1

INTRODUCTION TO SENSOR-BASED MEASUREMENT SYSTEMS

Measurements pervade our life. Industry, commerce, medicine, and science rely on measurements. Sensors enable measurements because they yield electric signals with embedded information about the measurand. Electronic circuits process those signals in order to extract that information. Hence, sensors are the basis of measurement systems. This chapter describes the basics of sensors, their static and dynamic characteristics, primary sensors for common quantities, and sensor materials and technology.

1.1 GENERAL CONCEPTS AND TERMINOLOGY

1.1.1 Measurement Systems

A system is a combination of two or more elements, subsystems, and parts necessary to carry out one or more functions. The function of a measurement system is the objective and empirical assignment of a number to a property or quality of an object or event in order to describe it. That is, the result of a measurement must be independent of the observer (objective) and experimentally based (empirical). Numerical quantities must fulfill the same relations fulfilled by the described properties. For example, if a given object has a property larger than the same property in another object, the numerical result when measuring the first object must exceed that when measuring the second object.

One objective of a measurement can be process monitoring: for example, ambient temperature measurement, gas and water volume measurement, and clinical monitoring. Another objective can be process control: for example, for temperature or level control in a tank. Another objective could be to assist

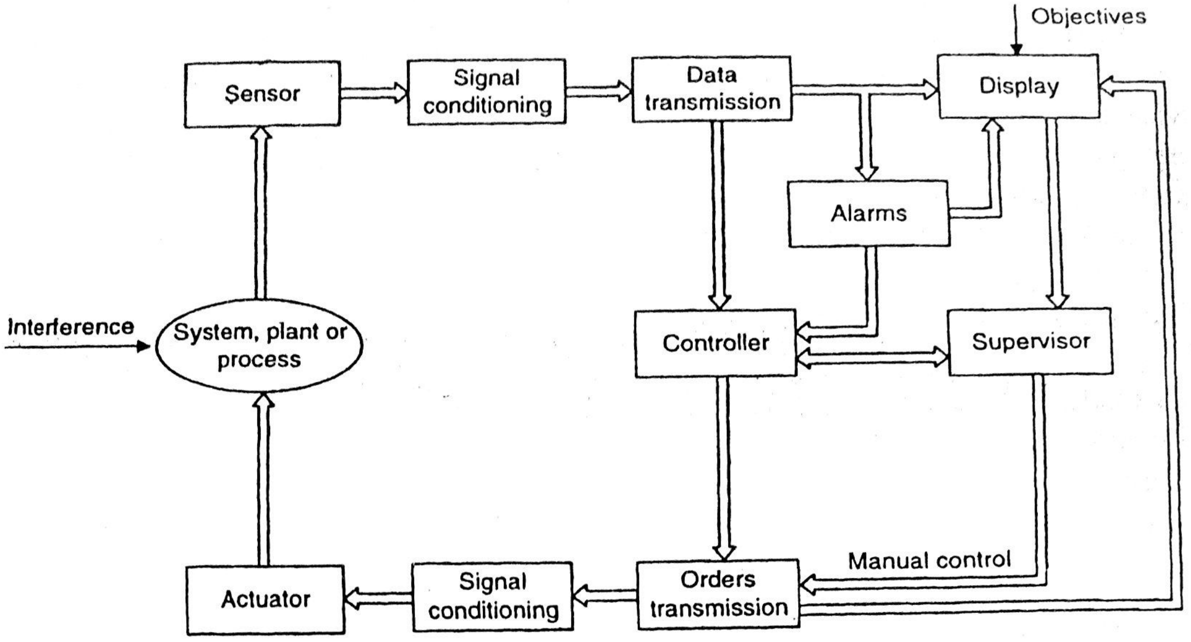


Figure 1.1 Functions and data flow in a measurement and control system. Sensors and actuators are transducers at the physical interface between electronic systems and processes or experiments.

experimental engineering: for example, to study temperature distribution inside an irregularly shaped object or to determine force distribution on a dummy driver in a car crash. Because of the nature of the desired information and its quantity, computer-aided design (CAD) does not yield complete data for these experiments. Thus measurements in prototypes are also necessary to verify the results of computer simulations.

Figure 1.1 shows the functions and data flow of a measurement and control system. In general, in addition to the acquisition of information carried out by a sensor, a measurement requires the processing of that information and the presentation of the result in order to make it perceptible to human senses. Any of these functions can be local or remote, but remote functions require information transmission. Modern measurement systems are not physically arranged according to the data flow in Figure 1.1 but are instead arranged according to their connection to the digital bus communicating different subsystems (Sections 8.6 and 8.7).

1.1.2 Transducers, Sensors, and Actuators

A *transducer* is a device that converts a signal from one physical form to a corresponding signal having a different physical form. Therefore, it is an energy converter. This means that the input signal always has energy or power; that is, signals consist of two component quantities whose product has energy or power dimension. But in measurement systems, one of the two components of the measured signal is usually so small that it is negligible, and thus only the remaining component is measured.

When measuring a force, for example, we assume that the displacement in the transducer is insignificant. That is, that there is no "loading" effect. Otherwise it might happen that the measured force is unable to deliver the needed energy to allow the movement. But there is always some power taken by the transducer, so we must ensure that the measured system is not perturbed by the measuring action.

Since there are six different kinds of signals—mechanical, thermal, magnetic, electric, chemical, and radiation (corpuscular and electromagnetic, including light)—any device converting signals of one kind to signals of a different kind is a transducer. The resulting signals can be of any useful physical form. Devices offering an electric output are called *sensors*. Most measurement systems use electric signals, and hence rely on sensors. Electronic measurement systems provide the following advantages:

1. Sensors can be designed for any nonelectric quantity, by selecting an appropriate material. Any variation in a nonelectric parameter implies a variation in an electric parameter because of the electronic structure of matter.
2. Energy does not need to be drained from the process being measured because sensor output signals can be amplified. Electronic amplifiers yield (low) power gains exceeding 10^{10} in a single stage. The energy of the amplifier output comes from its power supply. The amplifier input signal only controls (modulates) that energy.
3. There is a variety of integrated circuits available for electric signal conditioning or modification. Some sensors integrate these conditioners in a single package.
4. Many options exist for information display or recording by electronic means. These permit us to handle numerical data and text, graphics, and diagrams.
5. Signal transmission is more versatile for electric signals. Mechanical, hydraulic, or pneumatic signals may be appropriate in some circumstances, such as in environments where ionizing radiation or explosive atmospheres are present, but electric signals prevail.

Sensor and transducer are sometimes used as synonymous terms. However, sensor suggests the extension of our capacity to acquire information about physical quantities not perceived by human senses because of their subliminal nature or minuteness. Transducer implies that input and output quantities are not the same. A sensor may not be a transducer. The word *modifier* has been proposed for instances where input and output quantities are the same, but it has not been widely accepted.

The distinction between input-transducer (physical signal/electric signal) and output-transducer (electric signal/display or actuation) is seldom used at present. Nowadays, input transducers are termed *sensors*, or *detectors* for radiation, and output transducers are termed *actuators* or *effectors*. Sensors are intended to acquire information. Actuators are designed mainly for power conversion.

Sometimes, particularly when measuring mechanical quantities, a *primary sensor* converts the measurand into a measuring signal. Then a sensor would convert that signal into an electric signal. For example, a diaphragm is a primary sensor that stresses when subject to a pressure difference, and strain gages (Section 1.7.2 and Section 2.2) sense that stress. In this book we will designate as sensor the whole device, including the package and leads. We must realize, however, that we cannot directly perceive signals emerging from sensors unless they are further processed.

1.1.3 Signal Conditioning and Display

Signal conditioners are measuring system elements that start with an electric sensor output signal and then yield a signal suitable for transmission, display, or recording, or that better meet the requirements of a subsequent standard equipment or device. They normally consist of electronic circuits performing any of the following functions: amplification, level shifting, filtering, impedance matching, modulation, and demodulation. Some standards call the sensor plus signal conditioner subsystem a *transmitter*.

One of the stages of measuring systems is usually digital and the sensor output is analog. Analog-to-digital converters (ADCs) yield a digital code from an analog signal. ADCs have relatively low input impedance, and they require their input signal to be dc or slowly varying, with amplitude within specified margins, usually less than ± 10 V. Therefore, sensor output signals, which may have an amplitude in the millivolt range, must be conditioned before they can be applied to the ADC.

The display of measured results can be in an analog (optical, acoustic, or tactile) or in a digital (optical) form. The recording can be magnetic, electronic, or on paper, but the information to be recorded should always be in electrical form.

1.1.4 Interfaces, Data Domains, and Conversion

In measurement systems, the functions of signal sensing, conditioning, processing, and display are not always divided into physically distinct elements. Furthermore, the border between signal conditioning and processing may be indistinct. But generally there is a need for some signal processing of the sensor output signal before its end use. Some authors use the term *interface* to refer to signal-modifying elements that operate in the electrical domain, even when changing from one data domain to another, such as an ADC.

A *data domain* is the name of a quantity used to represent or transmit information. The concept of data domains and conversion between domains helps in describing sensors and electronic circuits associated with them [1]. Figure 1.2 shows some possible domains, most of which are electrical.

