Chapter 1

Introduction

1.1 Measurements

Measurements are made or measurement systems are set up for one or more of the following functions:

1. To *monitor* processes and operations
2. To *control* processes and operations
3. To carry-out some *analysis*.

The functions are elaborated below.

Monitoring

Thermometers, barometers, anemometers, water, gas, electricity meters only indicate certain quantities. Their readings do not perform any control functions in the ordinary sense. These measurements are made for monitoring purposes only.

Control

The thermostat in a refrigerator or geyser determines the temperature of the relevant environment and accordingly switches OFF or ON the cooling or heating mechanism to keep the temperature constant, i.e. to control the temperature. A single system sometimes may require many controls. For example, an aircraft needs controls from altimeters, gyroscopes, angle-of-attack sensors, thermocouples, accelerometer, etc.

Controlling a variable is rather an involved process and as such it is a subject of study by itself.

Analysis

Measurements are also made to

1. test the validity of predictions from theories,
2. build empirical models, i.e. relationships between parameters and quantities associated with a problem, and
3. characterise materials, devices and components.

In general, these requirements may be called *analysis*.

1.2 Instruments

Measurements are made with the help of instruments. Instruments, in general, consist of a few elements. But before we go into the contents of a generalised instrument, let us define what we mean by an instrument.

An instrument can be defined as a device or a system which is designed in such a way that it maintains a functional relationship between a prescribed property of a substance and a physical variable, and communicates this relationship to a human observer by some ways and means. For example, a mercury-in-glass thermometer is an instrument, because it maintains a linear relationship between thermal expansion of mercury (prescribed property) and temperature (physical variable) and communicates this relationship to us through a graduated scale.

A generalised instrument can be schematically represented as shown in Fig. 1.1. It consists of

1 A transducer
2 A signal conditioner and transmitter, and
3 A display/recording device.

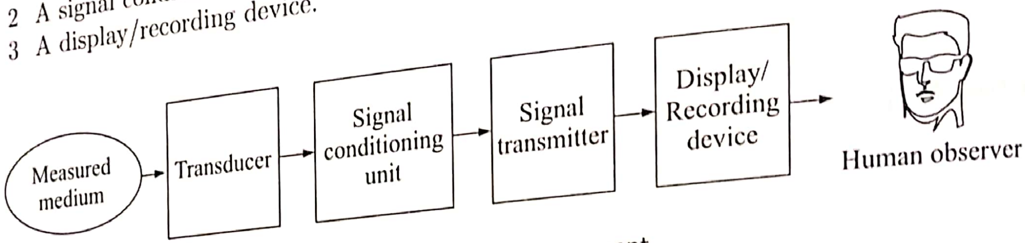

Fig. 1.1 A generalised instrument.

Transducer

A transducer senses the physical variable to be measured (i.e. measurand) and converts it to a suitable signal, preferably an electrical one.

One point has to be noted in this context. All transducers extract some energy from the measured medium which implies that the measurand is always disturbed by the measurement system. Therefore, *a perfect measurement is theoretically impossible.*

We will consider transducers in general in Chapter 5 and a few representative transducers for measurement of a few non-electrical quantities in Chapters 6 to 11.

Signal Conditioner and Transmitter

The signal generated by the transducer may need to be amplified, attenuated, integrated, differentiated, modulated, converted to a digital signal, and so on. The signal conditioner performs one or more such tasks. Since electrical signals have distinct advantages in this respect, more so with the development of electronics, a signal conditioner is now basically an electronic gadget. We will, however, discuss basics of signal conditioning in Chapter 16.

Signal transmitters are necessary for remote measurements. Remote measurements and control, called telemetry, is a highly-developed subject. We will exclude this topic from our consideration.

Display/Recording Device

The purpose of this element of an instrument is obvious—to communicate the information about the measurand to the human observer or to present it in an intelligible form. This aspect of instrumentation is discussed in Chapter 17.

