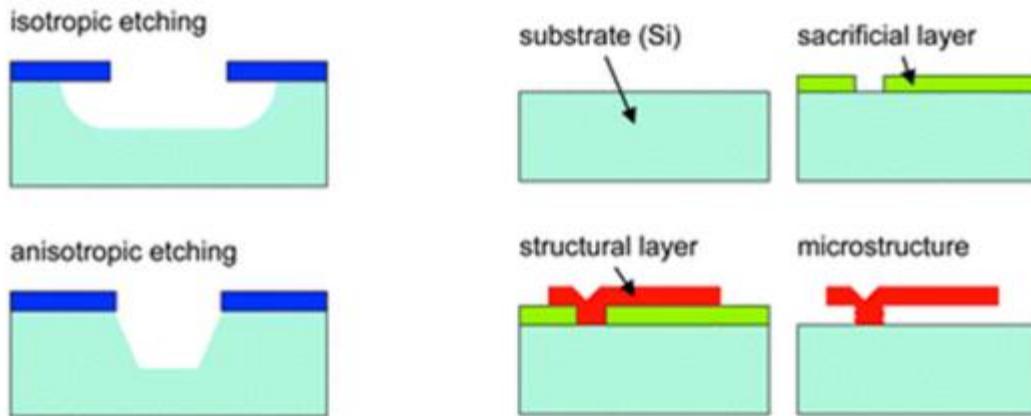


MEMS fabrication technologies: bulk micromachining

Silicon micromachining

Most MEMS fabrication techniques can be classified as

- *in the substrate, **bulk micromachining**, or*
- *above the substrate, **surface micromachining**.*



Bulk micromachining is a fabrication technique to selectively remove substrate to create MEMS devices

Surface micromachining is a fabrication technique for depositing various films on top of the substrate (substrate as a construction base material) and selectively remove parts of deposited films to create MEMS devices

Bulk micromachining processes

The process of bulk micromachining refers to creating structures by re-moving material from the substrate. This can occur in many ways. The most prevalent method is to chemically eat away the material in a process called chemical etching, or simply etching. Solutions containing the required components to achieve the desired chemical reaction(s) are appropriately called etchants. Historically most etchants have been in liquid form, and the process has therefore also been called wet chemical etching.

This also distinguishes the process from those in which the reactive components are contained in a gas or plasma, which is called dry etching.

The overall goal in bulk micromachining is to remove material from the substrate itself. The amount of material removed per unit time during etching is called the etch rate. It is usually measured in dimensions of length (signifying the depth into the substrate) per unit time. In general, higher etch rates are desirable. This allows for batch fabrication to proceed more quickly.

4.2.1 Wet chemical etching

As an example of wet etching, consider Fig. 4.1, which shows a two step process used to etch a pit into a (100) silicon wafer. In the first step a silicon dioxide layer is etched through a window of a photoresist mask. The resist layer is then stripped, leaving a window through the oxide layer to the silicon substrate below. The oxide layer now forms a mask for etching the substrate. (This is often referred to as a hard mask.) Using a different etchant, the exposed portion of the silicon wafer is etched next, forming the desired pit in the substrate.

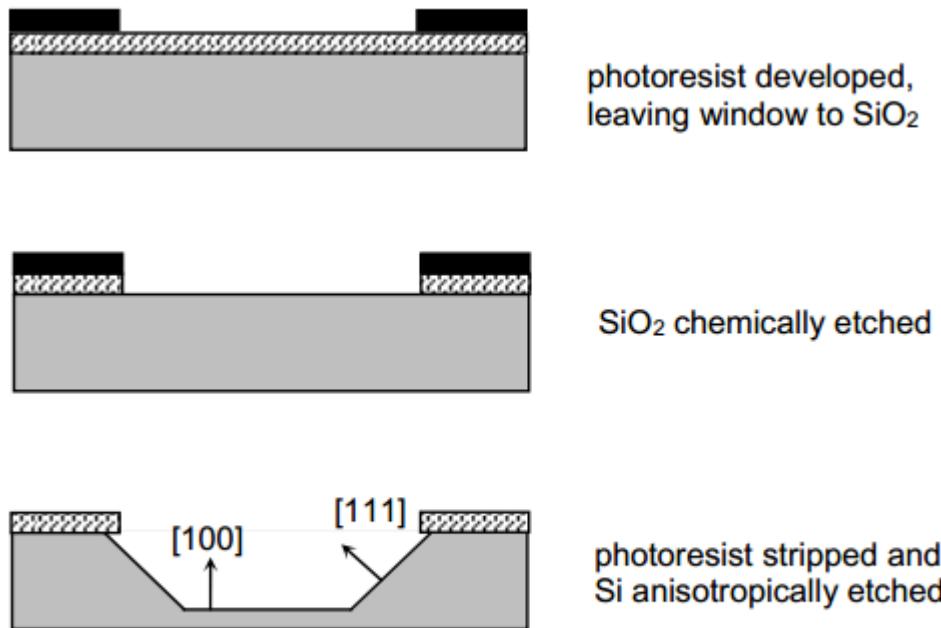
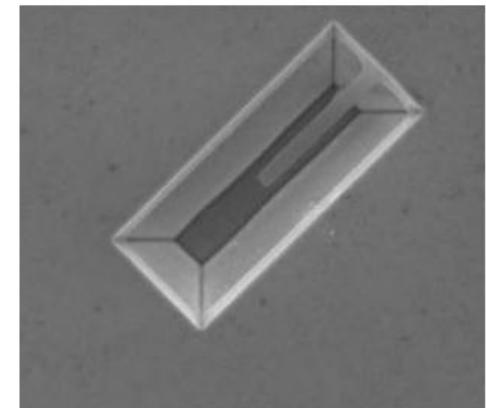


Fig. 4.1. Two step wet etching process to produce a pit in a (100) silicon wafer

It is important to realize that the etchants used in the above described process react with everything with which they come into contact, at least to a small degree. For the process to work, however, the etchants must react very slowly with some materials, while reacting very quickly with others. For example, we would like the etchant used to etch the oxide layer to react very quickly with the oxide itself, but very slowly with the photoresist layer as well as with the silicon substrate below. In this way we can ensure that the resulting pattern in the oxide layer closely resembles the photoresist pattern, and that the reaction stops when the substrate is reached. Likewise, the etchant used for creating the pit in the substrate should react quickly with the silicon, but very slowly with the oxide.

The relative etch rate of an etchant solution with one material compared to another is called **selectivity**. In essence we are playing games with selectivity in order to bring about the desired structures.

You may have noticed in Fig. 4.1 that the resulting window in the oxide layer does not resemble the photoresist mask perfectly. Specifically, the opening in the oxide is slightly wider than that of the photoresist. This is because the chemical reaction has proceeded in the horizontal direction to some degree as well as in the thickness direction of the oxide layer. This phenomenon is called **undercutting**.



You may also have noticed in two preceding figures that the silicon wafer pits have slanted sidewalls. This is because the etchant used not only is selective towards the silicon material, but to certain crystallographic directions as well. When an etching process proceeds at different rates in different directions, it is called an **anisotropic etch**. For the etchant and crystalline orientation of the wafer shown here, the resulting shape of the pit resembles an inverted pyramid with sidewalls corresponding to the {111} planes. When the etching process proceeds at the same rate in all directions, we have an **isotropic etch**. The resulting shape resembles the rounded pit shown in Fig. 4.3.

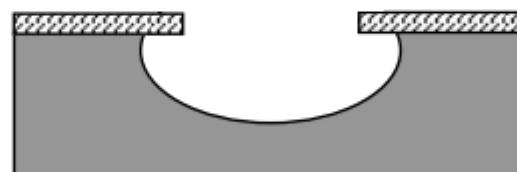


Fig. 4.3. Side view of an isotropically etched pit

Table 4.1 gives the etch rates of various commonly used etchants with respect to silicon and silicon dioxide. Etch rates for various materials are available from numerous sources including handbooks and online resources. They can differ significantly based on differences in conditions, such as concentration and temperature. From Table 4.1 we see that the first two etchants listed are highly selective towards SiO_2 over Si, whereas the last three are selective towards Si. Furthermore, basic etchants tend to etch silicon anisotropically, whereas acidic etchants are isotropic. Since most photoresists are soluble in basic solutions, it is therefore usually necessary to form a silicon dioxide hard mask to etch silicon anisotropically.

Table 4.1. Applications and properties of different etchants (Data taken from Madou)

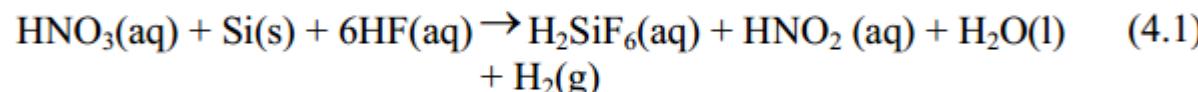
Etchant	Application	Etch Rate (s)	Notes
48% HF	SiO ₂	20-2000 nm/min 0.3 Å/min for Si	Isotropic
Buffered oxide etch (BOE) (28 mL HF/113 g NH ₄ F/170 mL H ₂ O)	SiO ₂	100-500 nm/min (25°C)	Isotropic
Poly etch HF/HNO ₃ /HC ₂ H ₃ O ₂ 8/75/17 (v/v/v)	Si	5 µm/min (25°C)	Isotropic
KOH (44 g/100 mL)	Si	1.4 µm/min (80°C) 28 Å/min SiO ₂	Anisotropic IC incompatible
Tetramethylammonium hydroxide (TMAH) (22 wt%)	Si	10 µm/min (90°C) SiO ₂ virtually unreac- tive	Anisotropic IC incompatible

In wet chemical etching, an etchant solution is brought into contact with the surface to be etched. The reactive component in the etchant solution first finds its way to the surface. Next, one or more chemical reactions take place involving the surface material and the etchant. Last, the products of the reaction(s) move away from the surface, and new etchant moves in to take the place of the products. This process continues until actively stopped in some way.

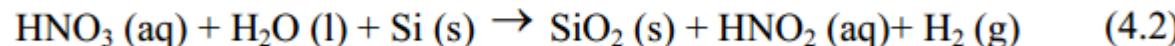
Isotropic etching

In isotropic wet etching the etch rate is the same in all directions. Isotropic etchants are typically acidic with the reactions carried out at room temperature. The isotropy is due to the fast chemical kinetics (fast reactions) of the etchant/material combination. Hence, isotropic etching tends to be diffusion limited. Isotropic etchants also have the fastest etch rates, on the order of microns to tens of microns per minute.

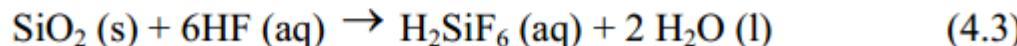
One of the most prevalent isotropic etchants used with silicon is *HNA* ($\text{HF}/\text{HNO}_3/\text{HC}_2\text{H}_3\text{O}_2$) also called *poly etch* for its frequent use with polycrystalline silicon¹. The overall reaction is given by



The etching process actually occurs in several steps. In the first step, nitric acid oxidizes the silicon.



In the second step, the newly formed silicon dioxide is etched by the hydrofluoric acid.



In the isotropic etching of glass or silicon dioxide, the oxidation step is not required. Hence, there is no need for nitric acid, and a buffered oxide etch (BOE) consisting of aqueous HF and NH_4F can be used. The reaction proceeds directly from the second step of Eq. (4.3).

In isotropic etching the amount of undercutting of a mask will be on the order of the depth to which the layer is etched. For example, the opening at the top of an oxide layer created using a photoresist mask with initial dimensions of $200\text{ }\mu\text{m} \times 200\text{ }\mu\text{m}$ will have dimensions of $210\text{ }\mu\text{m} \times 210\text{ }\mu\text{m}$ after etching to a depth of $10\text{ }\mu\text{m}$.

Anisotropic etching

Unlike isotropic etching, anisotropic etching occurs at different rates in different directions, specifically along different crystalline planes. The etchants are typically alkaline instead of acidic, and the reactions are carried out at slightly elevated temperatures, usually between 70°C and 90°C. Also unlike isotropic etching, the process is reaction limited, resulting in slower etch rates on the order of 1 μm/min. The anisotropy itself is due to different crystal planes etching at different rates.

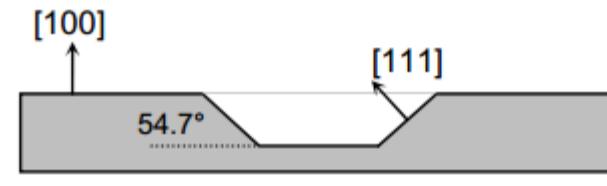
In the anisotropic etching of silicon, silicon dioxide is formed first. Then, the oxide is reacted with a strong base, producing $\text{Si}(\text{OH})_4$ (aq) or more likely, $\text{H}_2\text{SiO}_4^{2-}$ (aq). The {111} planes etch the slowest, and the {100} planes etch the fastest. In fact, the etch rate for the {100} planes is 100 times greater than that for {111} planes when using KOH as the etchant!

Common anisotropic etchants of silicon and their properties are summarized in Table 4.2

Table 4.2. Selected anisotropic etchants of silicon.

Etchant	Temperature	Si etch rate (μm/min)	{111}/{100} selectivity	SiO ₂ etch rate (nm/min)
KOH (40-50 wt%)	70°-90°C	0.5-3	100:1	10
EDP (750ml Ethyl- enediamine 120g Py- rochatechol, 100 ml water)	115°C	0.75	35:1	0.2
TMAH (Tetra- thylammonium hy- droxide 22 wt%)	90°C	0.5-1.5	50:1	0.1

MEMS designers use anisotropic etching techniques in order to create the uniquely shaped trenches often seen in MEMS devices. The most common of these is the inverted pyramid shape seen in the anisotropic etching of (100) silicon, the sidewalls of which make angles of 54.7° with the plane of the wafer. (Fig. 4.5.) If the window through the hard mask is made small enough, the etch will continue only until the intersection of the {111} planes is reached. The crystalline structure serves as its own etch stop in this case. However, this technique only works when the {111} planes form concave corners. Even the slightest deformity on a convex corner of the intersection of {111} planes will expose much faster etching planes, which leads to massive undercutting of the hard mask. Sometimes this is most undesirable and the designer must make use of cleverly designed sets of masks for “corner compensation”. However, this undercutting is often planned right into the process, as is the case in creating suspended structures over etched pits. (Fig. 4.6.)



(a)



(b)

Fig. 4.5. Anisotropic etching of (100) silicon. (a) Anisotropic etching exposes {111} planes, creating sidewalls at 54.7° angles. (b) Intersection of various {111} planes creating a self-limited etch.