In the *analog domain* the information is carried by signal amplitude (i.e.,

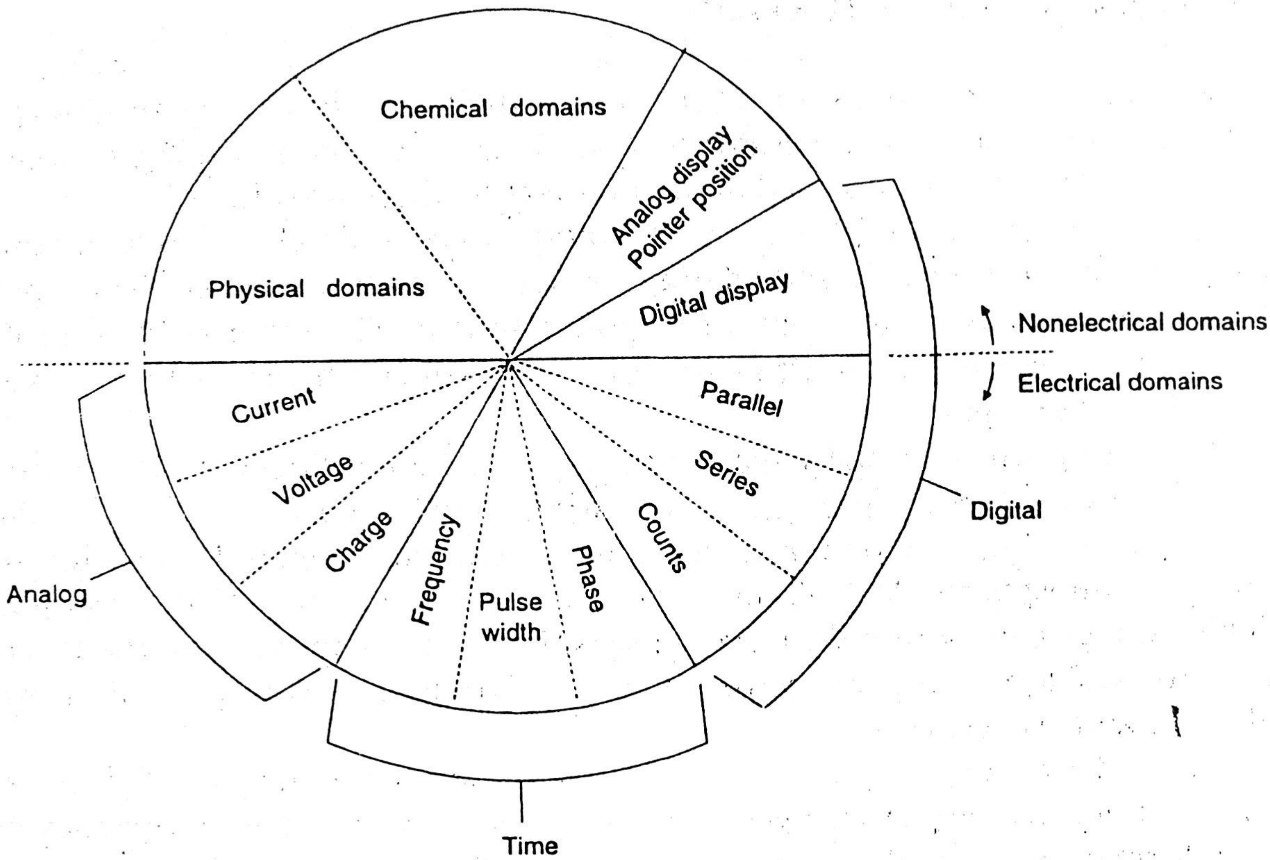


Figure 1.2 Data domains are quantities used to represent or transmit information [1]. (From H. V. Malmstadt, C. G. Enke, and S. R. Crouch, *Electronics and Instrumentation for Scientists*, copyright 1981. Reprinted by permission of Benjamin/Cummings, Menlo Park, CA.)

charge, voltage, current, or power). In the *time domain* the information is not carried by amplitude but by time relations (period or frequency, pulse width, or phase). In the *digital domain*, signals have only two values. The information can be carried by the number of pulses or by a coded serial or parallel word.

The analog domain is the most prone to electrical interference (Section 1.3.1). In the time domain, the coded variable cannot be measured—that is, converted to the numerical domain—in a continuous way. Rather, a cycle or pulse duration must elapse. In the digital domain, numbers are easily displayed.

The structure of a measurement system can be described then in terms of domain conversions and changes, depending on the direct or indirect nature of the measurement method.

Direct physical measurements yield quantitative information about a physical object or action by direct comparison with a reference quantity. This comparison is sometimes simply mechanical, as in a weighing scale.

In *indirect physical measurements* the quantity of interest is calculated by applying an equation that describes the law relating other quantities measured with a device, usually an electric one. For example, one measures the mechanical power transmitted by a shaft by multiplying the measured torque and speed of rotation, the electric resistance by dividing dc voltage by current, or the traveled distance by integrating the speed. Many measurements are indirect.

A great number of sensors are available for different physical quantities. In order to study them, it is advisable first to classify sensors according to some criterion. White [10] provides additional criteria to those used here.

In considering the need for a power supply, sensors are classified as modulating or self-generating. In modulating (or active) sensors, most of the output signal power comes from an auxiliary power source. The input only controls the output. Conversely, in self-generating (or passive) sensors, output power comes from the input.

Modulating sensors usually require more wires than self-generating sensors, because wires different from the signal wires supply power. Moreover, the presence of an auxiliary power source can increase the danger of explosion in explosive atmospheres. Modulating sensors have the advantage that the power supply voltage can modify their overall sensitivity. Some authors use the terms *active* for self-generating and *passive* for modulating. To avoid confusion, we will not use these terms.

In considering output signals, we classify sensors as analog or digital. In *analog sensors* the output changes in a continuous way at a macroscopic level. The information is usually obtained from the amplitude, although sensors with output in the time domain are usually considered as analog. Sensors whose output is a variable frequency are called *quasi-digital* because it is very easy to obtain a digital output from them (by counting for a time).

The output of *digital sensors* takes the form of discrete steps or states. Digital sensors do not require an ADC, and their output is easier to transmit than that of analog sensors. Digital output is also more repeatable and reliable and often more accurate. But regrettably, digital sensors cannot measure many physical quantities.

In considering the operating mode, sensors are classified in terms of their function in a deflection or a null mode. In *deflection sensors* the measured quantity produces a physical effect that generates in some part of the instrument a similar but opposing effect that is related to some useful variable. For example, a dynamometer to measure force is a sensor where the force to be measured deflects a spring to the point where the force it exerts, proportional to its deformation, balances the applied force.

Null-type sensors attempt to prevent deflection from the null point by applying a known effect that opposes that produced by the quantity being measured. There is an imbalance detector and some means to restore balance. In a weighing scale, for example, the placement of a mass on a pan produces an imbalance indicated by a pointer. The user has to place one or more calibrated weights on the other pan until a balance is reached, which can be observed from the pointer's position.

Null measurements are usually more accurate because the opposing known effect can be calibrated against a high-precision standard or a reference quantity. The imbalance detector only measures near zero; therefore it can be very

TABLE 1.1 Sensor Classifications According to Different Exhaustive Criteria

Criterion	Classes	Examples
Power supply	Modulating	Thermistor
	Self-generating	Thermocouple
Output signal	Analog	Potentiometer
	Digital	Position encoder
Operation mode	Deflection	Deflection accelerometer
	Null	Servo-accelerometer

sensitive and does not require any calibration. Nevertheless, null measurements are slow; and despite attempts at automation using a servomechanism, their response time is usually not as short as that of deflection systems.

In considering the input–output relationship, sensors can be classified as zero, first, second, or higher order (Section 1.5). The order is related to the number of independent energy-storing elements present in the sensor, and this affects its accuracy and speed. Such classification is important when the sensor is part of a closed-loop control system because excessive delay may lead to oscillation [6].

Table 1.1 compares the classification criteria above and gives examples for each type in different measurement situations. In order to study these myriad devices, it is customary to classify them according to the measurand. Consequently we speak of sensors for temperature, pressure, flow, level, humidity and moisture, pH, chemical composition, odor, position, velocity, acceleration, force, torque, density, and so forth. This classification, however, can hardly be exhaustive because of the seemingly unlimited number of measurable quantities. Consider, for example, the variety of pollutants in the air or the number of different proteins inside the human body whose detection is of interest.

Electronic engineers prefer to classify sensors according to the variable electrical quantity—resistance, capacity, inductance—and then to add sensors generating voltage, charge, or current, and other sensors not included in the preceding groups, mainly *p–n* junctions and radiation-based sensors. This approach reduces the number of groups and enables the direct study of the associated signal conditioners. Table 1.2 summarizes the usual sensors and sensing methods for common quantities.

1.7 PRIMARY SENSORS

Primary sensors convert measurands from physical quantities to other forms. We classify primary sensors here according to the measurand. Devices that

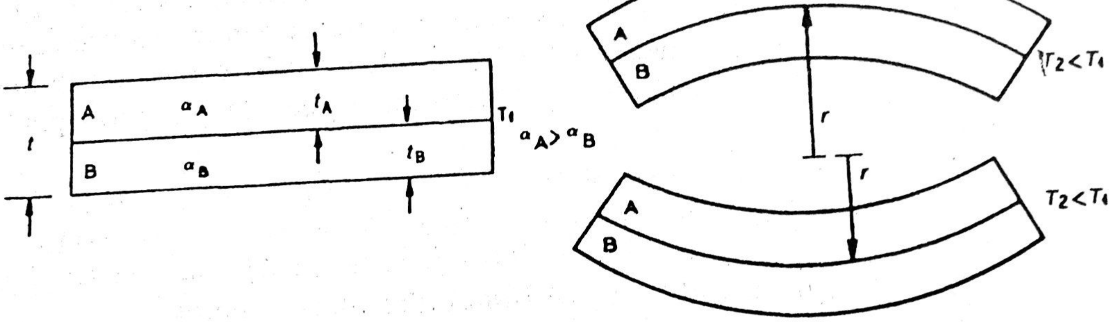


Figure 1.13 A bimetal consists of two metals with dissimilar thermal expansion coefficients, which deforms when temperature changes. Dimensions and curvature have been exaggerated to better illustrate the working principle. (From E. O. Doebelin, *Measurement Systems Application and Design*, 4th ed., copyright 1990. Reprinted by permission of McGraw-Hill, New York.)

have direct electric output are plain sensors and are discussed in Chapters 2, 4, and 6. Radiation-based measurement methods are described in Chapter 9. Khazan [11] and Fraden [12] describe additional primary sensors.

