Part 2 Measurement Sensors and Instruments

The second part of the book starts by discussing some of the main physical principles used in measurement sensors, and then it goes on to discuss the range of sensors and instruments that are available for measuring various physical quantities. In presenting each range of measurement devices, the aim has been to be as comprehensive as possible, so that the book can be used as a source of reference when choosing an instrument for any particular measurement situation. This method of treatment means that some quite rare and expensive instruments are included as well as very common and cheap ones, and therefore it is necessary to choose from the instrument ranges presented with care.

The first step in choosing an instrument for a particular measurement situation is to specify the static and dynamic characteristics required. These must always be defined carefully, as a high degree of accuracy, sensitivity etc. in the instrument specification inevitably involves a high cost. The characteristics specified should therefore be the minimum necessary to achieve the required level of performance in the system that the instrument is providing a measurement for. Once the static and dynamic characteristics required have been defined, the range of instruments presented in each of the following chapters can be reduced to a subset containing those instruments that satisfy the defined specifications.

The second step in choosing an instrument is to consider the working conditions in which the instrument will operate. Conditions demanding special consideration are those where the instrument will be exposed to mechanical shocks, vibration, fumes, dust or fluids. Such considerations further subdivide the subset of possible instruments already identified. If a choice still exists at this stage, then the final criterion is one of cost.

The relevant criteria in instrument choice for measuring particular physical quantities are considered further in the following chapters.

13

Sensor technologies

13.1 Capacitive and resistive sensors

Capacitive sensors consist of two parallel metal plates in which the dielectric between the plates is either air or some other medium. The capacitance C is given by $C = \varepsilon_0 \varepsilon_r A/d$, where ε_0 is the absolute permittivity, ε_r is the relative permittivity of the dielectric medium between the plates, A is the area of the plates and d is the distance between them. Capacitive devices are often used as displacement sensors, in which motion of a moveable capacitive plate relative to a fixed one changes the capacitance. Often, the measured displacement is part of instruments measuring pressure, sound or acceleration. Alternatively, fixed plate capacitors can also be used as sensors, in which the capacitance value is changed by causing the measured variable to change the dielectric constant of the material between the plates in some way. This principle is used in devices to measure moisture content, humidity values and liquid level, as discussed in later chapters.

Resistive sensors rely on the variation of the resistance of a material when the measured variable is applied to it. This principle is most commonly applied in temperature measurement using resistance thermometers or thermistors, and in displacement measurement using strain gauges or piezoresistive sensors. In addition, some moisture meters work on the resistance-variation principle.

13.2 Magnetic sensors

Magnetic sensors utilize the magnetic phenomena of inductance, reluctance and eddy currents to indicate the value of the measured quantity, which is usually some form of displacement.

Inductive sensors translate movement into a change in the mutual inductance between magnetically coupled parts. One example of this is the inductive displacement transducer shown in Figure 13.1. In this, the single winding on the central limb of an 'E'-shaped ferromagnetic body is excited with an alternating voltage. The displacement to be measured is applied to a ferromagnetic plate in close proximity to the 'E' piece. Movements of the plate alter the flux paths and hence cause a change in the current flowing in the winding. By Ohm's law, the current flowing in the winding is

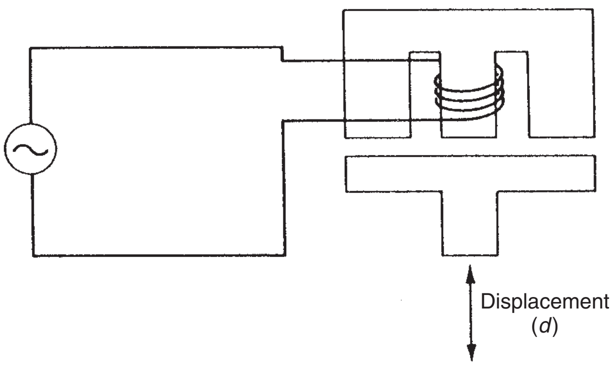


Fig. 13.1 Inductive displacement sensor.

given by $I = V/\omega L$. For fixed values of w and V, this equation becomes $I = 1/KL$, where K is a constant. The relationship between L and the displacement, d, applied to the plate is a non-linear one, and hence the output-current/displacement characteristic has to be calibrated.

The inductance principle is also used in differential transformers to measure translational and rotational displacements.

In *variable reluctance sensors*, a coil is wound on a permanent magnet rather than on an iron core as in variable inductance sensors. Such devices are commonly used to measure rotational velocities. Figure 13.2 shows a typical instrument in which a ferromagnetic gearwheel is placed next to the sensor. As the tip of each tooth on the gearwheel moves towards and away from the pick-up unit, the changing magnetic flux in the pick-up coil causes a voltage to be induced in the coil whose magnitude is proportional to the rate of change of flux. Thus, the output is a sequence of positive and negative pulses whose frequency is proportional to the rotational velocity of the gearwheel.

Eddy current sensors consist of a probe containing a coil, as shown in Figure 13.3, that is excited at a high frequency, which is typically 1 MHz. This is used to measure the displacement of the probe relative to a moving metal target. Because of the high frequency of excitation, eddy currents are induced only in the surface of the target,

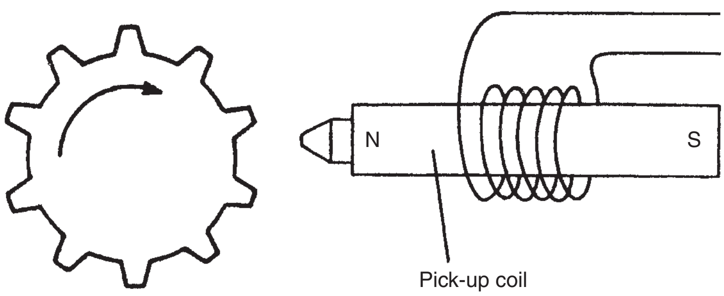


Fig. 13.2 Variable reluctance sensor.

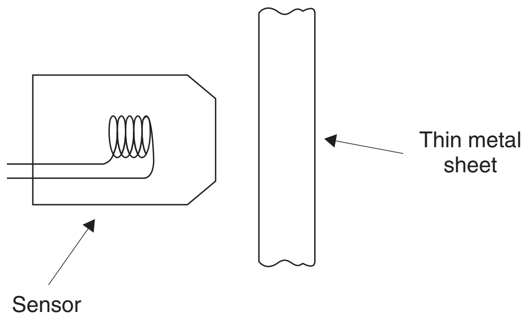


Fig. 13.3 Eddy current sensor.

and the current magnitude reduces to almost zero a short distance inside the target. This allows the sensor to work with very thin targets, such as the steel diaphragm of a pressure sensor. The eddy currents alter the inductance of the probe coil, and this change can be translated into a d.c. voltage output that is proportional to the distance between the probe and the target. Measurement resolution as high as 0.1 μm can be achieved. The sensor can also work with a non-conductive target if a piece of aluminium tape is fastened to it.

13.3 Hall-effect sensors

Basically, a Hall-effect sensor is a device that is used to measure the magnitude of a magnetic field. It consists of a conductor carrying a current that is aligned orthogonally with the magnetic field, as shown in Figure 13.4. This produces a transverse voltage difference across the device that is directly proportional to the magnetic field strength. For an excitation current I and magnetic field strength B, the output voltage is given by $V = KIB$, where K is known as the Hall constant.

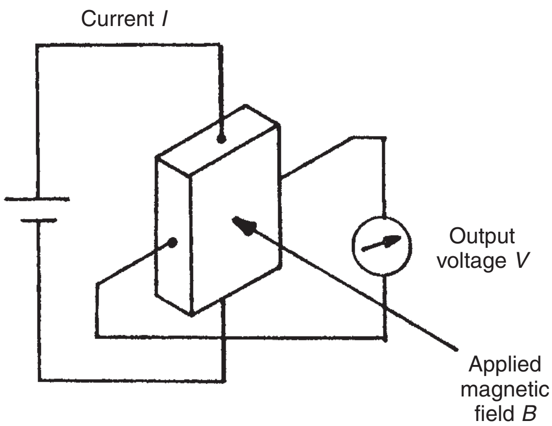


Fig. 13.4 Principles of Hall-effect sensor.

The conductor in Hall-effect sensors is usually made from a semiconductor material as opposed to a metal, because a larger voltage output is produced for a magnetic field of a given size. In one common use of the device as a proximity sensor, the magnetic field is provided by a permanent magnet that is built into the device. The magnitude of this field changes when the device becomes close to any ferrous metal object or boundary. The Hall effect is also commonly used in keyboard pushbuttons, in which a magnet is attached underneath the button. When the button is depressed, the magnet moves past a Hall-effect sensor. The induced voltage is then converted by a trigger circuit into a digital output. Such pushbutton switches can operate at high frequencies without contact bounce.

13.4 Piezoelectric transducers

Piezoelectric transducers produce an output voltage when a force is applied to them. They are frequently used as ultrasonic receivers and also as displacement transducers, particularly as part of devices measuring acceleration, force and pressure. In ultrasonic receivers, the sinusoidal amplitude variations in the ultrasound wave received are translated into sinusoidal changes in the amplitude of the force applied to the piezoelectric transducer. In a similar way, the translational movement in a displacement transducer is caused by mechanical means to apply a force to the piezoelectric transducer. Piezoelectric transducers are made from piezoelectric materials. These have an asymmetrical lattice of molecules that distorts when a mechanical force is applied to it. This distortion causes a reorientation of electric charges within the material, resulting in a relative displacement of positive and negative charges. The charge displacement induces surface charges on the material of opposite polarity between the two sides. By implanting electrodes into the surface of the material, these surface charges can be measured as an output voltage. For a rectangular block of material, the induced voltage is given by:

$$V = \frac{kFd}{A} \tag{13.1}$$

where F is the applied force in g, A is the area of the material in mm, d is the thickness of the material and k is the piezoelectric constant. The polarity of the induced voltage depends on whether the material is compressed or stretched.

The input impedance of the instrument used to measure the induced voltage must be chosen carefully. Connection of the measuring instrument provides a path for the induced charge to leak away. Hence, the input impedance of the instrument must be very high, particularly where static or slowly varying displacements are being measured.

Materials exhibiting piezoelectric behaviour include natural ones such as quartz, synthetic ones such as lithium sulphate and ferroelectric ceramics such as barium titanate. The piezoelectric constant varies widely between different materials. Typical values of k are 2.3 for quartz and 140 for barium titanate. Applying equation (13.1) for a force of 1 g applied to a crystal of area 100 mm^2 and thickness 1 mm gives an output of 23 µV for quartz and 1.4 mV for barium titanate.

Certain polymeric films such as polyvinylidine also exhibit piezoelectric properties. These have a higher voltage output than most crystals and are very useful in many applications where displacement needs to be translated into a voltage. However, they have very limited mechanical strength and are unsuitable for applications where resonance might be generated in the material.

The piezoelectric principle is invertible, and therefore distortion in a piezoelectric material can be caused by applying a voltage to it. This is commonly used in ultrasonic transmitters, where the application of a sinusoidal voltage at a frequency in the ultrasound range causes a sinusoidal variation in the thickness of the material and results in a sound wave being emitted at the chosen frequency. This is considered further in the section below on ultrasonic transducers.

13.5 Strain gauges

Strain gauges are devices that experience a change in resistance when they are stretched or strained. They are able to detect very small displacements, usually in the range 0–50 μm, and are typically used as part of other transducers, for example diaphragm pressure sensors that convert pressure changes into small displacements of the diaphragm. Measurement inaccuracies as low as ±0.15% of full-scale reading are achievable and the quoted life expectancy is usually three million reversals. Strain gauges are manufactured to various nominal values of resistance, of which 120 Ω, 350 Ω and 1000 Ω are very common. The typical maximum change of resistance in a 120 Ω device would be 5 Ω at maximum deflection.

The traditional type of strain gauge consists of a length of metal resistance wire formed into a zigzag pattern and mounted onto a flexible backing sheet, as shown in Figure 13.5(a). The wire is nominally of circular cross-section. As strain is applied to the gauge, the shape of the cross-section of the resistance wire distorts, changing the cross-sectional area. As the resistance of the wire per unit length is inversely proportional to the cross-sectional area, there is a consequential change in resistance. The input–output relationship of a strain gauge is expressed by the *gauge factor*, which is defined as the change in resistance (R) for a given value of strain (S), i.e.

$$\text{gauge factor} = \delta R/\delta S$$

In recent years, wire-type gauges have largely been replaced, either by metal-foil types as shown in Figure 13.5(b), or by semiconductor types. Metal-foil types are very

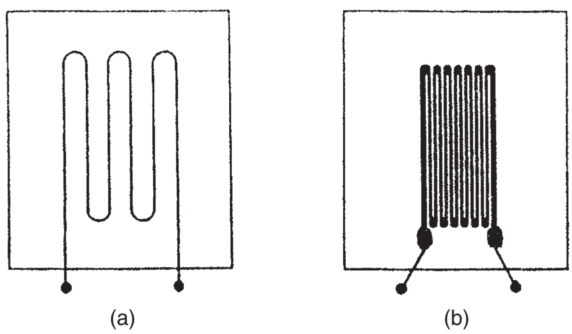


Fig. 13.5 Strain gauges: (a) wire type; (b) foil type.

similar to metal-wire types except the active element consists of a piece of metal foil cut into a zigzag pattern. Cutting a foil into the required shape is much easier than forming a piece of resistance wire into the required shape, and this makes the devices cheaper to manufacture. A popular material in metal strain gauge manufacture is a copper–nickel–manganese alloy, which is known by the trade name of ‘Advance’. Semiconductor types have piezoresistive elements, which are considered in greater detail in the next section. Compared with metal gauges, semiconductor types have a much superior gauge factor (up to 100 times better) but they are more expensive. Also, whilst metal gauges have an almost zero temperature coefficient, semiconductor types have a relatively high temperature coefficient.

In use, strain gauges are bonded to the object whose displacement is to be measured. The process of bonding presents a certain amount of difficulty, particularly for semiconductor types. The resistance of the gauge is usually measured by a d.c. bridge circuit and the displacement is inferred from the bridge output measured. The maximum current that can be allowed to flow in a strain gauge is in the region of 5 to 50 mA depending on the type. Thus, the maximum voltage that can be applied is limited and consequently, as the resistance change in a strain gauge is typically small, the bridge output voltage is also small and amplification has to be carried out. This adds to the cost of using strain gauges.

13.6 Piezoresistive sensors

A piezoresistive sensor is made from semiconductor material in which a p-type region has been diffused into an n-type base. The resistance of this varies greatly when the sensor is compressed or stretched. This is frequently used as a strain gauge, where it produces a significantly higher gauge factor than that given by metal wire or foil gauges. Also, measurement uncertainty can be reduced to $\pm 0.1\%$. It is also used in semiconductor-diaphragm pressure sensors and in semiconductor accelerometers.

It should also be mentioned that the term piezoresistive sensor is sometimes used to describe all types of strain gauge, including metal types. However, this is incorrect since only about 10% of the output from a metal strain gauge is generated by piezoresistive effects, with the remainder arising out of the dimensional cross-section change in the wire or foil. Proper piezoelectric strain gauges, which are alternatively known as *semiconductor strain gauges*, produce most (about 90%) of their output through piezoresistive effects, and only a small proportion of the output is due to dimensional changes in the sensor.

13.7 Optical sensors (air path)

Optical sensors are based on the modulation of light travelling between a light source and a light detector, as shown in Figure 13.6. The transmitted light can travel along either an air path or a fibre-optic cable. Either form of transmission gives immunity to electromagnetically induced noise, and also provides greater safety than electrical sensors when used in hazardous environments.

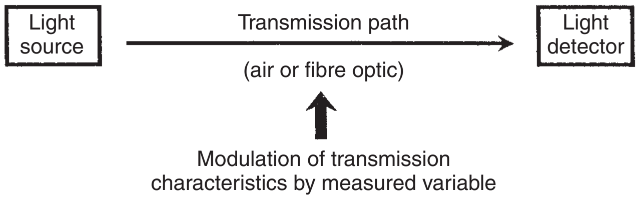


Fig. 13.6 Operating principles of optical sensors.

Light sources suitable for transmission across an air path include tungsten-filament lamps, laser diodes and light-emitting diodes (LEDs). However, as the light from tungsten lamps is usually in the visible part of the light frequency spectrum, it is prone to interference from the sun and other sources. Hence, infrared LEDs or infrared laser diodes are usually preferred. These emit light in a narrow frequency band in the infrared region and are not affected by sunlight.

The main forms of light detector used with optical systems are photocells (cadmium sulphide or cadmium selenide being the most common type of photocell), phototransistors and photodiodes. These are all photoconductive devices, whose resistance is reduced according to the intensity of light to which they are exposed. Photocells and phototransistors are particularly sensitive in the infrared region, and so are ideal partners for infrared LED and laser diode sources.

Air-path optical sensors are commonly used to measure proximity, translational motion, rotational motion and gas concentration. These uses are discussed in more detail in later chapters.

13.8 Optical sensors (fibre-optic)

As an alternative to using air as the transmission medium, optical sensors can use fibre-optic cable instead to transmit light between a source and a detector. In such sensors, the variable being measured causes some measurable change in the characteristics of the light transmitted by the cable. However, the problems and solutions that were described in Chapter 8 for fibre-optic signal transmission, in ensuring that the proportion of light entering the cable is maximized, apply equally when optical fibres are used as sensors.

The basis of operation of fibre-optic sensors is the translation of the physical quantity measured into a change in one or more parameters of a light beam. The light parameters that can be modulated are one or more of the following:

- intensity
- phase
- polarization
- wavelength
- transmission time.

Fibre-optic sensors usually incorporate either glass/plastic cables or all plastic cables. All glass types are rarely used because of their fragility. Plastic cables have particular advantages for sensor applications because they are cheap and have a relatively large diameter of 0.5–1.0 mm, making connection to the transmitter and receiver easy. However, plastic cables should not be used in certain hostile environments where they may be severely damaged. The cost of the fibre-optic cable itself is insignificant for sensing applications, as the total cost of the sensor is dominated by the cost of the transmitter and receiver.

Fibre-optic sensors characteristically enjoy long life. For example, the life expectancy of reflective fibre-optic switches is quoted at ten million operations. Their accuracy is also good, with, for instance, $\pm 1\%$ of full-scale reading being quoted as a typical inaccuracy level for a fibre-optic pressure sensor. Further advantages are their simplicity, low cost, small size, high reliability and capability of working in many kinds of hostile environment.

Two major classes of fibre-optic sensor exist, intrinsic sensors and extrinsic sensors. In *intrinsic sensors*, the fibre-optic cable itself is the sensor, whereas in *extrinsic sensors*, the fibre-optic cable is only used to guide light to/from a conventional sensor.

13.8.1 Intrinsic sensors

Intrinsic sensors can modulate either the intensity, phase, polarization, wavelength or transit time of light. Sensors that modulate light intensity tend to use mainly multimode fibres, but only monomode cables are used to modulate other light parameters. A particularly useful feature of intrinsic fibre-optic sensors is that they can, if required, provide distributed sensing over distances of up to 1 metre.

Light intensity is the simplest parameter to manipulate in intrinsic sensors because only a simple source and detector are required. The various forms of switches shown in Figure 13.7 are perhaps the simplest form of these, as the light path is simply blocked and unblocked as the switch changes state.

Modulation of the intensity of transmitted light takes place in various simple forms of proximity, displacement, pressure, pH and smoke sensors. Some of these are sketched in Figure 13.8. In proximity and displacement sensors (the latter are often given the special name *fotonic sensors*), the amount of reflected light varies with the distance between the fibre ends and a boundary. In pressure sensors, the refractive index of the fibre, and hence the intensity of light transmitted, varies according to the mechanical deformation of the fibres caused by pressure. In the pH probe, the amount of light reflected back into the fibres depends on the pH-dependent colour of the chemical indicator in the solution around the probe tip. Finally, in a form of smoke detector, two fibre-optic cables placed either side of a space detect any reduction in the intensity of light transmission between them caused by the presence of smoke.

A simple form of accelerometer can be made by placing a mass subject to the acceleration on a multimode fibre. The force exerted by the mass on the fibre causes a change in the intensity of light transmitted, hence allowing the acceleration to be determined. The typical inaccuracy quoted for this device is ± 0.02 g in the measurement range ± 5 g and $\pm 2\%$ in the measurement range up to 100 g.

A similar principle is used in probes that measure the internal diameter of tubes. The probe consists of eight strain-gauged cantilever beams that track changes in diameter, giving a measurement resolution of 20 μm.

A slightly more complicated method of effecting light intensity modulation is the variable shutter sensor shown in Figure 13.9. This consists of two fixed fibres with

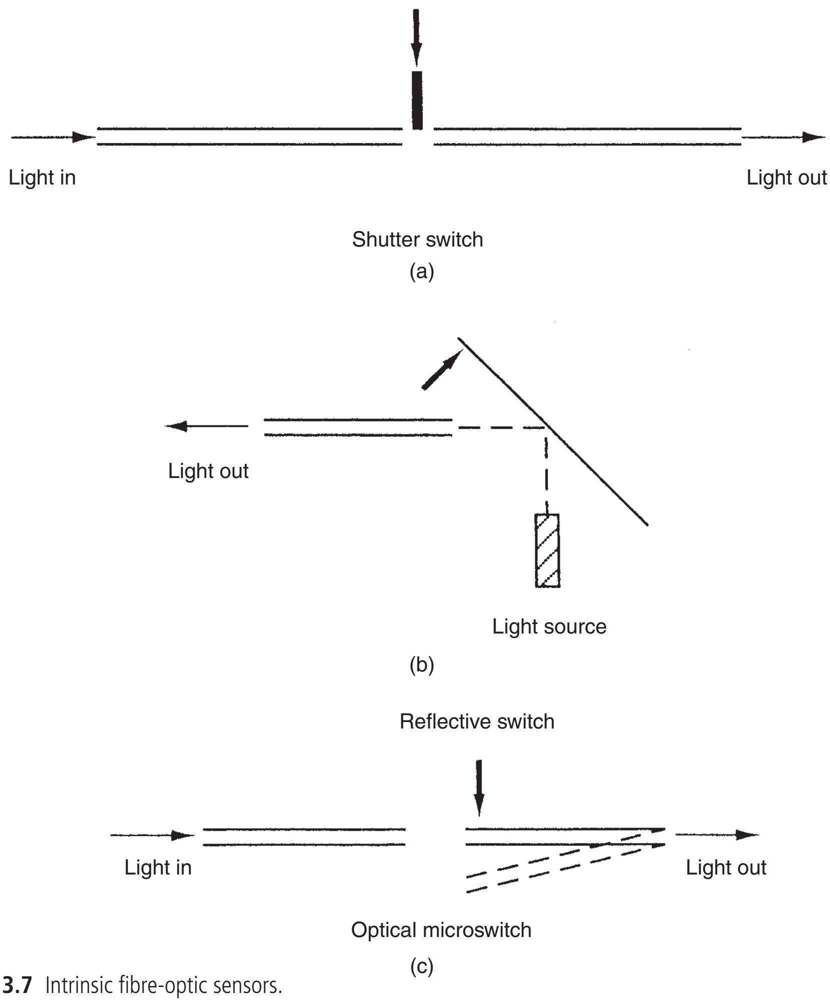


Fig. 13.7 Intrinsic fibre-optic sensors.

two collimating lenses and a variable shutter between them. Movement of the shutter changes the intensity of light transmitted between the fibres. This is used to measure the displacement of various devices such as Bourdon tubes, diaphragms and bimetallic thermometers.

Yet another type of intrinsic sensor uses cable where the core and cladding have similar refractive indices but different temperature coefficients. This is used as a temperature sensor. Temperature rises cause the refractive indices to become even closer together and losses from the core to increase, thus reducing the quantity of light transmitted.

Refractive index variation is also used in a form of intrinsic sensor used for cryogenic leak detection. The fibre used for this has a cladding whose refractive index becomes greater than that of the core when it is cooled to cryogenic temperatures. The fibre-optic cable is laid in the location where cryogenic leaks might occur. If any leaks do occur, light travelling in the core is transferred to the cladding, where it is attenuated.

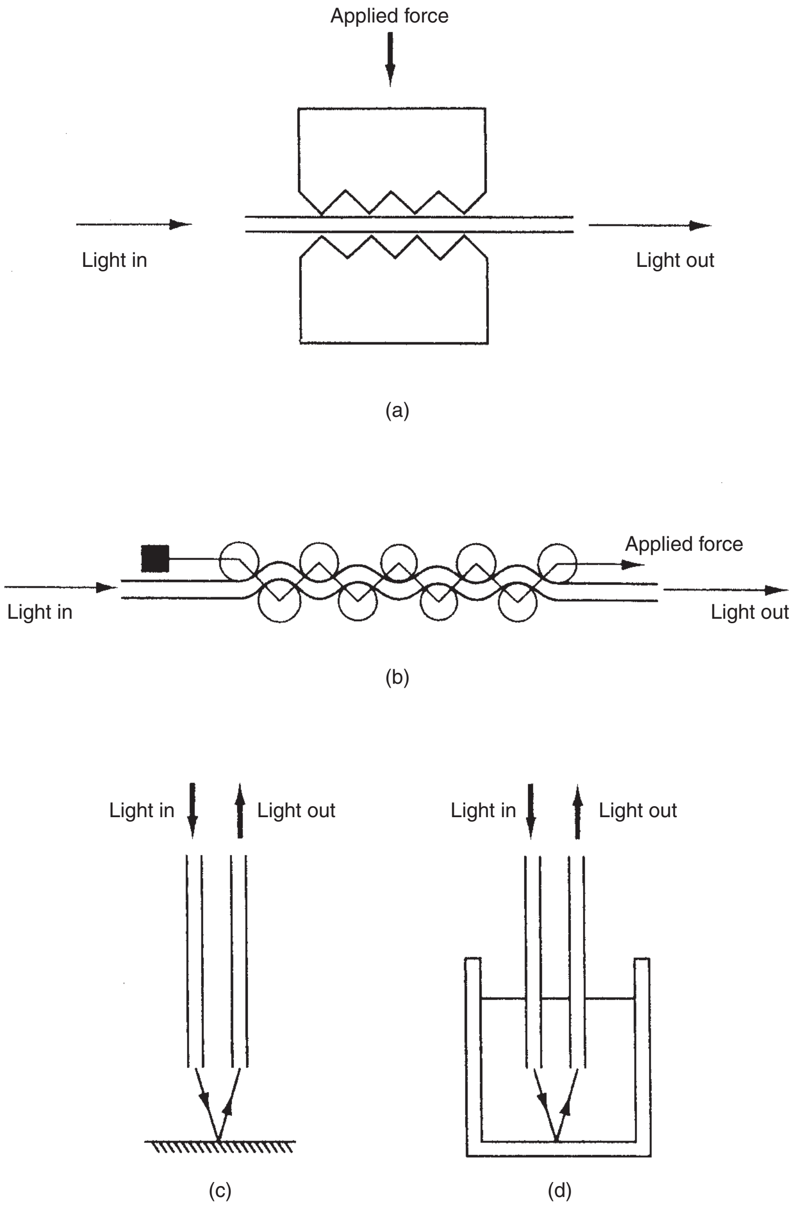


Fig. 13.8 Intensity modulating sensor: (a) simple pressure sensor; (b) roller-chain pressure sensor (microbend sensor); (c) proximity sensor; (d) pH sensor.

Cryogenic leakage is thus indicated by monitoring the light transmission characteristics of the fibre.

A further use of refractive index variation is found in devices that detect oil in water. These use a special form of cable where the cladding used is sensitive to oil. Any oil present diffuses into the cladding and changes the refractive index, thus increasing

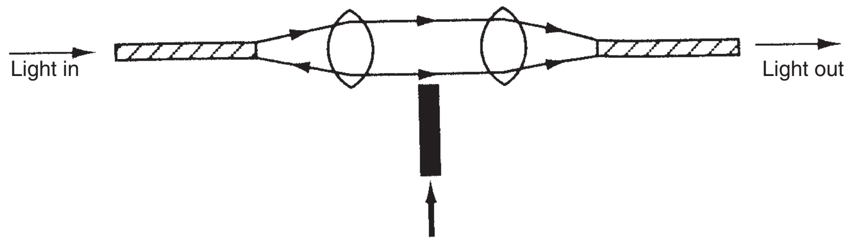


Fig. 13.9 Variable-shutter sensor.

light losses from the core. Unclad fibres are used in a similar way. In these, any oil present settles on the core and allows light to escape.

The *cross-talk sensor* measures several different variables by modulating the intensity of light transmitted. It consists of two parallel fibres that are close together and where one or more short lengths of adjacent cladding are removed from the fibres. When immersed in a transparent liquid, there are three different effects that each cause a variation in the intensity of light transmitted. Thus, the sensor can perform three separate functions. Firstly, it can measure temperature according to the temperature-induced variation in the refractive index of the liquid. Secondly, it can act as a level detector, as the transmission characteristics between the fibres change according to the depth of the liquid. Thirdly, it can measure the refractive index of the liquid itself when used under controlled temperature conditions.

The refractive index of a liquid can be measured in an alternative way by using an arrangement where light travels across the liquid between two cable ends that are fairly close together. The angle of the cone of light emitted from the source cable, and hence the amount of light transmitted into the detector, is dependent on the refractive index of the liquid.

The use of materials where the fluorescence varies according to the value of the measurand can also be used as part of intensity modulating intrinsic sensors. Fluorescence-modulating sensors can give very high sensitivity and are potentially very attractive in biomedical applications where requirements exist to measure very small quantities such as low oxygen and carbon monoxide concentrations, low blood pressure levels etc. Similarly, low concentrations of hormones, steroids etc. may be measured (Grattan, 1989).

Further examples of intrinsic fibre-optic sensors that modulate light intensity are described later in Chapter 17 (level measurement) and Chapter 19 (measuring small displacements).

As mentioned previously, light phase, polarization, wavelength and transit time can be modulated as well as intensity in intrinsic sensors. Monomode cables are used almost exclusively in these types of intrinsic sensor.

Phase modulation normally requires a coherent (laser) light source. It can provide very high sensitivity in displacement measurement but cross-sensitivity to temperature and strain degrades its performance. Additional problems are maintaining frequency stability of the light source and manufacturing difficulties in coupling the light source to the fibre. Various versions of this class of instrument exist to measure temperature, pressure, strain, magnetic fields and electric fields. Field-generated quantities such as electric current and voltage can also be measured. In each case, the measurand causes a phase change between a measuring and a reference light beam that is detected by an interferometer. Fuller details can be found in Harmer (1982) and Medlock (1986).

The principle of phase modulation has also been used in the fibre-optic accelerometer (where a mass subject to acceleration rests on a fibre), and in fibre strain gauges (where two fibres are fixed on the upper and lower surfaces of a bar under strain). These are discussed in more detail in Harmer (1982). The fibre-optic gyroscope described in Chapter 20 is a further example of a phase-modulating device.

Devices using polarization modulation require special forms of fibre that maintain polarization. Polarization changes can be effected by electrical fields, magnetic fields, temperature changes and mechanical strain. Each of these parameters can therefore be measured by polarization modulation.

Various devices that modulate the wavelength of light are used for special purposes, as described in Medlock (1986). However, the only common wavelength-modulating fibre-optic device is the form of laser Doppler flowmeter that uses fibre-optic cables, as described in Chapter 16.

Fibre-optic devices using modulation of the transit time of light are uncommon because of the speed of light. Measurement of the transit time for light to travel from a source, be reflected off an object, and travel back to a detector, is only viable for extremely large distances. However, a few special arrangements have evolved which use transit-time modulation, as described in Medlock (1986). These include instruments such as the optical resonator, which can measure both mechanical strain and temperature. Temperature-dependent wavelength variation also occurs in semiconductor crystal beads (e.g. aluminium gallium arsenide). This is bonded to the end of a fibre-optic cable and excited from an LED at the other end of the cable. Light from the LED is reflected back along the cable by the bead at a different wavelength. Measurement of the wavelength change allows temperatures in the range up to 200°C to be measured accurately. A particular advantage of this sensor is its small size, typically 0.5 mm diameter at the sensing tip. Finally, to complete the catalogue of transit-time devices, the frequency modulation in a piezoelectric quartz crystal used for gas sensing can also be regarded as a form of time domain modulation.

13.8.2 Extrinsic sensors

Extrinsic fibre-optic sensors use a fibre-optic cable, normally a multimode one, to transmit modulated light from a conventional sensor such as a resistance thermometer. A major feature of extrinsic sensors, which makes them so useful in such a large number of applications, is their ability to reach places that are otherwise inaccessible. One example of this is the insertion of fibre-optic cables into the jet engines of aircraft to measure temperature by transmitting radiation into a radiation pyrometer located remotely from the engine. Fibre-optic cable can be used in the same way to measure the internal temperature of electrical transformers, where the extreme electromagnetic fields present make other measurement techniques impossible.

An important advantage of extrinsic fibre-optic sensors is the excellent protection against noise corruption that they give to measurement signals. Unfortunately, the output of many sensors is not in a form that can be transmitted by a fibre-optic cable, and conversion into a suitable form must therefore take place prior to transmission. For example, a platinum resistance thermometer (PRT) translates temperature changes into resistance changes. The PRT therefore needs electronic circuitry to convert the resistance changes into voltage signals and thence into a modulated light form, and this in turn means that the device needs a power supply. This complicates the measurement process and means that low-voltage power cables must be routed with the fibre-optic cable to the transducer. One particular adverse effect of this is that the advantage of intrinsic safety is lost. One solution to this problem (Grattan, 1989) is to use a power source in the form of electronically generated pulses driven by a lithium battery. Alternatively (Johnson, 1994), power can be generated by transmitting light down the fibre-optic cable to a photocell. Both of these solutions provide intrinsically safe operation.

Piezoelectric sensors lend themselves particularly to use in extrinsic sensors because the modulated frequency of a quartz crystal can be readily transmitted into a fibre-optic cable by fitting electrodes to the crystal that are connected to a low power LED. Resonance of the crystal can be created either by electrical means or by optical means using the photothermal effect. The photothermal effect describes the principle where, if light is pulsed at the required oscillation frequency and directed at a quartz crystal, the localized heating and thermal stress caused in the crystal results in it oscillating at the pulse frequency. Piezoelectric extrinsic sensors can be used as part of various pressure, force and displacement sensors. At the other end of the cable, a phase-locked loop is typically used to measure the transmitted frequency.

Fibre-optic cables are also now commonly included in digital encoders, where the use of fibres to transmit light to and from the discs allows the light source and detectors to be located remotely. This allows the devices to be smaller, which is a great advantage in many applications where space is at a premium.

13.8.3 Distributed sensors

Current research is looking at ways of distributing a number of discrete sensors measuring different variables along a fibre-optic cable. Alternatively, sensors of the same type, which are located at various points along a cable, are being investigated as a means of providing distributed sensing of a single measured variable. For example, the use of a 2 km long cable to measure the temperature distribution along its entire length has been demonstrated, measuring temperature at 400 separate points to a resolution of 1°C.

13.9 Ultrasonic transducers

Ultrasonic devices are used in many fields of measurement, particularly for measuring fluid flow rates, liquid levels and translational displacements. Details of such applications can be found in later chapters. Uses of ultrasound in imaging systems will also be briefly described at the end of this section, although the coverage in this case will be brief since such applications are rather outside the scope of this text.

Ultrasound is a band of frequencies in the range above 20 kHz, that is, above the sonic range that humans can usually hear. Measurement devices that use ultrasound consist of one device that transmits an ultrasound wave and another device that receives the wave. Changes in the measured variable are determined either by measuring the change in time taken for the ultrasound wave to travel between the transmitter and receiver, or, alternatively, by measuring the change in phase or frequency of the transmitted wave.

The most common form of ultrasonic element is a piezoelectric crystal contained in a casing, as illustrated in Figure 13.10. Such elements can operate interchangeably as either a transmitter or receiver. These are available with operating frequencies that vary between 20 kHz and 15 MHz. The principles of operation, by which an alternating voltage generates an ultrasonic wave and vice versa, have already been covered in the section above on piezoelectric transducers.

For completeness, mention should also be made of capacitive ultrasonic elements. These consist of a thin, dielectric membrane between two conducting layers. The membrane is stretched across a backplate and a bias voltage is applied. When a varying voltage is applied to the element, it behaves as an ultrasonic transmitter and an ultrasound wave is produced. The system also works in the reverse direction as an ultrasonic receiver. Elements with resonant frequencies in the range between 30 kHz and 3 MHz can be obtained (Rafiq, 1991).

13.9.1 Transmission speed

The transmission speed of ultrasound varies according to the medium through which it travels. Transmission speeds for some common media are given in Table 13.1.

When transmitted through air, the speed of ultrasound is affected by environmental factors such as temperature, humidity and air turbulence. Of these, temperature has the largest effect. The velocity of sound through air varies with temperature according to:

$$V = 331.6 + 0.6T \text{ m/s} \tag{13.2}$$

where T is the temperature in °C. Thus, even for a relatively small temperature change of 20 degrees from 0°C to 20°C, the velocity changes from 331.6 m/s to 343.6 m/s.

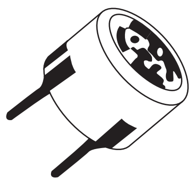


Fig. 13.10 Ultrasonic sensor.

Table 13.1 Transmission speed of ultrasound through different media

Medium	*Velocity* (m/s)
Air	331.6
Water	1440
Wood (pine)	3320
Iron	5130
Rock (granite)	6000

Humidity changes have a much smaller effect. If the relative humidity increases by 20%, the corresponding increase in the transmission velocity of ultrasound is 0.07% (corresponding to an increase from 331.6 m/s to 331.8 m/s at 0°C).

Changes in air pressure itself have negligible effect on the velocity of ultrasound. Similarly, air turbulence normally has no effect (though note that air turbulence may deflect ultrasound waves away from their original direction of travel). However, if turbulence involves currents of air at different temperatures, then random changes in ultrasound velocity occur according to equation (13.2).

13.9.2 Direction of travel of ultrasound waves

Air currents can alter the direction of travel of ultrasound waves. An air current moving with a velocity of 10 km/h has been shown experimentally to deflect an ultrasound wave by 8 mm over a distance of 1 m.

13.9.3 Directionality of ultrasound waves

Although it has perhaps been implied above that ultrasound waves travel in a narrow line away from the transmitter, this is not in fact what happens in practice. The ultrasound element actually emits a spherical wave of energy whose magnitude in any direction is a function of the angle made with respect to the direction that is normal to the face of the ultrasonic element. The peak emission always occurs along a line that is normal to the transmitting face of the ultrasonic element, and this is loosely referred to as the 'direction of travel' in the earlier paragraphs. At any angle other than the 'normal' one, the magnitude of transmitted energy is less than the peak value. Figure 13.11 shows the characteristics of the emission for a range of ultrasonic elements. This is shown in terms of the attenuation of the transmission magnitude (measured in dB) as the angle with respect to the 'normal' direction increases. For many purposes, it

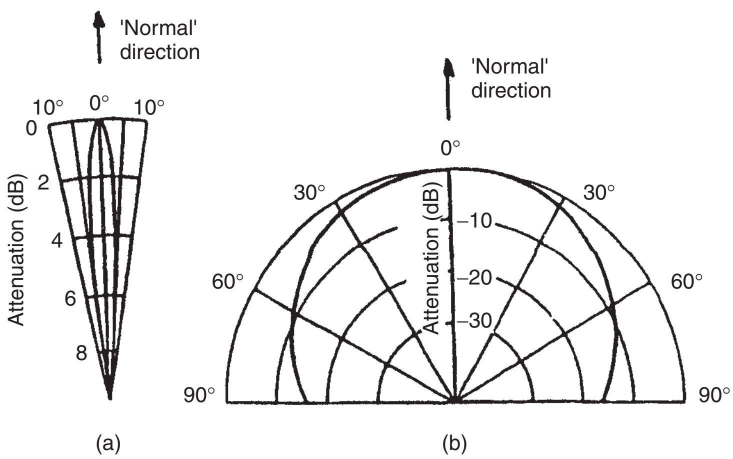


Fig. 13.11 Ultrasonic emission characteristics.

is useful to treat the transmission as a conical volume of energy, with the edges of the cone defined as the transmission angle where the amplitude of the energy in the transmission is −6 dB compared with the peak value (i.e. where the amplitude of the energy is half that in the normal direction). Using this definition, a 40 kHz ultrasonic element has a transmission cone of ±50° and a 400 kHz element has a transmission cone of ±3°.

13.9.4 Relationship between wavelength, frequency and directionality of ultrasound waves

The frequency and wavelength of ultrasound waves are related according to:

$$\lambda = v/f \tag{13.3}$$

where λ is the wavelength, v is the velocity and f is the frequency of the ultrasound waves.

This shows that the relationship between λ and f depends on the velocity of the ultrasound and hence varies according to the nature and temperature of the medium through which it travels. Table 13.2 compares the nominal frequencies, wavelengths and transmission cones (−6 dB limits) for three different types of ultrasonic element.

It is clear from Table 13.2 that the directionality (cone angle of transmission) reduces as the nominal frequency of the ultrasound transmitter increases. However, the cone angle also depends on factors other than the nominal frequency, particularly on the shape of the transmitting horn in the element, and different models of ultrasonic element with the same nominal frequency can have substantially different cone angles.

13.9.5 Attenuation of ultrasound waves

Ultrasound waves suffer attenuation in the amplitude of the transmitted energy according to the distance travelled. The amount of attenuation also depends on the nominal frequency of the ultrasound and the adsorption characteristics of the medium through which it travels. The amount of adsorption depends not only on the type of transmission medium but also on the level of humidity and dust in the medium.

The amplitude X_d of the ultrasound wave at a distance d from the emission point can be expressed as:

$$\frac{X_d}{X_0} = \frac{\sqrt{e^{-\alpha d}}}{fd} \tag{13.4}$$

where X_0 is the magnitude of the energy at the point of emission, f is the nominal frequency of the ultrasound and α is the attenuation constant that depends on the

Table 13.2 Comparison of frequency, wavelength and cone angle for various ultrasonic transmitters

Nominal frequency (kHz)	23	40	400
Wavelength (in air at 0°C)	14.4	8.3	0.83
Cone angle of transmission (−6 dB limits)	±80°	±50°	±3°

ultrasound frequency, the medium that the ultrasound travels through and any pollutants in the medium such as dust or water particles.

13.9.6 Ultrasound as a range sensor

The basic principles of an ultrasonic range sensor are to measure the time between transmission of a burst of ultrasonic energy from an ultrasonic transmitter and receipt of that energy by an ultrasonic receiver. Then, the distance d can be calculated from:

$$d = vt \tag{13.5}$$

where v is the ultrasound velocity and t is the measured energy transit time. An obvious difficulty in applying this equation is the variability of v with temperature according to equation (13.2). One solution to this problem is to include an extra ultrasonic transmitter/receiver pair in the measurement system in which the two elements are positioned a known distance apart. Measurement of the transmission time of energy between this fixed pair provides the necessary measurement of velocity and hence compensation for any environmental temperature changes.

The degree of directionality in the ultrasonic elements used for range measurement is unimportant as long as the receiver and transmitter are positioned carefully so as to face each other exactly (i.e. such that the 'normal' lines to their faces are coincident). Thus, directionality imposes no restriction on the type of element suitable for range measurement. However, element choice is restricted by the attenuation characteristics of different types of element, and relatively low-frequency elements have to be used for the measurement of large ranges.

Measurement resolution and accuracy

The best measurement resolution that can be obtained with an ultrasonic ranging system is equal to the wavelength of the transmitted wave. As wavelength is inversely proportional to frequency, high-frequency ultrasonic elements would seem to be preferable. For example, whilst the wavelength and hence resolution for a 40 kHz element is 8.6 mm at room temperature (20°C), it is only 0.86 mm for a 400 kHz element. However, choice of element also depends on the required range of measurement. The range of higher-frequency elements is much reduced compared with low-frequency ones due to the greater attenuation of the ultrasound wave as it travels away from the transmitter. Hence, choice of element frequency has to be a compromise between measurement resolution and range.

The best measurement accuracy obtainable is equal to the measurement resolution value, but this is only achieved if the electronic counter used to measure the transmission time starts and stops at exactly the same point in the ultrasound cycle (usually the point in the cycle corresponding to peak amplitude is used). However, the sensitivity of the ultrasonic receiver also affects measurement accuracy. The amplitude of the ultrasound wave that is generated in the transmitter ramps up to full amplitude in the manner shown in Figure 13.12. The receiver has to be sensitive enough to detect the peak of the first cycle, which can usually be arranged. However, if the range of measurement is large, attenuation of the ultrasound wave may cause the amplitude of the first cycle to become less than the threshold level that the receiver is set to detect.

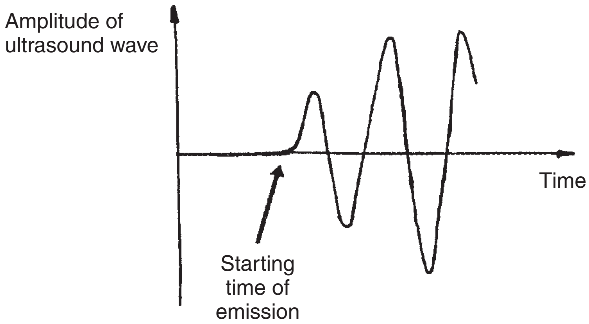


Fig. 13.12 Ramp-up of ultrasonic wave after emission.

In this case, only the second cycle will be detected and there will be an additional measurement error equal to one wavelength.

13.9.7 Use of ultrasound in tracking 3D object motion

An arrangement of the form shown in Figure 13.13 can be used to provide measurements of the position of an object moving in 3D space. In this, an ultrasonic transmitter mounted on the moving object (T) transmits bursts of energy to three receivers A, B, C located at the origin (A) and at distances q (to B) and p (to C) along the axes of an xyz co-ordinate system. If the transit times from T to A, B and C are measured, the distances a, b and c from T to the receivers can be calculated from equation (13.5). The position of T in spatial (xyz) co-ordinates can then be calculated by triangulation by solving the following set of equations:

$$x = \frac{a^2 + q^2 - b^2}{2q}; \quad y = \frac{a^2 + p^2 - c^2}{2p}; \quad z = \sqrt{a^2 - x^2 - y^2}$$

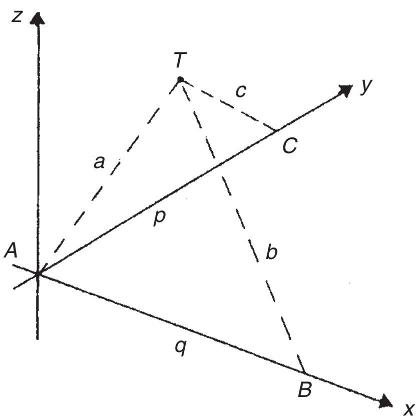


Fig. 13.13 Three-dimensional position measurement system.

13.9.8 Effect of noise in ultrasonic measurement systems

Signal levels at the output of ultrasonic measurement systems are usually of low amplitude and are therefore prone to contamination by electromagnetic noise. Because of this, it is necessary to use special precautions such as making ground (earth) lines thick, using shielded cables for transmission of the signal from the ultrasonic receiver and locating the signal amplifier as close to the receiver as possible.

Another potentially serious form of noise is background ultrasound produced by manufacturing operations in the typical industrial environment that many ultrasonic range measurement systems operate. Analysis of industrial environments has shown that ultrasound at frequencies up to 100 kHz is generated by many operations and some operations generate ultrasound at higher frequencies up to 200 kHz. There is not usually any problem if ultrasonic measurement systems operate at frequencies above 200 kHz, but these often have insufficient range for the needs of the measurement situation. In these circumstances, any objects that are likely to generate energy at ultrasonic frequencies should be covered in sound-absorbing material such that interference with ultrasonic measurement systems is minimized. The placement of sound-absorbing material around the path that the measurement ultrasound wave travels along contributes further towards reducing the effect of background noise. A natural solution to the problem is also partially provided by the fact that the same processes of distance travelled and adsorption that attenuate the amplitude of ultrasound waves travelling between the transmitter and receiver in the measurement system also attenuate ultrasound noise that is generated by manufacturing operations.

Because ultrasonic energy is emitted at angles other than the direction that is normal to the face of the transmitting element, a problem arises in respect of energy that is reflected off some object in the environment around the measurement system and back into the ultrasonic receiver. This has a longer path than the direct one between the transmitter and receiver and can cause erroneous measurements in some circumstances. One solution to this is to arrange for the transmission-time counter to stop as soon as the receiver first detects the ultrasound wave. This will usually be the wave that has travelled along the direct path, and so no measurement error is caused as long as the rate at which ultrasound pulses are emitted is such that the next burst isn't emitted until all reflections from the previous pulse have died down. However, in circumstances where the direct path becomes obstructed by some obstacle, the counter will only be stopped when the reflected signal is detected by the receiver, giving a potentially large measurement error.

13.9.9 Exploiting Doppler shift in ultrasound transmission

The Doppler effect is evident in all types of wave motion and describes the apparent change in frequency of the wave when there is relative motion between the transmitter and receiver. If a continuous ultrasound wave with velocity is v and frequency f takes t seconds to travel from a source S to a receiver R, then R will receive ft cycles of sound during time t (see Figure 13.14). Suppose now that R moves towards S at velocity r (with S stationary). R will receive rt/λ extra cycles of sound during time t, increasing the total number of sound cycles received to $(ft + rt/\lambda)$. With $(ft + rt/\lambda)$

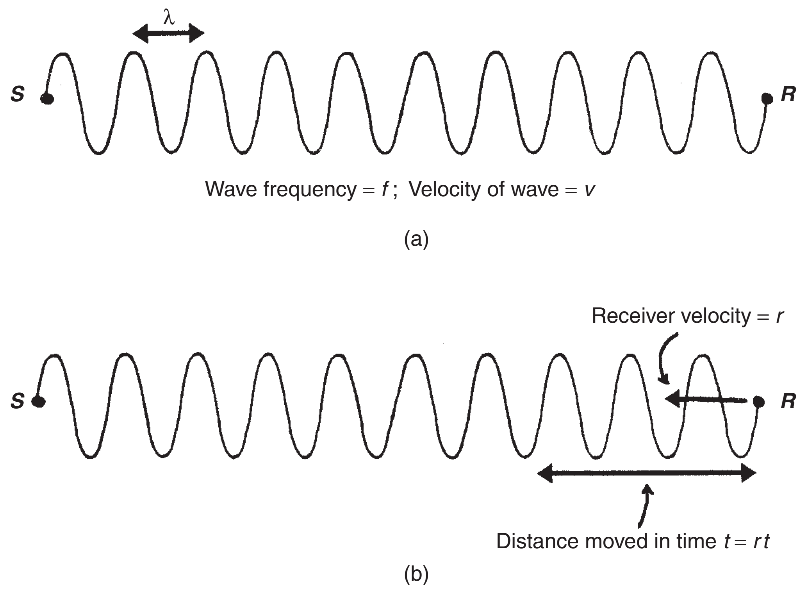


Fig. 13.14 Illustration of Doppler effect.

cycles received in t seconds, the apparent frequency f' is given by:

$$f' = \frac{ft + rt/\lambda}{t} = f + r/\lambda = f + \frac{rf}{v} = \frac{f(r+v)}{v}$$

$$\left(\text{using the relation } \frac{1}{\lambda} = \frac{f}{v} \text{ from equation 13.3}\right)$$

The frequency difference Δf can be expressed as:

$$\Delta f = f' - f = \frac{f(v+r)}{v} - f = \frac{fr}{v}$$

from which the velocity of the receiver r can be expressed as

$$r = v\Delta f / f.$$

Similarly, it can be shown that, if R moves away from S with velocity r, f' is given by:

$$f' = \frac{f(v-r)}{v}$$

and

$$\Delta f = -\frac{fr}{v}$$

If the ultrasound source moves towards the stationary receiver at velocity s, it will move a distance st in time t and the ft cycles that are emitted during time t will be compressed into a distance $(vt - st)$.

Hence, the apparent wavelength λ' will be given by:

$$\lambda' = \frac{vt - st}{ft} = \frac{v - s}{f}$$

Using equation (13.3), this can be expressed alternatively as:

$$f' = \frac{v}{\lambda'} = \frac{vf}{v - s}$$

Similarly, with S moving away from R, it can be shown that:

$$f' = \frac{vf}{v + s}$$

Thus, the velocity of an ultrasound receiver moving with respect to an ultrasound source can be calculated from the measured ratio between the real and apparent frequencies of the wave. This is used in devices like the Doppler shift flowmeter.

13.9.10 Ultrasonic imaging

The main applications of ultrasound in imaging are found in medical diagnosis and in industrial testing procedures. In both of these applications, a short burst of ultrasonic energy is transmitted from the ultrasonic element into the medium being investigated and the energy that is reflected back into the element is analysed. Ultrasonic elements in the frequency range 1 MHz to 15 MHz are used.

Ultrasound is reflected back at all interfaces between different materials, with the proportion of energy reflected being a function of the materials either side of the interface. The principal components inside a human body are water, fat, muscle and bone, and the interfaces between each of these have different reflectance characteristics. Measurement of the time between energy transmission and receipt of the reflected signal gives the depth of the interface according to equation (13.5). Therefore, in medical diagnosis procedures, the reflected energy appears as a series of peaks, with the magnitude of each peak corresponding to the type of interface that it is reflected from and the time of each peak corresponding to the depth of the interface in the body. Thus, a ‘map’ of fat, muscle and bone in the body is obtained. A fuller account can be found elsewhere (Webster, 1998).

Applications in industrial test procedures usually involve detecting internal flaws within components. Such flaws cause an interface between air and the material that the component is made of. By timing the reflections of ultrasound from the flaw, the depth of each flaw is determined.

13.10 Nuclear sensors

Nuclear sensors are uncommon measurement devices, partly because of the strict safety regulations that govern their use, and partly because they are usually expensive. Some very low-level radiation sources are now available that largely overcome the safety problems, but measurements are then prone to contamination by background radiation. The principle of operation of nuclear sensors is very similar to optical sensors in that radiation is transmitted between a source and a detector through some medium in which the magnitude of transmission is attenuated according to the value of the measured variable. Caesium-137 is commonly used as a gamma-ray source and a sodium iodide device is commonly used as a gamma-ray detector. The latter gives a voltage output that is proportional to the radiation incident upon it. One current use of nuclear sensors is in a non-invasive technique for measuring the level of liquid in storage tanks (see Chapter 17). They are also used in mass flow rate measurement (see Chapter 16) and in medical scanning applications (see Webster, 1998).

13.11 Microsensors

Microsensors are millimetre-sized two- and three-dimensional micromachined structures that have smaller size, improved performance, better reliability and lower production costs than many alternative forms of sensor. Currently, devices to measure temperature, pressure, force, acceleration, humidity, magnetic fields, radiation and chemical parameters are either in production or at advanced stages of research.

Microsensors are usually constructed from a silicon semiconductor material, but are sometimes fabricated from other materials such as metals, plastics, polymers, glasses and ceramics that are deposited on a silicon base. Silicon is an ideal material for sensor construction because of its excellent mechanical properties. Its tensile strength and Young's modulus is comparable to that of steel, whilst its density is less than that of aluminium. Sensors made from a single crystal of silicon remain elastic almost to the breaking point, and mechanical hysteresis is very small. In addition, silicon has a very low coefficient of thermal expansion and can be exposed to extremes of temperature and most gases, solvents and acids without deterioration.

Microengineering techniques are an essential enabling technology for microsensors, which are designed so that their electromechanical properties change in response to a change in the measured parameter. Many of the techniques used for integrated circuit (IC) manufacture are also used in sensor fabrication, common techniques being crystal growing and polishing, thin film deposition, ion implantation, wet and dry chemical and laser etching, and photolithography. However, apart from standard IC production techniques, some special techniques are also needed in addition to produce the 3D structures that are unique to some types of microsensor. The various manufacturing techniques are used to form sensors directly in silicon crystals and films. Typical structures have forms such as thin diaphragms, cantilever beams and bridges.

Whilst the small size of a microsensor is of particular benefit in many applications, it also leads to some problems that require special attention. For example, microsensors typically have very low capacitance. This makes the output signals very prone to noise contamination. Hence, it is usually necessary to integrate microelectronic circuits that perform signal processing in the device, which therefore becomes a *smart microsensor*. Another problem is that microsensors generally produce output signals of very low magnitude. This requires the use of special types of analogue-to-digital converter that can cope with such low-amplitude input signals. One suitable technique is sigma–delta conversion. This is based on charge balancing techniques and gives better than 16-bit accuracy in less than 20 ms (Riedijk, 1997). Special designs can reduce conversion time to less than 0.1 ms if necessary.

At present, almost all smart microsensors have an analogue output. However, a resonant-technology pressure-measuring device is now available with a digital output. This consists of a silicon crystal on which two H-shaped resonators are formed, one at the centre and one at the edge. If the pressure to be measured is applied to the crystal, the central resonator is compressed, changing the spring constant of the material and thus reducing its resonant frequency. At the same time, the outer resonator is stretched, increasing its resonant frequency. The resulting frequency difference produces a digital output signal that is proportional to the applied pressure. The device can also give a signal proportional to differential pressure if this is applied between the centre and periphery of the crystal.

Microsensors are used most commonly for measuring pressure, acceleration, force and chemical parameters. They are used in particularly large numbers in the automotive industry, where unit prices can be as low as £5–£10. Microsensors are also widely used in medical applications, particularly for blood pressure measurement, with unit prices down to £10.

Mechanical microsensors transform measured variables such as force, pressure and acceleration into a displacement. The displacement is usually measured by capacitive or piezoresistive techniques, although some devices use other technologies such as resonant frequency variation, resistance change, inductance change, the piezoelectric effect and changes in magnetic or optical coupling. The design of a cantilever silicon microaccelerometer is shown in Figure 13.15. The proof mass within this is about 100 μm across and the typical deflection measured is of the order of 1 micron (10^{-3} mm).

An alternative capacitive microaccelerometer provides a calibrated, compensated and amplified output. It has a capacitive silicon microsensor to measure displacement of the proof mass. This is integrated with a signal processing chip and protected by a plastic enclosure. The capacitive element has a 3D structure, which gives a higher measurement sensitivity than surface-machined elements.

Microsensors to measure many other physical variables are either in production or at advanced stages of research. Microsensors measuring magnetic field are based on a number of alternative technologies such as Hall-effect, magnetoresistors, magnetodiodes and magnetotransistors. Radiation microsensors are made from silicon p-n diodes or avalanche photodiodes and can detect radiation over wavelengths from the visible spectrum to infrared. Microsensors in the form of a micro thermistor, a p-n thermodiode

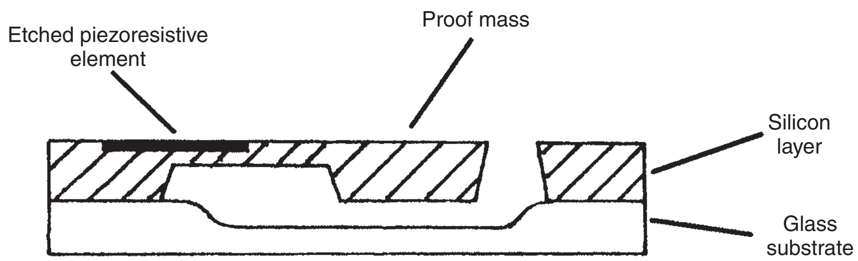


Fig. 13.15 Silicon microaccelerometer.

or a thermotransistor are used as digital thermometers. Microsensors have also enabled measurement techniques that were previously laboratory-based ones to be extended into field instruments. Examples are spectroscopic instruments and devices to measure viscosity.

References and further reading

Grattan, K.T.V. (1989) New developments in sensor technology – fibre-optics and electro-optics, *Measurement and Control*, **22**(6), pp. 165–175.

Harmer, A.L. (1982) Principles of optical fibre sensors and instrumentation, *Measurement and Control*, **15**(4), pp. 143–151.

Johnson, J.S. (1994) Optical sensors: the OSCA experience, *Measurement and Control*, 27, pp. 180–184.

Medlock, R.S. (1986) Review of modulating techniques for fibre optic sensors, *Measurement and Control*, **19**(1), pp. 4–17.

Rafiq, M. and Wykes, C. (1991) The performance of capacitive ultrasonic transducers, *J. Meas. Sci. Technology*, pp. 168–174.

Riedijk, F.R. and Huijsing, J.H. (1997) Sensor interface environment based on a serial bus interface, *Measurement and Control*, **30**, pp. 297–299.

Webster, J.G. (1998) *Medical Instrumentation*, John Wiley.

14

Temperature measurement

14.1 Principles of temperature measurement

Temperature measurement is very important in all spheres of life and especially so in the process industries. However, it poses particular problems, since temperature measurement cannot be related to a fundamental standard of temperature in the same way that the measurement of other quantities can be related to the primary standards of mass, length and time. If two bodies of lengths l_1 and l_2 are connected together end to end, the result is a body of length $l_1 + l_2$. A similar relationship exists between separate masses and separate times. However, if two bodies at the same temperature are connected together, the joined body has the same temperature as each of the original bodies.

This is a root cause of the fundamental difficulties that exist in establishing an absolute standard for temperature in the form of a relationship between it and other measurable quantities for which a primary standard unit exists. In the absence of such a relationship, it is necessary to establish fixed, reproducible reference points for temperature in the form of freezing and boiling points of substances where the transition between solid, liquid and gaseous states is sharply defined. The *International Practical Temperature Scale* (IPTS)* uses this philosophy and defines six *primary fixed points* for reference temperatures in terms of:

- the triple point of equilibrium hydrogen $-259.34°C$
- the boiling point of oxygen $-182.962°C$
- the boiling point of water 100.0°C
- the freezing point of zinc 419.58°C
- the freezing point of silver 961.93°C
- the freezing point of gold 1064.43°C

  (all at standard atmospheric pressure)

The freezing points of certain other metals are also used as *secondary fixed points* to provide additional reference points during calibration procedures.

* The IPTS is subject to periodic review and improvement as research produces more precise fixed reference points. The latest version was published in 1990.

Instruments to measure temperature can be divided into separate classes according to the physical principle on which they operate. The main principles used are:

- The thermoelectric effect
- Resistance change
- Sensitivity of semiconductor device
- Radiative heat emission
- Thermography
- Thermal expansion
- Resonant frequency change
- Sensitivity of fibre optic devices
- Acoustic thermometry
- Colour change
- Change of state of material.

14.2 Thermoelectric effect sensors (thermocouples)

Thermoelectric effect sensors rely on the physical principle that, when any two different metals are connected together, an e.m.f., which is a function of the temperature, is generated at the junction between the metals. The general form of this relationship is:

$$e = a_1 T + a_2 T^2 + a_3 T^3 + \cdots + a_n T^n \tag{14.1}$$

where e is the e.m.f. generated and T is the absolute temperature.

This is clearly non-linear, which is inconvenient for measurement applications. Fortunately, for certain pairs of materials, the terms involving squared and higher powers of T ($a_2 T^2$, $a_3 T^3$ etc.) are approximately zero and the e.m.f.–temperature relationship is approximately linear according to:

$$e \approx a_1 T \tag{14.2}$$

Wires of such pairs of materials are connected together at one end, and in this form are known as *thermocouples*. Thermocouples are a very important class of device as they provide the most commonly used method of measuring temperatures in industry.

Thermocouples are manufactured from various combinations of the base metals copper and iron, the base-metal alloys of alumel (Ni/Mn/Al/Si), chromel (Ni/Cr), constantan (Cu/Ni), nicrosil (Ni/Cr/Si) and nisil (Ni/Si/Mn), the noble metals platinum and tungsten, and the noble-metal alloys of platinum/rhodium and tungsten/rhenium. Only certain combinations of these are used as thermocouples and each standard combination is known by an internationally recognized type letter, for instance type K is chromel–alumel. The e.m.f.–temperature characteristics for some of these standard thermocouples are shown in Figure 14.1: these show reasonable linearity over at least part of their temperature-measuring ranges.

A typical thermocouple, made from one chromel wire and one constantan wire, is shown in Figure 14.2(a). For analysis purposes, it is useful to represent the thermocouple by its equivalent electrical circuit, shown in Figure 14.2(b). The e.m.f. generated at the point where the different wires are connected together is represented by a voltage

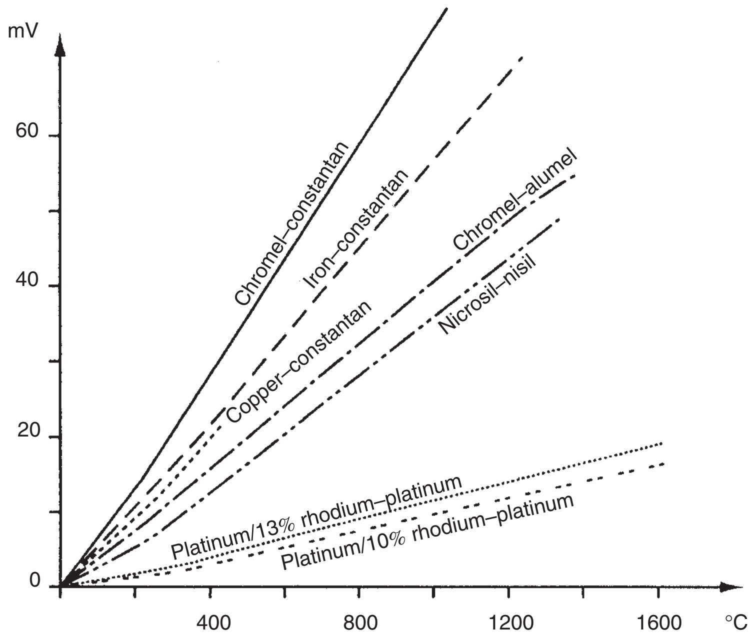


Fig. 14.1 E.m.f. temperature characteristics for some standard thermocouple materials.

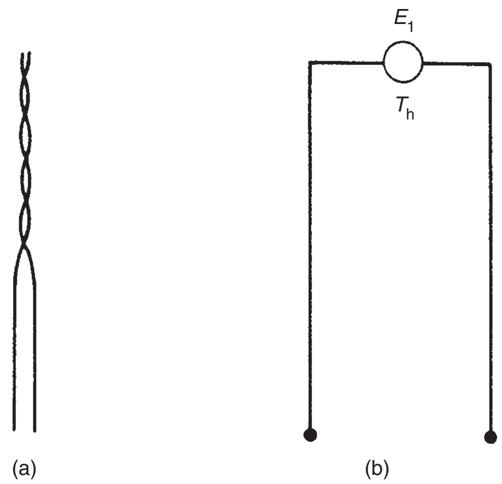


Fig. 14.2 (a) Thermocouple; (b) equivalent circuit.

source, E_1, and the point is known as the *hot junction*. The temperature of the hot junction is customarily shown as T_h on the diagram. The e.m.f. generated at the hot junction is measured at the open ends of the thermocouple, which is known as the *reference junction*.

In order to make a thermocouple conform to some precisely defined e.m.f.–temperature characteristic, it is necessary that all metals used are refined to a high degree of pureness and all alloys are manufactured to an exact specification. This makes the materials used expensive, and consequently thermocouples are typically only a few centimetres long. It is clearly impractical to connect a voltage-measuring instrument at the open end of the thermocouple to measure its output in such close proximity to the environment whose temperature is being measured, and therefore *extension leads* up to several metres long are normally connected between the thermocouple and the measuring instrument. This modifies the equivalent circuit to that shown in Figure 14.3(a). There are now three junctions in the system and consequently three voltage sources, E_1, E_2 and E_3, with the point of measurement of the e.m.f. (still called the reference junction) being moved to the open ends of the extension leads.

The measuring system is completed by connecting the extension leads to the voltage-measuring instrument. As the connection leads will normally be of different materials to those of the thermocouple extension leads, this introduces two further e.m.f.-generating junctions E_4 and E_5 into the system as shown in Figure 14.3(b). The net output e.m.f. measured (E_m) is then given by:

$$E_m = E_1 + E_2 + E_3 + E_4 + E_5 \tag{14.3}$$

and this can be re-expressed in terms of E_1 as:

$$E_1 = E_m - E_2 - E_3 - E_4 - E_5 \tag{14.4}$$

In order to apply equation (14.1) to calculate the measured temperature at the hot junction, E_1 has to be calculated from equation (14.4). To do this, it is necessary to calculate the values of E_2, E_3, E_4 and E_5.