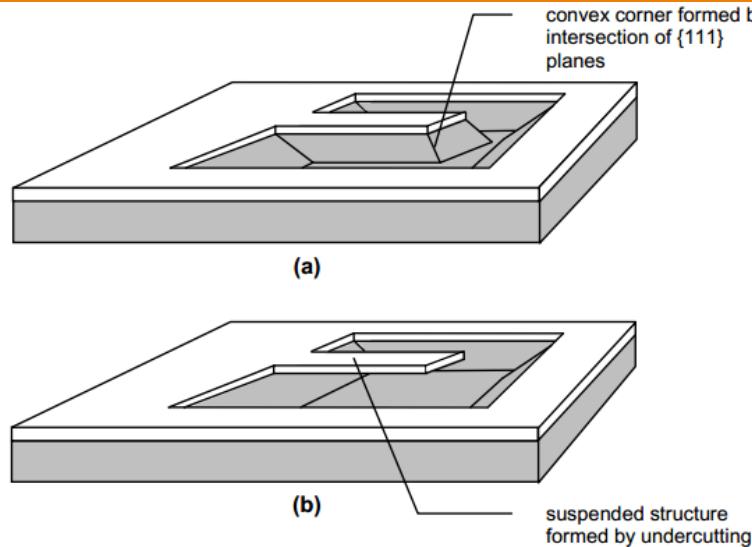


fig. 4.6. Purposely exposing convex corners formed by {111} planes (a) can be used to undercut hard masks, thereby creating suspended structures (b).

For (110) silicon anisotropic etching results in an opening in the plane of the wafer that resembles a lopsided hexagon. The sidewalls make different angles with the horizontal than in (100) silicon. (Fig. 4.7.) Four of these sidewalls are vertical, a feature that otherwise usually requires more expensive and involved techniques compared to wet chemical etching. The bottom of the etched pit is initially a {110} plane, and therefore flat. For long etch times, however, the two slanted {111} planes intersect each other, creating a self-limiting etch. This self-limiting etch is pronounced for mask openings with small aspect ratios, but not large ones. Hence, anisotropic etching of (110) silicon is often used to create long narrow trenches like those used in active liquid cooling of microelectronic components. As with (100) silicon, undercutting of hard masks is sometimes used to create suspended structures.

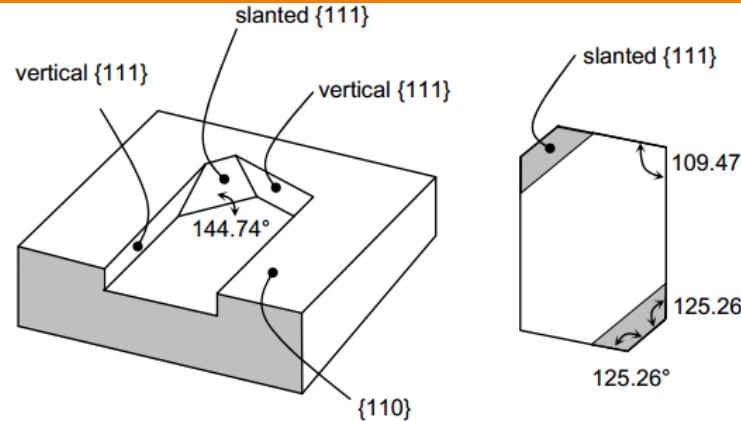


Fig. 4.7. Anisotropic etching of (110) silicon exposes various {111} planes, producing four vertical sidewalls and two slanted sidewalls.

For (111) silicon, the surface plane of the substrate is the one that etches the slowest. Since no other planes are exposed, a bare (111) wafer placed in an anisotropic etchant simply etches very, very slowly with no result other than the wafer becoming thinner. If a “starter hole” is pre-etched in the surface using some other technique, however, other planes will be exposed. Anisotropic etching can now proceed. This is most often used to create pits or trenches underneath a hard mask on the surface. (Fig. 4.8 (a).) When the sidewalls of the pre-etched shape are protected, deep trenches well below the surface can be formed. (Fig. 4.8 (b).) This technique is sometimes used to create microfluidic channels.

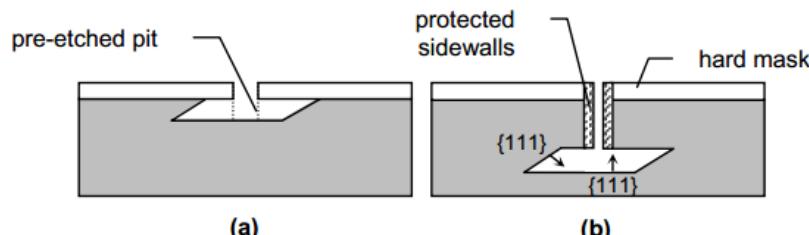


Fig. 4.8. Anisotropic etching of (111) silicon: (a) A pit must be pre-etched before anisotropic etching can proceed. (b) Protecting the sidewalls of the pre-etched shape can be used to produce deep trenches.

Etch stop

Getting the chemical reaction that makes up an etching process to stop can pose a difficult problem within microfabrication. One technique is to time the etch, simply pulling a wafer out of the wet etching solution when the desired depth has been reached. Variability of etch rates, however, makes this technique less than precise. The resolution for thickness is generally on the orders of microns. That is to say, if you attempt to leave a membrane in a silicon wafer less than 10 μm thick using this technique, it could result in a membrane 0 μm thick!

In an *insulator etch stop* the etching process stops when the etchant reaches an insulating layer. In the case of a silicon wafer, the insulator is usually an oxide layer grown on one side of the wafer. Due to the selectivity of the etchant, the etching essentially ceases upon reaching the silicon dioxide. In the case of an insulator etch stop, the insulator makes up part of the structure of the MEMS.

Other etch stop techniques involve varying etchant composition and temperature. Much more precise, however, is to make use of a wafer that has been previously doped at a certain depth. This can allow for a very precise control over etch depth, which has in large part led to microfabrication becoming a high yield process.

Anisotropic etchants do no attack heavily doped Si layers (often designated p⁺), and severe drops in etch rate result when the p-type layer is reached. As boron is most frequently used to create p-type silicon, this process is often called a *boron etch stop*. According to the anisotropic etching model of Siedel *et al.* the p-type layer is electron deficient, and there are no electrons available to react with water at the surface. (According to this model, the reduction of water is the rate-limiting step in the etching process.) One disadvantage of a boron etch stop is that the high

level of p-type doping, on the order of $5 \times 10^{18}/\text{cm}^3$, is not compatible with CMOS standards for integrated circuit fabrication.

One technique that is compatible with CMOS standards is called *electrochemical etch stop*. In this particular etch stop an n-type layer is grown on a p-type Si substrate, creating a p-n junction. The required level of doping is very light compared to the boron etch stop method. During etching, an electric potential is applied across the p-n junction creating a reverse-bias diode. The diode keeps current from flowing across the junction, and etching of the p-type substrate can occur readily. As soon as the etching reaches the n-type layer, however, the diode vanishes and current starts flowing. The newly freed electrons oxidize the Si at the surface, forming SiO_2 . Since silicon dioxide etches very slowly in an alkaline medium, the etching stops almost immediately.

4.2.2 Dry etching

In contrast to wet etching, which employs etchants in aqueous solution form, **dry etching** delivers the chemically reactive species to the etching surface in gaseous form, or within an ionized gas, called a *plasma*. In dry etching the gas or plasma bombards the etching surface, producing a sputtering-like effect. Hence, dry etching actually consists of a combination of physical and chemical etching mechanisms. This bombardment also gives the etching process a directional component, and dry etching techniques have become the standard for creating near vertical sidewalls with large aspect ratios.

An example of an etchant used in dry etching is xenon diflouride (XeF_2). XeF_2 is a highly selective etchant that can etch silicon or polysilicon without etching metals, silicon dioxide, or many other structural layers. This selectivity makes XeF_2 vapor a valuable etchant for the release step in surface micromachining processes in which polysilicon is the sacrificial layer.

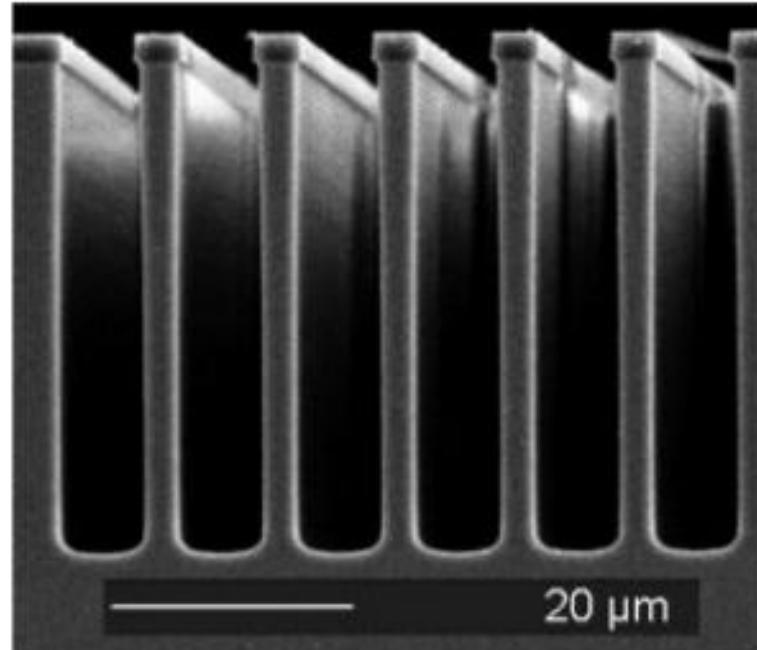
Dry etching

In contrast to wet etching, which employs etchants in aqueous solution form, dry etching delivers the chemically reactive species to the etching surface in gaseous form, or within an ionized gas, called a plasma. In dry etching the gas or plasma bombards the etching surface, producing a sputtering-like effect. Hence, dry etching actually consists of a combination of physical and chemical etching mechanisms. This bombardment also gives the etching process a directional component, and dry etching techniques have become the standard for creating near vertical sidewalls with large aspect ratios.

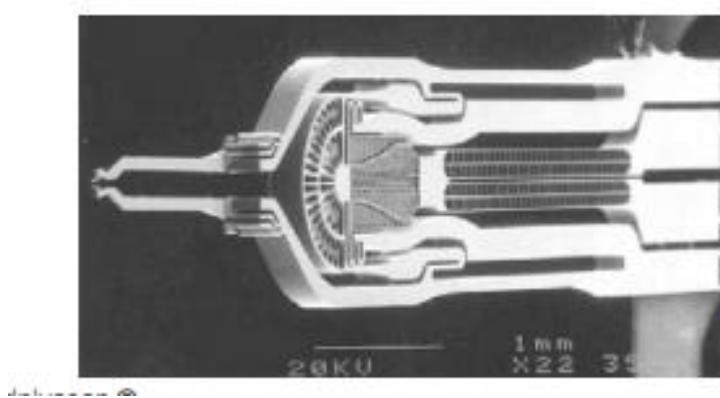
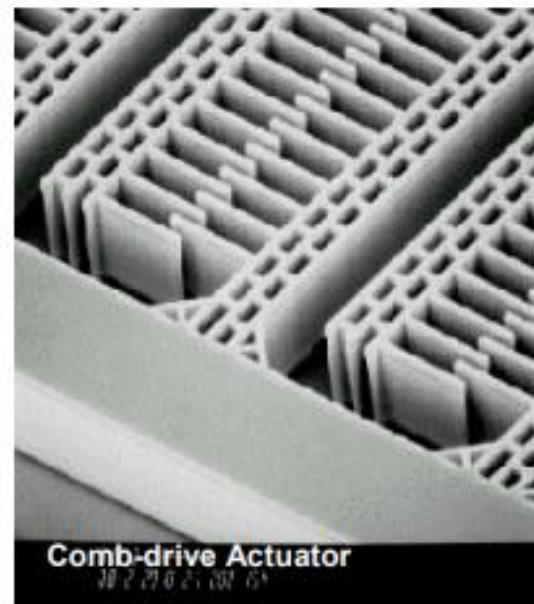
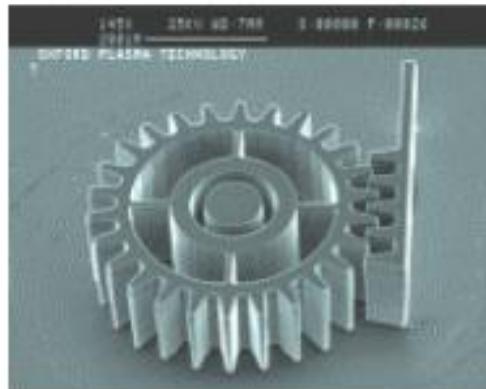
Reactive ion etching (RIE)

In the limit of low pressures, plasma etching can take on much higher degrees of directionality, especially when the ions are directed normal to the surface. This process has become known as reactive ion etching, or RIE.

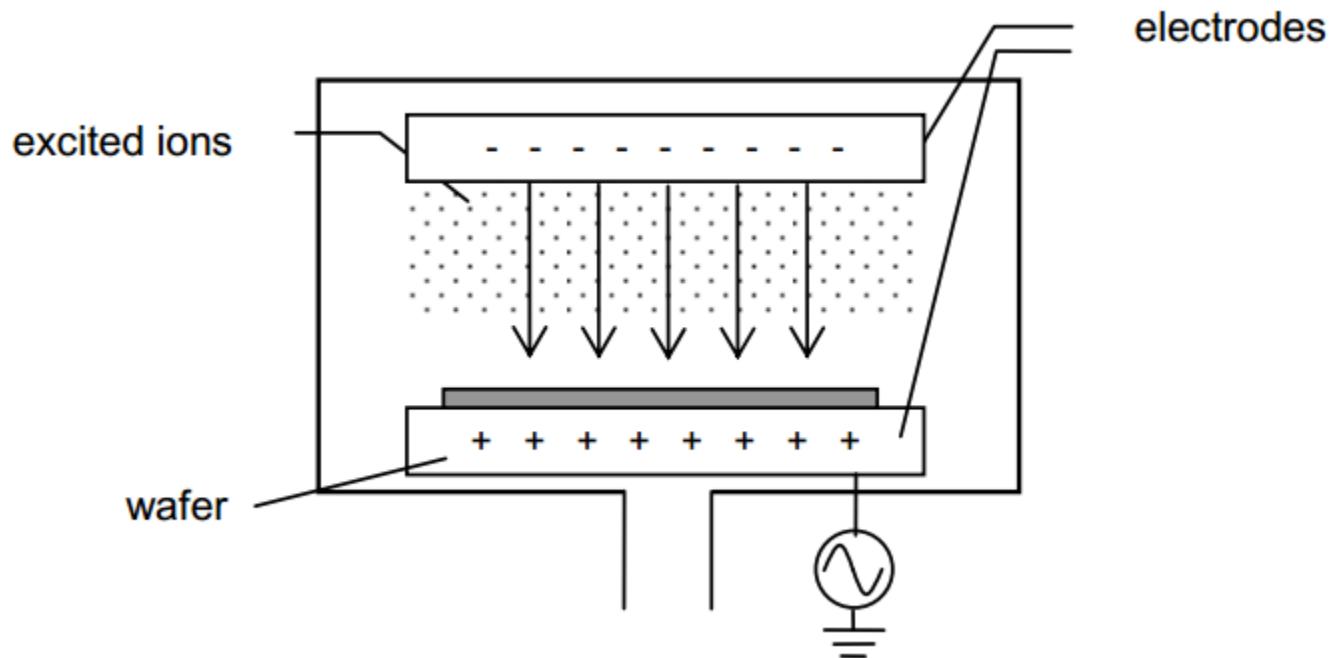
RIE is capable of creating features with high aspect ratios and nearly vertical sidewalls of high surface quality



DRIE Examples



RIE is typically done by bombarding wafers with heavy ions such as Ar ions in the presence of an energetic plasma. Since the ions are directed toward horizontal surfaces, they typically do not hit the sidewalls during etching. It is by this mechanism that vertical walls are produced.