We will study the subject element-wise. But before doing that we need to study the static (Chapter 2) and dynamic (Chapter 4) characteristics of instruments, and understand how to estimate errors (Chapter 3) because all these matters determine the performance of an instrumentation system.

CHAPTER 2

Static Characteristics of Instruments

By static characteristics we mean attributes associated with static measurements or measurement of quantities which are constant or vary very slowly with time. For example, the measurement of emf (electromotive force) of a cell or the resistance of a resistor at constant temperature are both static measurements.

Static characteristics of instruments can broadly be divided into two categories—desirable and undesirable—each consisting of a few characteristics as shown in Fig. 2.1.

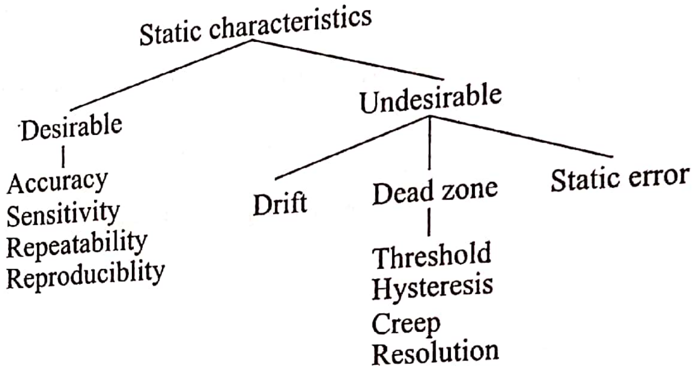

Fig. 2.1 Static characteristics tree.

2.1 Desirable Characteristics

The nature of these characteristics is discussed below.

Accuracy

Accuracy determines the closeness of an *instrument reading* to the true value of the measurand. Suppose, a known voltage of 200 V is being measured by a voltmeter and the successive readings are 204, 205, 203, 203 and 205 volts. So, the accuracy is about 2.5%. Here, though the repeatability of readings is not too bad, the accuracy is low because the instrument may be having a large calibration error. Hence, the accuracy can be improved upon by better calibration of the instrument.

Precision

Precision is another term which is often used in the same connotation as the accuracy. But in reality precision is different from accuracy. In the above example, the reading can be expressed as 204 ± 1 V, which means that the precision is a little less than 0.5% in this case.

Precision is, therefore, related to the repeatability[1] of the instrument reading and is a characteristic of the instrument itself. To improve the precision of an instrument, its design and construction have to be improved upon.

Symbolically, therefore, if a denotes accuracy, p the precision and c the calibration error then $a = p + c$.

Precision of a measurement also depends on what is called the number of significant figures. An example will perhaps make the point clear. Suppose the resistance of a conductor is being measured by an analogue ohmmeter. The ohmmeter indicates the true value, but the observer is unable to read the exact value because of lack of graduation beyond a certain number of decimals. Thus, though the instrument is showing the correct value, the precision of the measurement depends upon the number of significant figures to which the observer can read the value. And in an involved measurement where many measurands have to be combined, the number of significant figures plays a crucial role to determine the precision of the ultimate measurement. We discuss below this aspect in somewhat greater detail.

Significant figures

Significant figures convey information regarding the magnitude of precision of a quantity. For example, if a measurement reports that the line voltage is 220 V, it means that the line voltage is closer to 220 V than it is to 219 V or 221 V. Alternatively, if the reported value is 220.0 V, it means that the value is closer to 220.0 V than it is to 219.9 V or 220.1 V. Talking in terms of significant figures, it is 3 in the former case and 4 in the latter case. Significant figures play an important role in figuring out the final value in an involved measurement. Suppose, four resistors of values 28.4, 4.25, 56.605 and 0.76 ohms are connected in series.

$$\begin{array}{r} 28.\mathit{4} \\ 4.2\mathit{5} \\ 56.60\mathit{5} \\ 0.7\mathit{6} \\ \hline 90.\mathit{015} \end{array}$$

What should be the value for the total resistance? The general tendency is to report the result obtained by a straightforward addition, i.e. 90.015 ohms. But a close look reveals that this result conveys a wrong information regarding the precision of the measurement. If we signify doubtful figures by italics, it will be evident from the calculation shown above that the value when reported as 90.0 ohms would convey the right information regarding its precision. The simple rules arriving at such figures in mathematical manipulation of data are now enumerated.