It is usual to choose materials for the extension lead wires such that the magnitudes of E_2 and E_3 are approximately zero, irrespective of the junction temperature. This avoids the difficulty that would otherwise arise in measuring the temperature of the junction between the thermocouple wires and the extension leads, and also in determining the e.m.f./temperature relationship for the thermocouple–extension lead combination.

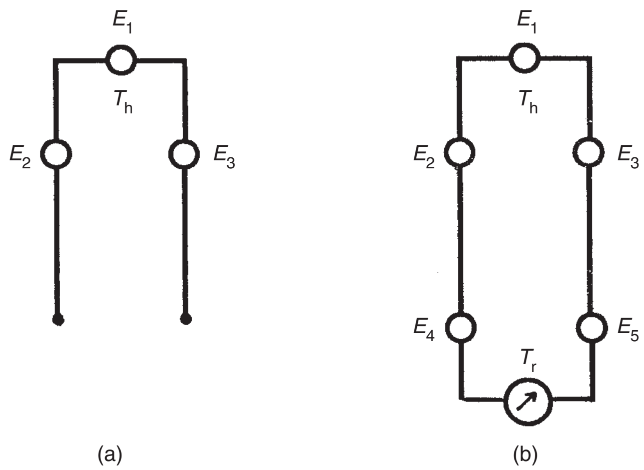


Fig. 14.3 (a) Equivalent circuit for thermocouple with extension leads; (b) equivalent circuit for thermocouple and extension leads connected to a meter.

A zero junction e.m.f. is most easily achieved by choosing the extension leads to be of the same basic materials as the thermocouple, but where their cost per unit length is greatly reduced by manufacturing them to a lower specification. However, such a solution is still prohibitively expensive in the case of noble metal thermocouples, and it is necessary in this case to search for base-metal extension leads that have a similar thermoelectric behaviour to the noble-metal thermocouple. In this form, the extension leads are usually known as *compensating leads*. A typical example of this is the use of nickel/copper–copper extension leads connected to a platinum/rhodium–platinum thermocouple. Copper compensating leads are also sometimes used with some types of base metal thermocouples and, in such cases, the law of intermediate metals can be applied to compensate for the e.m.f. at the junction between the thermocouple and compensating leads.

To analyse the effect of connecting the extension leads to the voltage-measuring instrument, a thermoelectric law known as the *law of intermediate metals* can be used. This states that the e.m.f. generated at the junction between two metals or alloys A and C is equal to the sum of the e.m.f. generated at the junction between metals or alloys A and B and the e.m.f. generated at the junction between metals or alloys B and C, where all junctions are at the same temperature. This can be expressed more simply as:

$$e_{AC} = e_{AB} + e_{BC} \tag{14.5}$$

Suppose we have an iron–constantan thermocouple connected by copper leads to a meter. We can express E_4 and E_5 in Figure 14.4 as:

$$E_4 = e_{\text{iron-copper}}; \quad E_5 = e_{\text{copper-constantan}}$$

The sum of E_4 and E_5 can be expressed as:

$$E_4 + E_5 = e_{\text{iron-copper}} + e_{\text{copper-constantan}}$$

Applying equation (14.5):

$$e_{\text{iron-copper}} + e_{\text{copper-constantan}} = e_{\text{iron-constantan}}$$

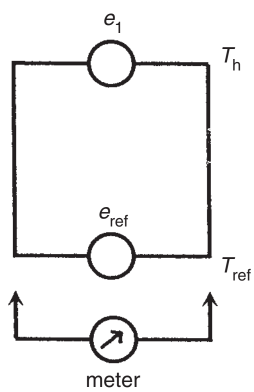


Fig. 14.4 Effective e.m.f. sources in a thermocouple measurement system.

Thus, the effect of connecting the thermocouple extension wires to the copper leads to the meter is cancelled out, and the actual e.m.f. at the reference junction is equivalent to that arising from an iron–constantan connection at the reference junction temperature, which can be calculated according to equation (14.1). Hence, the equivalent circuit in Figure 14.3(b) becomes simplified to that shown in Figure 14.4. The e.m.f. E_m measured by the voltage-measuring instrument is the sum of only two e.m.f.s, consisting of the e.m.f. generated at the hot junction temperature E_1 and the e.m.f. generated at the reference junction temperature E_{ref}. The e.m.f. generated at the hot junction can then be calculated as:

$$E_1 = E_m + E_{ref}$$

E_{ref} can be calculated from equation (14.1) if the temperature of the reference junction is known. In practice, this is often achieved by immersing the reference junction in an ice bath to maintain it at a reference temperature of 0°C. However, as discussed in the following section on thermocouple tables, it is very important that the ice bath remains exactly at 0°C if this is to be the reference temperature assumed, otherwise significant measurement errors can arise. For this reason, refrigeration of the reference junction at a temperature of 0°C is often preferred.

14.2.1 Thermocouple tables

Although the preceding discussion has suggested that the unknown temperature T can be evaluated from the calculated value of the e.m.f. E_1 at the hot junction using equation (14.1), this is very difficult to do in practice because equation (14.1) is a high order polynomial expression. An approximate translation between the value of E_1 and temperature can be achieved by expressing equation (14.1) in graphical form as in Figure 14.1. However, this is not usually of sufficient accuracy, and it is normal practice to use tables of e.m.f. and temperature values known as *thermocouple tables*. These include compensation for the effect of the e.m.f. generated at the reference junction (E_{ref}), which is assumed to be at 0°C. Thus, the tables are only valid when the reference junction is exactly at this temperature. Compensation for the case where the reference junction temperature is not at zero is considered later in this section.

Tables for a range of standard thermocouples are given in Appendix 4. In these tables, a range of temperatures is given in the left-hand column and the e.m.f. output for each standard type of thermocouple is given in the columns to the right. In practice, any general e.m.f. output measurement taken at random will not be found exactly in the tables, and interpolation will be necessary between the values shown in the table.

Example 14.1

If the e.m.f. output measured from a chromel–constantan thermocouple is 13.419 mV with the reference junction at 0°C, the appropriate column in the tables shows that this corresponds to a hot junction temperature of 200°C.

Example 14.2

If the measured output e.m.f. for a chromel–constantan thermocouple (reference junction at 0°C) was 10.65 mV, it is necessary to carry out linear interpolation between the temperature of 160°C corresponding to an e.m.f. of 10.501 mV shown in the tables and the temperature of 170°C corresponding to an e.m.f. of 11.222 mV. This interpolation procedure gives an indicated hot junction temperature of 162°C.

14.2.2 Non-zero reference junction temperature

If the reference junction is immersed in an ice bath to maintain it at a temperature of 0°C so that thermocouple tables can be applied directly, the ice in the bath must be in a state of just melting. This is the only state in which ice is exactly at 0°C, and otherwise it will be either colder or hotter than this temperature. Thus, maintaining the reference junction at 0°C is not a straightforward matter, particularly if the environmental temperature around the measurement system is relatively hot. In consequence, it is common practice in many practical applications of thermocouples to maintain the reference junction at a non-zero temperature by putting it into a controlled environment maintained by an electrical heating element. In order to still be able to apply thermocouple tables, correction then has to be made for this non-zero reference junction temperature using a second thermoelectric law known as the *law of intermediate temperatures*. This states that:

$$E_{(T_h,T_0)} = E_{(T_h,T_r)} + E_{(T_r,T_0)} \tag{14.6}$$

where: $E_{(T_h,T_0)}$ is the e.m.f. with the junctions at temperatures T_h and T_0, $E_{(T_h,T_r)}$ is the e.m.f. with the junctions at temperatures T_h and T_r, and $E_{(T_r,T_0)}$ is the e.m.f. with the junctions at temperatures T_r and T_0, T_h is the hot junction measured temperature, T_0 is 0°C and T_r is the non-zero reference junction temperature that is somewhere between T_0 and T_h.

Example 14.3
Suppose that the reference junction of a chromel–constantan thermocouple is maintained at a temperature of 80°C and the output e.m.f. measured is 40.102 mV when the hot junction is immersed in a fluid.
The quantities given are T_r = 80°C and $E_{(T_h,T_r)} = 40.102$ mV
From the tables, $E_{(T_r,T_0)} = 4.983$ mV
Now applying equation (14.6), $E_{(T_h,T_0)} = 40.102 + 4.983 = 45.085$ mV
Again referring to the tables, this indicates a fluid temperature of 600°C.

In using thermocouples, it is essential that they are connected correctly. Large errors can result if they are connected incorrectly, for example by interchanging the extension leads or by using incorrect extension leads. Such mistakes are particularly serious because they do not prevent some sort of output being obtained, which may look sensible even though it is incorrect, and so the mistake may go unnoticed for a long period of time. The following examples illustrate the sort of errors that may arise:

Example 14.4
This example is an exercise in the use of thermocouple tables, but it also serves to illustrate the large errors that can arise if thermocouples are used incorrectly. In a particular industrial situation, a chromel–alumel thermocouple with chromel–alumel

extension wires is used to measure the temperature of a fluid. In connecting up this measurement system, the instrumentation engineer responsible has inadvertently interchanged the extension wires from the thermocouple. The ends of the extension wires are held at a reference temperature of 0°C and the output e.m.f. measured is 14.1 mV. If the junction between the thermocouple and extension wires is at a temperature of 40°C, what temperature of fluid is indicated and what is the true fluid temperature?

Solution

The initial step necessary in solving a problem of this type is to draw a diagrammatical representation of the system and to mark on this the e.m.f. sources, temperatures etc., as shown in Figure 14.5. The first part of the problem is solved very simply by looking up in thermocouple tables what temperature the e.m.f. output of 12.1 mV indicates for a chromel–alumel thermocouple. This is 297.4°C. Then, summing e.m.f.s around the loop:

$$V = 12.1 = E_1 + E_2 + E_3 \quad \text{or} \quad E_1 = 12.1 - E_2 - E_3$$

$$E_2 = E_3 = \text{e.m.f.}_{(\text{alumel}-\text{chromel})_{40}} = -\text{e.m.f.}_{(\text{chromel}-\text{alumel})_{40}}{}^{*} = -1.611\,\text{mV}$$

Hence:

$$E_1 = 12.1 + 1.611 + 1.611 = 15.322\,\text{mV}$$

Interpolating from the thermocouple tables, this indicates that the true fluid temperature is 374.5°C.

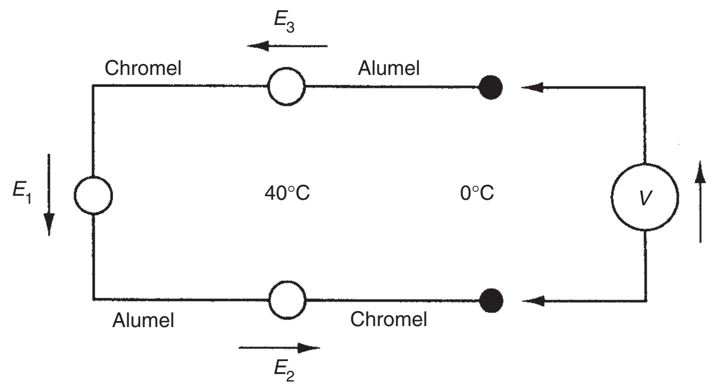


Fig. 14.5 Diagram for solution of example 14.4.

Example 14.5

This example also illustrates the large errors that can arise if thermocouples are used incorrectly. An iron–constantan thermocouple measuring the temperature of a fluid is connected by mistake with copper–constantan extension leads (such that the two constantan wires are connected together and the copper extension wire is connected to the iron thermocouple wire). If the fluid temperature was actually 200°C, and the

---

* The thermocouple tables quote e.m.f. using the convention that going from chromel to alumel is positive. Hence, the e.m.f. going from alumel to chromel is minus the e.m.f. going from chromel to alumel.

junction between the thermocouple and extension wires was at 50°C, what e.m.f. would be measured at the open ends of the extension wires if the reference junction is maintained at 0°C? What fluid temperature would be deduced from this (assuming that the connection mistake was not known about)?

Solution
Again, the initial step necessary is to draw a diagram showing the junctions, temperatures and e.m.f.s, as shown in Figure 14.6. The various quantities can then be calculated:

$$E_2 = \text{e.m.f.}_{(\text{iron}-\text{copper})_{50}}$$

By the law of intermediate metals:

$$\begin{aligned}\text{e.m.f.}_{(\text{iron}-\text{copper})_{50}} &= \text{e.m.f.}_{(\text{iron}-\text{constantan})_{50}} - \text{e.m.f.}_{(\text{copper}-\text{constantan})_{50}}\\ &= 2.585 - 2.035 \text{ (from thermocouple tables)} = 0.55\,\text{mV}\end{aligned}$$

$$E_1 = \text{e.m.f.}_{(\text{iron}-\text{constantan})_{200}} = 10.777 \text{ (from thermocouple tables)}$$

$$V = E_1 - E_2 = 10.777 - 0.55 = 10.227$$

Using tables and interpolating, 10.227 mV indicates a temperature of:

$$\left(\frac{10.227 - 10.222}{10.777 - 10.222}\right) 10 + 190 = 190.1^\circ\text{C}$$

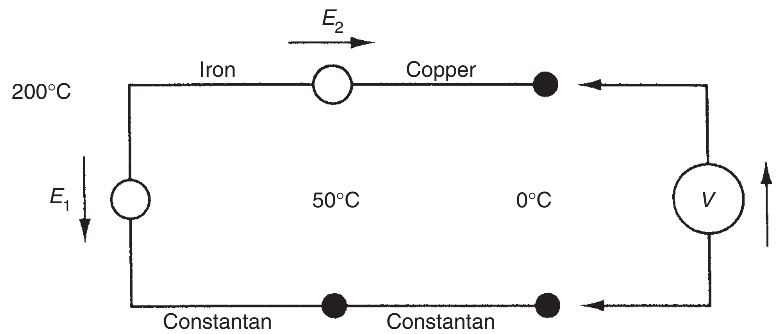


Fig. 14.6 Diagram for solution of example 14.5.

14.2.3 Thermocouple types

The five standard base-metal thermocouples are chromel–constantan (type E), iron–constantan (type J), chromel–alumel (type K), nicrosil–nisil (type N) and copper–constantan (type T). These are all relatively cheap to manufacture but they become inaccurate with age and have a short life. In many applications, performance is also affected through contamination by the working environment. To overcome this, the thermocouple can be enclosed in a *protective sheath*, but this has the adverse effect of introducing a significant time constant, making the thermocouple slow to respond to temperature changes. Therefore, as far as possible, thermocouples are used without protection.

Chromel–constantan devices give the highest measurement sensitivity of $80\,\mu V/^{\circ}C$, with an inaccuracy of ±0.5% and a useful measuring range of −200°C up to 900°C. Unfortunately, whilst they can operate satisfactorily in oxidizing environments when unprotected, their performance and life are seriously affected by reducing atmospheres. Iron–constantan thermocouples have a sensitivity of $60\,\mu V/^{\circ}C$ and are the preferred type for general-purpose measurements in the temperature range −150°C to +1000°C, where the typical measurement inaccuracy is ±0.75%. Their performance is little affected by either oxidizing or reducing atmospheres. Copper–constantan devices have a similar measurement sensitivity of $60\,\mu V/^{\circ}C$ and find their main application in measuring subzero temperatures down to −200°C, with an inaccuracy of ±0.75%. They can also be used in both oxidising and reducing atmospheres to measure temperatures up to 350°C. Chromel–alumel thermocouples have a measurement sensitivity of only $45\,\mu V/^{\circ}C$, although their characteristic is particularly linear over the temperature range between 700°C and 1200°C and this is therefore their main application. Like chromel–constantan devices, they are suitable for oxidizing atmospheres but not for reducing ones unless protected by a sheath. Their measurement inaccuracy is ±0.75%. Nicrosil–nisil thermocouples are a recent development that resulted from attempts to improve the performance and stability of chromel–alumel thermocouples. Their thermoelectric characteristic has a very similar shape to type K devices, with equally good linearity over a large temperature measurement range, measurement sensitivity of $40\,\mu V/^{\circ}C$ and measurement uncertainty of ±0.75%. The operating environment limitations are the same as for chromel–alumel devices but their long-term stability and life are at least three times better. A detailed comparison between type K and N devices can be found in Brooks, (1985).

Noble-metal thermocouples are always expensive but enjoy high stability and long life even when used at high temperatures, though they cannot be used in reducing atmospheres. Thermocouples made from platinum and a platinum–rhodium alloy (type R and type S) have a low inaccuracy of only ±0.5% and can measure temperatures up to 1500°C, but their measurement sensitivity is only $10\,\mu V/^{\circ}C$. Alternative devices made from tungsten and a tungsten/rhenium alloy have a better sensitivity of $20\,\mu V/^{\circ}C$ and can measure temperatures up to 2300°C, though they cannot be used in either oxidizing or reducing atmospheres.

14.2.4 Thermocouple protection

Thermocouples are delicate devices that must be treated carefully if their specified operating characteristics are to be maintained. One major source of error is induced strain in the hot junction. This reduces the e.m.f. output, and precautions are normally taken to minimize induced strain by mounting the thermocouple horizontally rather than vertically. It is usual to cover most of the thermocouple wire with thermal insulation, which also provides mechanical protection, although the tip is left exposed if possible to maximize the speed of response to changes in the measured temperature. However, thermocouples are prone to contamination in some operating environments. This means

Table 14.1 Common sheath materials for thermocouples

Material	*Maximum operating temperature* (°C)*
Mild steel	900
Nickel–chromium	900
Fused silica	1000
Special steel	1100
Mullite	1700
Recrystallized alumina	1850
Beryllia	2300
Magnesia	2400
Zirconia	2400
Thoria	2600

*The maximum operating temperatures quoted assume oxidizing or neutral atmospheres. For operation in reducing atmospheres, the maximum allowable temperature is usually reduced.

that their e.m.f.–temperature characteristic varies from that published in standard tables. Contamination also makes them brittle and shortens their life.

Where they are prone to contamination, thermocouples have to be protected by enclosing them entirely in an insulated sheath. Some common sheath materials and their maximum operating temperatures are shown in Table 14.1. Whilst the thermocouple is a device that has a naturally first order type of step response characteristic, the time constant is usually so small as to be negligible when the thermocouple is used unprotected. However, when enclosed in a sheath, the time constant of the combination of thermocouple and sheath is significant. The size of the thermocouple and hence the diameter required for the sheath has a large effect on the importance of this. The time constant of a thermocouple in a 1 mm diameter sheath is only 0.15 s and this has little practical effect in most measurement situations, whereas a larger sheath of 6 mm diameter gives a time constant of 3.9 s that cannot be ignored so easily.

14.2.5 Thermocouple manufacture

Thermocouples are manufactured by connecting together two wires of different materials, where each material is produced so as to conform precisely with some defined composition specification. This ensures that its thermoelectric behaviour accurately follows that for which standard thermocouple tables apply. The connection between the two wires is effected by welding, soldering or in some cases just by twisting the wire ends together. Welding is the most common technique used generally, with silver soldering being reserved for copper–constantan devices.

The diameter of wire used to construct thermocouples is usually in the range between 0.4 mm and 2 mm. The larger diameters are used where ruggedness and long life are required, although these advantages are gained at the expense of increasing the measurement time constant. In the case of noble-metal thermocouples, the use of large diameter wire incurs a substantial cost penalty. Some special applications have a requirement for a very fast response time in the measurement of temperature, and in such cases wire diameters as small as 0.1 μm (0.1 microns) can be used.

14.2.6 The thermopile

The thermopile is the name given to a temperature-measuring device that consists of several thermocouples connected together in series, such that all the reference junctions are at the same cold temperature and all the hot junctions are exposed to the temperature being measured, as shown in Figure 14.7. The effect of connecting n thermocouples together in series is to increase the measurement sensitivity by a factor of n. A typical thermopile manufactured by connecting together 25 chromel–constantan thermocouples gives a measurement resolution of 0.001°C.

14.2.7 Digital thermometer

Thermocouples are also used in digital thermometers, of which both simple and intelligent versions exist (see section 14.13 for a description of the latter). A simple digital thermometer is the combination of a thermocouple, a battery-powered, dual slope digital voltmeter to measure the thermocouple output, and an electronic display. This provides a low noise, digital output that can resolve temperature differences as small as 0.1°C. The accuracy achieved is dependent on the accuracy of the thermocouple element, but reduction of measurement inaccuracy to ±0.5% is achievable.

14.2.8 The continuous thermocouple

The continuous thermocouple is one of a class of devices that detect and respond to heat. Other devices in this class include the *line-type heat detector* and *heat-sensitive cable*. The basic construction of all these devices consists of two or more strands of wire separated by insulation within a long thin cable. Whilst they sense temperature, they do not in fact provide an output measurement of temperature. Their function is to respond to abnormal temperature rises and thus prevent fires, equipment damage etc.

The advantages of continuous thermocouples become more apparent if the problems with other types of heat detector are considered. The insulation in the line-type heat

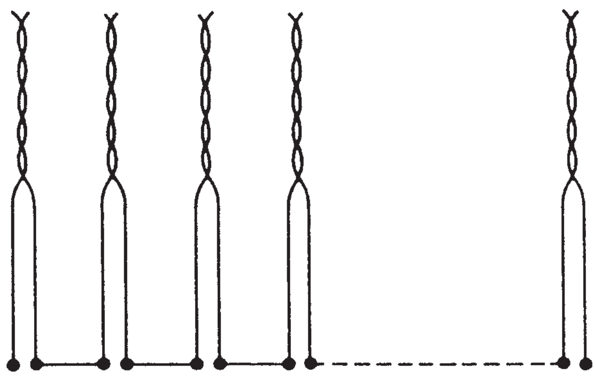


Fig. 14.7 Thermopile.

detector and heat-sensitive cable consists of plastic or ceramic material with a negative temperature coefficient (i.e. the resistance falls as the temperature rises). An alarm signal can be generated when the measured resistance falls below a certain level. Alternatively, in some versions, the insulation is allowed to break down completely, in which case the device acts as a switch. The major limitation of these devices is that the temperature change has to be relatively large, typically 50–200°C above ambient temperature, before the device responds. Also, it is not generally possible for such devices to give an output that indicates that an alarm condition is developing before it actually happens, and thus allow preventative action. Furthermore, after the device has generated an alarm it usually has to be replaced. This is particularly irksome because there is a large variation in the characteristics of detectors coming from different batches and so replacement of the device requires extensive on-site recalibration of the system.

In contrast, the continuous thermocouple suffers from very few of these problems. It differs from other types of heat detector in that the two strands of wire inside it are a pair of thermocouple materials* separated by a special, patented, mineral insulation and contained within a stainless steel protective sheath. If any part of the cable is subjected to heat, the resistance of the insulation at that point is reduced and a 'hot junction' is created between the two wires of dissimilar metals. An e.m.f. is generated at this hot junction according to normal thermoelectric principles.

The continuous thermocouple can detect temperature rises as small as 1°C above normal. Unlike other types of heat detector, it can also monitor abnormal rates of temperature rise and provide a warning of alarm conditions developing before they actually happen. Replacement is only necessary if a great degree of insulation breakdown has been caused by a substantial hot spot at some point along the detector's length. Even then, the use of thermocouple materials of standard characteristics in the detector means that recalibration is not needed if it is replaced. Calibration is not affected either by cable length, and so a replacement cable may be of a different length to the one it is replacing. One further advantage of continuous thermocouples over earlier forms of heat detector is that no power supply is needed, thus significantly reducing installation costs.

14.3 Varying resistance devices

Varying resistance devices rely on the physical principle of the variation of resistance with temperature. The devices are known as either resistance thermometers or thermistors according to whether the material used for their construction is a metal or a semiconductor, and both are common measuring devices. The normal method of measuring resistance is to use a d.c. bridge. The excitation voltage of the bridge has to be chosen very carefully because, although a high value is desirable for achieving high measurement sensitivity, the self-heating effect of high currents flowing in the temperature transducer creates an error by increasing the temperature of the device and so changing the resistance value.

* Normally type E, chromel–constantan, or type K, chromel–alumel.

14.3.1 Resistance thermometers (resistance temperature devices)

Resistance thermometers, which are alternatively known as *resistance temperature devices* (or RTDs), rely on the principle that the resistance of a metal varies with temperature according to the relationship:

$$R = R_0 \left(1 + a_1 T + a_2 T^2 + a_3 T^3 + \cdots + a_n T^n\right) \tag{14.7}$$

This equation is non-linear and so is inconvenient for measurement purposes. The equation becomes linear if all the terms in a_2T^2 and higher powers of T are negligible such that the resistance and temperature are related according to:

$$R \approx R_0 (1 + a_1 T)$$

This equation is approximately true over a limited temperature range for some metals, notably platinum, copper and nickel, whose characteristics are summarized in Figure 14.8. Platinum has the most linear resistance–temperature characteristic, and it also has good chemical inertness, making it the preferred type of resistance thermometer in most applications. Its resistance–temperature relationship is linear within ±0.4% over the temperature range between −200°C and +40°C. Even at +1000°C, the quoted inaccuracy figure is only ±1.2%. Platinum thermometers are made in two forms, as a coil wound on a mandrel and as a film deposited on a ceramic substrate. The nominal resistance at 0°C is typically 100 Ω or 1000 Ω, though 200 Ω and 500 Ω versions also exist. Sensitivity is 0.385 Ω/°C (100 Ω type) or 3.85 Ω/°C (1000 Ω type). A high nominal resistance is advantageous in terms of higher measurement sensitivity, and the resistance of connecting leads has less effect on measurement accuracy. However, cost goes up as the nominal resistance increases.

Besides having a less linear characteristic, both nickel and copper are inferior to platinum in terms of their greater susceptibility to oxidation and corrosion. This seriously limits their accuracy and longevity. However, because platinum is very expensive compared with nickel and copper, the latter are used in resistance thermometers when cost is important. Another metal, tungsten, is also used in resistance thermometers in some circumstances, particularly for high temperature measurements. The working range of each of these four types of resistance thermometer is as shown below:

Platinum: −270°C to +1000°C (though use above 650°C is uncommon)
Copper: −200°C to +260°C
Nickel: −200°C to +430°C
Tungsten: −270°C to +1100°C

In the case of non-corrosive and non-conducting environments, resistance thermometers are used without protection. In all other applications, they are protected inside a sheath. As in the case of thermocouples, such protection reduces the speed of response of the system to rapid changes in temperature. A typical time constant for a sheathed platinum resistance thermometer is 0.4 seconds. Moisture build-up within the sheath can also impair measurement accuracy.

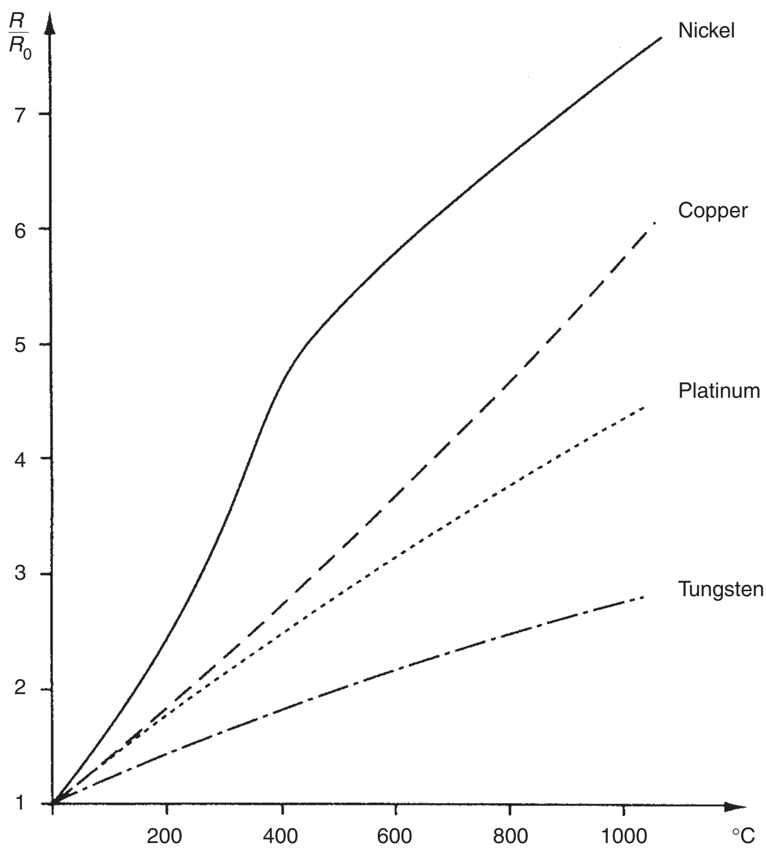


Fig. 14.8 Typical resistance–temperature characteristics of metals.

14.3.2 Thermistors

Thermistors are manufactured from beads of semiconductor material prepared from oxides of the iron group of metals such as chromium, cobalt, iron, manganese and nickel. Normally, thermistors have a negative temperature coefficient, i.e. the resistance decreases as the temperature increases, according to:

$$R = R_0 e^{[\beta(1/T - 1/T_0)]} \tag{14.8}$$

This relationship is illustrated in Figure 14.9. However, alternative forms of heavily doped thermistors are now available (at greater cost) that have a positive temperature coefficient. The form of equation (14.8) is such that it is not possible to make a linear approximation to the curve over even a small temperature range, and hence the thermistor is very definitely a non-linear sensor. However, the major advantages of thermistors are their relatively low cost and their small size. This size advantage means that the time constant of thermistors operated in sheaths is small, although the size reduction also decreases its heat dissipation capability and so makes the self-heating effect greater. In consequence, thermistors have to be operated at generally

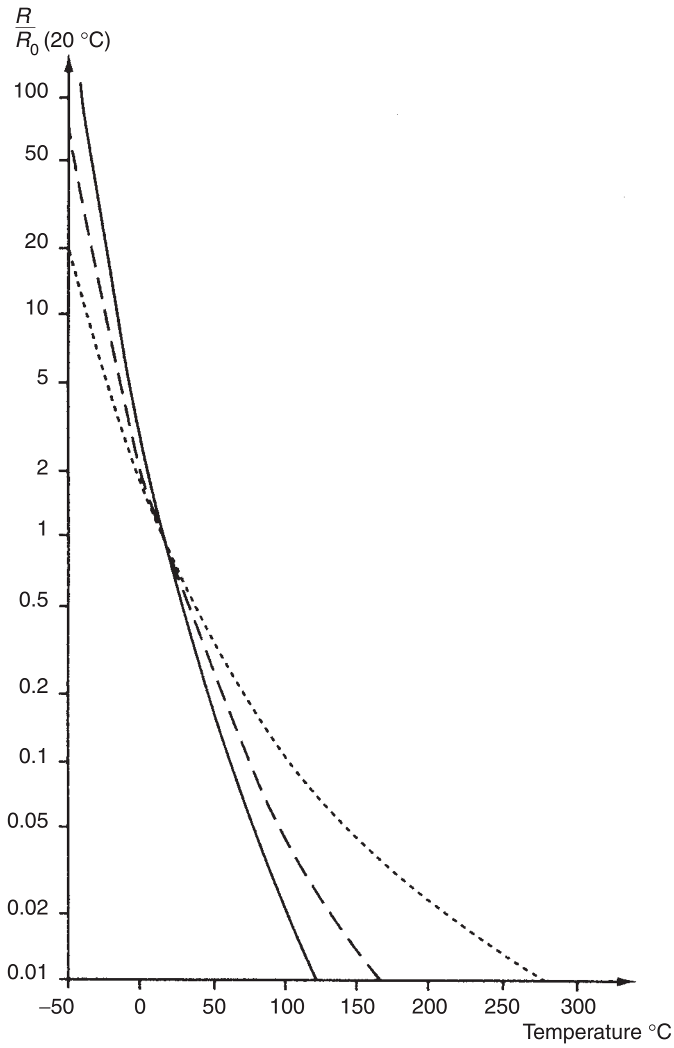


Fig. 14.9 Typical resistance–temperature characteristics of thermistor materials.

lower current levels than resistance thermometers and so the measurement sensitivity is less.

14.4 Semiconductor devices

Semiconductor devices, consisting of either diodes or integrated circuit transistors, have only been commonly used in industrial applications for a few years, but they were first invented several decades ago. They have the advantage of being relatively inexpensive, but one difficulty that affects their use is the need to provide an external power supply to the sensor.

Integrated circuit transistors produce an output proportional to the absolute temperature. Different types are configured to give an output in the form of either a varying current (typically 1 μA/K) or varying voltage (typically 10 mV/K). Current forms are normally used with a digital voltmeter that detects the current output in terms of the voltage drop across a 10 kΩ resistor. Although the devices have a very low cost (typically a few pounds) and a better linearity than either thermocouples or resistance thermometers, they only have a limited measurement range from −50°C to +150°C. Their inaccuracy is typically ±3%, which limits their range of application. However, they are widely used to monitor pipes and cables, where their low cost means that it is feasible to mount multiple sensors along the length of the pipe/cable to detect hot spots.

In diodes, the forward voltage across the device varies with temperature. Output from a typical diode package is in the microamp range. Diodes have a small size, with good output linearity and typical inaccuracy of only ±0.5%. Silicon diodes cover the temperature range from −50 to +200°C and germanium ones from −270 to +40°C.

14.5 Radiation thermometers

All objects emit electromagnetic radiation as a function of their temperature above absolute zero, and radiation thermometers (also known as radiation pyrometers) measure this radiation in order to calculate the temperature of the object. The total rate of radiation emission per second is given by:

$$E = KT^4 \tag{14.9}$$

The power spectral density of this emission varies with temperature in the manner shown in Figure 14.10. The major part of the frequency spectrum lies within the band of wavelengths between 0.3 μm and 40 μm, which corresponds to the visible (0.3–0.72 μm) and infrared (0.72–1000 μm) ranges. As the magnitude of the radiation varies with temperature, measurement of the emission from a body allows the temperature of the body to be calculated. Choice of the best method of measuring the emitted radiation depends on the temperature of the body. At low temperatures, the peak of the power spectral density function (Figure 14.10) lies in the infrared region, whereas at higher temperatures it moves towards the visible part of the spectrum. This phenomenon is observed as the red glow that a body begins to emit as its temperature is increased beyond 600°C.

Different versions of radiation thermometers are capable of measuring temperatures between −100°C and +10 000°C with measurement inaccuracy as low as ±0.05% (though this level of accuracy is not obtained when measuring very high temperatures). Portable, battery-powered, hand-held versions are also available, and these are particularly easy to use. The important advantage that radiation thermometers have over other types of temperature-measuring instrument is that there is no contact with the hot body while its temperature is being measured. Thus, the measured system is not disturbed in any way. Furthermore, there is no possibility of contamination, which is particularly important in food and many other process industries. They are especially suitable for measuring high temperatures that are beyond the capabilities of contact

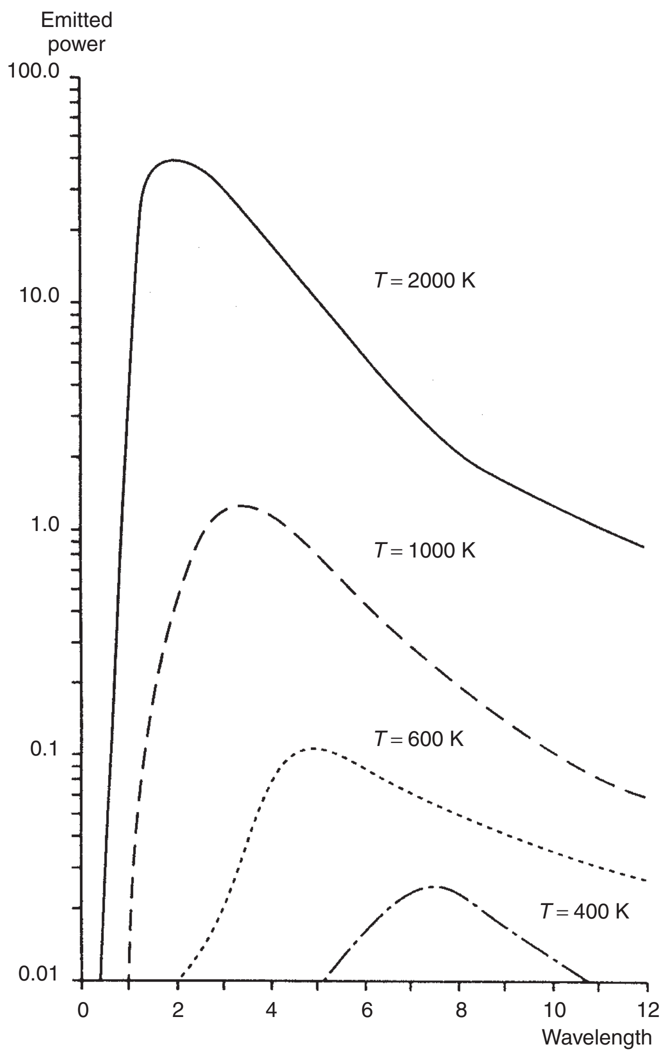


Fig. 14.10 Power spectral density of radiated energy emission at various temperatures.

instruments such as thermocouples, resistance thermometers and thermistors. They are also capable of measuring moving bodies, for instance the temperature of steel bars in a rolling mill. Their use is not as straightforward as the discussion so far might have suggested, however, because the radiation from a body varies with the composition and surface condition of the body as well as with temperature. This dependence on surface condition is quantified by the *emissivity* of the body. The use of radiation thermometers is further complicated by absorption and scattering of the energy between the emitting body and the radiation detector. Energy is scattered by atmospheric dust and water droplets and absorbed by carbon dioxide, ozone and water vapour molecules. Therefore, all radiation thermometers have to be carefully calibrated for each particular body whose temperature they are required to monitor.

Various types of radiation thermometer exist, as described below. The optical pyrometer can only be used to measure high temperatures, but various types of radiation pyrometers are available that between them cover the whole temperature spectrum. Intelligent versions (see section 14.13) also now provide full or partial solution to many of the problems described below for non-intelligent pyrometers.

14.5.1 Optical pyrometers

The optical pyrometer, illustrated in Figure 14.11, is designed to measure temperatures where the peak radiation emission is in the red part of the visible spectrum, i.e. where the measured body glows a certain shade of red according to the temperature. This limits the instrument to measuring temperatures above 600°C. The instrument contains a heated tungsten filament within its optical system. The current in the filament is increased until its colour is the same as the hot body: under these conditions the filament apparently disappears when viewed against the background of the hot body. Temperature measurement is therefore obtained in terms of the current flowing in the filament. As the brightness of different materials at any particular temperature varies according to the emissivity of the material, the calibration of the optical pyrometer must be adjusted according to the emissivity of the target. Manufacturers provide tables of standard material emissivities to assist with this.

The inherent measurement inaccuracy of an optical pyrometer is ±5°C. However, in addition to this error, there can be a further operator-induced error of ±10°C arising out of the difficulty in judging the moment when the filament 'just' disappears. Measurement accuracy can be improved somewhat by employing an optical filter within the instrument that passes a narrow band of frequencies of wavelength around 0.65 μm corresponding to the red part of the visible spectrum. This also extends the upper temperature measurable from 5000°C in unfiltered instruments up to 10 000°C.

The instrument cannot be used in automatic temperature control schemes because the eye of the human operator is an essential part of the measurement system. The

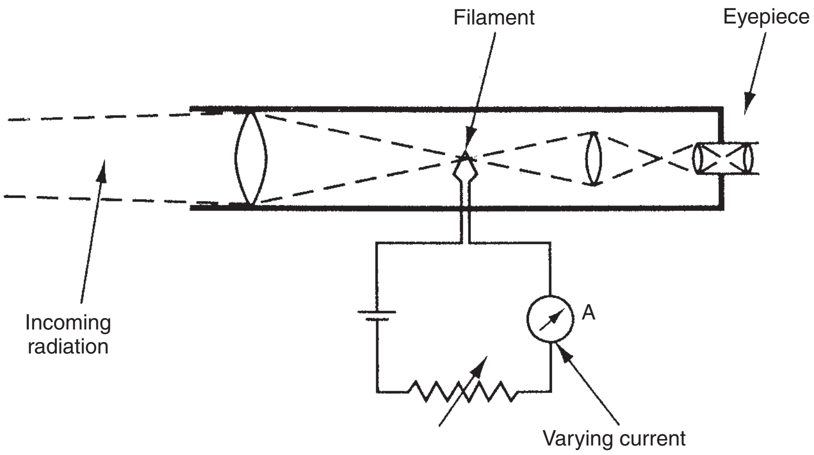


Fig. 14.11 Optical pyrometer.

reading is also affected by fumes in the sight path. Because of these difficulties and its low accuracy, hand-held radiation pyrometers are rapidly overtaking the optical pyrometer in popularity, although the instrument is still widely used in industry for measuring temperatures in furnaces and similar applications at present.

14.5.2 Radiation pyrometers

All the alternative forms of radiation pyrometer described below have an optical system that is similar to that in the optical pyrometer and focuses the energy emitted from the measured body. However, they differ by omitting the filament and eyepiece and having instead an energy detector in the same focal plane as the eyepiece was, as shown in Figure 14.12. This principle can be used to measure temperature over a range from $-100°C$ to $+3600°C$. The radiation detector is either a thermal detector, which measures the temperature rise in a black body at the focal point of the optical system, or a photon detector.

Thermal detectors respond equally to all wavelengths in the frequency spectrum, and consist of either thermopiles, resistance thermometers or thermistors. All of these typically have time constants of several milliseconds, because of the time taken for the black body to heat up and the temperature sensor to respond to the temperature change.

Photon detectors respond selectively to a particular band within the full spectrum, and are usually of the photoconductive or photovoltaic type. They respond to temperature changes very much faster than thermal detectors because they involve atomic processes, and typical measurement time constants are a few microseconds.

Fibre-optic technology is frequently used in high-temperature measurement applications to collect the incoming radiation and transmit it to a detector and processing electronics that are located remotely. This prevents exposure of the processing electronics to potentially damaging, high temperature. Fibre-optic cables are also used to apply radiation pyrometer principles in very difficult applications, such as measuring the temperature inside jet engines by collecting the radiation from inside the engine and transmitting it outside (see section 14.9).

The size of objects measured by a radiation pyrometer is limited by the optical resolution, which is defined as the ratio of target size to distance. A good ratio is 1:300, and this would allow temperature measurement of a 1 mm sized object at a range of 300 mm. With large distance/target size ratios, accurate aiming and focusing of the pyrometer at the target is essential. It is now common to find 'through the lens' viewing provided in pyrometers, using a principle similar to SLR camera technology,

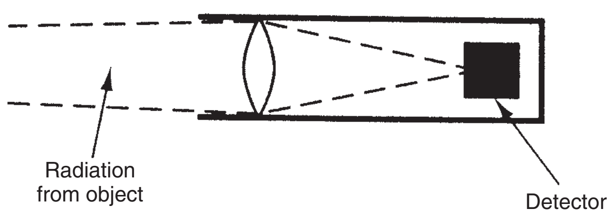


Fig. 14.12 Structure of the radiation thermometer.

as focusing and orientating the instrument for visible light automatically focuses it for infrared light. Alternatively, dual laser beams are sometimes used to ensure that the instrument is aimed correctly towards the target.

Various forms of electrical output are available from the radiation detector: these are functions of the incident energy on the detector and are therefore functions of the temperature of the measured body. Whilst this therefore makes such instruments of use in automatic control systems, their accuracy is often inferior to optical pyrometers. This reduced accuracy arises firstly because a radiation pyrometer is sensitive to a wider band of frequencies than the optical instrument and the relationship between emitted energy and temperature is less well defined. Secondly, the magnitude of energy emission at low temperatures gets very small, according to equation (14.9), increasing the difficulty of accurate measurement.

The forms of radiation pyrometer described below differ mainly in the technique used to measure the emitted radiation. They also differ in the range of energy wavelengths, and hence the temperature range, which each is designed to measure. One further difference is the material used to construct the energy-focusing lens. Outside the visible part of the spectrum, glass becomes almost opaque to infrared wavelengths, and other lens materials such as arsenic trisulphide are used.

Broad-band (unchopped) radiation pyrometers

The broadband radiation pyrometer finds wide application in industry and has a measurement inaccuracy that varies from $\pm 0.05\%$ of full scale in the best instruments to $\pm 0.5\%$ in the cheapest. However, their accuracy deteriorates significantly over a period of time, and an error of 10°C is common after 1–2 years' operation at high temperatures. As its name implies, the instrument measures radiation across the whole frequency spectrum and so uses a thermal detector. This consists of a blackened platinum disc to which a thermopile* is bonded. The temperature of the detector increases until the heat gain from the incident radiation is balanced by the heat loss due to convection and radiation. For high-temperature measurement, a two-couple thermopile gives acceptable measurement sensitivity and has a fast time constant of about 0.1 s. At lower measured temperatures, where the level of incident radiation is much less, thermopiles constructed from a greater number of thermocouples must be used to get sufficient measurement sensitivity. This increases the measurement time constant to as much as 2 s. Standard instruments of this type are available to measure temperatures between −20°C and +1800°C, although in theory much higher temperatures could be measured by this method.

Chopped broad-band radiation pyrometers

The construction of this form of pyrometer is broadly similar to that shown in Figure 14.12 except that a rotary mechanical device is included that periodically interrupts the radiation reaching the detector. The voltage output from the thermal detector thus becomes an alternating quantity that switches between two levels. This form of a.c. output can be amplified much more readily than the d.c. output coming from an unchopped instrument. This is particularly important when amplification is necessary to achieve an acceptable measurement resolution in situations where the

* Typically manganin–constantan.

level of incident radiation from the measured body is low. For this reason, this form of instrument is the more common when measuring body temperatures associated with peak emission in the infrared part of the frequency spectrum. For such chopped systems, the time constant of thermopiles is too long. Instead, thermistors are generally used, giving a time constant of 0.01 s. Standard instruments of this type are available to measure temperatures between $+20^\circ$C and $+1300^\circ$C. This form of pyrometer suffers similar accuracy drift to unchopped forms. Its life is also limited to about two years because of motor failures.

Narrow-band radiation pyrometers

Narrow-band radiation pyrometers are highly stable instruments that suffer a drift in accuracy that is typically only 1°C in 10 years. They are also less sensitive to emissivity changes than other forms of radiation pyrometer. They use photodetectors of either the photoconductive or photovoltaic form whose performance is unaffected by either carbon dioxide or water vapour in the path between the target object and the instrument. A photoconductive detector exhibits a change in resistance as the incident radiation level changes whereas a photovoltaic cell exhibits an induced voltage across its terminals that is also a function of the incident radiation level. All photodetectors are preferentially sensitive to a particular narrow band of wavelengths in the range $0.5\,\mu$m–$1.2\,\mu$m and all have a form of output that varies in a highly non-linear fashion with temperature, and thus a microcomputer inside the instrument is highly desirable. Four commonly used materials for photodetectors are cadmium sulphide, lead sulphide, indium antimonide and lead–tin telluride. Each of these is sensitive to a different band of wavelengths and therefore all find application in measuring the particular temperature ranges corresponding to each of these bands.

The output from the narrow-band radiation pyrometer is normally chopped into an a.c. signal in the same manner as used in the chopped broad-band pyrometer. This simplifies the amplification of the output signal, which is necessary to achieve an acceptable measurement resolution. The typical time constant of a photon detector is only $5\,\mu$s, which allows high chopping frequencies up to 20 kHz. This gives such instruments an additional advantage in being able to measure fast transients in temperature as short as $10\,\mu$s.

Two-colour pyrometer (ratio pyrometer)

As stated earlier, the emitted radiation–temperature relationship for a body depends on its emissivity. This is very difficult to calculate, and therefore in practice all pyrometers have to be calibrated to the particular body they are measuring. The two-colour pyrometer (alternatively known as a ratio pyrometer) is a system that largely overcomes this problem by using the arrangement shown in Figure 14.13. Radiation from the body is split equally into two parts, which are applied to separate narrow-band filters. The outputs from the filters consist of radiation within two narrow bands of wavelength λ_1 and λ_2. Detectors sensitive to these frequencies produce output voltages V_1 and V_2 respectively. The ratio of these outputs, (V_1/V_2), can be shown (see Dixon, 1987) to be a function of temperature and to be independent of the emissivity provided that the two wavelengths λ_1 and λ_2 are close together.

The theoretical basis of the two-colour pyrometer is that the output is independent of emissivity because the emissivities at the two wavelengths λ_1 and λ_2 are equal.

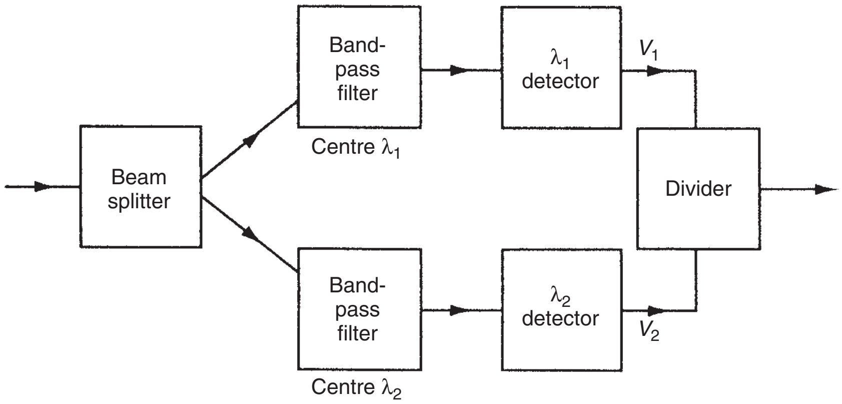


Fig. 14.13 Two-colour pyrometer system.

This is based on the assumption that λ_1 and λ_2 are very close together. In practice, this assumption does not hold and therefore the accuracy of the two-colour pyrometer tends to be relatively poor. However, the instrument is still of great use in conditions where the target is obscured by fumes or dust, which is a common problem in the cement and mineral processing industries. Two-colour pyrometers typically cost 50%–100% more than other types of pyrometer.

Selected waveband pyrometer

The selected waveband pyrometer is sensitive to one waveband only, e.g. 5 μm, and is dedicated to particular, special situations where other forms of pyrometer are inaccurate. One example of such a situation is measuring the temperature of steel billets that are being heated in a furnace. If an ordinary radiation pyrometer is aimed through the furnace door at a hot billet, it receives radiation from the furnace walls (by reflection off the billet) as well as radiation from the billet itself. If the temperature of the furnace walls is measured by a thermocouple, a correction can be made for the reflected radiation, but variations in transmission losses inside the furnace through fumes etc. make this correction inaccurate. However, if a carefully chosen selected-waveband pyrometer is used, this transmission loss can be minimized and the measurement accuracy is thereby greatly improved.

14.6 Thermography (thermal imaging)

Thermography, or thermal imaging, involves scanning an infrared radiation detector across an object. The information gathered is then processed and an output in the form of the temperature distribution across the object is produced. Temperature measurement over the range from −20°C up to +1500°C is possible. Elements of the system are shown in Figure 14.14.

The radiation detector uses the same principles of operation as a radiation pyrometer in inferring the temperature of the point that the instrument is focused on from a measurement of the incoming infrared radiation. However, instead of providing a

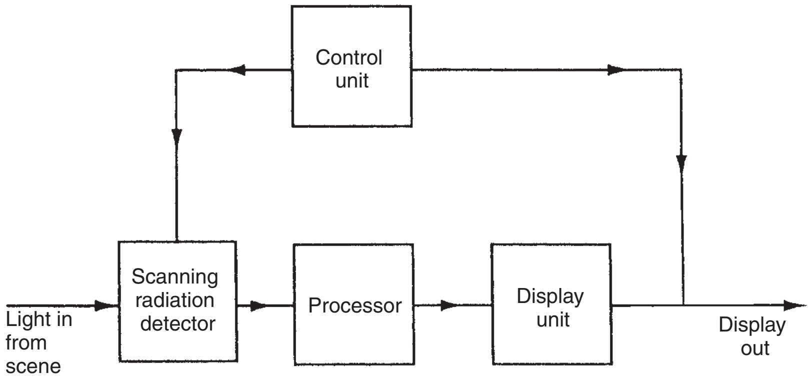


Fig. 14.14 Thermography (thermal imaging) system.

measurement of the temperature of a single point at the focal point of the instrument, the detector is scanned across a body or scene, and thus provides information about temperature distributions. Because of the scanning mode of operation of the instrument, radiation detectors with a very fast response are required, and only photoconductive or photovoltaic sensors are suitable. These are sensitive to the portion of the infrared spectrum between wavelengths of $2\,\mu m$ and $14\,\mu m$.

Simpler versions of thermal imaging instruments consist of hand-held viewers that are pointed at the object of interest. The output from an array of infrared detectors is directed onto a matrix of red light-emitting diodes assembled behind a glass screen, and the output display thus consists of different intensities of red on a black background, with the different intensities corresponding to different temperatures. Measurement resolution is high, with temperature differences as small as $0.1°C$ being detectable. Such instruments are used in a wide variety of applications such as monitoring product flows through pipework, detecting insulation faults, and detecting hot spots in furnace linings, electrical transformers, machines, bearings etc. The number of applications is extended still further if the instrument is carried in a helicopter, where uses include scanning electrical transmission lines for faults, searching for lost or injured people and detecting the source and spread pattern of forest fires.

More complex thermal imaging systems comprise a tripod-mounted detector connected to a desktop computer and display system. Multi-colour displays are commonly used in such systems, where up to 16 different colours represent different bands of temperature across the measured range. The heat distribution across the measured body or scene is thus displayed graphically as a contoured set of coloured bands representing the different temperature levels. Such colour-thermography systems find many applications such as inspecting electronic circuit boards and monitoring production processes. There are also medical applications in body scanning.

14.7 Thermal expansion methods

Thermal expansion methods make use of the fact that the dimensions of all substances, whether solids, liquids or gases, change with temperature. Instruments operating on this physical principle include the liquid-in-glass thermometer, the bimetallic thermometer and the pressure thermometer.

14.7.1 Liquid-in-glass thermometers

The liquid-in-glass thermometer is a well-known temperature-measuring instrument that is used in a wide range of applications. The fluid used is usually either mercury or coloured alcohol, and this is contained within a bulb and capillary tube, as shown in Figure 14.15(a). As the temperature rises, the fluid expands along the capillary tube and the meniscus level is read against a calibrated scale etched on the tube. The process of estimating the position of the curved meniscus of the fluid against the scale introduces some error into the measurement process and a measurement inaccuracy less than $\pm 1\%$ of full-scale reading is hard to achieve.

However, an inaccuracy of only $\pm 0.15\%$ can be obtained in the best industrial instruments. Industrial versions of the liquid-in-glass thermometer are normally used to measure temperature in the range between -200°C and $+1000^{\circ}$C, although instruments are available to special order that can measure temperatures up to 1500°C.

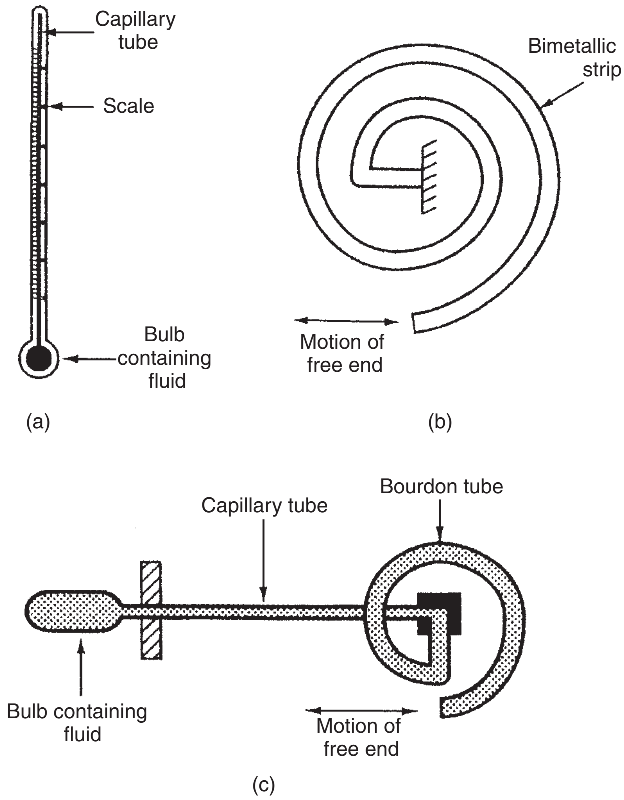


Fig. 14.15 Thermal expansion devices: (a) liquid-in-glass thermometer; (b) bimetallic thermometer; (c) Pressure thermometer.

14.7.2 Bimetallic thermometer

The bimetallic principle is probably more commonly known in connection with its use in thermostats. It is based on the fact that if two strips of different metals are bonded together, any temperature change will cause the strip to bend, as this is the only way in which the differing rates of change of length of each metal in the bonded strip can be accommodated. In the bimetallic thermostat, this is used as a switch in control applications. If the magnitude of bending is measured, the bimetallic device becomes a thermometer. For such purposes, the strip is often arranged in a spiral or helical configuration, as shown in Figure 14.15(b), as this gives a relatively large displacement of the free end for any given temperature change. The measurement sensitivity is increased further by choosing the pair of materials carefully such that the degree of bending is maximized, with Invar (a nickel–steel alloy) or brass being commonly used.

The system used to measure the displacement of the strip must be carefully designed. Very little resistance must be offered to the end of the strip, otherwise the spiral or helix will distort and cause a false reading in the measurement of the displacement. The device is normally just used as a temperature indicator, where the end of the strip is made to turn a pointer that moves against a calibrated scale. However, some versions produce an electrical output, using either a linear variable differential transformer (LVDT) or a fibre-optic shutter sensor to transduce the output displacement.

Bimetallic thermometers are used to measure temperatures between -75°C and $+1500^\circ$C. The inaccuracy of the best instruments can be as low as $\pm 0.5\%$ but such devices are quite expensive. Many instrument applications do not require this degree of accuracy in temperature measurements, and in such cases much cheaper bimetallic thermometers with substantially inferior accuracy specifications are used.

14.7.3 Pressure thermometers

Pressure thermometers have now been superseded by other alternatives in most applications, but they still remain useful in a few applications such as furnace temperature measurement when the level of fumes prevents the use of optical or radiation pyrometers. Examples can also still be found of their use as temperature sensors in pneumatic control systems. The sensing element in a pressure thermometer consists of a stainless-steel bulb containing a liquid or gas. If the fluid were not constrained, temperature rises would cause its volume to increase. However, because it is constrained in a bulb and cannot expand, its pressure rises instead. As such, the pressure thermometer does not strictly belong to the thermal expansion class of instruments but is included because of the relationship between volume and pressure according to Boyle's law: $PV = KT$.

The change in pressure of the fluid is measured by a suitable pressure transducer such as the Bourdon tube (see Chapter 15). This transducer is located remotely from the bulb and connected to it by a capillary tube as shown in Figure 14.15(c). The need to protect the pressure-measuring instrument from the environment where the temperature is being measured can require the use of capillary tubes up to 5 m long, and the temperature gradient, and hence pressure gradient, along the tube acts as a modifying input that can introduce a significant measurement error. Pressure thermometers can be used to measure temperatures in the range between −250°C and +2000°C and their typical inaccuracy is ±0.5% of full-scale reading. However, the instrument response has a particularly long time constant.

14.8 Quartz thermometers

The quartz thermometer makes use of the principle that the resonant frequency of a material such as quartz is a function of temperature, and thus enables temperature changes to be translated into frequency changes. The temperature-sensing element consists of a quartz crystal enclosed within a probe (sheath). The probe commonly consists of a stainless steel cylinder, which makes the device physically larger than devices like thermocouples and resistance thermometers. The crystal is connected electrically so as to form the resonant element within an electronic oscillator. Measurement of the oscillator frequency therefore allows the measured temperature to be calculated.

The instrument has a very linear output characteristic over the temperature range between −40°C and +230°C, with a typical inaccuracy of ±0.1%. Measurement resolution is typically 0.1°C but versions can be obtained with resolutions as small as 0.0003°C. The characteristics of the instrument are generally very stable over long periods of time and therefore only infrequent calibration is necessary. The frequency-change form of output means that the device is insensitive to noise. However, it is very expensive, with a typical cost of £3000 ($5000).

14.9 Fibre-optic temperature sensors

Fibre-optic cables can be used as either intrinsic or extrinsic temperature sensors, as discussed in Chapter 13, though special attention has to be paid to providing a suitable protective coating when high temperatures are measured. Cost varies from £1000 to £4000, according to type, and the normal temperature range covered is 250°C to 3000°C, though special devices can detect down to 100°C and others can detect up to 3600°C. Their main application is measuring temperatures in hard-to-reach locations, though they are also used when very high measurement accuracy is required. Some laboratory versions have an inaccuracy as low as ±0.01%, which is better than a type S thermocouple, although versions used in industry have a more typical inaccuracy of ±1.0%. Whilst it is often assumed that fibre-optic sensors are intrinsically safe, it has been shown (Johnson, 1994) that flammable gas might be ignited by the optical power levels available from some laser diodes. Thus, the power level used with optical fibres must be carefully chosen, and certification of intrinsic safety is necessary if such sensors are to be used in hazardous environments.

One type of intrinsic sensor uses cable where the core and cladding have similar refractive indices but different temperature coefficients. Temperature rises cause the refractive indices to become even closer together and losses from the core to increase, thus reducing the quantity of light transmitted. Other types of intrinsic temperature sensor include the cross-talk sensor, phase modulating sensor and optical resonator, as described in Chapter 13. Research into the use of distributed temperature sensing using fibre-optic cable has also been reported. This can be used to measure things like the temperature distribution along an electricity supply cable. It works by measuring the reflection characteristics of light transmitted down a fibre-optic cable that is bonded to the electrical cable. By analysing the back-scattered radiation, a table of temperature versus distance along the cable can be produced, with a measurement inaccuracy of only $\pm 0.5°C$.

A common form of extrinsic sensor uses fibre-optic cables to transmit light from a remote targeting lens into a standard radiation pyrometer. This technique can be used with all types of radiation pyrometer, including the two-colour version, and a particular advantage is that this method of measurement is intrinsically safe. However, it is not possible to measure very low temperatures, because the very small radiation levels that exist at low temperatures are badly attenuated during transmission along the fibre-optic cable. Therefore, the minimum temperature that can be measured is about 50°C, and the light guide for this must not exceed 600 mm in length. At temperatures exceeding 1000°C, lengths of fibre up to 20 m long can be successfully used as a light guide.

One extremely accurate device that uses this technique is known as the Accufibre sensor. This is a form of radiation pyrometer that has a black box cavity at the focal point of the lens system. A fibre-optic cable is used to transmit radiation from the black box cavity to a spectrometric device that computes the temperature. This has a measurement range 500°C to 2000°C, a resolution of 10^{-5}°C and an inaccuracy of only $\pm 0.0025\%$ of full scale.

Several other types of device that are marketed as extrinsic fibre-optic temperature sensors consist of a conventional temperature sensor (e.g. a resistance thermometer) connected to a fibre-optic cable so that the transmission of the signal from the measurement point is free of noise. Such devices must include an electricity supply for the electronic circuit that is needed to convert the sensor output into light variations in the cable. Thus, low-voltage power cables must be routed with the fibre-optic cable, and the device is therefore not intrinsically safe.

14.10 Acoustic thermometers

The principle of acoustic thermometry was discovered as long ago as 1873 and uses the fact that the velocity of sound through a gas varies with temperature according to the equation:

$$v = \sqrt{\alpha RT/M} \tag{14.10}$$

where v is the sound velocity, T is the gas temperature, M is the molecular weight of the gas and both R and α are constants. Until very recently, it had only been used for measuring cryogenic (very low) temperatures, but it is now also used for measuring higher temperatures and can potentially measure right up to 20 000°C. However, typical inaccuracy is $\pm 5\%$, and the devices are expensive (typically £6000 or $10 000). The various versions of acoustic thermometer that are available differ according to the technique used for generating sound and measuring its velocity in the gas. If ultrasonic generation is used, the instrument is often known as an *ultrasonic thermometer*. Further information can be found in Michalski, (1991).

14.11 Colour indicators

The colour of various substances and objects changes as a function of temperature. One use of this is in the optical pyrometer as discussed earlier. The other main use of colour change is in special colour indicators that are widely used in industry to determine whether objects placed in furnaces have reached the required temperature. Such colour indicators consist of special paints or crayons that are applied to an object before it is placed in a furnace. The colour-sensitive component within these is some form of metal salt (usually of chromium, cobalt or nickel). At a certain temperature, a chemical reaction takes place and a permanent colour change occurs in the paint or crayon, although this change does not occur instantaneously but only happens over a period of time.

Hence, the colour change mechanism is complicated by the fact that the time of exposure as well as the temperature is important. Such crayons or paints usually have a dual rating that specifies the temperature and length of exposure time required for the colour change to occur. If the temperature rises above the rated temperature, then the colour change will occur in less than the rated exposure time. This causes little problem if the rate of temperature rise is slow with respect to the specified exposure time required for colour change to occur. However, if the rate of rise of temperature is high, the object will be significantly above the rated change temperature of the paint/crayon by the time that the colour change happens. Besides wasting energy by leaving the object in the furnace longer than necessary, this can also cause difficulty if excess temperature can affect the required metallurgical properties of the heated object.

Paints and crayons are available to indicate temperatures between 50°C and 1250°C. A typical exposure time rating is 30 minutes, i.e. the colour change will occur if the paint/crayon is exposed to the rated temperature for this length of time. They have the advantage of low cost, typically a few pounds per application. However, they adhere strongly to the heated object, which can cause difficulty if they have to be cleaned off the object later.

Some liquid crystals also change colour at a certain temperature. According to the design of sensors using such liquid crystals, the colour change can either occur gradually during a temperature rise of perhaps 50°C or else change abruptly at some specified temperature. The latter kind of sensors are able to resolve temperature changes as small as 0.1°C and, according to type, are used over the temperature range from −20°C to +100°C.

14.12 Change of state of materials

Temperature-indicating devices known as Seger cones or pyrometric cones are commonly used in the ceramics industry. They consist of a fused oxide and glass material that is formed into a cone shape. The tip of the cone softens and bends over when a particular temperature is reached. Cones are available that indicate temperatures over the range from 600°C to +2000°C.

14.13 Intelligent temperature-measuring instruments

Intelligent temperature transmitters have now been introduced into the catalogues of most instrument manufacturers, and they bring about the usual benefits associated with intelligent instruments. Such transmitters are separate boxes designed for use with transducers that have either a d.c. voltage output in the mV range or an output in the form of a resistance change. They are therefore suitable for use in conjunction with thermocouples, thermopiles, resistance thermometers, thermistors and broad-band radiation pyrometers. All of the transmitters presently available have non-volatile memories where all constants used in correcting output values for modifying inputs etc. are stored, thus enabling the instrument to survive power failures without losing such information. Facilities in transmitters now available include adjustable damping, noise rejection, self-adjustment for zero and sensitivity drifts and expanded measurement range. These features allow an inaccuracy level of $\pm 0.05\%$ of full scale to be specified.

Mention must be made particularly of intelligent pyrometers, as some versions of these are able to measure the emissivity of the target body and automatically provide an emissivity-corrected output. This particular development provides an alternative to the two-colour pyrometer when emissivity measurement and calibration for other types of pyrometer pose difficulty.

Digital thermometers (see section 14.2) also exist in intelligent versions, where the inclusion of a microprocessor allows a number of alternative thermocouples and resistance thermometers to be offered as options for the primary sensor.

The cost of intelligent temperature transducers is significantly more than their non-intelligent counterparts, and justification purely on the grounds of their superior accuracy is hard to make. However, their expanded measurement range means immediate savings are made in terms of the reduction in the number of spare instruments needed to cover a number of measurement ranges. Their capability for self-diagnosis and self-adjustment means that they require attention much less frequently, giving additional savings in maintenance costs.

14.14 Choice between temperature transducers

The suitability of different instruments in any particular measurement situation depends substantially on whether the medium to be measured is a solid or a fluid. For measuring the temperature of solids, it is essential that good contact is made between the body and the transducer unless a radiation thermometer is used. This restricts the range of suitable transducers to thermocouples, thermopiles, resistance thermometers, thermistors, semiconductor devices and colour indicators. On the other hand, fluid temperatures can be measured by any of the instruments described in this chapter, with the exception of radiation thermometers.