1.7.1 Temperature Sensors: Bimetals

A bimetal consists of two welded metal strips having different thermal expansion coefficients that are exposed to the same temperature. As temperature changes, the strip warps according to a uniform circular arc (Figure 1.13). If the metals have similar moduli of elasticity and thicknesses, the radius of curvature r, when changing from temperature T_1 to T_2, is [2]

$$r \cong \frac{2t}{3(\alpha_A - \alpha_B)(T_2 - T_1)} \tag{1.42}$$

where t is the total thickness of the piece and where α_A and α_B are the respective thermal expansion coefficients. Therefore the radius of curvature is inversely proportional to the temperature difference. A position or displacement sensor would yield a corresponding electric signal. Alternatively, the force exerted by a total or partially bonded or clamped element can be measured.

The thickness of common bimetal strips ranges from 10 μm to 3 mm. A metal having $\alpha_B < 0$ would yield a small r, hence high sensitivity. Because useful metals have $\alpha_B > 0$, bimetal strips combine a high-coefficient metal (proprietary iron–nickel–chrome alloys) with invar (steel and nickel alloy) that shows $\alpha = 1.7 \times 10^{-6}/°C$. Micromachined actuators (microvalves) use silicon and aluminum.

Bimetal strips are used in the range from −75 °C to +540 °C, and mostly from 0 °C to +300 °C. They are manufactured in the form of cantilever, spiral, helix, diaphragm, and so on, normally with a pointer fastened to one end of the strip, which indicates temperature on a dial. Bimetal strips are also used as actuators to directly open or close contacts (thermostats, on–off controls, starters for fluorescent lamps) and for overcurrent protection in electric circuits: The current along the bimetal heats it by Joule effect until reaching a temperature high enough to exert a mechanical force on a trigger device that opens the circuit and interrupts the current.

Other nonmeasurement applications of bimetal strips are the thermal compensation of temperature-sensitive devices and fire alarms. Their response is slow because of their large mass. Each October issue of *Measurements & Control* lists the manufacturers and types of bimetallic thermometers.

1.7.2 Pressure Sensors

Pressure measurement in liquids or gases is common, particularly in process control and in electronic engine control. Blood pressure measurement is very common for patient diagnosis and monitoring. Pressure is defined as the force per unit area. *Differential pressure* is the difference in pressure between two measurement points. *Gage pressure* is measured relative to atmospheric pressure. *Absolute pressure* is measured relative to a perfect vacuum. To measure a pressure, it is either compared with a known force or its effect on an elastic element is measured (deflection measurement). Table 1.7 shows some sensing

TABLE 1.7 Some Common Methods to Measure Fluid Pressure in Its Normal Range

1. Liquid column + level detection	
2. Elastic element	
2.1. Bourdon tube + displacement measurement:	Potentiometer
	LVDT
	Inductive sensor
	Digital encoder
2.2. Diaphragm + deformation measurement	
2.2.1. Central deformation[a]:	Potentiometer
	LVDT
	Inductive sensor
	Unbonded strain gages
	Cantilever and strain gages
	Vibrating wire
2.2.2. Global deformation:	Variable reluctance
	Capacitive sensor
	Optical sensor
	Piezoelectric sensor
2.2.3. Local deformation: strain gages:	Bonded foil
	Bonded semiconductor
	Deposited
	Sputtered (thin film)
	Diffused/implanted semiconductor

[a] Capsules and bellows yield larger displacements than diaphragms but suit only static pressures.

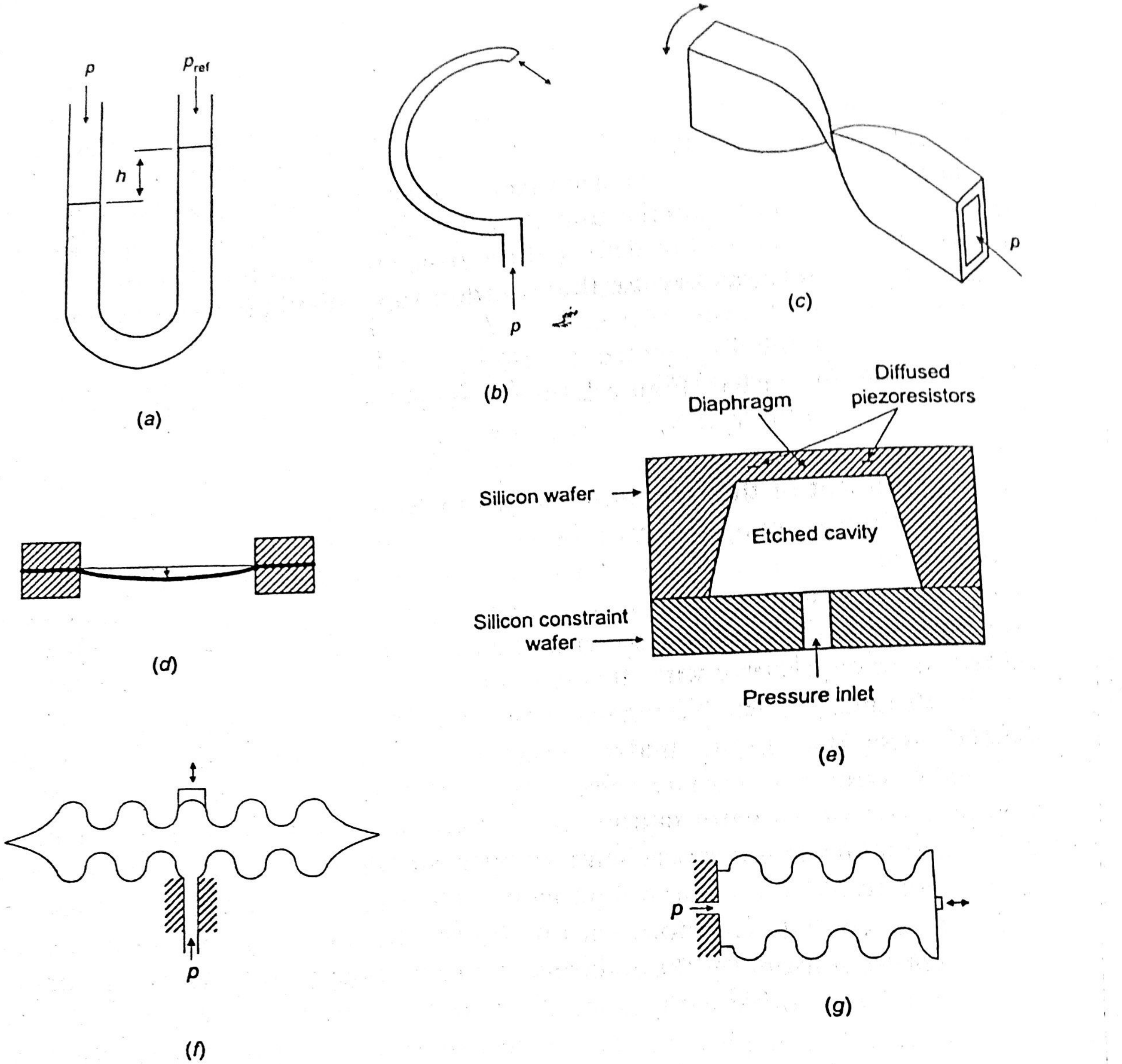


Figure 1.14 Primary pressure sensors. (*a*) Liquid-column U-tube manometer. The liquid must be compatible with the fluid for which pressure is to be measured, and the tube must withstand the mechanical stress. (*b*) C-shaped Bourdon tube. (*c*) Twisted Bourdon tube. (*d*) Membrane diaphragm. (*e*) Micromachined diaphragm. (*f*) Capsule. (*g*) Bellows. The area of the diaphragm in (*e*) is less than 1 mm^2. All other devices can measure up to several centimeters.

methods. Each issue of *Measurements & Control* lists the manufacturers of different pressure sensors: potentiometric (January); strain-gage and piezoresistive (April); capacitive (June); digital and reluctive (September); piezoelectric and liquid-column (October); and bellows, Bourdon tube, and diaphragm (December).

A liquid-column manometer such as the U-tube in Figure 1.14*a* compares the pressure to be measured with a reference pressure and yields a difference h of liquid level. When second order effects are disregarded, the result is

$$h = \frac{p - p_{\text{ref}}}{\rho g} \tag{1.43}$$

where ρ is the density of the liquid and g is the acceleration of gravity. A level sensor (photoelectric, float, etc.) yields an electric output signal.

Elastic elements deform under pressure until the internal stress balances the applied pressure. The material and its geometry determine the amplitude of the resulting displacement or deformation, hence the appropriate sensor (Table 1.7). Usual pressure sensors use the Bourdon tube, diaphragms, capsules, and bellows.

The Bourdon tube—patented by Eugene Bourdon in 1849—is a curved (Figure 1.14*b*) or twisted (Figure 1.14*c*), flattened metallic tube with one closed end. The tube is obtained by deforming a tube having a circular cross section. When pressure is applied through the open end, the tube tends to straighten. The displacement of the free end indicates the pressure applied. This displacement is not linear along its entire range, but is linear enough in short ranges. Displacement sensors yield an electric output signal. Tube configurations with greater displacements (spiral, helical) have large compliance and length that result in a small-frequency passband. The tube metal (brass, monel, steel) is selected to be compatible with the medium.