A schematic of an RIE chamber. Electrodes produce an RF electric field that excites ions and directs them toward the surface of the wafers

Plasma etching

In **plasma etching**, the chemically reactive gas is ignited by an RF (radio frequency) electric field, usually in the range of 10-15 MHz. The reactive chemical species is contained within a plasma inside a vacuum chamber where the surface to be etched also resides. The plasma provides the necessary energy for, or “excites” the reactive gas in order to etch the wafer.

One variant of plasma etching called **plasma ashing** is commonly used to remove photoresist from wafers after completing photolithography steps. In this process an oxygen plasma converts the polymer that makes up the photoresist into carbon dioxide and water vapor, which are then removed by the vacuum pump of the ashing system.

Backside processing

Many MEMS devices are fabricated using both sides of the substrate. A good example is a bulk micromachined MEMS pressure sensor. (Fig. 4.19.) In this device a thin diaphragm is created by etching through the bottom of the wafer. Etching through the bottom of the wafer in this way is an example of **backside processing**. Piezoresistive sensing elements and metal interconnects are also needed for the device, but these are processed from the top.

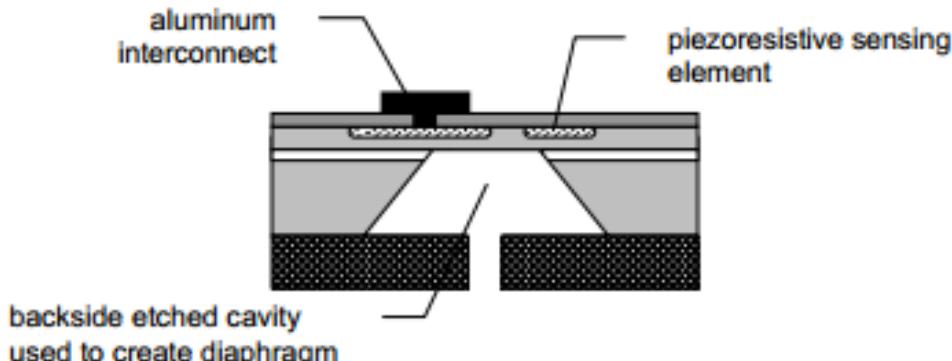
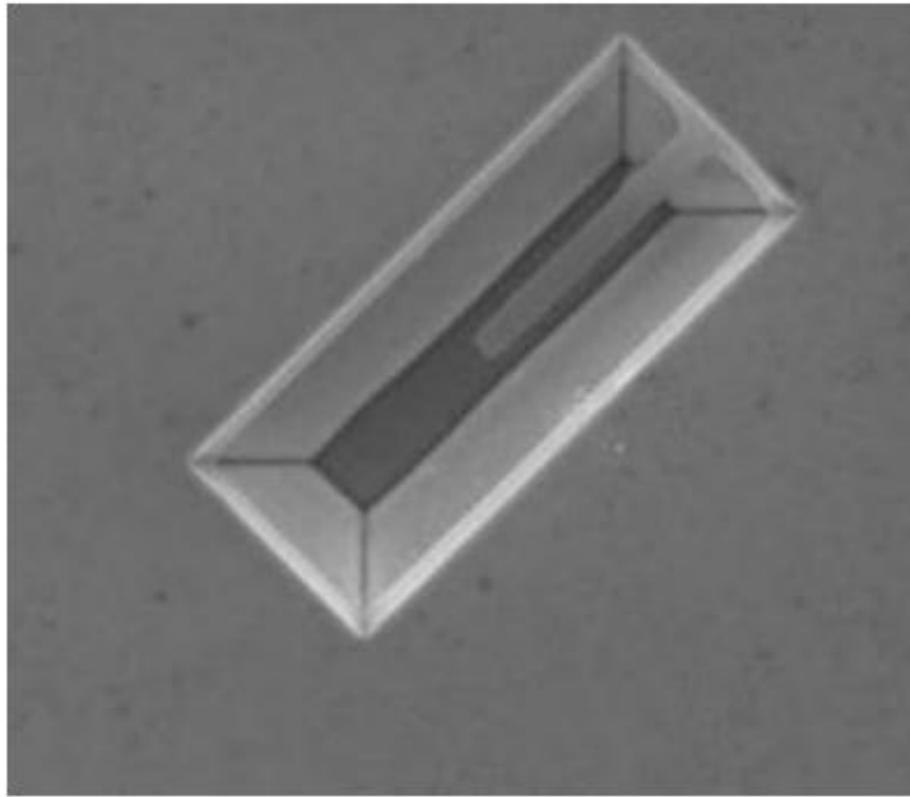


Fig. 4.19. Bulk micromachined pressure sensors are examples of devices requiring backside processing.



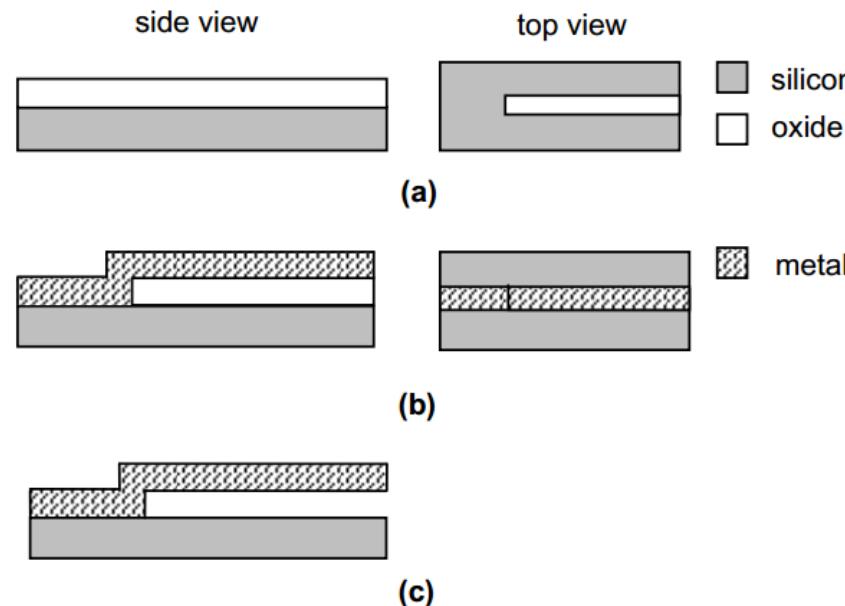
Scanning electron microscope image of a silicon dioxide cantilever formed by undercutting (Courtesy: S. Mohana Sundaram and A. Ghosh, Department of Physics, Indian Institute of Science, Bangalore)

Surface micromachining



Surface micromachining (SMM) is a process that uses thin-film deposition, patterning via photolithography, and chemical etching to build mechanical structures on top of a substrate, typically a silicon wafer. It is a layered fabrication process in which some layers form structural elements and others are etched away.

The layers are referred to as structural layers and sacrificial layers, respectively, and form the building blocks of the process. The typical process of SMM is first to deposit a layer, then pattern it, and finally chemically etch away unwanted material. This set of steps can be repeated several times in order to create complicated structures, often with moving parts.



Process steps in fabrication of a cantilever structure using surface micromachining

High-aspect-ratio (LIGA and LIGA-Like) technology



- LIGA is a German acronym that stands for Lithographie, Galvanoformung and Abformung.
- When translated it means lithography, electroplating and molding.

- Two main types of LIGA Technology: X-ray LIGA and Extreme Ultraviolet (EUV) LIGA.
- X-ray LIGA can fabricate with great precision high aspect ratio microstructures.
- EUV LIGA can fabricate lower quality microstructures.

- LIGA is a hybrid fabrication technique
- The LIGA Process
 - Lithography
 - ✖ Electron beam lithography
 - ✖ Focused ion beam lithography
 - ✖ Optical and exciter laser lithography
 - ✖ Deep X-ray lithography using synchrotron radiation
 - Electroplating
 - ✖ metalized layer (seed layer)
 - Molding
 - ✖ Machining process to remove overplated metal region

3.3.3 High-Aspect-Ratio (LIGA and LIGA-Like) Technology

There is a critical need to develop the fabrication technologies allowing one to fabricate high-aspect-ratio microstructures and microdevices. The LIGA process, which denotes Lithography–Galvanoforming–Molding (in German Lithografie–Galvanik–Abformung), is capable of producing three-dimensional microstructures of a few centimeters high with the aspect ratio (depth versus lateral dimension) of more than 100. This ratio can be achieved only through bulk micromachining using wet *anisotropic* etching.

The LIGA technology is based on the x-ray lithography which guarantees shorter wavelength (from few to ten Å which lead to negligible diffraction effects) and larger depth of focus compared with optical lithography. The ability to fabricate microstructures and microdevices in the centimeter range is particularly important in the actuators applications since the specifications are imposed on the rated force and torque developed by the microdevices. Due to the limited force and torque densities, the designer faces the need to increase the actuator dimensions. The LIGA and LIGA-like processes are based on deep x-ray lithography and electroplating of metal and alloy structures, allowing one to achieve structural heights in the centimeter range [32-35]. This type of processing expands the material base significantly and allows the fabrication of new high-performance electromechanical microtransducers.

Figure 3.3.7 illustrates the basic sequential processes (steps) in LIGA technology. Here, the x-ray lithography is used to produce patterns in very thick layers of photoresist. The x-rays from a synchrotron source are shone through a special mask onto a thick photoresist layer (sensitive to x-rays) which covers a conductive substrate (step 1). This photoresist is then developed (step 2). The pattern formed is electroplated with metal (step 3). The metal structures produced can be the final product, however, it is common to produce a metal mould (step 4). This mould can then be filled with a suitable media (e.g., metal, alloy, polymer, etc.) as shown in step 5. The final structure is released (step 6).

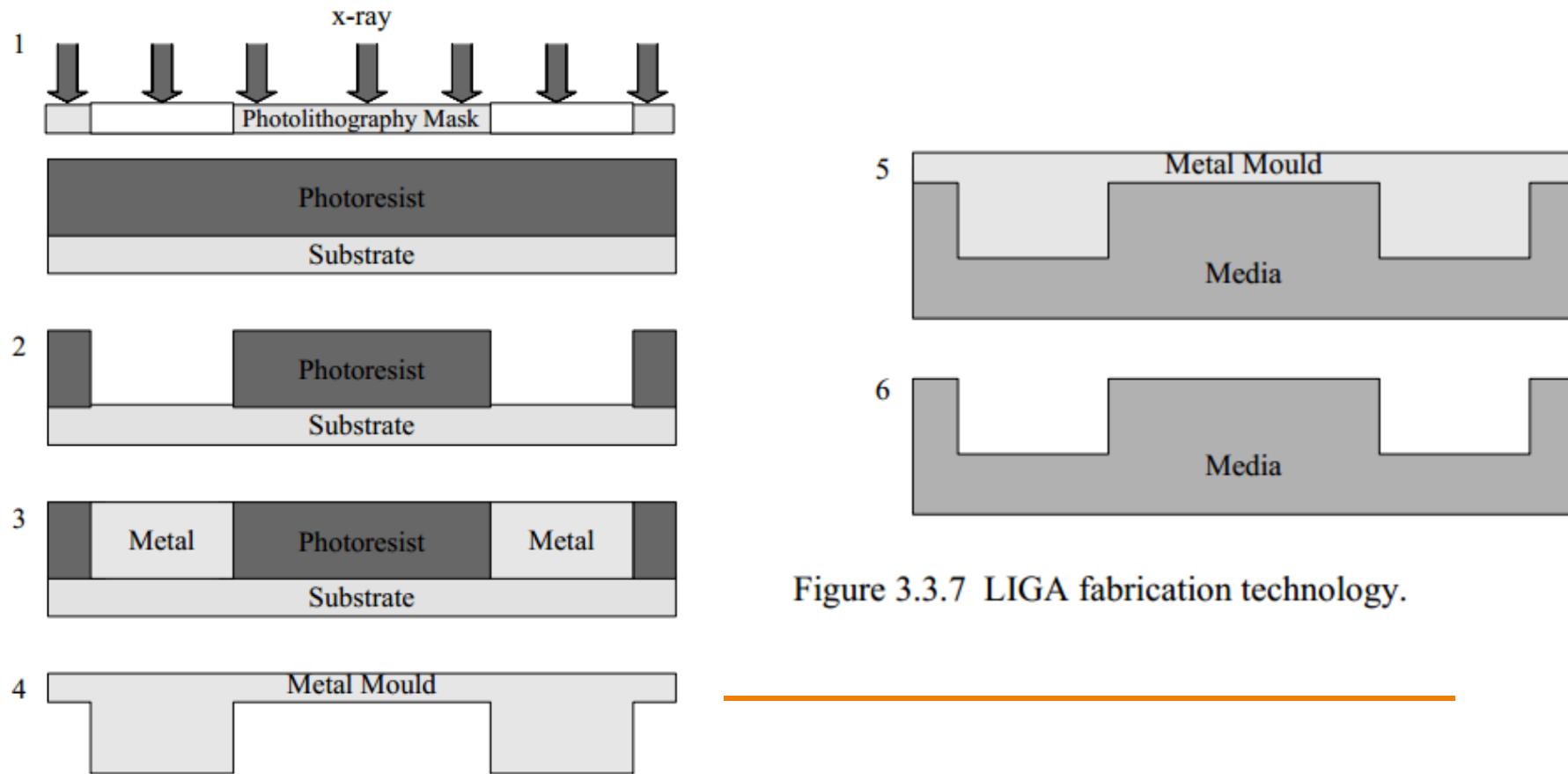


Figure 3.3.7 LIGA fabrication technology.

