CHAPTER 9

Display, Recording, and Presentation of Measurement Data

Chapter Outline

Measurement and Instrumentation. http://dx.doi.org/10.1016/B978-0-12-800884-3.00009-5

9.1 Introduction

The earlier chapters in this book have been essentially concerned with describing ways of producing high quality, error-free data at the output of a measurement system. Having obtained the data, the next consideration is how to present it in a form where it can be readily used and analyzed. This chapter therefore starts by covering the techniques available to either display measurement data for immediate use or to record and store it for future use. Following this, standards of good practice for presenting data in either graphical or tabular form are covered, using either paper or a computer monitor screen as the display medium. This leads on to a discussion of mathematical regression techniques for fitting the best lines through data points on a graph. Confidence tests to assess the correctness of the line fitted are also described. Finally, correlation tests are described that determine the degree of association between two sets of data when they are both subject to random fluctuations.

9.2 Display of Measurement Signals

The traditional ways of displaying measurement signals are to use either an *electrical meter* or an *oscilloscope*. However, newer display options now exist as well, such as electronic output displays or using a computer monitor. All of these apart from oscilloscopes are designed to display the magnitude of measurement signals that are in the form of a varying electrical voltage. If a measurement signal exists in some other form, such as changes in the frequency, phase, or current in a signal, conversion to a varying voltage form has to be carried out first using one of the techniques explained earlier in Chapter 7. However, this is not necessary for oscilloscopes if the measurement signal involves changes in frequency or phase, since an oscilloscope is able to display such frequency and phase changes. It should also be noted that electronic displays and computer monitors are digital devices, and analog signals must be converted to digital form before being inputted to these.

We will start this section by looking at electrical meters. These exist in both digital and analog forms, although the use of the analog form now tends to be restricted to panel meters, where the analog form of the output display means that abnormal conditions of monitored systems are more readily identified than is the case with the numeric form of output given by digital meters. As well as signal-level voltages (i.e., low voltages typically in the range up to about 5 V in magnitude), many of the meters available can also measure higher magnitude voltages, and this is indicated where appropriate in the discussion of electrical meters that follows.

Having covered meters, we will go on to look at the oscilloscope as a display device. The oscilloscope is particularly useful for interpreting instrument outputs that exist in the form of a varying phase or frequency of an electrical signal. It is a very versatile measuring instrument that is widely used for signal measurement in spite of the measurement accuracy provided being inferior to that of most meters. Although existing in both analog and digital forms, most instruments used professionally are now digital, with analog versions being limited to cheap, low-specification instruments intended for use in educational establishments. Although of little use to professional users, the features of analog instruments are covered in this chapter because students are quite likely to meet these when doing practical work associated with their course. As far as digital oscilloscopes are concerned, the basic type of instrument used is known as a digital storage oscilloscope. More recently, digital phosphor oscilloscopes have been introduced, which have a capability of detecting and recording rapid transients in voltage signals. A third type is the digital sampling oscilloscope, which is able to measure very high-frequency signals. A fourth and final type is the PC-based oscilloscope, which is effectively an add-on unit to a standard PC. All of these different types of oscilloscope are discussed in Section 9.2.3.

This chapter ends by looking at the newer forms of output display that have emerged in recent years. These new techniques include electronic displays and computer monitor displays.

9.2.1 Digital Meters

All types of digital meters are basically modified forms of the *digital voltmeter* (DVM), irrespective of the quantity that they are designed to measure. Digital meters designed to measure quantities other than voltage are in fact DVMs that contain appropriate electrical circuits to convert current or resistance measurement signals into voltage signals. *Digital multimeters* are also essentially DVMs that contain several conversion circuits, thus allowing the measurement and display of voltage, current, and resistance magnitudes within one instrument.