Addition and subtraction. After performing the operation write the result rounded to the same number of *decimal places* as the least accurate figure.

[1] See Section 2.1 at page 9.

Multiplication and division. After the operation round the result to the same number of *significant figures* as the least accurate number.

Example 2.1
Four capacitors of values 45.1, 3.22, 89.309 and 0.48 μF, are connected in parallel. Find the value of the equivalent capacitor to the appropriate number of significant figures.

Solution
The straightforward addition yields 138.109 μF. Rounding it off to the same number of decimal places as the least accurate figure, namely 45.1 μF, the acceptable value is 138.1 μF.

Example 2.2
A current of 3.12 A is flowing through a resistor of 53.635 Ω. Find the value of the voltage drop across the resistor to the appropriate number of significant figures.

Solution
The straightforward multiplication yields 167.3412 V. Rounding it off to three significant figures—the same as that of 3.12—the value to be reported is 167 V.

Sensitivity

Sensitivity is defined as the absolute ratio of the increment of the output signal (or response) to that of the input signal (or measurand). Stated mathematically, $S = \Delta q_o/\Delta q_i$ where q_i and q_o are input and output quantities respectively.

Suppose in a mercury-in-glass thermometer the meniscus moves by 1 cm when the temperature changes by 10°C. Its sensitivity is, therefore, 1 mm/°C.

The sensitivity of a voltmeter, however, is expressed in ohm/volt. A voltmeter is considered to be more sensitive if it draws less current from the circuit which, in turn, is ensured by the high resistance of the voltmeter that has to be connected in parallel with the circuit. For this reason, the sensitivity of a voltmeter varies inversely with the current required for full-scale deflection (FSD). Thus,

$$\text{Sensitivity} = \frac{1\ (\text{V})}{I_{FSD}\ (\text{A})}\ \text{ohm/volt}$$

where, I_{FSD} is the current required for FSD of the meter movement.

Example 2.3
What is the sensitivity of a voltmeter having 50 μA FSD?

Solution
The required sensitivity is given by

$$\text{Sensitivity} = \frac{1}{50 \times 10^{-6}} = 20{,}000\ \text{ohm/volt}$$

Lab quality voltmeters should have a minimum sensitivity of 20 kΩ/volt.

Linearity and nonlinearity

If the functional relationship between the input quantity and the output reading of an instrument is linear, we call it a linear instrument. For example, a mercury-in-glass thermometer is a linear instrument while a simple thermocouple arrangement for measuring temperature is nonlinear.

The sensitivity of a linear instrument is constant while that of a nonlinear one varies from range to range as will be evident from Fig. 2.2.

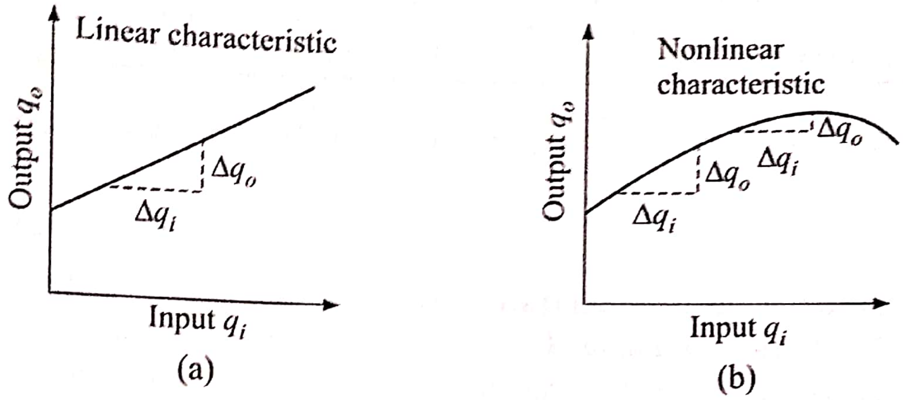

Fig. 2.2 Sensitivity: (a) linear instrument where the sensitivity is constant over the entire range, (b) nonlinear instrument where sensitivity varies from one range to another.