A diaphragm is a flexible circular plate consisting of a taut membrane or a clamped sheet that strains under the action of the pressure difference to be measured (Figure 1.14*d*). The sensor detects the deflection of the center of the diaphragm, its global deformation, or the local strain (by strain gages, Section 2.2). Some metals used are beryllium–copper, stainless steel, and nickel–copper alloys. A micromachined diaphragm is an etched silicon wafer with diffused or implanted gages that sense local strain (Figure 1.14*e*). Cars and hospitals use silicon pressure sensors by the millions. The diaphragm and elements bonded on it must be compatible with the medium and withstand the required temperature. Stainless steel diaphragms can protect sensing diaphragms from corrosive media, but in order to couple both diaphragms we need to interpose a fluid, which increases the sensor compliance and thermal sensitivity. Ceramic (96% Al_2O_3, 4% SiO_2) and sapphire (Al_2O_3) are highly immune to corrosive attack; but because they are very expensive, their use is restricted to the more demanding applications involving aggressive media, high temperature, or both.

For a thin plate with thickness t and radius R experiencing a pressure difference Δp across it, if the center deflection is $z < t/3$, we have [2]

$$z \cong \frac{3(1 - \nu^2)R^4}{16Et^3}\Delta p \tag{1.44}$$

where E is Young's modulus and ν the Poisson's ratio for the plate material. Large, flexible diaphragms undergo large deflection but have large compliance. Thin plates yield large deflections but are fragile. An alternative to sense the central deflection is to use a rod to transmit force to a cantilever beam with bonded strain gages, away from media temperature. Ceramic and some silicon pressure sensors rely on the capacitance change between an electrode applied on the diaphragm and one fixed electrode.

Piezoresistive sensors distributed on the diaphragm can sense radial and tangential strain. They are connected in a measurement bridge to add their signal and compensate temperature interference (Section 3.4.4).

Capsules and bellows yield larger displacements than diaphragms. A capsule (Figure 1.14*f*) consists of twin corrugated diaphragms joined by their external border and placed on opposite sides of the same chamber. A bellows (Figure 1.14*g*) is a flexible chamber with axial elongation that undergoes deflections larger than capsules, up to 10% of its length. Capsules and bellows are vibration- and acceleration-sensitive, do not withstand high overpressures, and have high compliance, hence poor dynamic response. Their displacement, however, can be sensed by an inexpensive potentiometer.

Pressure between contacting surfaces can be measured by a thin plastic film (Fuji Prescale Film) whose color increases for increasing pressure.

1.7.3 Flow Velocity and Flow-Rate Sensors

Flow is the movement of a fluid in a channel or in open or closed conduits. The flow rate is the quantity of matter, in volume or weight, that flows in a unit time. Flow rate is measured in all energy and mass transport processes to control or monitor those processes and for metering purposes—for example, water, gas, gasoline, diesel, and crude oil. Table 1.8 lists some measurement principles used in flowmeters. Chapters 28 and 29 in reference 13 discuss them. Each issue of *Measurements & Control* lists the manufacturers of different flowmeters: turbine (February); electromagnetic (April); anemometers and vortex (June); differential pressure, rotameters, and mass (September); positive displacement and ultrasonic (October); and open-channel, target, and flowmeters based on laminar flow elements (December).

A viscous or laminar flow is that of a fluid flowing along a straight smooth-walled and uniform transverse section conduit, where all particles have a trajectory parallel to the conduit walls and move in the same direction, each following a streamline. In turbulent flow, in contrast, some of the fluid particles have longitudinal and transverse velocity components—thus resulting in whirls—and only the average velocity is parallel to the axis of the conduit. In laminar flow, the fluid velocity profile across the conduit is parabolic. In turbulent flow, the fluid velocity profile is flatter.

The commonest flowmeters measure the drop in pressure across an obstruction inserted in the pressurized pipe in which we wish to measure the flow rate. Bernoulli's theorem relates fluid pressure, velocity, and height. It applies to an incompressible fluid experiencing only gravity as internal force (i.e., without friction) flowing in stationary movement and with no heat entering or leaving it. Any change in velocity produces an opposite change in pressure that equals the change of kinetic energy per unit of volume added to the change due to any

TABLE 1.8 Measurement Principles Used in Flowmeters

Input Quantity	Measurement Principle	Output Signal
Fluid velocity: local	Pitot probe	Differential pressure
	Thermal (hot wire anemometry)	Temperature
	Laser anemometry	Frequency shift
Fluid velocity: average	Electromagnetic	Voltage
	Ultrasound: transit time	Time
	Ultrasound: Doppler	Frequency
Volume flow rate[a]	Orifice plate	Differential pressure
	Venturi tube	Differential pressure
	Pitot probe	Differential pressure
	Flow nozzle and tube	Differential pressure
	Elbow	Differential pressure
	Laminar flow element	Differential pressure
	Impeller (paddlewheel)	Cycles, revolutions
	Positive displacement	Cycles, revolutions
	Target (drag force)	Force
	Turbine	Cycles, revolutions
	Variable area (rotameter)	Float displacement
	Variable area (weir, flume)	Level
	Vortex shedding	Frequency shift
Mass flow rate	Coriolis effect	Force
	Thermal transport	Temperature

[a] Volume flow rate can also be calculated by multiplying the average fluid velocity by the pipe cross section.

difference in level. That is, along a flow streamline we have

$$p + \rho g h + \frac{\rho v^2}{2} = \text{constant} \tag{1.45}$$

where p is the static pressure, ρ is the fluid density (incompressible), g is the acceleration of gravity, h is the height with respect to a reference level, and v is the fluid velocity at the point considered. When studying actual fluid flows, (1.45) is corrected by experimental coefficients.

The primary sensor in obstruction flowmeters is a restriction having constant cross section that obstructs the flow. For example, if we introduce in a pipe a plate having a hole, the fluid vein contracts, thereby changing from a cross section A_1 (that of the pipe) to a cross section A_2 (that of the hole) (Figure 1.15). Because of the principle of mass conservation, a cross-section change results in a corresponding change of velocity,

$$Q = A_1 v_1 = A_2 v_2 \tag{1.46}$$

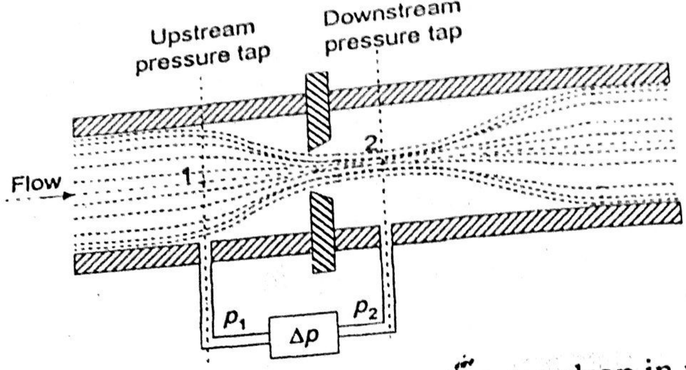


Figure 1.15 An orifice plate inserted in a pipe produces a drop in pressure related to the flow rate.

At the same time, from (1.45) we have

$$p_1 + \rho g h_1 + \frac{\rho v_1^2}{2} = p_2 + \rho g h_2 + \frac{\rho v_2^2}{2} \tag{1.47}$$

If $h_1 = h_2$, these two equations yield

$$v_2 = \sqrt{\frac{2(p_1 - p_2)}{\rho\left[1 - \left(\frac{A_1}{A_2}\right)^2\right]}} \tag{1.48}$$

Therefore, we can calculate the velocity from the drop in pressure across the plate, and we can determine the theoretical volumetric flow rate from $Q = A_2 v_2$. The real flow rate is somewhat lower and it is determined by experimentally calculating a correction coefficient, called a discharge factor, C_d, that depends on A_1, A_2, and other parameters. Then, $Q_r = C_d Q$. Tables in standards (ASME, ISO) give C_d for different pipe diameter and hole position and size, flow regimes, and pressure ports placement. For orifice plates we have $C_d \approx 0.6$.

Orifice meters produce a loss in pressure and cannot easily measure fluctuating flows unless the differential pressure sensor is fast enough, including the effects of the hydraulic connections. Flow nozzles and Venturi tubes (Figure 1.16) are based on the same principles but their internal shapes are not so blunt, thus reducing the loss of pressure (C_d can reach 0.97).

Variable-area flowmeters are primary sensors that apply Bernoulli's theorem and the principle of mass conservation in a way reciprocal to that described. They make the fluid pass section variable and keep the difference in pressure between both sides of the obstruction constant. The measured flow rate is then related to the area of the pass section.

The rotameter in Figure 1.17 applies this method. It consists of a uniform conic section tube and a grooved float inside it that is dragged by the fluid to a

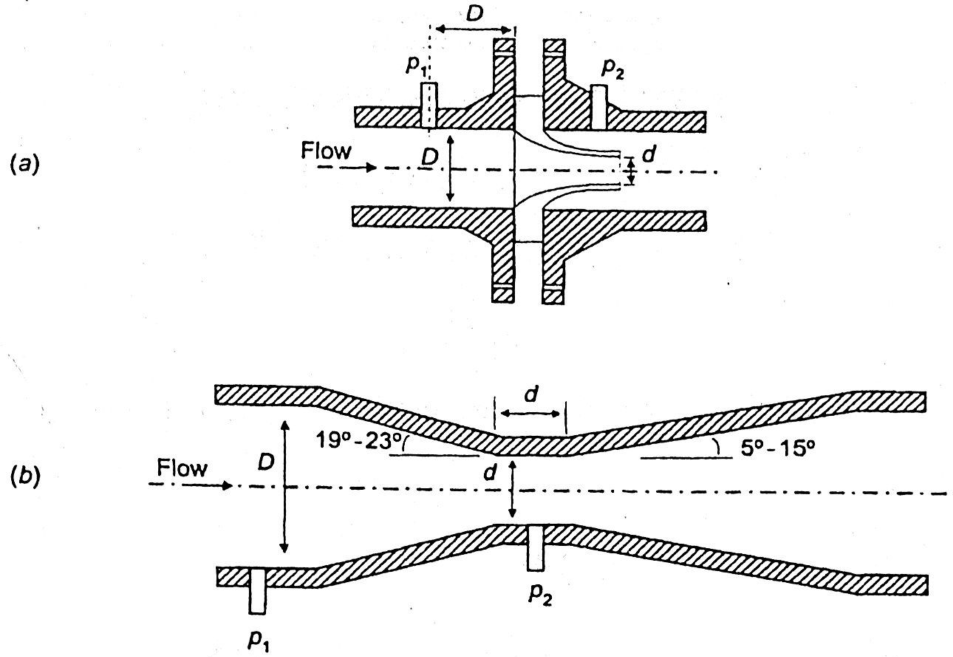


Figure 1.16 Flow nozzles (*a*) and Venturi tubes (*b*) inserted in pipes yield a lower drop in pressure than orifice plates, hence saving energy.

height determined by its weight and the flow. The fluid—gas or liquid—flows upward. When the flow increases, the float rises, thus allowing an increased annular pass section and keeping the pressure difference between both ends constant. The displacement of the float indicates the fluid flow rate. For pressures lower than 3.5 kPa and nonopaque liquids, the tube can be of glass and include the scale to read the float position. For higher pressures and flows the tube must be of metal, and the position of the float is detected magnetically. There are also inexpensive plastic tubes for low-pressure, high flow rates. Adding a solenoid outside the tube enables us to apply the null-measurement method. A photoelectric detector measures the float position. The flow is determined from the amplitude of the current supplied to the solenoid in order to reposition the float at zero.