Digital meters have been developed to satisfy a need for higher measurement accuracies and a faster speed of response to voltage changes than can be achieved with analog instruments. They are technically superior to analog meters in almost every respect. The binary nature of the output reading from a digital instrument can be readily applied to a display that is in the form of discrete numerals. Where human operators are required to measure and record signal voltage levels, this form of output makes an important contribution to measurement reliability and accuracy, since the problem of analog meter parallax error is eliminated and the possibility of gross error through misreading the meter output is greatly reduced. The availability in many instruments of a direct output in digital form is also very useful in the rapidly expanding range of computer control applications. Quoted inaccuracy values are between ±0.005% (measuring DC voltages) and ±2%. Digital meters also have very high input impedance (10 MΩ compared with 1−20 kΩ for analog meters), which avoids the measurement system loading problem (see Chapter 3) that frequently occurs when analog meters are used. Additional advantages of digital meters are their ability to measure signals of frequency up to 1 MHz and the common inclusion of features such as automatic ranging, which prevents overload and reverse polarity connection, etc.

The major part of a DVM is the circuitry that converts the analog voltage being measured into a digital quantity. As the instrument only measures DC quantities in its basic mode, another necessary component within it is one that performs AC−DC conversion and thereby gives it the capacity to measure AC signals. After conversion, the voltage value is displayed by means of indicating tubes or a set of solid-state light-emitting diodes. Four-, five-, or even six-figure output displays are commonly used, and although the instrument itself may not be inherently more accurate than some analog types, this form of display enables measurements to be recorded with much greater accuracy than that obtainable by reading an analog meter scale.

DVMs differ mainly in the technique used to effect the analog-to-digital conversion between the measured analog voltage and the output digital reading. As a general rule, the more expensive and complicated conversion methods achieve a faster conversion speed. Some common types of DVM are discussed below.

Voltage-to-time conversion DVM: This is the simplest form of DVM and is a ramp type of instrument. When an unknown voltage signal is applied to the input terminals of the instrument, a negative-slope ramp waveform is generated internally and compared with the input signal. When the two are equal, a pulse is generated that opens a gate, and at a later point in time a second pulse closes the gate when the negative ramp voltage reaches zero. The length of time between the gate opening and closing is monitored by an electronic counter, which produces a digital display according to the level of the input voltage signal. Its main drawbacks are nonlinearities in the shape of the ramp waveform used and lack of noise rejection, and these problems lead to a typical inaccuracy of ±0.05%. It is relatively cheap, however.

Potentiometric DVM: This uses a servo principle, in which the error between the unknown input voltage level and a reference voltage is applied to a servo-driven potentiometer that adjusts the reference voltage until it balances the unknown voltage. The output reading is produced by a mechanical drum-type digital display driven by the potentiometer. This is also a relatively cheap form of DVM that gives excellent performance for its price.

Dual-slope integration DVM: This is another relatively simple form of DVM that has better noise-rejection capabilities than many other types and give correspondingly better measurement accuracy (inaccuracy as low as ±0.005%). Unfortunately, it is quite expensive. The unknown voltage is applied to an integrator for a fixed time T_1, following which a reference voltage of opposite sign is applied to the integrator, which discharges down to a zero output in an interval T_2 measured by a counter. The output time relationship for the integrator is shown in Figure 9.1, from which the unknown voltage V_i can be calculated geometrically from the triangle as:

$$V_i = V_{\text{ref}}(T_1/T_2) \tag{9.1}$$

Voltage-to-frequency conversion DVM: In this instrument, the unknown voltage signal is fed via a range switch and an amplifier into a converter circuit whose output is in the form of a train of voltage pulses at a frequency proportional to the magnitude of the input signal. The main advantage of this type of DVM is its ability to reject AC noise.

Digital multimeter: This is an extension of the DVM. It can measure both AC and DC voltages over a number of ranges through inclusion within a set of switchable amplifiers and attenuators. It is widely used in circuit test applications as an alternative to the analog multimeter, and includes protection circuits that prevent damage if high voltages are applied to the wrong range.

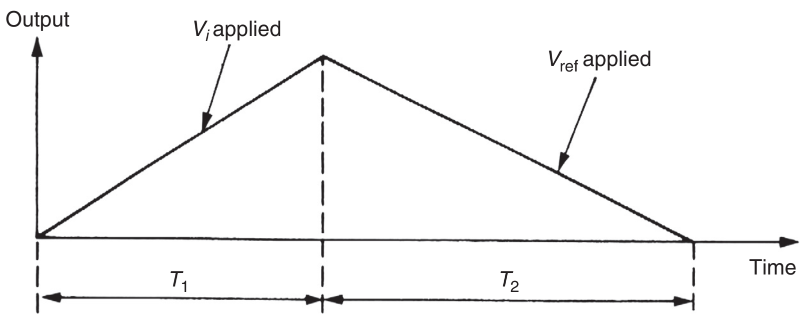

Figure 9.1
Output—time relationship for integrator in a dual-slope digital voltmeter.