The described LIGA technology (frequently referred to as the high-aspect-ratio technique) allows one to fabricate microstructures with small lateral dimensions compared with thickness. Thick and narrow microstructures guarantee high ruggedness in the direction perpendicular to the substrate and compliance in the lateral directions. For actuators, high-aspect-ratio technology offers the possibility to fabricate high torque and force density microtransducers. As was emphasized, high-intensity, low-divergence, and hard x-rays are used as the exposure source for the lithography.

Due to short exposure wavelength, the desired features size is achieved.

These x-rays are usually produced by a synchrotron radiation source [1, 32-35]. Polymethyl-methacrylate (PMMA) and polylactides are used as the x-ray resists because PMMA (PlexiglasTM or LuciteTM) and polylactides photoresists have high sensitivity to x-rays, thermal stability, desired absorption, as well as high resolution and resistance to chemical, ion, and plasma etching. Polyglycidyl-methacryl-atecoethylacrylate (PGMA) is used as the negative x-ray resist.

The exposure wavelength varies depending upon the x-ray radiation source used. For example, the 0.2 nm x-ray wavelength allows one to transfer the pattern from the high-contrast x-ray mask into the photoresist layer with a few centimeters thickness so that the photoresist relief may be fabricated with an extremely high depth to width ratio. The sidewalls of the plated structures are vertical and smooth (polished), and therefore, they can be used as optical surfaces.

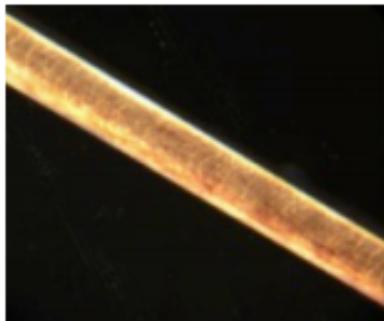
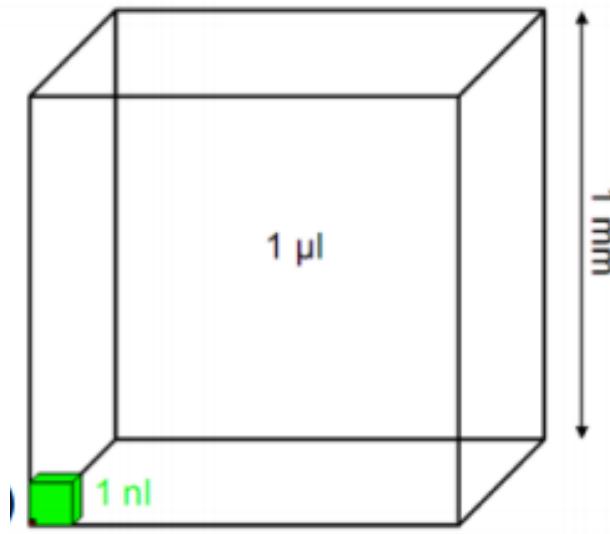
A critical part of the high-aspect-ratio processes is plating to form the metallic electromechanical microstructures in the mold. Using plating, metal is deposited from ions in a solution following the shape of the plating mold. This is the additive process, and the thickness of the plated metal can be large since the plating rate can be high. A variety of metals (Al, Au, Cu, Fe, Ni, and W) and alloys (NiCo, NiFe, and NiSi) can be deposited or codeposited. It is important that roughness (smoothness) of the reflective metal surfaces with the desired shape can be achieved even for optical applications. Electroplating (well-known from chemistry and covered in Chapter 8) and electroless plating (reduction of the metal ions occurs by the chemical reaction between a reducing agent and metal ion on a properly activated substrate) are the commonly used plating processes. The metal seed layer can be deposited and removed from the substrate or sacrificial layer. The plating rate and the grain size are controlled by the current density, temperature, duty cycle, etc. (see Chapter 8).

Microfluidics microsystem components



Microfluidics

- **Fluidics:** handling of liquids and/or gases
- **Micro:** has at least one of the following features:
 - Small volumes
 - Small size
 - Low energy consumption
 - Use of special phenomena



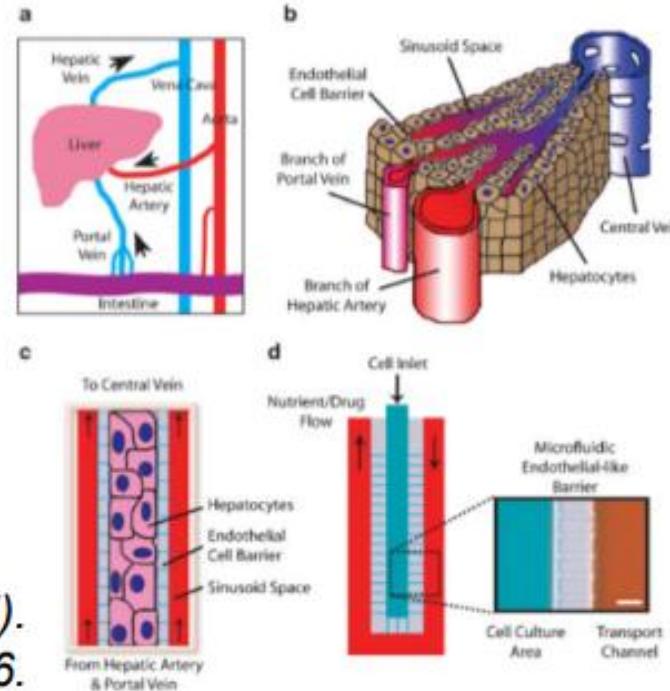
A microfluidic channel is about the same width as a human hair, 70 μm

Advantages of Microfluidics

- Low sample and reagent consumption; fluid volumes (μl ; nl ; pl ; fL)
- Small physical and economic footprint
- Parallelization and high throughput experimentation
- Unique physical phenomena: use of effects in the micro-domain:
 - Laminar flow
 - Capillary forces
 - Diffusion



P.J. Lee et al. (2007).
Biotech. Bioeng. 97: 1340-6.



- Types of fluids:
 - Newtonian fluids
 - Non-Newtonian fluids

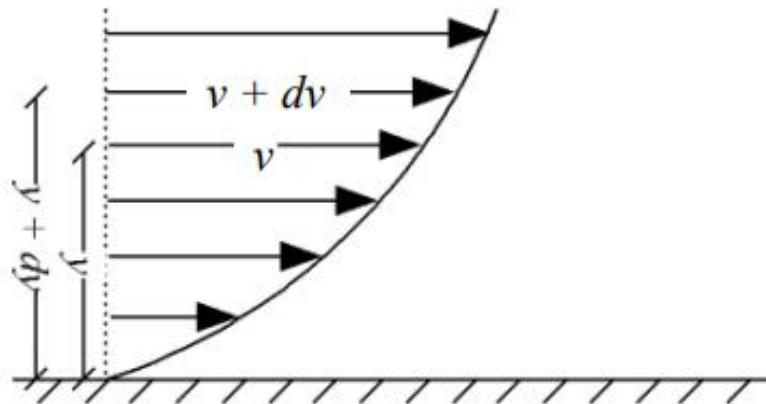


- Types of fluid flow:
 - Laminar
 - Turbulent



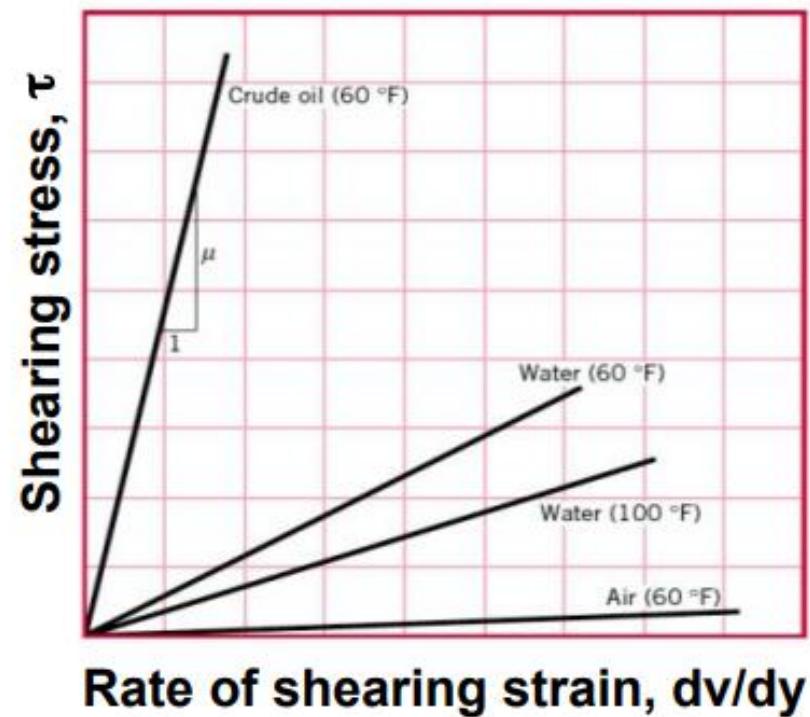
Newtonian Fluids

- Linear relationship between stress and strain, i.e. viscosity is independent of stress and velocity



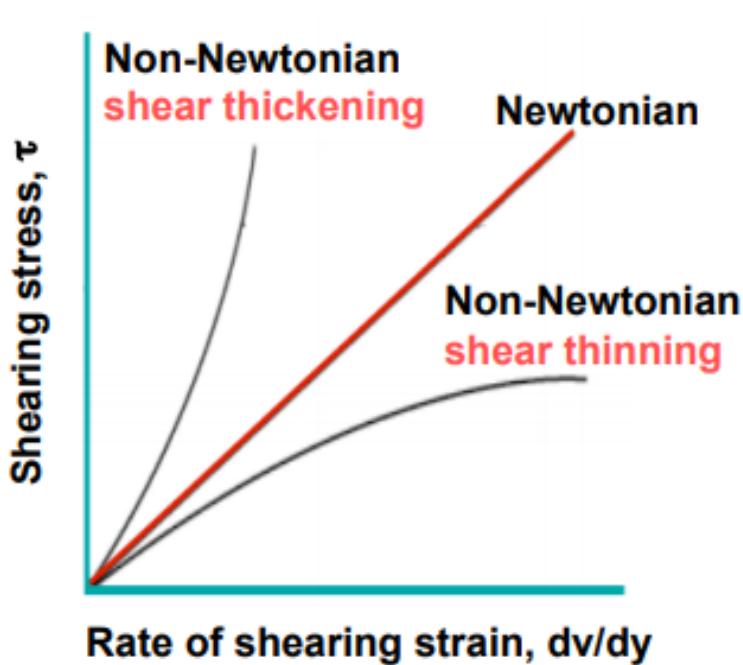
www.engineeringtoolbox.com

$$\tau = \mu \frac{dv}{dy}$$



Non-Newtonian Fluids

- Non-linear relationship between shear stress and shear strain
- Examples: paint, blood, ketchup, cornstarch solution



Viscosity

- Viscosity is a measure of internal friction (resistance) to flow

Substance	Viscosity (mPa·s)
Air	0.017
Acetone	0.3
Water	0.9
Mercury	1.5
Olive oil	80
Honey	2,000 – 10,000

Laminar and Turbulent Flow

- Laminar flow:
 - Fluid particles move along smooth paths in layers
 - Most of energy losses are due to viscous effects
 - Viscous forces are the key players and inertial forces are negligible
- Turbulent flow:
 - An unsteady flow where fluid particles move along irregular paths
 - Inertial forces are the key players and viscous forces are negligible
- Reynolds number:
 - Measure of flow turbulence
 - $Re < 2000$ for laminar
 - Due to small dimensions
 $Re < 1$ in microfluidic systems

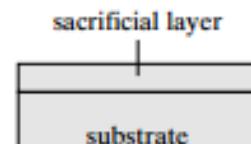
$$Re = \frac{\rho v L}{\mu} \text{ where } L = \frac{4A}{P}$$

DESIGN AND FABRICATION OF MICROFLUDIC COMPONENTS

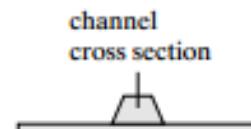
Microfluid channels are the most important components in a microfluid system, despite its relatively simple form and function compared with others (such as pumps, valves). The selection of the channel material is the starting point for any development efforts of microfluidic systems. There are several important aspects that must be taken into consideration when selecting channel materials and subsequent fabrication methods. These include:

1. *Hydrophobicity of the channel wall.* Liquid moves freely in channels with hydrophilic walls by capillary action, simplifying sample loading and priming. Glass, for example, is hydrophilic to many liquids and its properties are well known. Introducing liquid into channels with hydrophobic walls is considerably more difficult.
2. *Biocompatibility and chemical compatibility.* Ideally, the channel wall should not react with the fluid, particles or gases within. Glass, the material for beakers and test tubes, is perhaps the most established biocompatible material and is well liked by researchers in the biological and chemistry community. However, there is a lack of micromachining methods for glass.
3. *Permeability of channel material to air and liquid.* High permeability will cause excessive loss of fluid or, in the case multiple channels are placed close to each other, cross-contamination. Permeability of air or gas is often taken advantage of for air escape and removing trapped air bubbles.
4. *Retention of chemicals on walls.* Walls that retain chemicals may cause cross-contamination during repeated use.
5. *Optical transparency.* Optically transparent walls facilitate observation and quantitative assay analysis.
6. *Temperature of the processing.* Low temperature processing is always desirable. High-temperature processes would narrow the choice of structural and surface coating materials.
7. *Functional complexity and development cost.* The materials for channels should be amendable to integration of active components such as pumps and valves. The barrier to prototyping and manufacturing should be low.