The most commonly used device in industry for temperature measurement is the base-metal thermocouple. This is relatively cheap, with prices varying widely from a few pounds upwards according to the thermocouple type and sheath material used. Typical inaccuracy is $\pm 0.5\%$ of full scale over the temperature range $-250°C$ to $+1200°C$. Noble metal thermocouples are much more expensive, but are chemically inert and can measure temperatures up to 2300°C with an inaccuracy of ±0.2% of full scale. However, all types of thermocouple have a low-level output voltage, making them prone to noise and therefore unsuitable for measuring small temperature differences.

Resistance thermometers are also in common use within the temperature range −270°C to +650°C, with a measurement inaccuracy of ±0.5%. Whilst they have a smaller temperature range than thermocouples, they are more stable and can measure small temperature differences. The platinum resistance thermometer is generally regarded as offering the best ratio of price to performance for measurement in the temperature range −200°C to +500°C, with prices starting from £15.

Thermistors are another relatively common class of devices. They are small and cheap, with a typical cost of around £5. They give a fast output response to temperature changes, with good measurement sensitivity, but their measurement range is quite limited.

Dual diverse sensors are a new development that include a thermocouple and a resistance thermometer inside the same sheath. Both of these devices are affected by various factors in the operating environment, but each tends to be sensitive to different things in different ways. Thus, comparison of the two outputs means that any change in characteristics is readily detected, and appropriate measures to replace or recalibrate the sensors can be taken.

Pulsed sensors are a further recent development. They consist of a water-cooled thermocouple or resistance thermometer, and enable temperature measurement to be made well above the normal upper temperature limit for these devices. At the measuring instant, the water-cooling is temporarily stopped, causing the temperature in the sensor to rise towards the process temperature. Cooling is restarted before the sensor temperature rises to the level where the sensor would be damaged, and the process temperature is then calculated by extrapolating from the measured temperature according to the exposure time.

Semiconductor devices have a better linearity than thermocouples and resistance thermometers and a similar level of accuracy. Thus they are a viable alternative to these in many applications. Integrated circuit transistor sensors are particularly cheap (from £10 each), although their accuracy is relatively poor and they have a very limited measurement range (−50°C to +150°C). Diode sensors are much more accurate and have a wider temperature range (−270°C to +200°C), though they are also more expensive (typical costs are anywhere from £50 to £500).

A major virtue of radiation thermometers is their non-contact, non-invasive mode of measurement. Costs vary from £250 up to £3000 according to type. Although calibration for the emissivity of the measured object often poses difficulties, some instruments now provide automatic calibration. Optical pyrometers are used to monitor temperatures above 600°C in industrial furnaces etc., but their inaccuracy is typically ±5%. Various forms of radiation pyrometer are used over the temperature range between −20°C and +1800°C and can give measurement inaccuracies as low as ±0.05%. One particular merit of narrow-band radiation pyrometers is their ability to measure fast temperature transients of duration as small as 10 μs. No other instrument can measure transients anywhere near as fast as this.

The range of instruments working on the thermal expansion principle are mainly used as temperature indicating devices rather than as components within automatic control schemes. Temperature ranges and costs are: mercury-in-glass thermometers up to +1000°C (cost from a few pounds), bi-metallic thermometers up to +1500°C (cost £50 to £100) and pressure thermometers up to +2000°C (cost £100 to £500). The usual measurement inaccuracy is in the range ±0.5% to ±1.0%. The bimetallic thermometer is more rugged than liquid-in-glass types but less accurate (however, the greater inherent accuracy of liquid-in-glass types can only be realized if the liquid meniscus level is read carefully).

Fibre optic devices are more expensive than most other forms of temperature sensor (costing up to £4000) but provide a means of measuring temperature in very inaccessible locations. Inacccuracy varies from ±1% down to ±0.01% in some laboratory versions. Measurement range also varies with type, but up to +3600°C is possible.

The quartz thermometer provides very high resolution (0.0003°C is possible with special versions) but is expensive because of the complex electronics required to analyse the frequency-change form of output. A typical price is £3000 ($5000). It only operates over the limited temperature range of −40°C to +230°C, but gives a low measurement inaccuracy of ±0.1% within this range.

Acoustic thermometers provide temperature measurement over a very wide range (−150°C to +20 000°C). However, their inaccuracy is relatively high (typically ±5%) and they are very expensive (typically £6000 or $10 000).

Colour indicators are widely used to determine when objects in furnaces have reached the required temperature. These indicators work well if the rate of rise of temperature of the object in the furnace is relatively slow but, because temperature indicators only change colour over a period of time, the object will be above the required temperature by the time that the indicator responds if the rate of rise of temperature is large. Cost is low, for example a crayon typically costs £3.

14.15 Self-test questions

14.1 The output e.m.f. from a chromel–alumel thermocouple (type K), with its reference junction maintained at 0°C, is 12.207 mV. What is the measured temperature?

14.2 The output e.m.f. from a nicrosil–nisil thermocouple (type N), with its reference junction maintained at 0°C, is 4.21 mV. What is the measured temperature?

14.3 A platinum/10% rhodium–platinum (type S) thermocouple is used to measure the temperature of a furnace. The output e.m.f., with the reference junction maintained at 50°C, is 5.975 mV. What is the temperature of the furnace?

14.4 In a particular industrial situation, a nicrosil–nisil thermocouple with nicrosil–nisil extension wires is used to measure the temperature of a fluid. In connecting up this measurement system, the instrumentation engineer responsible has inadvertently interchanged the extension wires from the thermocouple. The ends of the extension wires are held at a reference temperature of 0°C and the output e.m.f. measured is 21.0 mV. If the junction between the thermocouple and extension wires is at a temperature of 50°C, what temperature of fluid is indicated and what is the true fluid temperature?

14.5 A chromel–constantan thermocouple measuring the temperature of a fluid is connected by mistake with copper–constantan extension leads (such that the two constantan wires are connected together and the copper extension wire is connected to the chromel thermocouple wire). If the fluid temperature was actually 250°C, and the junction between the thermocouple and extension wires was at 80°C, what e.m.f. would be measured at the open ends of the extension wires if the reference junction is maintained at 0°C? What fluid temperature would be deduced from this (assuming that the connection mistake was not known about)? (Hint: apply the law of intermediate metals for the thermocouple-extension lead junction.)

References and further reading

Brookes, C. (1985) Nicrosil–nisil thermocouples, *Journal of Measurement and Control*, **18**(7), pp. 245–248.

Dixon, J. (1987) Industrial radiation thermometry, *Journal of Measurement and Control*, **20**(6), pp. 11–16.

Editorial (1996) Control Engineering, September, p. 93.

Johnson, J.S. (1994) Optical sensors: the OCSA experience, *Measurement and Control*, **27**(7), pp. 180–184.

Michalski, L., Eckersdorf, K. and McGhee, J. (1991) *Temperature Measurement*, John Wiley.

15

Pressure measurement

Pressure measurement is a very common requirement for most industrial process control systems and many different types of pressure-sensing and pressure-measurement systems are available. However, before considering these in detail, it is important to explain some terms used in pressure measurement and to define the difference between absolute pressure, gauge pressure and differential pressure.

Absolute pressure: This is the difference between the pressure of the fluid and the absolute zero of pressure.

Gauge pressure: This describes the difference between the pressure of a fluid and atmospheric pressure. Absolute and gauge pressure are therefore related by the expression:

$$\text{Absolute pressure} = \text{Gauge pressure} + \text{Atmospheric pressure}$$

Thus, gauge pressure varies as the atmospheric pressure changes and is therefore not a fixed quantity.

Differential pressure: This term is used to describe the difference between two absolute pressure values, such as the pressures at two different points within the same fluid (often between the two sides of a flow restrictor in a system measuring volume flow rate).

In most applications, the typical values of pressure measured range from 1.013 bar (the mean atmospheric pressure) up to 7000 bar. This is considered to be the 'normal' pressure range, and a large number of pressure sensors are available that can measure pressures in this range. Measurement requirements outside this range are much less common. Whilst some of the pressure sensors developed for the 'normal' range can also measure pressures that are either lower or higher than this, it is preferable to use special instruments that have been specially designed to satisfy such low- and high-pressure measurement requirements.

The discussion below summarizes the main types of pressure sensor that are in use. This discussion is primarily concerned only with the measurement of static pressure, because the measurement of dynamic pressure is a very specialized area that is not of general interest. In general, dynamic pressure measurement requires special instruments, although modified versions of diaphragm-type sensors can also be used if they contain a suitable displacement sensor (usually either a piezoelectric crystal or a capacitive element).

15.1 Diaphragms

The diaphragm, shown schematically in Figure 15.1, is one of three types of elastic-element pressure transducer. Applied pressure causes displacement of the diaphragm and this movement is measured by a displacement transducer. Different versions of diaphragm sensors can measure both absolute pressure (up to 50 bar) and gauge pressure (up to 2000 bar) according to whether the space on one side of the diaphragm is respectively evacuated or is open to the atmosphere. A diaphragm can also be used to measure differential pressure (up to 2.5 bar) by applying the two pressures to the two sides of the diaphragm. The diaphragm can be either plastic, metal alloy, stainless steel or ceramic. Plastic diaphragms are cheapest, but metal diaphragms give better accuracy. Stainless steel is normally used in high temperature or corrosive environments. Ceramic diaphragms are resistant even to strong acids and alkalis, and are used when the operating environment is particularly harsh.

The typical magnitude of diaphragm displacement is 0.1 mm, which is well suited to a strain-gauge type of displacement-measuring transducer, although other forms of displacement measurement are also used in some kinds of diaphragm-based sensors. If the displacement is measured with strain gauges, it is normal to use four strain gauges arranged in a bridge circuit configuration. The output voltage from the bridge is a function of the resistance change due to the strain in the diaphragm. This arrangement automatically provides compensation for environmental temperature changes. Older pressure transducers of this type used metallic strain gauges bonded to a diaphragm typically made of stainless steel. However, apart from manufacturing difficulties arising from the problem of bonding the gauges, metallic strain gauges have a low gauge factor, which means that the low output from the strain gauge bridge has to be amplified by an expensive d.c. amplifier. The development of semiconductor (piezoresistive) strain gauges provided a solution to the low-output problem, as they have gauge factors up

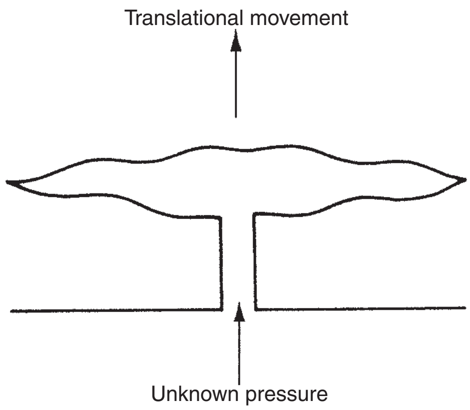


Fig. 15.1 Schematic representation of diaphragm pressure sensor.

to one hundred times greater than metallic gauges. However, the difficulty of bonding gauges to the diaphragm remained and a new problem emerged regarding the highly non-linear characteristic of the strain–output relationship.

The problem of strain-gauge bonding was solved with the emergence of monolithic piezoresistive pressure transducers. These have a typical measurement uncertainty of ±0.5% and are now the most commonly used type of diaphragm pressure transducer. The monolithic cell consists of a diaphragm made of a silicon sheet into which resistors are diffused during the manufacturing process. Such pressure transducers can be made to be very small and are often known as *micro-sensors*. Also, besides avoiding the difficulty with bonding, such monolithic silicon measuring cells have the advantage of being very cheap to manufacture in large quantities. Although the inconvenience of a non-linear characteristic remains, this is normally overcome by processing the output signal with an active linearization circuit or incorporating the cell into a microprocessor-based intelligent measuring transducer. The latter usually provides analogue-to-digital conversion and interrupt facilities within a single chip and gives a digital output that is readily integrated into computer control schemes. Such instruments can also offer automatic temperature compensation, built-in diagnostics and simple calibration procedures. These features allow measurement inaccuracy to be reduced to a figure as low as ±0.1% of full-scale reading.

15.2 Capacitive pressure sensor

A capacitive pressure sensor is simply a diaphragm-type device in which the diaphragm displacement is determined by measuring the capacitance change between the diaphragm and a metal plate that is close to it. Such devices are in common use. It is also possible to fabricate capacitive elements in a silicon chip and thus form very small *micro-sensors*. These have a typical measurement uncertainty of ±0.2%.

15.3 Fibre-optic pressure sensors

Fibre-optic sensors provide an alternative method of measuring displacements in diaphragm and Bourdon tube pressure sensors by optoelectronic means, and enable the resulting sensors to have lower mass and size compared with sensors in which the displacement is measured by other methods. The shutter sensor described earlier in Chapter 13 is one form of fibre-optic displacement sensor. Another form is the Fotonic sensor shown in Figure 15.2 in which light travels from a light source, down an optical fibre, is reflected back from a diaphragm, and then travels back along a second fibre to a photodetector. There is a characteristic relationship between the light reflected and the distance from the fibre ends to the diaphragm, thus making the amount of reflected light dependent upon the diaphragm displacement and hence the measured pressure.

Apart from the mass and size advantages of fibre-optic displacement sensors, the output signal is immune to electromagnetic noise. However, the measurement accuracy is usually inferior to that provided by alternative displacement sensors, and choice of such sensors also incurs a cost penalty. Thus, sensors using fibre optics to measure diaphragm or Bourdon tube displacement tend to be limited to applications where

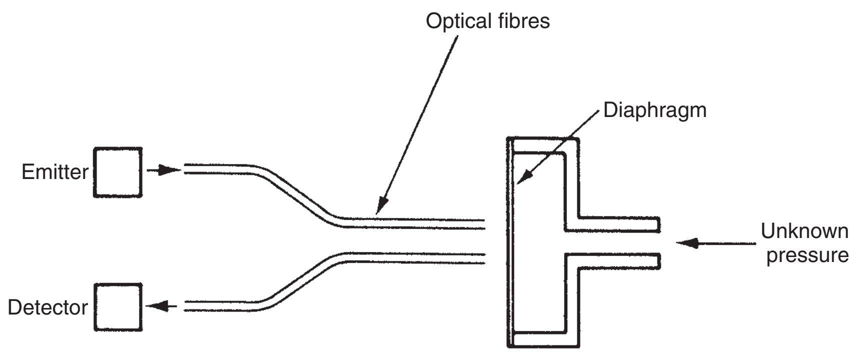


Fig. 15.2 Fotonic sensor.

their small size, low mass and immunity to electromagnetic noise are particularly advantageous.

Apart from the limited use above within diaphragm and Bourdon tube sensors, fibre-optic cables are also used in several other ways to measure pressure. A form of fibre-optic pressure sensor known as a *microbend sensor* is sketched in Figure 13.7(a). In this, the refractive index of the fibre (and hence of the intensity of light transmitted) varies according to the mechanical deformation of the fibre caused by pressure. The sensitivity of pressure measurement can be optimized by applying the pressure via a roller chain such that the bending is applied periodically (see Figure 13.7(b)). The optimal pitch for the chain varies according to the radius, refractive index and type of cable involved. Microbend sensors are typically used to measure the small pressure changes generated in vortex shedding flowmeters. When fibre-optic sensors are used in this flow-measurement role, the alternative arrangement shown in Figure 15.3 can be used, where a fibre-optic cable is merely stretched across the pipe. This often simplifies the detection of vortices.

Phase-modulating fibre-optic pressure sensors also exist. The mode of operation of these was discussed in Chapter 13.

15.4 Bellows

The bellows, schematically illustrated in Figure 15.4, is another elastic-element type of pressure sensor that operates on very similar principles to the diaphragm pressure sensor. Pressure changes within the bellows, which is typically fabricated as a seamless tube of either metal or metal alloy, produce translational motion of the end of the bellows that can be measured by capacitive, inductive (LVDT) or potentiometric transducers. Different versions can measure either absolute pressure (up to 2.5 bar) or gauge pressure (up to 150 bar). Double-bellows versions also exist that are designed to measure differential pressures of up to 30 bar.

Bellows have a typical measurement uncertainty of only ±0.5%, but they have a relatively high manufacturing cost and are prone to failure. Their principal attribute in the past has been their greater measurement sensitivity compared with diaphragm sensors. However, advances in electronics mean that the high-sensitivity requirement

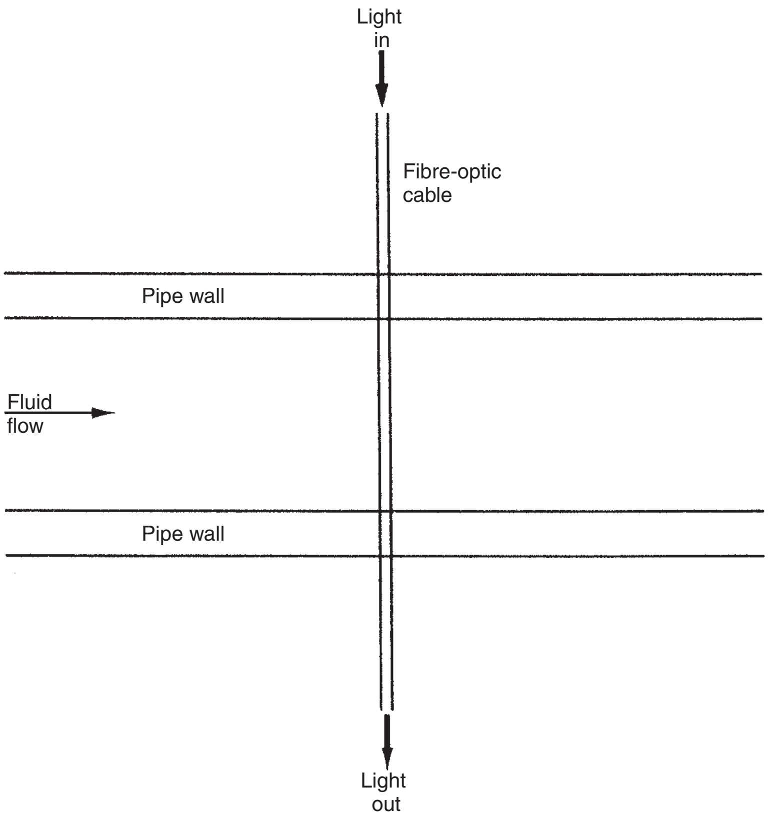


Fig. 15.3 Simple fibre-optic vortex detector.

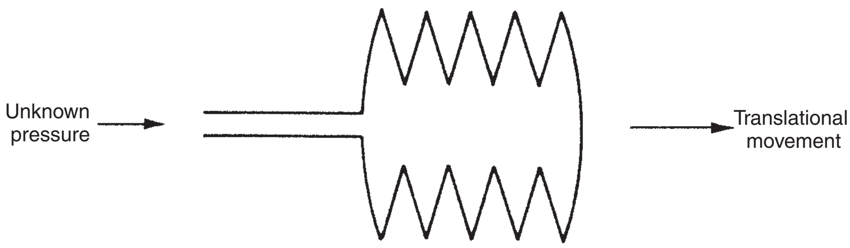


Fig. 15.4 Bellows.

can usually be satisfied now by diaphragm-type devices, and usage of bellows is therefore falling.

15.5 Bourdon tube

The Bourdon tube is also an elastic element type of pressure transducer. It is relatively cheap and is commonly used for measuring the gauge pressure of both gaseous and liquid fluids. It consists of a specially shaped piece of oval-section, flexible, metal tube that is fixed at one end and free to move at the other end. When pressure is applied at the open, fixed end of the tube, the oval cross-section becomes more circular. In consequence, there is a displacement of the free end of the tube. This displacement is measured by some form of displacement transducer, which is commonly a potentiometer or LVDT. Capacitive and optical sensors are also sometimes used to measure the displacement.

The three common shapes of Bourdon tube are shown in Figure 15.5. The maximum possible deflection of the free end of the tube is proportional to the angle subtended by the arc through which the tube is bent. For a C-type tube, the maximum value for this arc is somewhat less than 360°. Where greater measurement sensitivity and resolution are required, spiral and helical tubes are used. These both give a much greater deflection at the free end for a given applied pressure. However, this increased measurement performance is only gained at the expense of a substantial increase in manufacturing difficulty and cost compared with C-type tubes, and is also associated with a large decrease in the maximum pressure that can be measured. Spiral and helical types are sometimes provided with a rotating pointer that moves against a scale to give a visual indication of the measured pressure.

C-type tubes are available for measuring pressures up to 6000 bar. A typical C-type tube of 25 mm radius has a maximum displacement travel of 4 mm, giving a moderate level of measurement resolution. Measurement inaccuracy is typically quoted at $\pm 1\%$ of full-scale deflection. Similar accuracy is available from helical and spiral types, but whilst the measurement resolution is higher, the maximum pressure measurable is only 700 bar.

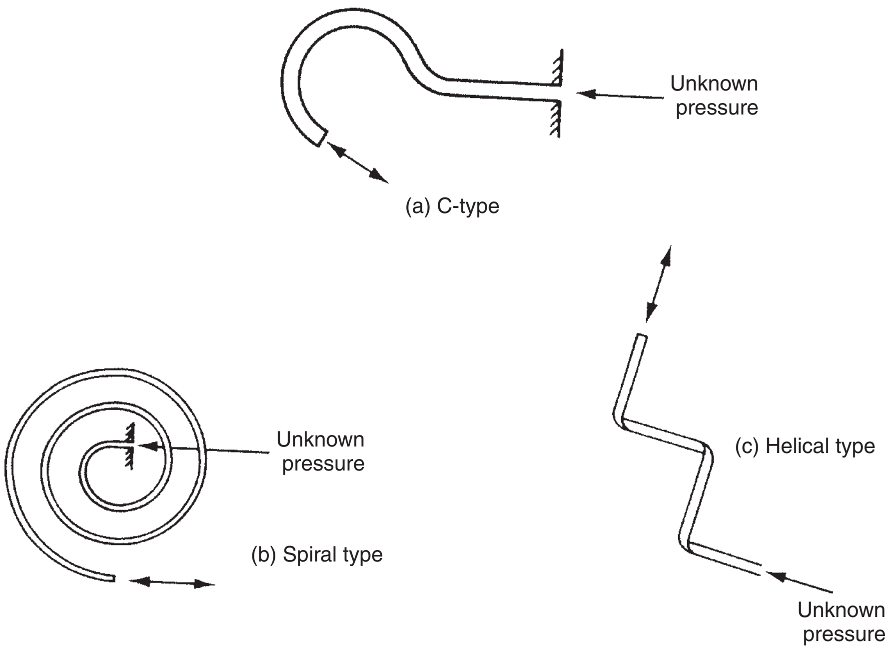


Fig. 15.5 Bourdon tubes.

The existence of one potentially major source of error in Bourdon tube pressure measurement has not been widely documented, and few manufacturers of Bourdon tubes make any attempt to warn users of their products appropriately. The problem is concerned with the relationship between the fluid being measured and the fluid used for calibration. The pointer of Bourdon tubes is normally set at zero during manufacture, using air as the calibration medium. However, if a different fluid, especially a liquid, is subsequently used with a Bourdon tube, the fluid in the tube will cause a non-zero deflection according to its weight compared with air, resulting in a reading error of up to 6%. This can be avoided by calibrating the Bourdon tube with the fluid to be measured instead of with air, assuming of course that the user is aware of the problem. Alternatively, correction can be made according to the calculated weight of the fluid in the tube. Unfortunately, difficulties arise with both of these solutions if air is trapped in the tube, since this will prevent the tube being filled completely by the fluid. Then, the amount of fluid actually in the tube, and its weight, will be unknown.

In conclusion, therefore, Bourdon tubes only have guaranteed accuracy limits when measuring gaseous pressures. Their use for accurate measurement of liquid pressures poses great difficulty unless the gauge can be totally filled with liquid during both calibration and measurement, a condition that is very difficult to fulfil practically.

15.6 Manometers

Manometers are passive instruments that give a visual indication of pressure values. Various types exist.

The *U-tube manometer*, shown in Figure 15.6(a), is the most common form of manometer. Applied pressure causes a displacement of liquid inside the U-shaped glass tube, and the output pressure reading P is made by observing the difference h between the level of liquid in the two halves of the tube A and B, according to the equation $P = h\rho g$, where ρ is the specific gravity of the fluid. If an unknown pressure is applied to side A, and side B is open to the atmosphere, the output reading is gauge pressure. Alternatively, if side B of the tube is sealed and evacuated, the output reading is absolute pressure. The U-tube manometer also measures the differential pressure $(p_1 - p_2)$, according to the expression $(p_1 - p_2) = h\rho g$, if two unknown pressures p_1 and p_2 are applied respectively to sides A and B of the tube.

Output readings from U-tube manometers are subject to error, principally because it is very difficult to judge exactly where the meniscus levels of the liquid are in the two halves of the tube. In absolute pressure measurement, an addition error occurs because it is impossible to totally evacuate the closed end of the tube.

U-tube manometers are typically used to measure gauge and differential pressures up to about 2 bar. The type of liquid used in the instrument depends on the pressure and characteristics of the fluid being measured. Water is a cheap and convenient choice, but it evaporates easily and is difficult to see. Nevertheless, it is used extensively, with the major obstacles to its use being overcome by using coloured water and by regularly topping up the tube to counteract evaporation. However, water is definitely not used when measuring the pressure of fluids that react with or dissolve in water. Water is also unsuitable when high-pressure measurements are required. In such circumstances, liquids such as aniline, carbon tetrachloride, bromoform, mercury or transformer oil are used instead.

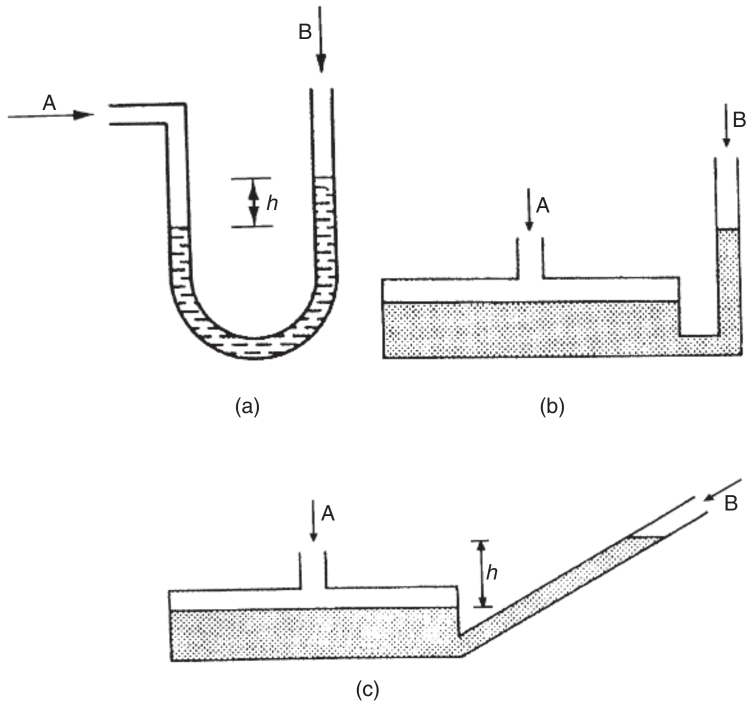


Fig. 15.6 Manometers: (a) U-tube; (b) well type; (c) inclined type.

The *well-type or cistern manometer*, shown in Figure 15.6(b), is similar to a U-tube manometer but one half of the tube is made very large so that it forms a well. The change in the level of the well as the measured pressure varies is negligible. Therefore, the liquid level in only one tube has to be measured, which makes the instrument much easier to use than the U-tube manometer. If an unknown pressure p_1 is applied to port A, and port B is open to the atmosphere, the gauge pressure is given by $p_1 = h\rho$. It might appear that the instrument would give a better measurement accuracy than the U-tube manometer because the need to subtract two liquid level measurements in order to arrive at the pressure value is avoided. However, this benefit is swamped by errors that arise due to the typical cross-sectional area variations in the glass used to make the tube. Such variations do not affect the accuracy of the U-tube manometer to the same extent.

The *inclined manometer* or *draft gauge*, shown in Figure 15.6(c), is a variation on the well-type manometer in which one leg of the tube is inclined to increase measurement sensitivity. However, similar comments to those above apply about accuracy.

15.7 Resonant-wire devices

A typical resonant-wire device is shown schematically in Figure 15.7. Wire is stretched across a chamber containing fluid at unknown pressure subjected to a magnetic field.

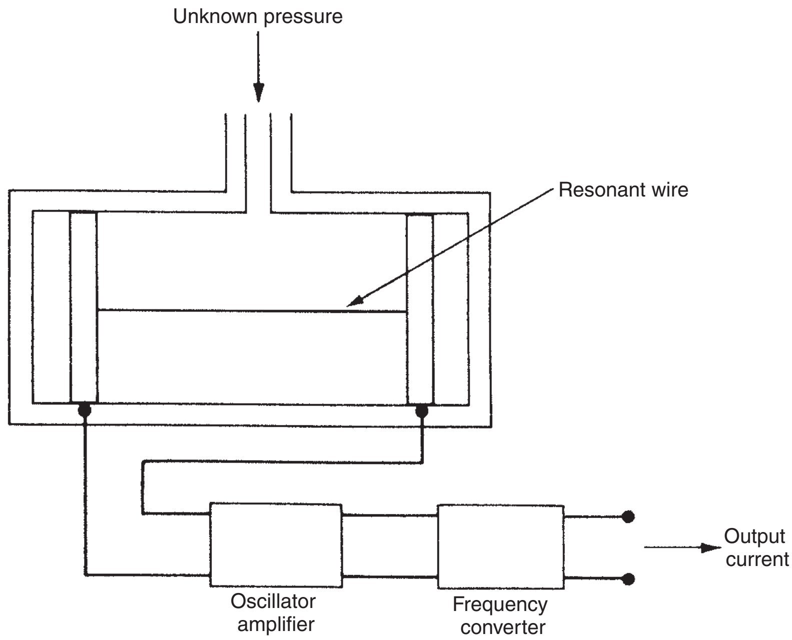


Fig. 15.7 Resonant-wire device.

The wire resonates at its natural frequency according to its tension, which varies with pressure. Thus pressure is calculated by measuring the frequency of vibration of the wire. Such frequency measurement is normally carried out by electronics integrated into the cell. These devices are highly accurate, with a typical inaccuracy figure being $\pm 0.2\%$ full-scale reading. They are also particularly insensitive to ambient condition changes and can measure pressures between 5 mbar and 2 bar.

15.8 Dead-weight gauge

The dead-weight gauge, as shown in Figure 2.3, is a null-reading type of measuring instrument in which weights are added to the piston platform until the piston is adjacent to a fixed reference mark, at which time the downward force of the weights on top of the piston is balanced by the pressure exerted by the fluid beneath the piston. The fluid pressure is therefore calculated in terms of the weight added to the platform and the known area of the piston. The instrument offers the ability to measure pressures to a high degree of accuracy but is inconvenient to use. Its major application is as a reference instrument against which other pressure-measuring devices are calibrated. Various versions are available that allow measurement of gauge pressures up to 7000 bar.

15.9 Special measurement devices for low pressures

A number of special devices have been developed for measurement of pressures in the vacuum range below atmospheric pressure (<1.013 bar). These special devices include the thermocouple gauge, the Pirani gauge, the thermistor gauge, the McLeod gauge and the ionization gauge, and they are covered in more detail below. Unfortunately, all of these specialized instruments are quite expensive.

The *thermocouple gauge* is one of a group of gauges working on the thermal conductivity principal. The paranoia and thermistor gauges also belong to this group. At low pressure, the kinematic theory of gases predicts a linear relationship between pressure and thermal conductivity. Thus measurement of thermal conductivity gives an indication of pressure. Figure 15.8 shows a sketch of a thermocouple gauge. Operation of the gauge depends on the thermal conduction of heat between a thin hot metal strip in the centre and the cold outer surface of a glass tube (that is normally at room temperature). The metal strip is heated by passing a current through it and its temperature is measured by a thermocouple. The temperature measured depends on the thermal conductivity of the gas in the tube and hence on its pressure. A source of error in this instrument is the fact that heat is also transferred by radiation as well as conduction. This error is of a constant magnitude, independent of pressure. Hence, it can be measured, and thus correction can be made for it. However, it is usually more convenient to design for low radiation loss by choosing a heated element with low emissivity. Thermocouple gauges are typically used to measure pressures in the range 10^{-4} mbar up to 1 mbar.

A typical form of *Pirani gauge* is shown in Figure 15.9(a). This is similar to a thermocouple gauge but has a heated element that consists of four coiled tungsten wires connected in parallel. Two identical tubes are normally used, connected in a bridge circuit as shown in Figure 15.9(b), with one containing the gas at unknown pressure and the other evacuated to a very low pressure. Current is passed through the tungsten element, which attains a certain temperature according to the thermal conductivity of the gas. The resistance of the element changes with temperature and causes an imbalance of the measurement bridge. Thus, the Pirani gauge avoids the use

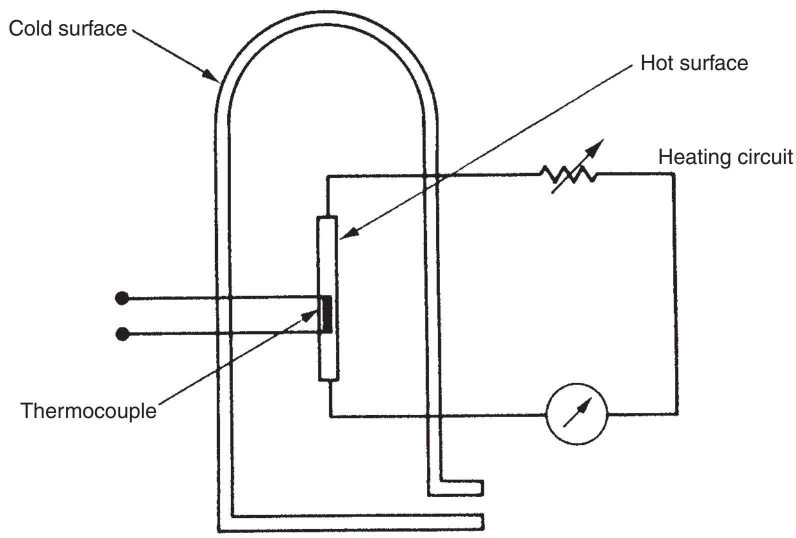


Fig. 15.8 Thermocouple gauge.

of a thermocouple to measure temperature (as in the thermocouple gauge) by effectively using a resistance thermometer as the heated element. Such gauges cover the pressure range 10^{-5} mbar to 1 mbar.

The *thermistor gauge* operates on identical principles to the Pirani gauge but uses semiconductor materials for the heated elements instead of metals. The normal pressure range covered is 10^{-4} mbar to 1 mbar.

Figure 15.10(a) shows the general form of a *McLeod gauge*, in which low-pressure fluid is compressed to a higher pressure that is then read by manometer techniques. In

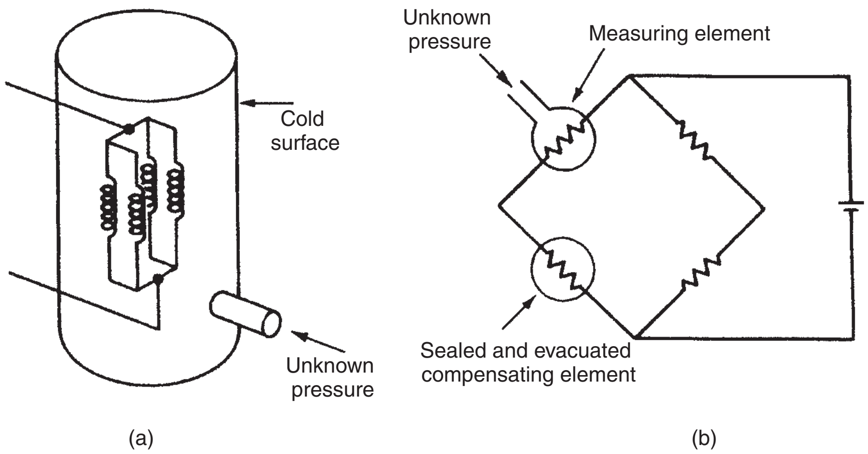


Fig. 15.9 (a) Pirani gauge; (b) Wheatstone bridge circuit to measure output.

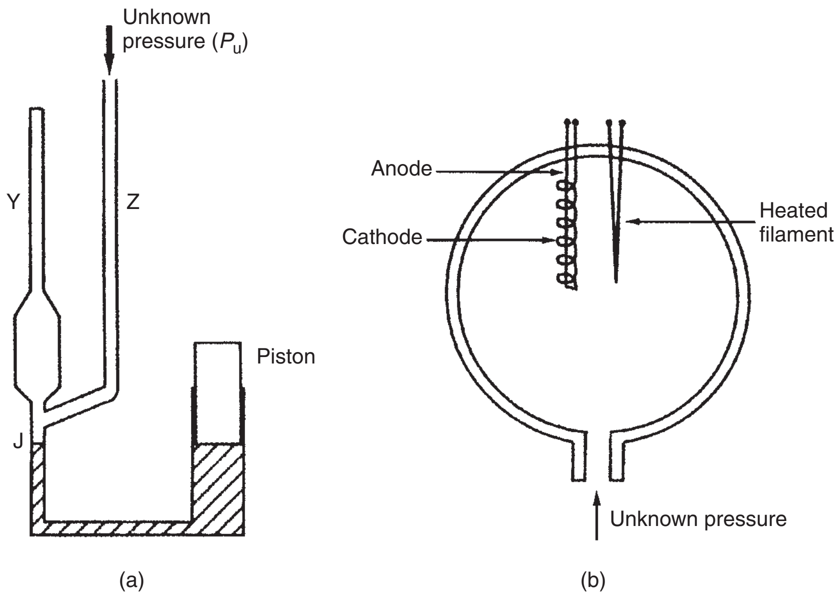


Fig. 15.10 Other low-pressure gauges: (a) McLeod gauge; (b) ionization gauge.

essence, the gauge can be visualized as a U-tube manometer that is sealed at one end, and where the bottom of the U can be blocked at will. To operate the gauge, the piston is first withdrawn. This causes the level of mercury in the lower part of the gauge to fall below the level of the junction J between the two tubes marked Y and Z in the gauge. Fluid at unknown pressure P_u is then introduced via the tube marked Z, from where it also flows into the tube of cross-sectional area A marked Y. Next, the piston is pushed in, moving the mercury level up to block the junction J. At the stage where J is just blocked, the fluid in tube Y is at pressure P_u and is contained in a known volume V_u. Further movement of the piston compresses the fluid in tube Y and this process continues until the mercury level in tube Z reaches a zero mark. Measurement of the height (h) above the mercury column in tube Y then allows calculation of the compressed volume of the fluid V_c as $V_c = hA$.
Then, by Boyle's law:

$$P_u V_u = P_c V_c$$

Also, applying the normal manometer equation:

$$P_c = P_u + h\rho g$$

where ρ is the mass density of mercury, the pressure P_u can be calculated as:

$$P_u = \frac{Ah^2 \rho g}{V_u - Ah} \tag{15.1}$$

The compressed volume V_c is often very much smaller than the original volume, in which case equation (15.1) approximates to:

$$P_u = \frac{Ah^2 \rho g}{V_u} \quad \text{for} \quad Ah \ll V_u \tag{15.2}$$

Although the smallest inaccuracy achievable with McLeod gauges is $\pm 1\%$, this is still better than that which is achievable with most other gauges that are available for measuring pressures in this range. Therefore, the McLeod gauge is often used as a standard against which other gauges are calibrated. The minimum pressure normally measurable is 10^{-4} bar, although lower pressures can be measured if pressure-dividing techniques are applied.

The *ionization gauge* is a special type of instrument used for measuring very low pressures in the range 10^{-13} to 10^{-3} bar. Gas of unknown pressure is introduced into a glass vessel containing free electrons discharged from a heated filament, as shown in Figure 15.10(b). Gas pressure is determined by measuring the current flowing between an anode and cathode within the vessel. This current is proportional to the number of ions per unit volume, which in turn is proportional to the gas pressure. Ionization gauges are normally only used in laboratory conditions.

15.10 High-pressure measurement (greater than 7000 bar)

Measurement of pressures above 7000 bar is normally carried out electrically by monitoring the change of resistance of wires of special materials. Materials having

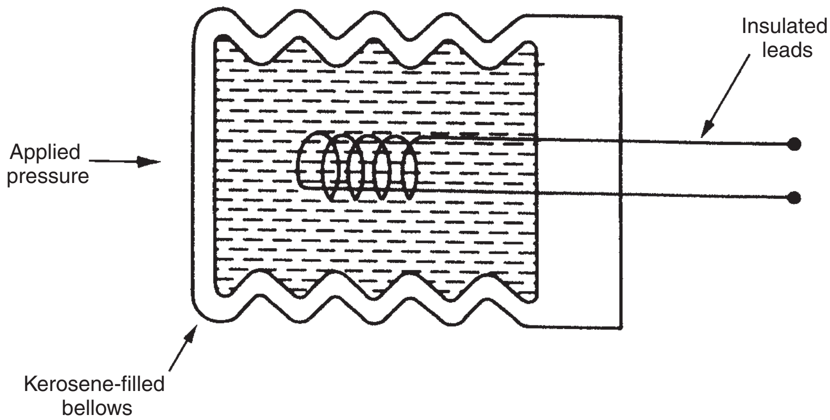


Fig. 15.11 High-pressure measurement–wire coil in bellows.

resistance-pressure characteristics that are suitably linear and sensitive include manganin and gold–chromium alloys. A coil of such wire is enclosed in a sealed, kerosene filled, flexible bellows, as shown in Figure 15.11. The unknown pressure is applied to one end of the bellows, which transmits the pressure to the coil. The magnitude of the applied pressure is then determined by measuring the coil resistance. Pressures up to 30 000 bar can be measured by devices like the manganin-wire pressure sensor, with a typical inaccuracy of ±0.5%.

15.11 Intelligent pressure transducers

Adding microprocessor power to pressure transducers brings about substantial improvements in their characteristics. Measurement sensitivity improvement, extended measurement range, compensation for hysteresis and other non-linearities, and correction for ambient temperature and pressure changes are just some of the facilities offered by intelligent pressure transducers. For example, inaccuracy figures as low as ±0.1% can be achieved with silicon piezoresistive-bridge devices.

Inclusion of microprocessors has also enabled the use of novel techniques of displacement measurement, for example the optical method of displacement measurement shown in Figure 15.12. In this, the motion is transmitted to a vane that progressively shades one of two monolithic photodiodes that are exposed to infrared radiation. The second photodiode acts as a reference, enabling the microprocessor to compute a ratio signal that is linearized and is available as either an analogue or digital measurement of pressure. The typical measurement inaccuracy is ±0.1%. Versions of both diaphragms and Bourdon tubes that use this technique are available.

15.12 Selection of pressure sensors

Choice between the various types of instrument available for measuring mid-range pressures (1.013–7000 bar) is usually strongly influenced by the intended application.

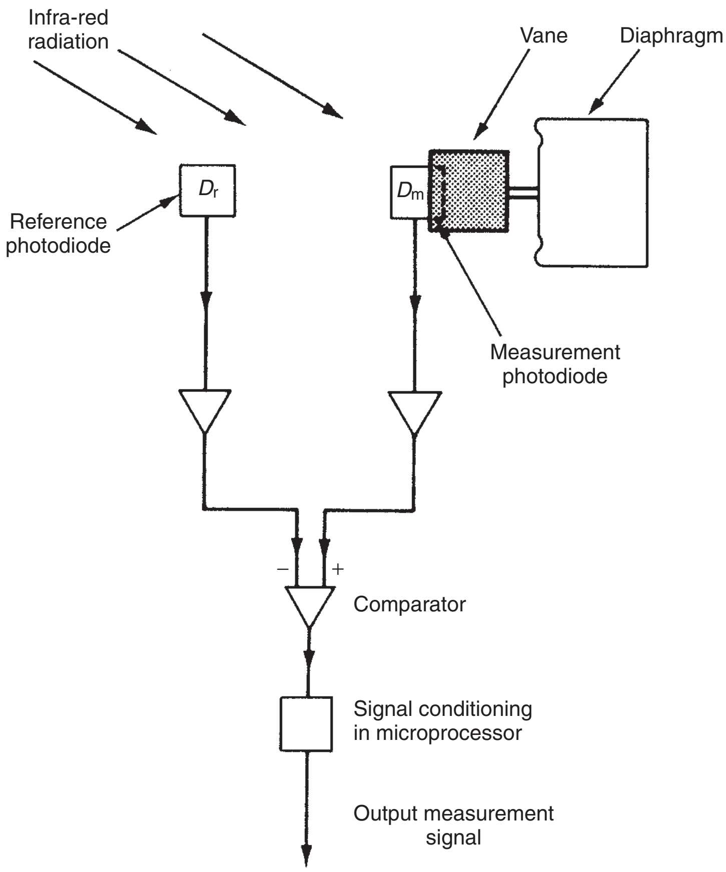


Fig. 15.12 Example of intelligent pressure-measuring instrument.

Manometers are commonly used when just a visual indication of pressure level is required, and deadweight gauges, because of their superior accuracy, are used in calibration procedures of other pressure-measuring devices. When an electrical form of output is required, the choice is usually either one of the several types of diaphragm sensor (strain gauge, capacitive or fibre optic) or, less commonly, a Bourdon tube. Bellows-type instruments are also sometimes used for this purpose, but much less frequently. If very high measurement accuracy is required, the resonant-wire device is a popular choice.

In the case of pressure measurement in the vacuum range (less than atmospheric pressure, i.e. below 1.013 bar), adaptations of most of the types of pressure transducer described earlier can be used. Special forms of Bourdon tubes measure pressures down to 10 mbar, manometers and bellows-type instruments measure pressures down to 0.1 mbar, and diaphragms can be designed to measure pressures down to 0.001 mbar. However, a number of more specialized instruments have also been developed to measure vacuum pressures, as discussed in section 15.9. These generally give better measurement accuracy and sensitivity compared with instruments that are primarily designed for measuring mid-range pressures. This improved accuracy is particularly evident at low pressures. Therefore, only the special instruments described in section 15.9 are used to measure pressures below 10^{-4} mbar.

At high pressures (>7000 bar), the only devices in common use are the manganin-wire sensor and similar devices based on alternative alloys to manganin.

For differential pressure measurement, diaphragm-type sensors are the preferred option, with double-bellows sensors being used occasionally. Manometers are also sometimes used to give visual indication of differential pressure values (especially in liquid flow-rate indicators). These are passive instruments that have the advantage of not needing a power supply.

16

Flow measurement

The rate at which fluid flows through a closed pipe can be quantified by either measuring the mass flow rate or measuring the volume flow rate. Of these alternatives, mass flow measurement is more accurate, since mass, unlike volume, is invariant. In the case of the flow of solids, the choice is simpler, since only mass flow measurement is appropriate.

16.1 Mass flow rate

The method used to measure mass flow rate is largely determined by whether the measured quantity is in a solid, liquid or gaseous state. The main techniques available are summarized below. A more comprehensive discussion can be found in Medlock (1990).

16.1.1 Conveyor-based methods

These methods are concerned with measurement of the flow of solids that are in the form of small particles. Such particles are usually produced by crushing or grinding procedures in process industries, and the particles are usually transported by some form of conveyor. This mode of transport allows the mass flow rate to be calculated in terms of the mass of material on a given length of conveyor multiplied by the speed of the conveyor. Figure 16.1 shows a typical measurement system. A load cell measures the mass M of material distributed over a length L of the conveyor. If the conveyor velocity is v, the mass flow rate, Q, is given by:

$$Q = Mv/L$$

As an alternative to weighing the flowing material, a *nuclear mass-flow sensor* can be used, in which a gamma-ray source is directed at the material being transported along the conveyor. The material absorbs some radiation, and the amount of radiation received by a detector on the other side of the material indicates the amount of material on the conveyor. This technique has obvious safety concerns, and is therefore subject to licensing and strict regulation.

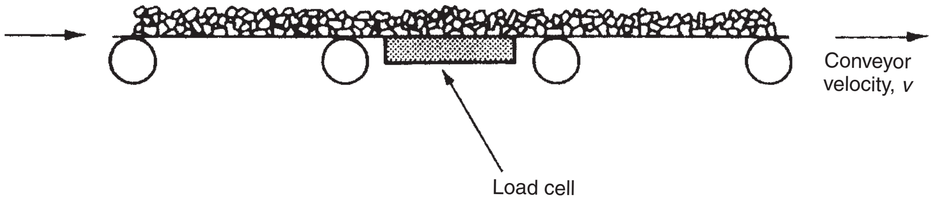


Fig. 16.1 Conveyor-based mass flow rate measurement.

16.1.2 Coriolis flowmeter

The Coriolis flowmeter is primarily used to measure the mass flow rate of liquids, although it has also been successfully used in some gas-flow measurement applications. The flowmeter consists of either a pair of parallel vibrating tubes or else a single vibrating tube that is formed into a configuration that has two parallel sections.

The two vibrating tubes (or the two parallel sections of a single tube) deflect according to the mass flow rate of the measured fluid that is flowing inside. Tubes are made of various materials, of which stainless steel is the most common. They are also manufactured in different shapes such as B-shaped, D-shaped, U-shaped, triangular-shaped, helix-shaped and straight. These alternative shapes are sketched in Figure 16.2(a) and a U-shaped tube is shown in more detail in Figure 16.2(b). The tubes are anchored at two points. An electromechanical drive unit, positioned midway between the two anchors, excites vibrations in each tube at the tube resonant frequency. The vibrations in the two tubes, or the two parallel sections of a single tube, are 180 degrees out of phase. The vibratory motion of each tube causes forces on the particles in the flowing fluid. These forces induce motion of the fluid particles in a direction that is orthogonal to the direction of flow, and this produces a Coriolis force. This Coriolis force causes a deflection of the tubes that is superimposed on top of the vibratory motion. The net deflection of one tube relative to the other is given by $d = kfR$, where k is a constant, f is the frequency of the tube vibration and R is the mass flow rate of the fluid inside the tube. This deflection is measured by a suitable sensor. A full account of the theory of operation can be found in Figliola (1995).

Coriolis meters give excellent accuracy, with measurement uncertainties of $\pm 0.2\%$ being typical. They also have low maintenance requirements. However, apart from being expensive (typical cost is £4000), they suffer from a number of operational problems. Failure may occur after a period of use because of mechanical fatigue in the tubes. Tubes are also subject to both corrosion caused by chemical interaction with the measured fluid and abrasion caused by particles within the fluid. Diversion of the flowing fluid around the flowmeter causes it to suffer a significant pressure drop, though this is much less evident in straight tube designs.

16.1.3 Thermal mass flow measurement

Thermal mass flowmeters are primarily used to measure the flow rate of gases. The principle of operation is to direct the flowing material past a heated element. The mass flow rate is inferred in one of two ways, (a) by measuring the temperature rise in the

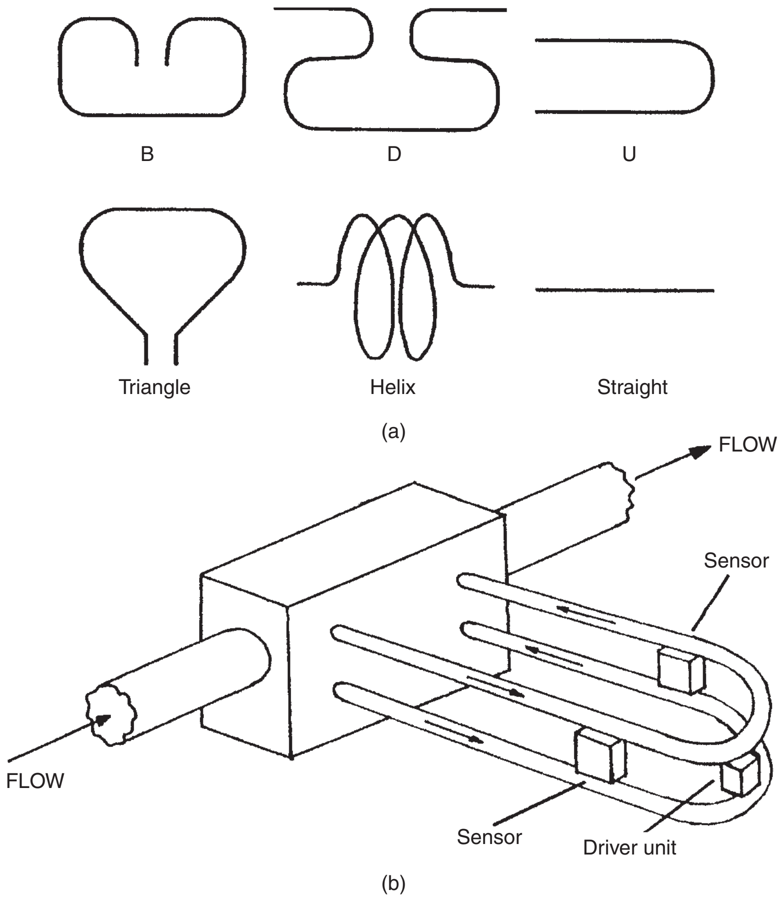


Fig. 16.2 (a) Coriolis flowmeter shapes; (b) detail of U-shaped Coriolis flowmeter.

flowing material or (b) by measuring the heater power required to achieve a constant set temperature in the flowing material. Typical measurement uncertainty is ±2%.

16.1.4 Joint measurement of volume flow rate and fluid density

Before the advent of the Coriolis meter, the usual way of measuring mass flow rate was to compute this from separate, simultaneous measurements of the volume flow rate and the fluid density. In many circumstances, this is still the cheapest option, although measurement accuracy is substantially inferior to that provided by a Coriolis meter.

16.2 Volume flow rate

Volume flow rate is an appropriate way of quantifying the flow of all materials that are in a gaseous, liquid or semi-liquid slurry form (where solid particles are suspended in a liquid host), although measurement accuracy is inferior to mass flow measurement as noted earlier. Materials in these forms are carried in pipes, and various instruments can be used to measure the volume flow rate as described below.

16.2.1 Differential pressure (obstruction-type) meters

Differential pressure meters involve the insertion of some device into a fluid-carrying pipe that causes an obstruction and creates a pressure difference on either side of the device. Such meters are sometimes known as obstruction-type meters or flow-restriction meters. Devices used to obstruct the flow include the *orifice plate*, the *Venturi tube*, the *flow nozzle* and the *Dall flow tube*, as illustrated in Figure 16.3. When such a restriction is placed in a pipe, the velocity of the fluid through the restriction increases and the pressure decreases. The volume flow rate is then proportional to the square root of the pressure difference across the obstruction. The manner in which this pressure difference is measured is important. Measuring the two pressures with different instruments and calculating the difference between the two measurements is not satisfactory because of the large measurement error which can arise when the pressure difference is small, as explained in Chapter 3. Therefore, the normal procedure is to use a differential pressure transducer, which is commonly a diaphragm type.

The *Pitot static tube* is a further device that measures flow by creating a pressure difference within a fluid-carrying pipe. However, in this case, there is negligible obstruction of flow in the pipe. The Pitot tube is a very thin tube that obstructs only a small part of the flowing fluid and thus measures flow at a single point across the cross-section of the pipe. This measurement only equates to average flow velocity in the pipe for the case of uniform flow. The *Annubar* is a type of multi-port Pitot tube that does measure the average flow across the cross-section of the pipe by forming the mean value of several local flow measurements across the cross-section of the pipe.

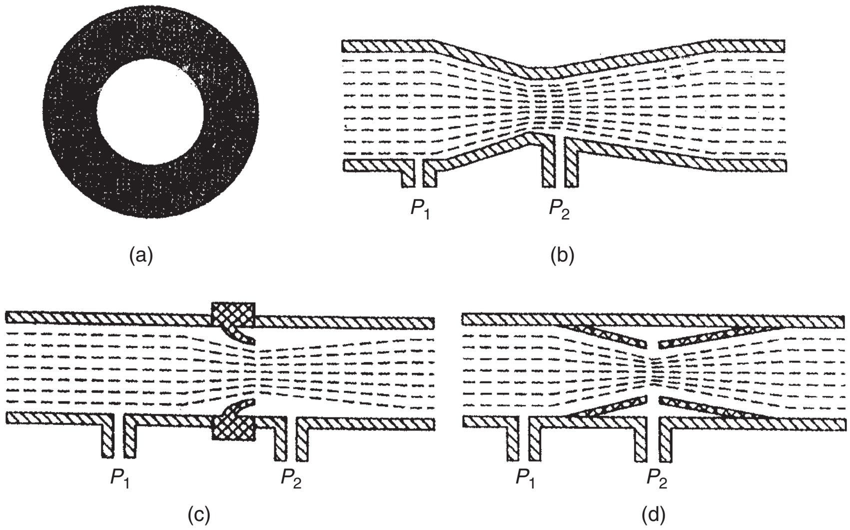


Fig. 16.3 Obstuction devices: (a) orifice plate; (b) venturi; (c) flow nozzle; (d) Dall flow tube.

All applications of this method of flow measurement assume that flow conditions upstream of the obstruction device are in steady state, and a certain minimum length of straight run of pipe ahead of the flow measurement point is specified to ensure this. The minimum lengths required for various pipe diameters are specified in British Standards tables (and also in alternative but equivalent national standards used in other countries), but a useful rule of thumb widely used in the process industries is to specify a length of ten times the pipe diameter. If physical restrictions make this impossible to achieve, special flow smoothing vanes can be inserted immediately ahead of the measurement point.

Flow-restriction type instruments are popular because they have no moving parts and are therefore robust, reliable and easy to maintain. One disadvantage of this method is that the obstruction causes a permanent loss of pressure in the flowing fluid. The magnitude and hence importance of this loss depends on the type of obstruction element used, but where the pressure loss is large, it is sometimes necessary to recover the lost pressure by an auxiliary pump further down the flow line. This class of device is not normally suitable for measuring the flow of slurries as the tappings into the pipe to measure the differential pressure are prone to blockage, although the Venturi tube can be used to measure the flow of dilute slurries.

Figure 16.4 illustrates approximately the way in which the flow pattern is interrupted when an orifice plate is inserted into a pipe. The other obstruction devices also have a similar effect to this. Of particular interest is the fact that the minimum cross-sectional area of flow occurs not within the obstruction but at a point downstream of there. Knowledge of the pattern of pressure variation along the pipe, as shown in Figure 16.5, is also of importance in using this technique of volume flow rate measurement. This shows that the point of minimum pressure coincides with the point of minimum cross-section flow, a little way downstream of the obstruction. Figure 16.5 also shows that there is a small rise in pressure immediately before the obstruction. It is therefore important not only to position the instrument measuring P_2 exactly at the point of minimum pressure, but also to measure the pressure P_1 at a point upstream of the point where the pressure starts to rise before the obstruction.

In the absence of any heat transfer mechanisms, and assuming frictionless flow of an incompressible fluid through the pipe, the theoretical volume flow rate of the fluid,

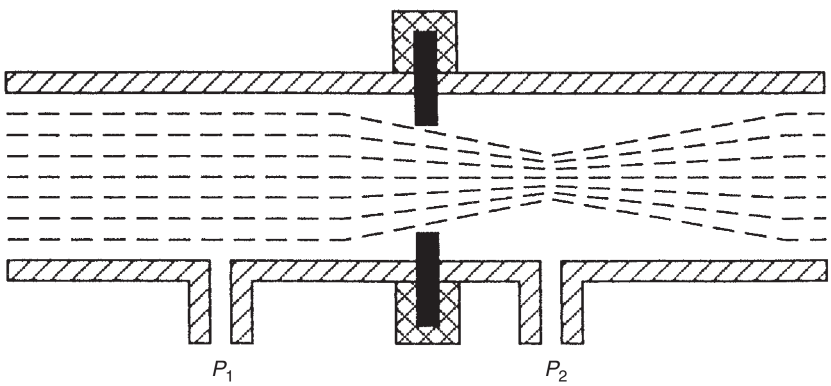


Fig. 16.4 Profile of flow across orifice plate.

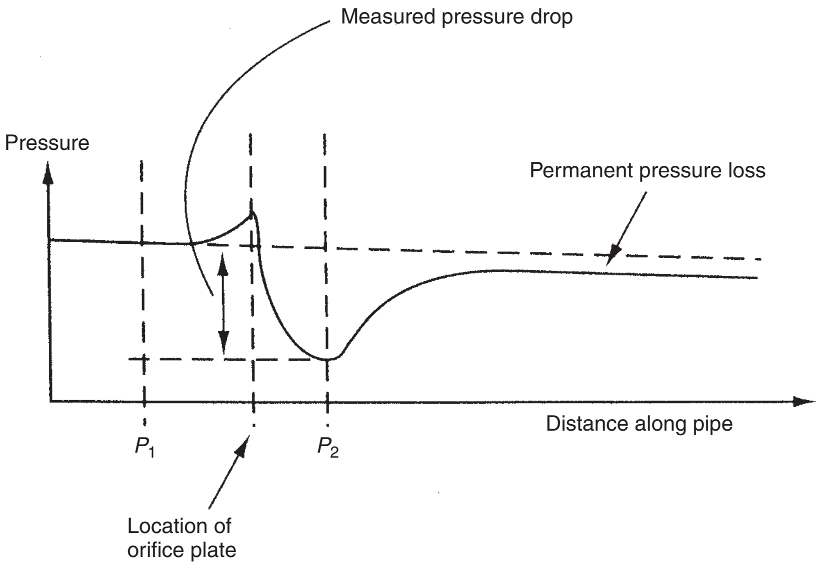


Fig. 16.5 Pattern of pressure variation either side of orifice plate.

Q, is given by:

$$Q = \left[\frac{A_2}{\sqrt{1 - (A_2/A_1)^2}} \right] \left[\sqrt{\frac{2(P_1 - P_2)}{\rho}} \right] \tag{16.1}$$

where A_1 and P_1 are the cross-sectional area and pressure of the fluid flow before the obstruction, A_2 and P_2 are the cross-sectional area and pressure of the fluid flow at the narrowest point of the flow beyond the obstruction, and ρ is the fluid density.

Equation (16.1) is never applicable in practice for several reasons. Firstly, frictionless flow is never achieved. However, in the case of turbulent flow through smooth pipes, friction is low and it can be adequately accounted for by a variable called the Reynolds number, which is a measurable function of the flow velocity and the viscous friction. The other reasons for the nonapplicability of equation (16.1) are that the initial cross-sectional area of the fluid flow is less than the diameter of the pipe carrying it and that the minimum cross-sectional area of the fluid is less than the diameter of the obstruction. Therefore, neither A_1 nor A_2 can be measured. These problems are taken account of by modifying equation (16.1) to the following:

$$Q = \left[\frac{C_D A'_2}{\sqrt{1 - (A'_2/A'_1)^2}} \right] \left[\sqrt{\frac{2(P_1 - P_2)}{\rho}} \right] \tag{16.2}$$

where A'_1 and A'_2 are the pipe diameters before and at the obstruction and C_D is a constant, known as the discharge coefficient, which accounts for the Reynolds number and the difference between the pipe and flow diameters.

Before equation (16.2) can be evaluated, the discharge coefficient must be calculated. As this varies between each measurement situation, it would appear at first sight that the discharge coefficient must be determined by practical experimentation in each case. However, provided that certain conditions are met, standard tables can be used to obtain the value of the discharge coefficient appropriate to the pipe diameter and fluid involved.

One particular problem with all flow restriction devices is that the pressure drop $(P_1 - P_2)$ varies as the square of the flow rate Q according to equation (16.2). The difficulty of measuring small pressure differences accurately has already been noted earlier. In consequence, the technique is only suitable for measuring flow rates that are between 30% and 100% of the maximum flow rate that a given device can handle. This means that alternative flow measurement techniques have to be used in applications where the flow rate can vary over a large range that can drop to below 30% of the maximum rate.

Orifice plate

The orifice plate is a metal disc with a concentric hole in it, which is inserted into the pipe carrying the flowing fluid. Orifice plates are simple, cheap and available in a wide range of sizes. In consequence, they account for almost 50% of the instruments used in industry for measuring volume flow rate. One limitation of the orifice plate is that its inaccuracy is typically at least $\pm 2\%$ and may approach $\pm 5\%$. Also, the permanent pressure loss caused in the measured fluid flow is between 50% and 90% of the magnitude of the pressure difference $(P_1 - P_2)$. Other problems with the orifice plate are a gradual change in the discharge coefficient over a period of time as the sharp edges of the hole wear away, and a tendency for any particles in the flowing fluid to stick behind the hole and thereby gradually reduce its diameter as the particles build up. The latter problem can be minimized by using an orifice plate with an eccentric hole. If this hole is close to the bottom of the pipe, solids in the flowing fluid tend to be swept through, and build-up of particles behind the plate is minimized. A very similar problem arises if there are any bubbles of vapour or gas in the flowing fluid when liquid flow is involved. These also tend to build up behind an orifice plate and distort the pattern of flow. This difficulty can be avoided by mounting the orifice plate in a vertical run of pipe.

Venturis and similar devices

A number of obstruction devices are available that are specially designed to minimize the pressure loss in the measured fluid. These have various names such as Venturi, flow nozzle and Dall flow tube. They are all much more expensive than an orifice plate but have better performance. The smooth internal shape means that they are not prone to solid particles or bubbles of gas sticking in the obstruction, as is likely to happen in an orifice plate. The smooth shape also means that they suffer much less wear, and consequently have a longer life than orifice plates. They also require less maintenance and give greater measurement accuracy.

The **Venturi** has a precision-engineered tube of a special shape. This offers measurement uncertainty of only $\pm 1\%$. However, the complex machining required to manufacture it means that it is the most expensive of all the obstruction devices discussed. Permanent pressure loss in the measured system is 10–15% of the pressure difference $(P_1 - P_2)$ across it.

The **Dall flow tube** consists of two conical reducers inserted into the fluid-carrying pipe. It has a very similar internal shape to the Venturi, except that it lacks a throat. This construction is much easier to manufacture and this gives the Dall flow tube an advantage in cost over the Venturi, although the typical measurement inaccuracy is a little higher ($\pm 1.5\%$). Another advantage of the Dall flow tube is its shorter length, which makes the engineering task of inserting it into the flow line easier. The Dall tube has one further operational advantage, in that the permanent pressure loss imposed on the measured system is only about 5% of the measured pressure difference ($P_1 - P_2$).

The **flow nozzle** is of simpler construction still, and is therefore cheaper than either a Venturi or a Dall flow tube, but the pressure loss imposed on the flowing fluid is 30–50% of the measured pressure difference ($P_1 - P_2$).

Pitot static tube

The Pitot static tube is mainly used for making temporary measurements of flow, although it is also used in some instances for permanent flow monitoring. It measures the local velocity of flow at a particular point within a pipe rather than the average flow velocity as measured by other types of flowmeter. This may be very useful where there is a requirement to measure local flow rates across the cross-section of a pipe in the case of non-uniform flow. Multiple Pitot tubes are normally used to do this.

The instrument depends on the principle that a tube placed with its open end in a stream of fluid, as shown in Figure 16.6, will bring to rest that part of the fluid which impinges on it, and the loss of kinetic energy will be converted to a measurable increase in pressure inside the tube. This pressure (P_1), as well as the static pressure of the undisturbed free stream of flow (P_2), is measured. The flow velocity can then be calculated from the formula:

$$v = C\sqrt{2g(P_1 - P_2)}$$

The constant C, known as the Pitot tube coefficient, is a factor which corrects for the fact that not all fluid incident on the end of the tube will be brought to rest: a proportion will slip around it according to the design of the tube. Having calculated v, the volume flow rate can then be calculated by multiplying v by the cross-sectional area of the flow pipe, A.

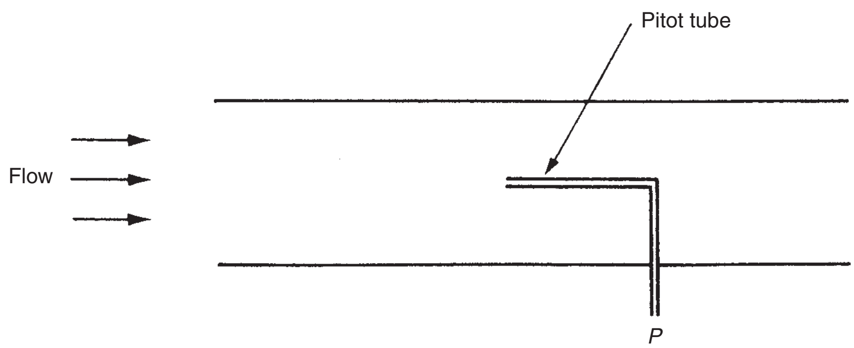


Fig. 16.6 Pitot tube.