9.2.2 Analog Meters

Despite the technical superiority of digital meters, particularly in terms of their greater accuracy and much higher input impedance, analog meters continue to be used in a significant number of applications. Firstly, they are often preferred as indicators in system control panels. This is because deviations of controlled parameters away from the normal expected range are spotted more easily by a pointer moving against a scale in an analog meter rather than by variations in the numeric output display of a digital meter. Analog instruments also tend to suffer less from noise and isolation problems, which favors their use in some applications. In addition, because analog instruments are usually passive instruments that do not need a power supply, this is often very useful in measurement applications where a suitable mains power supply is not readily available. Many examples of analog meter also remain in use for historical reasons. A typical, commercially available analog panel meter is shown in Figure 9.2.

Analog meters are electromechanical devices that drive a pointer against a scale. They are prone to measurement errors from a number of sources that include inaccurate scale marking during manufacture, bearing friction, bent pointers, and ambient temperature variations. Further human errors are introduced through parallax error (not reading the scale from directly above) and mistakes in interpolating between scale markings. Quoted inaccuracy values are between ±0.1% and ±3%. Various types of analog meter are used as discussed below.

Moving coil meter

A moving coil meter is a very common form of analog voltmeter because of its sensitivity, accuracy, and linear scale, although it only responds to DC signals. As shown schematically in Figure 9.3, it consists of a rectangular coil wound round a soft iron core that is suspended in the field of a permanent magnet. The signal being measured is applied

Figure 9.2
Eltime analog panel meter. *Reproduced by kind permission of Eltime Controls.*

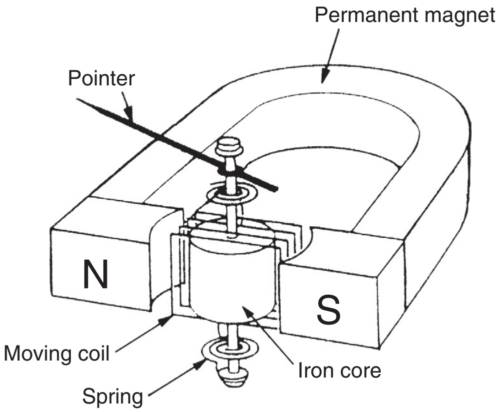

Figure 9.3
Mechanism of moving coil meter.

to the coil and this produces a radial magnetic field. Interaction between this induced field and the field produced by the permanent magnet causes a torque, which results in rotation of the coil. The amount of rotation of the coil is measured by attaching a pointer to it that moves past a graduated scale. The theoretical torque produced is given by:

$$T = BIhwN \tag{9.2}$$

where B is the flux density of the radial field, I is the current flowing in the coil, h is the height of the coil, w is the width of the coil, and N is the number of turns in the coil. If the iron core is cylindrical and the air gap between the coil and pole faces of the permanent magnet is uniform, then the flux density B is constant, and Eqn (9.2) can be rewritten as:

$$T = KI \tag{9.3}$$

i.e., the torque is proportional to the coil current and the instrument scale is linear.

The torque due to the coil current is opposed by the reaction torque of a torsional spring such that the final rotational displacement of the coil is proportional to the measured current.

As the basic instrument operates at low current levels of 1 mA or so, it is only suitable for measuring voltages up to around 2 V. If there is a requirement to measure higher voltages, the measuring range of the instrument can be increased by placing a resistance in series with the coil, such that only a known proportion of the applied voltage is measured by the meter. In this situation, the added resistance is known as a *shunting resistor*.

While Figure 9.3 shows the traditional moving coil instrument with a long U-shaped permanent magnet, many newer instruments employ much shorter magnets made from recently developed magnetic materials such as Alnico and Alcomax. These materials produce a substantially greater flux density, which, besides allowing the magnet to be smaller, has additional advantages in allowing reductions to be made in the size of the coil and in increasing the usable range of deflection of the coil to about 120°. Some versions of the instrument also have either a specially shaped core or specially shaped magnet pole-faces to cater for special situations where a nonlinear scale such as a logarithmic one is required.