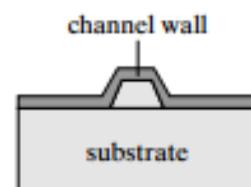
Process 1



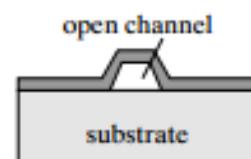
(a)



(b)



(c)



(d)

Process 2

etch trenches in wafer



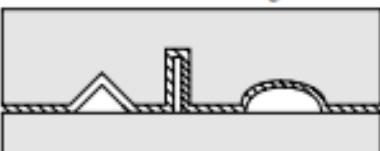
(a)

conformal coating



(b)

wafer bonding



(c)

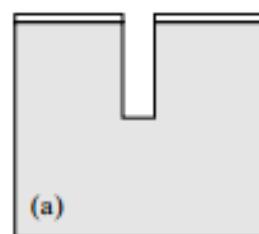
dissolution of wafers



(d)

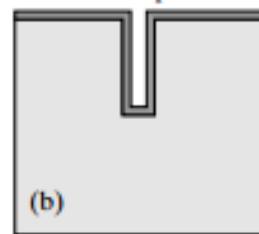
Process 3

deep reactive ion etching



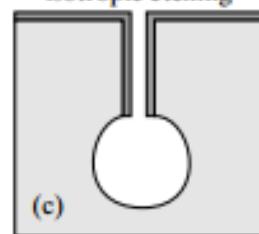
(a)

uniform deposition



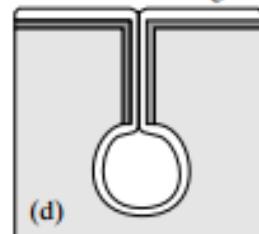
(b)

isotropic etching



(c)

channel sealing



(d)

FIGURE 14.6

Representative fabrication processes of channels on silicon substrates.

Valves

Valves are important elements in a microfluid system. They provide complex system-level functionalities to a laboratory-on-a-chip system. The following factors are generally considered when selecting or developing a micromachined valve:

- The reliability of valve operation. Ideally, a valve should be leak free during off state and open during on state.
- The repeatability of valve operation.
- The ability to withstand large pressure.
- The simplicity of valve construction.
- The simplicity of valve operation and control.
- Biocompatibility with the fluid and biological particles.

Valves can be classified according to the mode of operations into several categories:

- Cyclic valves can be operated multiple times. They can be constant on, meaning the valve holds its open position without active input of power, or constant off, meaning the valve maintains sealed position without active power.
- One-time valves are operated only once during the life of operation. Constant-on valves will seal off a channel permanently when activated. Constant-off valves will open once it is activated, for applications such as collection of environmental samples.

Since the valve is critical for the performance of a microfluid system and for enabling miniaturization, many valve designs have been developed in the past. Generally, valve structures fall into the following categories:

- Hard-membrane valves
- Soft-membrane valves
- Plug valves
- Threshold valves

Among the most important fluid-handling elements are pumps and valves. Pumps are devices to set fluids into motion, and valves are designed to control this motion and, for example, define a preferred direction for the motion of the fluid. The engineering community has developed a wealth of pumps and valves based on a large variety of principles. Many of these principles and implementations can also be used in the microdomain, but additionally new designs have appeared making clever use of the different behavior of liquids and mechanics on small scales.

A valve is a device within a fluidic system in which flow is allowed in one direction but suppressed in the opposite direction, thus introducing directionality into the flow. Valves are often classified by whether they work by themselves (i.e., by utilizing energy from the flow) or whether they need external energy to function. The former are called passive or check valves, whereas the latter are called active valves. The wish-list of ideal characteristics that a valve should have includes zero leakage (when it's closed it's closed, when it's open it's open), zero power consumption (obviously not true for active valves), zero dead volume (should not introduce extra volumes, which negatively affect the performance), infinite differential pressure capability (a little bit of extra pressure from one side should open it, a little bit of extra pressure from the other side should close it), zero response time (no delays), insensitivity to particulate contamination, ability to operate with any fluid, and so on. Unfortunately, no single valve meets all these requirements, in other words, all valves have some flaws, which need to be taken into account when operating them.

An example of a passive valve is the cantilever type (Fig. 4.1), here, a thin strip of silicon that bends when enough pressure is applied from one side. From the other direction bending is restricted by the valve seat. Depending on the exact design and operating conditions, these kinds of valves are more or less leaky. However, they are relatively simple to make, do not require external energy for operation, and have a fairly fast response time. There are also passive valves, which do not rely on mechanical action, but rather make use of forces such as surface tension. These types of valves are typically one-time-use or burst valves, because their functioning depends on an air–liquid interface, which is typically present only when first filling a microchip. A simple implementation of such a valve is a restriction in the channel. As the air–liquid interface approaches the restriction the

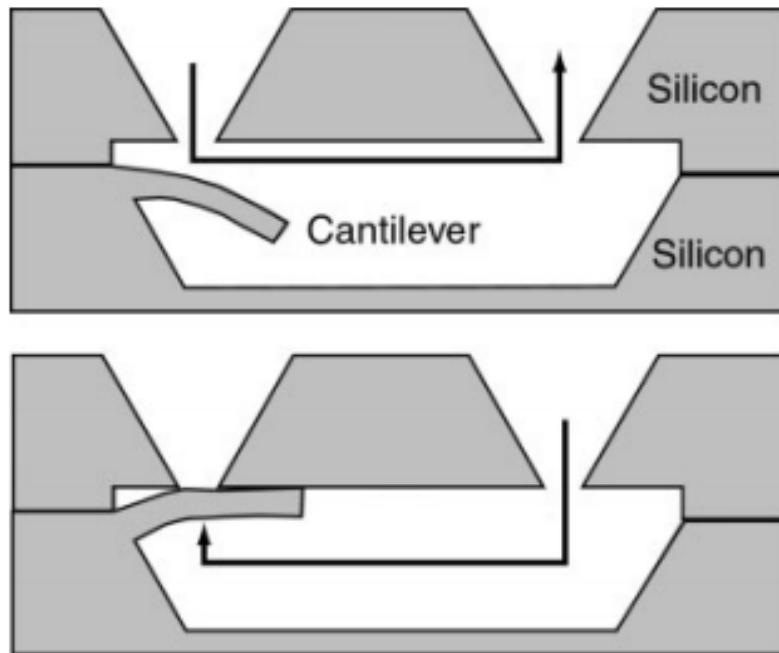


Fig. 4.1 A check valve using a cantilever as a valve flap (adapted from [1]).

Active valves require an actuator to provide a mechanical action, which moves a part of the valve to close or open the flow passage. Valves can be designed to be ‘normally open’ or ‘normally closed’, meaning that this is the state they are in without actuation. For reaching and maintaining the opposite state, energy needs to be applied. A few designs have two stable states (open and closed) and need energy only for transition between the two states (so-called bistable valves). The actuation principles used are various, including pneumatic (using compressed air), thermopneumatic (using heated fluids), piezoelectric (using special materials that expand when an electrical voltage is applied), electrostatic (using electrical attraction or repulsion), shape memory alloy (metals that change shape under temperature changes but ‘remember’ and revert to their original shape), and electromagnetic actuation, to name but a few. Actuators are typically characterized by four criteria: the pressure they can build up, the stroke displacement they generate, their response time, and their reliability. Again, some actuators perform better in one category but less well in another. For example, piezoelectric stacks can produce large pressures, but not very impressive strokes. Electromagnetic actuators, on the other hand, can give large displacements but only relatively small pressures. A valve with a thermopneumatic actuator is depicted in Fig. 4.3.

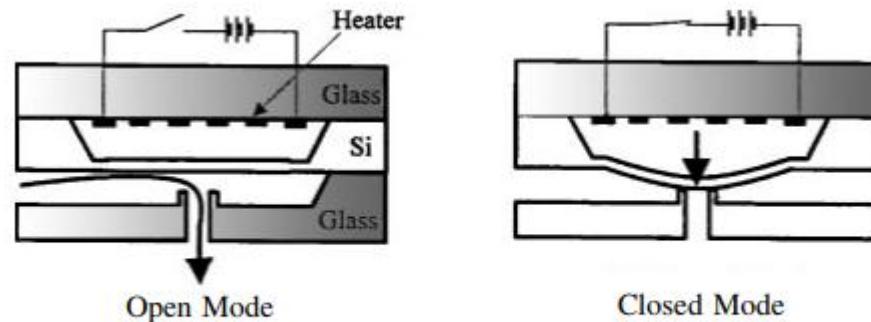


Fig. 4.3 An active valve using a heating chamber with a working fluid and a diaphragm (reprinted with permission from [3]. Copyright (1998) Springer Verlag).

The requirements for a microfabricated pump can be as manifold as its intended uses. A typical list of requirements for a microfabricated pump to be used in a microsystem for analyzing wastewater contained the following points [7]:

- as little pulsation as possible
 - flow rate controllable over a certain range (e.g., $0.25\text{--}10 \mu\text{L min}^{-1}$)
 - leak rates under 1 nL min^{-1} at counterpressures of $1\text{--}10 \text{ kPa}$
 - sensitivity to counterpressures of $1\text{--}10\% \text{ kPa}^{-1}$
 - flow rate precision over several days within $1\text{--}3\%$
 - resistance to aggressive chemicals over a longer period of time at elevated temperatures
 - production costs as low as possible
-

Example of shock waves inducing streaming of a fluid

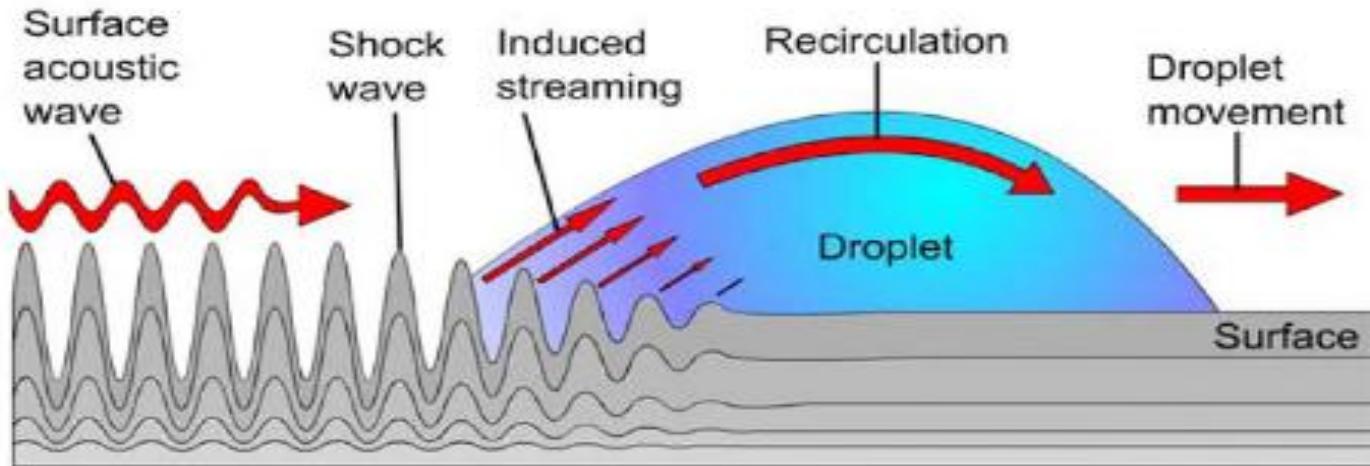
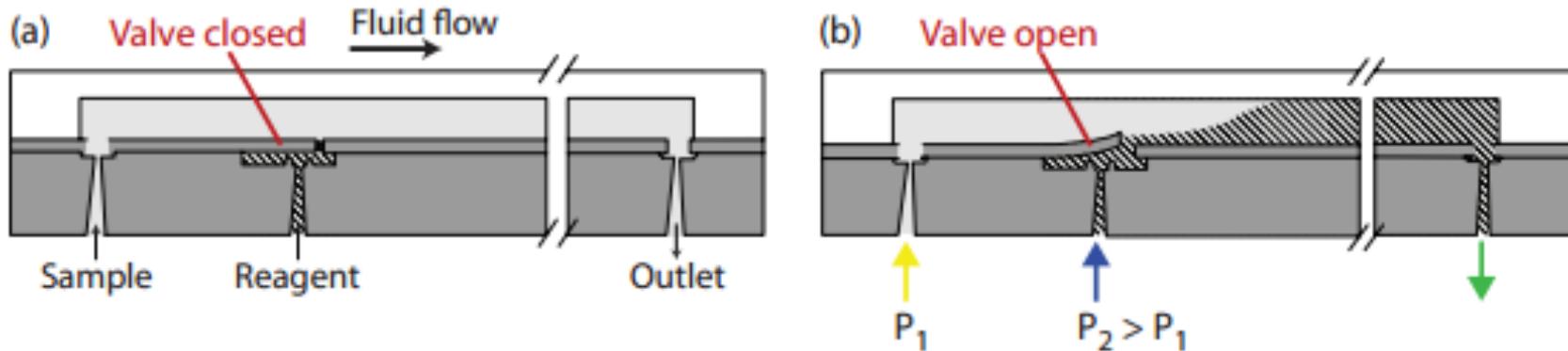
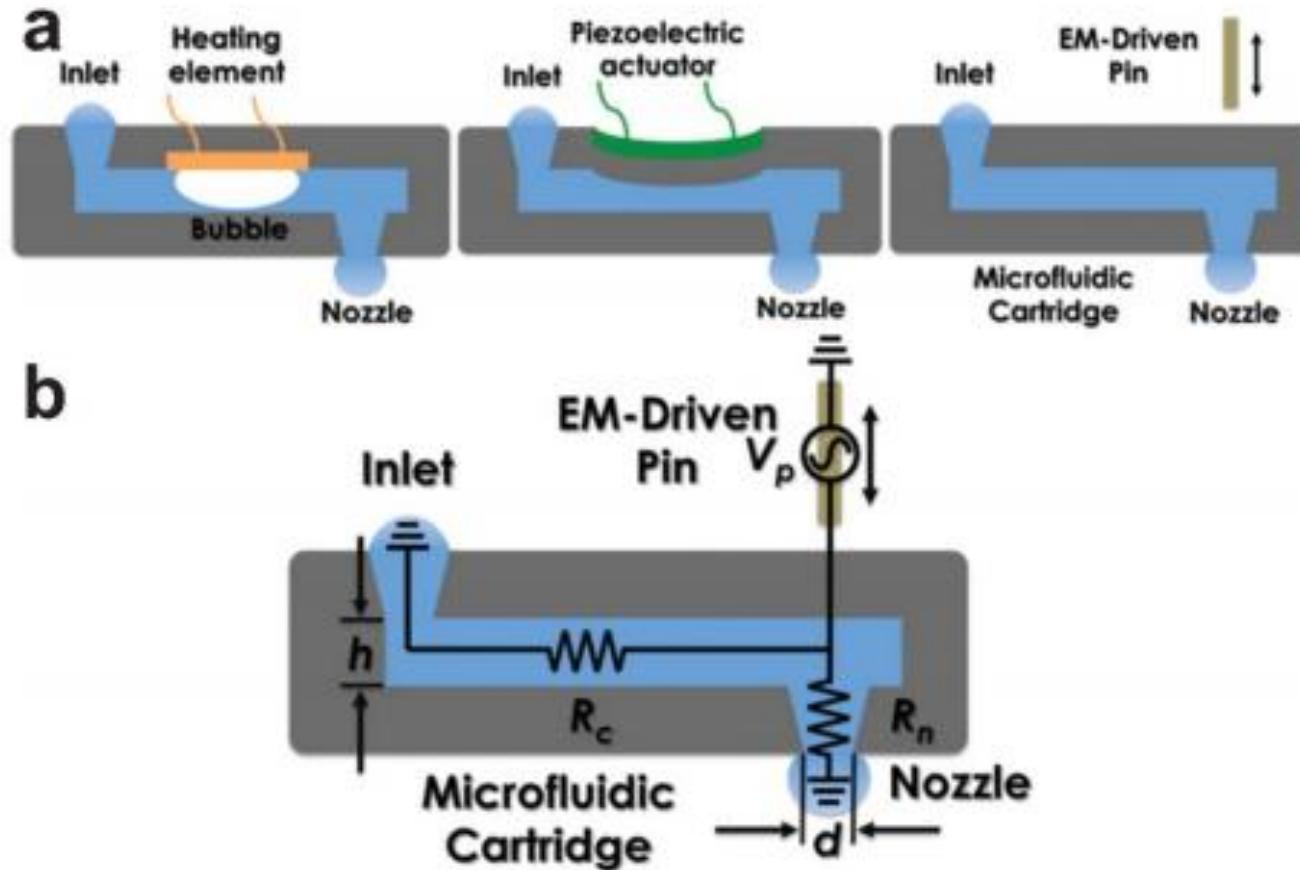


Figure 17. Surface acoustic wave (SAW). (According to Tan et al. [277].) The shock waves induce a stream on the solid–liquid interface and lead finally to a movement of the droplet (amplitude of acoustic wave not to scale).