Pitot tubes have the advantage that they cause negligible pressure loss in the flow. They are also cheap, and the installation procedure consists of the very simple process of pushing them down a small hole drilled in the flow-carrying pipe. Their main failing is that the measurement inaccuracy is typically about ±5%, although more expensive versions can reduce inaccuracy down to ±1%. The *annubar* is a development of the Pitot tube that has multiple sensing ports distributed across the cross-section of the pipe. It thus provides only an approximate measurement of the mean flow rate across the pipe.

16.2.2 Variable area flowmeters (Rotameters)

In the variable area flowmeter (which is also sometimes known as a Rotameter), the differential pressure across a variable aperture is used to adjust the area of the aperture. The aperture area is then a measure of the flow rate. The instrument is reliable and cheap and used extensively throughout industry, accounting for about 20% of all flowmeters sold. Normally, this type of instrument only gives a visual indication of flow rate, and so it is of no use in automatic control schemes. However, special versions of variable area flowmeters are now available that incorporate fibre optics. In these, a row of fibres detects the position of the float by sensing the reflection of light from it, and an electrical signal output can be derived from this.

In its simplest form, shown in Figure 16.7, the instrument consists of a tapered glass tube containing a float which takes up a stable position where its submerged weight is balanced by the upthrust due to the differential pressure across it. The position of the float is a measure of the effective annular area of the flow passage and hence of

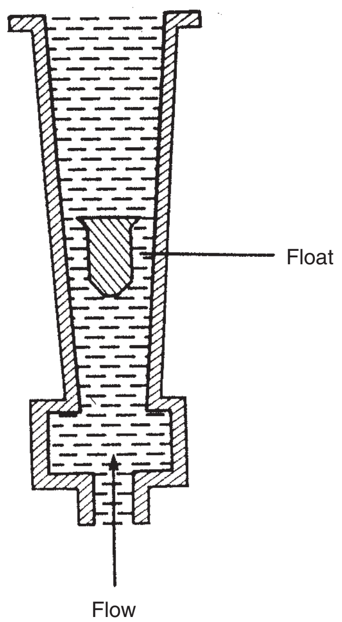


Fig. 16.7 Variable area flowmeter.

the flow rate. The inaccuracy of the cheapest instruments is typically ±5%, but more expensive versions offer measurement inaccuracies as low as ±0.5%.

16.2.3 Positive displacement flowmeters

Positive displacement flowmeters account for nearly 10% of the total number of flowmeters used in industry and are used in large numbers for metering domestic gas and water consumption. The cheapest instruments have a typical inaccuracy of about ±2%, but the inaccuracy in more expensive ones can be as low as ±0.5%. These higher quality instruments are used extensively within the oil industry, as such applications can justify the high cost of such instruments.

All positive displacement meters operate by using mechanical divisions to displace discrete volumes of fluid successively. Whilst this principle of operation is common, many different mechanical arrangements exist for putting the principle into practice. However, all versions of positive displacement meter are low friction, low maintenance and long-life devices, although they do impose a small permanent pressure loss on the flowing fluid. Low friction is especially important when measuring gas flows, and meters with special mechanical arrangements to satisfy this requirement have been developed.

The *rotary piston meter* is a common type of positive displacement meter, and the principles of operation of this are shown in Figure 16.8. It consists of a slotted cylindrical piston moving inside a cylindrical working chamber that has an inlet port and an outlet port. The piston moves round the chamber such that its outer surface maintains contact with the inner surface of the chamber, and, as this happens, the piston slot slides up and down a fixed division plate in the chamber. At the start of each piston motion cycle, liquid is admitted to volume B from the inlet port. The fluid

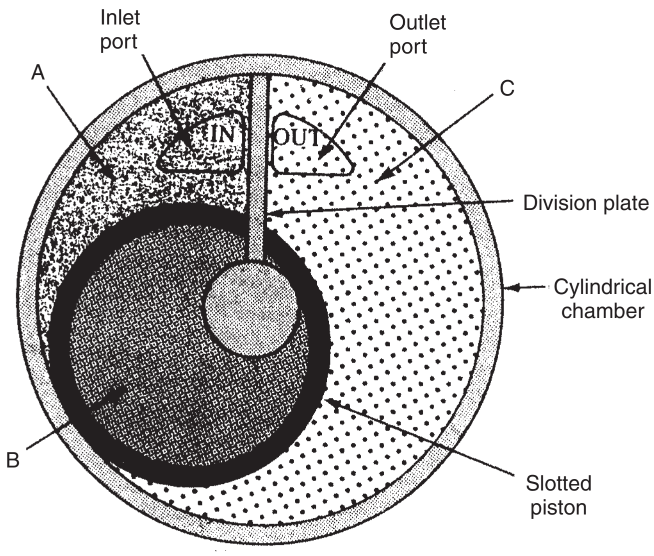


Fig. 16.8 Rotaty piston form of positive displacement flowmeter.

pressure causes the piston to start to rotate around the chamber, and, as this happens, liquid in volume C starts to flow out of the outlet port, and also liquid starts to flow from the inlet port into volume A. As the piston rotates further, volume B becomes shut off from the inlet port, whilst liquid continues to be admitted into A and pushed out of C. When the piston reaches the endpoint of its motion cycle, the outlet port is opened to volume B, and the liquid which has been transported round inside the piston is expelled. After this, the piston pivots about the contact point between the top of its slot and the division plate, and volume A effectively becomes volume C ready for the start of the next motion cycle. A peg on top of the piston causes a reciprocating motion of a lever attached to it. This is made to operate a counter, and the flow rate is therefore determined from the count in unit time multiplied by the quantity (fixed) of liquid transferred between the inlet and outlet ports for each motion cycle.

16.2.4 Turbine meters

A turbine flowmeter consists of a multi-bladed wheel mounted in a pipe along an axis parallel to the direction of fluid flow in the pipe, as shown in Figure 16.9. The flow of fluid past the wheel causes it to rotate at a rate that is proportional to the volume flow rate of the fluid. This rate of rotation has traditionally been measured by constructing the flowmeter such that it behaves as a variable reluctance tachogenerator. This is achieved by fabricating the turbine blades from a ferromagnetic material and placing a permanent magnet and coil inside the meter housing. A voltage pulse is induced in the coil as each blade on the turbine wheel moves past it, and if these pulses are measured by a pulse counter, the pulse frequency and hence flow rate can be deduced. In recent instruments, fibre optics are also now sometimes used to count the rotations by detecting reflections off the tip of the turbine blades.

Provided that the turbine wheel is mounted in low friction bearings, measurement inaccuracy can be as low as ±0.2%. However, turbine flowmeters are less rugged and

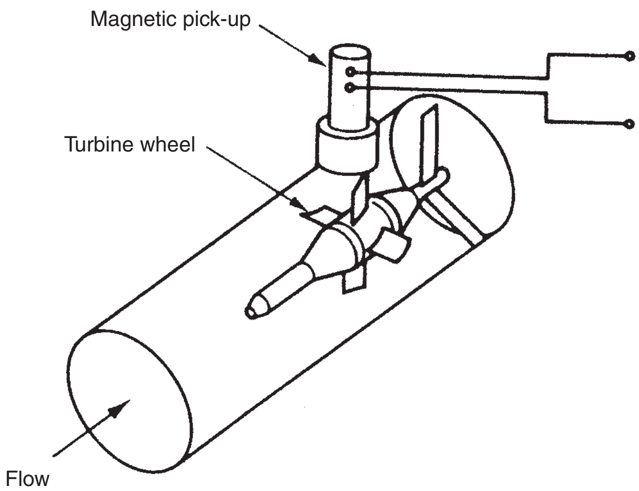


Fig. 16.9 Turbine flowmeter.

reliable than flow-restriction type instruments, and are badly affected by any particulate matter in the flowing fluid. Bearing wear is a particular problem and they also impose a permanent pressure loss on the measured system. Turbine meters are particularly prone to large errors when there is any significant second phase in the fluid measured. For instance, using a turbine meter calibrated on pure liquid to measure a liquid containing 5% air produces a 50% measurement error. As an important application of the turbine meter is in the petrochemical industries, where gas/oil mixtures are common, special procedures are being developed to avoid such large measurement errors. The most promising approach is to homogenize the two gas/oil phases prior to flow measurement (King, 1988).

Turbine meters have a similar cost and market share to positive displacement meters, and compete for many applications, particularly in the oil industry. Turbine meters are smaller and lighter than the latter and are preferred for low-viscosity, high-flow measurements. However, positive-displacement meters are superior in conditions of high viscosity and low flow rate.

16.2.5 Electromagnetic flowmeters

Electromagnetic flowmeters are limited to measuring the volume flow rate of electrically conductive fluids. The typical measurement inaccuracy of around $\pm 1\%$ is acceptable in many applications, but the instrument is expensive both in terms of the initial purchase cost and also in running costs, mainly due to its electricity consumption. A further reason for high cost is the need for careful calibration of each instrument individually during manufacture, as there is considerable variation in the properties of the magnetic materials used.

The instrument, shown in Figure 16.10, consists of a stainless steel cylindrical tube, fitted with an insulating liner, which carries the measured fluid. Typical lining materials used are Neoprene, polytetrafluoroethylene (PTFE) and polyurethane. A magnetic field is created in the tube by placing mains-energized field coils either side of it, and the voltage induced in the fluid is measured by two electrodes inserted into opposite sides of the tube. The ends of these electrodes are usually flush with the inner surface of the cylinder. The electrodes are constructed from a material which is unaffected by most types of flowing fluid, such as stainless steel, platinum–iridium alloys, Hastelloy, titanium and tantalum. In the case of the rarer metals in this list, the electrodes account for a significant part of the total instrument cost.

By Faraday's law of electromagnetic induction, the voltage, E, induced across a length, L, of the flowing fluid moving at velocity, v, in a magnetic field of flux density, B, is given by:

$$E = BLv \tag{16.3}$$

L is the distance between the electrodes, which is the diameter of the tube, and B is a known constant. Hence, measurement of the voltage E induced across the electrodes allows the flow velocity v to be calculated from equation (16.3). Having thus calculated v, it is a simple matter to multiply v by the cross-sectional area of the tube to obtain a value for the volume flow rate. The typical voltage signal measured across the electrodes is 1 mV when the fluid flow rate is 1 m/s.

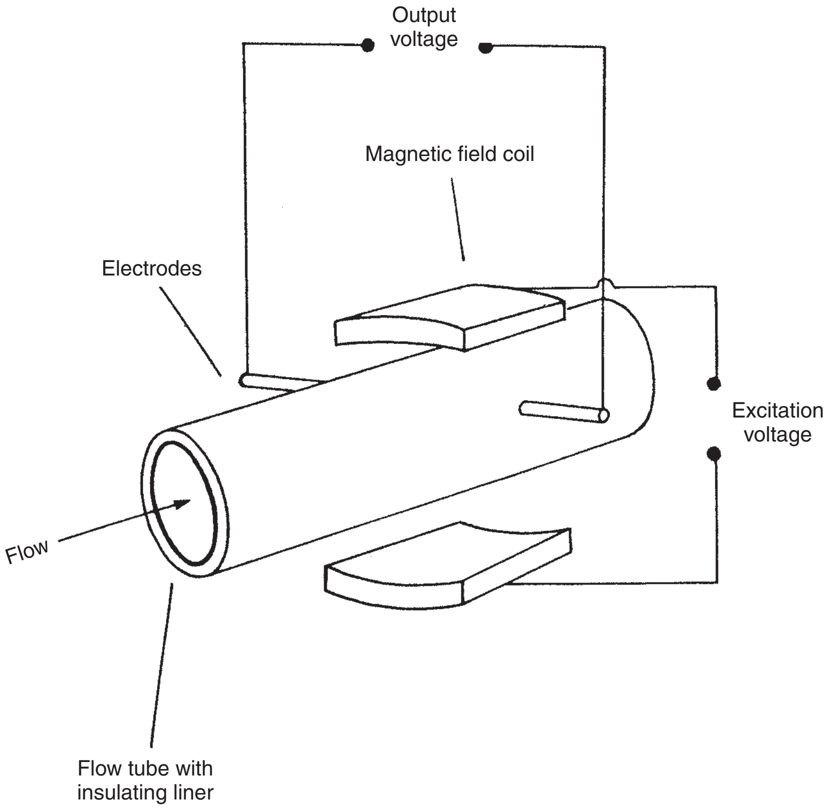


Fig. 16.10 Electromagnetic flowmeter.

The internal diameter of magnetic flowmeters is normally the same as that of the rest of the flow-carrying pipework in the system. Therefore, there is no obstruction to the fluid flow and consequently no pressure loss associated with measurement. Like other forms of flowmeter, the magnetic type requires a minimum length of straight pipework immediately prior to the point of flow measurement in order to guarantee the accuracy of measurement, although a length equal to five pipe diameters is usually sufficient.

Whilst the flowing fluid must be electrically conductive, the method is of use in many applications and is particularly useful for measuring the flow of slurries in which the liquid phase is electrically conductive. Corrosive fluids can be handled providing a suitable lining material is used. At the present time, magnetic flowmeters account for about 15% of the new flowmeters sold and this total is slowly growing. One operational problem is that the insulating lining is subject to damage when abrasive fluids are being handled, and this can give the instrument a limited life.

Current new developments in electromagnetic flowmeters are producing physically smaller instruments and employing better coil designs which reduce electricity consumption and make battery-powered versions feasible (these are now commercially available). Also, whereas conventional electromagnetic flowmeters require a minimum fluid conductivity of 10 μmho/cm^3, new versions can cope with fluid conductivities as low as 1 μmho/cm^3.

16.2.6 Vortex-shedding flowmeters

The vortex-shedding flowmeter is a relatively new type of instrument which is rapidly gaining in popularity and is being used as an alternative to traditional differential pressure meters in more and more applications. The operating principle of the instrument is based on the natural phenomenon of vortex shedding, created by placing an unstreamlined obstacle (known as a bluff body) in a fluid-carrying pipe, as indicated in Figure 16.11. When fluid flows past the obstacle, boundary layers of viscous, slow-moving fluid are formed along the outer surface. Because the obstacle is not streamlined, the flow cannot follow the contours of the body on the downstream side, and the separate layers become detached and roll into eddies or vortices in the low-pressure region behind the obstacle. The shedding frequency of these alternately shed vortices is proportional to the fluid velocity past the body. Various thermal, magnetic, ultrasonic and capacitive vortex detection techniques are employed in different instruments.

Such instruments have no moving parts, operate over a wide flow range, have a low power consumption, require little maintenance and have a similar cost to measurement using an orifice plate. They can measure both liquid and gas flows and a common inaccuracy figure quoted is $\pm 1\%$ of full-scale reading, though this can be seriously downgraded in the presence of flow disturbances upstream of the measurement point and a straight run of pipe before the measurement point of 50 pipe diameters is recommended. Another problem with the instrument is its susceptibility to pipe vibrations, although new designs are becoming available which have a better immunity to such vibrations.

16.2.7 Ultrasonic flowmeters

The ultrasonic technique of volume flow rate measurement is, like the magnetic flowmeter, a non-invasive method. It is not restricted to conductive fluids, however, and

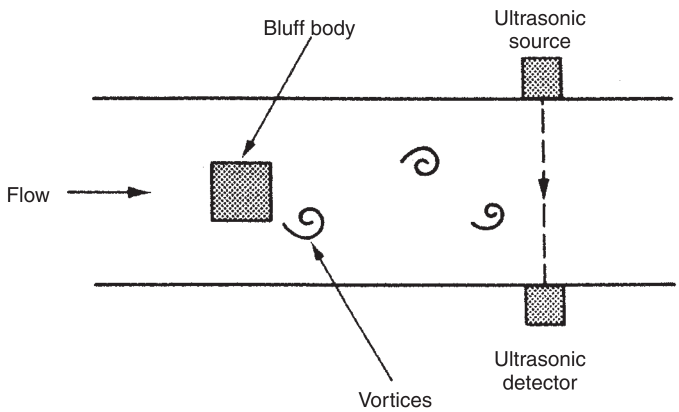


Fig. 16.11 Vortex-shedding flowmeter.

is particularly useful for measuring the flow of corrosive fluids and slurries. Besides its high reliability and low maintenance requirements, a further advantage of an ultrasonic flowmeter over a magnetic flowmeter is that the instrument can be clamped externally onto existing pipework rather than being inserted as an integral part of the flow line. As the procedure of breaking into a pipeline to insert a flowmeter can be as expensive as the cost of the flowmeter itself, the ultrasonic flowmeter has enormous cost advantages. Its clamp-on mode of operation has significant safety advantages in avoiding the possibility of personnel installing flowmeters coming into contact with hazardous fluids such as poisonous, radioactive, flammable or explosive ones. Also, any contamination of the fluid being measured (e.g. food substances and drugs) is avoided. Ultrasonic meters are still less common than differential pressure or electromagnetic flowmeters, though usage continues to expand year by year.

Two different types of ultrasonic flowmeter exist which employ distinct technologies, one based on Doppler shift and the other on transit time. In the past, the existence of these alternative technologies has not always been readily understood, and has resulted in ultrasonic technology being rejected entirely when one of these two forms has been found to be unsatisfactory in a particular application. This is unfortunate, because the two technologies have distinct characteristics and areas of application, and many situations exist where one form is very suitable and the other not suitable. To reject both, having only tried out one, is therefore a serious mistake.

Particular care has to be taken to ensure a stable flow profile in ultrasonic flowmeter applications. It is usual to increase the normal specification of the minimum length of straight pipe-run prior to the point of measurement, expressed as a number of pipe diameters, from a figure of 10 up to 20 or in some cases even 50 diameters. Analysis of the reasons for poor performance in many instances of ultrasonic flowmeter application has shown failure to meet this stable flow-profile requirement to be a significant factor.

Doppler shift ultrasonic flowmeter

The principle of operation of the Doppler shift flowmeter is shown in Figure 16.12. A fundamental requirement of these instruments is the presence of scattering elements within the flowing fluid, which deflect the ultrasonic energy output from the transmitter such that it enters the receiver. These can be provided by either solid particles, gas bubbles or eddies in the flowing fluid. The scattering elements cause a frequency shift between the transmitted and reflected ultrasonic energy, and measurement of this shift enables the fluid velocity to be inferred.

The instrument consists essentially of an ultrasonic transmitter–receiver pair clamped onto the outside wall of a fluid-carrying vessel. Ultrasonic energy consists of a train of short bursts of sinusoidal waveforms at a frequency between 0.5 MHz and 20 MHz. This frequency range is described as ultrasonic because it is outside the range of human hearing. The flow velocity, v, is given by:

$$v = \frac{c(f_t - f_r)}{2f_t \cos(\theta)} \qquad (16.4)$$

where f_t and f_r are the frequencies of the transmitted and received ultrasonic waves respectively, c is the velocity of sound in the fluid being measured, and θ is the angle that the incident and reflected energy waves make with the axis of flow in the pipe.

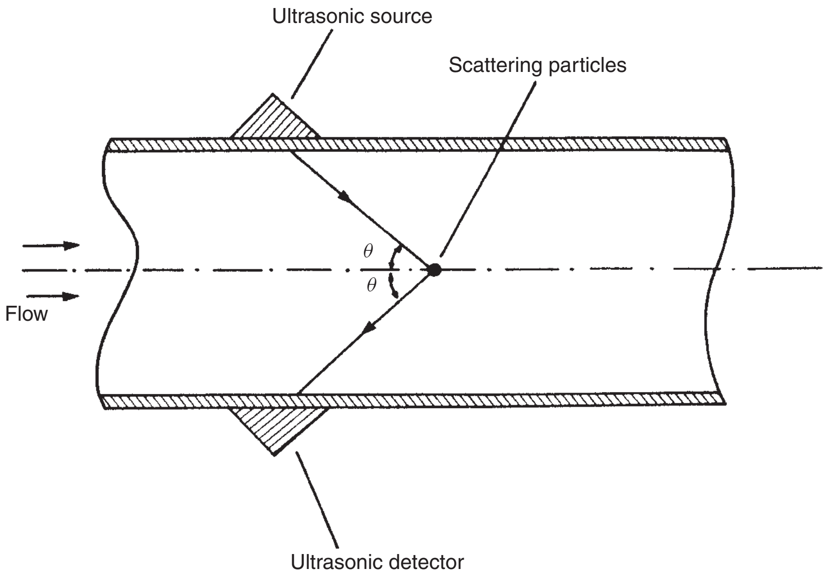


Fig. 16.12 Doppler shift ultrasonic flowmeter.

Volume flow rate is then readily calculated by multiplying the measured flow velocity by the cross-sectional area of the fluid-carrying pipe.

The electronics involved in Doppler-shift flowmeters is relatively simple and therefore cheap. Ultrasonic transmitters and receivers are also relatively inexpensive, being based on piezoelectric oscillator technology. As all of its components are cheap, the Doppler shift flowmeter itself is inexpensive. The measurement accuracy obtained depends on many factors such as the flow profile, the constancy of pipe-wall thickness, the number, size and spatial distribution of scatterers, and the accuracy with which the speed of sound in the fluid is known. Consequently, accurate measurement can only be achieved by the tedious procedure of carefully calibrating the instrument in each particular flow measurement application. Otherwise, measurement errors can approach $\pm 10\%$ of the reading, and for this reason Doppler shift flowmeters are often used merely as flow indicators, rather than for accurate quantification of the volume flow rate.

Versions are now available which avoid the problem of variable pipe thickness by being fitted inside the flow pipe, flush with its inner surface. A low inaccuracy level of $\pm 0.5\%$ is claimed for such devices. Other recent developments are the use of multiple-path ultrasonic flowmeters that use an array of ultrasonic elements to obtain an average velocity measurement that substantially reduces the error due to non-uniform flow profiles. There is a substantial cost penalty involved in this, however.

Transit-time ultrasonic flowmeter

The transit-time ultrasonic flowmeter is an instrument designed for measuring the volume flow rate in clean liquids or gases. It consists of a pair of ultrasonic transducers mounted along an axis aligned at an angle θ with respect to the fluid-flow axis, as shown in Figure 16.13. Each transducer consists of a transmitter–receiver pair, with the transmitter emitting ultrasonic energy which travels across to the receiver on the opposite

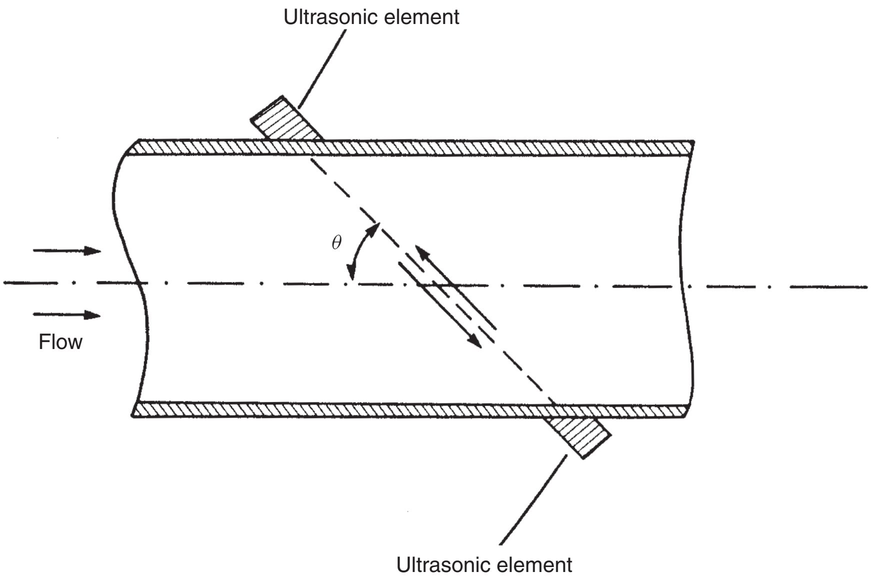


Fig. 16.13 Transit-time ultrasonic flowmeter.

side of the pipe. These ultrasonic elements are normally piezoelectric oscillators of the same type as used in Doppler shift flowmeters. Fluid flowing in the pipe causes a time difference between the transit times of the beams travelling upstream and downstream, and measurement of this difference allows the flow velocity to be calculated. The typical magnitude of this time difference is 100 ns in a total transit time of 100 μs, and high-precision electronics are therefore needed to measure it. There are three distinct ways of measuring the time shift. These are direct measurement, conversion to a phase change and conversion to a frequency change. The third of these options is particularly attractive, as it obviates the need to measure the speed of sound in the measured fluid as required by the first two methods. A scheme applying this third option is shown in Figure 16.14. This also multiplexes the transmitting and receiving functions, so that only one ultrasonic element is needed in each transducer. The forward and backward transit times across the pipe, T_f and T_b, are given by:

$$T_f = \frac{L}{c + v\cos(\theta)}; \quad T_b = \frac{L}{c - v\cos(\theta)}$$

where c is the velocity of sound in the fluid, v is the flow velocity, L is the distance between the ultrasonic transmitter and receiver, and θ is the angle of the ultrasonic beam with respect to the fluid flow axis.

The time difference δT is given by:

$$\delta T = T_b - T_f = \frac{2vL\cos(\theta)}{c^2 - v^2\cos^2(\theta)}$$

This requires knowledge of c before it can be solved. However, a solution can be found much more simply if the receipt of a pulse is used to trigger the transmission of the

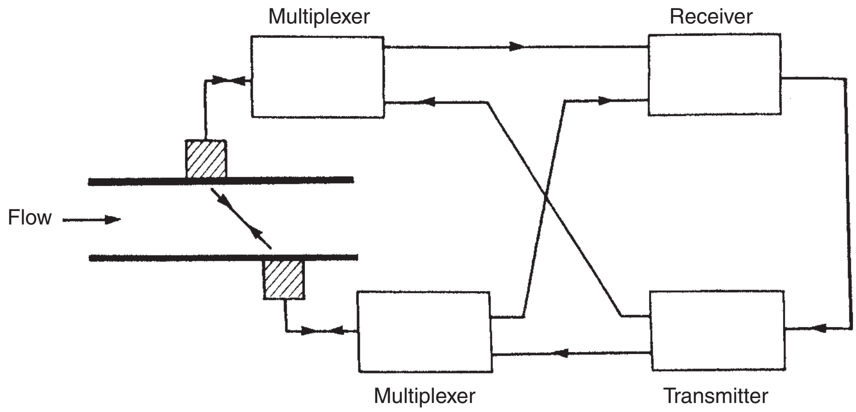


Fig. 16.14 Transit-time measurement system.

next ultrasonic energy pulse. Then, the frequencies of the forward and backward pulse trains are given by:

$$F_{\mathrm{f}} = \frac{1}{T_{\mathrm{f}}} = \frac{c - v\cos(\theta)}{L}; \quad F_{\mathrm{b}} = \frac{1}{T_{\mathrm{b}}} = \frac{c + v\cos(\theta)}{L}$$

If the two frequency signals are now multiplied together, the resulting beat frequency is given by:

$$\delta F = F_{\mathrm{b}} - F_{\mathrm{f}} = \frac{2v\cos(\theta)}{L}$$

c has now been eliminated and v can be calculated from a measurement of δF as:

$$v = \frac{L\delta F}{2\cos(\theta)}$$

This is often known as the *sing-around flowmeter*.

Transit-time flowmeters are of more general use than Doppler shift flowmeters, particularly where the pipe diameter involved is large and hence the transit time is consequently sufficiently large to be measured with reasonable accuracy. It is possible then to reduce the inaccuracy figure to $\pm 0.5\%$. The instrument costs more than a Doppler shift flowmeter, however, because of the greater complexity of the electronics needed to make accurate transit-time measurements.

16.2.8 Other types of flowmeter for measuring volume flow rate

The **gate meter** consists of a spring-loaded, hinged flap mounted at right angles to the direction of fluid flow in the fluid-carrying pipe. The flap is connected to a pointer outside the pipe. The fluid flow deflects the flap and pointer and the flow rate is indicated by a graduated scale behind the pointer. The major difficulty with such devices is in preventing leaks at the hinge point. A variation on this principle is the *air-vane meter*, which measures deflection of the flap by a potentiometer inside the pipe. This is commonly used to measure airflow within automotive fuel-injection systems. Another similar device is the *target meter*. This consists of a circular disc-shaped flap in the pipe. Fluid flow rate is inferred from the force exerted on the disc measured by strain gauges bonded to it. This meter is very useful for measuring the flow of dilute slurries but it does not find wide application elsewhere as it has a relatively high cost. Measurement uncertainty in all of these types of meter varies between 1% and 5% according to cost and design of each instrument.

The **cross-correlation flowmeter** has not yet achieved widespread practical use in industry. Much development work is still going on, and it therefore mainly only exists in prototype form in research laboratories. However, it is included here because use is likely to become much more widespread in the future. The instrument requires some detectable random variable to be present in the flowing fluid. This can take forms such as velocity turbulence and temperature fluctuations. When such a stream of variables is detected by a sensor, the output signal generated consists of noise with a wide frequency spectrum.

Cross-correlation flowmeters use two such sensors placed a known distance apart in the fluid-carrying pipe and cross-correlation techniques are applied to the two output signals from these sensors. This procedure compares one signal with progressively time-shifted versions of the other signal until the best match is obtained between the two waveforms. If the distance between the sensors is divided by this time shift, a measurement of the flow velocity is obtained. A digital processor is an essential requirement to calculate the cross-correlation function, and therefore the instrument must be properly described as an intelligent one.

In practice, the existence of random disturbances in the flow is unreliable, and their detection is difficult. To answer this problem, ultrasonic cross-correlation flowmeters are under development. These use ultrasonic transducers to inject disturbances into the flow and also to detect the disturbances further downstream.

Further information about cross-correlation flowmeters can be found in Medlock (1985).

The **Laser Doppler flowmeter** gives direct measurements of flow velocity for liquids containing suspended particles flowing in a transparent pipe. Light from a laser is focused by an optical system to a point in the flow, with fibre-optic cables being commonly used to transmit the light. The movement of particles causes a Doppler shift of the scattered light and produces a signal in a photodetector that is related to the fluid velocity. A very wide range of flow velocities between 10 μm/s and 105 m/s can be measured by this technique.

Sufficient particles for satisfactory operation are normally present naturally in most liquid and gaseous fluids, and the introduction of artificial particles is rarely needed. The technique is advantageous in measuring flow velocity directly rather than inferring it from a pressure difference. It also causes no interruption in the flow and, as the instrument can be made very small, it can measure velocity in confined areas. One limitation is that it measures local flow velocity in the vicinity of the focal point of the light beam, which can lead to large errors in the estimation of mean volume flow rate if the flow profile is not uniform. However, this limitation is often used constructively in applications of the instrument where the flow profile across the cross-section of a pipe is determined by measuring the velocity at a succession of points.

Whilst the **Coriolis meter** is primarily intended to be a mass flow measuring instrument, it can also be used to measure volume flow rate when high measurement accuracy is required. However, its high cost means that alternative instruments are normally used for measuring volume flow rate.

16.3 Intelligent flowmeters

All the usual benefits associated with intelligent instruments are applicable to most types of flowmeter. Indeed, all types of mass flowmeter routinely have intelligence as an integral part of the instrument. For volume flow rate measurement, intelligent differential pressure measuring instruments can be used to good effect in conjunction with obstruction type flow transducers. One immediate benefit of this in the case of the commonest flow restriction device, the orifice plate, is to extend the lowest flow measurable with acceptable accuracy down to 20% of the maximum flow value. In positive displacement meters, intelligence allows compensation for thermal expansion of meter components and temperature-induced viscosity changes. Correction for variations in flow pressure is also provided for. Intelligent electromagnetic flowmeters are also available, and these have a self-diagnosis and self-adjustment capability. The usable instrument range is typically from 3% to 100% of the full-scale reading and the quoted maximum inaccuracy is ±0.5%. It is also normal to include a non-volatile memory to protect constants used for correcting for modifying inputs, etc., against power supply failures. Intelligent turbine meters are able to detect their own bearing wear and also report deviations from initial calibration due to blade damage, etc. Some versions also have self-adjustment capability.

The trend is now moving towards total flow computers which can process inputs from almost any type of transducer. Such devices allow user input of parameters like specific gravity, fluid density, viscosity, pipe diameters, thermal expansion coefficients, discharge coefficients, etc. Auxiliary inputs from temperature transducers are also catered for. After processing the raw flow transducer output with this additional data, flow computers are able to produce measurements of flow to a very high degree of accuracy.

16.4 Choice between flowmeters for particular applications

The number of relevant factors to be considered when specifying a flowmeter for a particular application is very large. These include the temperature and pressure of the fluid, its density, viscosity, chemical properties and abrasiveness, whether it contains particles, whether it is a liquid or gas, etc. This narrows the field to a subset of instruments that are physically capable of making the measurement. Next, the required performance factors of accuracy, rangeability, acceptable pressure drop, output signal characteristics, reliability and service life must be considered. Accuracy requirements vary widely across different applications, with measurement uncertainty of ±5% being acceptable in some and less than ±0.5% being demanded in others.

Finally, the economic viability must be assessed and this must take account not only of purchase cost, but also of reliability, installation difficulties, maintenance requirements and service life.

Where only a visual indication of flow rate is needed, the variable-area meter is popular. Where a flow measurement in the form of an electrical signal is required, the choice of available instruments is very large. The orifice plate is used extremely commonly for such purposes and accounts for almost 50% of the instruments currently in use in industry. Other forms of differential pressure meter and electromagnetic flowmeters are used in significant numbers. Currently, there is a trend away from rotating devices such as turbine meters and positive displacement meters. At the same time, usage of ultrasonic and vortex meters is expanding. A survey of the current market share enjoyed by different types can be found in Control Engineering (1998).

References and further reading

Control Engineering (Editorial) (April 1998), pp. 119–128.

Figiola, R.S. and Beasley, D.E. (1995) *Theory and Design of Mechanical Measurements*, John Wiley.

Instrument Society of America (1988) *Flowmeters – a comprehensive survey and guide to selection*, Pittsburgh.

King, N.W. (1988) Multi-phase flow measurement at NEL, *Measurement and Control*, **21**, pp. 237–239.

Medlock, R.S. (1985) Cross-correlation flow measurement, *Measurement and Control*, **18**(8), pp. 293–298.

Medlock, R.S. and Furness, R.A. (1990) Mass flow measurement – a state of the art review, *Measurement and Control*, **23**(4), pp. 100–113.

17

Level measurement

A wide variety of instruments are available for measuring the level of liquids. Some of these can also be used to measure the levels of solids that are in the form of powders or small particles. In some applications, only a rough indication of level is needed, and simple devices such as dipsticks or float systems are adequate. However, in other cases where high accuracy is demanded, other types of instrument must be used. The sections below cover the various kinds of level-measuring device available.

17.1 Dipsticks

Dipsticks offer a simple means of measuring level approximately. The *ordinary dipstick* is the cheapest device available. This consists of a metal bar on which a scale is etched, as shown in Figure 17.1(a). The bar is fixed at a known position in the liquid-containing vessel. A level measurement is made by removing the instrument from the vessel and reading off how far up the scale the liquid has wetted. As a human operator is required to remove and read the dipstick, this method can only be used in relatively small and shallow vessels.

The *optical dipstick*, illustrated in Figure 17.1(b), is an alternative form that allows a reading to be obtained without removing the dipstick from the vessel, and so is applicable to larger, deeper tanks. Light from a source is reflected from a mirror, passes round the chamfered end of the dipstick, and enters a light detector after reflection by a second mirror. When the chamfered end comes into contact with liquid, its internal reflection properties are altered and light no longer enters the detector. By using a suitable mechanical drive system to move the instrument up and down and measure its position, the liquid level can be monitored.

17.2 Float systems

Float systems, whereby the position of a float on the surface of a liquid is measured by means of a suitable transducer, have a typical measurement inaccuracy of $\pm 1\%$. This method is also simple, cheap and widely used. The system using a potentiometer, shown earlier in Figure 2.2, is very common, and is well known for its application

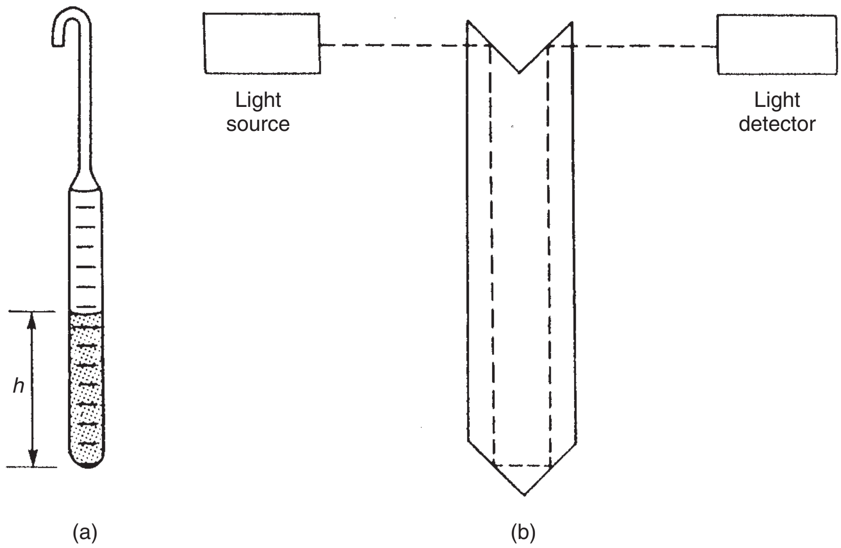


Fig. 17.1 Dipsticks: (a) simple dipstick; (b) optical dipstick.

to monitoring the level in motor vehicle fuel tanks. An alternative system, which is used in greater numbers, is called the *float and tape gauge* (or *tank gauge*). This has a tape attached to the float that passes round a pulley situated vertically above the float. The other end of the tape is attached to either a counterweight or a negative-rate counter-spring. The amount of rotation of the pulley, measured by either a synchro or a potentiometer, is then proportional to the liquid level. These two essentially mechanical systems of measurement are popular in many applications, but the maintenance requirements of them are always high.

17.3 Pressure-measuring devices (hydrostatic systems)

The hydrostatic pressure due to a liquid is directly proportional to its depth and hence to the level of its surface. Several instruments are available that use this principle, and they are widely used in many industries, particularly in harsh chemical environments. In the case of open-topped vessels (or covered ones that are vented to the atmosphere), the level can be measured by inserting a pressure sensor at the bottom of the vessel, as shown in Figure 17.2(a). The liquid level h is then related to the measured pressure P according to $h = P/\rho g$, where ρ is the liquid density and g is the acceleration due to gravity. One source of error in this method can be imprecise knowledge of the liquid density. This can be a particular problem in the case of liquid solutions and mixtures (especially hydrocarbons), and in some cases only an estimate of density is available. Even with single liquids, the density is subject to variation with temperature, and therefore temperature measurement may be required if very accurate level measurements are needed.

Where liquid-containing vessels are totally sealed, the liquid level can be calculated by measuring the differential pressure between the top and bottom of the tank, as

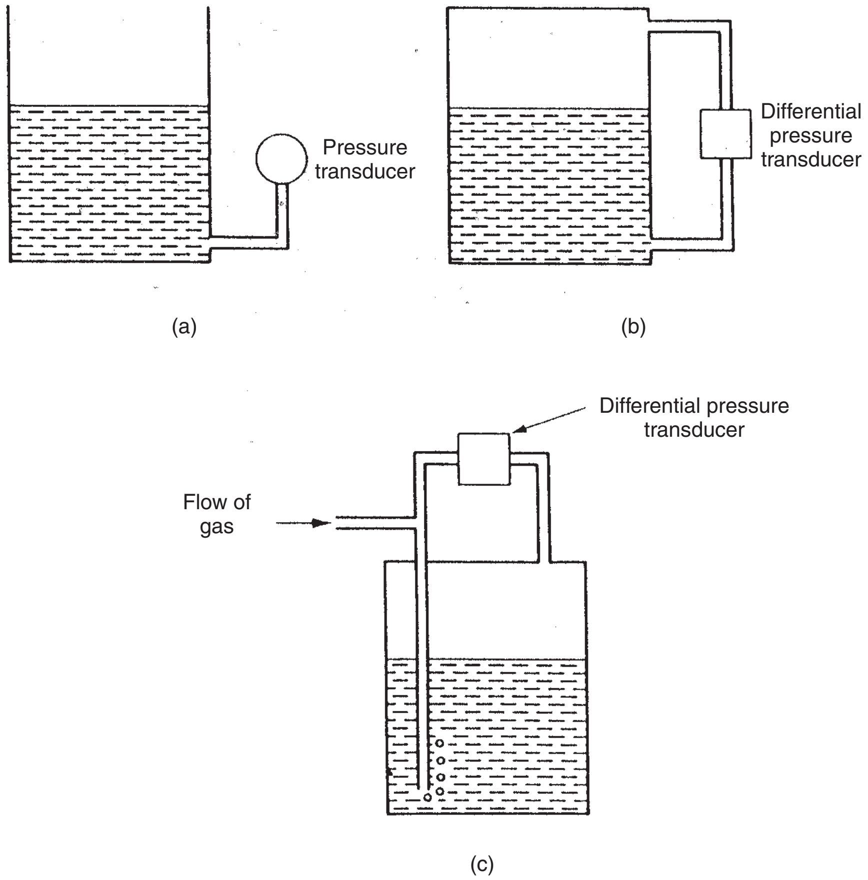


Fig. 17.2 Hydrostatic systems: (a) open-topped vessel; (b) sealed vessel; (c) bubbler unit.

shown in Figure 17.2(b). The differential pressure transducer used is normally a standard diaphragm type, although silicon-based microsensors are being used in increasing numbers. The liquid level is related to the differential pressure measured, δP, according to $h = \delta P/\rho g$. The same comments as for the case of the open vessel apply regarding uncertainty in the value of ρ. An additional problem that can occur is an accumulation of liquid on the side of the differential pressure transducer that is measuring the pressure at the top of the vessel. This can arise because of temperature fluctuations, which allow liquid to alternately vaporize from the liquid surface and then condense in the pressure tapping at the top of the vessel. The effect of this on the accuracy of the differential pressure measurement is severe, but the problem is easily avoided by placing a drain pot in the system.

A final pressure-related system of level measurement is the *bubbler unit* shown in Figure 17.2(c). This uses a dip pipe that reaches to the bottom of the tank and is purged free of liquid by a steady flow of gas through it. The rate of flow is adjusted until gas bubbles are just seen to emerge from the end of the tube. The pressure in the tube, measured by a pressure transducer, is then equal to the liquid pressure at the bottom of the tank. It is important that the gas used is inert with respect to the liquid in the vessel. Nitrogen, or sometimes just air, is suitable in most cases. Gas consumption is low, and a cylinder of nitrogen may typically last for six months. The method is suitable for measuring the liquid pressure at the bottom of both open and sealed tanks. It is particularly advantageous in avoiding the large maintenance problem associated with leaks at the bottom of tanks at the site of the pressure tappings required by alternative methods.

Measurement uncertainty varies according to the application and the condition of the measured material. A typical value would be ±0.5% of full-scale reading, although ±0.1% can be achieved in some circumstances.

17.4 Capacitive devices

Capacitive devices are widely used for measuring the level of both liquids and solids in powdered or granular form. They perform well in many applications, but become inaccurate if the measured substance is prone to contamination by agents that change the dielectric constant. Ingress of moisture into powders is one such example of this. They are also suitable for use in extreme conditions measuring liquid metals (high temperatures), liquid gases (low temperatures), corrosive liquids (acids, etc.) and high-pressure processes. Two versions are used according to whether the measured substance

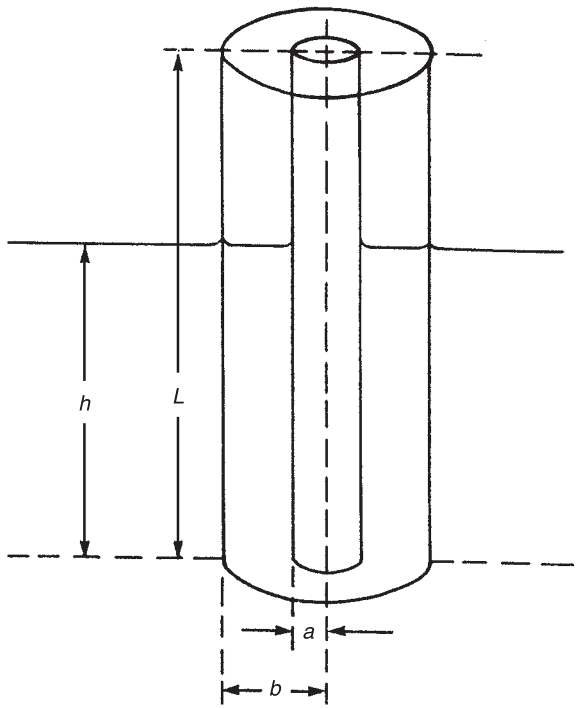


Fig. 17.3 Capacitive level sensor.

is conducting or not. For non-conducting substances (less than 0.1 μmho/cm^3), two bare-metal capacitor plates in the form of concentric cylinders are immersed in the substance, as shown in Figure 17.3. The substance behaves as a dielectric between the plates according to the depth of the substance. For concentric cylinder plates of radius a and b ($b > a$), and total height L, the depth of the substance h is related to the measured capacitance C by:

$$h = \frac{C \log_e (b/a) - 2\pi\varepsilon_0}{2\pi\varepsilon_0 (\varepsilon - 1)} \tag{17.1}$$

where ε is the relative permittivity of the measured substance and ε_0 is the permittivity of free space. In the case of conducting substances, exactly the same measurement techniques are applied, but the capacitor plates are encapsulated in an insulating material. The relationship between C and h in equation (17.1) then has to be modified to allow for the dielectric effect of the insulator. Measurement uncertainty is typically 1–2%.

17.5 Ultrasonic level gauge

Ultrasonic level measurement is one of a number of non-contact techniques available. The principle of the ultrasonic level gauge is that energy from an ultrasonic source above the liquid is reflected back from the liquid surface into an ultrasonic energy detector, as illustrated in Figure 17.4. Measurement of the time of flight allows the liquid level to be inferred. In alternative versions, the ultrasonic source is placed at the bottom of the vessel containing the liquid, and the time of flight between emission, reflection off the liquid surface and detection back at the bottom of the vessel is measured.

Ultrasonic techniques are especially useful in measuring the position of the interface between two immiscible liquids contained in the same vessel, or measuring the sludge or precipitate level at the bottom of a liquid-filled tank. In either case, the method employed is to fix the ultrasonic transmitter–receiver transducer at a known height in the upper liquid, as shown in Figure 17.5. This establishes the level of the liquid/liquid or liquid/sludge level in absolute terms. When using ultrasonic instruments, it is essential that proper compensation is made for the working temperature if this differs from the calibration temperature, since the speed of ultrasound through air varies with temperature (see Chapter 13). Ultrasound speed also has a small sensitivity to humidity, air pressure and carbon dioxide concentration, but these factors are usually insignificant. Temperature compensation can be achieved in two ways. Firstly, the operating temperature can be measured and an appropriate correction made. Secondly, and preferably, a comparison method can be used in which the system is calibrated each time it is used by measuring the transit time of ultrasonic energy between two known reference points. This second method takes account of humidity, pressure and carbon dioxide concentration variations as well as providing temperature compensation. With appropriate care, measurement uncertainty can be reduced to about ±1%.

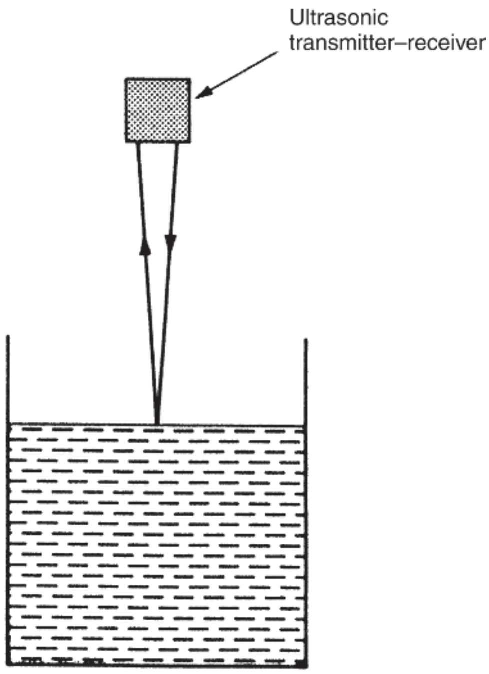


Fig. 17.4 Ultrasonic level gauge.

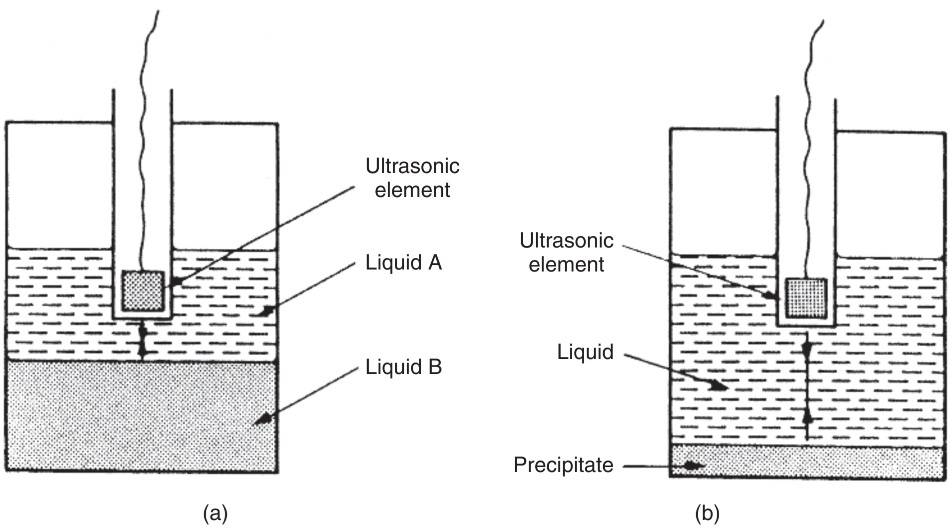


Fig. 17.5 Measuring interface positions: (a) liquid/liquid interface; (b) liquid/precipitate interface.

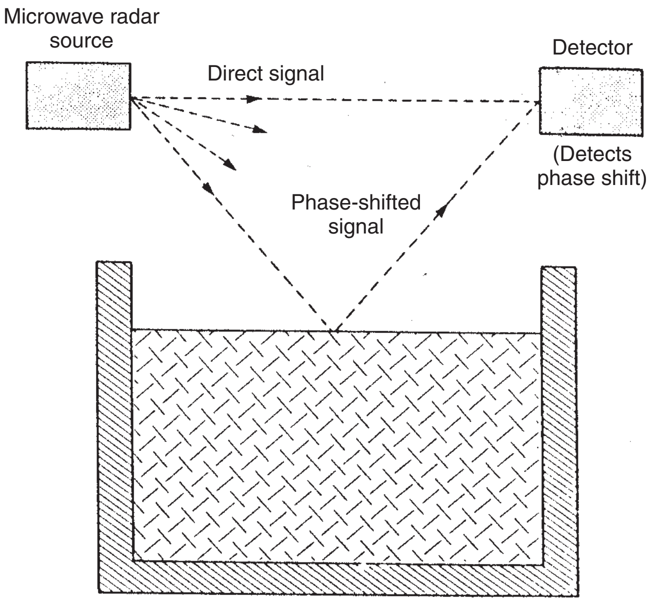


Fig. 17.6 Radar level detector.

17.6 Radar (microwave) methods

Level-measuring instruments using microwave radar are an alternative technique for non-contact measurement. Currently, they are still very expensive (~£3000), but prices are falling and usage is expanding rapidly. They are able to provide successful level measurement in applications that are otherwise very difficult, such as measurement in closed tanks, measurement where the liquid is turbulent, and measurement in the presence of obstructions and steam condensate. The technique involves directing a constant-amplitude, frequency-modulated microwave signal at the liquid surface. A receiver measures the phase difference between the reflected signal and the original signal transmitted directly through air to it, as shown in Figure 17.6. This measured phase difference is linearly proportional to the liquid level. The system is similar in principle to ultrasonic level measurement, but has the important advantage that the transmission time of radar through air is almost totally unaffected by ambient temperature and pressure fluctuations. However, as the microwave frequency is within the band used for radio communications, strict conditions on amplitude levels have to be satisfied, and the appropriate licences have to be obtained.

17.7 Radiation methods

The radiation method is an expensive technique, which uses a radiation source and detector system located outside a liquid-filled tank in the manner shown in Figure 17.6. The non-invasive nature of this technique in using a source and detector system outside

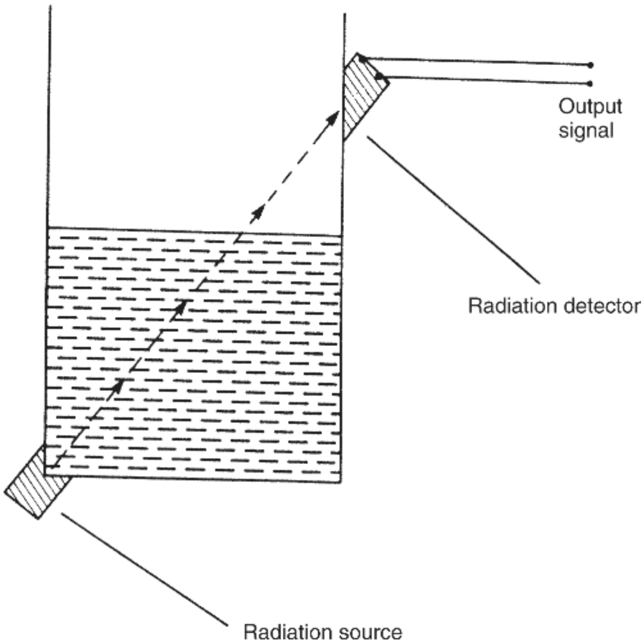


Fig. 17.7 Using a radiation source to measure level.

the tank is particularly attractive. The absorption of both beta rays and gamma rays varies with the amount of liquid between the source and detector, and hence is a function of liquid level. Caesium-137 is a commonly used gamma-ray source. The radiation level measured by the detector I is related to the length of liquid in the path x according to:

$$I = I_0 \exp(-\mu \rho x) \tag{17.2}$$

where I_0 is the intensity of radiation that would be received by the detector in the absence of any liquid, μ is the mass absorption coefficient for the liquid and ρ is the mass density of the liquid.

In the arrangement shown in Figure 17.7, the radiation follows a diagonal path across the liquid, and therefore some trigonometrical manipulation has to be carried out to determine the liquid level h from x. In some applications, the radiation source can be located in the centre of the bottom of the tank, with the detector vertically above it. Where this is possible, the relationship between the radiation detected and liquid level is obtained by directly substituting h in place of x in equation (17.2). Apart from use with liquids at normal temperatures, this method is commonly used for measuring the level of hot, liquid metals. However, because of the obvious dangers associated with using radiation sources, very strict safety regulations have to be satisfied when applying this technique. Very low activity radiation sources are used in some systems to overcome safety problems but the system is then sensitive to background radiation and special precautions have to be taken regarding the provision of adequate shielding. Because of the many difficulties in using this technique, it is only used in special applications.

17.8 Other techniques

17.8.1 Vibrating level sensor

The principle of the vibrating level sensor is illustrated in Figure 17.8. The instrument consists of two piezoelectric oscillators fixed to the inside of a hollow tube that generate flexural vibrations in the tube at its resonant frequency. The resonant frequency of the tube varies according to the depth of its immersion in the liquid. A phase-locked loop circuit is used to track these changes in resonant frequency and adjust the excitation frequency applied to the tube by the piezoelectric oscillators. Liquid level measurement is therefore obtained in terms of the output frequency of the oscillator when the tube is resonating.

17.8.2 Hot-wire elements/carbon resistor elements

Figure 17.9 shows a level measurement system that uses a series of hot-wire elements or carbon resistors placed at regular intervals along a vertical line up the side of a tank. The heat transfer coefficient of such elements differs substantially depending upon whether the element is immersed in air or in the liquid in the tank. Consequently, elements in the liquid have a different temperature and therefore a different resistance to those in air. This method of level measurement is a simple one, but the measurement resolution is limited to the distance between sensors.

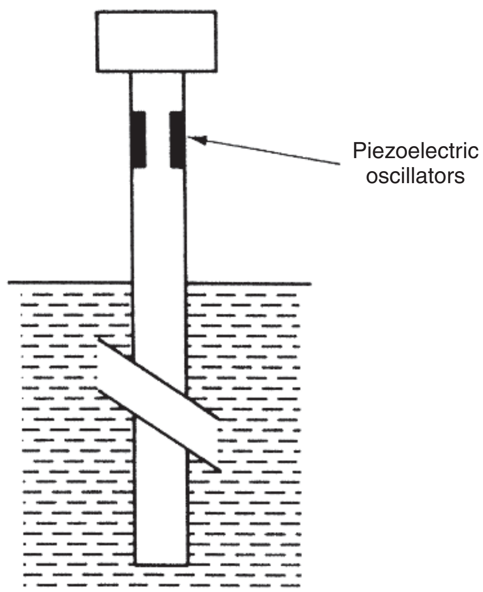


Fig. 17.8 Vibrating level sensor.

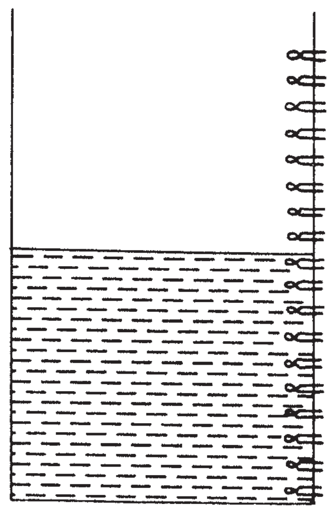


Fig. 17.9 Hot-wire-element level sensor.

17.8.3 Laser methods

One laser-based method is the *reflective level sensor*. This sensor uses light from a laser source that is reflected off the surface of the measured liquid into a line array of charge-coupled devices, as shown in Figure 17.10. Only one of these will sense light, according to the level of the liquid. An alternative, laser-based technique operates on the same general principles as the radar method described above but uses laser-generated pulses of infrared light directed at the liquid surface. This is immune to environmental conditions, and can be used with sealed vessels provided that a glass window is provided in the top of the vessel.

17.8.4 Fibre-optic level sensors

The *fibre-optic cross-talk sensor*, as described in Chapter 13, is one example of a fibre-optic sensor that can be used to measure liquid level. Another light-loss fibre-optic level sensor is the simple *loop sensor* shown in Figure 17.11. The amount of light loss depends on the proportion of cable that is submerged in the liquid. This effect is magnified if the alternative arrangement shown in Figure 17.12 is used, where light is reflected from an input fibre, round a prism, and then into an output fibre. Light is lost from this path into the liquid according to the depth of liquid surrounding the prism.

17.8.5 Thermography

Thermal imaging instruments, as discussed in Chapter 14, are a further means of detecting the level of liquids in tanks. Such instruments are capable of discriminating

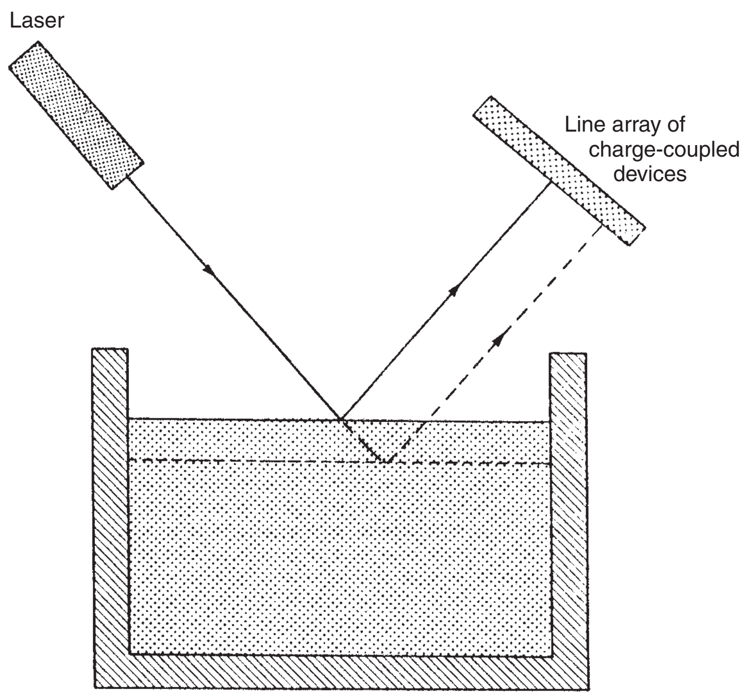


Fig. 17.10 Reflective level sensor.

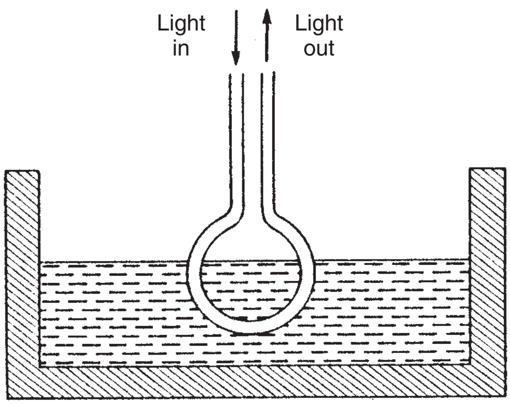


Fig. 17.11 Loop level sensor.

temperature differences as small as 0.1°C. Differences of this magnitude will normally be present at the interface between the liquid, which tends to remain at a constant temperature, and the air above, which constantly fluctuates in temperature by small amounts. The upper level of solids stored in hoppers is often detectable on the same principles.

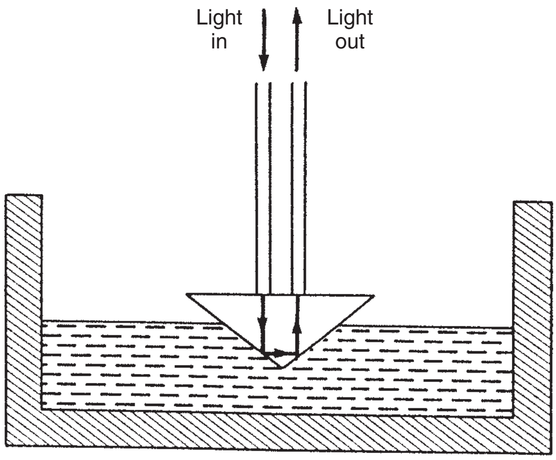


Fig. 17.12 Prism level sensor.

17.9 Intelligent level-measuring instruments

Most types of level gauge are now available in intelligent form. The pressure-measuring devices (section 17.3) are obvious candidates for inclusion within intelligent level-measuring instruments, and versions claiming $\pm 0.05\%$ accuracy are now on the market. Such instruments can also carry out additional functions, such as providing automatic compensation for liquid density variations. Microprocessors are also used to simplify installation and set-up procedures.

17.10 Choice between different level sensors

Two separate classes of level sensors can be distinguished according to whether they make contact or not with the material whose level is being measured. Contact devices are less reliable for a number of reasons, and therefore non-contact devices such as radar, laser, radiation or ultrasonic devices are preferred when there is a particular need for high reliability. According to the application, sensors that are relatively unaffected by changes in the temperature, composition, moisture content or density of the measured material may be preferred. In these respects, radar (microwave) and radiation sensors have the best immunity to such changes. Further guidance can be found in Liptak, (1995).

References and further reading

Liptak, B.G. (1995) *Instrument Engineers Handbook: Process Measurement and Analysis*, Chilton, Pennsylvania.

18

Mass, force and torque measurement

18.1 Mass (weight) measurement

Mass describes the quantity of matter that a body contains. Load cells are the most common instrument used to measure mass, especially in industrial applications. Most load cells are now electronic, although pneumatic and hydraulic types also exist. The alternatives to load cells are either mass-balance instruments or the spring balance.

18.1.1 Electronic load cell (electronic balance)

In an electronic load cell, the gravitational force on the body being measured is applied to an elastic element. This deflects according to the magnitude of the body mass. Mass measurement is thereby translated into a displacement measurement task. Electronic load cells have significant advantages over most other forms of mass-measuring instrument in terms of their relatively low cost, wide measurement range, tolerance of dusty and corrosive environments, remote measurement capability, tolerance of shock loading and ease of installation. The electronic load cell uses the physical principle that a force applied to an elastic element produces a measurable deflection. The elastic elements used are specially shaped and designed, some examples of which are shown in Figure 18.1. The design aims are to obtain a linear output relationship between the applied force and the measured deflection and to make the instrument insensitive to forces that are not applied directly along the sensing axis. Load cells exist in both compression and tension forms. In the compression type, the measured mass is placed on top of a platform resting on the load cell, which therefore compresses the cell. In the alternative tension type, the mass is hung from the load cell, thereby putting the cell into tension.

One problem that can affect the performance of load cells is the phenomenon of creep. Creep describes the permanent deformation that an elastic element undergoes after it has been under load for a period of time. This can lead to significant measurement errors in the form of a bias on all readings if the instrument is not recalibrated from time to time. However, careful design and choice of materials can largely eliminate the problem.

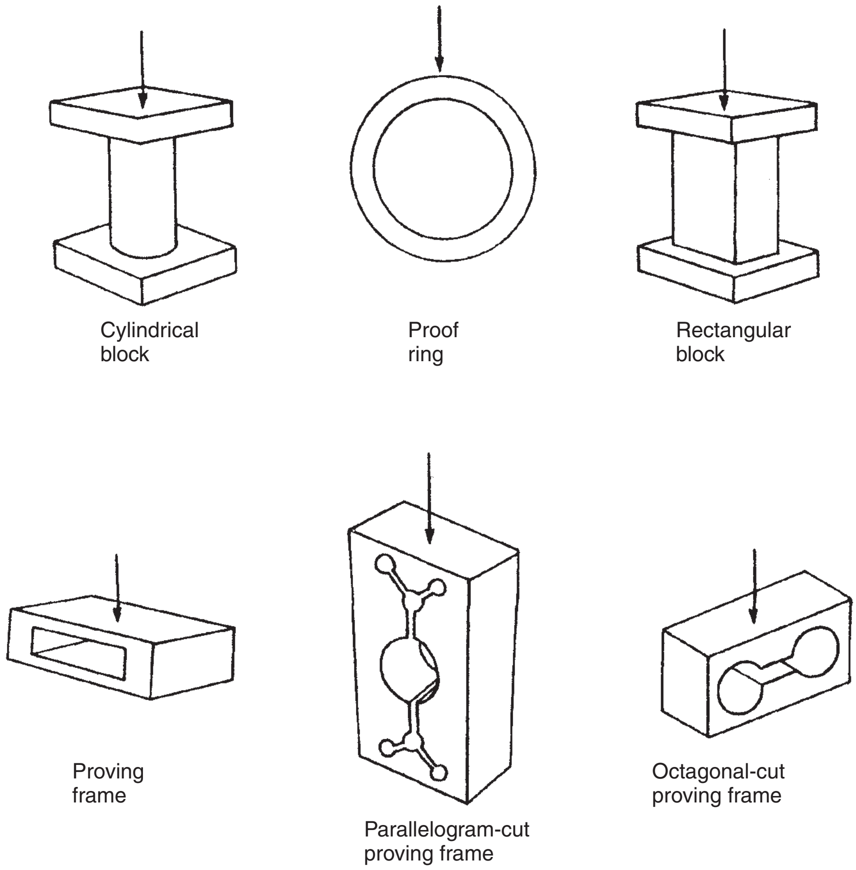


Fig. 18.1 Elastic elements used in load cells.

Various types of displacement transducer are used to measure the deflection of the elastic elements. Of these, the strain gauge is used most commonly, since this gives the best measurement accuracy, with an inaccuracy figure less than $\pm 0.05\%$ of full-scale reading being obtainable. Load cells including strain gauges are used to measure masses over a very wide range between 0 and 3000 tonnes. The measurement capability of an individual instrument designed to measure masses at the bottom end of this range would typically be 0.1–5 kg, whereas instruments designed for the top of the range would have a typical measurement span of 10–3000 tonnes.