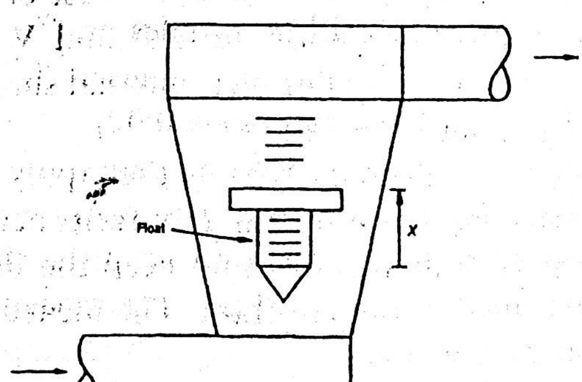


Figure 1.17 A rotameter is a variable area flowmeter in which the position of a float indicates the flow rate.

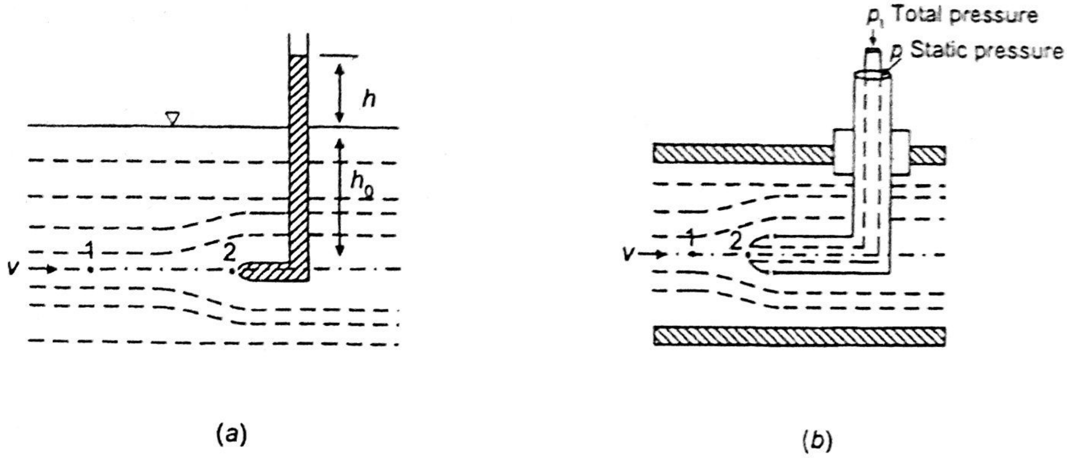


Figure 1.18 Pitot tube for point velocity flow measurement. (*a*) In an open conduit the velocity is indicated by the emerging fluid height. (*b*) In a closed conduit the velocity is calculated from the difference between total pressure and static pressure.

The Pitot tube used to measure the velocity of a fluid at a point also relies on Bernoulli's principle. If a bent open tube is introduced into an open conduit where an incompressible and frictionless fluid flows in a given known direction, and its open end is oriented against the flow (Figure 1.18*a*), the liquid enters into the tube and rises until the pressure exerted by the fluid column balances the force produced by the impacting velocity on the open end. Because in front of the opening the velocity is zero, flow lines distribute around the end, thereby creating a stagnation point. It holds therefore that

$$\frac{v^2}{2g} + \frac{p_1}{\rho g} = \frac{p_2}{\rho g} = h_0 + h \tag{1.49}$$

Also the static pressure in an open conduit comes from the weight of the fluid column, $p_1 = \rho g h_0$. Therefore

$$v = \sqrt{2gh} \tag{1.50}$$

We can thus infer the fluid velocity at the measurement point from the height of the column emerging above the surface.

If the Pitot tube is placed in a pressurized pipe, from (1.45) we obtain

$$v = \sqrt{\frac{2(p_t - p)}{\rho}} \tag{1.51}$$

Therefore, in order to determine the velocity we need to measure the difference between the total or stagnation pressure p_t and the static pressure p, which can be obtained from a port which faces perpendicular to the flow—for example, through a coaxial tube (Figure 1.18*b*). Pitot tubes are very common in laboratories and also for air speed measurement in avionics—in this last case using a modified version of (1.51) that includes temperature and specific heat because air is compressible.

Laminar flowmeters, also called laminar resistance flowmeters, rely on the Poiseuille's law. Jean M. Poiseuille—a physician—established in 1840 that for laminar flow in a tube much longer than wide, the volumetric flow rate is a linear function of the pressure drop according to

$$\Delta p = Q\frac{8\eta L}{\pi r^4} \tag{1.52}$$

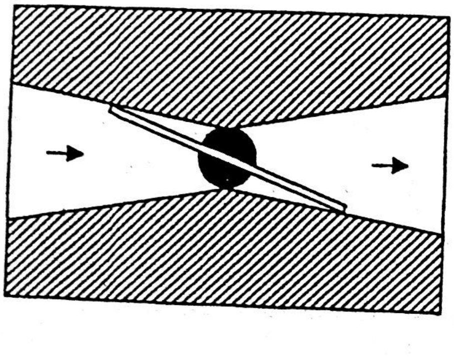
where η is the fluid viscosity, L is the tube length, and r is its radius. Laminar flowmeters consist of a bundle or a matrix of capillary tubes, or one or more fine mesh screens and two pressure connections. They are used for leak testing, for calibration work, and in respiratory pneumotachometers.

Target flowmeters sense the fluid force on a target or drag-disk suspended in the flow stream by a sensing tube. The force exerted on the target is measured by strain gages (Section 2.2) placed on the tube, outside the pipe. Target flowmeters can be applied to dirty or corrosive liquids.

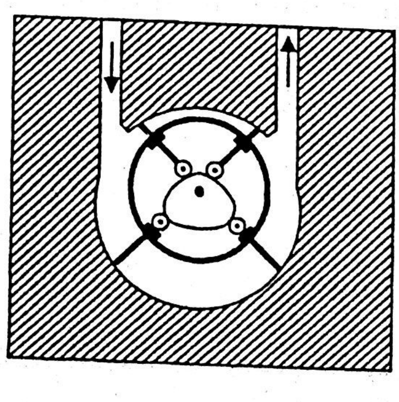
Turbine flowmeters consist of a bladed rotor suspended in a moving (clean) fluid that makes it turn at a speed proportional to the volumetric flow rate when it is high enough. The turning velocity is detected by a variable reluctance pickup. Vane flowmeters rely on the same principle.

Positive displacement flowmeters continuously separate the liquid stream into known volumes based on the physical dimensions of the meter, and register flow by counting cycles or revolutions. Figure 1.19 shows two different flow-segmentation methods. In the nutating disk meter, as the liquid attempts to flow through the meter, the pressure drop from inlet to outlet causes the disk to wobble. The sliding vane flowmeter has retractile vanes that seal a volume of liquid between the rotor and the casing and transport it from the inlet to the outlet, where it is discharged.

Weirs and flumes are calibrated restrictions used in open channel flows and



(a)



(b)

Figure 1.19 Two flow-segmentation methods used in positive displacement flowmeters: (*a*) Nutating disk and (*b*) sliding vane.

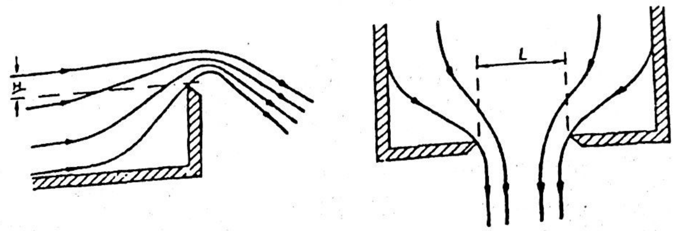


Figure 1.20 A weir is a channel restriction that raises the fluid to a height that depends on the flow rate.

in nonfilled conduits. A *weir* is a dam with a gorge in the top, built perpendicular to the flow direction. The liquid rises to a certain height, and then it flows through the gorge. This device converts part of the kinetic energy of the fluid into potential energy, and the fluid rises to a height relative to the lower point of the gorge that depends on the flow rate. If the gorge is rectangular, as in Figure 1.20, then we obtain

$$Q = kL\sqrt[3]{H^2} \tag{1.53}$$

where Q is the volumetric flow rate, H the height raised by the fluid, L is the weir width, and k is a constant. H can be measured using an upstream fluid level sensor. A *flume* is a channel restriction in area, slope, or both, based in the same principle as weirs.