Flap microvalve Example

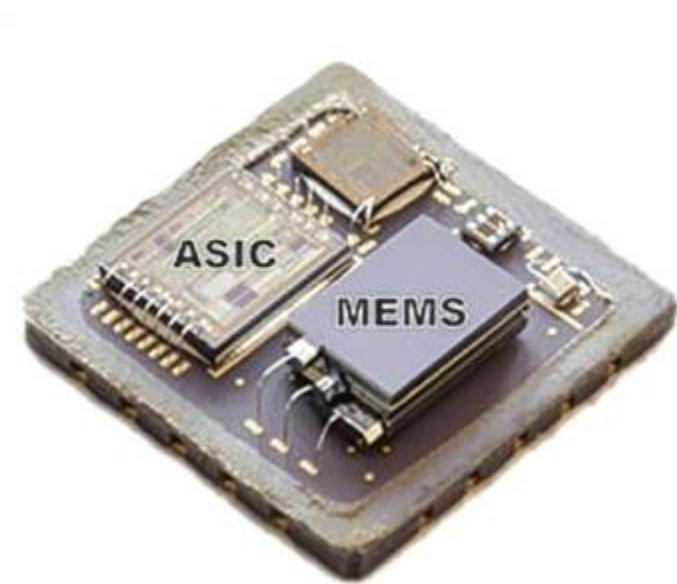
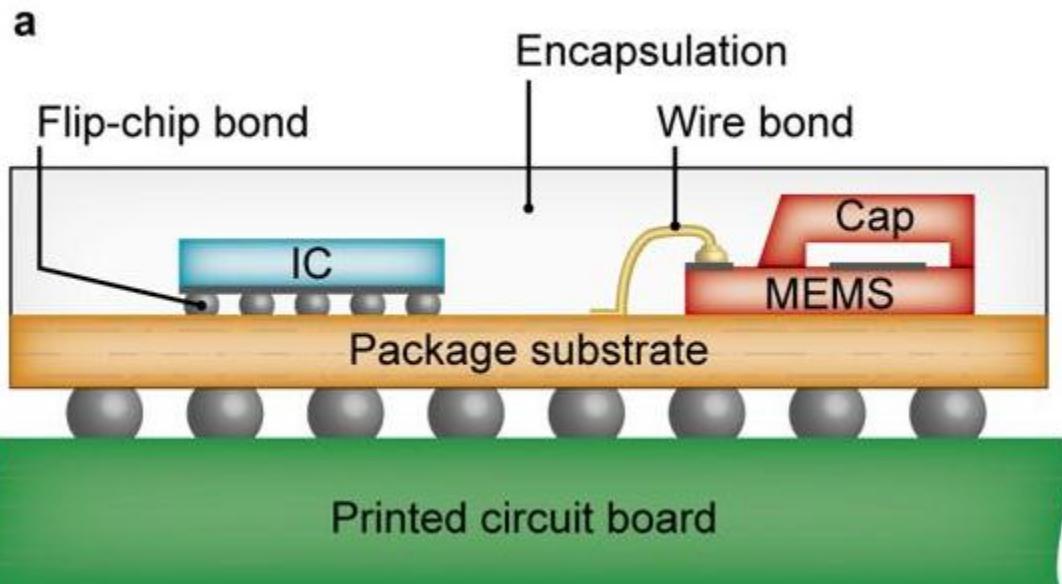
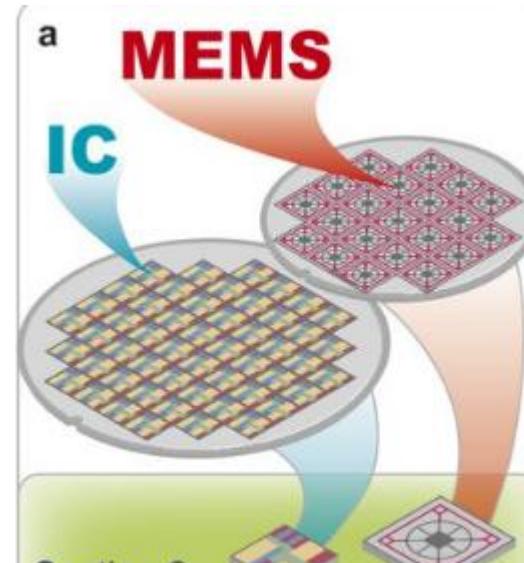




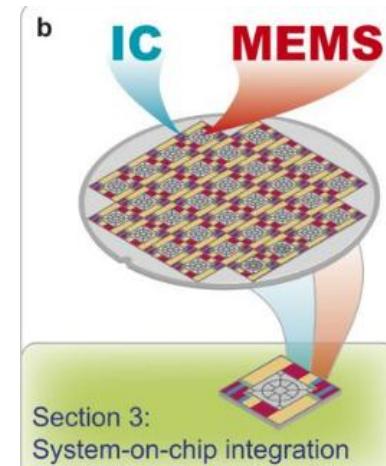
(a) **Comparison among various printing principles** including inkjet and MI-printing; (b) Illustration of the principle of the MI-printing process(The MI-printing process utilizes the kinetic energy from a moving pin, driven by an electromagnetic actuator, to strike on an elastic membrane above the microfluidic channel with an aligned nozzle

How to Integrate MEMS and ICs

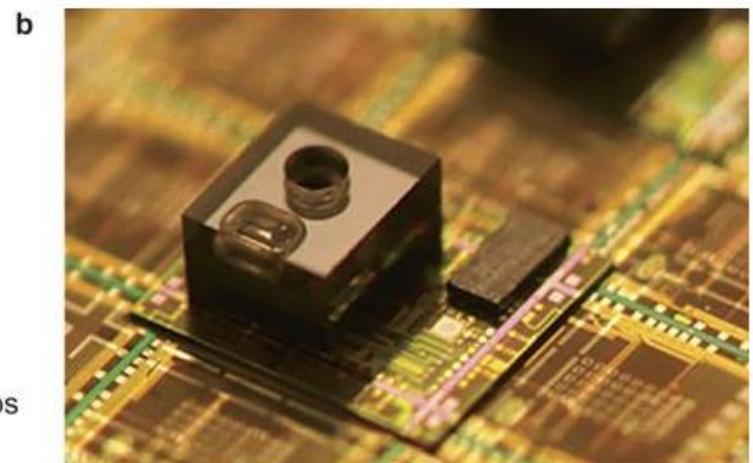
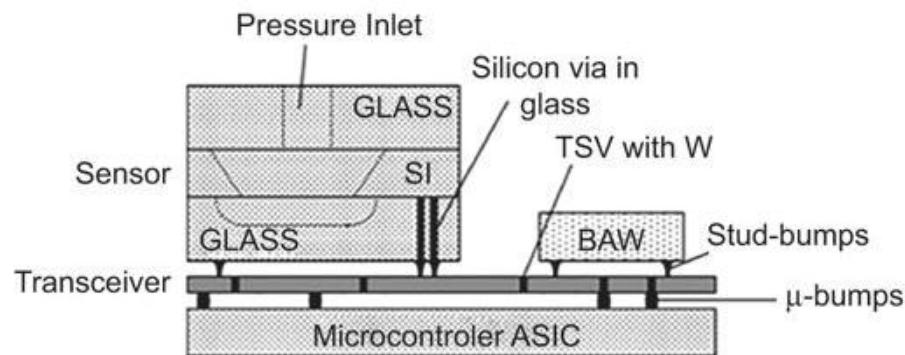
Fabricate MEMS and ICs separately and package as given below



Fabricate MEMS and IC on same substrate(silicon)



3D sensor node stack



(a) Sketch of a 3D sensor node stack fabricated using through-silicon vias. (b) Photograph of a 3D sensor node stack fabricated using through-silicon vias at the wafer level. From Ref 62.



Nanotechnology: product prospects - application trends



MOORE'S LAW

"Transistor density on integrated circuits doubles about every two years." *

1950s

Silicon
Transistor



1
Transistor

1960s

TTL
Quad Gate



16
Transistors

1970s

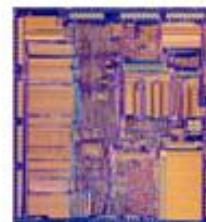
8-bit
Microprocessor



4500
Transistors

1980s

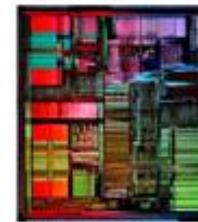
32-bit
Microprocessor



275,000
Transistors

1990s

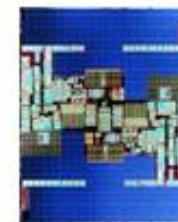
32-bit
Microprocessor



3,100,000
Transistors

2000s

64-bit
Microprocessor



592,000,000
Transistors



Qualcomm
snapdragon

Qualcomm Snapdragon 855

8,500,000,000

Principles of Self-organization

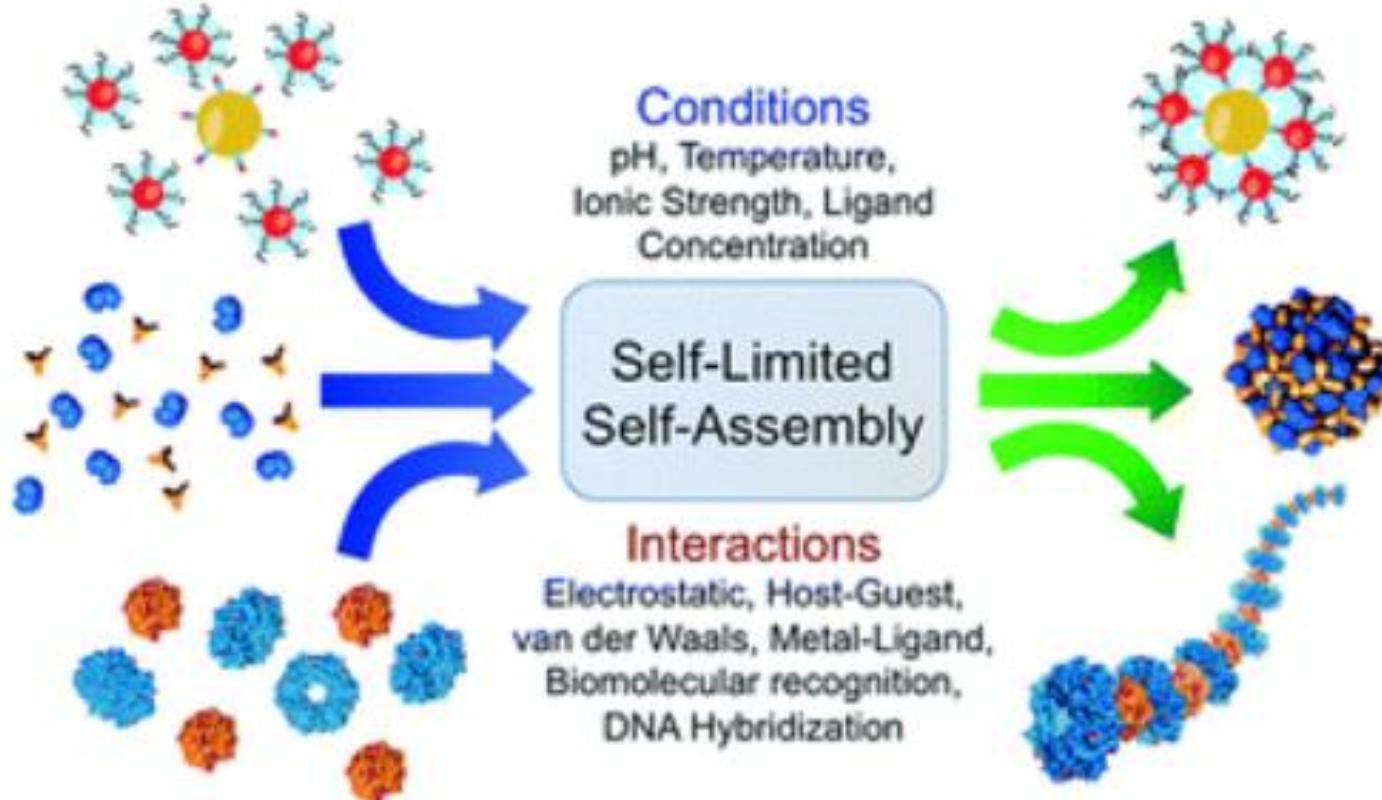
With the use of atomic and molecular architectures, new material structures are being provided target-selectively for the execution of well defined chemical, physical, and biological functions. These systems have new capabilities. They are among those intelligent products that record and evaluate data and execute actions resulting from this.

For example, **molecules can become electric switches or can be used for information reception, storage, and transmission.** Molecular architectures can also be used for pattern recognition, self-structurization, self- organization, self-reproduction, for the selective linking together of atoms, or for the creation or separation of molecular groups (clusters).