Elastic force transducers based on differential transformers (LVDTs) to measure defections are used to measure masses up to 25 tonnes. Apart from having a lower maximum measuring capability, they are also inferior to strain gauge-based instruments in terms of their $\pm 0.2\%$ inaccuracy figure. Their major advantage is their longevity and almost total lack of maintenance requirements.

The final type of displacement transducer used in this class of instrument is the piezoelectric device. Such instruments are used to measure masses in the range 0 to 1000 tonnes. Piezoelectric crystals replace the specially designed elastic member

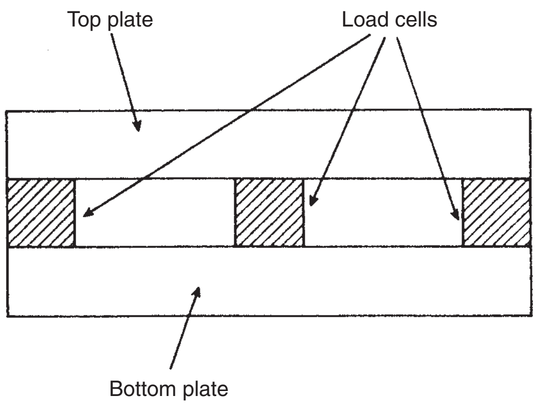


Fig. 18.2 Load-cell-based electronic balance.

normally used in this class of instrument, allowing the device to be physically small. As discussed previously, such devices can only measure dynamically changing forces because the output reading results from an induced electrical charge whose magnitude leaks away with time. The fact that the elastic element consists of the piezoelectric crystal means that it is very difficult to design such instruments to be insensitive to forces applied at an angle to the sensing axis. Therefore, special precautions have to be taken in applying these devices. Although such instruments are relatively cheap, their lowest inaccuracy is $\pm 1\%$ of full-scale reading, and they also have a high temperature coefficient.

The *electronic balance* is a device that contains several compression-type load cells, as illustrated in Figure 18.2. Commonly, either three or four load cells are used in the balance, with the output mass measurement being formed from the sum of the outputs of each cell. Where appropriate, the upper platform can be replaced by a tank for weighing liquids, powders etc.

18.1.2 Pneumatic/hydraulic load cells

Pneumatic and hydraulic load cells translate mass measurement into a pressure measurement task. A pneumatic load cell is shown schematically in Figure 18.3. Application of a mass to the cell causes deflection of a diaphragm acting as a variable restriction in a nozzle–flapper mechanism. The output pressure measured in the cell is approximately proportional to the magnitude of the gravitational force on the applied mass. The instrument requires a flow of air at its input of around $0.25\,m^3$/hour at a pressure of 4 bar. Standard cells are available to measure a wide range of masses. For measuring small masses, instruments are available with a full-scale reading of 25 kg, whilst at the top of the range, instruments with a full-scale reading of 25 tonnes are obtainable. Inaccuracy is typically $\pm 0.5\%$ of full scale in pneumatic load cells.

The alternative, hydraulic load cell is shown in Figure 18.4. In this, the gravitational force due to the unknown mass is applied, via a diaphragm, to oil contained within an enclosed chamber. The corresponding increase in oil pressure is measured by a suitable pressure transducer. These instruments are designed for measuring much larger masses

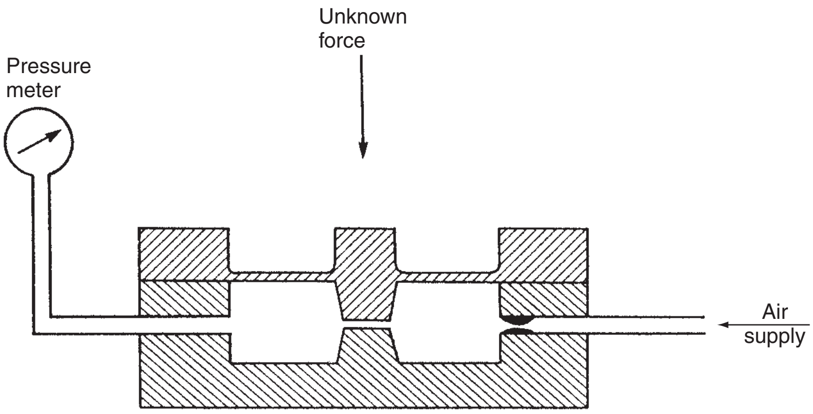


Fig. 18.3 Pneumatic load cell.

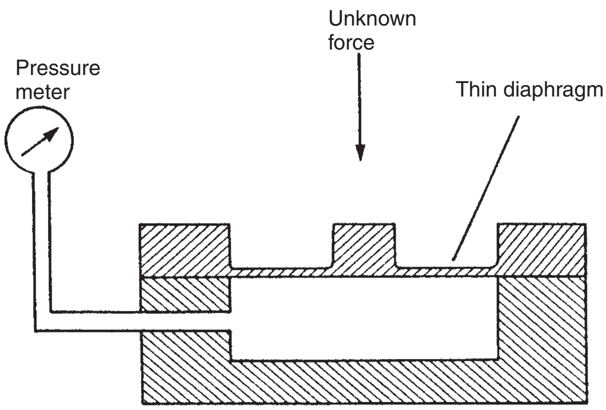


Fig. 18.4 Hydraulic load cell.

than pneumatic cells, with a load capacity of 500 tonnes being common. Special units can be obtained to measure masses as large as 50 000 tonnes. Besides their much greater measuring range, hydraulic load cells are much more accurate than pneumatic cells, with an inaccuracy figure of ±0.05% of full scale being typical. However, in order to obtain such a level of accuracy, correction for the local value of g (acceleration due to gravity) is necessary. A measurement resolution of 0.02% is attainable.

18.1.3 Intelligent load cells

Intelligent load cells are formed by adding a microprocessor to a standard cell. This brings no improvement in accuracy because the load cell is already a very accurate device. What it does produce is an intelligent weighing system that can compute total cost from the measured weight, using stored cost per unit weight information, and provide an output in the form of a digital display. Cost per weight figures can be pre-stored for a large number of substances, making such instruments very flexible in their operation.

In applications where the mass of an object is measured by several load cells used together (for example, load cells located at the corners of a platform in an electronic balance), the total mass can be computed more readily if the individual cells have a microprocessor providing digital output. In addition, it is also possible to use significant differences in the relative readings between different load cells as a fault detection mechanism in the system.

18.1.4 Mass-balance (weighing) instruments

Mass-balance instruments are based on comparing the gravitational force on the measured mass with the gravitational force on another body of known mass. This principle of mass measurement is commonly known as *weighing*, and is used in instruments like the beam balance, weigh beam, pendulum scale and electromagnetic balance.

Beam balance (equal-arm balance)

In the beam balance, shown in Figure 18.5, standard masses are added to a pan on one side of a pivoted beam until the magnitude of the gravity force on them balances the magnitude of the gravitational force on the unknown mass acting at the other end of the beam. This equilibrium position is indicated by a pointer that moves against a calibrated scale.

Instruments of this type are capable of measuring a wide span of masses. Those at the top of the range can typically measure masses up to 1000 grams whereas those at the bottom end of the range can measure masses of less than 0.01 gram. Measurement resolution can be as good as 1 part in 10^7 of the full-scale reading if the instrument is designed and manufactured very carefully. The lowest measurement inaccuracy figure attainable is $\pm 0.002\%$.

One serious disadvantage of this type of instrument is its lack of ruggedness. Continuous use and the inevitable shock loading that will occur from time to time both cause

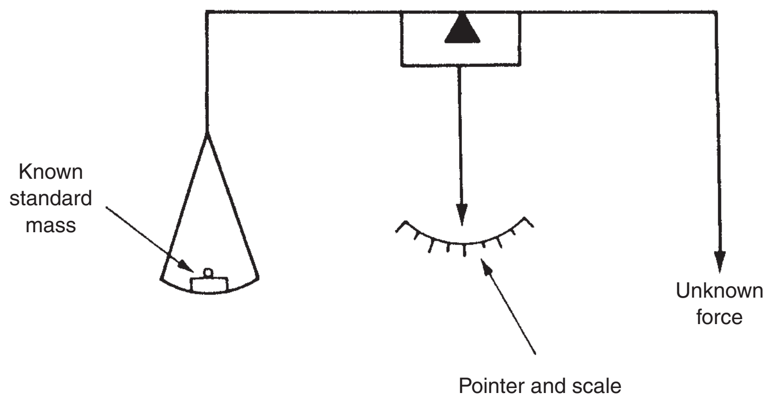


Fig. 18.5 Beam balance (equal-arm balance).

damage to the knife edges, leading to problems in measurement accuracy and resolution. A further problem in industrial use is the relatively long time needed to make each measurement. For these reasons, the beam balance is normally reserved as a calibration standard and is not used in day-to-day production environments.

Weigh beam

The weigh beam, sketched in two alternative forms in Figure 18.6, operates on similar principles to the beam balance but is much more rugged. In the first form, standard masses are added to balance the unknown mass and fine adjustment is provided by a known mass that is moved along a notched, graduated bar until the pointer is brought to the null, balance point. The alternative form has two or more graduated bars (three bars shown in Figure 18.6). Each bar carries a different standard mass and these are moved to appropriate positions on the notched bar to balance the unknown mass. Versions of these instruments are used to measure masses up to 50 tonnes.

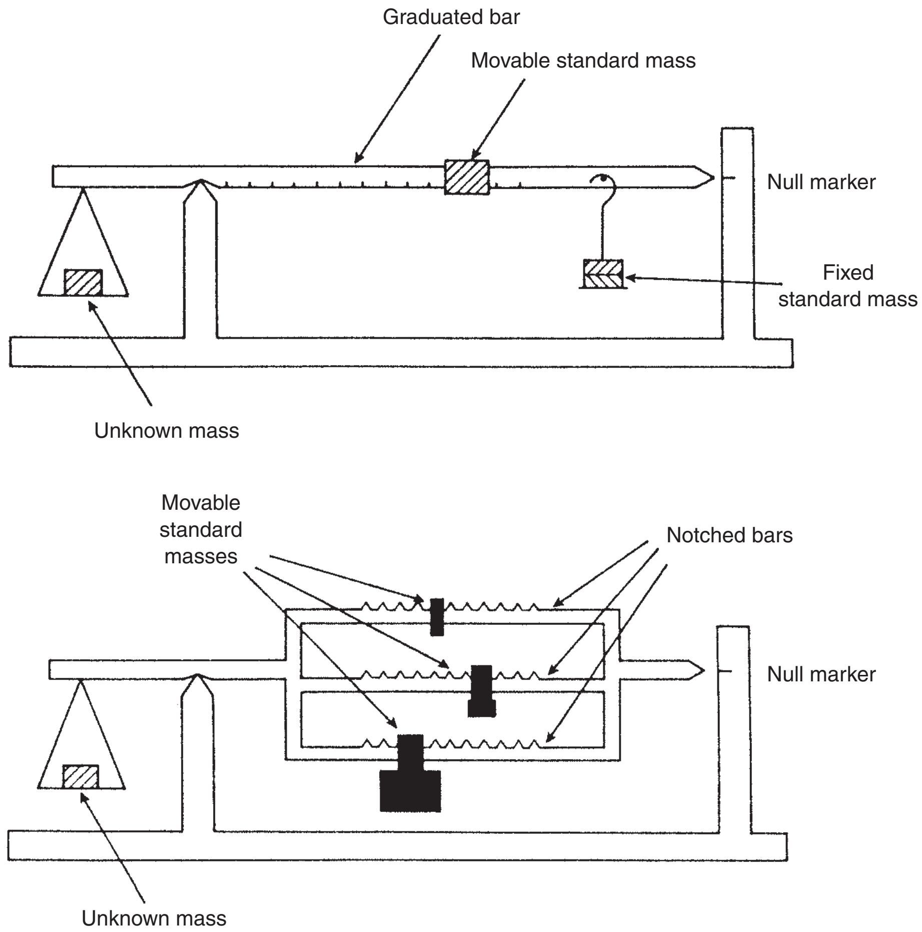


Fig. 18.6 Two alternative forms of weigh beam.

Pendulum scale

The pendulum scale, sketched in Figure 18.7, is another instrument that works on the mass-balance principle. The unknown mass is put on a platform that is attached by steel tapes to a pair of cams. Downward motion of the platform, and hence rotation of the cams, under the influence of the gravitational force on the mass, is opposed by the gravitational force acting on two pendulum type masses attached to the cams. The amount of rotation of the cams when the equilibrium position is reached is determined by the deflection of a pointer against a scale. The shape of the cams is such that this output deflection is linearly proportional to the applied mass.

This instrument is particularly useful in some applications because it is a relatively simple matter to replace the pointer and scale system by a rotational displacement transducer that gives an electrical output. Various versions of the instrument can measure masses in the range between 1 kg and 500 tonnes, with a typical measurement inaccuracy of ±0.1%.

One potential source of difficulty with the instrument is oscillation of the weigh platform when the mass is applied. Where necessary, in instruments measuring larger masses, dashpots are incorporated into the cam system to damp out such oscillations. A further possible problem can arise, mainly when measuring large masses, if the mass is not placed centrally on the platform. This can be avoided by designing a second platform to hold the mass, which is hung from the first platform by knife edges. This lessens the criticality of mass placement.

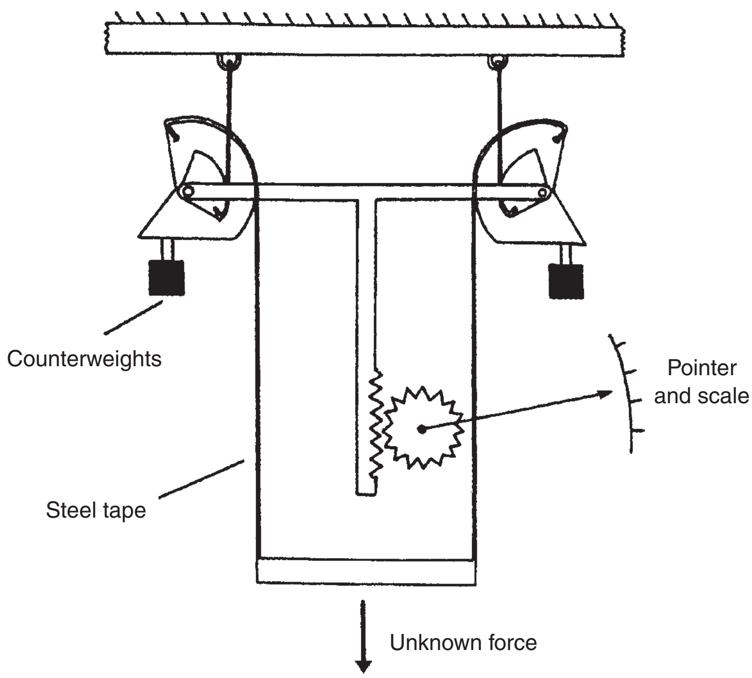


Fig. 18.7 Pendulum scale.

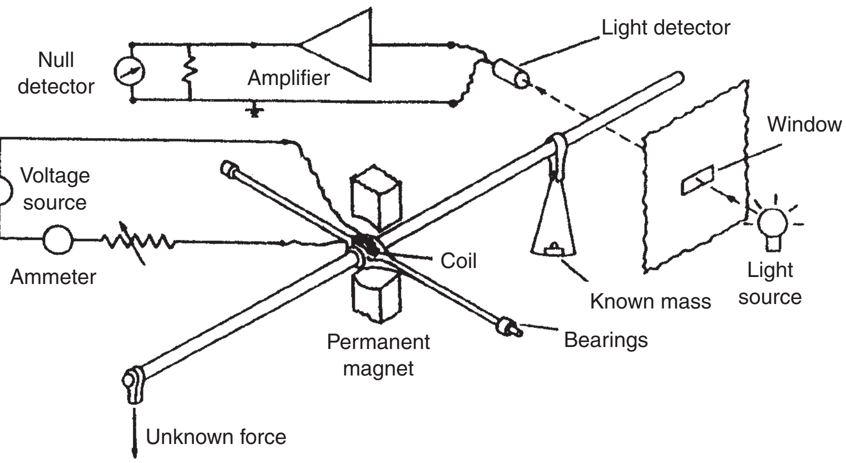


Fig. 18.8 Electromagnetic balance.

Electromagnetic balance

The electromagnetic balance uses the torque developed by a current-carrying coil suspended in a permanent magnetic field to balance the unknown mass against the known gravitational force produced on a standard mass, as shown in Figure 18.8. A light source and detector system is used to determine the null balance point. The voltage output from the light detector is amplified and applied to the coil, thus creating a servosystem where the deflection of the coil in equilibrium is proportional to the applied force. Its advantages over beam balances, weigh beams and pendulum scales include its smaller size, its insensitivity to environmental changes (modifying inputs) and its electrical form of output.

18.1.5 Spring balance

Spring balances provide a method of mass measurement that is both simple and cheap. The mass is hung on the end of a spring and the deflection of the spring due to the downwards gravitational force on the mass is measured against a scale. Because the characteristics of the spring are very susceptible to environmental changes, measurement accuracy is usually relatively poor. However, if compensation is made for the changes in spring characteristics, then a measurement inaccuracy less than $\pm 0.2\%$ is achievable. According to the design of the instrument, masses between 0.5 kg and 10 tonnes can be measured.

18.2 Force measurement

If a force of magnitude, F, is applied to a body of mass, M, the body will accelerate at a rate, A, according to the equation:

$$F = MA$$

The standard unit of force is the *Newton*, this being the force that will produce an acceleration of one metre per second squared in the direction of the force when it is applied to a mass of one kilogram. One way of measuring an unknown force is therefore to measure the acceleration when it is applied to a body of known mass. An alternative technique is to measure the variation in the resonant frequency of a vibrating wire as it is tensioned by an applied force.

18.2.1 Use of accelerometers

The technique of applying a force to a known mass and measuring the acceleration produced can be carried out using any type of accelerometer. Unfortunately, the method is of very limited practical value because, in most cases, forces are not free entities but are part of a system (from which they cannot be decoupled) in which they are acting on some body that is not free to accelerate. However, the technique can be of use in measuring some transient forces, and also for calibrating the forces produced by thrust motors in space vehicles.

18.2.2 Vibrating wire sensor

This instrument, illustrated in Figure 18.9, consists of a wire that is kept vibrating at its resonant frequency by a variable-frequency oscillator. The resonant frequency of a wire under tension is given by:

$$f = \frac{0.5}{L}\sqrt{\left(\frac{M}{T}\right)}$$

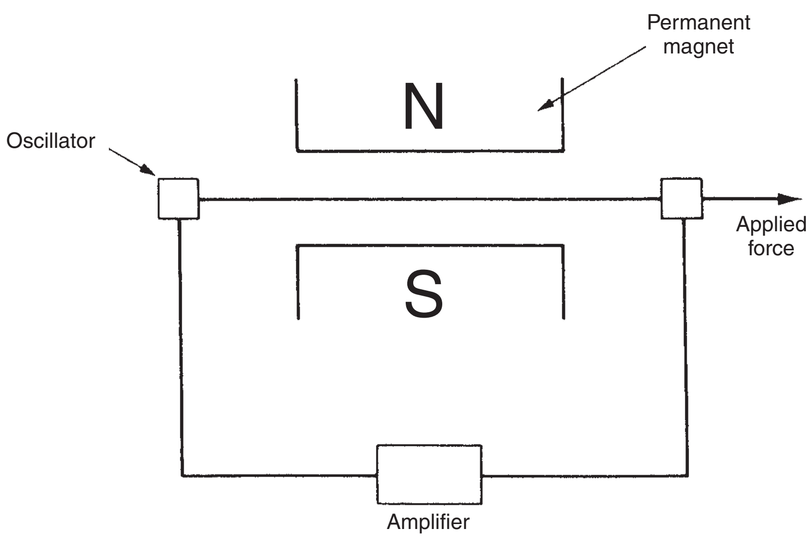


Fig. 18.9 Vibrating-wire sensor.

where M is the mass per unit length of the wire, L is the length of the wire, and T is the tension due to the applied force, F. Thus, measurement of the output frequency of the oscillator allows the force applied to the wire to be calculated.

18.3 Torque measurement

Measurement of applied torques is of fundamental importance in all rotating bodies to ensure that the design of the rotating element is adequate to prevent failure under shear stresses. Torque measurement is also a necessary part of measuring the power transmitted by rotating shafts. The three traditional methods of measuring torque consist of (i) measuring the reaction force in cradled shaft bearings, (ii) the 'Prony brake' method and (iii) measuring the strain produced in a rotating body due to an applied torque. However, recent developments in electronics and optic-fibre technology now offer an alternative method as described in paragraph 18.3.4 below.

18.3.1 Reaction forces in shaft bearings

Any system involving torque transmission through a shaft contains both a power source and a power absorber where the power is dissipated. The magnitude of the transmitted torque can be measured by cradling either the power source or the power absorber end of the shaft in bearings, and then measuring the reaction force, F, and the arm length L, as shown in Figure 18.10. The torque is then calculated as the simple product, FL. Pendulum scales are very commonly used for measuring the reaction force. Inherent errors in the method are bearing friction and windage torques.

18.3.2 Prony brake

The principle of the Prony brake is illustrated in Figure 18.11. It is used to measure the torque in a rotating shaft and consists of a rope wound round the shaft. One end of the rope is attached to a spring balance and the other end carries a load in the form of a standard mass, m. If the measured force in the spring balance is F_s, then the effective force, F_e, exerted by the rope on the shaft is given by:

$$F_e = mg - F_s$$

If the radius of the shaft is R_s and that of the rope is R_r, then the effective radius, R_e, of the rope and drum with respect to the axis of rotation of the shaft is given by:

$$R_e = R_s + R_r$$

The torque in the shaft, T, can then be calculated as:

$$T = F_e R_e$$

Whilst this is a well-known method of measuring shaft torque, a lot of heat is generated because of friction between the rope and shaft, and water cooling is usually necessary.

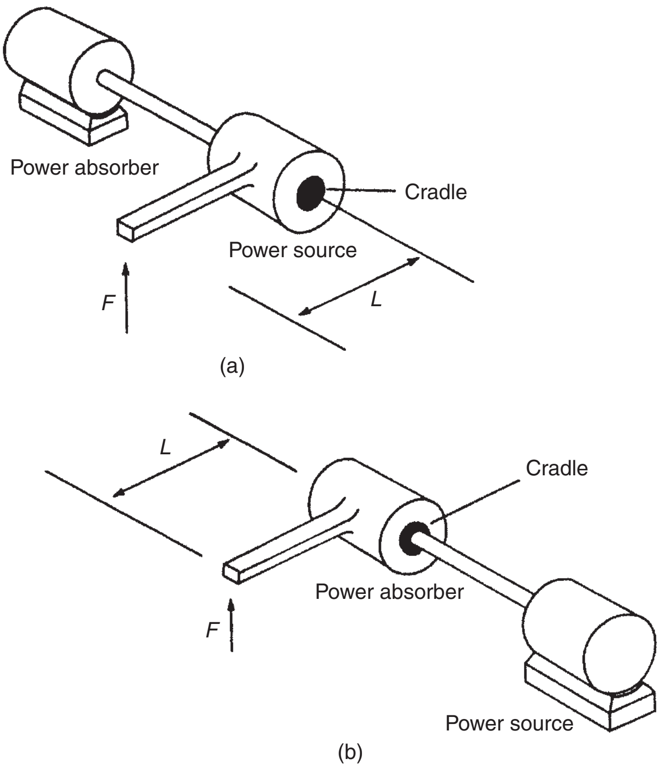


Fig. 18.10 Measuring reaction forces in cradled shaft bearings.

18.3.3 Measurement of induced strain

Measuring the strain induced in a shaft due to an applied torque has been the most common method used for torque measurement in recent years. It is a very attractive method because it does not disturb the measured system by introducing friction torques in the same way as the last two methods described do. The method involves bonding four strain gauges onto the shaft as shown in Figure 18.12, where the strain gauges are arranged in a d.c. bridge circuit. The output from the bridge circuit is a function of the strain in the shaft and hence of the torque applied. It is very important that the positioning of the strain gauges on the shaft is precise, and the difficulty in achieving this makes the instrument relatively expensive.

The technique is ideal for measuring the stalled torque in a shaft before rotation commences. However, a problem is encountered in the case of rotating shafts because a suitable method then has to be found for making the electrical connections to the strain gauges. One solution to this problem found in many commercial instruments is to use a system of slip rings and brushes for this, although this increases the cost of the instrument still further.

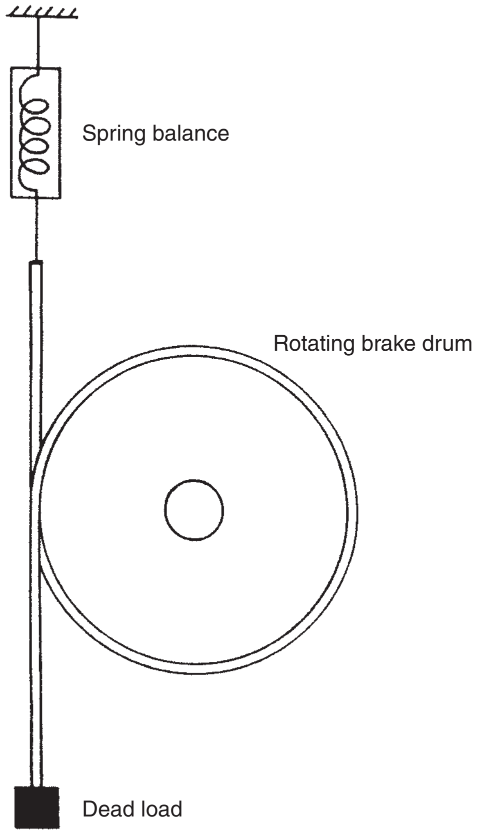


Fig. 18.11 The Prony brake.

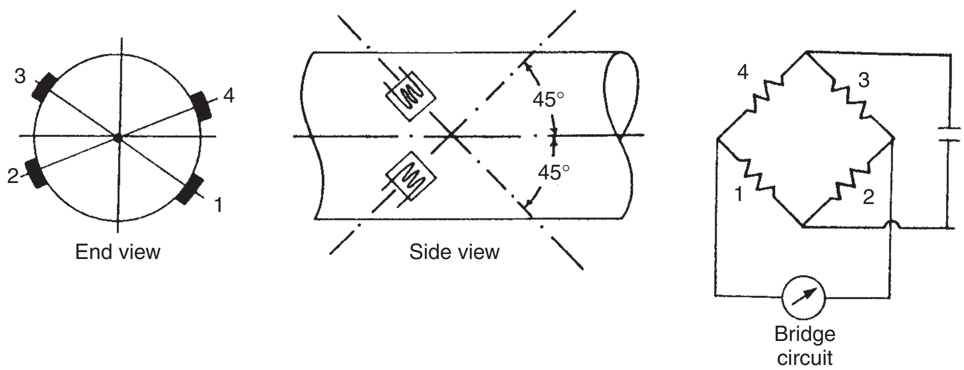


Fig. 18.12 Position of torque-measuring strain gauges on shaft.

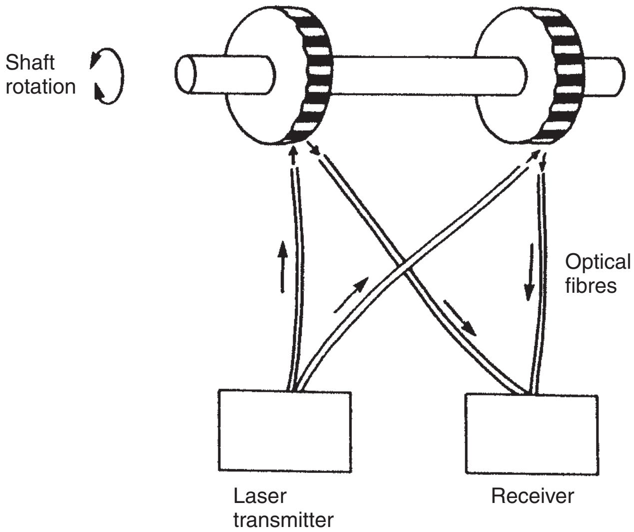


Fig. 18.13 Optical torque measurement.

18.3.4 Optical torque measurement

Optical techniques for torque measurement have become available recently with the development of laser diodes and fibre-optic light transmission systems. One such system is shown in Figure 18.13. Two black-and-white striped wheels are mounted at either end of the rotating shaft and are in alignment when no torque is applied to the shaft. Light from a laser diode light source is directed by a pair of optic-fibre cables onto the wheels. The rotation of the wheels causes pulses of reflected light and these are transmitted back to a receiver by a second pair of fibre-optic cables. Under zero torque conditions, the two pulse trains of reflected light are in phase with each other. If torque is now applied to the shaft, the reflected light is modulated. Measurement by the receiver of the phase difference between the reflected pulse trains therefore allows the magnitude of torque in the shaft to be calculated. The cost of such instruments is relatively low, and an additional advantage in many applications is their small physical size.

19

Translational motion transducers

19.1 Displacement

Translational displacement transducers are instruments that measure the motion of a body in a straight line between two points. Apart from their use as a primary transducer measuring the motion of a body, translational displacement transducers are also widely used as a secondary component in measurement systems, where some other physical quantity such as pressure, force, acceleration or temperature is translated into a translational motion by the primary measurement transducer. Many different types of translational displacement transducer exist and these, along with their relative merits and characteristics, are discussed in the following sections of this chapter. The factors governing the choice of a suitable type of instrument in any particular measurement situation are considered in the final section at the end of the chapter.

19.1.1 The resistive potentiometer

The resistive potentiometer is perhaps the best-known displacement-measuring device. It consists of a resistance element with a movable contact as shown in Figure 19.1. A voltage V_s is applied across the two ends A and B of the resistance element and an output voltage V_0 is measured between the point of contact C of the sliding element and the end of the resistance element A. A linear relationship exists between the output voltage V_0 and the distance AC, which can be expressed by:

$$\frac{V_0}{V_s} = \frac{\mathrm{AC}}{\mathrm{AB}} \tag{19.1}$$

The body whose motion is being measured is connected to the sliding element of the potentiometer, so that translational motion of the body causes a motion of equal magnitude of the slider along the resistance element and a corresponding change in the output voltage V_0.

Three different types of potentiometer exist, wire-wound, carbon-film and plastic-film, so named according to the material used to construct the resistance element. Wire-wound

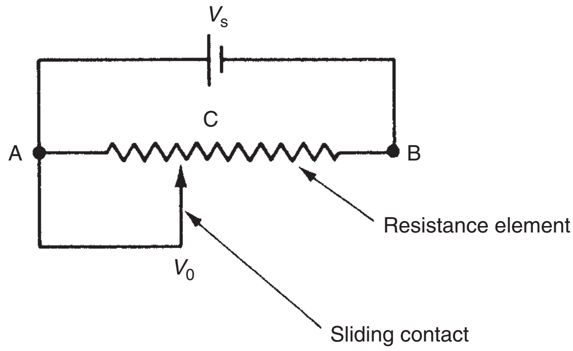


Fig. 19.1 The resistive potentiometer.

potentiometers consist of a coil of resistance wire wound on a non-conducting former. As the slider moves along the potentiometer track, it makes contact with successive turns of the wire coil. This limits the resolution of the instrument to the distance from one coil to the next. Much better measurement resolution is obtained from potentiometers using either a carbon film or a conducting plastic film for the resistance element. Theoretically, the resolution of these is limited only by the grain size of the particles in the film, suggesting that measurement resolutions up to 10^{-4} ought to be attainable. In practice, the resolution is limited by mechanical difficulties in constructing the spring system that maintains the slider in contact with the resistance track, although these types are still considerably better than wire-wound types.

Operational problems of potentiometers all occur at the point of contact between the sliding element and the resistance track. The most common problem is dirt under the slider, which increases the resistance and thereby gives a false output voltage reading, or in the worst case causes a total loss of output. High-speed motion of the slider can also cause the contact to bounce, giving an intermittent output. Friction between the slider and the track can also be a problem in some measurement systems where the body whose motion is being measured is moved by only a small force of a similar magnitude to these friction forces.

The life expectancy of potentiometers is normally quoted as a number of reversals, i.e. as the number of times the slider can be moved backwards and forwards along the track. The figures quoted for wire-wound, carbon-film and plastic-film types are respectively 1 million, 5 million and 30 million. In terms of both life expectancy and measurement resolution, therefore, the carbon and plastic film types are clearly superior, although wire-wound types do have one advantage in respect of their lower temperature coefficient. This means that wire-wound types exhibit much less variation in their characteristics in the presence of varying ambient temperature conditions.

A typical inaccuracy figure that is quoted for translational motion resistive potentiometers is $\pm 1\%$ of full-scale reading. Manufacturers produce potentiometers to cover a large span of measurement ranges. At the bottom end of this span, instruments with a range of ± 2 mm are available whilst at the top end, instruments with a range of ± 1 m are produced.

The resistance of the instrument measuring the output voltage at the potentiometer slider can affect the value of the output reading, as discussed in Chapter 3. As the slider moves along the potentiometer track, the ratio of the measured resistance to that of the measuring instrument varies, and thus the linear relationship between the measured displacement and the voltage output is distorted as well. This effect is minimized when the potentiometer resistance is small relative to that of the measuring instrument. This is achieved firstly by using a very high-impedance measuring instrument and secondly by keeping the potentiometer resistance as small as possible. Unfortunately, the latter is incompatible with achieving high measurement sensitivity since this requires a high potentiometer resistance. A compromise between these two factors is therefore necessary. The alternative strategy of obtaining high measurement sensitivity by keeping the potentiometer resistance low and increasing the excitation voltage is not possible in practice because of the power rating limitation. This restricts the allowable power loss in the potentiometer to its heat dissipation capacity.

The process of choosing the best potentiometer from a range of instruments that are available, taking into account power rating and measurement linearity considerations, is illustrated in the example below.

Example

The output voltage from a translational motion potentiometer of stroke length 0.1 metre is to be measured by an instrument whose resistance is $10\,\text{k}\Omega$. The maximum measurement error, which occurs when the slider is positioned two-thirds of the way along the element (i.e. when $\text{AC} = 2\text{AB}/3$ in Figure 19.1), must not exceed 1% of the full-scale reading. The highest possible measurement sensitivity is also required. A family of potentiometers having a power rating of 1 watt per 0.01 metre and resistances ranging from $100\,\Omega$ to $10\,\text{k}\Omega$ in $100\,\Omega$ steps is available. Choose the most suitable potentiometer from this range and calculate the sensitivity of measurement that it gives.

Solution

Referring to the labelling used in Figure 19.1, let the resistance of portion AC of the resistance element R_i and that of the whole length AB of the element be R_t. Also, let the resistance of the measuring instrument be R_m and the output voltage measured by it be V_m. When the voltage-measuring instrument is connected to the potentiometer, the net resistance across AC is the sum of two resistances in parallel (R_i and R_m) given by:

$$R_{AC} = \frac{R_i R_m}{R_i + R_m}$$

Let the excitation voltage applied across the ends AB of the potentiometer be V and the resultant current flowing between A and B be I. Then I and V are related by:

$$I = \frac{V}{R_{AC} + R_{CB}} = \frac{V}{\left[R_i R_m / R_i + R_m\right] + R_t - R_i}$$

V_m can now be calculated as:

$$V_m = IR_{AC} = \frac{VR_i R_m}{\left\{\left[R_i R_m / \left(R_i + R_m\right)\right] + R_t - R_i\right\}\left\{R_i + R_m\right\}}$$

If we express the voltage that exists across AC in the absence of the measuring instrument as V_0, then we can express the error due to the loading effect of the measuring instrument as Error $= V_0 - V_m$. From equation (19.1), $V_0 = (R_i V)/R_t$. Thus,

$$\text{Error} = V_0 - V_m = V\left(\frac{R_i}{R_t}\right)\left(\frac{R_i R_m}{\left\{\left[R_i R_m / R_i + R_m\right] + R_t - R_i\right\}\left\{R_i + R_m\right\}}\right)$$
$$\times\left(\frac{R_i^2\left(R_i - R_t\right)}{R_t\left[R_i R_t + R_m R_t - R_i^2\right]}\right) \tag{19.2}$$

Substituting $R_i = 2R_t/3$ into equation (19.2) to find the maximum error:

$$\text{Maximum error} = \frac{2R_t}{2R_t + 9R_m}$$

For a maximum error of 1%:

$$\frac{2R_t}{2R_t + 9R_m} = 0.01 \tag{19.3}$$

Substituting $R_m = 10\,000\,\Omega$ into the above expression (19.3) gives $R_t = 454\,\Omega$. The nearest resistance values in the range of potentiometers available are 400 Ω and 500 Ω. The value of 400 Ω has to be selected, as this is the only one that gives a maximum measurement error of less than 1%. The thermal rating of the potentiometers is quoted as 1 watt/0.01 m, i.e. 10 watts for the total length of 0.1 m. By Ohm's law, maximum supply voltage $= \sqrt{\text{power} \times \text{resistance}} = \sqrt{10 \times 400} = 63.25$ Volts.

Thus, the measurement sensitivity = 63.25/0.1 V/m = 632.5 V/m

19.1.2 Linear variable differential transformer (LVDT)

The linear variable differential transformer, which is commonly known by the abbreviation LVDT, consists of a transformer with a single primary winding and two secondary windings connected in the series opposing manner shown in Figure 19.2. The object whose translational displacement is to be measured is physically attached to the central iron core of the transformer, so that all motions of the body are transferred to the core. For an excitation voltage V_s given by $V_s = V_p \sin(\omega t)$, the e.m.f.s induced in the secondary windings V_a and V_b are given by:

$$V_a = K_a \sin(\omega t - \phi)\,; \quad V_b = K_b \sin(\omega t - \phi)$$

The parameters K_a and K_b depend on the amount of coupling between the respective secondary and primary windings and hence on the position of the iron core. With the core in the central position, $K_a = K_b$, and we have:

$$V_a = V_b = K \sin(\omega t - \phi)$$

Because of the series opposition mode of connection of the secondary windings, $V_0 = V_a - V_b$, and hence with the core in the central position, $V_0 = 0$. Suppose now that the core is displaced upwards (i.e. towards winding A) by a distance x. If then $K_a = K_1$ and $K_b = K_2$, we have:

$$V_0 = (K_1 - K_2)\sin(\omega t - \phi)$$

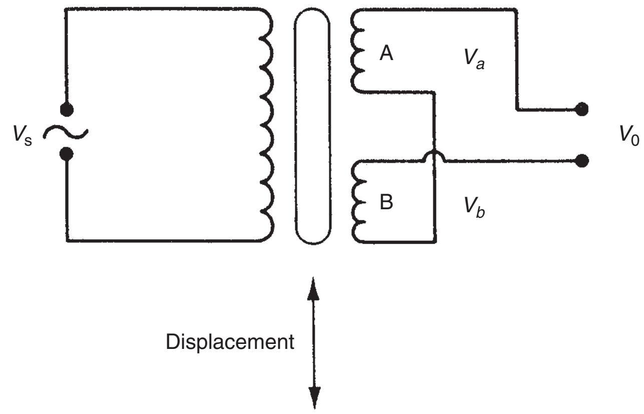


Fig. 19.2 The linear variable differential transformer (LVDT).

If, alternatively, the core were displaced downwards from the null position (i.e. towards winding B) by a distance x, the values of K_a and K_b would then be $K_a = K_2$ and $K_b = K_1$, and we would have:

$$V_0 = (K_2 - K_1)\sin(\omega t - \phi) = (K_1 - K_2)\sin(\omega t + [\pi - \phi])$$

Thus for equal magnitude displacements $+x$ and $-x$ of the core away from the central (null) position, the magnitude of the output voltage V_0 is the same in both cases. The only information about the direction of movement of the core is contained in the phase of the output voltage, which differs between the two cases by 180°. If, therefore, measurements of core position on both sides of the null position are required, it is necessary to measure the phase as well as the magnitude of the output voltage. The relationship between the magnitude of the output voltage and the core position is approximately linear over a reasonable range of movement of the core on either side of the null position, and is expressed using a constant of proportionality C as $V_0 = Cx$. The only moving part in an LVDT is the central iron core. As the core is only moving in the air gap between the windings, there is no friction or wear during operation. For this reason, the instrument is a very popular one for measuring linear displacements and has a quoted life expectancy of 200 years. The typical inaccuracy is ±0.5% of full-scale reading and measurement resolution is almost infinite. Instruments are available to measure a wide span of measurements from ±100 μm to ±100 mm. The instrument can be made suitable for operation in corrosive environments by enclosing the windings within a non-metallic barrier, which leaves the magnetic flux paths between the core and windings undisturbed. An epoxy resin is commonly used to encapsulate the coils for this purpose. One further operational advantage of the instrument is its insensitivity to mechanical shock and vibration.

Some problems that affect the accuracy of the LVDT are the presence of harmonics in the excitation voltage and stray capacitances, both of which cause a non-zero output of low magnitude when the core is in the null position. It is also impossible in practice to produce two identical secondary windings, and the small asymmetry that invariably exists between the secondary windings adds to this non-zero null output. The magnitude of this is always less than 1% of the full-scale output and in many measurement situations is of little consequence. Where necessary, the magnitude of these effects can be measured by applying known displacements to the instrument. Following this, appropriate compensation can be applied to subsequent measurements.

19.1.3 Variable capacitance transducers

Like variable inductance, the principle of variable capacitance is used in displacement measuring transducers in various ways. The three most common forms of variable capacitance transducer are shown in Figure 19.3. In Figure 19.3(a), the capacitor

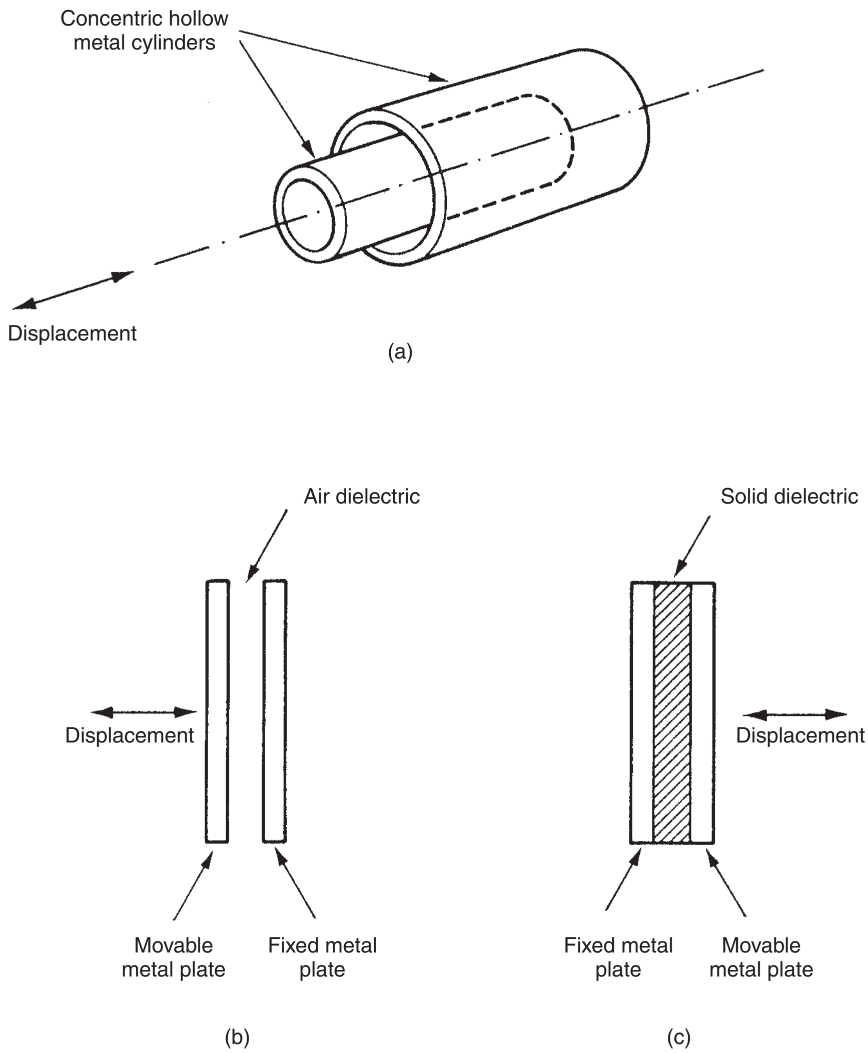


Fig. 19.3 Variable capacitance transducer.

plates are formed by two concentric, hollow, metal cylinders. The displacement to be measured is applied to the inner cylinder, which alters the capacitance. The second form, Figure 19.3(b), consists of two flat, parallel, metal plates, one of which is fixed and one of which is movable. Displacements to be measured are applied to the movable plate, and the capacitance changes as this moves. Both of these first two forms use air as the dielectric medium between the plates. The final form, Figure 19.3(c), has two flat, parallel, metal plates with a sheet of solid dielectric material between them. The displacement to be measured causes a capacitance change by moving the dielectric sheet.

Inaccuracies as low as $\pm 0.01\%$ are possible with these instruments, with measurement resolutions of 1 micron. Individual devices can be selected from manufacturers' ranges that measure displacements as small as 10^{-11} m or as large as 1 m. The fact that such instruments consist only of two simple conducting plates means that it is possible to fabricate devices that are tolerant to a wide range of environmental hazards such as extreme temperatures, radiation and corrosive atmospheres. As there are no contacting moving parts, there is no friction or wear in operation and the life expectancy quoted is 200 years. The major problem with variable capacitance transducers is their high impedance. This makes them very susceptible to noise and means that the length and position of connecting cables need to be chosen very carefully. In addition, very high impedance instruments need to be used to measure the value of the capacitance. Because of these difficulties, use of these devices tends to be limited to those few applications where the high accuracy and measurement resolution of the instrument are required.

19.1.4 Variable inductance transducers

One simple type of variable inductance transducer was shown earlier in Figure 13.4. This has a typical measurement range of 0–10 mm. An alternative form of variable inductance transducer shown in Figure 19.4(a) has a very similar size and physical appearance to the LVDT, but has a centre-tapped single winding. The two halves of the winding are connected, as shown in Figure 19.4(b), to form two arms of a bridge circuit that is excited with an alternating voltage. With the core in the central position, the output from the bridge is zero. Displacements of the core either side of the null position cause a net output voltage that is approximately proportional to the displacement for small movements of the core. Instruments in this second form are available to cover a wide span of displacement measurements. At the lower end of this span, instruments with a range of 0–2 mm are available, whilst at the top end, instruments with a range of 0–5 m can be obtained.

19.1.5 Strain gauges

The principles of strain gauges were covered earlier in Chapter 13. Because of their very small range of measurement (typically 0–50 μm), strain gauges are normally only used to measure displacements within devices like diaphragm-based pressure sensors rather than as a primary sensor in their own right for direct displacement measurement. However, strain gauges can be used to measure larger displacements if the range of

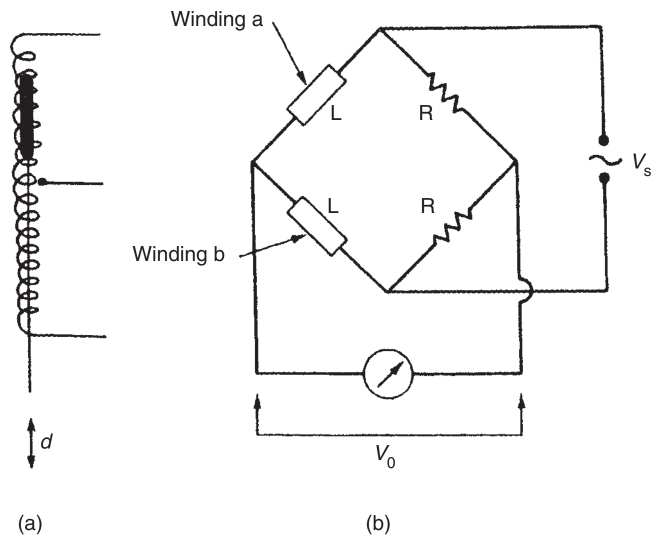


Fig. 19.4 (a) Variable inductance transducers; (b) connection in bridge circuit.

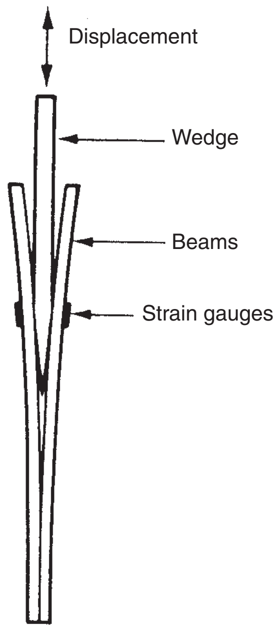


Fig. 19.5 Strain gauges measuring large displacements.

displacement measurement is extended by the scheme illustrated in Figure 19.5. In this, the displacement to be measured is applied to a wedge fixed between two beams carrying strain gauges. As the wedge is displaced downwards, the beams are forced apart and strained, causing an output reading on the strain gauges. Using this method, displacements up to about 50 mm can be measured.

19.1.6 Piezoelectric transducers

The piezoelectric transducer is effectively a force-measuring device that is used in many instruments measuring force, or the force-related quantities of pressure and acceleration. It is included within this discussion of linear displacement transducers because its mode of operation is to generate an e.m.f. that is proportional to the distance by which it is compressed. The device is manufactured from a crystal, which can be either a natural material such as quartz or a synthetic material such as lithium sulphate. The crystal is mechanically stiff (i.e. a large force is required to compress it), and consequently piezoelectric transducers can only be used to measure the displacement of mechanical systems that are stiff enough themselves to be unaffected by the stiffness of the crystal. When the crystal is compressed, a charge is generated on the surface that is measured as the output voltage. As is normal with any induced charge, the charge leaks away over a period of time. Consequently, the output voltage–time characteristic is as shown in Figure 19.6. Because of this characteristic, piezoelectric transducers are not suitable for measuring static or slowly varying displacements, even though the time constant of the charge–decay process can be lengthened by adding a shunt capacitor across the device.

As a displacement-measuring device, the piezoelectric transducer has a very high sensitivity, about one thousand times better than the strain gauge. Its typical inaccuracy is $\pm 1\%$ of full-scale reading and its life expectancy is three million reversals.

19.1.7 Nozzle flapper

The nozzle flapper is a displacement transducer that translates displacements into a pressure change. A secondary pressure-measuring device is therefore required within the instrument. The general form of a nozzle flapper is shown schematically in Figure 19.7. Fluid at a known supply pressure, P_s, flows through a fixed restriction and then through

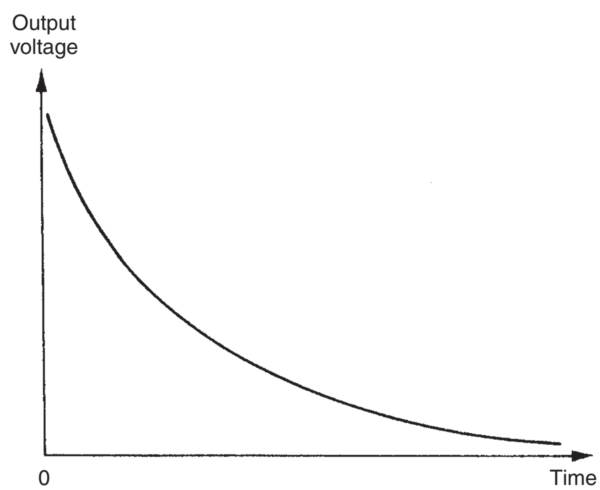


Fig. 19.6 Voltage–time characteristic of piezoelectric transducer following step displacement.

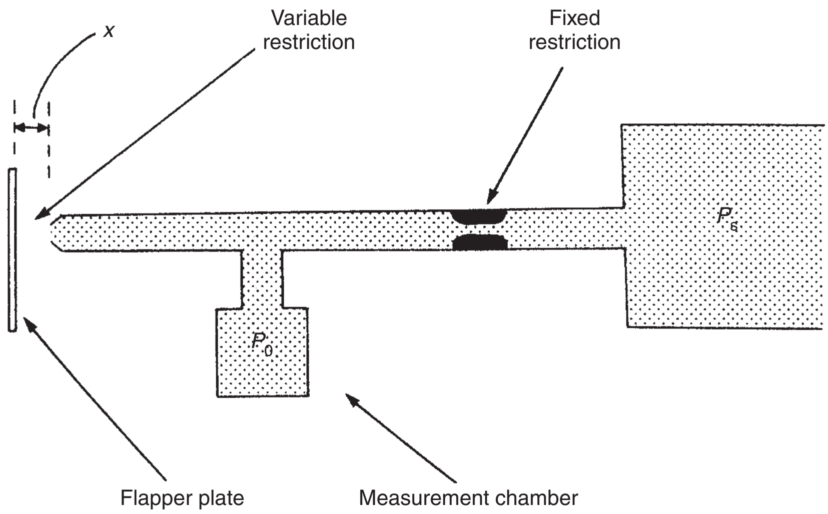


Fig. 19.7 Nozzle flapper.

a variable restriction formed by the gap, x, between the end of the main vessel and the flapper plate. The body whose displacement is being measured is connected physically to the flapper plate. The output measurement of the instrument is the pressure P_o in the chamber shown in Figure 19.7, and this is almost proportional to x over a limited range of movement of the flapper plate. The instrument typically has a first order response characteristic. Air is very commonly used as the working fluid and this gives the instrument a time constant of about 0.1 seconds. The instrument has extremely high sensitivity but its range of measurement is quite small. A typical measurement range is ± 0.05 mm with a measurement resolution of ± 0.01 µm. One very common application of nozzle flappers is measuring the displacements within a load cell, which are typically very small.

19.1.8 Other methods of measuring small displacements

Apart from the methods outlined above, several other techniques for measuring small translational displacements exist, as discussed below. Some of these involve special instruments that have a very limited sphere of application, for instance in measuring machine tool displacements. Others are very recent developments that may potentially gain wide use in the future but have few applications at present.

Linear inductosyn

The linear inductosyn is an extremely accurate instrument that is widely used for axis measurement and control within machine tools. Typical measurement resolution is 2.5 microns. The instrument consists of two magnetically coupled parts that are separated by an air gap, typically 0.125 mm wide, as shown in Figure 19.8. One part, the track, is attached to the axis along which displacements are to be measured. This would

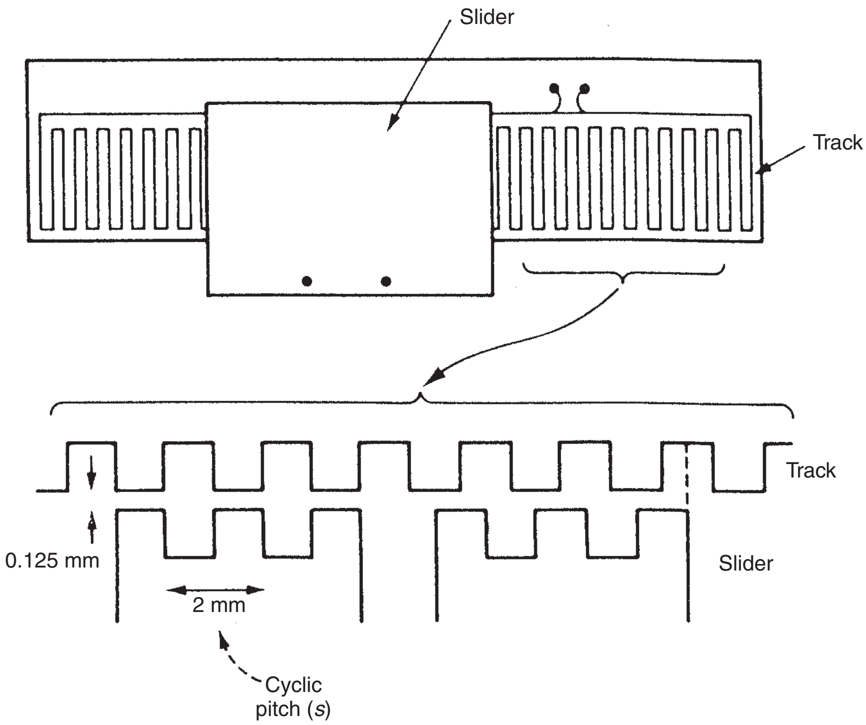


Fig. 19.8 Linear inductosyn.

generally be the bed of a machine tool. The other part, the slider, is attached to the body that is to be measured or positioned. This would usually be a cutting tool.

The track, which may be several metres long, consists of a fine metal wire formed into the pattern of a continuous rectangular waveform and deposited onto a glass base. The typical pitch (cycle length), s, of the pattern is 2 mm, and this extends over the full length of the track. The slider is usually about 50 mm wide and carries two separate wires formed into continuous rectangular waveforms that are displaced with respect to each other by one-quarter of the cycle pitch, i.e. by 90 electrical degrees. The wire waveform on the track is excited by an applied voltage given by:

$$V_s = V \sin(\omega t)$$

This excitation causes induced voltages in the slider windings. When the slider is positioned in the null position such that its first winding is aligned with the winding on the track, the output voltages on the two slider windings are given by:

$$V_1 = 0; \quad V_2 = V \sin(\omega t)$$

For any other position, the slider winding voltages are given by:

$$V_1 = V \sin(\omega t) \sin(2\pi x/s); \quad V_2 = V \sin(\omega t) \cos(2\pi x/s)$$

where x is the displacement of the slider away from the null position.

Consideration of these equations for the slider winding outputs shows that the pattern of output voltages repeats every cycle pitch. Therefore, the instrument can only discriminate displacements of the slider within one cycle pitch of the windings. This means that the typical measurement range of an inductosyn is only 2 mm. This is of no use in normal applications, and therefore an additional displacement transducer with coarser resolution but larger measurement range has to be used as well. This coarser measurement is commonly made by translating the linear displacements by suitable gearing into rotary motion, which is then measured by a rotational displacement transducer such as a synchro or resolver.

One slight problem with the inductosyn is the relatively low level of coupling between the track and slider windings. Compensation for this is made by using a high-frequency excitation voltage (5–10 kHz is common).

Translation of linear displacements into rotary motion

In some applications, it is inconvenient to measure linear displacements directly, either because there is insufficient space to mount a suitable transducer or because it is inconvenient for other reasons. A suitable solution in such cases is to translate the translational motion into rotational motion by suitable gearing. Any of the rotational displacement transducers discussed in Chapter 20 can then be applied.

Integration of output from velocity transducers and accelerometers

If velocity transducers or accelerometers already exist in a system, displacement measurements can be obtained by integration of the output from these instruments. This, however, only gives information about the relative position with respect to some arbitrary starting point. It does not yield a measurement of the absolute position of a body in space unless all motions away from a fixed starting point are recorded.

Laser interferometer

This recently developed instrument is shown in Figure 19.9. In this particular design, a dual-frequency helium–neon (He–Ne) laser is used that gives an output pair of light waves at a nominal frequency of 5×10^{14} Hz. The two waves differ in frequency by 2×10^{6} Hz and have opposite polarization. This dual-frequency output waveform is split into a measurement beam and a reference beam by the first beam splitter.

The reference beam is sensed by the polarizer and photodetector, A, which converts both waves in the light to the same polarization. The two waves interfere constructively and destructively alternately, producing light–dark flicker at a frequency of 2×10^{6} Hz. This excites a 2 MHz electrical signal in the photodetector.

The measurement beam is separated into the two component frequencies by a polarizing beam splitter. Light of the first frequency, f_1, is reflected by a fixed reflecting cube into a photodetector and polarizer, B. Light of the second frequency, f_2, is reflected by a movable reflecting cube and also enters B. The displacement to be measured is applied to the movable cube. With the movable cube in the null position, the light waves entering B produce an electrical signal output at a frequency of 2 MHz, which is the same frequency as the reference signal output from A. Any displacement of the movable cube causes a Doppler shift in the frequency f_2 and changes the output from B. The frequency of the output signal from B varies between 0.5 MHz and 3.5 MHz according to the speed and direction of movement of the movable cube.

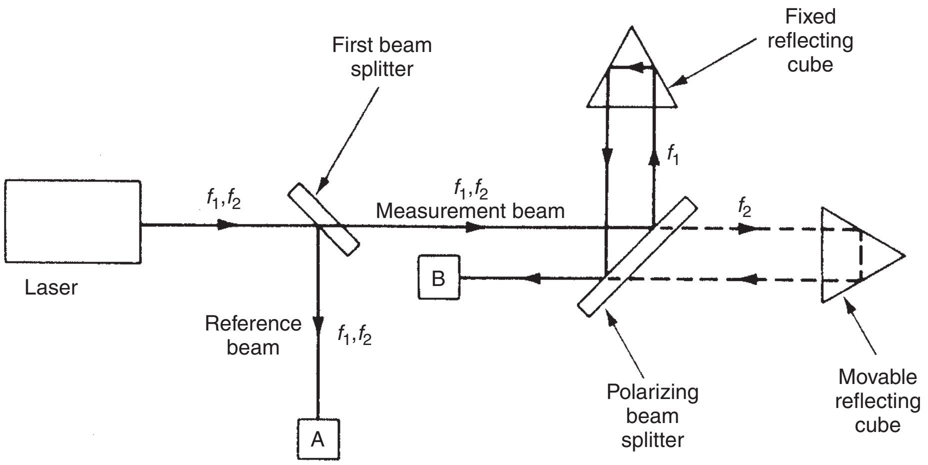


Fig. 19.9 Laser interferometer.

The outputs from A and B are amplified and subtracted. The resultant signal is fed to a counter whose output indicates the magnitude of the displacement in the movable cube and whose rate of change indicates the velocity of motion.

This technique is used in applications requiring high-accuracy measurement, such as machine tool control. Such systems can measure displacements over ranges of up to 2 m with an inaccuracy of only a few parts per million. They are therefore an attractive alternative to the inductosyn, in having both high measurement resolution and a large measurement range within one instrument.

Fotonic sensor

The Fotonic sensor is one of many instruments developed recently that make use of fibre-optic techniques. It consists of a light source, a light detector, a fibre-optic light transmission system and a plate that moves with the body whose displacement is being measured, as shown in Figure 19.10. Light from the outward fibre-optic cable travels across the air gap to the plate and some of it is reflected back into the return fibre-optic

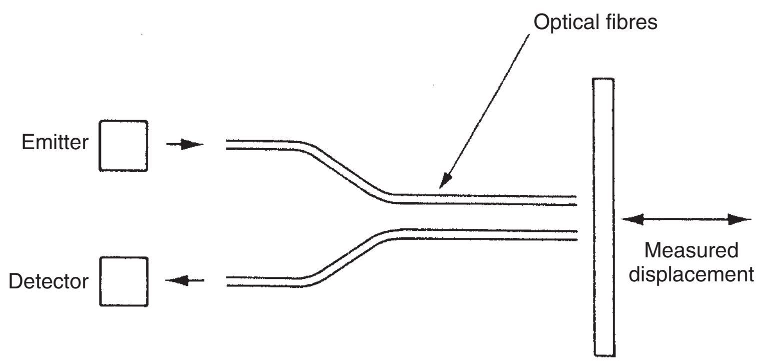


Fig. 19.10 Fotonic sensor.

cable. The amount of light reflected back from the plate is a function of the air gap length, x, and hence of the plate displacement. Measurement of the intensity of the light carried back along the return cable to the light detector allows the displacement of the plate to be calculated. Common applications of Fotonic sensors are measuring diaphragm displacements in pressure sensors and measuring the movement of bimetallic temperature sensors.

Evanescent-field fibre-optic sensors

This sensor consists of a prism and a light source/detector system, as shown in Figure 19.11. The amount of light reflected into the detector depends on the proximity of a movable silver surface to the prism. Reflection varies from 96% when the surface is touching the prism to zero when it is 1 µm away. This provides a means of measuring very tiny displacements over the range between 0 and 1 µm (1 micron).

Non-contacting optical sensor

Figure 19.12 shows an optical technique that is used to measure small displacements. The motion to be measured is applied to a vane, whose displacement progressively shades one of a pair of monolithic photodiodes that are exposed to infrared radiation. A displacement measurement is obtained by comparing the output of the reference (unshaded) photodiode with that of the shaded one. The typical range of measurement is ±0.5 mm with an inaccuracy of ±0.1% of full scale. Such sensors are used in some intelligent pressure-measuring instruments based on Bourdon tubes or diaphragms as described in Chapter 15.

19.1.9 Measurement of large displacements (range sensors)

One final class of instruments that has not been mentioned so far consists of those designed to measure relatively large translational displacements. Most of these are known as range sensors and measure the motion of a body with respect to some fixed datum point.

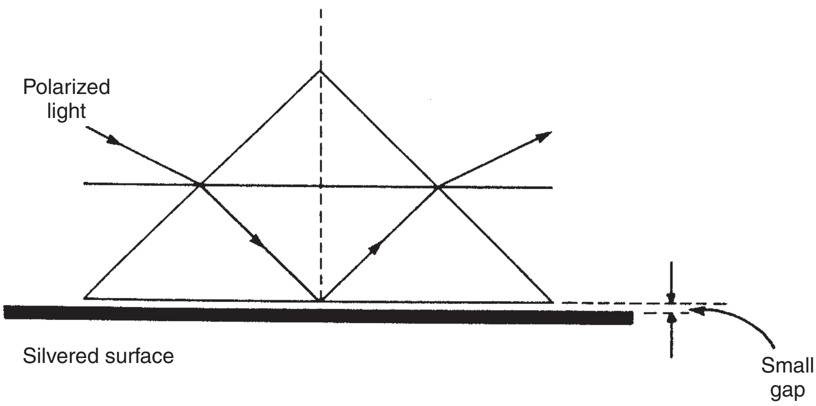


Fig. 19.11 Evanescent field fibre-optic displacement sensor.

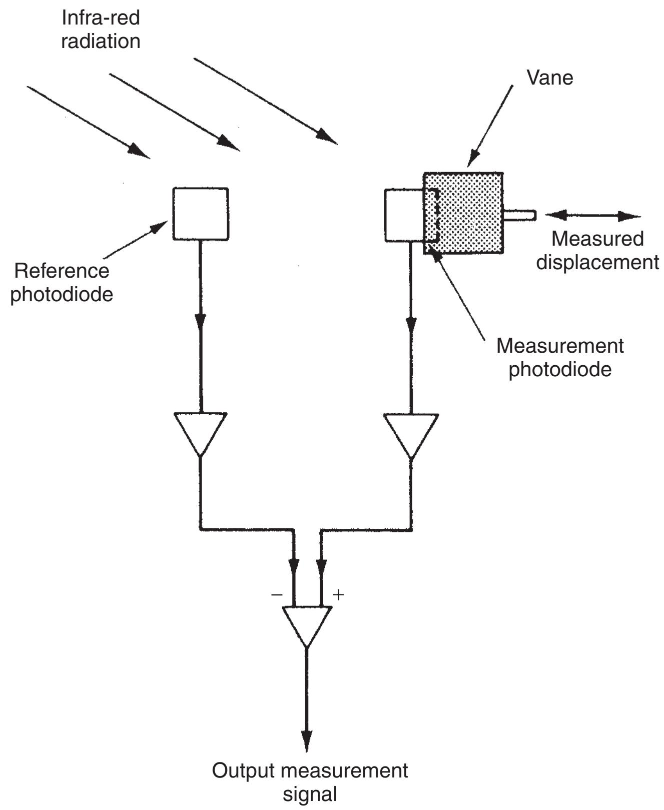


Fig. 19.12 Non-contacting optical sensor.

Rotary potentiometer and spring-loaded drum

One scheme for measuring large displacements that are beyond the measurement range of common displacement transducers is shown in Figure 19.13. This consists of a steel wire attached to the body whose displacement is being measured: the wire passes round a pulley and on to a spring-loaded drum whose rotation is measured by a rotary potentiometer. A multi-turn potentiometer is usually required for this to give an adequate measurement resolution. With this measurement system, it is possible to reduce measurement uncertainty to as little as $\pm 0.01\%$ of full-scale reading.

Range sensors

Range sensors provide a well-used technique of measuring the translational displacement of a body with respect to some fixed boundary. The common feature of all range sensing systems is an energy source, an energy detector and an electronic means of timing the time of flight of the energy between the source and detector. The form of energy used is either ultrasonic or light. In some systems, both energy source and detector are fixed on the moving body and operation depends on the energy being reflected back from the fixed boundary as in Figure 19.14(a). In other systems, the energy source is attached to the moving body and the energy detector is located within the fixed boundary, as shown in Figure 19.14(b).