A perfectly linear instrument is rather difficult to realise, because almost all the so-called linear instruments show some deviation from linearity. This deviation may assume one of the following three forms as illustrated in Fig. 2.3.

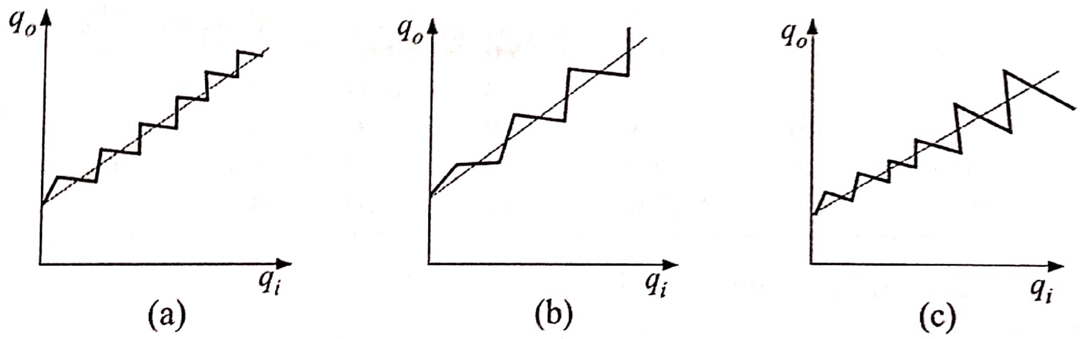

Fig. 2.3 Deviation from linearity: (a) oscillation with fixed amplitude, (b) oscillation with varying amplitude, and (c) combined type oscillation around the best-fit straight line.

1. The actual output of the instrument may oscillate with the same amplitude around the best-fit straight line. In this case, the nonlinearity is expressed in terms of the amplitude (or maximum deviation). The amplitude is calculated as the ± of the full scale deflection (FSD).
2. The actual output of the instrument may oscillate around the best-fit straight line, but the amplitude of oscillation varies with the input value. Here, the nonlinearity is expressed as a function of the input value. Actually, the slopes of lines connecting positive and negative deviations are determined and the one with a higher deviation from the best-fit line is used to express the per cent nonlinearity with respect to the input value.
3. The actual output may oscillate with a fixed amplitude around the best-fit straight line over a certain range and then the amplitude may become a function of the input over the rest. In that case, two computations are made—one for the fixed amplitude part and expressed as ±% of the FSD, and another for the varying amplitude part, expressed as ±% of the input value. Nonlinearity is expressed in terms of the higher value.

Example 2.4

The output of a temperature transducer is recorded over its full-scale range of 25°C as shown below:

Calibration temperature (°C)	0.0	5.0	10.0	15.0	20.0	25.0
Output reading (°C)	0.0	5.0	9.8	14.8	19.9	25.0

Determine (a) the static sensitivity of the device, and (b) the maximum nonlinearity of the device.

Solution

Let q_i be the calibration temperature in °C
q_o be the output reading in °C
S be the sensitivity $= \Delta q_o / \Delta q_i$
D be the deviation from the calibration temperature
Δl be the nonlinearity $= 100D/\text{FSD} = 4D$ since FSD $= 25$°C.