Mass flow rate can be indirectly measured from volumetric flow rate and density. However, density depends on pressure and temperature, and any error in their measurement will propagate into the calculated flow rate. Thermal and Coriolis flowmeters (Section 8.2.5) yield better accuracy. There are two thermal flowmeters: hot-wire probes and heat transfer flow meters.

Hot wire probes (Figure 1.21*a*) measure the rate of heat loss to the flowing

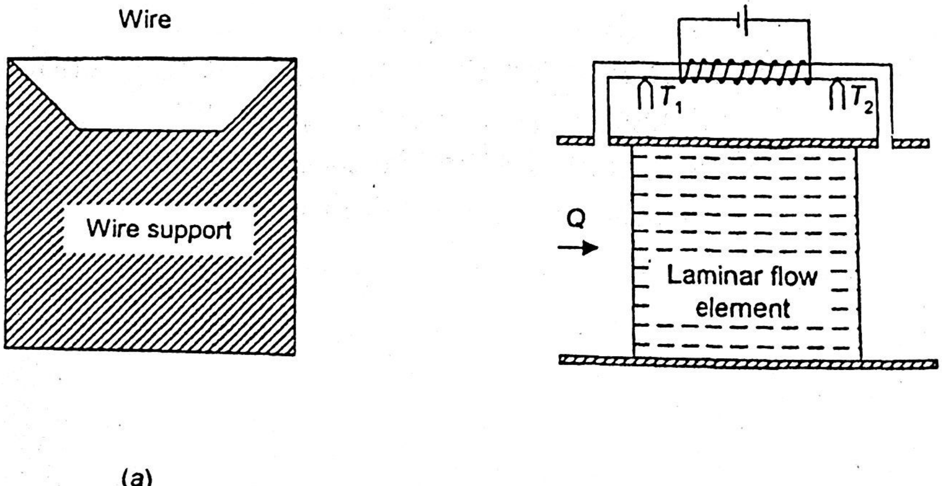


Figure 1.21 Thermal mass flowmeters convert the flow into a temperature change. In hot-wire anemometers (*a*), the rate of heat loss from a heated wire to the fluid depends on the local velocity. In heat-transfer flowmeters (*b*), the temperature rise on the downstream sensor depends on the mass flow.

fluid from a hot body—a resistive wire (Section 2.3), a thermistor (Section 2.4) or a thermopile (Section 6.1)—held perpendicular to the fluid flow. The heat flow rate from the wire to the fluid is proportional to the heat interchanging area A, to the difference in temperature between the wire and the fluid, and to the film coefficient of heat transfer h. The power dissipated by Joule effect is I^2R, and therefore, in equilibrium, we have

$$I^2R = khA(T_w - T_f) \tag{1.54}$$

where k is a unit-conversion constant. The coefficient of heat transfer depends on fluid velocity according to

$$h = c_0 + c_1\sqrt{v} \tag{1.55}$$

where c_0 and c_1 are factors that include the dependence on the dimensions of the wire and on the density, viscosity, specific heat, and thermal conductivity of the fluid. A large mass flow cools the wire to a lower temperature. If the current supply to the wire is constant, the resistance of the wire—or the generated voltage in a thermopile—indicate the mass flow. Alternatively, we can measure the current necessary to keep the wire at constant temperature.

Heat transfer flowmeters measure the rise in temperature of the fluid after a known amount of heat has been added to it. The primary sensor is a capillary tube with a wound heater and two temperature sensors symmetrically mounted upstream and downstream of the heater on the tube surface (Figure 1.21*b*). When there is no flow, both sensors have the same temperature. As flow increases, the incoming fluid removes heat from the tube and cools the upstream end while it heats the downstream end when passing through it. For low flows, the difference in temperature between sensors is proportional to the mass flow rate. Large flows remove heat even from the hottest point in the tube, and the proportionality is lost. To measure large flows, a laminar flow element in the main pipe causes a drop in pressure proportional to the volumetric flow rate—equation (1.52)—that forces through the capillary a small fraction of the flow. There are micromachined silicon flowmeters that use diffused resistors as heaters and resistor bridges and that use thermodiodes or thermocouples as temperature sensors. They consume low power, their response time is less than 3 ms, and their mass is about 10 g.

1.7.4 Level Sensors

Dipsticks are simple level sensors, but cannot easily provide an electric signal. Floats, based on Archimedes' buoyancy principle, convert liquid level to force or displacement (Figures 1.22*a* and 1.22*b*). In sealed or high-pressure containers, the position of the float can be detected magnetically. Build-up and deposits on the float surface limit performance.

The pressure of liquid or solid is proportional to level (Figure 1.22*c*),

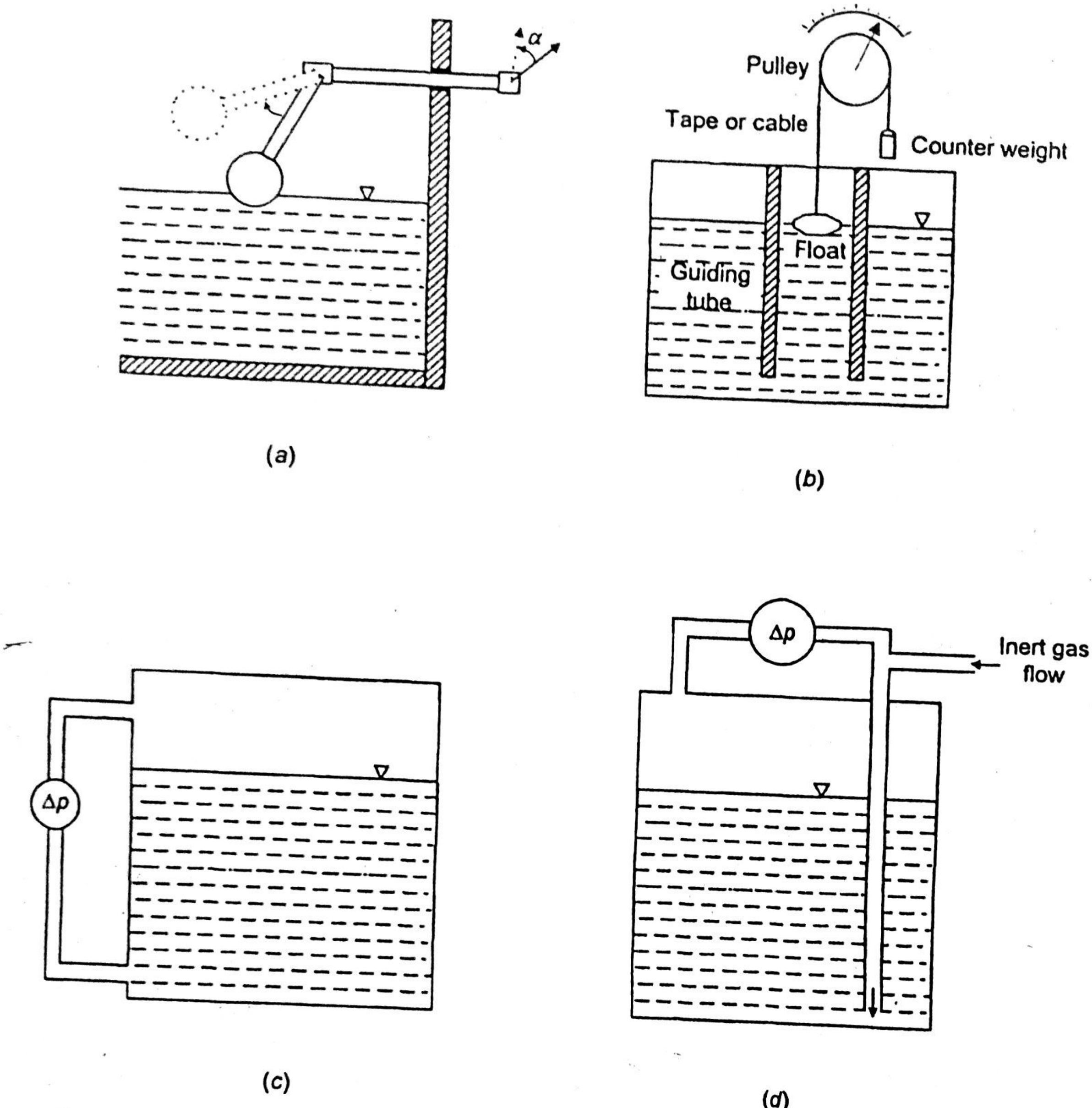


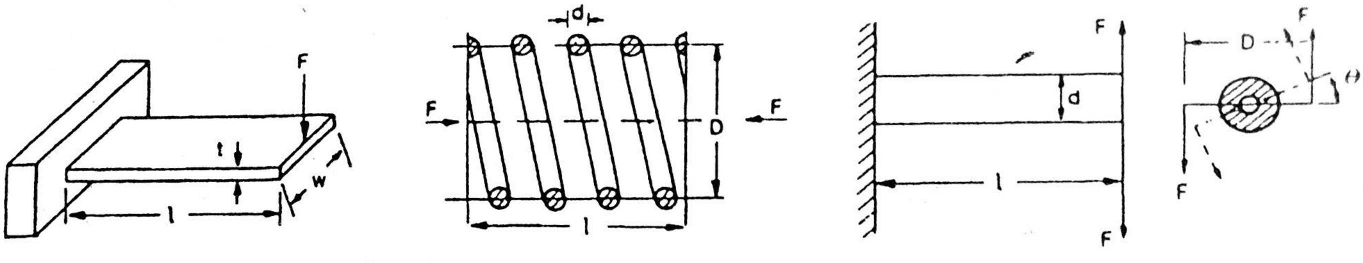
Figure 1.22 Primary level sensors. (*a*) and (*b*) Based on a float. (*c*) and (*d*) Based on differential pressure measurement.

according to

$$h = \frac{\Delta p}{\rho g} \tag{1.56}$$

where ρ is density and g is the acceleration of gravity. This method is suitable for both pressurized and open containers. Temperature interferes because it varies density.