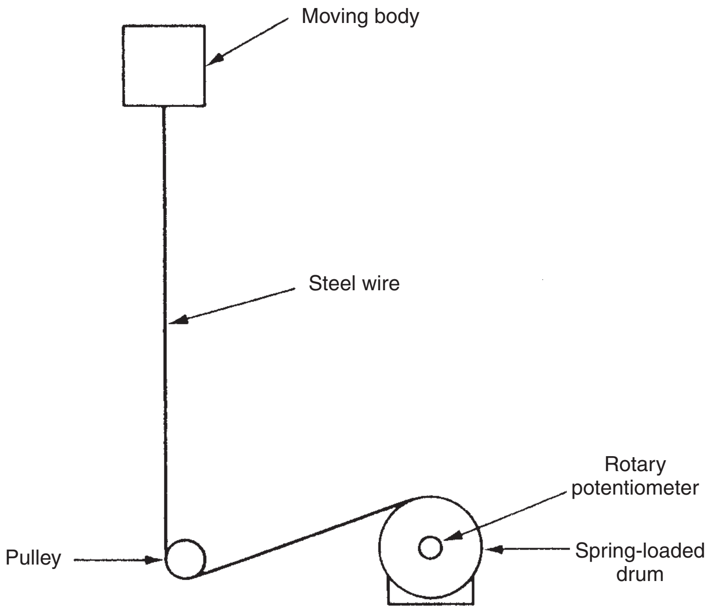


Fig. 19.13 System for measuring large displacements.

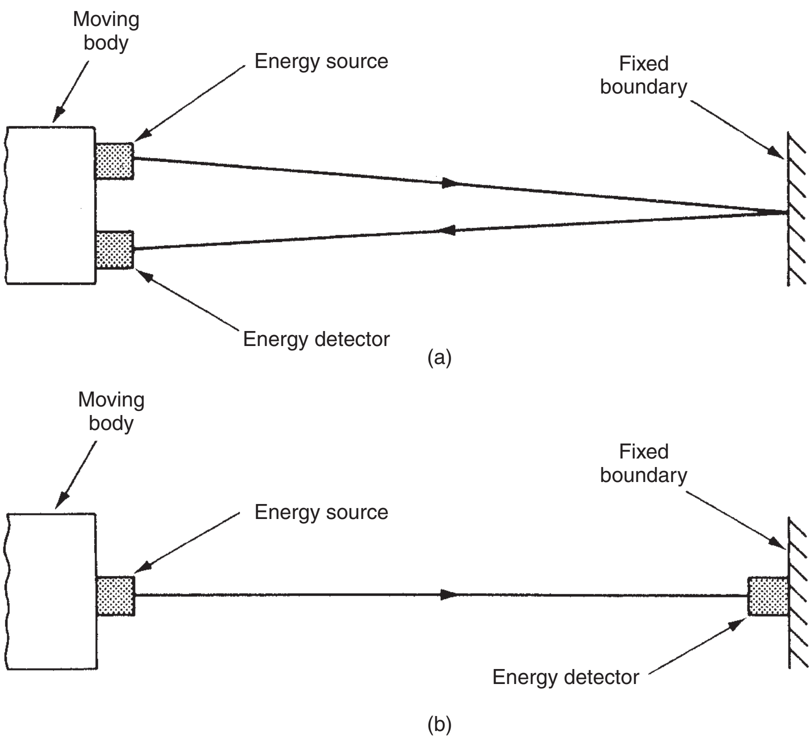


Fig. 19.14 Range sensors.

In ultrasonic systems, the energy is transmitted from the source in high-frequency bursts. A frequency of at least 20 kHz is usual, and 40 kHz is common for measuring distances up to 5 m. By measuring the time of flight of the energy, the distance of the body from the fixed boundary can be calculated, using the fact that the speed of sound in air is 340 m/s. Because of difficulties in measuring the time of flight with sufficient accuracy, ultrasonic systems are not suitable for measuring distances of less than about 300 mm. Measurement resolution is limited by the wavelength of the ultrasonic energy and can be improved by operating at higher frequencies. At higher frequencies, however, attenuation of the magnitude of the ultrasonic wave as it passes through air becomes significant. Therefore, only low frequencies are suitable if large distances are to be measured. The typical inaccuracy of ultrasonic range finding systems is $\pm 0.5\%$ of full scale.

Optical range finding systems generally use a laser light source. The speed of light in air is about 3×10^8 m/s, so that light takes only a few nanoseconds to travel a metre. In consequence, such systems are only suitable for measuring very large displacements where the time of flight is long enough to be measured with reasonable accuracy.

19.1.10 Proximity sensors

For the sake of completeness, it is proper to conclude this chapter on translational displacement transducers with consideration of proximity sensors. Proximity detectors provide information on the displacement of a body with respect to some boundary, but only insofar as to say whether the body is less than or greater than a certain distance away from the boundary. The output of a proximity sensor is thus binary in nature: the body is or is not close to the boundary.

Like range sensors, proximity detectors make use of an energy source and detector. The detector is a device whose output changes between two states when the magnitude of the incident reflected energy exceeds a certain threshold level. A common form of proximity sensor uses an infrared light-emitting diode (LED) source and a phototransistor. Light triggers the transistor into a conducting state when the LED is within a certain distance from a reflective boundary and the reflected light exceeds a threshold level. This system is physically small, occupying a volume of only a few cubic centimetres. If even this small volume is obtrusive, then fibre-optic cables can be used to transmit light from a remotely mounted LED and phototransistor. The threshold displacement detected by optical proximity sensors can be varied between 0 and 2 m.

Another form of proximity sensor uses the principle of varying inductance. Such devices are particularly suitable for operation in aggressive environmental conditions and they can be made vibration and shock resistant by vacuum encapsulation techniques. The sensor contains a high-frequency oscillator whose output is demodulated and fed via a trigger circuit to an amplifier output stage. The oscillator output radiates through the surface of the sensor and, when the sensor surface becomes close to an electrically or magnetically conductive boundary, the output voltage is reduced because of the interference with the flux paths. At a certain point, the output voltage is reduced sufficiently for the trigger circuit to change state and reduce the amplifier output to zero. Inductive sensors can be adjusted to change state at displacements in the range of 1 to 20 mm.

A third form of proximity sensor uses the capacitive principle. These can operate in similar conditions to inductive types. The threshold level of displacement detected can be varied between 5 and 40 mm.

Fibre-optic proximity sensors also exist where the amount of reflected light varies with the proximity of the fibre ends to a boundary, as shown in Figure 13.2(c).

19.1.11 Selection of translational measurement transducers

Choice between the various translational motion transducers available for any particular application depends mainly on the magnitude of the displacement to be measured, although the operating environment is also relevant. Displacements larger than five metres can only be measured by a range sensor, or possibly by the method of using a wire described in section 19.1.9 (Figure 19.14). Such methods are also used for displacements in the range between 2 and 5 metres, except where the expense of a variable inductance transducer can be justified.

For measurements within the range of 2 mm to 2 m, the number of suitable instruments grows. Both the relatively cheap potentiometer and the LVDT, which is somewhat more expensive, are commonly used for such measurements. Variable-inductance and variable-capacitance transducers are also used in some applications. Additionally, strain gauges measuring the strain in two beams forced apart by a wedge (see section 19.1.5) can measure displacements up to 50 mm. If very high measurement resolution is required, either the linear inductosyn or the laser interferometer is used.

The requirement to measure displacements of less than 2 mm usually occurs as part of an instrument that is measuring some other physical quantity such as pressure, and several types of device have evolved to fulfil this task. The LVDT, strain gauges, the Fotonic sensor, variable-capacitance transducers and the non-contacting optical transducer all find application in measuring diaphragm or Bourdon-tube displacements within pressure transducers. Load cell displacements are also very small, and these are commonly measured by nozzle flapper devices.

If the environmental operating conditions are severe (for example, hot, radioactive or corrosive atmospheres), devices that can be easily protected from these conditions must be chosen, such as the LVDT, variable inductance and variable capacitance instruments.

19.2 Velocity

Translational velocity cannot be measured directly and therefore must be calculated indirectly by other means as set out below.

19.2.1 Differentiation of displacement measurements

Differentiation of position measurements obtained from any of the translational displacement transducers described in section 19.1 can be used to produce a translational velocity signal. Unfortunately, the process of differentiation always amplifies noise in a measurement system. Therefore, if this method has to be used, a low-noise instrument such as a d.c. excited carbon film potentiometer or laser interferometer should be chosen. In the case of potentiometers, a.c. excitation must be avoided because of the problem that harmonics in the power supply would cause.

19.2.2 Integration of the output of an accelerometer

Where an accelerometer is already included within a system, integration of its output can be performed to yield a velocity signal. The process of integration attenuates rather than amplifies measurement noise and this is therefore an acceptable technique.

19.2.3 Conversion to rotational velocity

Conversion from translational to rotational velocity is the final measurement technique open to the system designer and is the one most commonly used. This enables any of the rotational velocity measuring instruments described in Chapter 20 to be applied.

19.3 Acceleration

The only class of device available for measuring acceleration is the accelerometer. These are available in a wide variety of types and ranges designed to meet particular measurement requirements. They have a frequency response between zero and a high value, and have a form of output that can be readily integrated to give displacement and velocity measurements. The frequency response of accelerometers can be improved by altering the level of damping in the instrument. Such adjustment must be done carefully, however, because frequency response improvements are only achieved at the expense of degrading the measurement sensitivity. Besides their use for general-purpose motion measurement, accelerometers are widely used to measure mechanical shocks and vibrations.

Most forms of accelerometer consist of a mass suspended by a spring and damper inside a housing, as shown in Figure 19.15. The accelerometer is rigidly fastened to the body undergoing acceleration. Any acceleration of the body causes a force, F_a, on the mass, M, given by:

$$F_a = M\ddot{x}$$

This force is opposed by the restraining effect, F_s, of a spring with spring constant K, and the net result is that the mass is displaced by a distance x from its starting position such that:

$$F_s = Kx$$

In steady state, when the mass inside is accelerating at the same rate as the case of the accelerometer, $F_a = F_s$ and so:

$$Kx = M\ddot{x} \quad \text{or} \quad \ddot{x} = (Kx)/M \tag{19.4}$$

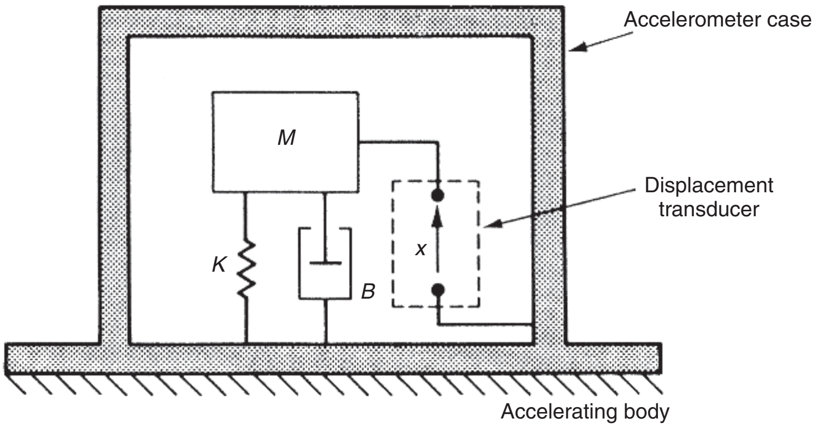


Fig. 19.15 Structure of an accelerometer.

This is the equation of motion of a second order system, and, in the absence of damping, the output of the accelerometer would consist of non-decaying oscillations. A damper is therefore included within the instrument, which produces a damping force, F_d, proportional to the velocity of the mass M given by:

$$F_d = B\dot{x}$$

This modifies the previous equation of motion (19.4) to the following:

$$Kx + B\dot{x} = M\ddot{x} \tag{19.5}$$

One important characteristic of accelerometers is their sensitivity to accelerations at right angles to the sensing axis (the direction along which the instrument is designed to measure acceleration). This is defined as the *cross-sensitivity* and is specified in terms of the output, expressed as a percentage of the full-scale output, when an acceleration of some specified magnitude (e.g. $30g$) is applied at 90° to the sensing axis.

The acceleration reading is obtained from the instrument by measurement of the displacement of the mass within the accelerometer. Many different displacement-measuring techniques are used in the various types of accelerometer that are commercially available. Different types of accelerometer also vary in terms of the type of spring element and form of damping used.

Resistive potentiometers are one such displacement-measuring instrument used in accelerometers. These are used mainly for measuring slowly varying accelerations and low-frequency vibrations in the range $0-50g$. The measurement resolution obtainable is about 1 in 400 and typical values of cross-sensitivity are $\pm 1\%$. Inaccuracy is about $\pm 1\%$ and life expectancy is quoted at two million reversals. A typical size and weight are $125\,cm^3$ and 500 grams.

Strain gauges and piezoresistive sensors are also used in accelerometers for measuring accelerations up to $200g$. These serve as the spring element as well as measuring mass displacement, thus simplifying the instrument's construction. Their typical characteristics are a resolution of 1 in 1000, inaccuracy of $\pm 1\%$ and cross-sensitivity of 2%. They have a major advantage over potentiometer-based accelerometers in terms of their much smaller size and weight ($3\,cm^3$ and 25 grams).

Another displacement transducer found in accelerometers is the LVDT. This device can measure accelerations up to $700g$ with a typical inaccuracy of $\pm 1\%$ of full scale. They are of a similar physical size to potentiometer-based instruments but are lighter in weight (100 grams).

Accelerometers based on variable inductance displacement measuring devices have extremely good characteristics and are suitable for measuring accelerations up to $40g$. Typical specifications of such instruments are inaccuracy $\pm 0.25\%$ of full scale, resolution 1 in 10 000 and cross-sensitivity 0.5%. Their physical size and weight are similar to potentiometer-based devices. Instruments with an output in the form of a varying capacitance also have similar characteristics.

The other common displacement transducer used in accelerometers is the piezoelectric type. The major advantage of using piezoelectric crystals is that they also act as the spring and damper within the instrument. In consequence, the device is quite small ($15\,cm^3$) and very low mass (50 grams), but because of the nature of piezoelectric crystal operation, such instruments are not suitable for measuring constant or slowly time-varying accelerations. As the electrical impedance of a piezoelectric crystal is itself high, the output voltage must be measured with a very high-impedance instrument to avoid loading effects. Many recent piezoelectric crystal-based accelerometers incorporate a high impedance charge amplifier within the body of the instrument. This simplifies the signal conditioning requirements external to the accelerometer but can lead to problems in certain operational environments because these internal electronics are exposed to the same environmental hazards as the rest of the accelerometer. Typical measurement resolution of this class of accelerometer is 0.1% of full scale with an inaccuracy of $\pm 1\%$. Individual instruments are available to cover a wide range of measurements from $0.03g$ full scale up to $1000g$ full scale. *Intelligent accelerometers* are also now available that give even better performance through inclusion of processing power to compensate for environmentally induced errors.

Recently, very small microsensors have become available for measuring acceleration. These consist of a small mass subject to acceleration that is mounted on a thin silicon membrane. Displacements are measured either by piezoresistors deposited on the membrane or by etching a variable capacitor plate into the membrane.

Two forms of fibre-optic-based accelerometer also exist. One form measures the effect on light transmission intensity caused by a mass subject to acceleration resting on a multimode fibre. The other form measures the change in phase of light transmitted through a monomode fibre that has a mass subject to acceleration resting on it.

19.3.1 Selection of accelerometers

In choosing between the different types of accelerometer for a particular application, the mass of the instrument is particularly important. This should be very much less than that of the body whose motion is being measured, in order to avoid loading effects that affect the accuracy of the readings obtained. In this respect, instruments based on strain gauges are best.

19.4 Vibration

19.4.1 Nature of vibration

Vibrations are very commonly encountered in machinery operation, and therefore measurement of the accelerations associated with such vibrations is extremely important in industrial environments. The peak accelerations involved in such vibrations can be of $100g$ or greater in magnitude, whilst both the frequency of oscillation and the magnitude of displacements from the equilibrium position in vibrations have a tendency to vary randomly. Vibrations normally consist of linear harmonic motion that can be expressed mathematically as:

$$X = X_0 \sin(\omega t) \tag{19.6}$$

where X is the displacement from the equilibrium position at any general point in time, X_0 is the peak displacement from the equilibrium position, and ω is the angular frequency of the oscillations. By differentiating equation (19.6) with respect to time, an expression for the velocity v of the vibrating body at any general point in time is obtained as:

$$v = -\omega X_0 \cos(\omega t) \tag{19.7}$$

Differentiating equation (19.7) again with respect to time, we obtain an expression for the acceleration, α, of the body at any general point in time as:

$$\alpha = -\omega^2 X_0 \sin(\omega t) \tag{19.8}$$

Inspection of equation (19.8) shows that the peak acceleration is given by:

$$\alpha_{\mathrm{peak}} = \omega^2 X_0 \tag{19.9}$$

This square law relationship between peak acceleration and oscillation frequency is the reason why high values of acceleration occur during relatively low-frequency oscillations. For example, an oscillation at 10 Hz produces peak accelerations of $2g$.

Example
A pipe carrying a fluid vibrates at a frequency of 50 Hz with displacements of 8 mm from the equilibrium position. Calculate the peak acceleration.

Solution
From equation (19.9),

$$\alpha_{\mathrm{peak}} = \omega^2 X_0 = (2\pi 50)^2 \times (0.008) = 789.6\ \mathrm{m/s^2}$$

Using the fact that the acceleration due to gravity, g, is 9.81 m/s^2, this answer can be expressed alternatively as:

$$\alpha_{\mathrm{peak}} = 789.6/9.81 = 80.5g$$

19.4.2 Vibration measurement

It is apparent that the intensity of vibration can be measured in terms of either displacement, velocity or acceleration. Acceleration is clearly the best parameter to measure at high frequencies. However, because displacements are large at low frequencies according to equation (19.9), it would seem that measuring either displacement or velocity would be best at low frequencies. The amplitude of vibrations can be measured by various forms of displacement transducer. Fibre-optic-based devices are particularly attractive and can give measurement resolution as high as 1 μm. Unfortunately, there are considerable practical difficulties in mounting and calibrating displacement and velocity transducers and therefore they are rarely used. Thus, vibration is usually measured by accelerometers at all frequencies. The most common type of transducer used is the piezoaccelerometer, which has typical inaccuracy levels of ±2%.

The frequency response of accelerometers is particularly important in vibration measurement in view of the inherently high-frequency characteristics of the measurement situation. The bandwidth of both potentiometer-based accelerometers and accelerometers using variable-inductance type displacement transducers goes up to 25 Hz only. Accelerometers including either the LVDT or strain gauges can measure frequencies up to 150 Hz and the latest instruments using piezoresistive strain gauges have bandwidths up to 2 kHz. Finally, inclusion of piezoelectric crystal displacement transducers yields an instrument with a bandwidth that can be as high as 7 kHz.

When measuring vibration, consideration must be given to the fact that attaching an accelerometer to the vibrating body will significantly affect the vibration characteristics if the body has a small mass. The effect of such 'loading' of the measured system can be quantified by the following equation:

$$a_1 = a_b \left(\frac{m_b}{m_b + m_a} \right)$$

where a_1 is the acceleration of the body with accelerometer attached, a_b is the acceleration of the body without the accelerometer, m_a is the mass of the accelerometer and m_b is the mass of the body. Such considerations emphasize the advantage of piezoaccelerometers, as these have a lower mass than other forms of accelerometer and so contribute least to this system-loading effect.

As well as an accelerometer, a vibration measurement system requires other elements, as shown in Figure 19.16, to translate the accelerometer output into a recorded signal. The three other necessary elements are a signal-conditioning element, a signal analyser and a signal recorder. The signal-conditioning element amplifies the relatively weak output signal from the accelerometer and also transforms the high output impedance of the accelerometer to a lower impedance value. The signal analyser then converts the signal into the form required for output. The output parameter may be either displacement, velocity or acceleration and this may be expressed as either the peak value, r.m.s. value or average absolute value. The final element of the measurement system is the signal recorder. All elements of the measurement system, and especially the signal recorder, must be chosen very carefully to avoid distortion of the vibration waveform. The bandwidth should be such that it is at least a factor of ten better than the bandwidth of the vibration frequency components at both ends. Thus its lowest frequency limit should be less than or equal to 0.1 times the fundamental frequency of vibration and its upper frequency limit should be greater than or equal to ten times the highest significant vibration frequency component.

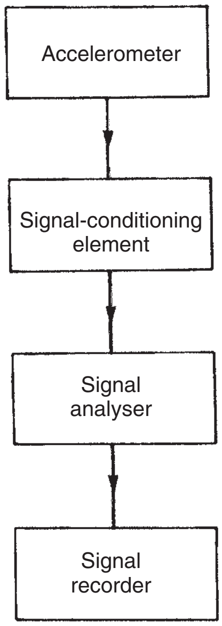


Fig. 19.16 Vibration measurement system.

If the frequency of vibration has to be known, the stroboscope is a suitable instrument to measure this. If the stroboscope is made to direct light pulses at the body at the same frequency as the vibration, the body will apparently stop vibrating.

19.5 Shock

Shock describes a type of motion where a moving body is brought suddenly to rest, often because of a collision. This is very common in industrial situations and usually involves a body being dropped and hitting the floor. Shocks characteristically involve large-magnitude deceleration (e.g. $500g$) that last for a very short time (e.g. 5 ms). An instrument having a very high-frequency response is required for shock measurement, and for this reason, piezoelectric crystal-based accelerometers are commonly used. Again, other elements for analysing and recording the signal are required as shown in Figure 19.16 and described in the last section. A storage oscilloscope is a suitable instrument for recording the output signal, as this allows the time duration as well as the acceleration levels in the shock to be measured. Alternatively, if a permanent record is required, the screen of a standard oscilloscope can be photographed. A further option is to record the output on magnetic tape, which facilitates computerized signal analysis.

Example
A body is dropped from a height of 10 m and suffers a shock when it hits the ground. If the duration of the shock is 5 ms, calculate the magnitude of the shock in terms of g.

Solution
The equation of motion for a body falling under gravity gives the following expression for the terminal velocity, v:

$$v = \sqrt{2gx}$$

where x is the height through which the body falls. Having calculated v, the average deceleration during the collision can be calculated as:

$$\alpha = v/t$$

where t is the time duration of the shock. Substituting the appropriate numerical values into these expressions:

$$v = \sqrt{(2 \times 9.81 \times 10)} = 14.0\,\text{m/s}; \quad \alpha = 14.0/0.005 = 2801\,\text{m/s} = 286g.$$

20

Rotational motion transducers

20.1 Rotational displacement

Rotational displacement transducers measure the angular motion of a body about some rotation axis. They are important not only for measuring the rotation of bodies such as shafts, but also as part of systems that measure translational displacement by converting the translational motion to a rotary form. The various devices available for measuring rotational displacements are presented below, and the arguments for choosing a particular form in any given measurement situation are considered at the end of the chapter.

20.1.1 Circular and helical potentiometers

The circular potentiometer is the cheapest device available for measuring rotational displacements. It works on almost exactly the same principles as the translational motion potentiometer, except that the track is bent round into a circular shape. The measurement range of individual devices varies from $0-10^\circ$ to $0-360^\circ$ depending on whether the track forms a full circle or only part of a circle. Where greater measurement range than $0-360^\circ$ is required, a helical potentiometer is used, with some devices being able to measure up to 60 full turns. The helical potentiometer accommodates multiple turns of the track by forming the track into a helix shape. However, its greater mechanical complexity makes the device significantly more expensive than a circular potentiometer. The two forms of device are shown in Figure 20.1.

Both kinds of device give a linear relationship between the measured quantity and the output reading because the output voltage measured at the sliding contact is proportional to the angular displacement of the slider from its starting position. However, as with linear track potentiometers, all rotational potentiometers can give performance problems due to dirt on the track causing loss of contact. They also have a limited life because of wear between the sliding surfaces. The typical inaccuracy of this class of devices varies from $\pm1\%$ of full scale for circular potentiometers down to $\pm0.002\%$ of full scale for the best helical potentiometers.

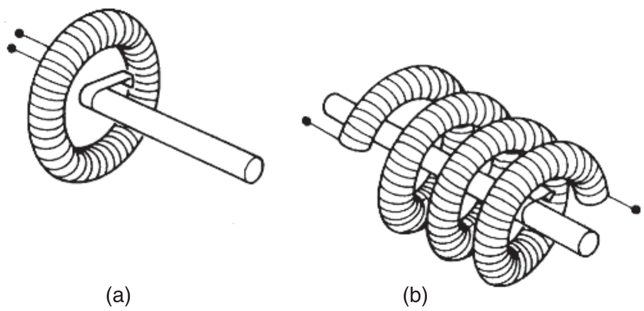


Fig. 20.1 Rotary motion potentiometers: (a) circular; (b) helical.

20.1.2 Rotational differential transformer

This is a special form of differential transformer that measures rotational rather than translational motion. The method of construction and connection of the windings is exactly the same as for the linear variable differential transformer (LVDT), except that a specially shaped core is used that varies the mutual inductance between the windings as it rotates, as shown in Figure 20.2. Like its linear equivalent, the instrument suffers no wear in operation and therefore has a very long life with almost no maintenance requirements. It can also be modified for operation in harsh environments by enclosing the windings inside a protective enclosure. However, apart from the difficulty of avoiding some asymmetry between the secondary windings, great care has to be taken in these instruments to machine the core to exactly the right shape. In consequence, the inaccuracy cannot be reduced below $\pm 1\%$, and even this level of accuracy is only obtained for limited excursions of the core of $\pm 40^\circ$ away from the null position. For angular displacements of $\pm 60^\circ$, the typical inaccuracy rises to $\pm 3\%$, and the instrument is unsuitable for measuring displacements greater than this.

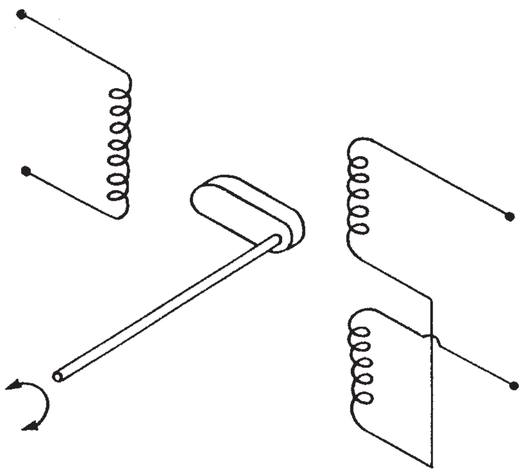


Fig. 20.2 Rotary differential transformer.

20.1.3 Incremental shaft encoders

Incremental shaft encoders are one of a class of encoder devices that give an output in digital form. They measure the instantaneous angular position of a shaft relative to some arbitrary datum point, but are unable to give any indication about the absolute position of a shaft. The principle of operation is to generate pulses as the shaft whose displacement is being measured rotates. These pulses are counted and the total angular rotation inferred from the pulse count. The pulses are generated either by optical or by magnetic means and are detected by suitable sensors. Of the two, the optical system is considerably cheaper and therefore much more common. Such instruments are very convenient for computer control applications, as the measurement is already in the required digital form and therefore the usual analogue to digital signal conversion process is avoided.

An example of an optical incremental shaft encoder is shown in Figure 20.3. It can be seen that the instrument consists of a pair of discs, one of which is fixed and one of which rotates with the body whose angular displacement is being measured. Each disc is basically opaque but has a pattern of windows cut into it. The fixed disc has only one window and the light source is aligned with this so that the light shines through all the time. The second disc has two tracks of windows cut into it that are equidistantly spaced around the disc, as shown in Figure 20.4. Two light detectors are positioned beyond the second disc so that one is aligned with each track of windows. As the second disc rotates, light alternately enters and does not enter the detectors, as windows and then opaque regions of the disc pass in front of them. These pulses of light are fed to a counter, with the final count after motion has ceased corresponding to the angular position of the moving body relative to the starting position. The primary information about the magnitude of rotation is obtained by the detector aligned with the

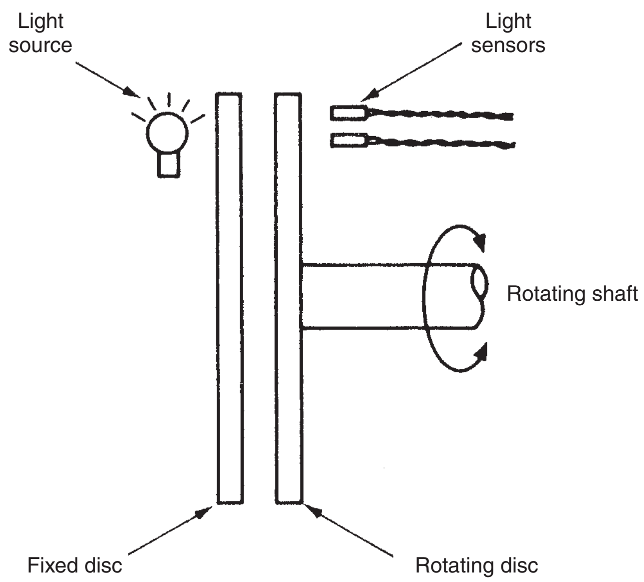


Fig. 20.3 Optical incremental shaft encoder.

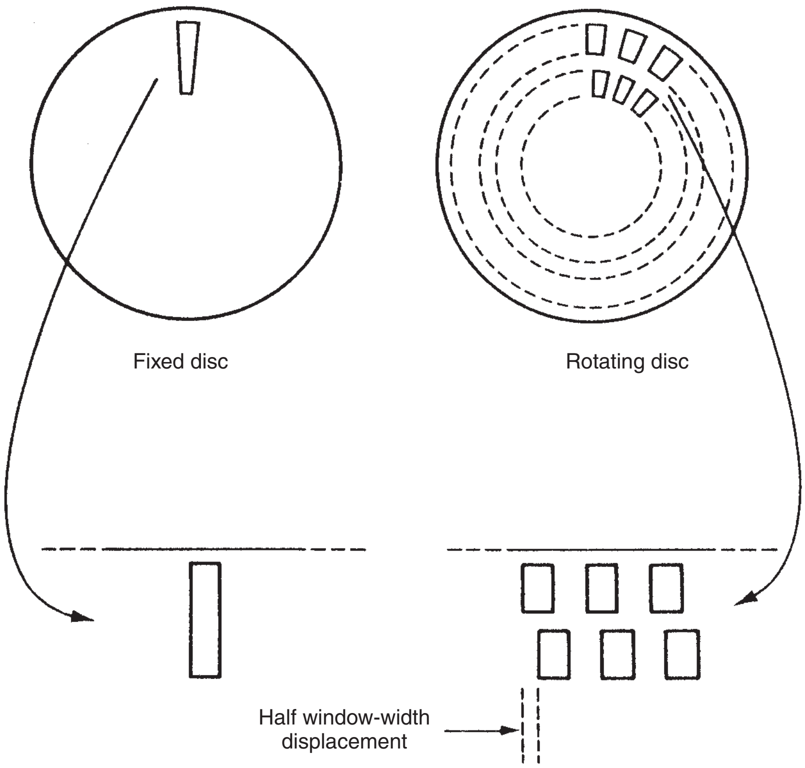


Fig. 20.4 Window arrangement in incremental shaft encoder.

outer track of windows. The pulse count obtained from this gives no information about the direction of rotation, however. Direction information is provided by the second, inner track of windows, which have an angular displacement with respect to the outer set of windows of half a window width. The pulses from the detector aligned with the inner track of windows therefore lag or lead the primary set of pulses according to the direction of rotation.

The maximum measurement resolution obtainable is limited by the number of windows that can be machined onto a disc. The maximum number of windows per track for a 150 mm-diameter disc is 5000, which gives a basic angular measurement resolution of 1 in 5000. By using more sophisticated circuits that increment the count on both the rising and falling edges of the pulses through the outer track of windows, it is possible to double the resolution to a maximum of 1 in 10 000. At the expense of even greater complexity in the counting circuit, it is possible also to include the pulses from the inner track of windows in the count, so giving a maximum measurement resolution of 1 in 20 000.

Optical incremental shaft encoders are a popular instrument for measuring relative angular displacements and are very reliable. Problems of noise in the system giving false counts can sometimes cause difficulties, although this can usually be eliminated by squaring the output from the light detectors. Such instruments are found in many applications where rotational motion has to be measured. Incremental shaft encoders are also commonly used in circumstances where a translational displacement has been transformed to a rotational one by suitable gearing. One example of this practice is in measuring the translational motions in numerically controlled (NC) drilling machines. Typical gearing used for this would give one revolution per mm of translational displacement. By using an incremental shaft encoder with 1000 windows per track in such an arrangement, a measurement resolution of 1 micron is obtained.

20.1.4 Coded-disc shaft encoders

Unlike the incremental shaft encoder that gives a digital output in the form of pulses that have to be counted, the digital shaft encoder has an output in the form of a binary number of several digits that provides an absolute measurement of shaft position. Digital encoders provide high accuracy and reliability. They are particularly useful for computer control applications, but they have a significantly higher cost than incremental encoders. Three different forms exist, using optical, electrical and magnetic energy systems respectively.

Optical digital shaft encoder

The optical digital shaft encoder is the cheapest form of encoder available and is the one used most commonly. It is found in a variety of applications, and one where it is particularly popular is in measuring the position of rotational joints in robot manipulators. The instrument is similar in physical appearance to the incremental shaft encoder. It has a pair of discs (one movable and one fixed) with a light source on one side and light detectors on the other side, as shown in Figure 20.5. The fixed

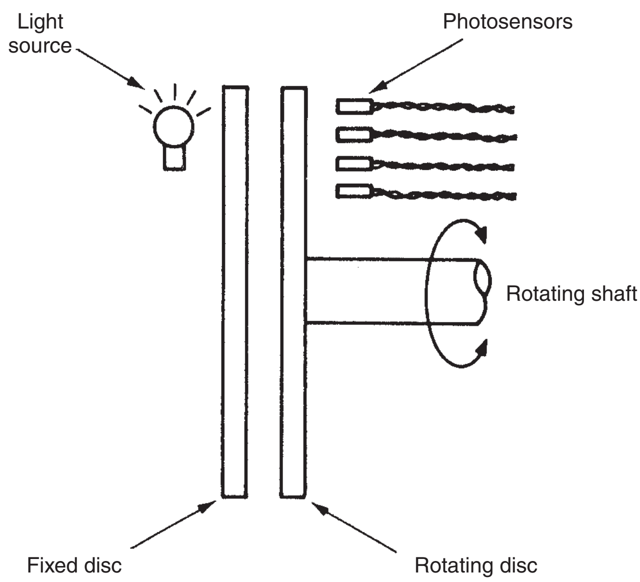


Fig. 20.5 Coded disc shaft encoder.

disc has a single window, and the principal way in which the device differs from the incremental shaft encoder is in the design of the windows on the movable disc, as shown in Figure 20.6. These are cut in four or more tracks instead of two and are arranged in sectors as well as tracks. An energy detector is aligned with each track, and these give an output of '1' when energy is detected and an output of '0' otherwise. The measurement resolution obtainable depends on the number of tracks used. For a four-track version, the resolution is 1 in 16, with progressively higher measurement resolution being attained as the number of tracks is increased. These binary outputs from the detectors are combined together to give a binary number of several digits. The number of digits corresponds to the number of tracks on the disc, which in the example shown in Figure 20.6 is four. The pattern of windows in each sector is cut such that, as that particular sector passes across the window in the fixed disc, the four energy detector outputs combine to give a unique binary number. In the binary-coded example shown in Figure 20.6, the binary number output increments by one as each sector in the rotating disc passes in turn across the window in the fixed disc. Thus the output from sector 1 is 0001, from sector 2 is 0010, from sector 3 is 0011, etc.

Whilst this arrangement is perfectly adequate in theory, serious problems can arise in practice due to the manufacturing difficulty involved in machining the windows of the movable disc such that the edges of the windows in each track are exactly aligned with each other. Any misalignment means that, as the disc turns across the boundary between one sector and the next, the outputs from each track will switch at slightly different instants of time, and therefore the binary number output will be incorrect over small angular ranges corresponding to the sector boundaries. The worst error can occur at the boundary between sectors seven and eight, where the output is switching from 0111 to 1000. If the energy sensor corresponding to the first digit switches before the others, then the output will be 1111 for a very small angular range of movement, indicating that sector 15 is aligned with the fixed disc rather than sector seven or eight. This represents an error of 100% in the indicated angular position.

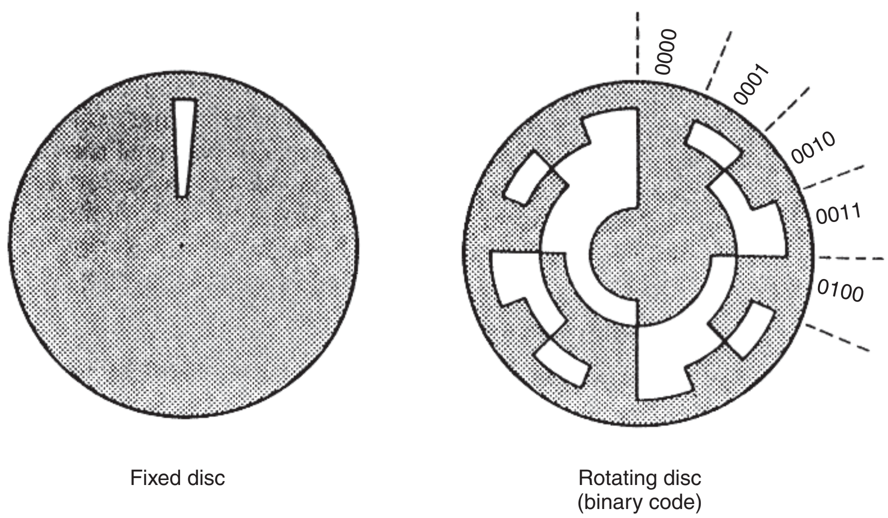


Fig. 20.6 Window arrangement for coded disc shaft encoder.

There are two ways used in practice to overcome this difficulty, which both involve an alteration to the manner in which windows are machined on the movable disc, as shown in Figure 20.7. The first of these methods adds an extra outer track on the disc, known as an *anti-ambiguity track*, which consists of small windows that span a small angular range on either side of each sector boundary of the main track system. When energy sensors associated with this extra track sense energy, this is used to signify that the disc is aligned on a sector boundary and the output is unreliable.

The second method is somewhat simpler and cheaper, because it avoids the expense of machining the extra anti-ambiguity track. It does this by using a special code, known as the Gray code, to cut the tracks in each sector on the movable disc. The Gray code is a special binary representation, where only one binary digit changes in moving from one decimal number representation to the next, i.e. from one sector to the next in the digital shaft encoder. The code is illustrated in Table 20.1.

It is possible to manufacture optical digital shaft encoders with up to 21 tracks, which gives a measurement resolution of 1 part in 10^6 (about one second of arc). Unfortunately, there is a high cost involved in the special photolithography techniques used to cut the windows in order to achieve such a measurement resolution, and very high-quality mounts and bearings are needed. Hence, such devices are very expensive.

Contacting (electrical) digital shaft encoder

The contacting digital shaft encoder consists of only one disc that rotates with the body whose displacement is being measured. The disc has conducting and non-conducting segments rather than the transparent and opaque areas found on the movable disc of the optical form of instrument, but these are arranged in an identical pattern of sectors and tracks. The disc is charged to a low potential by an electrical brush in contact with one side of the disc, and a set of brushes on the other side of the disc measures the potential in each track. The output of each detector brush is interpreted as a binary value of '1' or '0' according to whether the track in that particular segment is conducting or not and hence whether a voltage is sensed or not. As for the case of the optical

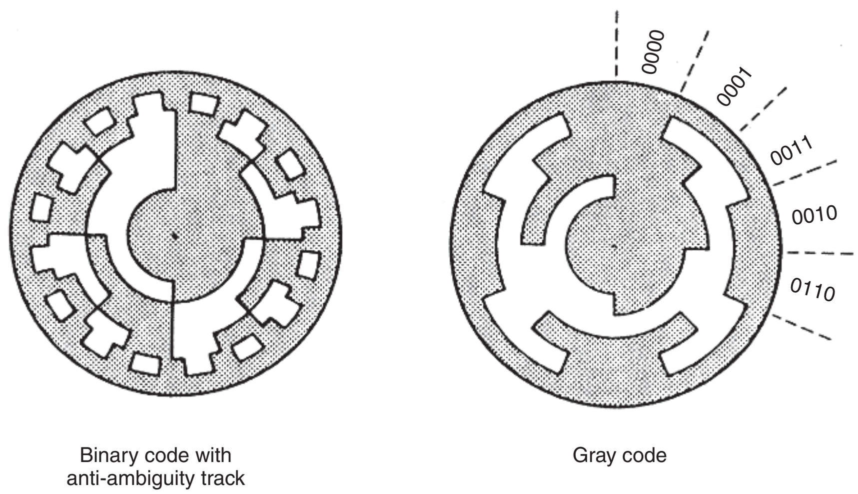


Fig. 20.7 Modified window arrangements for the rotating disc.

Table 20.1 The Gray code

Decimal	*Binary*	*Gray*
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100
9	1001	1101
10	1010	1111
11	1011	1110
12	1100	1010
13	1101	1011
14	1110	1001
15	1111	1000

form of instrument, these outputs are combined together to give a multi-bit binary number. Contacting digital shaft encoders have a similar cost to the equivalent optical instruments and have operational advantages in severe environmental conditions of high temperature or mechanical shock. They suffer from the usual problem of output ambiguity at the sector boundaries but this problem is overcome by the same methods as used in optical instruments.

A serious problem in the application of contacting digital shaft encoders arises from their use of brushes. These introduce friction into the measurement system, and the combination of dirt and brush wear causes contact problems. Consequently, problems of intermittent output can occur, and such instruments generally have limited reliability and a high maintenance cost. Measurement resolution is also limited because of the lower limit on the minimum physical size of the contact brushes. The maximum number of tracks possible is ten, which limits the resolution to 1 part in 1000. Thus, contacting digital shaft encoders are only used where the environmental conditions are too severe for optical instruments.

Magnetic digital shaft encoder

Magnetic digital shaft encoders consist of a single rotatable disc, as in the contacting form of encoder discussed in the previous section. The pattern of sectors and tracks consists of magnetically conducting and non-conducting segments, and the sensors aligned with each track consist of small toroidal magnets. Each of these sensors has a coil wound on it that has a high or low voltage induced in it according to the magnetic field close to it. This field is dependent on the magnetic conductivity of that segment of the disc that is closest to the toroid.

These instruments have no moving parts in contact and therefore have a similar reliability to optical devices. Their major advantage over optical equivalents is an ability to operate in very harsh environmental conditions. Unfortunately, the process of manufacturing and accurately aligning the toroidal magnet sensors required makes such instruments very expensive. Their use is therefore limited to a few applications where both high measurement resolution and also operation in harsh environments are required.

20.1.5 The resolver

The resolver, also known as a *synchro-resolver*, is an electromechanical device that gives an analogue output by transformer action. Physically, resolvers resemble a small a.c. motor and have a diameter ranging from 10 mm to 100 mm. They are frictionless and reliable in operation because they have no contacting moving surfaces, and consequently they have a long life. The best devices give measurement resolutions of 0.1%.

Resolvers have two stator windings, which are mounted at right angles to one another, and a rotor, which can have either one or two windings. As the angular position of the rotor changes, the output voltage changes. The simpler configuration of a resolver with only one winding on the rotor is illustrated in Figure 20.8. This exists in two separate forms that are distinguished according to whether the output voltage changes in amplitude or changes in phase as the rotor rotates relative to the stator winding.

Varying amplitude output resolver

The stator of this type of resolver is excited with a single-phase sinusoidal voltage of frequency ω, where the amplitudes in the two windings are given by:

$$V_1 = V \sin(\beta); \quad V_2 = V \cos(\beta)$$

where $V = V_s \sin(\omega t)$

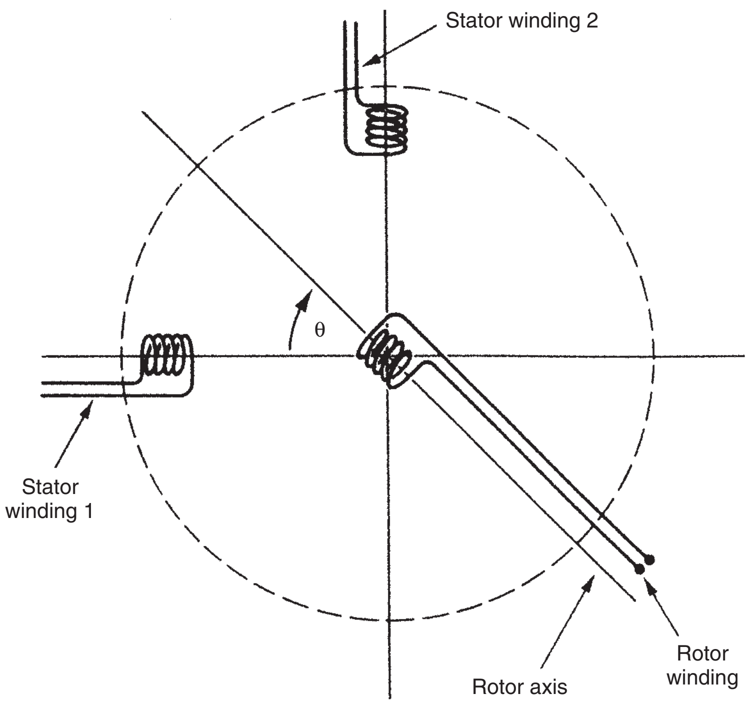


Fig. 20.8 Schematic representation of resolver windings.

The effect of this is to give a field at an angle of $(\beta + \pi/2)$ relative to stator winding 1. (A full proof of this can be found in Healey, (1975).)

Suppose that the angle of the rotor winding relative to that of the stator winding is given by θ. Then the magnetic coupling between the windings is a maximum for $\theta = (\beta + \pi/2)$ and a minimum for $(\theta = \beta)$. The rotor output voltage (see Healey (1975) for proof) is of fixed frequency and varying amplitude given by:

$$V_0 = KV_s \sin(\beta - \theta) \sin(\omega t)$$

This relationship between shaft angle position and output voltage is non-linear, but approximate linearity is obtained for small angular motions where $|\beta - \theta| < 15°$.

An intelligent version of this type of resolver is now available that uses a microprocessor to process the sine and cosine outputs, giving a measurement resolution of 2 minutes of arc (Analogue Devices, 1988).

Varying phase output resolver

This is a less common form of resolver but it is used in a few applications. The stator windings are excited with a two-phase sinusoidal voltage of frequency ω, and the instantaneous voltage amplitudes in the two windings are given by:

$$V_1 = V_s \sin(\omega t); \quad V_2 = V_s \sin(\omega t + \pi/2) = V_s \cos(\omega t)$$

The net output voltage in the rotor winding is the sum of the voltages induced due to each stator winding. This is given by:

$$\begin{aligned} V_0 &= KV_s \sin(\omega t)\cos(\theta) + KV_s \cos(\omega t)\cos(\pi/2 - \theta) \\ &= KV_s[\sin(\omega t)\cos(\theta) + \cos(\omega t)\sin(\theta)] \\ &= KV_s \sin(\omega t + \theta) \end{aligned}$$

This represents a linear relationship between shaft angle and the phase shift of the rotor output relative to the stator excitation voltage. The accuracy of shaft rotation measurement depends on the accuracy with which the phase shift can be measured. This can be improved by increasing the excitation frequency, ω, and it is possible to reduce inaccuracy to $\pm 0.1\%$. However, increasing the excitation frequency also increases magnetizing losses. Consequently, a compromise excitation frequency of about 400 Hz is used.

20.1.6 The synchro

Like the resolver, the synchro is a motor-like, electromechanical device with an analogue output. Apart from having three stator windings instead of two, the instrument is similar in appearance and operation to the resolver and has the same range of physical dimensions. The rotor usually has a dumb-bell shape and, like the resolver, can have either one or two windings.

Synchros have been in use for many years for the measurement of angular positions, especially in military applications, and achieve similar levels of accuracy and measurement resolution to digital encoders. One common application is axis measurement in machine tools, where the translational motion of the tool is translated into a rotational displacement by suitable gearing. Synchros are tolerant to high temperatures, high humidity, shock and vibration and are therefore suitable for operation in such harsh environmental conditions. Some maintenance problems are associated with the slip ring and brush system used to supply power to the rotor. However, the only major source of error in the instrument is asymmetry in the windings, and reduction of measurement inaccuracy down to ±0.5% is easily achievable.

Figure 20.9 shows the simpler form of synchro with a single rotor winding. If an a.c. excitation voltage is applied to the rotor via slip rings and brushes, this sets up a certain pattern of fluxes and induced voltages in the stator windings by transformer action. For a rotor excitation voltage, V_r, given by:

$$V_r = V \sin(\omega t)$$

the voltages induced in the three stator windings are:

$$V_1 = V \sin(\omega t) \sin(\beta);\ \ V_2 = V \sin(\omega t) \sin(\beta + 2\pi/3;\ \ V_3 = V \sin(\omega t) \sin(\beta - 2\pi/3)$$

where β is the angle between the rotor and stator windings.

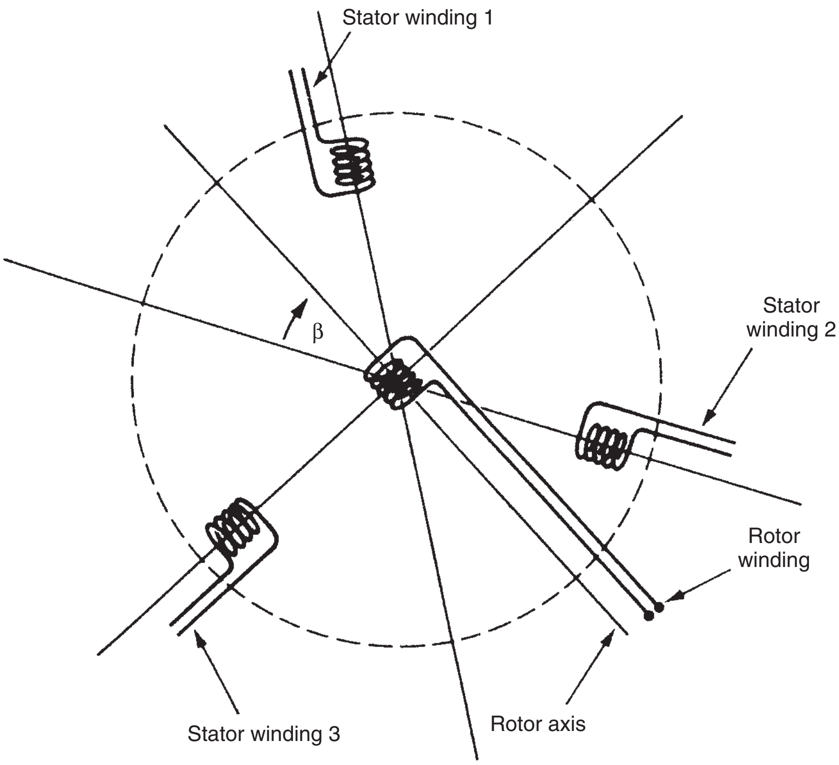


Fig. 20.9 Schematic representation of synchro windings.

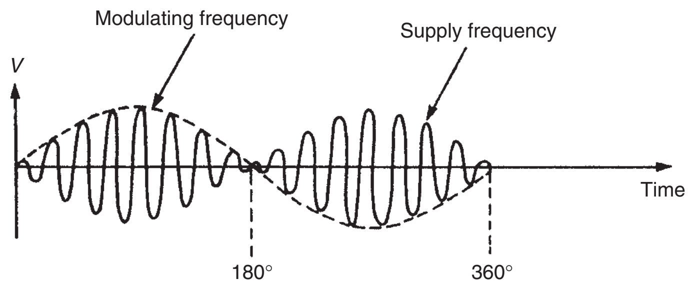


Fig. 20.10 Synchro stator voltage waveform.

If the rotor is turned at constant velocity through one full revolution, the voltage waveform induced in each stator winding is as shown in Figure 20.10. This has the form of a carrier-modulated waveform, in which the carrier frequency corresponds to the excitation frequency, ω. It follows that if the rotor is stopped at any particular angle, β', the peak-to-peak amplitude of the stator voltage is a function of β'. If therefore the stator winding voltage is measured, generally as its root-mean-squared (r.m.s.) value, this indicates the magnitude of the rotor rotation away from the null position. The direction of rotation is determined by the phase difference between the stator voltages, which is indicated by their relative instantaneous magnitudes.

Although a single synchro thus provides a means of measuring angular displacements, it is much more common to find a pair of them used for this purpose. When used in pairs, one member of the pair is known as the synchro transmitter and the other as the synchro transformer, and the two sets of stator windings are connected together, as shown in Figure 20.11. Each synchro is of the form shown in Figure 20.9, but the rotor of the transformer is fixed for displacement-measuring applications. A sinusoidal excitation voltage is applied to the rotor of the transmitter, setting up a pattern of fluxes and induced voltages in the transmitter stator windings. These voltages are transmitted to the transformer stator windings where a similar flux pattern is established. This in turn causes a sinusoidal voltage to be induced in the fixed transformer rotor winding. For an excitation voltage, $V\sin(\omega t)$, applied to the transmitter rotor, the voltage measured in the transformer rotor is given by:

$$V_0 = V\sin(\omega t)\sin(\theta)$$

where θ is the relative angle between the two rotor windings.

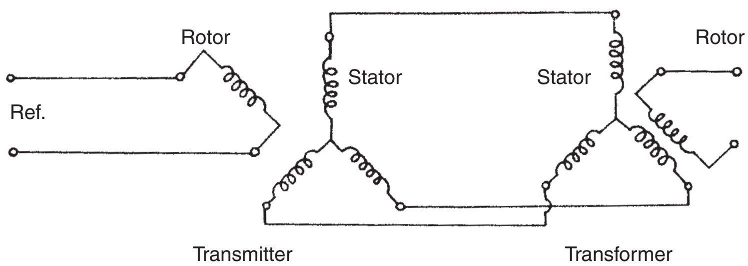


Fig. 20.11 Synchro transmitter-transformer pair.

Apart from their use as a displacement transducer, such synchro pairs are commonly used to transmit angular displacement information over some distance, for instance to transmit gyro compass measurements in an aircraft to remote meters. They are also used for load positioning, allowing a load connected to the transformer rotor shaft to be controlled remotely by turning the transmitter rotor. For these applications, the transformer rotor is free to rotate and is also damped to prevent oscillatory motions. In the simplest arrangement, a common sinusoidal excitation voltage is applied to both rotors. If the transmitter rotor is turned, this causes an imbalance in the magnetic flux patterns and results in a torque on the transformer rotor that tends to bring it into line with the transmitter rotor. This torque is typically small for small displacements, and so this technique is only useful if the load torque on the transformer shaft is very small. In other circumstances, it is necessary to incorporate the synchro pair within a servomechanism, where the output voltage induced in the transformer rotor winding is amplified and applied to a servomotor that drives the transformer rotor shaft until it is aligned with the transmitter shaft.

20.1.7 The induction potentiometer

These instruments belong to the same class as resolvers and synchros but have only one rotor winding and one stator winding. They are of a similar size and appearance to other devices in the class. A single-phase sinusoidal excitation is applied to the rotor winding and this causes an output voltage in the stator winding through the mutual inductance linking the two windings. The magnitude of this induced stator voltage varies with the rotation of the rotor. The variation of the output with rotation is naturally sinusoidal if the coils are wound such that their field is concentrated at one point, and only small excursions can be made away from the null position if the output relationship is to remain approximately linear. However, if the rotor and stator windings are distributed around the circumference in a special way, an approximately linear relationship for angular displacements of up to $\pm 90^\circ$ can be obtained.

20.1.8 The rotary inductosyn

This instrument is similar in operation to the linear inductosyn, except that it measures rotary displacements and has tracks that are arranged radially on two circular discs, as shown in Figure 20.12. Typical diameters of the instrument vary between 75 mm and 300 mm. The larger versions give a measurement resolution of up to 0.05 seconds of arc. Like its linear equivalent, however, the rotary inductosyn has a very small measurement range, and a lower-resolution, rotary displacement transducer with a larger measurement range must be used in conjunction with it.

20.1.9 Gyroscopes

Gyroscopes measure both absolute angular displacement and absolute angular velocity. The predominance of mechanical, spinning-wheel gyroscopes in the market place is now being challenged by recently introduced optical gyroscopes.

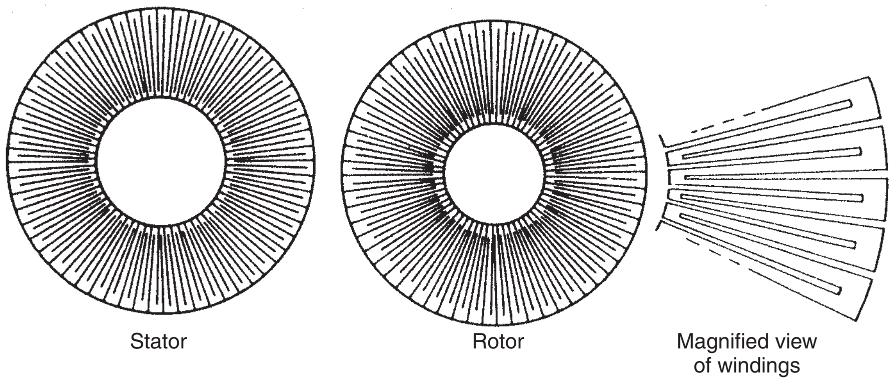


Fig. 20.12 Rotary inductosyn.

Mechanical gyroscopes

Mechanical gyroscopes consist essentially of a large, motor driven wheel whose angular momentum is such that the axis of rotation tends to remain fixed in space, thus acting as a reference point. The gyro frame is attached to the body whose motion is to be measured. The output is measured in terms of the angle between the frame and the axis of the spinning wheel. Two different forms of mechanical gyroscope are used for measuring angular displacement, the free gyro and the rate-integrating gyro. A third type of mechanical gyroscope, the rate gyro, measures angular velocity and is described in section 20.2.

Free gyroscope

The free gyroscope is illustrated in Figure 20.13. This measures the absolute angular rotation of the body to which its frame is attached about two perpendicular axes. Two alternative methods of driving the wheel are used in different versions of the instrument. One of these is to enclose the wheel in stator-like coils that are excited with a sinusoidal voltage. A voltage is applied to the wheel via slip rings at both ends of the spindle carrying the wheel. The wheel behaves as a rotor and motion is produced by motor action. The other, less common, method is to fix vanes on the wheel that is then driven by directing a jet of air onto the vanes.

The instrument can measure angular displacements of up to 10° with a high accuracy. For greater angular displacements, interaction between the measurements on the two perpendicular axes starts to cause a serious loss of accuracy. The physical size of the coils in the motor-action driven system also limits the measurement range to 10°. For these reasons, this type of gyroscope is only suitable for measuring rotational displacements of up to 10°. A further operational problem of free gyroscopes is the presence of angular drift (precession) due to bearing friction torque. This has a typical magnitude of 0.5° per minute and means that the instrument can only be used over short time intervals of say, 5 minutes. This time duration can be extended if the angular momentum of the spinning wheel is increased.

A major application of gyroscopes is in inertial navigation systems. Only two free gyros mounted along orthogonal axes are needed to monitor motions in three dimensions, because each instrument measures displacement about two axes. The limited

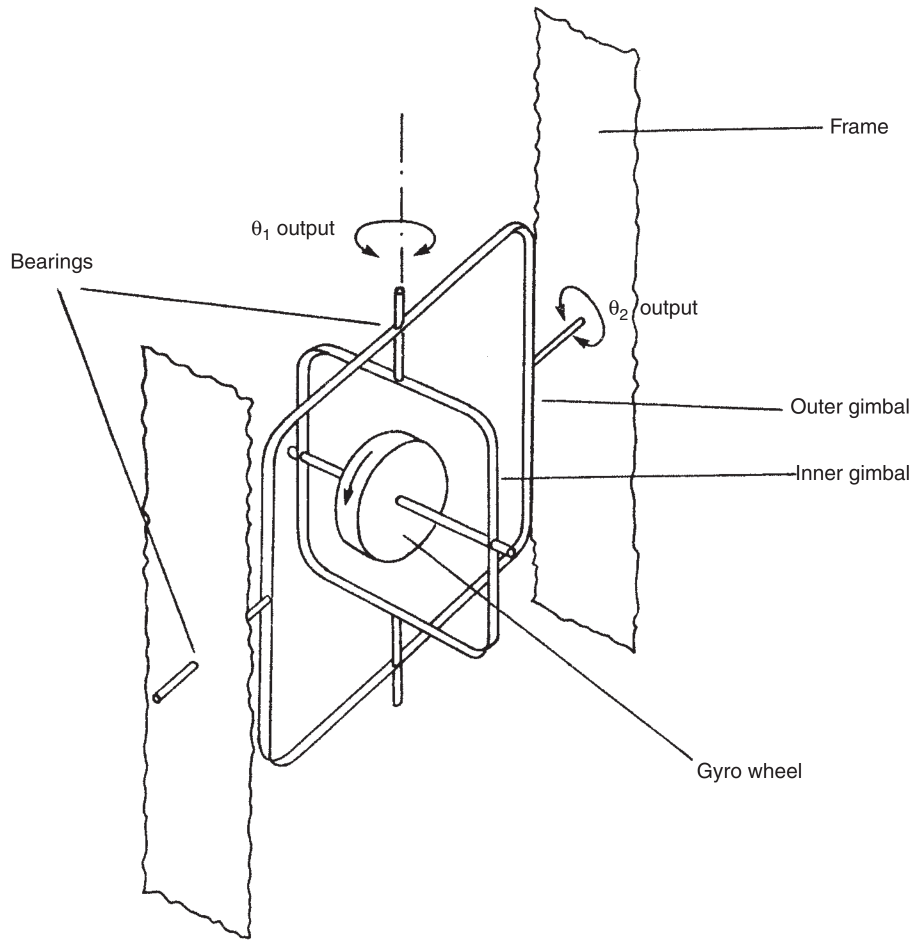


Fig. 20.13 Free gyroscope.

angular range of measurement is not usually a problem in such applications, as control action prevents the error in the direction of motion about any axis ever exceeding one or two degrees. Precession is a much greater problem, however, and for this reason, the rate-integrating gyro is used much more commonly.

Rate integrating gyroscope

The rate-integrating gyroscope, or *integrating gyro* as it is commonly known, is illustrated in Figure 20.14. It measures angular displacements about a single axis only, and therefore three instruments are required in a typical inertial navigation system. The major advantage of the instrument over the free gyro is the almost total absence of precession, with typical specifications quoting drifts of only 0.01°/hour. The instrument has a first order type of response given by:

$$\frac{\theta_o}{\theta_i}(D) = \frac{K}{\tau D + 1} \tag{20.1}$$

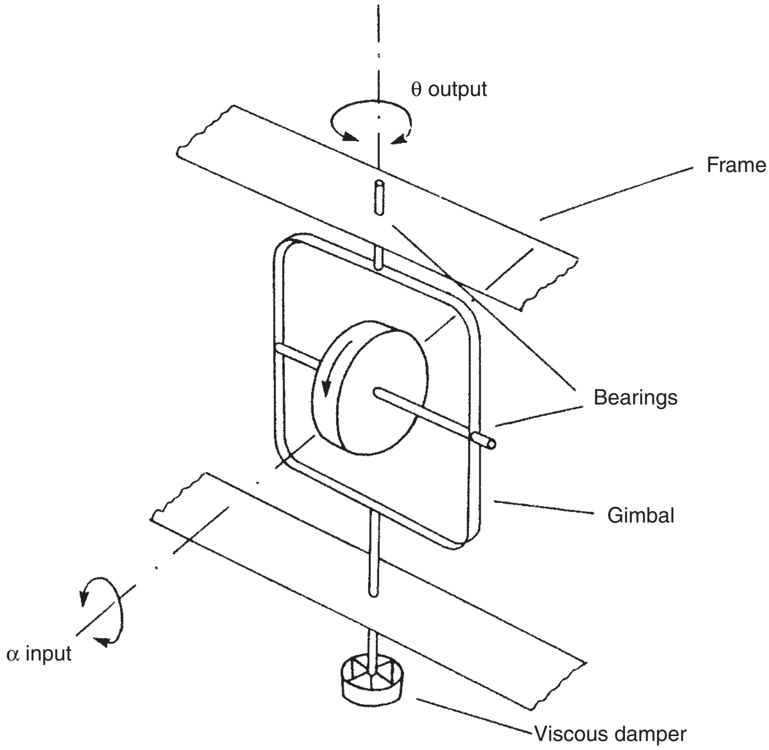


Fig. 20.14 Rate-integrating gyroscope.

where $K = H/\beta$, $\tau = M/\beta$, θ_i is the input angle, θ_o is the output angle, D is the D-operator, H is the angular momentum, M is the moment of inertia of the system about the measurement axis and β is the damping coefficient.

Inspection of equation (20.1) shows that to obtain a high value of measurement sensitivity, K, a high value of H and low value of β are required. A large H is normally obtained by driving the wheel with a hysteresis-type motor revolving at high speeds of up to 24 000 rpm. The damping coefficient β can only be reduced so far, however, because a small value of β results in a large value for the system time constant, τ, and an unacceptably low speed of system response. Therefore, the value of β has to be chosen as a compromise between these constraints.

Besides their use as a fixed reference in inertial guidance systems, integrating gyros are also commonly used within aircraft autopilot systems and in military applications such as stabilizing weapon systems in tanks.

Optical gyroscopes

Optical gyroscopes have been developed only recently and come in two forms, the ring laser gyroscope and the fibre-optic gyroscope.

The *ring laser gyroscope* consists of a glass ceramic chamber containing a helium–neon gas mixture in which two laser beams are generated by a single anode/twin cathode system, as shown in Figure 20.15. Three mirrors, supported by the

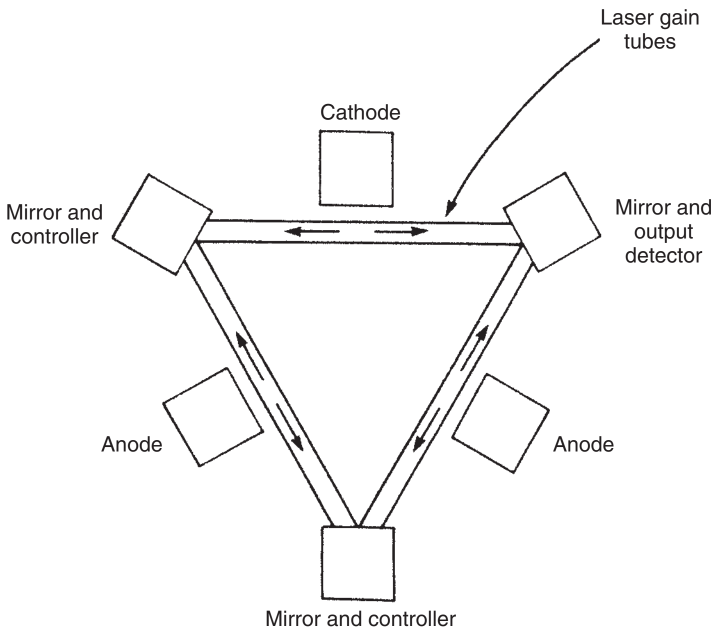


Fig. 20.15 Ring laser gyroscope.

ceramic block and mounted in a triangular arrangement, direct the pair of laser beams around the cavity in opposite directions. Any rotation of the ring affects the coherence of the two beams, raising one in frequency and lowering the other. The clockwise and anticlockwise beams are directed into a photodetector that measures the beat frequency according to the frequency difference, which is proportional to the angle of rotation. A more detailed description of the mode of operation can be found elsewhere (Nuttall, 1987). The advantages of the ring laser gyroscope are considerable. The measurement accuracy obtained is substantially better than that afforded by mechanical gyros in a similar price range. The device is also considerably smaller physically, which is of considerable benefit in many applications.

The *fibre-optic gyroscope* measures angular velocity and is described in section 20.2.