Then, we have

q_i (°C)	Δq_i (°C)	q_o (°C)	Δq_o (°C)	S	D (°C)	Δl (%)
0.0		0.0			0.0	0.0
5.0	5.0	5.0	5.0	1.0	0.0	0.0
10.0	5.0	9.8	4.8	0.96	−0.2	0.8
15.0	5.0	14.8	5.0	1.0	−0.2	0.8
20.0	5.0	19.9	5.1	1.02	−0.1	0.4
25.0	5.0	25.0	5.1	1.02	0.0	0.0

Thus $S_{\text{min}} = 0.96$ and maximum nonlinearity $= 0.8\%$

So far we discussed about the nonlinearity of an output when it is expected to be linear even from theoretical considerations. But there are cases where the output is not expected to be linear, and we have to resort to linear approximations for convenience. For example, the emf-temperature relation of a thermocouple can be written to a first approximation as

$$E = \alpha T + \beta T^2 \tag{2.1}$$

where, α and β are constants for a given thermocouple and T is the temperature of the hot junction, the cold junction being kept at 0°C. For such a thermocouple, if we assume emfs are E_1 and E_2 at temperatures T_1 and T_2 and that a linear relationship exists between emf and temperature, then

$$E_{\text{linear}} = \left(\frac{E_1 - E_2}{T_1 - T_2}\right) T \equiv \alpha' T \tag{2.2}$$

where, $\alpha' = \dfrac{E_1 - E_2}{T_1 - T_2}$. In this case, the nonlinearity N may be expressed as

$$N = E_{\text{actual}} - E_{\text{linear}} = (\alpha - \alpha')T + \beta T^2 \tag{2.3}$$

Alternatively, the nonlinearity may be defined in terms of the nonlinear term βT^2 in expression (2.1). We will consider a case in Example 10.1.

Repeatability

Repeatability is defined broadly as the measure of agreement between the results of successive measurements of the output of a measurement system for repeated applications of a given input in the same way and within the range of calibration of the measurement system. The tests should be made by the same observer, with the same measuring equipment, on the same occasion (i.e. successive measurements should be made in a relatively short span of time), without mechanical or electrical disturbance, and calibration conditions such as temperature, alignment of loading, and the timing of readings held constant as far as possible.

Reproducibility

Reproducibility is defined as the closeness of the agreement between the results of measurements of the same physical quantity carried out under *changed conditions of measurement.* A valid statement of reproducibility requires specification of the particular conditions changed and typically refers to measurements made weeks, months, or years apart. It would also measure, for example, changes caused by dismantling and re-assembling the equipment.

Reproducibility, therefore, determines precision of an instrument. The related undesirable characteristic is *drift.*

2.2 Undesirable Characteristics

As discussed before, the undesirable characteristics of instruments can be divided into three categories—drift, dead zone and static errors.

Drift

Drift denotes the change in the indicated reading of an instrument over time when the value of the measurand remains constant. If there is no drift, the reproducibility is 100%.

Several causes contribute to the drift. Stray electromagnetic fields, mechanical vibrations, changes in superincumbent temperature or pressure, Joule heating of the components of the instrument, etc. are some of the causes. In the case of suspended coil permanent magnet moving coil (PMMC) instruments the release of internal strain of the suspension wire causes drift of the zero-setting.

Dead Zone

Four phenomena —hysteresis, threshold, creep and resolution—contribute to the dead zone.

Hysteresis

Not all the energy put into a system while loading is recoverable upon unloading. For example, a spring balance may show one set of readings when the weight is increased in steps and another set of readings when the weight is decreased in steps. As a result the pointer reading vs weight plot may have the appearance of Fig. 2.4.

The loading and unloading curves do not coincide because of consumption of [illegible] by the internal friction of the solid, and also because of the external sliding fri[illegible]

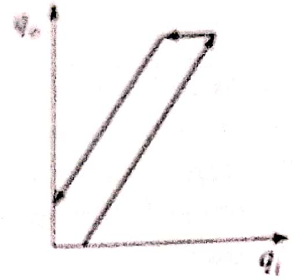

Fig. 2.4 Hysteresis effects shown in an exaggerated way.

components of the instrument. This phenomenon, which is akin to the one experienced during magnetising and demagnetising a magnetic material, is called 'hysteresis'.