By the fabrication of new structurally defined molecular species, metals can be transformed into semiconductors or into optical or “switch- able” industrial materials. Filters or membranes become controllable, since they close their pores upon contact with certain molecules.

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

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



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

Creation of Clusters and Nanocrystalline Materials

Clusters are agglomerates consisting of about 10- 1000 atoms. With these, the dependence of their physical and chemical properties on cluster size is directly correlated with their major surface-to-volume ratio. There are many different techniques that can be used to make clusters in the condensed phase, in the gaseous phase, or in vacuum. The scientific and technical interest in clusters is in essence based on their physical and chemical properties, which specifically for clusters consisting of up to about 100 atoms may depend greatly on their size.

Thus, for example, structure, stability, electronic state density, optical absorption behavior, ionization potential, melting-point lowering, and also catalytic, magnetic, and chemical properties of cluster materials depend significantly on the number of particles in the cluster.

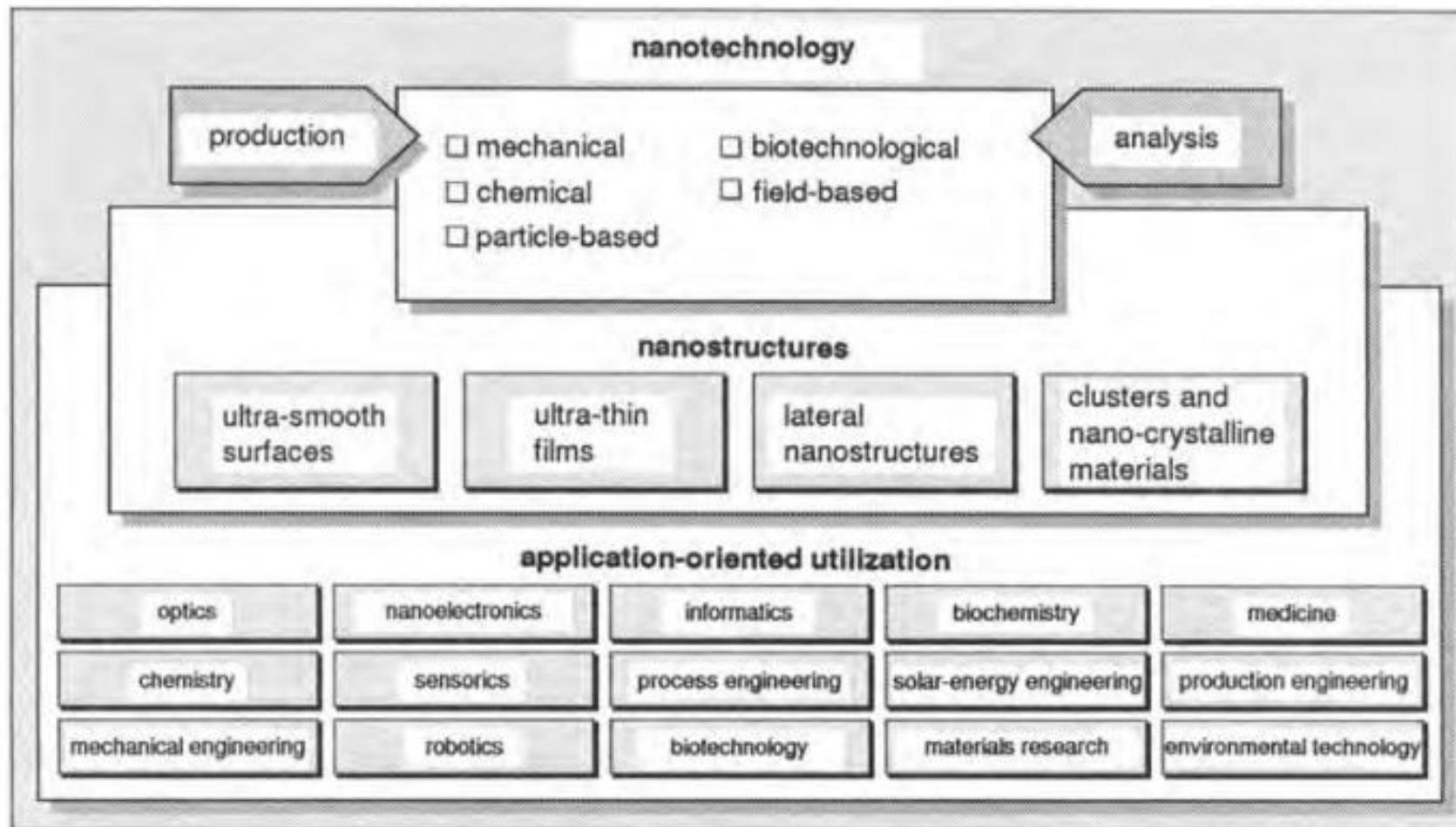
The Making of Ultrathin Films

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

The requirements placed on the making of films in nanotechnology are characterized by the demand for sharp atomic interfaces and for controlling depositions in atomically thin layers.

These are mostly vacuum methods, which are based either on **molecular beam epitaxy (MBE)** or **precipitation from the gaseous phase** (eg, **atomic layer epitaxy (ALE)**, **chemical beam epitaxy (CBE)**). For organic films the **Langmuir-Blodgett method** can be mentioned as an important preparation technique.

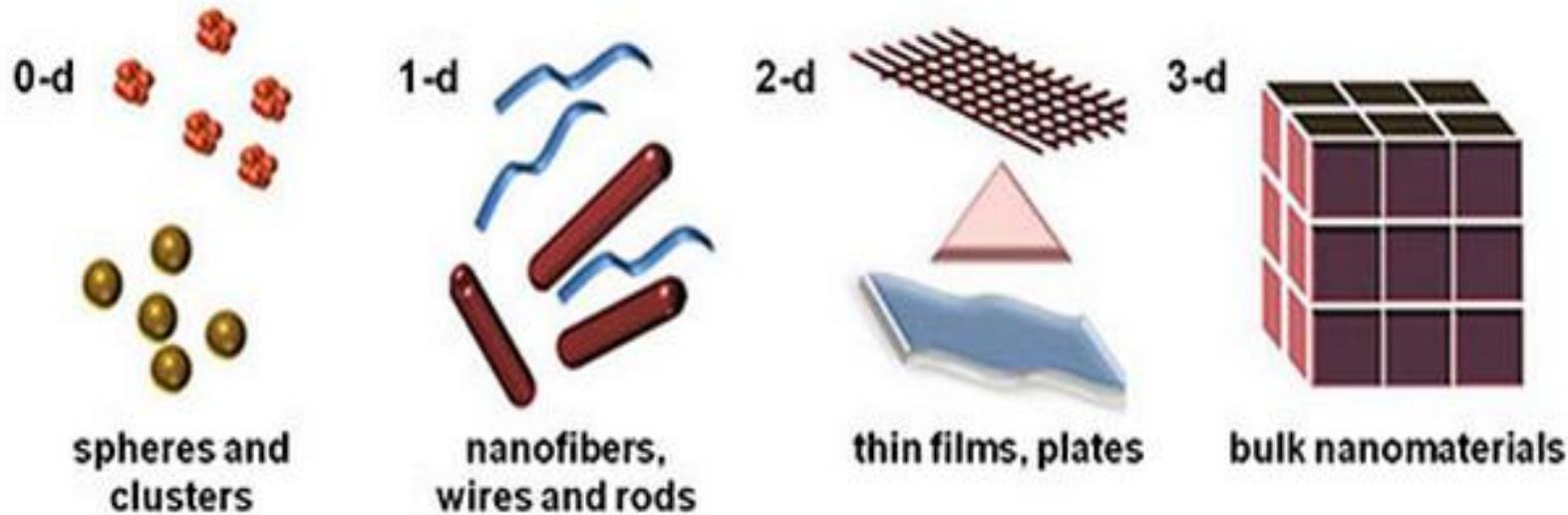
THE PRODUCTION, ANALYSIS, AND APPLICATION OF NANOTECHNOLOGY STRUCTURES



Clusters and Nanocrystalline Materials



Nanomaterials Classification



Three-dimensional (3D) structure or bulk structure

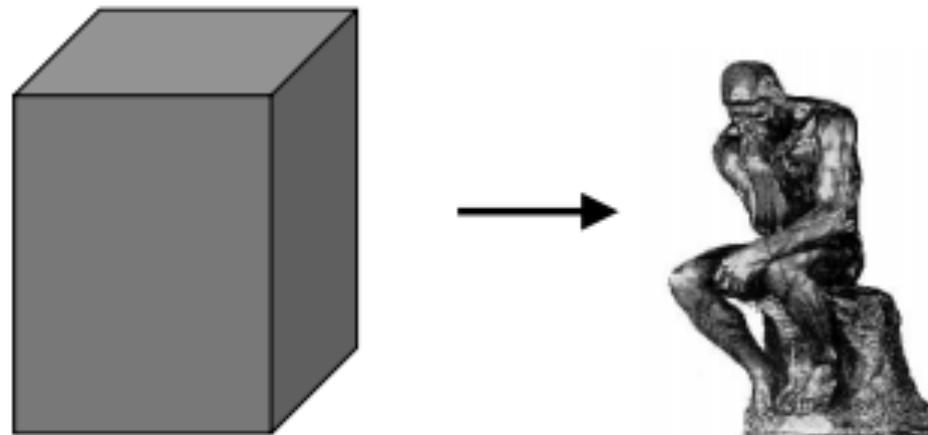
Two-dimensional (2D) structure or quantum well

One-dimensional (1D) structure or quantum wire

Zero-dimensional (0D) structure or quantum dot

How do you build something so small?

“Top-down” – building something by starting with a larger piece and carving away material (like a sculpture)



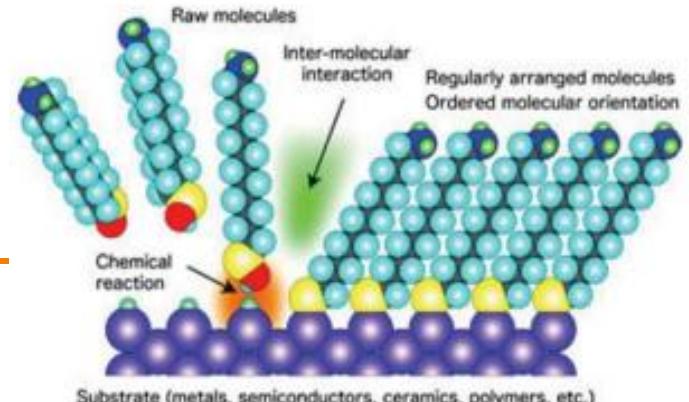
“Bottom-up” – building something by putting together smaller pieces (like building a car engine)



Bottom-up fabrication

- Adding atoms to atoms, molecules to molecules
- “Self-assembly” of atoms and molecules
- Use of chemical and biological processes

Self-assemble of organic monolayers for molecular transistors, etc



Types of Clusters

The clusters can be classified according to the type of chemical bonding between the atoms forming the aggregate

Van der Waals clusters

The interactions between inert gas atoms are weak and can be described accurately by central pair forces.

Metal clusters

The interatomic forces in metals are not simple. Many metals have non close-packed structures because the interatomic forces are partially directional

Clusters of ionic materials

Ionic materials are composed of closed shell ions, for instance Na^+ cations and Cl^- anions in the NaCl salt. In general these materials are formed from electropositive metals on the left side of the Periodic Table and electronegative elements on the right side. The cohesion in NaCl and similar clusters can be described by potentials composed of an attractive part due to electrostatic monopole forces and a repulsive part from the quantum mechanical overlap of the electronic clouds of ions with filled electronic shells.

Network clusters

Covalent bonding leads to the formation of atomic networks in clusters of materials like Si, Ge and C. Networks also form in the corresponding solid crystals and in the amorphous forms of those elements. Since many atoms in small clusters are on the cluster surface, these surface atoms have dangling bonds and important structural reconstruction is expected, as in the case of solid surfaces.

CLUSTERS AND NANOCRYSTALLINE MATERIALS

Introduction to Clusters

Atomic clusters are aggregates of atoms containing from few to a few thousand atoms. Due to their small size, the properties of the clusters are, in general, different from those of the corresponding material in the macroscopic bulk phase. The differences result from the fact that the number of atoms forming the surface is a substantial fraction of the number of atoms forming the cluster, while this fraction is negligible in the case of a macroscopic solid.

Cluster assembled solids

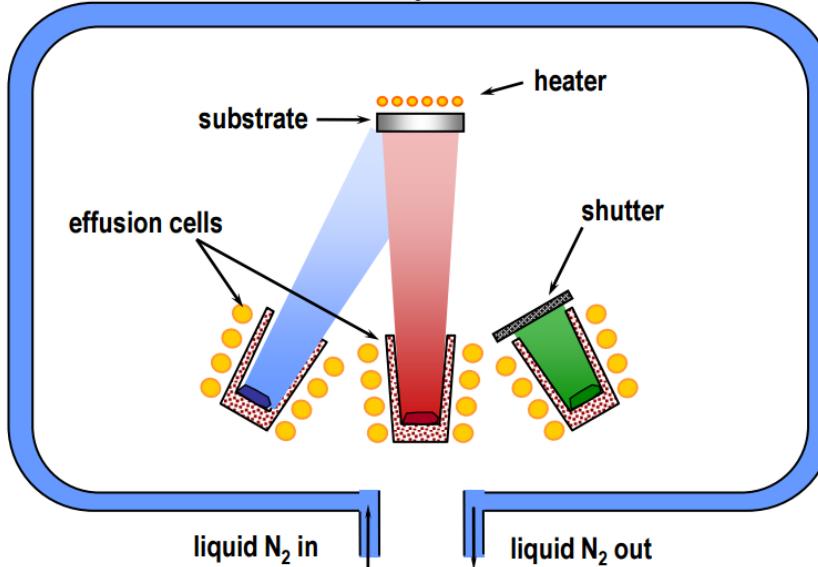
The possibility of building highly stable and symmetrical nanostructures makes the class of network clusters the most promising one for the purposes of cluster assembling. In fact the self-assembling of **C₆₀ clusters** to form the fullerite solid provides the best example.