The bubble tube in Figure 1.22*d* overcomes the need for a pressure port near the container bottom, which is a potential leak source. The dip tube has an open end close to the bottom of the tank. An inert gas flows through the dip tube and when gas bubbles escape from the open end, the gas pressure in the tube equals the hydraulic pressure from the liquid. The level can be calculated from (1.56).



(a) (b) (c)

Figure 1.23 (*a*) A cantilever, (*b*) a helical spring, and (*c*) a torsion bar deflect in response to an applied force or torque.

Each April issue of *Measurements & Control* lists manufacturers and types of level measurement and control devices.

1.7.5 Force and Torque Sensors

A method to measure force (or torque) is to compare it with a well-known force, as is done on scales. Another method measures the effect of the force on an elastic element, called a *load cell.* In electric load cells, that effect is a deformation or a displacement. In hydraulic and pneumatic load cells it is an increase in the pressure of, respectively, a liquid or a gas. Each October issue of *Measurements & Control* lists the manufacturers and types of mass/force sensors and load cells.

When a mechanical force is applied to a fixed elastic element, it strains until the strain-generated stresses balance those due to the applied force. The result is a change in the dimensions of the element that is proportional to the applied force, if the shape is appropriate.

Figure 1.23 shows three suitable arrangements. Table 1.9 lists the corre-

TABLE 1.9 Deflection x or θ and Maximal Stress s_M or τ_M for the Elastic Elements Shown in Figure 1.23

Element	Deflection	Maximal Stress
Cantilever	$x = \dfrac{4Fl^3}{Ewt^3} = \dfrac{2\sigma l^2}{3Et}$	$s_M = \dfrac{6Fl}{wt^2} = \dfrac{3Etx}{2l^2}$
Helical spring	$x = \dfrac{8FnD^3}{Gd^4} = \dfrac{\pi n D^2 \tau}{Gdk_1}$	$\tau_M = \dfrac{8k_1 DF}{\pi d^3} = \dfrac{Gdxk_1}{\pi n D^2}$
Torsion bar	$\theta = \dfrac{32FDl}{\pi d^4 G} = \dfrac{2\tau l}{dG}$	$\tau_M = \dfrac{16FD}{\pi d^3} = \dfrac{dG\theta}{2l}$

Source: From H. K. P. Neubert, *Instrument transducers*, copyright 1975. Reprinted by permission of Oxford University Press, Fair Law, NJ.

Note: All quantities are in SI units (lengths in meters, forces in newtons, angles in radians). E = longitudinal modulus of elasticity (Young's modulus), G = modulus of rigidity (torsion elasticity modulus), k_1 = stress factor (function of D/d, valued from 1.1 to 1.6), n = number of turns.

sponding equations. Neubert [14] gives additional shapes and their corresponding equations. Most load cells are underdamped second-order systems (Section 1.5.3), which limits the maximal frequency of dynamic forces that can be accurately measured to a frequency range well below the load cell's natural frequency.

1.7.6 Acceleration and Inclination Sensors

The primary sensor for acceleration is the seismic mass–spring system (Figure 1.10). The output signal is displacement, strain, or capacitance change. Acceleration is measured for structural model verification, engine vibration level measurement in aircraft, machine monitoring, and inertial measuring units (to guide ammunition to a target). It is also used in experimental modal analysis, which is the empirical characterization of structures in terms of their damping, resonant frequencies, and vibration mode shapes; the larger the structure, the lower the frequency of the first vibrating mode. Micromachined accelerometers have found their way in automotive air bags, automotive suspension systems, stabilization systems for video equipment, transportation shock recorders, and activity responsive pacemakers. Each December issue of *Measurements & Control* lists the manufacturers and types of accelerometers and vibration sensors.

Inclinometers measure the attitude of orientation with respect to a reference axis. If the reference axis is defined by gravity (vertical axis), accelerometers work as inclinometers because they sense the acceleration applied along their sensitive axes. Alternatively, the liquid bubble inclinometer works the same as the level vial used by carpenters. In tilt sensors there is a curved tube with a trapped bubble that displaces when the tube tilts (Figure 1.24*a*). Resistive (Section 2.1) or capacitive (Section 4.1) sensors can sense the bubble position. The suspended pendulum (Figure 1.24*b*) is a weight attached to a ball bearing that can rotate. If the case rotates, the mass remains vertical, so that it under-

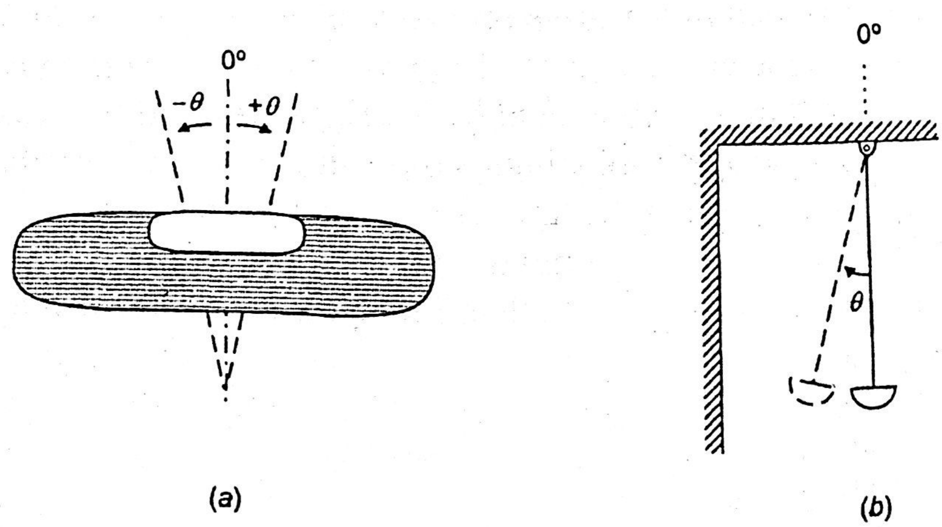


Figure 1.24 Inclination sensors. (*a*) The bubble inside a partially filled vial displaces when the vial tilts. (*b*) A mass suspended within a case rotates when the case rotates.

goes an angular displacement relative to the case, equal to the case rotation angle. Horizontal accelerations interfere with both sensors.

A compass senses the inclination with respect to a reference axis defined by a magnetic field. A potentiometer (Section 2.1), reluctive sensor (Section 4.2.1), or variable transformer (Section 4.2.4) can yield an electric signal corresponding to the rotation of the needle.

The spinning wheel of a gyroscope (Section 1.7.7) also defines a reference axis. If the frame in which the wheel rotates is fixed to a vehicle, the change of attitude of the vehicle results in a change in angle between the frame and the axis of rotation of the wheel. Each September issue of *Measurements & Control* lists the manufacturers and types of inclinometers.

1.7.7 Velocity Sensors

Linear velocity can be measured by integrating acceleration or differentiating displacement. Linear velocity can also be converted into rotational velocity by attaching a rack to the moving object and coupling it to a pinion gear that drives a rotor—as in car speedometers.

The seismic sensor in Figure 1.10 can be applied to linear velocity sensing without any link between the moving object and the reference respect to which the velocity is sensed. Integrating the mass displacement, which according to (1.31) is proportional to the input acceleration, yields the input velocity. Alternatively, if we sense the velocity of the mass relative to its housing, by manipulating (1.31) we obtain

$$\frac{\dot{X}_o(s)}{\dot{X}_i(s)} = \frac{sX_o(s)}{sX_i(s)} = \frac{s^2X_o(s)}{s^2X_i(s)} = \frac{M}{K}\frac{s^2(K/M)}{s^2 + s(B/M) + K/M} \tag{1.57}$$

Therefore, at frequencies above the natural frequency of the mass–spring system, the output of the internal velocity sensor is proportional to the input speed $\dot{x}_i$—relative to an inertial reference.

Absolute angular velocity measurement often relies on *gyroscopes* (or *gyros*). In a classic single-axis mechanical gyro, a motor-driven spinning mass (disk or wheel) is supported within a gimbal, held by bearings attached to a case (Figure 1.25*a*). In a two-axis gyro, the bearings supporting the inner gimbal are attached to an outer gimbal able to rotate with respect to the case.

A rate gyro is a single-axis gyro having an elastic restraint of the spin axis about the output axis (Figure 1.25*b*). When the gyroscope is rotated around the axis (y-axis) perpendicular to the spinning mass (x-axis), an angular momentum is developed around the z-axis, perpendicular to the x- and y-axes. That momentum is proportional to the angular speed around the y-axis and can be sensed by torque or force sensors [2].

Micromachined gyros have no rotating parts, and thus no bearings. They sense rotation from the Coriolis effect on vibrating mechanical elements [15].