20.1.10 Choice between rotational displacement transducers

Choice between the various rotational displacement transducers that might be used in any particular measurement situation depends first of all upon whether absolute measurement of angular position is required or whether the measurement of rotation relative to some arbitrary starting point is acceptable. Other factors affecting the choice between instruments are the required measurement range, the resolution of the transducer and the measurement accuracy afforded.

Where only measurement of relative angular position is required, the incremental encoder is a very suitable instrument. The best commercial instruments of this type can measure rotations to a resolution of 1 part in 20 000 of a full revolution, and the measurement range is an infinite number of revolutions. Instruments with such a high measurement resolution are very expensive, but much cheaper versions are available according to what lower level of measurement resolution is acceptable.

All the other instruments presented in this chapter provide an absolute measurement of angular position. The required measurement range is a dominant factor in the choice between these. If this exceeds one full revolution, then the only instrument available is the helical potentiometer. Such devices can measure rotations of up to 60 full turns, but they are expensive because the procedure involved in manufacturing a helical resistance element to a reasonable standard of accuracy is difficult.

For measurements of less than one full revolution, the range of available instruments widens. The cheapest one available is the circular potentiometer, but much better measurement accuracy and resolution is obtained from coded-disc encoders. The cheapest of these is the optical form, but certain operating environments necessitate the use of the alternative contacting (electrical) and magnetic versions. All types of coded-disc encoder are very reliable and are particularly attractive in computer control schemes, as the output is in digital form. A varying phase output resolver is yet another instrument that can measure angular displacements up to one full revolution in magnitude. Unfortunately, this instrument is expensive because of the complicated electronics incorporated to measure the phase variation and convert it to a varying-amplitude output signal, and hence use is not common.

An even greater range of instruments becomes available as the required measurement range is reduced further. These include the synchro ($\pm 90°$), the varying amplitude output resolver ($\pm 90°$), the induction potentiometer ($\pm 90°$) and the differential transformer ($\pm 40°$). All these instruments have a high reliability and a long service life.

Finally, two further instruments are available for satisfying special measurement requirements, the rotary inductosyn and the gyroscope. The rotary inductosyn is used in applications where very high measurement resolution is required, although the measurement range afforded is extremely small and a coarser-resolution instrument must be used in parallel to extend the measurement range. Gyroscopes, in both mechanical and optical forms, are used to measure small angular displacements up to $\pm 10°$ in magnitude in inertial navigation systems and similar applications.

20.2 Rotational velocity

The main application of rotational velocity transducers is in speed control systems. They also provide the usual means of measuring translational velocities, which are transformed into rotational motions for measurement purposes by suitable gearing. Many different instruments and techniques are available for measuring rotational velocity as presented below.

20.2.1 Digital tachometers

Digital tachometers, or to give them their proper title, digital *tachometric generators*, are usually non-contact instruments that sense the passage of equally spaced marks on the surface of a rotating disc or shaft. Measurement resolution is governed by the number of marks around the circumference. Various types of sensor are used, such as optical, inductive and magnetic ones. As each mark is sensed, a pulse is generated and input to an electronic pulse counter. Usually, velocity is calculated in terms of the pulse count in unit time, which of course only yields information about the mean velocity. If the velocity is changing, instantaneous velocity can be calculated at each instant of time that an output pulse occurs, using the scheme shown in Figure 20.16. In this circuit, the pulses from the transducer gate the train of pulses from a 1 MHz clock into a counter. Control logic resets the counter and updates the digital output value after receipt of each pulse from the transducer. The measurement resolution of this system is highest when the speed of rotation is low.

Optical sensing

Digital tachometers with optical sensors are often known as *optical tachometers*. Optical pulses can be generated by one of the two alternative photoelectric techniques illustrated in Figure 20.17. In Figure 20.17(a), the pulses are produced as the windows in a slotted disc pass in sequence between a light source and a detector. The alternative form, Figure 20.17(b), has both light source and detector mounted on the same side of a reflective disc which has black sectors painted onto it at regular angular intervals. Light sources are normally either lasers or LEDs, with photodiodes and phototransistors being used as detectors. Optical tachometers yield better accuracy than other forms of digital tachometer but are not as reliable because dust and dirt can block light paths.

Inductive sensing

Variable reluctance velocity transducers, also known as *induction tachometers*, are a form of digital tachometer that use inductive sensing. They are widely used in the automotive industry within anti-skid devices, anti-lock braking systems (ABS) and traction control. One relatively simple and cheap form of this type of device was

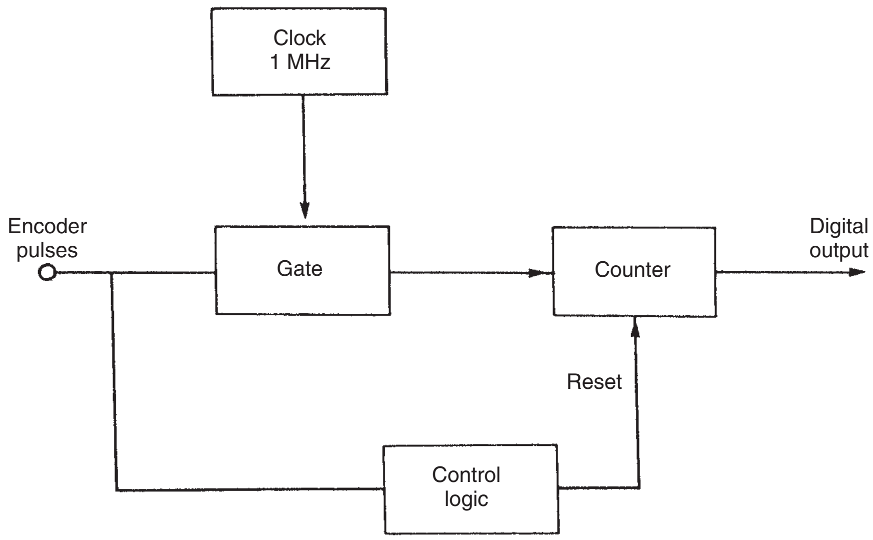


Fig. 20.16 Scheme to measure instantaneous angular velocities.

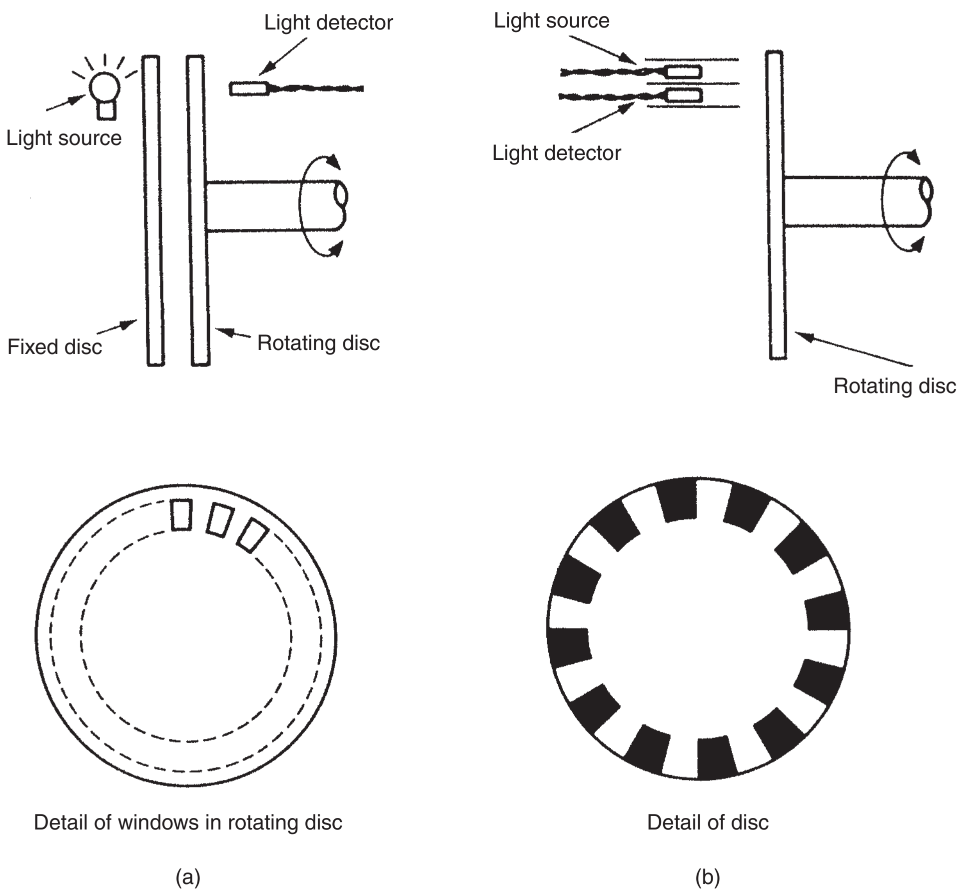


Fig. 20.17 Photoelectric pulse generation techniques.

described earlier in section 13.2 (Figure 13.2). A more sophisticated version shown in Figure 20.18 has a rotating disc that is constructed from a bonded-fibre material into which soft iron poles are inserted at regular intervals around its periphery. The sensor consists of a permanent magnet with a shaped pole piece, which carries a wound coil. The distance between the pick-up and the outer perimeter of the disc is around 0.5 mm. As the disc rotates, the soft iron inserts on the disc move in turn past the pick-up unit. As each iron insert moves towards the pole piece, the reluctance of the magnetic circuit increases and hence the flux in the pole piece also increases. Similarly, the flux in the pole piece decreases as each iron insert moves away from the sensor. The changing magnetic flux inside the pick-up coil causes a voltage to be induced in the coil whose magnitude is proportional to the rate of change of flux. This voltage is positive whilst the flux is increasing and negative whilst it is decreasing. Thus, the output is a sequence of positive and negative pulses whose frequency is proportional to the rotational velocity of the disc. The maximum angular velocity that the instrument can measure is limited to about 10 000 rpm because of the finite width of the induced pulses. As the velocity increases, the distance between the pulses is

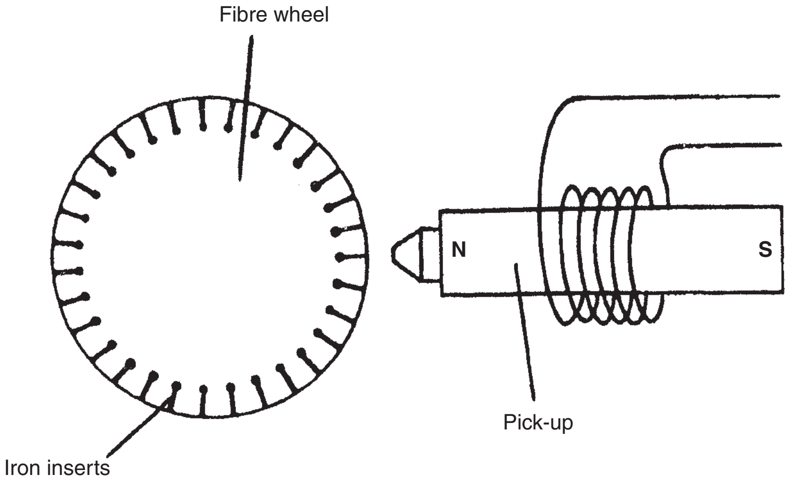


Fig. 20.18 Variable reluctance transducer.

reduced, and at a certain velocity, the pulses start to overlap. At this point, the pulse counter ceases to be able to distinguish the separate pulses. The optical tachometer has significant advantages in this respect, since the pulse width is much narrower, allowing measurement of higher velocities.

A simpler and cheaper form of variable reluctance transducer also exists that uses a ferromagnetic gear wheel in place of a fibre disc. The motion of the tip of each gear tooth towards and away from the pick-up unit causes a similar variation in the flux pattern to that produced by the iron inserts in the fibre disc. The pulses produced by these means are less sharp, however, and consequently the maximum angular velocity measurable is lower.

Magnetic (Hall-effect) sensing

The rotating element in *Hall-effect* or *magnetostrictive tachometers* has a very simple design in the form of a toothed metal gearwheel. The sensor is a solid-state, Hall-effect device that is placed between the gear wheel and a permanent magnet. When an inter-tooth gap on the gear wheel is adjacent to the sensor, the full magnetic field from the magnet passes through it. Later, as a tooth approaches the sensor, the tooth diverts some of the magnetic field, and so the field through the sensor is reduced. This causes the sensor to produce an output voltage that is proportional to the rotational speed of the gear wheel.

20.2.2 Stroboscopic methods

The stroboscopic technique of rotational velocity measurement operates on a similar physical principle to digital tachometers except that the pulses involved consist of flashes of light generated electronically and whose frequency is adjustable so that it can be matched with the frequency of occurrence of some feature on the rotating body being measured. This feature can either be some naturally occurring one such as gear teeth or the spokes of a wheel, or it can be an artificially created pattern of black and white stripes. In either case, the rotating body appears stationary when the frequencies of the light pulses and body features are in synchronism. Flashing rates available in commercial stroboscopes vary from 110 up to 150 000 per minute according to the range of velocity measurement required, and typical measurement inaccuracy is $\pm 1\%$ of the reading. The instrument is usually in the form of a hand-held device that is pointed towards the rotating body.

It must be noted that measurement of the flashing rate at which the rotating body appears stationary does not automatically indicate the rotational velocity, because synchronism also occurs when the flashing rate is some integral sub-multiple of the rotational speed. The practical procedure followed is therefore to adjust the flashing rate until synchronism is obtained at the largest flashing rate possible, R_1. The flashing rate is then carefully decreased until synchronism is again achieved at the next lower flashing rate, R_2. The rotational velocity is then given by:

$$V = \frac{R_1 R_2}{R_1 - R_2}$$

20.2.3 Analogue tachometers

Analogue tachometers are less accurate than digital tachometers but are nevertheless still used successfully in many applications. Various forms exist.

The *d.c. tachometer* has an output that is approximately proportional to its speed of rotation. Its basic structure is identical to that found in a standard d.c. generator used for producing power, and is shown in Figure 20.19. Both permanent-magnet types and separately excited field types are used. However, certain aspects of the design are optimized to improve its accuracy as a speed-measuring instrument. One significant design modification is to reduce the weight of the rotor by constructing the windings on a hollow fibreglass shell. The effect of this is to minimize any loading effect of the instrument on the system being measured. The d.c. output voltage from the instrument is of a relatively high magnitude, giving a high measurement sensitivity that is typically 5 volts per 1000 rpm. The direction of rotation is determined by the polarity of the output voltage. A common range of measurement is 0–6000 rpm. Maximum non-linearity is usually about $\pm 1\%$ of the full-scale reading. One problem with these devices that can cause difficulties under some circumstances is the presence of an a.c. ripple in the output signal. The magnitude of this can be up to 2% of the output d.c. level.

The *a.c. tachometer* has an output approximately proportional to rotational speed like the d.c. tachogenerator. Its mechanical structure takes the form of a two-phase induction motor, with two stator windings and (usually) a drag-cup rotor, as shown in Figure 20.20. One of the stator windings is excited with an a.c. voltage and the measurement signal is taken from the output voltage induced in the second winding. The magnitude of this output voltage is zero when the rotor is stationary, and otherwise proportional to the angular velocity of the rotor. The direction of rotation is determined by the phase of the output voltage, which switches by 180° as the direction reverses. Therefore, both the phase and magnitude of the output voltage have to be measured. A

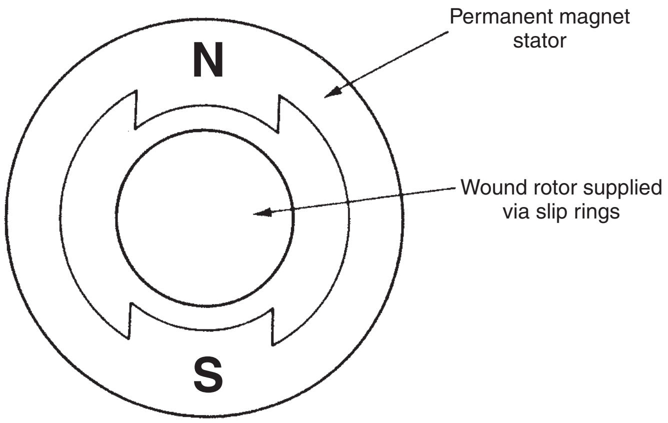


Fig. 20.19 D.c. tachometer.

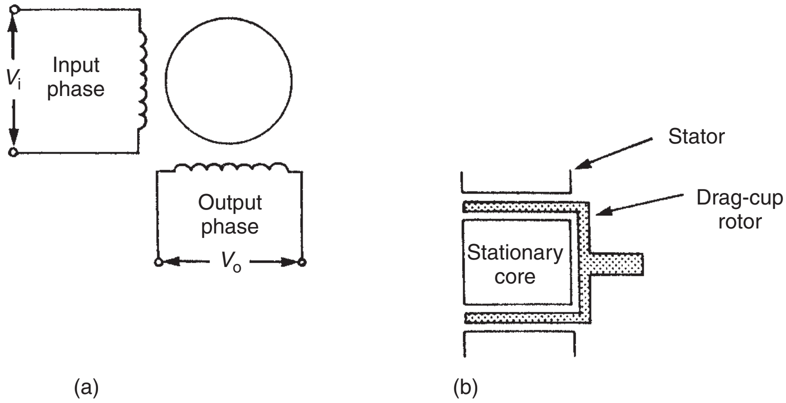


Fig. 20.20 A.c. tachometer.

typical range of measurement is 0–4000 rpm, with an inaccuracy of ±0.05% of full-scale reading. Cheaper versions with a squirrel-cage rotor also exist, but measurement inaccuracy in these is typically ±0.25%.

The *drag-cup tachometer*, also known as an *eddy-current tachometer*, has a central spindle carrying a permanent magnet that rotates inside a non-magnetic drag-cup consisting of a cylindrical sleeve of electrically conductive material, as shown in Figure 20.21. As the spindle and magnet rotate, a voltage is induced which causes circulating eddy currents in the cup. These currents interact with the magnetic field from the permanent magnet and produce a torque. In response, the drag-cup turns until the induced torque is balanced by the torque due to the restraining springs connected to the cup. When equilibrium is reached, the angular displacement of the cup is proportional to the rotational velocity of the central spindle. The instrument has a typical measurement inaccuracy of ±0.5% and is commonly used in the speedometers of motor vehicles and as a speed indicator for aero-engines. It is capable of measuring velocities up to 15 000 rpm.

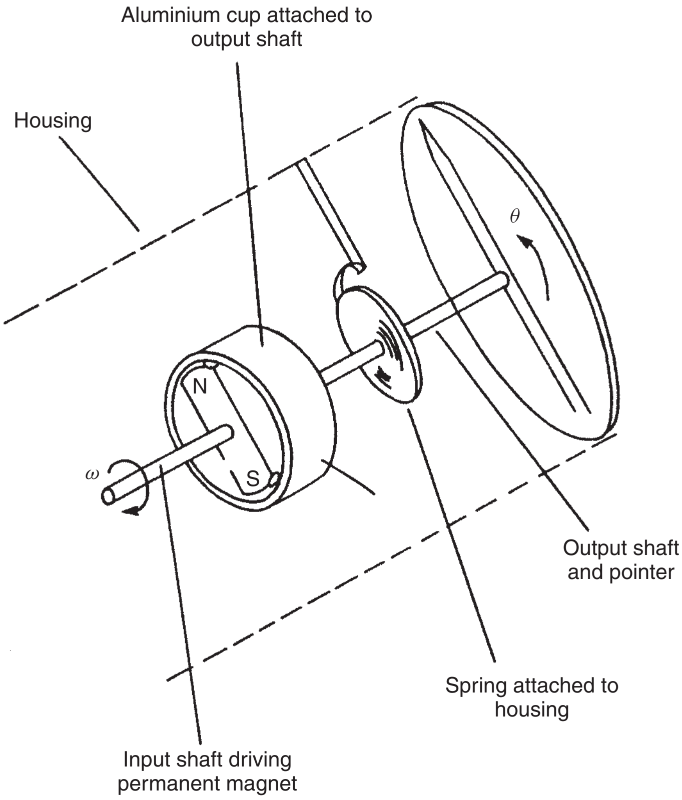


Fig. 20.21 Drag-cup tachometer.

Analogue-output forms of the *variable reluctance velocity transducer* (see section 20.2.1) also exist in which the output voltage pulses are converted into an analogue, varying-amplitude, d.c. voltage by means of a frequency-to-voltage converter circuit. However, the measurement accuracy is inferior to digital output forms.

20.2.4 Mechanical flyball

The mechanical flyball (alternatively known as a *centrifugal tachometer*) is a velocity-measuring instrument that was invented in 1817 and so might now be regarded as being old-fashioned. However, because it can act as a control actuator as well as a measuring instrument, it still finds substantial use in speed-governing systems for engines and turbines in which the measurement output is connected via a system of mechanical links to the throttle. The output is linear, typical measurement inaccuracy is $\pm 1\%$, and velocities up to 40 000 rpm can be measured. As shown in Figure 20.22, the device consists of a pair of spherical balls pivoted on the rotating shaft. These balls move outwards under the influence of centrifugal forces as the rotational velocity of the shaft increases and lift a pointer against the resistance of a spring. The pointer can be arranged to give a visual indication of speed by causing it to move in front

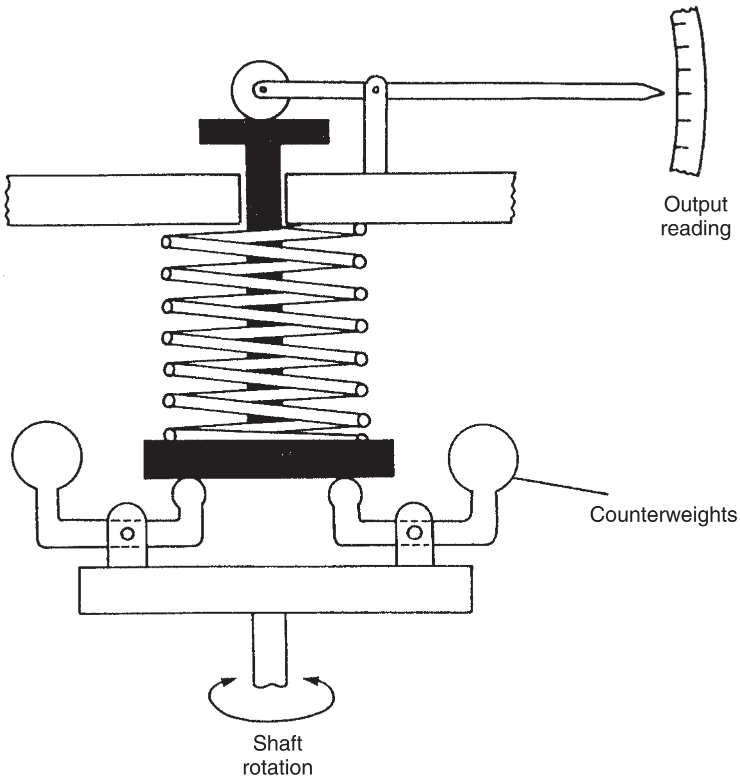


Fig. 20.22 Mechanical flyball.

of a calibrated scale, or its motion can be converted by a translational displacement transducer into an electrical signal.

In equilibrium, the centrifugal force, F_c, is balanced by the spring force, F_s, where:

$$F_c = K_c\omega^2; \quad F_s = K_s x$$

and K_c and K_s are constants, ω is the rotational velocity and x is the displacement of the pointer.

Thus:

$$K_c\omega^2 = K_s x \quad \text{or} \quad \omega = \sqrt{\left(\frac{K_s x}{K_c}\right)}$$

This is inconvenient because it involves a non-linear relationship between the pointer displacement and the rotational velocity. If this is not acceptable, a linear relationship can be obtained by using a spring with a non-linear characteristic such that $F_s = K_s' x^2$.

Then, equating expressions for F_c and F_s as before gives:

$$\omega = x\sqrt{\left(\frac{K_s'}{K_c}\right)}$$

20.2.5 The rate gyroscope

The rate gyro, illustrated in Figure 20.23, has an almost identical construction to the rate integrating gyro (Figure 20.14), and differs only by including a spring system which acts as an additional restraint on the rotational motion of the frame. The instrument measures the absolute angular velocity of a body, and is widely used in generating stabilizing signals within vehicle navigation systems. The typical measurement resolution given by the instrument is 0.01°/s and rotation rates up to 50°/s can be measured. The angular velocity, α, of the body is related to the angular deflection of the gyroscope, θ, by the equation:

$$\frac{\theta}{\alpha}(D) = \frac{H}{MD^2 + \beta D + K} \tag{20.2}$$

where H is the angular momentum of the spinning wheel, M is the moment of inertia of the system, β is the viscous damping coefficient, K is the spring constant, and D is the D-operator.

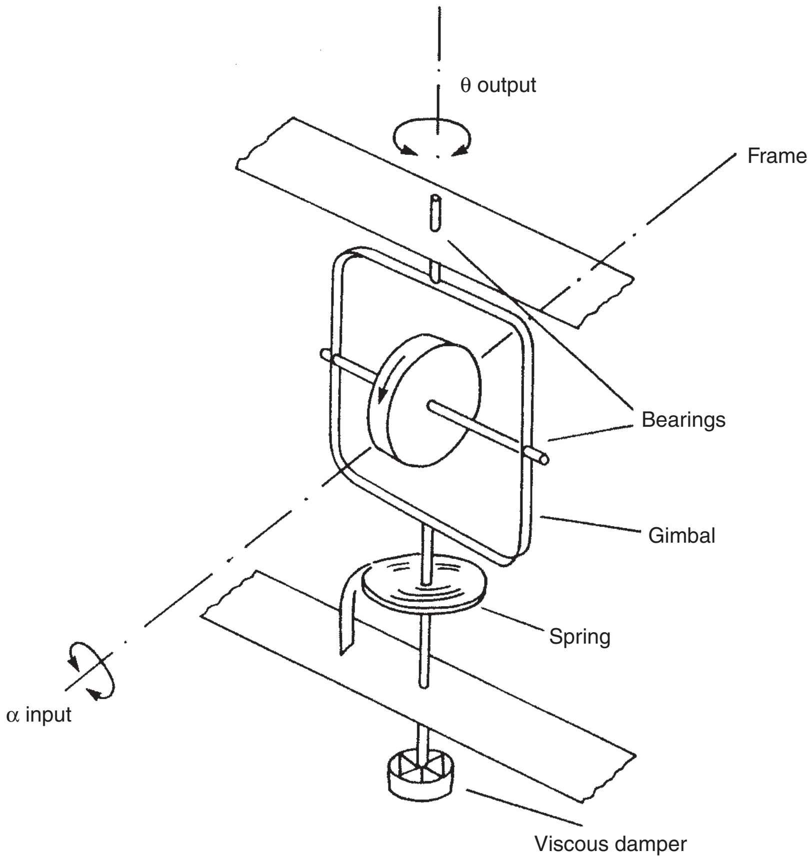


Fig. 20.23 Rate gyroscope.

This relationship (20.2) is a second order differential equation and therefore we must expect the device to have a response typical of second order instruments, as discussed in Chapter 2. The instrument must therefore be designed carefully so that the output response is neither oscillatory nor too slow in reaching a final reading. To assist in the design process, it is useful to re-express equation (20.2) in the following form:

$$\frac{\theta}{\alpha}(D) = \frac{K'}{D^2/\omega^2 + 2\xi D/\omega + 1} \tag{20.3}$$

where

$$K' = H/K, \omega = \sqrt{K/M} \quad \text{and} \quad \xi = \frac{\beta}{2\sqrt{KM}}$$

The static sensitivity of the instrument, K', is made as large as possible by using a high-speed motor to spin the wheel and so make H high. Reducing the spring constant K further improves the sensitivity but this cannot be reduced too far as it makes the resonant frequency ω of the instrument too small. The value of β is chosen such that the damping ratio ξ is close to 0.7.

20.2.6 Fibre-optic gyroscope

This is a relatively new instrument that makes use of fibre-optic technology. Incident light from a source is separated by a beam splitter into a pair of beams a and b, as shown in Figure 20.24. These travel in opposite directions around an optic-fibre coil (which may be several hundred metres long) and emerge from the coil as the beams marked a′ and b′. The beams a′ and b′ are directed by the beam splitter into an interferometer. Any motion of the coil causes a phase shift between a′ and b′ which is detected by the interferometer. Further details can be found in Nuttall (1987).

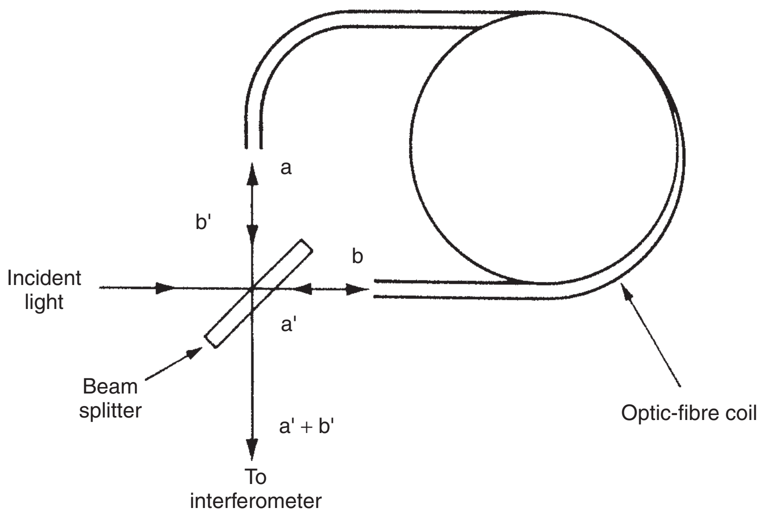


Fig. 20.24 Fibre-optic gyroscope.

20.2.7 Differentiation of angular displacement measurements

Angular velocity measurements can be obtained by differentiating the output signal from angular displacement transducers. Unfortunately, the process of differentiation amplifies any noise in the measurement signal, and therefore this technique has only rarely been used in the past. The technique has become more feasible with the advent of intelligent instruments, and one such instrument which processes the output of a resolver claims a maximum velocity measurement inaccuracy of $\pm1\%$ (Analogue Devices, 1988).

20.2.8 Integration of the output from an accelerometer

In measurement systems that already contain an angular acceleration transducer, it is possible to obtain a velocity measurement by integrating the acceleration measurement signal. This produces a signal of acceptable quality, as the process of integration attenuates any measurement noise. However, the method is of limited value in many measurement situations because the measurement obtained is the average velocity over a period of time, rather than a profile of the instantaneous velocities as motion takes place along a particular path.

20.2.9 Choice between rotational velocity transducers

Choice between different rotational velocity transducers is influenced strongly by whether an analogue or digital form of output is required. Digital output instruments are now widely used and a choice has to be made between the variable reluctance transducer, devices using electronic light pulse counting methods, and the stroboscope. The first two of these are used to measure angular speeds up to about 10 000 rpm and the last one can measure speeds up to 25 000 rpm.

Probably the most common form of analogue output device used is the d.c. tachometer. This is a relatively simple device that measures speeds up to about 5000 rpm with a maximum inaccuracy of $\pm1\%$. Where better accuracy is required within a similar range of speed measurement, a.c. tachometers are used. The squirrel-cage rotor type has an inaccuracy of only $\pm0.25\%$ and drag-cup rotor types can have inaccuracies as low as $\pm0.05\%$.

Other devices with an analogue output that are also sometimes used are the drag-cup tachometer and the mechanical flyball. The drag-cup tachometer has a typical inaccuracy of $\pm5\%$ but it is cheap and therefore very suitable for use in vehicle speedometers. The Mechanical flyball has a better accuracy of $\pm1\%$ and is widely used in speed governors, as noted earlier.

20.3 Measurement of rotational acceleration

Rotational accelerometers work on very similar principles to translational motion accelerometers. They consist of a rotatable mass mounted inside a housing that is attached to the accelerating, rotating body. Rotation of the mass is opposed by a torsional spring and damping. Any acceleration of the housing causes a torque $J\ddot{\theta}$ on the mass. This torque is opposed by a backward torque due to the torsional spring and in equilibrium:

$$J\ddot{\theta} = K\theta \quad \text{and hence:} \quad \ddot{\theta} = k\theta/J$$

A damper is usually included in the system to avoid undying oscillations in the instrument. This adds an additional backward torque $B\dot{\theta}$ to the system and the equation of motion becomes:

$$J\ddot{\theta} = B\dot{\theta} + K\theta$$

References and further reading

Analogue Devices (1988) Resolver to digital converter, *Measurement and Control*, **21**(10), p. 291.

Healey, M. (1975) *Principles of Automatic Control*, Hodder and Stoughton, London.

Nuttall, J.D. (1987) Optical gyroscopes, *Electronics and Power*, **33**(11), pp. 703–707.

21

Summary of other measurements

21.1 Dimension measurement

Dimension measurement includes measurement of the length, width and height of components and also the depth of holes and slots. Tapes and rules are commonly used to give approximate measurements, and various forms of calliper and micrometer are used where more accurate measurements are required. Gauge blocks and length bars are also used when very high accuracy is required, although these are primarily intended for calibration duties.

A flat and level *reference plane*, on which components being measured are placed, is often essential in dimension measurement. Such reference planes are available in a range of standard sizes, and a means of adjusting the feet is always provided to ensure that the surface is exactly level. Smaller sizes exist as a *surface plate* resting on a supporting table, whereas larger sizes take the form of free standing tables that usually have a projection at the edge to facilitate the clamping of components. They are normally used in conjunction with box cubes and vee blocks (see Figure 21.1) that locate components in a fixed position. In modern tables, granite has tended to supersede iron as the preferred material for the plate, although iron plates are still available. Granite is ideal for this purpose as it does not corrode, is dimensionally very stable and does not form burrs when damaged. Iron plates, on the other hand, are prone to rusting and susceptible to damage: this results in burrs on the surface that interfere with measurement procedures.

21.1.1 Rules and tapes

Rules and tapes are the simplest way of measuring larger dimensions. Steel rules are generally only available to measure dimensions up to 1 metre. Beyond this, steel tapes (measuring to 30 m) or an ultrasonic rule (measuring to 10 m) are used.

The *steel rule* is undoubtedly the simplest instrument available for measuring length. Measurement accuracy is only modest using standard rules, which typically have rulings at 0.5 mm intervals, but the best rules have rulings at 0.05 mm intervals and a measurement resolution of 0.02 mm. When used by placing the rule against an object,

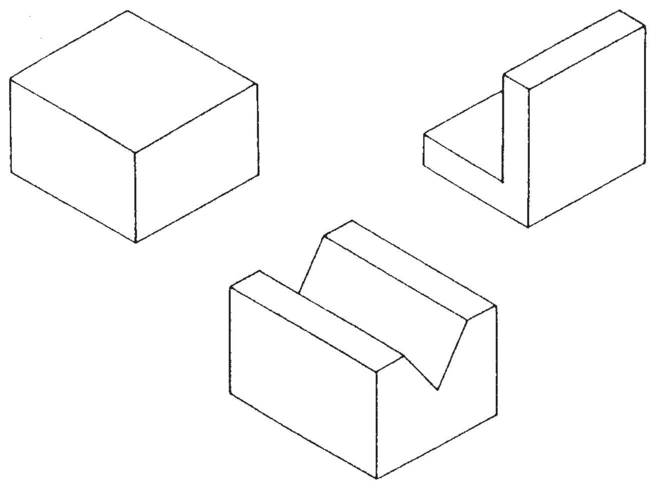


Fig. 21.1 Box cubes and vee blocks.

the measurement accuracy is much dependent upon the skill of the human measurer and, at best, the inaccuracy is likely to be at least $\pm 0.5\%$.

The retractable *steel tape* is another well-known instrument. The end of the tape is usually provided with a flat hook that is loosely fitted so as to allow for automatic compensation of the hook thickness when the rule is used for internal measurements. Again, measurement accuracy is governed by human skill, but, with care, the measurement inaccuracy can be made to be as low as $\pm 0.01\%$ of full-scale reading.

The *ultrasonic rule* consists of an ultrasonic energy source, an ultrasonic energy detector and battery-powered, electronic circuitry housed within a hand-held box, as shown in Figure 21.2. Both source and detector often consist of the same type of piezoelectric crystal excited at a typical frequency of 40 kHz. Energy travels from the source to a target object and is then reflected back into the detector. The time of flight of this energy is measured and this is converted into a distance reading by the enclosed electronics. Maximum measurement inaccuracy of $\pm 1\%$ of the full-scale reading is claimed. This is only a modest level of accuracy, but it is sufficient for such purposes as measuring rooms by estate agents prior to producing sales literature, where the ease and speed of making measurements is of great value.

A fundamental problem in the use of ultrasonic energy of this type is the limited measurement resolution (7 mm) imposed by the 7 mm wavelength of sound at this frequency. Further problems are caused by the variation in the speed of sound with humidity (variations of $\pm 0.5\%$ possible) and the temperature-induced variation of 0.2% per °C. Therefore, the conditions of use must be carefully controlled if the claimed accuracy figure is to be met.

21.1.2 Callipers

Callipers are generally used in situations where measurement of dimensions using a rule or tape is not accurate enough. Two types exist, the standard calliper and the vernier calliper.

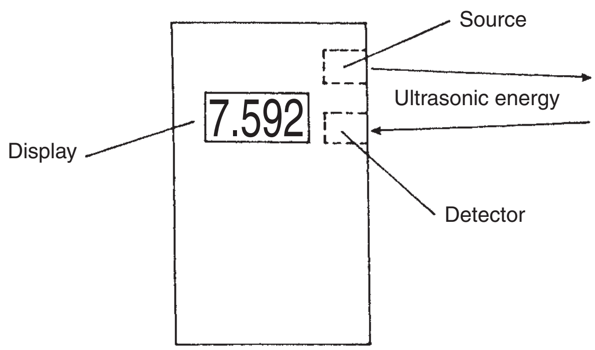


Fig. 21.2 Ultrasonic rule.

Figure 21.3 shows two types of *standard calliper*. The range of measurement, according to the version used, is up to 600 mm. These are used to transfer the measured dimension from the workpiece to a steel rule. This avoids the necessity to align the end of the rule exactly with the edge of the workpiece and reduces the measurement inaccuracy by a factor of two. In the basic calliper, careless use can allow the setting of the calliper to be changed during transfer from the workpiece to the rule. Hence, the spring-loaded type, which prevents this happening, is preferable.

The *vernier calliper*, shown in Figure 21.4(a), is a combination of a standard calliper and a steel rule. The main body of the instrument includes the main scale with a fixed anvil at one end. This carries a sliding anvil that is provided with a second, vernier scale. This second scale is shorter than the main scale and is divided into units that are slightly smaller than the main scale units but related to them by a fixed factor. Determination of the point where the two scales coincide enables very accurate measurements to be made, with typical inaccuracy levels down to $\pm 0.01\%$.

Figure 21.4(b) shows details of a typical combination of main and vernier scales. The main scale is ruled in 1 mm units. The vernier scale is 49 mm long and divided into 50 units, thereby making each unit 0.02 mm smaller than the main scale units. Each group of five units on the vernier scale thus differs from the main scale by 0.1 mm and the numbers marked on the scale thus refer to these larger units of 0.1 mm. In

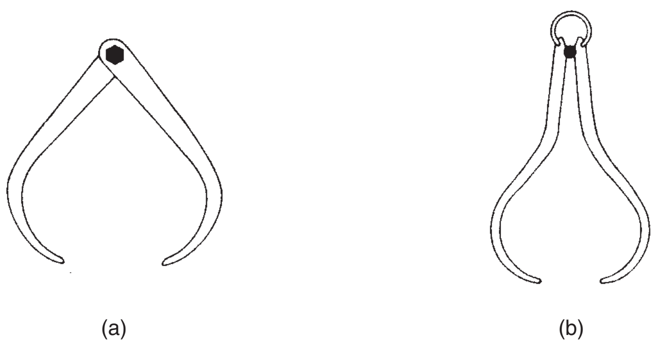
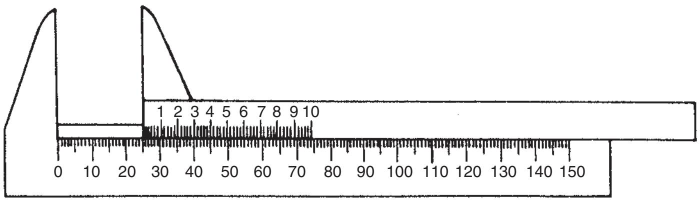


Fig. 21.3 (a) Standard calliper; (b) spring-loaded calliper.



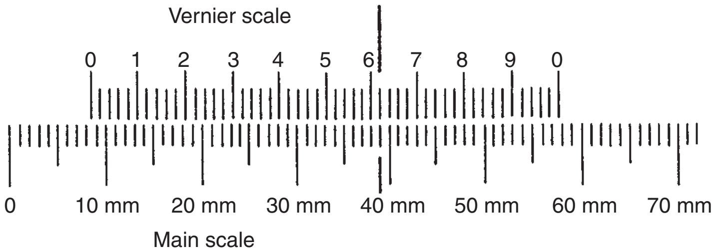


Fig. 21.4 Vernier calliper: (a) basic instrument; (b) details of scale.

the particular position shown in the figure, the zero on the vernier scale is indicating a measurement between 8 and 9 mm. Both scales coincide at a position of 6.2 (large units). This defines the interval between 8 and 9 mm to be $6.2 \times 0.1 = 0.62$ mm, i.e. the measurement is 8.62 mm.

Intelligent digital callipers are now available that give a measurement resolution of 0.01 mm and a low inaccuracy of ± 0.03 mm. These have automatic compensation for wear, and hence calibration checks have to be very infrequent. In some versions, the digital display can be directly interfaced to an external computer monitoring system.

21.1.3 Micrometers

Micrometers provide a means of measuring dimensions to high accuracy. Different forms provide measurement of both internal and external dimensions of components, and of holes, slots etc. within components. In the *standard micrometer*, shown in Figure 21.5(a), measurement is made between two anvils, one fixed and one that is moved along by the rotation of an accurately machined screw thread. One complete rotation of the screw typically moves the anvil by a distance of 0.5 mm. Such movements of the anvil are measured using a scale marked with divisions every 0.5 mm along the barrel of the instrument. A scale marked with 50 divisions is etched around the circumference of the spindle holder: each division therefore corresponds to an axial movement of 0.01 mm. Assuming that the user is able to judge the position of the spindle on this circular scale against the datum mark to within one-fifth of a division, a measurement resolution of 0.002 mm is possible.

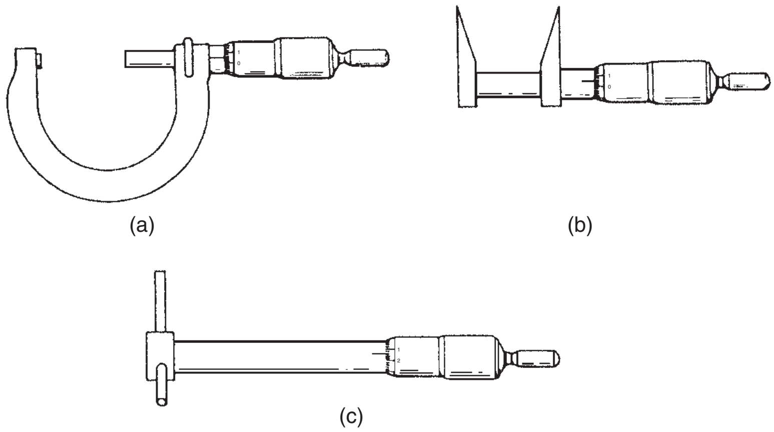


Fig. 21.5 Micrometers: (a) standard (external) micrometer; (b) internal micrometer; (c) bore micrometer.

The most common measurement ranges are either 0–25 mm or 25–50 mm, with inaccuracy levels down to ±0.003%. However, a whole family of micrometers is available, where each has a measurement span of 25 mm, but with the minimum distance measured varying from 0 mm up to 575 mm. Thus, the last instrument in this family measures the range from 575 to 600 mm. Some manufacturers also provide micrometers with two or more interchangeable anvils, which extend the span measurable with one instrument to between 50 mm and 100 mm according to the number of anvils supplied. Therefore, an instrument with four anvils might for instance measure the range from 300 mm to 400 mm, by making appropriate changes to the anvils.

The *internal micrometer* (see Figure 21.5(b)) is able to measure internal dimensions such as the diameters of holes. In the case of measuring holes, micrometers are inaccurate if there is any ovality in the hole, unless the diameter is measured at several points. An alternative solution to this problem is to use a special type of instrument known as a *bore micrometer* (Figure 21.5(c)). In this, three probes move out radially from the body of the instrument as the spindle is turned. These probes make contact with the sides of the hole at three equidistant points, thus averaging out any ovality.

Intelligent micrometers in the form of the electronic *digital micrometer* are now available. These have a self-calibration capability and a digital readout, with a measurement resolution of 0.001 mm (1 micron).

21.1.4 Gauge blocks (slip gauges) and length bars

Gauge blocks, also known as *slip gauges* (see Figure 21.6(a)), consist of rectangular blocks of hardened steel that have flat and parallel end faces. These faces are machined to very high standards of accuracy in terms of their surface finish and flatness. The purpose of gauge blocks is to provide a means of checking whether a particular dimension in a component is within the allowable tolerance rather than actually measuring what the dimension is. To do this, a number of gauge blocks are joined together to make up the required dimension to be checked. Gauge blocks are available in five grades of accuracy known as calibration, 00, 0, 1 and 2. Grades 1 and 2 are used for

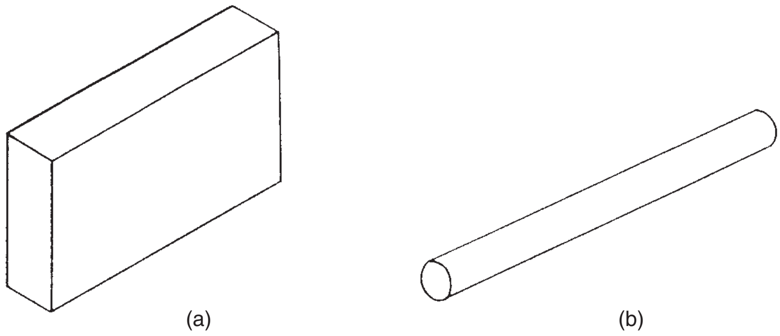


Fig. 21.6 (a) Gauge block; (b) length bar.

normal production and inspection measurements, with the other grades being intended only for calibration procedures at various levels.

Gauge blocks are available in boxed sets containing a range of block sizes, which allows any dimension up to 200 mm to be constructed by joining together an appropriate number of blocks. Whilst 200 mm is the maximum dimension that should be set up with gauge blocks alone, they can be used in conjunction with length bars to set up much greater standard dimensions. Blocks are joined by ‘wringing’, a procedure in which the two end faces are rotated slowly against each other. This removes the air film and allows adhesion to develop by intermolecular attraction. Adhesion is so good in fact that, if groups of blocks were not separated within a few hours, the molecular diffusion process would continue to the point where the blocks would be permanently welded together. The typical interblock gap resulting from wringing has been measured as 0.001 μm, which is effectively zero. Thus, any number of blocks can be joined without creating any significant measurement error. It is fairly common practice with blocks of grades 0, 1 and 2 to include an extra pair of 2 mm thick blocks in the set that are made from wear-resisting tungsten carbide. These are marked with a letter P and are designed to protect the other blocks from wear during use. Where such protector blocks are used, due allowance has to be made for their thickness (4 mm) in calculating the sizes of block needed to make up the required length.

A necessary precaution when using gauge blocks is to avoid handling them more than is necessary. The length of a bar that was 100 mm long at 20°C would increase to 100.02 mm at 37°C (body temperature). Hence, after wringing bars together, they should be left to stabilize back to the ambient room temperature before use. This wait might need to be several hours if the blocks have been handled to any significant extent.

Where a greater dimension than 200 mm is required, gauge blocks are used in conjunction with *length bars* (Figure 21.6(b)). Length bars consist of straight, hardened, high-quality steel bars of a uniform 22 mm diameter and in a range of lengths between 100 mm and 1200 mm. They are available in four grades of accuracy, *reference*, *calibration*, *grade 1* and *grade 2*. Reference and calibration grades have accurately flat end faces, which allows a number of bars to be wrung together to obtain the required standard length. Bars of grades 1 and 2 have threaded ends that allow them to be screwed together. Grade 2 bars are used for general measurement duties, with grade 1 bars being reserved for inspection duties. By combining length bars with gauge blocks, any dimension up to about 2 m can be set up with a resolution of 0.0005 mm.

21.1.5 Height and depth measurement

The height of objects and the depth of holes, slots etc. are measured by the height gauge and depth gauge respectively. A dial gauge is often used in conjunction with these instruments to improve measurement accuracy. The *height gauge*, shown in Figure 21.7(a), effectively consists of a vernier calliper mounted on a flat base. Measurement inaccuracy levels down to $\pm 0.015\%$ are possible. The *depth gauge* (Figure 21.7(b)) is a further variation on the standard vernier calliper principle that has the same measurement accuracy capabilities as the height gauge.

In practice, certain difficulties can arise in the use of these instruments where either the base of the instrument is not properly located on the measuring table or where the point of contact between the moving anvil and the workpiece is unclear. In such cases, a dial gauge, which has a clearly defined point of contact with the measured object, is used in conjunction with the height or depth gauge to avoid these possible sources of error. These instruments can also be obtained in intelligent versions that give a digital display and have self-calibration capabilities.

The *dial gauge*, shown in Figure 21.8(a), consists of a spring-loaded probe that drives a pointer around a circular scale via rack and pinion gearing. Typical measurement resolution is 0.01 mm. When used to measure the height of objects, it is clamped in a retort stand and a measurement taken of the height of the unknown component. Then it is put in contact with a height gauge (Figure 21.8(b)) that is adjusted until the reading on the dial gauge is the same. At this stage, the height gauge is set to the height

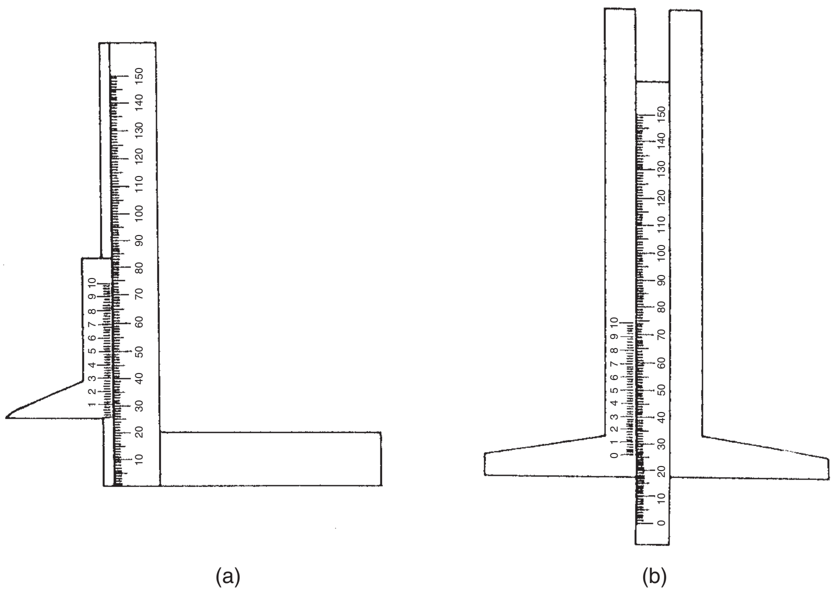
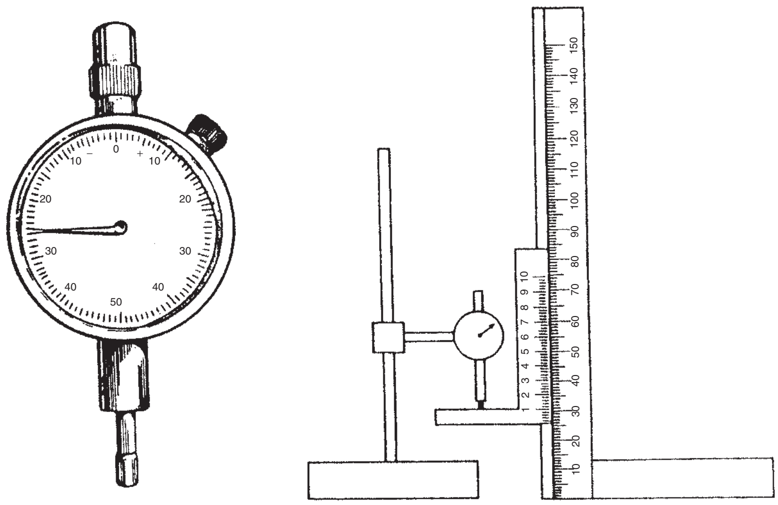


Fig. 21.7 (a) Height gauge; (b) depth gauge.



(a) (b)

Fig. 21.8 Dial gauge: (a) basic instrument; (b) use in conjunction with height gauge.

of the object. The dial gauge is also used in conjunction with the depth gauge in an identical manner. (Gauge blocks can be used instead of height/depth gauges in such measurement procedures if greater accuracy is required.)

21.2 Angle measurement

Measurement of angles is one of the less common measurement requirements that instrumentation technologists are likely to meet. However, angle measurement is required in some circumstances, such as when the angle between adjoining faces on a component must be checked. The main instruments used are protractors and a form of angle-measuring spirit level.

In some circumstances, a simple protractor of the sort used in school for geometry exercises can be used. However, the more sophisticated form of angle protractor shown in Figure 21.9 provides better measurement accuracy. This consists of two straight edges, one of which is able to rotate with respect to the other. Referring to Figure 21.9, the graduated circular scale A attached to the straight edge C rotates inside a fixed circular housing attached to the other straight edge B. The relative angle between the two straight edges in contact with the component being measured is determined by the position of the moving scale with respect to a reference mark on the fixed housing B. With this type of instrument, measurement inaccuracy is at least $\pm 1\%$. An alternative form, the *bevel protractor*, is similar to this form of angle protractor, but it has a vernier scale on the fixed housing. This allows the inaccuracy level to be reduced to ± 10 minutes of arc.

The *spirit level* shown in Figure 21.10 is an alternative angle-measuring instrument. It consists of a standard spirit level attached to a rotatable circular scale that is mounted

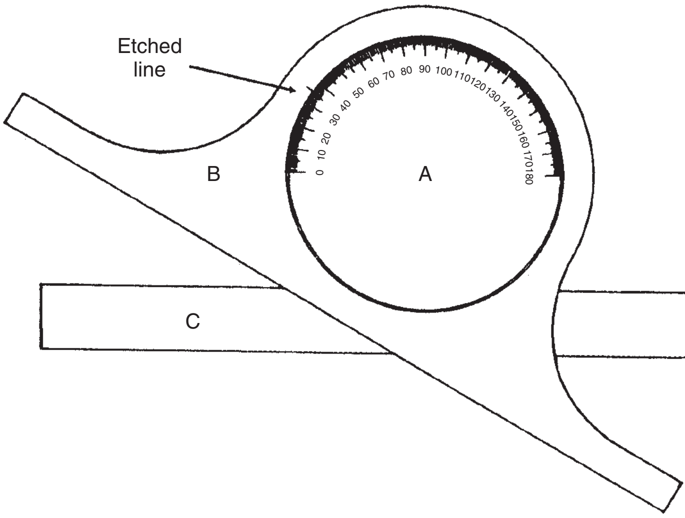


Fig. 21.9 Angle protractor.

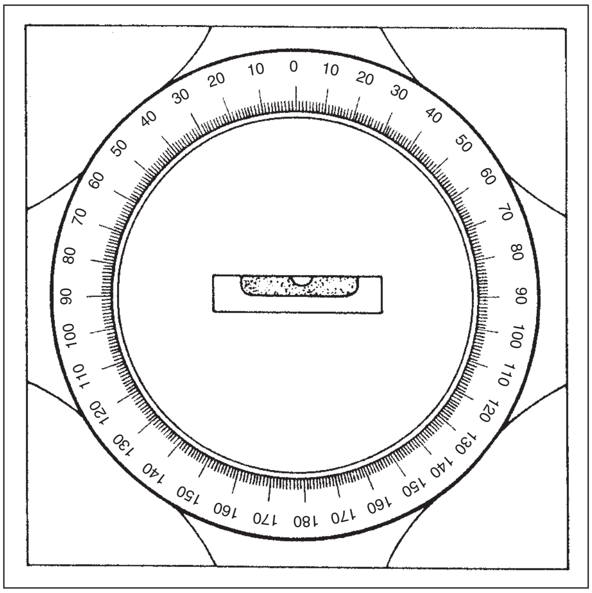


Fig. 21.10 Angle-measuring spirit level.

inside an accurately machined square frame. When placed on the sloping surfaces of components, rotation of the scale to centralize the bubble in the spirit level allows the angle of slope to be measured. Again, measuring inaccuracies down to ±10 minutes of arc are possible if a vernier scale is incorporated in the instrument.

The *electronic spirit level* contains a pendulum whose position is sensed electrically. Measurement resolution as good as 0.2 seconds of arc is possible.

21.3 Flatness measurement

The only dimensional parameter not so far discussed where a measurement requirement sometimes exists is the flatness of the surface of a component. This is measured by a *variation gauge*. As shown in Figure 21.11, this has four feet, three of which are fixed and one of which floats in a vertical direction. Motion of the floating foot is measured by a dial gauge that is calibrated such that its reading is zero when the floating foot is exactly level with the fixed feet. Thus, any non-zero reading on the dial gauge indicates non-flatness at the point of contact of the floating foot. By moving the variation gauge over the surface of a component and taking readings at various points, a contour map of the flatness of the surface can be obtained.

21.4 Volume measurement

Volume measurement is required in its own right as well as being required as a necessary component in some techniques for the measurement and calibration of other quantities such as volume flow rate and viscosity. The volume of vessels of a regular shape, where the cross-section is circular or oblong in shape, can be readily calculated from the dimensions of the vessel. Otherwise, for vessels of irregular shape, it is necessary to use either gravimetric techniques or a set of calibrated volumetric measures.

In the gravimetric technique, the dry vessel is weighed and then is completely filled with water and weighed again. The volume is then simply calculated from this weight difference and the density of water.

The alternative technique involves transferring the liquid from the vessel into an appropriate number of volumetric measures taken from a standard-capacity, calibrated set. Each vessel in the set has a mark that shows the volume of liquid contained when the vessel is filled up to the mark. Special care is needed to ensure that the meniscus of the water is in the correct position with respect to the reference mark on the vessel when it is deemed to be full. Normal practice is to set the water level such that the

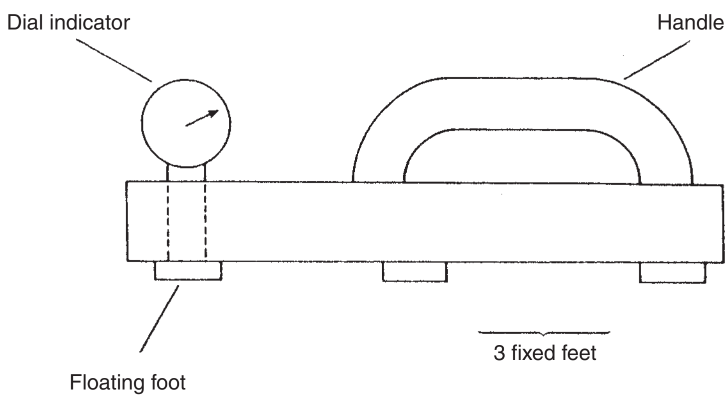


Fig. 21.11 Variation gauge.

Table 21.1 Typical measurement uncertainty of calibrated volumetric measures

Capacity	*Volumetric uncertainty*
1 ml	±4%
10 ml	±0.8%
100 ml	±0.2%
1 l	±0.1%
10 l	±0.05%
100 l	±0.02%
1000 l	±0.02%

reference mark forms a smooth tangent with the convex side of the meniscus. This is made easier to achieve if the meniscus is viewed against a white background and the vessel is shaded from stray illumination.

The measurement uncertainty using calibrated volumetric measures depends on the number of measures used for any particular measurement. The total error is a multiple of the individual error of each measure, typical values of which are shown in Table 21.1.

21.5 Viscosity measurement

Viscosity measurement is important in many process industries. In the food industry, the viscosity of raw materials such as dough, batter and ice cream has a direct effect on the quality of the product. Similarly, in other industries such as the ceramic one, the quality of raw materials affects the final product quality. Viscosity control is also very important in assembly operations that involve the application of mastics and glue flowing through tubes. Clearly, successful assembly requires such materials to flow through tubes at the correct rate and therefore it is essential that their viscosity is correct.

Viscosity describes the way in which a fluid flows when it is subject to an applied force. Consider an elemental cubic volume of fluid and a shear force F applied to one of its faces of area A. If this face moves a distance L and at a velocity V relative to the opposite face of the cube under the action of F, the shear stress (s) and shear rate (r) are given by:

$$s = F/A; \quad r = V/L$$

The *coefficient of viscosity* (C_V) is the ratio of shear stress to shear rate, i.e.

$$C_V = s/r.$$

C_V is often described simply as the 'viscosity'. A further term, *kinematic viscosity*, is also sometimes used, given by $K_V = C_V/\rho$, where K_V is the kinematic viscosity and ρ is the fluid density. To avoid confusion, C_V is often known as the *dynamic viscosity*, to distinguish it from K_V. C_V is measured in units of poise or Ns/m^2 and K_V is measured in units of stokes or m^2/s.

Viscosity was originally defined by Newton, who assumed that it was constant with respect to shear rate. However, it has since been shown that the viscosity of many fluids varies significantly at high shear rates and the viscosity of some varies even at low shear rates. The worst non-Newtonian characteristics tend to occur with emulsions, pastes and slurries. For non-Newtonian fluids, subdivision into further classes can also be made according to the manner in which the viscosity varies with shear rate, as shown in Figure 21.12.

The relationship between the input variables and output measurement for instruments that measure viscosity normally assumes that the measured fluid has Newtonian characteristics. For non-Newtonian fluids, a correction must be made for shear rate variations (see Miller, 1975a). If such a correction is not made, the measurement obtained is known as the *apparent viscosity*, and this can differ from the true viscosity by a large factor. The true viscosity is often called the *absolute viscosity* to avoid ambiguity. Viscosity also varies with fluid temperature and density.

Instruments for measuring viscosity work on one of three physical principles:

- Rate of flow of the liquid through a tube
- Rate of fall of a body through the liquid
- Viscous friction force exerted on a rotating body.

21.5.1 Capillary and tube viscometers

These are the most accurate types of viscometer, with typical measurement inaccuracy levels down to ±0.3%. Liquid is allowed to flow, under gravity from a reservoir, through a tube of known cross-section. In different instruments, the tube can vary from capillary-sized to a large diameter. The pressure difference across the ends of the tube and the time for a given quantity of liquid to flow are measured, and then the

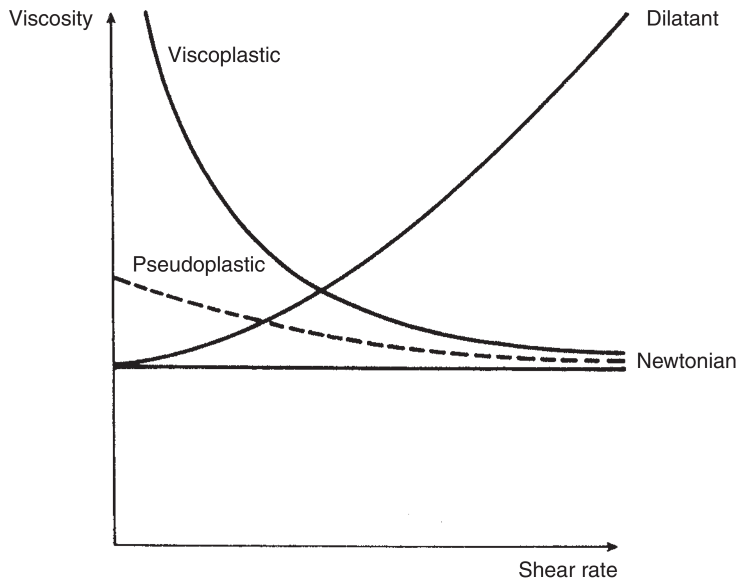


Fig. 21.12 Different viscosity/shear-rate relationships.

liquid viscosity for Newtonian fluids can be calculated as (in units of poise):

$$C_V = \frac{1.25\pi R^4 PT}{LV}$$

where R is the radius (m) of the tube, L is its length (m), P is the pressure difference (N/m^2) across the ends and V is the volume of liquid flowing in time T (m^3/s).

For non-Newtonian fluids, corrections must be made for shear rate variations (Miller, 1975a). For any given viscometer, R, L and V are constant and equation (21.1) can be written as:

$$C_V = KPT$$

where K is known as the viscometer constant.

21.5.2 Falling body viscometer

The falling body viscometer is particularly recommended for the measurement of high-viscosity fluids. It can give measurement uncertainty levels down to $\pm 1\%$. It involves measuring the time taken for a spherical body to fall a given distance through the liquid. The viscosity for Newtonian fluids is then given by Stoke's formula as (in units of poise):

$$C_V = \frac{R^2 g(\rho_s - \rho_l)}{450\,V}$$

where R is the radius (m) of the sphere, g is the acceleration due to gravity (m/s^2), ρ_s and ρ_l are the specific gravities (g/m^3) of the sphere and liquid respectively and V is the velocity (m/s) of the sphere.

For non-Newtonian fluids, correction for the variation in shear rate is very difficult.

21.5.3 Rotational viscometers

Rotational viscometers are relatively easy to use but their measurement inaccuracy is at least $\pm 10\%$. All types have some form of element rotating inside the liquid at a constant rate. One common version has two coaxial cylinders with the fluid to be measured contained between them. One cylinder is driven at a constant angular velocity by a motor and the other is suspended by torsion wire. After the driven cylinder starts from rest, the suspended cylinder rotates until an equilibrium position is reached where the force due to the torsion wire is just balanced by the viscous force transmitted through the liquid. The viscosity (in poise) for Newtonian fluids is then given by:

$$C_V = 2.5G\left(\frac{1/R_1^2 - 1/R_2^2}{\pi h\omega}\right)$$

where G is the couple (Nm) formed by the force exerted by the torsion wire and its deflection, R_1 and R_2 are the radii (m) of the inner and outer cylinders, h is the length of the cylinder (m) and ω is the angular velocity (rad/s) of the rotating cylinder. Again, corrections have to be made for non-Newtonian fluids.

21.6 Moisture measurement

There are many industrial requirements for the measurement of the moisture content. This can be required in solids, liquids or gases. The physical properties and storage stability of most solid materials is affected by their water content. There is also a statutory requirement to limit the moisture content in the case of many materials sold by weight. In consequence, the requirement for moisture measurement pervades a large number of industries involved in the manufacture of foodstuffs, pharmaceuticals, cement, plastics, textiles and paper.

Measurement of the water content in liquids is commonly needed for fiscal purposes, but is also often necessary to satisfy statutory requirements. The petrochemical industry has wide-ranging needs for moisture measurement in oil etc. The food industry also needs to measure the water content of products such as beer and milk.

In the case of moisture in gases, the most common measurement is the amount of moisture in air. This is usually known as the humidity level. Humidity measurement and control is an essential requirement in many buildings, greenhouses and vehicles.

As there are several ways in which humidity can be defined, three separate terms have evolved so that ambiguity can be avoided. *Absolute humidity* is the mass of water in a unit volume of moist air; *specific humidity* is the mass of water in a unit mass of moist air; *relative humidity* is the ratio of the actual water vapour pressure in air to the saturation vapour pressure, usually expressed as a percentage.

21.6.1 Industrial moisture measurement techniques

Industrial methods for measuring moisture are based on the variation of some physical property of the material with moisture content. Many different properties can be used and therefore the range of available techniques, as listed below, is large.

Electrical methods

Measuring the amount of absorption of *microwave energy* beamed through the material is the most common technique for measuring moisture content and is described in detail in Anderson (1989), and Thompson (1989). Microwaves at wavelengths between 1 mm and 1 m are absorbed to a much greater extent by water than most other materials. Wavelengths of 30 mm or 100 mm are commonly used because 'off-the-shelf' equipment to produce these is readily available from instrument suppliers. The technique is suitable for moisture measurement in solids, liquids and gases at moisture-content levels up to 45% and measurement uncertainties down to $\pm 0.3\%$ are possible.