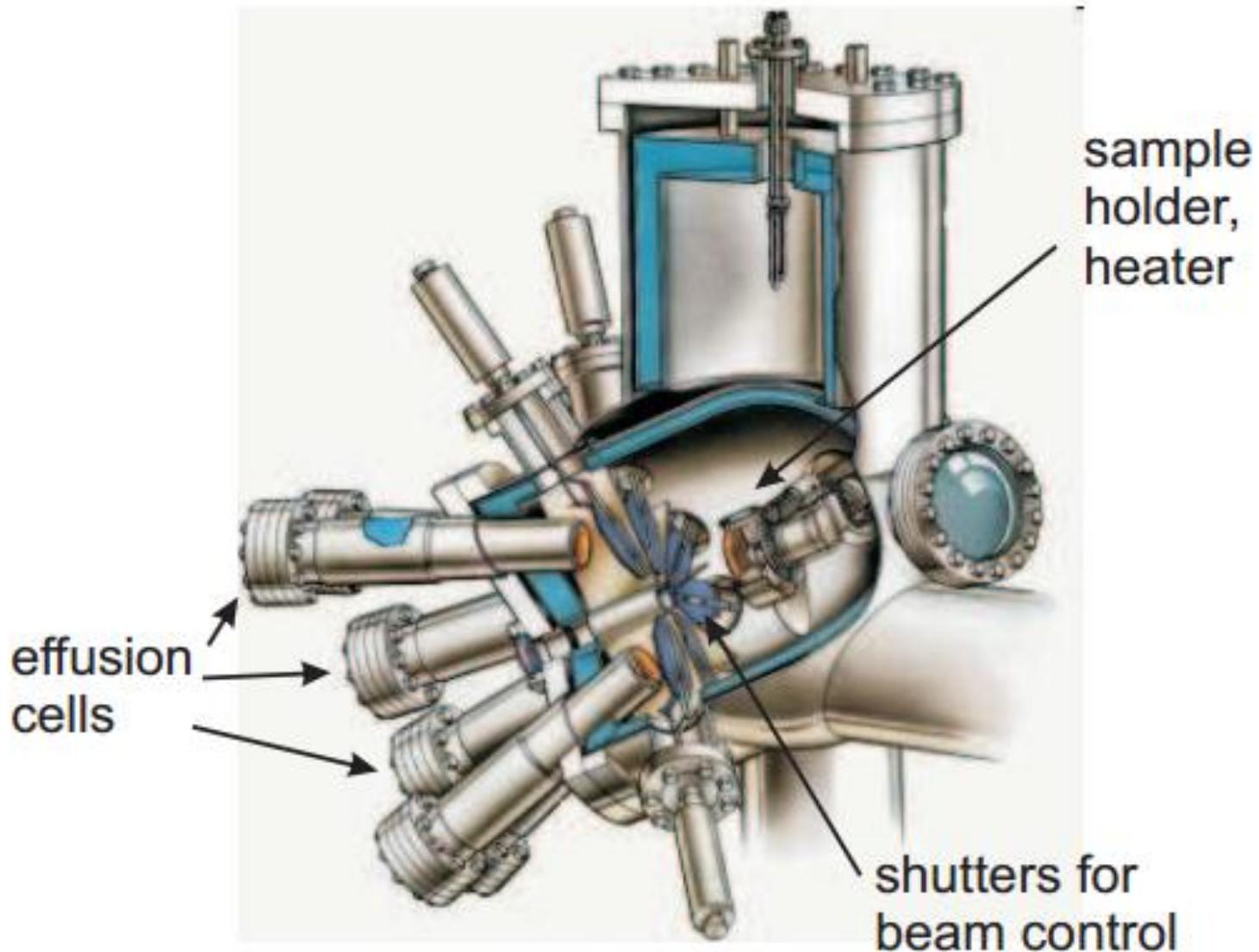
Procedures and techniques: the making of ultrathin films



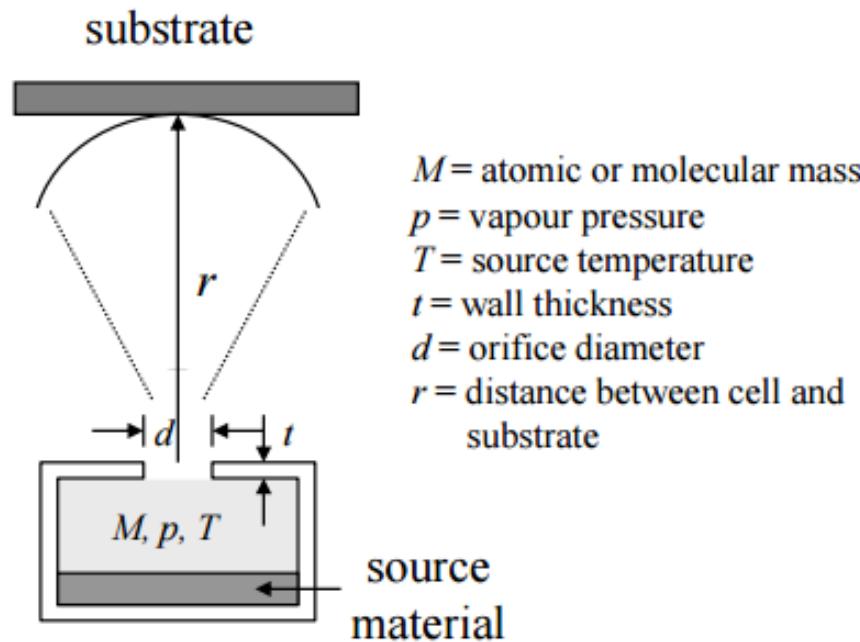
Molecular Beam Epitaxy



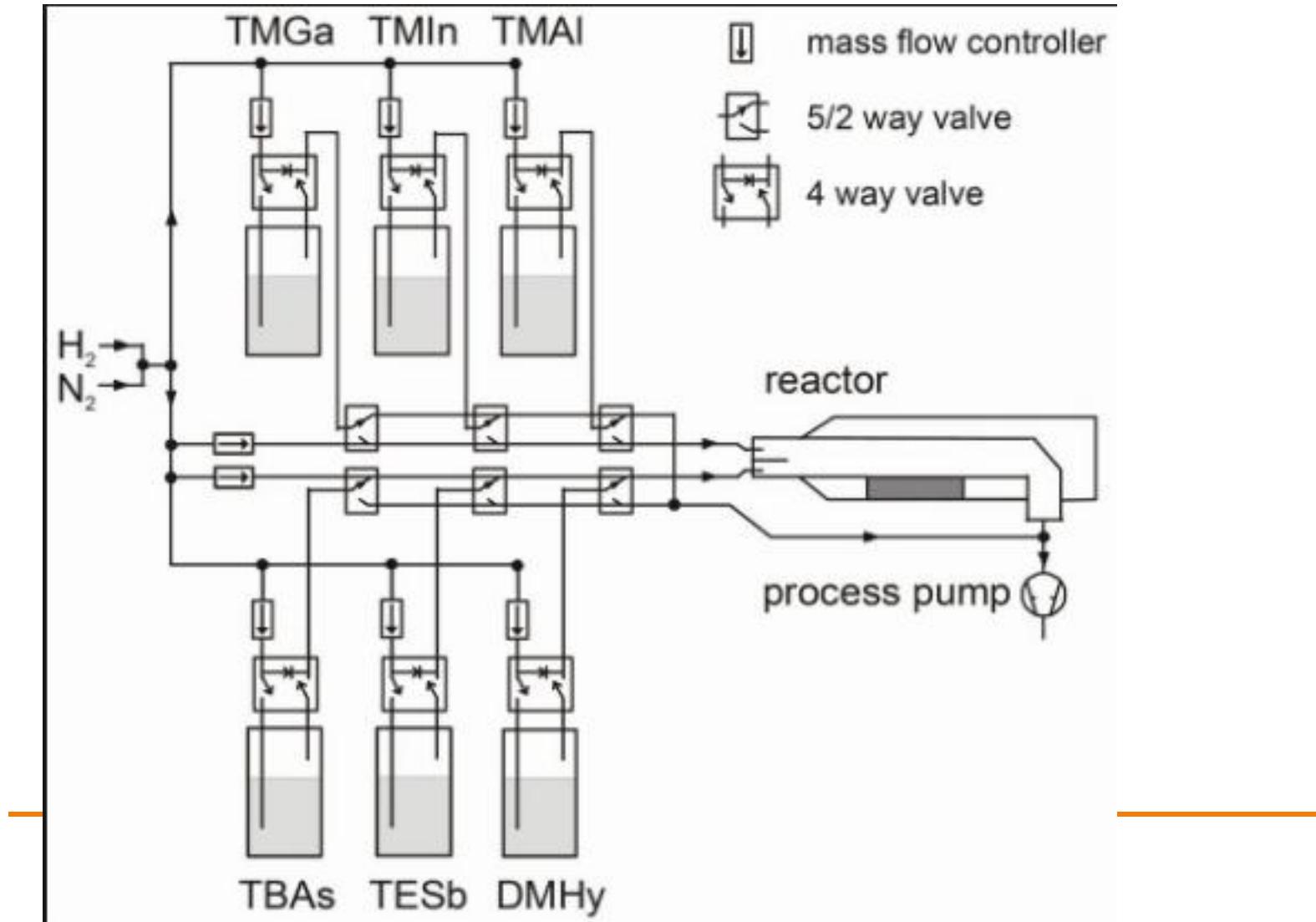
- Ultra-pure elements such as gallium and arsenic are heated in separate effusion cells until they begin to slowly sublime from the solid or evaporate from the liquid phase. The effusion cell temperature is typically used to control the flux of the atomic beam. For some molecules/atoms (e.g. nitrogen), gas cells are used as well.
- The composition of the different materials is controlled, using mechanical shutters in front of the output ports of the cells.
- The gaseous elements then condense on the substrate, where they may react with each other. In order to obtain a high mobility of the adatoms, the substrate is heated to high temperatures (typ. $\approx 300^\circ \text{ C}$ for II-VI materials, such as ZnSe, and $\approx 600^\circ \text{ C}$ for III-V materials, such as GaAs).
- In order to maintain the crystal structure, the substrates have to be mono-crystalline wafers with carefully cleaned surfaces.



*When molecules that are in a container, in which the flow is **molecular**, pass through a small orifice into another container at lower pressure, they are said to be **effusing** through the orifice.* The first study was made by Knudsen in 1909.



Metal-Organic Chemical Vapour Deposition (MOCVD)



In contrast to MBE, the growth of crystals is by chemical reaction instead of physical deposition. This takes place not in a vacuum, but from the gas phase at moderate pressures (2 to 100 kPa).

Metal-Organic Chemical Vapour Deposition (MOCVD), also known as Metal-Organic Vapour Phase Epitaxy (MOVPE), is a chemical vapour deposition method of epitaxial growth of materials, especially compound semiconductors from the surface reaction of organic compounds or metalorganics and metal hydrides containing the required chemical elements.

Metal Organic Chemical Vapor Phase Deposition (MOCVD) is a highly complex process for growing crystalline layers.

MOCVD is used in manufacturing light-emitting diodes (LEDs), lasers, transistors, solar cells and other electronic and opto-electronic devices, and is the key enabling technology for future markets with high growth potential.

It is the most significant manufacturing MOCVD process for III-V compound semiconductors, especially for those based on Gallium Nitride (GaN).

MOCVD is a process for manufacturing complex semiconductor multilayer structures used in electronic or optoelectronic components such as LEDs, lasers, high speed transistors or solar cells.

Unlike the better-known Silicon (used in the production of computer chips, for example), these semiconductors consist of not just one element, but rather of two or even more. They are therefore referred to as “compound semiconductors”. They include Gallium Arsenide (GaAs), Indium Phosphide (InP), Gallium Nitride (GaN) and related alloys.

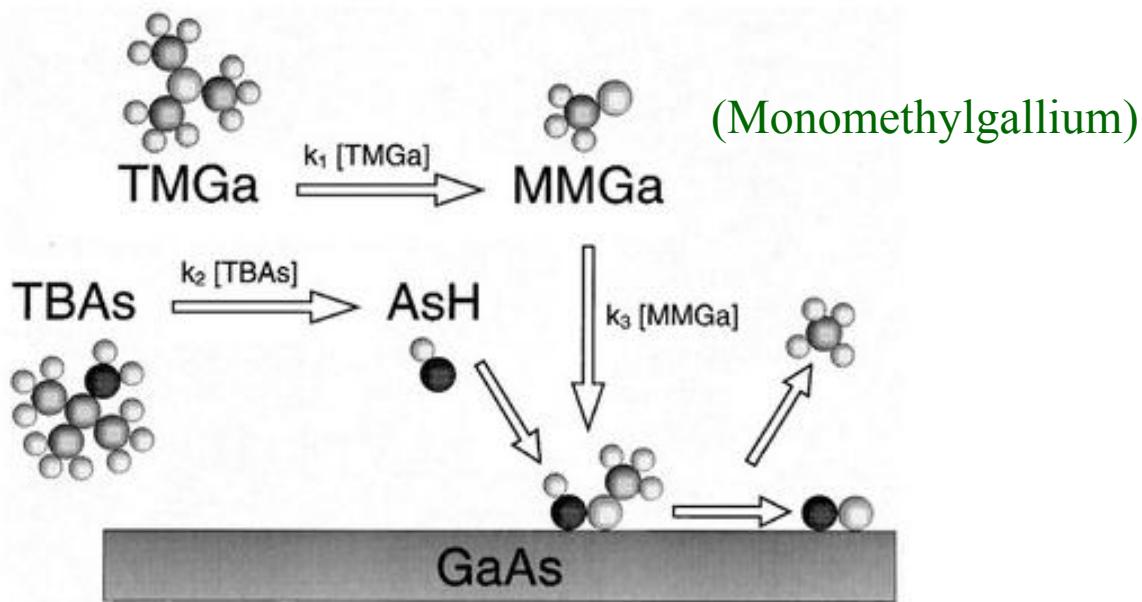
They are also called “III-V semiconductors” because they are made from elements of group III and V of the Periodic Table and can interact to form crystalline compounds.

Compound semiconductors have significant advantages over Silicon. Because electrons can move very fast in III-V materials, those devices containing III-V semiconductors can “process” the very high frequencies in mobile phones, for example.

Moreover, they can also function even at very high temperatures. Most importantly, they are highly efficient at converting light into electric power and vice-versa – this is what high-performance solar cells and all LEDs are based on.

To produce compound semiconductors, the chemicals are vaporized and transported into the reactor together with other gases. There, the critical chemical reaction takes place that turns the chemicals into the desired crystal (the compound semiconductor). In MOCVD the injected gases are ultra-pure and can be finely dosed.

Below is a schematic example of how GaAs is grown. Note that we use **tertiarybutylarsine (TBAs)** instead of AsH₃, and **tertiarybutylphosphine (TBP)** instead of PH₃ for lower toxicity and higher decomposition efficiency in this system.



For more detailed study refer the books mentioned in the syllabus