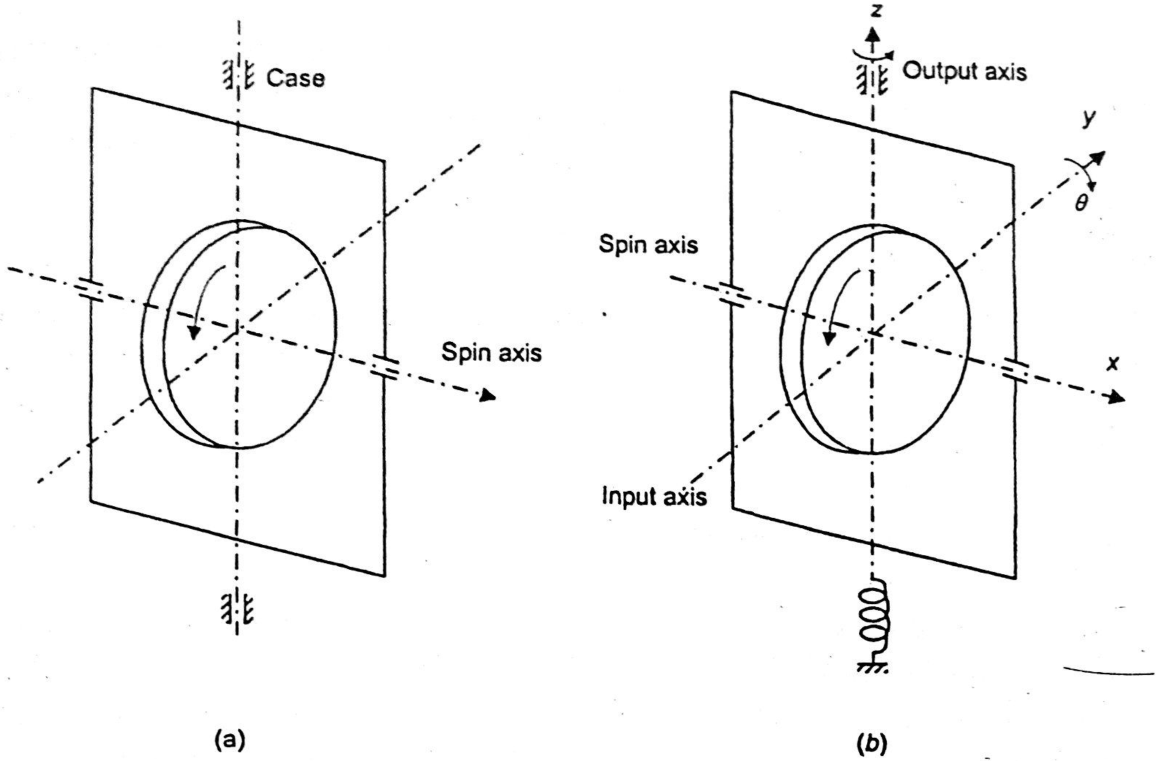


Figure 1.25 Single-axis mechanical gyroscope. (*a*) A spinning wheel defines the *x*-axis. (*b*) A rotation around the *y*-axis, perpendicular to the *x*-axis, yields a torque around the *z*-axis, perpendicular to both *x*- and *y*-axes.

The Coriolis effect is an apparent acceleration that arises in a moving element in a rotating body. Consider a traveling particle with velocity $\boldsymbol{v}$ (Figure 1.26) and an observer placed on the *x*-axis watching the particle. If the coordinate system (including the observer) rotates around the *z*-axis with angular velocity Ω, the observer thinks that the particle is moving toward the *x*-axis with acceleration

$$a_{\mathrm{Cor}} = 2\Omega \times \boldsymbol{v} \tag{1.58}$$

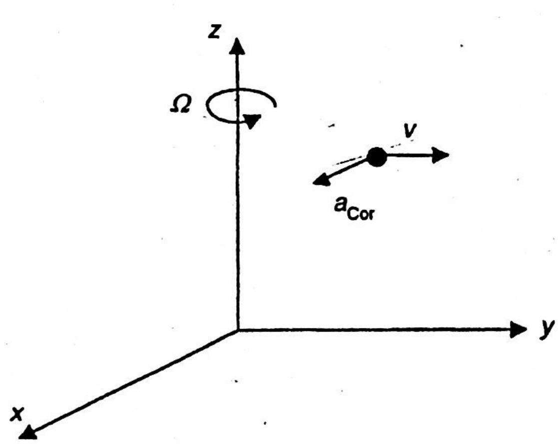


Figure 1.26 The Coriolis acceleration appears on a traveling particle when the coordinate system rotates with angular velocity $\boldsymbol{\Omega}$.

Therefore, when a mechanical element (tuning fork, disk, plate, etc.) is made to oscillate by the application of an alternating force, and this oscillating body is placed in a rotating reference frame, the Coriolis force produces a secondary oscillation perpendicular to the primary oscillation motion. The vibrating structure can be driven by electrostatic, electromagnetic, or piezoelectric force. Capacitive, piezoresistive, or piezoelectric sensors can detect the Coriolis-induced vibrations.

Fiber-optic and laser gyros do not use angular momentum but use the optical heterodyning of counterrotating optical or laser beams produced by Sagnac's effect (Section 9.4).

1.8 MATERIALS FOR SENSORS

Sensors rely on physical or chemical phenomena and materials where those phenomena appear usefully—that is, with high sensitivity, repeatability and specificity. Those phenomena may concern the material itself or its geometry, and most of them have been known for a long time. Major changes in sensors come from new materials, new fabrication techniques, or both.

Solids, liquids, and gases consist of atoms, molecules, or ions—atoms or group of atoms that have lost or gained one or more electrons. Atoms consist of a positive nucleus and electrons orbiting around it in shells. If the outer electron shell is not full, atoms try to gain extra electrons and become bonded in the process, forming molecules or agglomerates. There are four main bond types: ionic, metallic, covalent, and van der Waals [16]. Ionic bonds result from the electrostatic attraction between ions of different polarity. Ionic bonds form crystals—solids whose atoms are arranged in a long-range three-dimensional pattern in a way that reduces the overall energy and maintains electrical neutrality. Ionic crystals, such as NaCl and CsCl, have low electrical conductivity (because there are no free charges), relatively high fusion temperature, and good mechanical resistance, all resulting from the strong cohesion between ions.

The metallic bond also arises from electrostatic forces. But unlike the ionic bond, those forces are not between charges occupying a fixed position but between fixed positive charges and a cloud of electrons moving around the fixed positive ions. Mobile electrons in metals come from the outermost electron shell (valence electrons) of their atoms. Hence, metals have a regular structure (i.e., form crystals), but there is no need for a particular atom arrangement in those crystals to ensure electric neutrality. The swarming electron cloud (or *electron gas*) maintains electroneutrality. The crystal structure is then determined by the packing capability of atoms. Smaller atoms can diffuse through the lattice of higher-radius atoms, such as copper in germanium. Free electrons confer to metals their high electrical and thermal conductivity. The ubiquity of electrostatic forces along the lattice makes metals highly ductile and malleable.

Covalent bonds come from atoms sharing electrons with nearby atoms, so that they "believe" their respective outer electron shell is full. This bond may keep together atoms in a molecule (e.g., chlorine) or in a crystal [e.g., diamond (carbon), silicon, and germanium]. Shared electrons cannot move from their positions, and therefore they are not available to conduct electricity. Hence, materials with covalent bonds have low electrical conductivity.

Van der Waals bonds appear between molecules with intramolecular covalent bonds that have a small dipolar moment because of the lack of coincidence between the centers of positive and negative charge as a result of continuous electron movement. Van der Waals bonds keep together organic molecules to form crystals with low cohesion energy and whose structure depends on how well the molecules can pack together. Because of the low cohesion, materials with van der Waals bonds have low melting and boiling points.

Electrons in atoms can occupy only defined states, or energy levels, even when excited. The gap between the energy level corresponding to a nonexcited state and that corresponding to an excited state equals the amount of energy needed for one electron to jump from the base to the excited state. In a mass of atoms there are many energy levels. Close energy levels form an energy band. We distinguish three energy bands: the saturated or valence band, the conduction or excited band, and the forbidden band between them. Valence electrons cannot leave their positions. Excited electrons are nearly free to move around inside the material.

The relative separation between energy bands determines the electrical conductivity of materials, which is a useful property for sensors. Figure 1.27 shows that valence and conduction bands overlap in conductors, so that there are always free electrons and the electrical conductivity is high. Insulators have

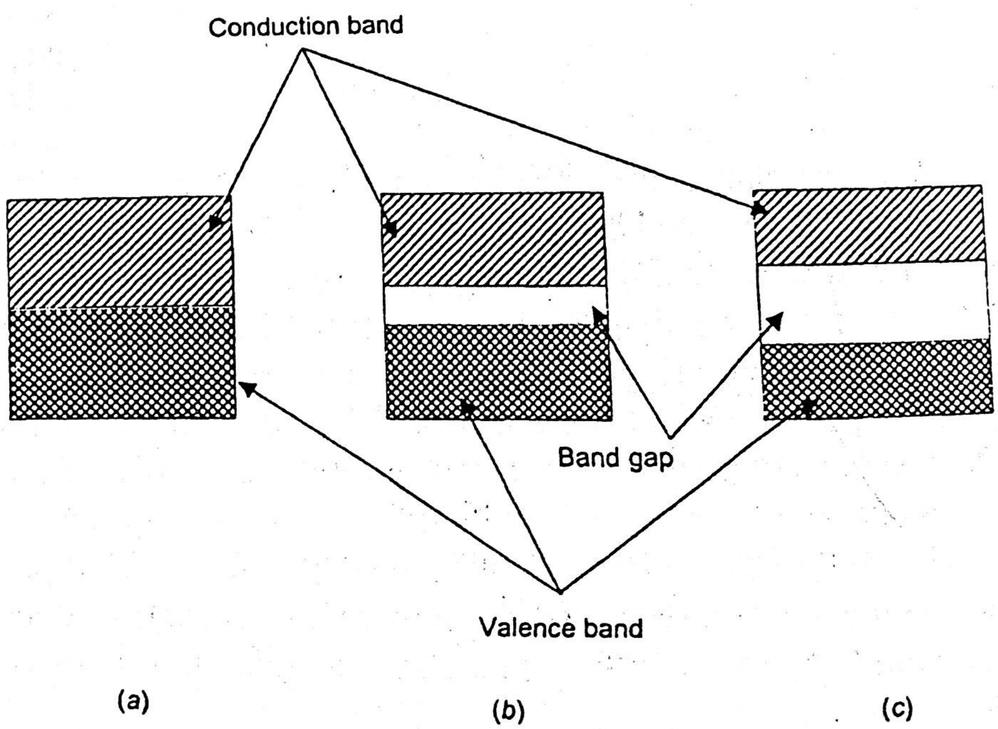


Figure 1.27 Energy bands for (*a*) a conductor, (*b*) a semiconductor, and (*c*) an electrical insulator.