The *capacitance moisture meter* uses the principle that the dielectric constant of materials varies according to their water content. Capacitance measurement is therefore related to moisture content. The instrument is useful for measuring moisture-content levels up to 30% in both solids and liquids, and measurement uncertainty down to $\pm 0.3\%$ has been claimed for the technique (Slight, 1989). Drawbacks of the technique include (a) limited measurement resolution owing to the difficulty in measuring small changes in a relatively large standing capacitance value and (b) difficulty when the sample has a high electrical conductivity. An alternative capacitance charge transfer technique has been reported (Gimson, 1989) that overcomes these problems by measuring the charge carrying capacity of the material. In this technique, wet and dry samples of the material are charged to a fixed voltage and then simultaneously discharged into charge-measuring circuits.

The *electrical conductivity* of most materials varies with moisture content and this therefore provides another means of measurement. Techniques using electrical conductivity variation are cheap and can measure moisture levels up to 25%. However, the presence of other conductive substances in the material such as salts or acids affects the measurement.

A further technique is to measure the frequency change in a *quartz crystal* that occurs as it takes in moisture.

Neutron moderation

Neutron moderation measures moisture content using a radioactive source and a neutron counter. Fast neutrons emitted from the source are slowed down by hydrogen nuclei in the water, forming a cloud whose density is related to the moisture content. Measurements take a long time because the output density reading may take up to a few minutes to reach steady state, according to the nature of the materials involved. Also, the method cannot be used with any materials that contain hydrogen molecules, such as oils and fats, as these slow down neutrons as well. Specific humidities up to 15% ($\pm 1\%$ error) can be measured.

Low resolution nuclear magnetic resonance (NMR)

Low-resolution nuclear magnetic resonance involves subjecting the sample to both an unidirectional and an alternating radio-frequency (RF) magnetic field. The amplitude of the unidirectional field is varied cyclically, which causes resonance once per cycle in the coil producing the RF field. Under resonance conditions, protons are released from the hydrogen content of the water in the sample. These protons cause a measurable moderation of the amplitude of the RF oscillator waveform that is related to the moisture content of the sample. The technique is described more fully in Young (1989).

Materials that naturally have a hydrogen content cannot normally be measured. However, pulsed NMR techniques have been developed that overcome this problem by taking advantage of the different relaxation times of hydrogen nuclei in water and oil. In such pulsed techniques, the dependence on the relaxation time limits the maximum fluid flow rate for which moisture can be measured.

Optical methods

The *refractometer* is a well-established instrument that is used for measuring the water content of liquids. It measures the refractive index of the liquid, which changes according to the moisture content.

Moisture-related *energy absorption* of near-infra-red light can be used for measuring the moisture content of solids, liquids and gases. At a wavelength of 1.94 μm, energy absorption due to moisture is high, whereas at 1.7 μm, absorption due to moisture is zero. Therefore, measuring absorption at both 1.94 μm and 1.7 μm allows absorption due to components in the material other than water to be compensated for, and the resulting measurement is directly related to energy content. The latest instruments use multiple-frequency infra-red energy and have an even greater capability for eliminating the effect of components in the material other than water that absorb energy. Such multi-frequency instruments also cope much better with variations in particle size in the measured material.

In alternative versions of this technique, energy is either transmitted through the material or reflected from its surface. In either case, materials that are either very dark or highly reflective give poor results. The technique is particularly attractive, where applicable, because it is a non-contact method that can be used to monitor moisture content continuously at moisture levels up to 50%, with inaccuracy as low as ±0.1% in the measured moisture level. A deeper treatment can be found in Benson (1989).

Ultrasonic methods

The presence of water changes the speed of propagation of ultrasonic waves through liquids. The moisture content of liquids can therefore be determined by measuring the transmission speed of ultrasound. This has the inherent advantage of being a non-invasive technique but temperature compensation is essential because the velocity of ultrasound is particularly affected by temperature changes. The method is best suited to measurement of high moisture levels in liquids that are not aerated or of high viscosity. Typical measurement uncertainty is ±1% but measurement resolution is very high, with changes in moisture level as small as 0.05% being detectable. Further details can be found in Wiltshire (1989).

Mechanical properties

Density changes in many liquids and slurries can be measured and related to moisture content, with good measurement resolution up to 0.2% moisture. Moisture content can also be estimated by measuring the moisture level-dependent viscosity of liquids, pastes and slurries.

21.6.2 Laboratory techniques for moisture measurement

Laboratory techniques for measuring moisture content generally take much longer to obtain a measurement than the industrial techniques described above. However, the measurement accuracy obtained is usually much better.

Water separation

Various laboratory techniques are available that enable the moisture content of liquids to be measured accurately by separating the water from a sample of the host liquid. Separation is effected either by titration (Karl Fischer technique), distillation (Dean and Stark technique) or a centrifuge. Any of these methods can measure water content in a liquid with measurement uncertainty levels down to ±0.03%.

Gravimetric methods

Moisture content in solids can be measured accurately by weighing the moist sample, drying it and then weighing again. Great care must be taken in applying this procedure, as many samples rapidly take up moisture again if they are removed from the drier and exposed to the atmosphere before being weighed. Normal procedure is to put the sample in an open container, dry it in an oven and then screw an airtight top onto the container before it is removed from the oven.

Phase-change methods

The boiling and freezing point of materials is altered by the presence of moisture, and therefore the moisture level can be determined by measuring the phase-change temperature. This technique is used for measuring the moisture content in many food products and in some oil and alcohol products.

Equilibrium relative humidity measurement

This technique involves placing a humidity sensor in close proximity to the sample in an airtight container. The water vapour pressure close to the sample is related to the moisture content of the sample. The moisture level can therefore be determined from the humidity measurement.

21.6.3 Humidity measurement

The three major instruments used for measuring humidity in industry are the electrical hygrometer, the psychrometer and the dew point meter. The dew point meter is the most accurate of these and is commonly used as a calibration standard. The various types of hygrometer are described more fully in Miller (1975b).

The electrical hygrometer

The electrical hygrometer measures the change in capacitance or conductivity of a hygroscopic material as its moisture level changes. Conductivity types use two noble metal electrodes either side of an insulator coated in a hygroscopic salt such as calcium chloride. Capacitance types have two plates either side of a hygroscopic dielectric such as aluminium oxide.

These instruments are suitable for measuring moisture levels between 15% and 95%, with typical measurement uncertainty of ±3%. Atmospheric contaminants and operation in saturation conditions both cause characteristics drift, and therefore the recalibration frequency has to be determined according to the conditions of use.

The psychrometer (wet and dry bulb hygrometer)

The psychrometer, also known as the wet and dry bulb hygrometer, has two temperature sensors, one exposed to the atmosphere and one enclosed in a wet wick. Air is blown across the sensors, which causes evaporation and a reduction in temperature in the wet sensor. The temperature difference between the sensors is related to the humidity level. The lowest measurement uncertainty attainable is ±4%.

Dew point meter

The elements of the dew point meter, also known as the dew point hygrometer, are shown in Figure 21.13. The sample is introduced into a vessel with an electrically cooled mirror surface. The mirror surface is cooled until a light source-light detector system detects the formation of dew on the mirror, and the condensation temperature is measured by a sensor bonded to the mirror surface. The dew point is the temperature at which the sample becomes saturated with water. Therefore, this temperature is related to the moisture level in the sample. A microscope is also provided in the instrument

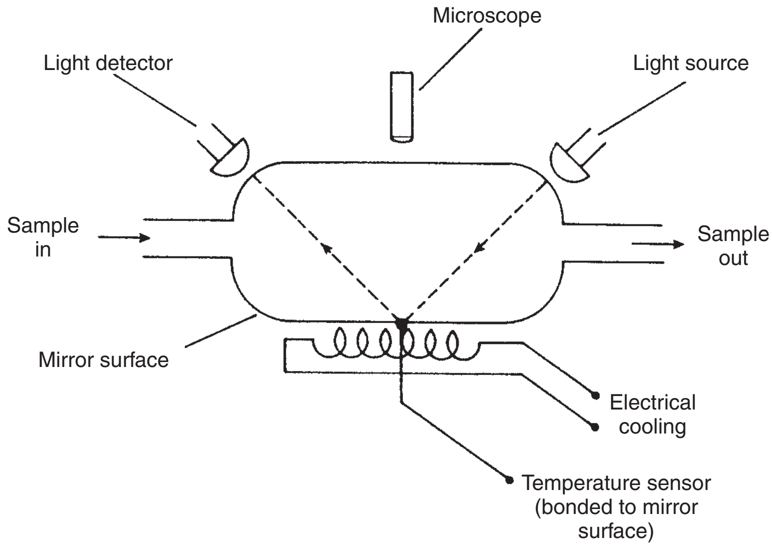


Fig. 21.13 Dew point meter.

so that the thickness and nature of the condensate can be observed. The instrument is described in greater detail in Pragnell (1989).

Even small levels of contaminants on the mirror surface can cause large changes in the dew point and therefore the instrument must be kept very clean. When necessary, the mirror should be cleaned with deionized or distilled water applied with a lint-free swab. Any contamination can be detected by a skilled operator, as this makes the condensate look 'blotchy' when viewed through the microscope. The microscope also shows up other potential problems such as large ice crystals in the condensate that cause temperature gradients between the condensate and the temperature sensor. When used carefully, the instrument is very accurate and is often used as a reference standard.

21.7 Sound measurement

Noise can arise from many sources in both industrial and non-industrial environments. Even low levels of noise can cause great annoyance to the people subjected to it and high levels of noise can actually cause hearing damage. Apart from annoyance and possible hearing loss, noise in the workplace also causes loss of output where the persons subjected to it are involved in tasks requiring high concentration. Extreme noise can even cause material failures through fatigue stresses set up by noise-induced vibration.

Various items of legislation exist to control the creation of noise. Court orders can be made against houses or factories in a neighbourhood that create noise exceeding a certain acceptable level. In extreme cases, where hearing damage may be possible, health and safety legislation comes into effect. Such legislation clearly requires the existence of accurate methods of quantifying sound levels. Sound is measured in terms

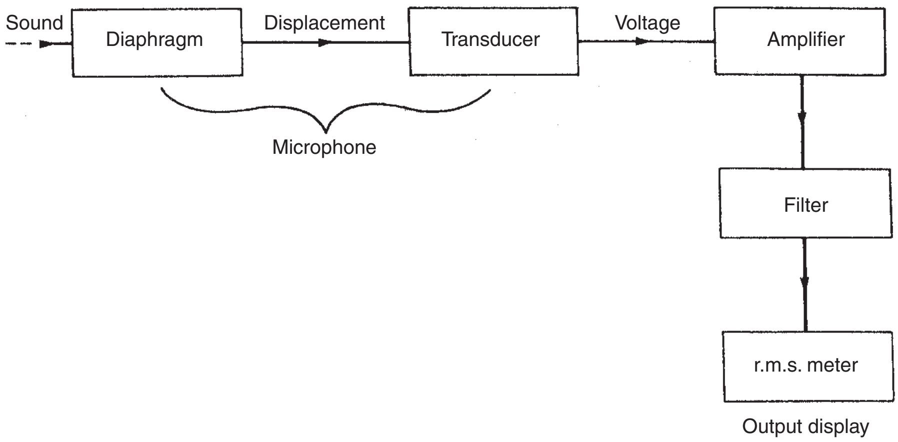


Fig. 21.14 Sound meter.

of the *sound pressure level*, S_P, which is defined as:

$$S_P = 20\ \log_{10}\left(\frac{P}{0.0002}\right) \text{ decibels (dB)}$$

where P is the r.m.s. sound pressure in μbar.

The quietest sound that the average human ear can detect is a tone at a frequency of 1 kHz and sound pressure level of 0 dB (2×10^{-4}μbar). At the upper end, sound pressure levels of 144 dB (3.45 mbar) cause physical pain.

Sound is usually measured with a sound meter. This essentially processes the signal collected by a microphone, as shown in Figure 21.14. The microphone is a diaphragm-type pressure-measuring device that converts sound pressure into a displacement. The displacement is applied to a displacement transducer (normally capacitive, inductive or piezoelectric type) which produces a low magnitude voltage output. This is amplified, filtered and finally gives an output display on an r.m.s. meter. The filtering process has a frequency response approximating that of the human ear so that the sound meter 'hears' sounds in the same way as a human ear. In other words, the meter selectively attenuates frequencies according to the sensitivity of the human ear at each frequency, so that the sound level measurement output accurately reflects the sound level heard by humans. If sound level meters are being used to measure sound to predict vibration levels in machinery, then they are used without filters so that the actual rather than the human-perceived sound level is measured.

21.8 pH measurement

pH is a parameter that quantifies the level of acidity or akalinity in a chemical solution. It defines the concentration of hydrogen atoms in the solution in grams/litre and is expressed as:

$$\text{pH} = \log_{10}[1/H^+]$$

where H^+ is the hydrogen ion concentration in the solution.

The value of pH can range from 0, which describes extreme acidity, to 14, which describes extreme akalinity. Pure water has a pH of 7. pH measurement is required in many process industries, and especially those involving food and drink production. The most universally known method of measuring pH is to use litmus paper or some similar chemical indicator that changes colour according to the pH value. Unfortunately, this method gives only a very approximate indication of pH unless used under highly controlled laboratory conditions. Much research is ongoing into on-line pH sensors and the various activities are described later. However, at the present time, the device known as the glass electrode is by far the most common on-line sensor used.

21.8.1 The glass electrode

The glass electrode consists of a glass probe containing two electrodes, a measuring one and a reference one, separated by a solid glass partition. Neither of the electrodes is in fact glass. The reference electrode is a screened electrode, immersed in a buffer solution, which provides a stable reference e.m.f. that is usually 0 V. The tip of the measuring electrode is surrounded by a pH-sensitive glass membrane at the end of the probe, which permits the diffusion of ions according to the hydrogen ion concentration in the fluid outside the probe. The measuring electrode therefore generates an e.m.f. proportional to pH that is amplified and fed to a display meter. The characteristics of the glass electrode are very dependent on ambient temperature, with both zero drift and sensitivity drift occurring. Thus, temperature compensation is essential. This is normally achieved through calibrating the system output before use by immersing the probe in solutions at reference pH values. Whilst being theoretically capable of measuring the full range of pH values between 0 and 14, the upper limit in practice is generally a pH value of about 12 because electrode contamination at very high alkaline concentrations becomes a serious problem and also glass starts to dissolve at such high pH values. Glass also dissolves in acid solutions containing fluoride, and this represents a further limitation in use. If required, the latter problems can be overcome to some extent by using special types of glass.

Great care is necessary in the use of the glass electrode type of pH probe. Firstly, the measuring probe has a very high resistance (typically 108 Ω) and a very low output. Hence, the output signal from the probes must be electrically screened to prevent any stray pick-up and electrical insulation of the assembly must be very high. The assembly must also be very efficiently sealed to prevent the ingress of moisture.

A second problem with the glass electrode is the deterioration in accuracy that occurs as the glass membrane becomes coated with various substances it is exposed to in the measured solution. Cleaning at prescribed intervals is therefore necessary and this must be carried out carefully, using the correct procedures, to avoid damaging the delicate glass membrane at the end of the probe. The best cleaning procedure varies according to the nature of the contamination. In some cases, careful brushing or wiping is adequate, whereas in other cases spraying with chemical solvents is necessary. Ultrasonic cleaning is often a useful technique, though it tends to be expensive. Steam cleaning should not be attempted, as this damages the pH-sensitive membrane. Mention must also be made about storage. The glass electrode must not be allowed to dry out during storage, as this would cause serious damage to the pH-sensitive layer.

Finally, caution must be taken about the response time of the instrument. The glass electrode has a relatively large time constant of one to two minutes, and so it must be left to settle for a long time before the reading is taken. If this causes serious difficulties, special forms of low-resistivity glass electrode are now available that have smaller time constants.

21.8.2 Other methods of pH measurement

Whilst the glass electrode predominates at present in pH measurement, several other devices and techniques exist. Whilst most of these are still under development and unproven in long-term use, a few are in practical use, especially for special measurement situations.

One alternative, which is in current use, is the antimony electrode. This is of a similar construction to the glass electrode but uses antimony instead of glass. The device is more robust than the glass electrode and can be cleaned by rubbing it with emery cloth. However, its time constant is very large and its output response is grossly non-linear, limiting its application to environments where the glass electrode is unsuitable. Such applications include acidic environments containing fluoride and environments containing very abrasive particles. The normal measurement range is pH 1 to 11.

A fibre-optic pH sensor is another available device, as described earlier in Chapter 13, in which the pH level is indicated by the intensity of light reflected from the tip of a probe coated in a chemical indicator whose colour changes with pH. Unfortunately, this device only has the capability to measure over a very small range of pH (typically 2 pH) and it has a short life.

21.9 Gas sensing and analysis

Gas sensing and analysis is required in many applications. A primary role of gas sensing is in hazard monitoring to predict the onset of conditions where flammable gases are reaching dangerous concentrations. Danger is quantified in terms of the *lower explosive level*, which is usually reached when the concentration of gas in air is in the range of between 1% and 5%.

Gas sensing also provides a fire detection and prevention function. When materials burn, a variety of gaseous products result. Most sensors that are used for fire detection measure carbon monoxide concentration, as this is the most common combustion product. Early fire detection enables fire extinguishing systems to be triggered, preventing serious damage from occurring in most cases. However, fire prevention is even better than early fire detection, and solid-state sensors, based on a sintered mass of polycrystalline tin oxide, can now detect the gaseous products (generally various types of hydrocarbon) that are generated when materials become hot but before they actually burn.

Health and safety legislation creates a further requirement for gas sensors. Certain gases, such as carbon monoxide, hydrogen sulphide, chlorine and nitrous oxide, cause fatalities above a certain concentration and sensors must provide warning of impending danger. For other gases, health problems are caused by prolonged exposure and so the sensors in this case must integrate gas concentration over time to determine whether the allowable exposure limit over a given period of time has been exceeded. Again, solid-state sensors are now available to fulfil this function.

Concern about general environmental pollution is also making the development of gas sensors necessary in many new areas. Legislation is growing rapidly to control the emission of everything that is proven or suspected to cause health problems or environmental damage. The present list of controlled emissions includes nitrous oxide, oxides of sulphur, carbon monoxide and dioxide, CFCs, ammonia and hydrocarbons. Sensors are required both at the source of these pollutants, where concentrations are high, and also to monitor the much lower concentrations in the general environment. Oxygen concentration measurement is often of great importance also in pollution control, as the products of combustion processes are greatly affected by the air/fuel ratio.

Sensors associated with pollution monitoring and control often have to satisfy quite stringent specifications, particularly where the sensors are located at the pollutant source. Robustness is usually essential, as such sensors are subjected to bombardment from a variety of particulate matters, and they must also endure conditions of high humidity and temperature. They are also frequently located in inaccessible locations, such as in chimneys and flues, which means that they must have stable characteristics over long periods of time without calibration checks being necessary. The need for such high-specification sensors makes such pollutant-monitoring potentially very expensive if there are several problem gases involved. However, because the concentration of all output gases tends to vary to a similar extent according to the condition of filters etc., it is frequently only necessary to measure the concentration of one gas, from which the concentration of other gases can be predicted reliably. This greatly reduces the cost involved in such monitoring.

A number of devices that sense, measure the concentration of or analyse gases exist. In terms of frequency of usage, they vary from those that have been in use for a number of years, to those that have appeared recently, and finally to those that are still under research and development. In the following list of devices, their status in terms of current usage will be indicated. Fuller information can be found in Jones (1989).

21.9.1 Catalytic (calorimetric) sensors

Catalytic sensors, otherwise known as calorimetric sensors, have widespread use for measuring the concentration of flammable gases. Their principle of operation is to measure the heat evolved during the catalytic oxidation of reducing gases. They are cheap and robust but are unsuitable for measuring either very low or very high gas concentrations. The catalysts that have been commonly used in these devices in the past are adversely affected by many common industrial substances such as lead, phosphorus, silicon and sulphur, and this catalyst poisoning has previously prevented this type of device being used in many applications. However, new types of poison-resistant catalyst are now becoming available that are greatly extending the applicability of this type of device.

21.9.2 Paper tape sensors

By moving a paper tape impregnated with a reagent sensitive to a specific gas (e.g. lead acetate tape to detect hydrogen sulphide) through an air stream, the time history of the concentration of gas is indicated by the degree of colour change in the tape. This is used as a low accuracy but reliable and cheap means of detecting the presence of hydrogen sulphide and ammonia.

21.9.3 Liquid electrolyte electrochemical cells

These consist of two electrodes separated by electrolyte, to which the measured air supply is directed through a permeable membrane, as shown in Figure 21.15. The gas in the air to which the cell is sensitive reacts at the electrodes to form ions in the solution. This produces a voltage output from the cell.

Electrochemical cells have stable characteristics and give good measurement sensitivity. However, they are expensive and their durability is relatively poor, with life being generally limited to about one or two years at most. A further restriction is that they cannot be used above temperatures of about 50°C, as their performance deteriorates rapidly at high temperatures because of interference from other atmospheric substances.

The main use of such cells is in measuring toxic gases in satisfaction of health and safety legislation. Versions of the cell for this purpose are currently available to measure carbon monoxide, chlorine, nitrous oxide, hydrogen sulphide and ammonia. Cells to measure other gases are currently under development.

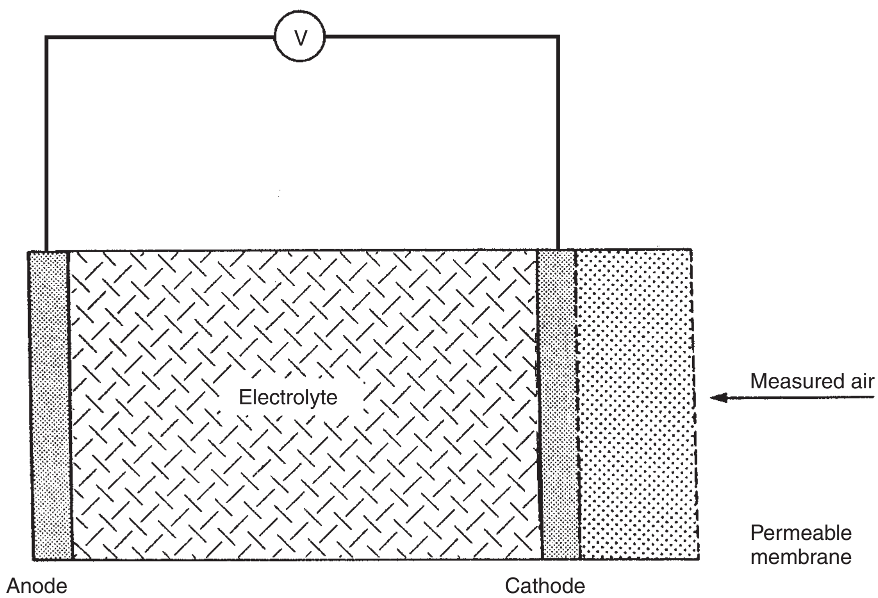


Fig. 21.15 Liquid electrolyte electrochemical cell.

In addition, electrochemical cells are also used to a limited extent to monitor carbon monoxide emissions in flue gases for environmental control purposes. Pre-cooling of the emitted gases is a necessary condition for this application.

21.9.4 Solid-state electrochemical cells (zirconia sensor)

At present, these cells are used only for measuring oxygen concentration, but ways of extending their use to other gases are currently in progress. The oxygen-measurement cell consists of two chambers separated by a zirconia wall. One chamber contains gas with a known oxygen concentration and the other contains the air being measured. Ions are conducted across the zirconia wall according to the difference in oxygen concentration across it and this produces an output e.m.f. The device is rugged but requires high temperatures to operate efficiently. It is, however, well proven and a standard choice for oxygen measurement. In industrial uses, it is often located in chimneystacks, where quite expensive mounting and protection systems are needed. However, very low cost versions (around £200) are now used in some vehicle exhaust systems as part of the engine management system.

21.9.5 Catalytic gate FETs

These consist of field effect transistors with a catalytic, palladium gate that is sensitive to hydrogen ions in the environment. The gate voltage, and hence characteristics of the device, change according to the hydrogen concentration. They can be made sensitive to gases such as hydrogen sulphide, ammonia and hydrocarbons as well as hydrogen. They are cheap and find application in workplace monitoring, in satisfaction of health and safety legislation, and in fire detection (mainly detecting hydrocarbon products).

21.9.6 Semiconductor (metal oxide) sensors

In these devices use is made of the fact that the surface conductivity of semiconductor metal oxides (generally tin or zinc oxides) changes according to the concentration of certain gases with which they are in contact. Unfortunately, they have a similar response for the range of gases to which they are sensitive. Hence, they show that a gas is present but not which one. Such sensors are cheap, robust, very durable and sensitive to very low gas concentrations. However, because their discrimination between gases is low and their accuracy in quantitative measurement is poor, they are mainly used only for qualitative indication of gas presence. In this role, they are particularly useful for fire prevention in detecting the presence of the combustion products that occur in low concentrations when the temperature starts to rise due to a fault.

21.9.7 Organic sensors

These work on similar principles to metal oxide semiconductors but use an organic surface layer that is designed to respond selectively to only one gas. At present, these devices are still the subjects of ongoing research, but industrial exploitation is anticipated in the near future. They promise to be cheap and have high stability and sensitivity.

21.9.8 Piezoelectric devices

In these devices, piezoelectric crystals are coated with an absorbent layer. As this layer absorbs gases, the crystal undergoes a change in resonant frequency that can be measured. There is no discrimination in this effect between different gases but the technique potentially offers a high sensitivity mechanism for detecting gas presence. At the present time, problems of finding a suitable type of coating material where absorption is reversible have not been generally solved, and the device only finds limited application at present for measuring moisture concentrations.

21.9.9 Infra-red absorption

This technique uses infra-red light at a particular wavelength that is directed across a chamber between a source and detector. The amount of light absorption is a function of the unknown gas concentration in the chamber. The instrument normally has a second chamber containing gas at a known concentration across which infra-red light at the same wavelength is directed to provide a reference. Sensitivity to carbon monoxide, carbon dioxide, ammonia or hydrocarbons can be provided according to the wavelength used. Microcomputers are now routinely incorporated in the instrument to reduce its sensitivity to gases other than the one being sensed and so improve measurement accuracy. The instrument finds widespread use in chimney/flue emission monitoring and in general process measurements.

21.9.10 Mass spectrometers

The mass spectrometer is a laboratory device for analysing gases. It first reduces a gas sample to a very low pressure. The sample is then ionized, accelerated and separated into its constituent components according to the respective charge-to-mass ratios. Almost any mixture of gases can be analysed and the individual components quantified, but the instrument is very expensive and requires a skilled user. Mass spectrometers have existed for over half a century but recent advances in electronic data processing techniques have greatly improved their performance.

21.9.11 Gas chromatography

This is also a laboratory instrument in which a gaseous sample is passed down a packed column. This separates the gas into its components, which are washed out of the column in turn and measured by a detector. Like the mass spectrometer, the instrument is versatile but expensive and it requires skilled use.

References and further reading

Dimension, angles and flatness:

Anthony, D.M. (1986) *Engineering Metrology*, Pergamon, Oxford, UK.

Hume, K.J. (1970) *Engineering Metrology*, McDonald, London.

Viscosity:

Miller, J.T. (ed.) (1975a) *The Instrument Manual*, United Trade Press, London, pp. 62–106.

Moisture and humidity:

Anderson, J.G. (1989) Paper moisture measurement using microwaves, *Measurement and Control*, **22**, pp. 82–84.

Benson, I.B. (1989) Industrial applications of near infrared reflectance for the measurement of moisture, *Measurement and Control*, **22**, pp. 45–49.

Gimson, C. (1989) Using the capacitance charge transfer principle for water content measurement, *Measurement Control*, **22**, pp. 79–81.

Miller, J.T. (ed.) (1975b) *The Instrument Manual*, United Trade Press, London, pp. 180–209.

Pragnell, R.F. (1989) The modern condensation dewpoint hygrometer, *Measurement and Control*, **22**, pp. 74–77.

Slight, H.A. (1989) Further thoughts on moisture measurement, *Measurement and Control*, **22**, pp. 85–86.

Thompson, F. (1989) Moisture measurement using microwaves, *Measurement and Control*, **22**, pp. 210–215.

Wiltshire, M.P. (1989) Ultrasonic moisture measurement, *Measurement and Control*, **22**, pp. 51–53.

Young, L. (1989) Moisture measurement using low resolution nuclear magnetic resonance, *Measurement and Control*, **22**, pp. 54–55.

Gas sensing and analysis:

Jones, T.A. (1989) Trends in the development of gas sensors, *Measurement and Control*, **22**(6), pp. 176–182.

Appendix 1 Imperial–metric–SI conversion tables

Length

SI units: mm, m, km
Imperial units: in, ft, mile

	mm	m	km	in	ft	mile
mm	1	10^{-3}	10^{-6}	0.039 3701	3.281×10^{-3}	–
m	1000	1	10^{-3}	39.3701	3.280 84	6.214×10^{-4}
km	10^6	10^3	1	39 370.1	3280.84	0.621 371
in	25.4	0.0254	–	1	0.083 333	–
ft	304.8	0.3048	3.048×10^{-4}	12	1	1.894×10^{-4}
mile	–	1609.34	1.609 34	63 360	5280	1

Area

SI units: mm^2, m^2, km^2
Imperial units: in^2, ft^2, $mile^2$

	mm^2	m^2	km^2	in^2	ft^2	$mile^2$
mm^2	1	10^{-6}	–	1.550×10^{-3}	1.076×10^{-5}	–
m^2	10^6	1	10^{-6}	1550	10.764	–
km^2	–	10^6	1	–	1076×10^7	0.3861
in^2	645.16	6.452×10^{-4}	–	1	6.944×10^{-3}	–
ft^2	92 903	0.092 90	–	144	1	–
$mile^2$	–	2.590×10^6	2.590	–	2.788×10^7	1

Second moment of area

SI units: mm^4, m^4
Imperial units: in^4, ft^4

	mm^4	m^4	in^4	ft^4
mm^4	1	10^{-12}	2.4025×10^{-6}	1.159×10^{-10}
m^4	10^{12}	1	2.4025×10^{6}	115.86
in^4	416 231	4.1623×10^{-7}	1	4.8225×10^{-5}
ft^4	8.631×10^{9}	8.631×10^{-3}	20 736	1

Volume

SI units: mm^3, m^3
Metric units: ml, l
Imperial units: in^3, ft^3, UK gallon

	mm^3	ml	l	m^3	in^3	ft^3	UK gallon
mm^3	1	10^{-3}	10^{-6}	10^{-9}	6.10×10^{-5}	–	–
ml	10^{3}	1	10^{-3}	10^{-6}	0.061 024	3.53×10^{-5}	2.2×10^{-4}
l	10^{6}	10^{3}	1	10^{-3}	61.024	0.035 32	0.22
m^3	10^{9}	10^{6}	10^{3}	1	61 024	35.31	220
in^3	16 387	16.39	0.0164	1.64×10^{-5}	1	5.79×10^{-4}	3.61×10^{-3}
ft^3	–	2.83×10^{4}	28.32	0.028 32	1728	1	6.229
UK gallon	–	4546	4.546	4.55×10^{-3}	277.4	0.1605	1

Note: Additional unit: 1 US gallon = 0.8327 UK gallon.

Density

SI unit: kg/m^3
Metric unit: g/cm^3
Imperial units: lb/ft^3, lb/in^3

	kg/m^3	g/cm^3	lb/ft^3	lb/in^3
kg/m^3	1	10^{-3}	0.062 428	3.605×10^{-5}
g/cm^3	1000	1	62.428	0.036 127
lb/ft^3	16.019	0.016 019	1	5.787×10^{-4}
lb/in^3	27 680	27.680	1728	1

Mass

SI units: g, kg, t
Imperial units: lb, cwt, ton

	g	kg	t	lb	cwt	ton
g	1	10^{-3}	10^{-6}	2.205×10^{-3}	1.968×10^{-5}	9.842×10^{-7}
kg	10^{3}	1	10^{-3}	2.204 62	0.019 684	9.842×10^{-4}
t	10^{6}	10^{3}	1	2204.62	19.6841	0.984 207
lb	453.592	0.453 59	4.536×10^{-4}	1	8.929×10^{-3}	4.464×10^{-4}
cwt	50 802.3	50.8023	0.050 802	112	1	0.05
ton	1.016×10^{6}	1016.05	1.016 05	2240	20	1

Force

SI units: N, kN
Metric unit: kg_f
Imperial units: pdl (poundal), lb_f, UK ton_f

	N	kg_f	kN	pdl	lb_f	UK ton_f
N	1	0.1020	10^{-3}	7.233	0.2248	1.004×10^{-4}
kg_f	9.807	1	9.807×10^{-3}	70.93	2.2046	9.842×10^{-4}
kN	1000	102.0	1	7233	224.8	0.1004
pdl	0.1383	0.0141	1.383×10^{-4}	1	0.0311	1.388×10^{-5}
lb_f	4.448	0.4536	4.448×10^{-3}	32.174	1	4.464×10^{-4}
UK ton_f	9964	1016	9.964	72 070	2240	1

Note: Additional unit: 1 dyne = 10^{-5}N = 7.233×10^{-5} pdl.

Torque (moment of force)

SI unit: N m
Metric unit: kg_f m
Imperial units: pdl ft, lb_f ft

	N m	kg_f m	pdl ft	lb_f ft
N m	1	0.1020	23.73	0.7376
kg_f m	9.807	1	232.7	7.233
pdl ft	0.042 14	4.297×10^{-3}	1	0.031 08
lb_f ft	1.356	0.1383	32.17	1

Inertia

SI unit: N m^2
Imperial unit: lb_f ft^2

$$1\,lb_f\,ft^2 = 0.4132\,N\,m^2$$

$$1\,N\,m^2 = 2.420\,lb_f\,ft^2$$

Pressure

SI units: mbar, bar, N/m^2 (pascal)
Imperial units: lb/in^2, in Hg, atm

	mbar	bar	N/m^2	lb/in^2	in Hg	atm
mbar	1	10^{-3}	100	0.014 50	0.029 53	9.869×10^{-4}
bar	1000	1	10^5	14.50	29.53	0.9869
N/m^2	0.01	10^{-5}	1	1.450×10^{-4}	2.953×10^{-4}	9.869×10^{-6}
lb/in^2	68.95	0.068 95	6895	1	2.036	0.068 05
in Hg	33.86	0.033 86	3386	0.4912	1	0.033 42
atm	1013	1.013	1.013×10^5	14.70	29.92	1

Additional conversion factors

1 inch water = 0.073 56 in Hg = 2.491 mbar
1 torr = 1.333 mbar
1 pascal = 1 N/m^2

Energy, work, heat

SI unit: J
Metric units: kg_f m, kW h
Imperial units: ft lb_f, cal, Btu

	J	kg$_f$ m	kW h	ft lb$_f$	cal	Btu
J	1	0.1020	2.778×10^{-7}	0.7376	0.2388	9.478×10^{-4}
kg$_f$ m	9.8066	1	2.724×10^{-6}	7.233	2.342	9.294×10^{-3}
kW h	3.600×10^6	367 098	1	2.655×10^6	859 845	3412.1
ft lb$_f$	1.3558	0.1383	3.766×10^{-7}	1	0.3238	1.285×10^{-3}
cal	4.1868	0.4270	1.163×10^{-6}	3.0880	1	3.968×10^{-3}
Btu	1055.1	107.59	2.931×10^{-4}	778.17	252.00	1

Additional conversion factors

1 therm = 10^5 Btu = 1.0551×10^8 J
1 thermie = 4.186×10^6 J
1 hp h = 0.7457 kW h = 2.6845×10^6 J
1 ft pdl = 0.042 14 J
1 erg = 10^{-7} J

Power

SI units W, kW
Imperial units: HP, ft lb_f/s

	W	kW	HP	ft lb_f/s
W	1	10^{-3}	1.341×10^{-3}	0.735 64
kW	10^3	1	1.341 02	735.64
HP	745.7	0.7457	1	548.57
ft lb_f/s	1.359 35	1.359×10^{-3}	1.823×10^{-3}	1

Velocity

SI units: mm/s, m/s
Metric unit: km/h
Imperial units: ft/s, mile/h

	mm/s	m/s	km/h	ft/s	mile/h
mm/s	1	10^{-3}	3.6×10^{-3}	3.281×10^{-3}	2.237×10^{-3}
m/s	1000	1	3.6	3.280 84	2.236 94
km/h	277.778	0.277 778	1	0.911 344	0.621 371
ft/s	304.8	0.3048	1.097 28	1	0.681 818
mile/h	447.04	0.447 04	1.609 344	1.466 67	1

Acceleration

SI unit: m/s^2
Other metric unit: cm/s^2
Imperial unit: ft/s^2
Other unit: g

	m/s^2	cm/s^2	ft/s^2	g
m/s^2	1	100	3.281	0.102
cm/s^2	0.01	1	0.0328	0.001 02
ft/s^2	0.3048	30.48	1	0.031 09
g	9.81	981	32.2	1

Mass flow rate

SI unit: g/s
Metric units: kg/h, tonne/d
Imperial units: lb/s, lb/h, ton/d

	g/s	kg/h	tonne/d	lb/s	lb/h	ton/d
g/s	1	3.6	0.086 40	2.205×10^{-3}	7.937	0.085 03
kg/h	0.2778	1	0.024 00	6.124×10^{-4}	2.205	0.023 62
tonne/d	11.57	41.67	1	0.025 51	91.86	0.9842
lb/s	453.6	1633	39.19	1	3600	38.57
lb/h	0.1260	0.4536	0.010 89	2.788×10^{-4}	1	0.010 71
ton/d	11.76	42.34	1.016	0.025 93	93.33	1

Volume flow rate

SI unit: m^3/s
Metric units: l/h, ml/s
Imperial units: gal/h, ft^3/s, ft^3/h

	l/h	ml/s	m^3/s	gal/h	ft^3/s	ft^3/h
l/h	1	0.2778	2.778×10^{-7}	0.2200	9.810×10^{-6}	0.035 316
ml/s	3.6	1	10^{-6}	0.7919	3.532×10^{-5}	0.127 14
m^3/s	3.6×10^6	10^6	1	7.919×10^5	35.31	1.271×10^5
gal/h	4.546	1.263	1.263×10^{-6}	1	4.460×10^{-5}	0.160 56
ft^3/s	1.019×10^5	2.832×10^4	0.028 32	2.242×10^4	1	3600
ft^3/h	28.316	7.8653	7.865×10^{-6}	6.2282	2.778×10^{-4}	1

Specific energy (heat per unit volume)

SI units: J/m^3, kJ/m^3, MJ/m^3
Imperial units: $kcal/m^3$, Btu/ft^3, therm/UK gal

	J/m^3	kJ/m^3	MJ/m^3	$kcal/m^3$	Btu/ft^3	therm/UK gal
J/m^3	1	10^{-3}	10^{-6}	1.388×10^{-4}	2.684×10^{-5}	–
kJ/m^3	1000	1	10^{-3}	0.2388	0.026 84	–
MJ/m^3	10^6	1000	1	238.8	26.84	4.309×10^{-5}
$kcal/m^3$	4187	4.187	4.187×10^{-3}	1	0.1124	1.804×10^{-7}
Btu/ft^3	3.726×10^4	37.26	0.037 26	8.899	1	1.605×10^{-6}
therm/gal	–	–	2.321×10^4	5.543×10^6	6.229×10^5	1

Dynamic viscosity

SI unit: N s/m^2
Metric unit: cP (centipoise), P (poise) [1 P = 100 g/m s]
Imperial unit: lb_m/ft h

	lb_m/ft h	P	cP	N s/m^2
lb_m/ft h	1	4.133×10^{-3}	0.4134	4.134×10^{-4}
P	241.9	1	100	0.1
cP	2.419	0.01	1	10^{-3}
N s/m^2	2419	10	1000	1

Note: Additional unit: 1 pascal second = 1 N s/m^2.

Kinematic viscosity

SI unit: m^2/s
Metric unit: cSt (centistokes), St (stokes)
Imperial unit: ft^2/s

	ft^2/s	m^2/s	cSt	St
ft^2/s	1	0.0929	9.29×10^4	929
m^2/s	10.764	1	10^6	10^4
cSt	1.0764×10^{-5}	10^{-6}	1	0.01
St	1.0764×10^{-3}	10^{-4}	100	1

Appendix 2 Thévenin's theorem

Thévenin's theorem is extremely useful in the analysis of complex electrical circuits. It states that any network which has two accessible terminals A and B can be replaced, as far as its external behaviour is concerned, by a single e.m.f. acting in series with a single resistance between A and B. The single equivalent e.m.f. is that e.m.f. which is measured across A and B when the circuit external to the network is disconnected. The single equivalent resistance is the resistance of the network when all current and voltage sources within it are reduced to zero. To calculate this internal resistance of the network, all current sources within it are treated as open circuits and all voltage sources as short circuits. The proof of Thévenin's theorem can be found in Skilling (1967).

Figure A2.1 shows part of a network consisting of a voltage source and four resistances. As far as its behaviour external to the terminals A and B is concerned, this can be regarded as a single voltage source V_t and a single resistance R_t. Applying Thévenin's theorem, R_t is found first of all by treating V_1 as a short circuit, as shown in Figure A2.2. This is simply two resistances, R_1 and $(R_2 + R_4 + R_5)$ in parallel. The

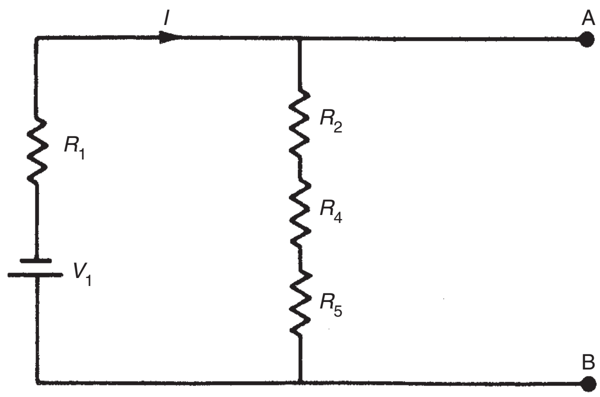


Fig. A2.1

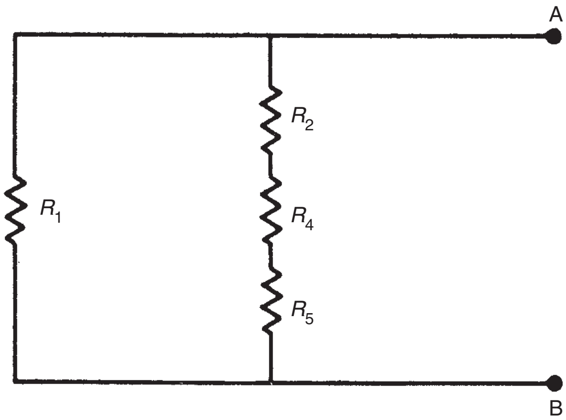


Fig. A2.2

equivalent resistance R_t is thus given by:

$$R_t = \frac{R_1(R_2 + R_4 + R_5)}{R_1 + R_2 + R_4 + R_5}$$

V_t is the voltage drop across AB. To calculate this, it is necessary to carry out an intermediate step of working out the current flowing, I. Referring to Figure A2.1, this is given by:

$$I = \frac{V_1}{R_1 + R_2 + R_4 + R_5}$$

Now, V_t can be calculated from:

$$\begin{aligned} V_t &= I(R_2 + R_4 + R_5) \\ &= \frac{V_1(R_2 + R_4 + R_5)}{R_1 + R_2 + R_4 + R_5} \end{aligned}$$

The network of Figure A2.1 has thus been reduced to the simpler equivalent network shown in Figure A2.3.

Let us now proceed to the typical network problem of calculating the current flowing in the resistor R_3 of Figure A2.4. R_3 can be regarded as an external circuit or load on

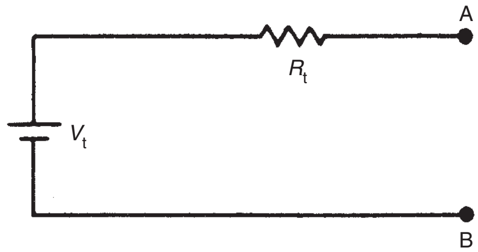


Fig. A2.3

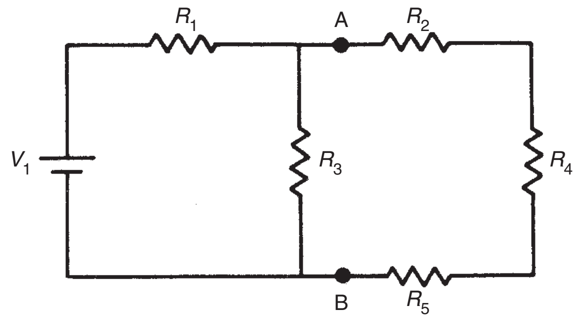


Fig. A2.4

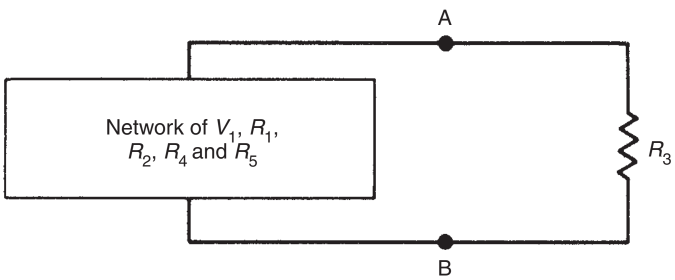


Fig. A2.5

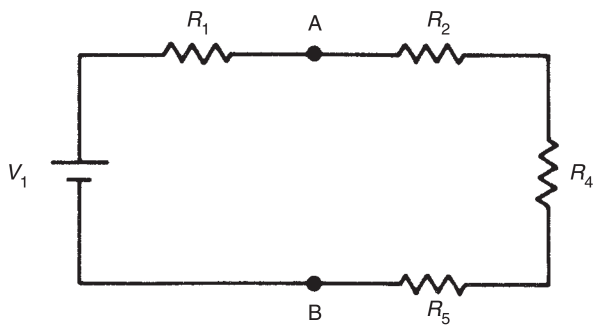


Fig. A2.6

the rest of the network consisting of V_1, R_1, R_2, R_4 and R_5, as shown in Figure A2.5. This network of V_1, R_1, R_2, R_4 and R_5 is that shown in Figure A2.6. This can be rearranged to the network shown in Figure A2.1, which is equivalent to the single voltage source and resistance, V_t and R_t, calculated above. The whole circuit is then equivalent to that shown in Figure A2.7, and the current flowing through R_3 can be written as:

$$I_{AB} = \frac{V_t}{R_t + R_3}$$

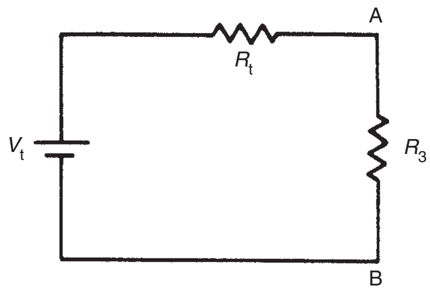


Fig. A2.7

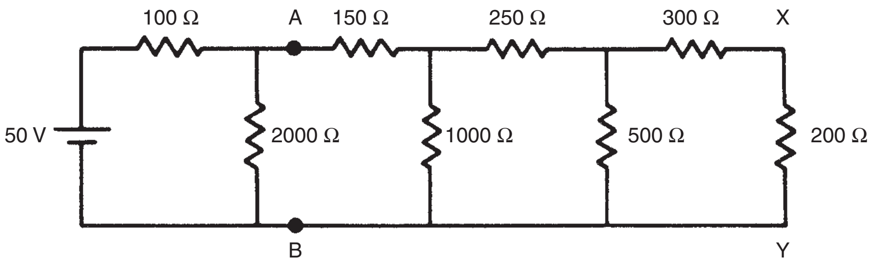


Fig. A2.8

Thévenin's theorem can be applied successively to solve ladder networks of the form shown in Figure A2.8. Suppose in this network that it is required to calculate the current flowing in branch XY.

The first step is to imagine two terminals in the circuit A and B and regard the network to the right of AB as a load on the circuit to the left of AB. The circuit to the left of AB can be reduced to a single equivalent voltage source, E_{AB}, and resistance, R_{AB}, by Thévenin's theorem. If the 50 V source is replaced by its zero internal resistance (i.e. by a short circuit), then R_{AB} is given by:

$$\frac{1}{R_{AB}} = \frac{1}{100} + \frac{1}{2000} = \frac{2000 + 100}{200\,000}$$

Hence:

$$R_{AB} = 95.24\,\Omega$$

When AB is open circuit, the current flowing round the loop to the left of AB is given by:

$$I = \frac{50}{100 + 2000}$$

Hence, E_{AB}, the open-circuit voltage across AB, is given by:

$$E_{AB} = I \times 2000 = 47.62\,\text{V}$$

We can now replace the circuit shown in Figure A2.8 by the simpler equivalent circuit shown in Figure A2.9.

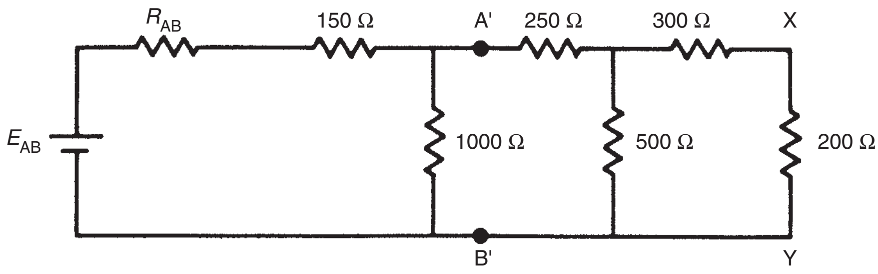


Fig. A2.9

The next stage is to apply an identical procedure to find an equivalent circuit consisting of voltage source $E_{A'B'}$ and resistance $R_{A'B'}$ for the network to the left of points A′ and B′ in Figure A2.9:

$$\frac{1}{R_{A'B'}} = \frac{1}{R_{AB} + 150} + \frac{1}{1000} = \frac{1}{245.24} + \frac{1}{1000} = \frac{1245.24}{245\,240}$$

Hence:

$$R_{A'B'} = 196.94\,\Omega$$

$$E_{A'B'} = \frac{1000}{R_{AB} + 150 + 1000} E_{AB} = 38.24\,\text{V}$$

The circuit can now be represented in the yet simpler form shown in Figure A2.10. Proceeding as before to find an equivalent voltage source and resistance, $E_{A''B''}$ and $R_{A''B''}$, for the circuit to the left of A″B″ in Figure A2.10:

$$\frac{1}{R_{A''B''}} = \frac{1}{R_{A'B'} + 250} + \frac{1}{500} = \frac{500 + 446.94}{223\,470}$$

Hence:

$$R_{A''B''} = 235.99\,\Omega$$

$$E_{A''B''} = \frac{500}{R_{A'B'} + 250 + 500} E_{A'B'} = 20.19\,\text{V}$$

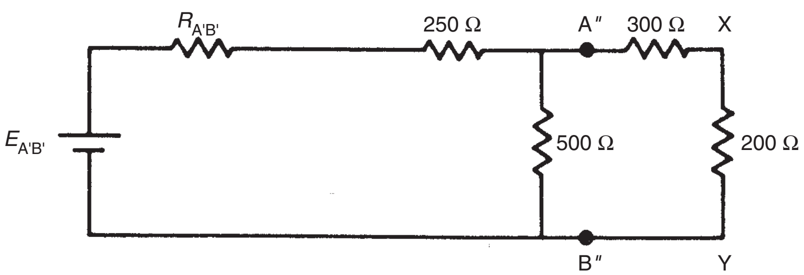


Fig. A2.10

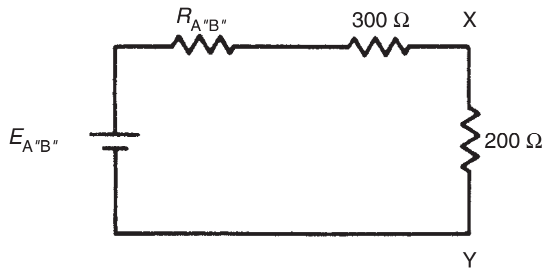


Fig. A2.11

The circuit has now been reduced to the form shown in Figure A2.11, where the current through branch XY can be calculated simply as:

$$I_{XY} = \frac{E_{A''B''}}{R_{A''B''} + 300 + 200} = \frac{20.19}{735.99} = 27.43\,\text{mA}$$

References and further reading

Skilling, H.H. (1967) *Electrical Engineering Circuits*, Wiley: New York.

Appendix 3
Thermocouple tables

Type E: chromel–constantan
Type J: iron–constantan
Type K: chromel–alumel
Type N: nicrosil–nisil
Type S: platinum/10% rhodium–platinum
Type T: copper–constantan

Temp. (°C)	*Type E*	*Type J*	*Type K*	*Type N*	*Type S*	*Type T*
−270	−9.834		−6.458	−4.345		
−260	−9.795		−6.441	−4.336		
−250	−9.719		−6.404	−4.313		
−240	−9.604		−6.344	−4.277		−6.105
−230	−9.456		−6.262	−4.227		−6.003
−220	−9.274		−6.158	−4.162		−5.891
−210	−9.063	−8.096	−6.035	−4.083		−5.753
−200	−8.824	−7.890	−5.891	−3.990		−5.603
−190	−8.561	−7.659	−5.730	−3.884		−5.438
−180	−8.273	−7.402	−5.550	−3.766		−5.261
−170	−7.963	−7.122	−5.354	−3.634		−5.070
−160	−7.631	−6.821	−5.141	−3.491		−4.865
−150	−7.279	−6.499	−4.912	−3.336		−4.648
−140	−6.907	−6.159	−4.669	−3.170		−4.419
−130	−6.516	−5.801	−4.410	−2.994		−4.177
−120	−6.107	−5.426	−4.138	−2.807		−3.923
−110	−5.680	−5.036	−3.852	−2.612		−3.656
−100	−5.237	−4.632	−3.553	−2.407		−3.378
−90	−4.777	−4.215	−3.242	−2.193		−3.089
−80	−4.301	−3.785	−2.920	−1.972		−2.788
−70	−3.811	−3.344	−2.586	−1.744		−2.475
−60	−3.306	−2.892	−2.243	−1.509		−2.152
−50	−2.787	−2.431	−1.889	−1.268	−0.236	−1.819

Temp. (°C)	*Type E*	*Type J*	*Type K*	*Type N*	*Type S*	*Type T*
−40	−2.254	−1.960	−1.527	−1.023	−0.194	−1.475
−30	−1.709	−1.481	−1.156	−0.772	−0.150	−1.121
−20	−1.151	−0.995	−0.777	−0.518	−0.103	−0.757
−10	−0.581	−0.501	−0.392	−0.260	−0.053	−0.383
0	0.000	0.000	0.000	0.000	0.000	0.000
10	0.591	0.507	0.397	0.261	0.055	0.391
20	1.192	1.019	0.798	0.525	0.113	0.789
30	1.801	1.536	1.203	0.793	0.173	1.196
40	2.419	2.058	1.611	1.064	0.235	1.611
50	3.047	2.585	2.022	1.339	0.299	2.035
60	3.683	3.115	2.436	1.619	0.365	2.467
70	4.329	3.649	2.850	1.902	0.432	2.908
80	4.983	4.186	3.266	2.188	0.502	3.357
90	5.646	4.725	3.681	2.479	0.573	3.813
100	6.317	5.268	4.095	2.774	0.645	4.277
110	6.996	5.812	4.508	3.072	0.719	4.749
120	7.683	6.359	4.919	3.374	0.795	5.227
130	8.377	6.907	5.327	3.679	0.872	5.712
140	9.078	7.457	5.733	3.988	0.950	6.204
150	9.787	8.008	6.137	4.301	1.029	6.702
160	10.501	8.560	6.539	4.617	1.109	7.207
170	11.222	9.113	6.939	4.936	1.190	7.718
180	11.949	9.667	7.338	5.258	1.273	8.235
190	12.681	10.222	7.737	5.584	1.356	8.757
200	13.419	10.777	8.137	5.912	1.440	9.286
210	14.161	11.332	8.537	6.243	1.525	9.820
220	14.909	11.887	8.938	6.577	1.611	10.360
230	15.661	12.442	9.341	6.914	1.698	10.905
240	16.417	12.998	9.745	7.254	1.785	11.456
250	17.178	13.553	10.151	7.596	1.873	12.011
260	17.942	14.108	10.560	7.940	1.962	12.572
270	18.710	14.663	10.969	8.287	2.051	13.137
280	19.481	15.217	11.381	8.636	2.141	13.707
290	20.256	15.771	11.793	8.987	2.232	14.281
300	21.033	16.325	12.207	9.340	2.323	14.860
310	21.814	16.879	12.623	9.695	2.414	15.443
320	22.597	17.432	13.039	10.053	2.506	16.030
330	23.383	17.984	13.456	10.412	2.599	16.621
340	24.171	18.537	13.874	10.772	2.692	17.217
350	24.961	19.089	14.292	11.135	2.786	17.816
360	25.754	19.640	14.712	11.499	2.880	18.420
370	26.549	20.192	15.132	11.865	2.974	19.027
380	27.345	20.743	15.552	12.233	3.069	19.638

Temp. (°C)	*Type E*	*Type J*	*Type K*	*Type N*	*Type S*	*Type T*
390	28.143	21.295	15.974	12.602	3.164	20.252
400	28.943	21.846	16.395	12.972	3.260	20.869
410	29.744	22.397	16.818	13.344	3.356	
420	30.546	22.949	17.241	13.717	3.452	
430	31.350	23.501	17.664	14.091	3.549	
440	32.155	24.054	18.088	14.467	3.645	
450	32.960	24.607	18.513	14.844	3.743	
460	33.767	25.161	18.938	15.222	3.840	
470	34.574	25.716	19.363	15.601	3.938	
480	35.382	26.272	19.788	15.981	4.036	
490	36.190	26.829	20.214	16.362	4.135	
500	36.999	27.388	20.640	16.744	4.234	
510	37.808	27.949	21.066	17.127	4.333	
520	38.617	28.511	21.493	17.511	4.432	
530	39.426	29.075	21.919	17.896	4.532	
540	40.236	29.642	22.346	18.282	4.632	
550	41.045	30.210	22.772	18.668	4.732	
560	41.853	30.782	23.198	19.055	4.832	
570	42.662	31.356	23.624	19.443	4.933	
580	43.470	31.933	24.050	19.831	5.034	
590	44.278	32.513	24.476	20.220	5.136	
600	45.085	33.096	24.902	20.609	5.237	
610	45.891	33.683	25.327	20.999	5.339	
620	46.697	34.273	25.751	21.390	5.442	
630	47.502	34.867	26.176	21.781	5.544	
640	48.306	35.464	26.599	22.172	5.648	
650	49.109	36.066	27.022	22.564	5.751	
660	49.911	36.671	27.445	22.956	5.855	
670	50.713	37.280	27.867	23.348	5.960	
680	51.513	37.893	28.288	23.740	6.064	
690	52.312	38.510	28.709	24.133	6.169	
700	53.110	39.130	29.128	24.526	6.274	
710	53.907	39.754	29.547	24.919	6.380	
720	54.703	40.382	29.965	25.312	6.486	
730	55.498	41.013	30.383	25.705	6.592	
740	56.291	41.647	30.799	26.098	6.699	
750	57.083	42.283	31.214	26.491	6.805	
760	57.873	42.922	31.629	26.885	6.913	
770	58.663	43.563	32.042	27.278	7.020	
780	59.451	44.207	32.455	27.671	7.128	
790	60.237	44.852	32.866	28.063	7.236	
800	61.022	45.498	33.277	28.456	7.345	
810	61.806	46.144	33.686	28.849	7.454	

Temp. (°C)	*Type E*	*Type J*	*Type K*	*Type N*	*Type S*	*Type T*
820	62.588	46.790	34.095	29.241	7.563	
830	63.368	47.434	34.502	29.633	7.672	
840	64.147	48.076	34.908	30.025	7.782	
850	64.924	48.717	35.314	30.417	7.892	
860	65.700	49.354	35.718	30.808	8.003	
870	66.473	49.989	36.121	31.199	8.114	
880	67.245	50.621	36.524	31.590	8.225	
890	68.015	51.249	36.925	31.980	8.336	
900	68.783	51.875	37.325	32.370	8.448	
910	69.549	52.496	37.724	32.760	8.560	
920	70.313	53.115	38.122	33.149	8.673	
930	71.075	53.729	38.519	33.538	8.786	
940	71.835	54.341	38.915	33.926	8.899	
950	72.593	54.949	39.310	34.315	9.012	
960	73.350	55.553	39.703	34.702	9.126	
970	74.104	56.154	40.096	35.089	9.240	
980	74.857	56.753	40.488	35.476	9.355	
990	75.608	57.349	40.879	35.862	9.470	
1000	76.357	57.942	41.269	36.248	9.585	
1010		58.533	41.657	36.633	9.700	
1020		59.121	42.045	37.018	9.816	
1030		59.708	42.432	37.402	9.932	
1040		60.293	42.817	37.786	10.048	
1050		60.877	43.202	38.169	10.165	
1060		61.458	43.585	38.552	10.282	
1070		62.040	43.968	38.934	10.400	
1080		62.619	44.349	39.315	10.517	
1090		63.199	44.729	39.696	10.635	
1100		63.777	45.108	40.076	10.754	
1110		64.355	45.486	40.456	10.872	
1120		64.933	45.863	40.835	10.991	
1130		65.510	46.238	41.213	11.110	
1140		66.087	46.612	41.590	11.229	
1150		66.664	46.985	41.966	11.348	
1160		67.240	47.356	42.342	11.467	
1170		67.815	47.726	42.717	11.587	
1180		68.389	48.095	43.091	11.707	
1190		68.963	48.462	43.464	11.827	
1200		69.536	48.828	43.836	11.947	
1210			49.192	44.207	12.067	
1220			49.555	44.577	12.188	
1230			49.916	44.947	12.308	
1240			50.276	45.315	12.429	

Temp. (°C)	*Type E*	*Type J*	*Type K*	*Type N*	*Type S*	*Type T*
1250			50.633	45.682	12.550	
1260			50.990	46.048	12.671	
1270			51.344	46.413	12.792	
1280			51.697	46.777	12.913	
1290			52.049	47.140	13.034	
1300			52.398	47.502	13.155	
1310			52.747		13.276	
1320			53.093		13.397	
1330			53.438		13.519	
1340			53.782		13.640	
1350			54.125		13.761	
1360			54.467		13.883	
1370			54.807		14.004	
1380					14.125	
1390					14.247	
1400					14.368	
1410					14.489	
1420					14.610	
1430					14.731	
1440					14.852	
1450					14.973	
1460					15.094	
1470					15.215	
1480					15.336	
1490					15.456	
1500					15.576	
1510					15.697	
1520					15.817	
1530					15.937	
1540					16.057	
1550					16.176	
1560					16.296	
1570					16.415	
1580					16.534	
1590					16.653	
1600					16.771	
1610					16.890	
1620					17.008	
1630					17.125	
1640					17.243	
1650					17.360	
1660					17.477	
1670					17.594	

Temp. (°C)	*Type E*	*Type J*	*Type K*	*Type N*	*Type S*	*Type T*
1680					17.711	
1690					17.826	
1700					17.942	
1710					18.056	
1720					18.170	
1730					18.282	
1740					18.394	
1750					18.504	
1760					18.612	

Appendix 4 Solutions to self-test questions

Chapter 2

Q5. 0.0175 mV/°C

Q7. (a) 2.62 ; (b) 2.94 ; 0.32

Q8. (a) 20 μm/kg; 22 μm/kg; (b) 200 μm; 2 μm/kg; (c) 14.3 μm/°C; 0.143 μm $(°C)^{-1}$ $(kg)^{-1}$

Q9. (a)

Time	*Depth*	*Temp. reading*	*Temp. error*
0	0	20.0	0.0
100	50	19.716	0.216
200	100	19.245	0.245
300	150	18.749	0.249
400	200	18.250	0.250
500	250	17.750	0.250

(b) 10.25°C

Chapter 3

Q3. 3.9%

Q5. 5.0%; 24 750 Ω

Q6. 10.0%

Q9. mean 31.1; median 30.5; standard deviation 3.0

Q10. mean 1.537; standard deviation 0.021; accuracy of mean value = ±0.007 i.e. mean value = 1.537 ± 0.007; for 1000 measurements, accuracy would be improved by a factor of 10

Q11. Mean value = 21.8 mA

Measurement	21.5	22.1	21.3	21.7	22.0	22.2	21.8	21.4	21.9	22.1
Deviation from mean	−0.3	+0.3	−0.5	−0.1	+0.2	+0.4	0.0	−0.4	+0.1	+0.3
(deviations)^2	0.09	0.09	0.25	0.01	0.04	0.16	0.0	0.16	0.01	0.09

Standard deviation = 0.32

Q12. 86.6%

Q13. 97.7%

Q14. ±0.7%

Q15. ±4.7%

Q16. ±2.2%

Q17. 46.7ohm ± 2.5%

Q18. 2.5%

Q19. (a) 0.31 m^3/ min; (b) ±4.1%

Chapter 7

Q1. 81.9 mV

Q2. 378 mV

Q3. (a) 0.82 mV/°C; (b) indicated temperature 101.9°C; error 1.9°C

Q4. 24 V; 1.2 W

Q6. 85.9 mV

Q7. (a) 69.6 Ω, 930.4 Ω; (b) 110.3 Ω

Q8. (a) $R_u = R_2R_3/R_1$; $L_u = R_2R_3C$; (b) 1.57 Ω; 100 mH; (c) 20

Q9. 2.538 V r.m.s.

Q10. 50 µF

Q11. (a) At balance $\dfrac{R_1 + j\omega L}{R_3} = \dfrac{R_2}{R_4 - j/\omega C}$. Then, by taking real and imaginary parts and manipulating $L = \dfrac{R_1}{\omega^2 R_4 C}$ and $R_1 = \dfrac{R_2R_3}{R_4(1 + 1/\omega^2 R_4^2 C^2)}$. Hence, at balance $L = \dfrac{R_2R_3C}{1 + \omega^2 R_4^2 C^2}$

(b) $Q = \omega L/R_1 = 1/\omega R_4 C$ using the equations developed in part (a) above. For large Q, $\omega^2 R_4^2 C^2 \ll 1$, and the equation for L above becomes $L = R_2R_3C$. This is independent of frequency because there is no ω term in the expression

(c) 20 mH

Chapter 9

Q2. *One's compl.* *Two's compl.*

(a)	01010000	01010001
(b)	10001000	10001001
(c)	10011010	10011011
(d)	00101001	00101010
(e)	00010011	00010100

Q3. (a) 111001 71 39
(b) 1100101 145 65
(c) 10101111 257 AF
(d) 100000011 403 103
(e) 1111100111 1747 3E7
(f) 10011010010 2322 4D2

Q4. (a) 7515; (b) F4D

Q5. (a) 130645; (b) B1A5

Q6. (a) 1214; (b) 28C; (c) 3352

Chapter 11

Q1. (b) $\omega_n = \sqrt{K_s/J}$; $\beta = \sqrt{K_I/2RJK_S}$; sensitivity $= K_I/K_S R$
(d) 0.7; (e) typical bandwidth 100 Hz; maximum frequency 30 Hz

Q3. $a = 12.410$; $b = 40.438$

Q4. 9.8 Ω

Q5. $a = 1.12$; $b = 2.00$

Q6. (a) $C = 5.77 \times 10^{-7}$; $T_0 = 11027$; (b) $T = 428°$K

Chapter 12

Q1. 30.4 days

Q2. 40.6 days

Q3. 222 days

Q4. 0.988

Q5. 0.61 or 61%

Q6. 0.95 or 95%

Q7. 0.86

Q8. 0.9975

Q9. 0.92

Q10. 24

Q11. 39
Q12. (a) 95.2%; (b) 49 seeded errors
Q13. 97.2%
Q14. 94.2%

Chapter 14

Q1. 300°C
Q2. 147.1°C
Q3. 700°C
Q4. 610°C; 678.4°C
Q5. 15.55 mV; 228.5°C

Index