Threshold

Suppose an instrument is in its zero position, i.e. there is no input to it. If now an input is gradually applied to it, the instrument will require some minimum value of input before it shows any output (Fig. 2.5).

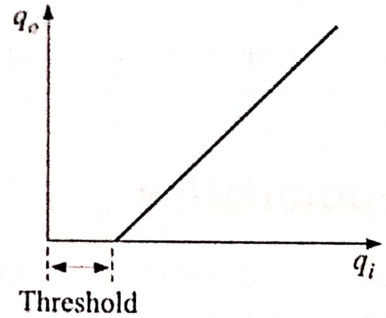

Fig. 2.5 Threshold effect.

This minimum input which is necessary to activate an instrument to produce an output is termed its *threshold*.

Creep

A measurement system may take some time to adjust fully to a change in the applied input, and the creep of a transducer is usually defined as the change of output with time following a step increase in the input from one value to another. Many instrument manufacturers specify the creep as the maximum change of output over a specified time after increasing the input from zero to the rated maximum input. Figure 2.6 shows an example of a creep curve where the transducer exhibits a change in output from R_1 to R_2 over a period of time from t_1 to t_2 after a step change between 0 and t_1. In figures this might be, say, 0.03% of the rated output over 30 minutes.

Creep recovery is the change of output following a step decrease in the applied input to the transducer, usually from the rated maximum input to zero. For both creep and creep recovery, the results will depend on how long the applied input has been at zero or the rated value respectively before the input is changed.

Resolution

Even above the threshold input, an instrument needs a minimum *increment* in input to produce a perceptible output. This minimum necessary increment is called the *resolution* of

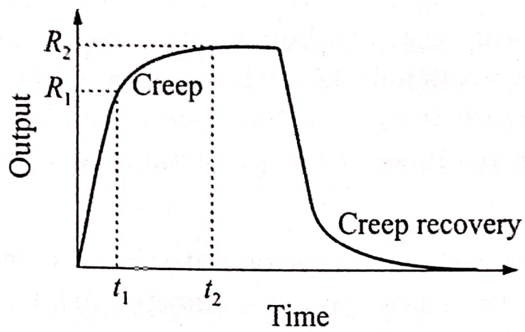

Fig. 2.6 Creep and creep recovery.

the instrument. Thus, resolution which denotes the smallest measurable *change in input* is similar to sliding friction while threshold signifies the smallest *initial input* resembling the static friction.

Example 2.5

An analogue ammeter has a linear scale of 50 divisions. Its full-scale reading is 10 A and half a scale division can be read. What is the resolution of the instrument?

Solution

1 scale division = 10/50 A = 0.2 A. Thus, resolution = 1/2 scale division = (0.2/2) A = 0.1 A.

Example 2.6

The dead-zone in a pyrometer is 0.125% of the span. The instrument is calibrated from 800 to 1800°C. What temperature change must occur before it is detected?

Solution

The span is (1800 – 800) = 1000°C. The dead zone is 0.125% of 1000°C, i.e. 1.25°C. Hence, no change in temperature below 1.25°C can be detected.

Static Errors

The tree in Fig. 2.7 depicts the classification of errors.

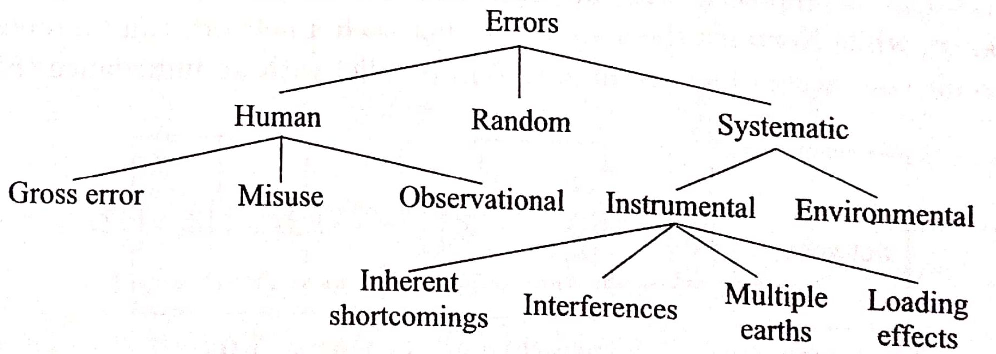

Fig. 2.7 The error tree.

Human errors

One may see from the tree that these errors can be subdivided into the following three classes—gross errors, misuse and observational errors.

Gross errors. Gross errors are basically human mistakes in reading or recording values. Suppose an instrument shows a value of 47.0 while the observer reads it as 42.0. Or, even if he reads the correct value, records it as 41.0. Such errors can be eliminated by automation or minimised by taking multiple readings of the same value at different times and by different observers.

Misuse. A casual approach on the part of the operator is the cause of this error. For example, in electrical measurements, if the leads are not connected firmly, or an ohmic contact[2] is not established or if the initial adjustment such as zero-checking is not done properly, or for a microvolt order measurement proper care is not taken to avoid thermo-emfs arising out of junctions of dissimilar metals, etc. errors will creep in. Alertness and perception on the part of the operator are the only remedy for such errors.

Observational errors. As distinct from gross errors or errors arising out of misuse, observational errors are caused by the observer's lack of knowledge in measurement methods. Parallax is one such error. There may be many more sources of observational errors from set-ups which depend on the so called eye estimation or human reflexes. One such example is the measurement of time period of a pendulum by a stopwatch. Here the precision of the measurement depends on the reflexes of the observer who clicks the stopwatch ON or OFF by noting the position of the bob of the pendulum visually.

Systematic errors

As shown in Fig. 2.7, this error may have two possible origins—instrumental and environmental.

Instrumental error. The instrumental error, in turn, may originate from four different causes—inherent shortcomings, interference, multiple earths and loading effects.

Inherent shortcomings. As the name implies, this error creeps in owing to malfunctioning of the components of instruments due to ageing, etc. For example, the spring of a galvanometer may become weak, thus changing its calibration. Therefore, to avoid this error the calibration of the instrument should be checked from time to time.

Interference. Thevenin's theorem states that a network consisting of linear impedances and voltage sources can be replaced with an equivalent circuit having a voltage source and a series impedance, while Norton's theorem states that such a network can be replaced with an equivalent circuit consisting of a current source in parallel with an impedance (Fig. 2.8).

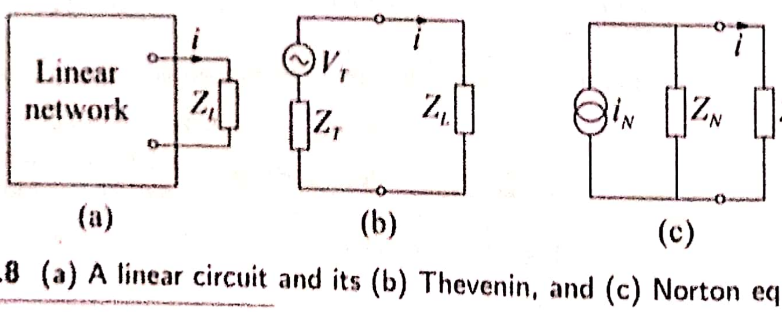

Fig. 2.8 (a) A linear circuit and its (b) Thevenin, and (c) Norton equivalents.

[2] If materials to be contacted are metals, it is easy to establish an ohmic contact between them through a proper cleaning of their surfaces. But if the contact is between a metal and a semiconductor, it is necessary to consider their Fermi levels or else, a rectifying contact may result. For a discussion on this, see *Solid State Electronic Devices*, 4th ed., by B G Streetman, Prentice-Hall of India (1993), pp 187-189.