Moving iron meter

As well as measuring DC signals, the moving iron meter can also measure AC signals at frequencies up to 125 Hz. It is the cheapest form of meter available and is used in similar numbers to moving coil meters. The signal to be measured is applied to a stationary coil, and the associated field produced is often amplified by the presence of an iron structure associated with the fixed coil. The moving element in the instrument consists of an iron vane that is suspended within the field of the fixed coil. When the fixed coil is excited, the iron vane turns in a direction that increases the flux through it.

The majority of moving iron instruments are either of the attraction type or of the repulsion type. A few instruments belong to a third combination type. The attraction type, where the iron vane is drawn into the field of the coil as the current is increased, is shown schematically in Figure 9.4(a). The alternative repulsion type is sketched in Figure 9.4(b). For an excitation current I, the torque produced that causes the vane to turn is given by:

$$T = \frac{I^2 dM}{2d\theta}$$

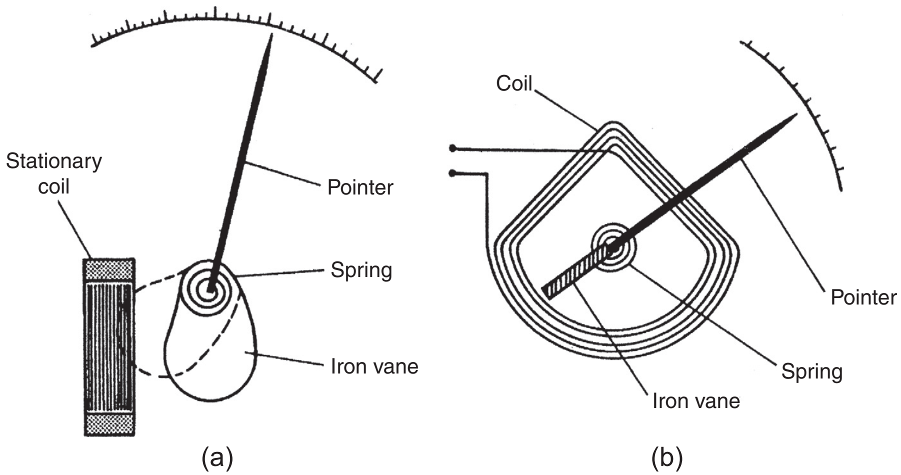

Figure 9.4
Mechanisms of moving iron meters (a) Attraction-type; (b) repulsion-type.

where M is the mutual inductance and θ is the angular deflection. Rotation is opposed by a spring that produces a backward torque given by:

$$T_s = K\theta$$

At equilibrium, $T = T_s$, and θ is therefore given by:

$$\theta = \frac{I^2 dM}{2K d\theta} \tag{9.4}$$

The instrument thus has a square-law response where the deflection is proportional to the square of the signal being measured, i.e., the output reading is a root-mean-squared (rms) quantity.

The instrument can typically measure voltages in the range of 0−30 V. However, it can be modified to measure higher voltages by placing a resistance in series with it, as in the case of moving coil meters. A series resistance is particularly beneficial in AC signal measurements because it compensates for the effect of coil inductance by reducing the total resistance/inductance ratio, and hence measurement accuracy is improved. A switchable series resistance is often provided within the casing of the instrument to facilitate range extension. However, when the voltage measured exceeds about 300 V, it becomes impractical to use a series resistance within the case of the instrument because of heat-dissipation problems, and an external resistance is used instead.

Clamp-on meters

These are used for measuring circuit currents and voltages in a noninvasive manner that avoids having to break the circuit being measured. The meter clamps onto a current-carrying conductor, and the output reading is obtained by transformer action. The principle of operation is illustrated in Figure 9.5, where it can be seen that the clamp-on jaws of the instrument act as a transformer core and the current-carrying conductor acts as a primary winding. Current induced in the secondary winding is rectified and applied to a moving coil meter. Although it is a very convenient instrument to use, the clamp-on meter has low sensitivity and the minimum current measurable is usually about 1 A.

Analog multimeter

The analog multimeter is now less common than its counterpart, the digital multimeter, but is still widely available. It is a multifunction instrument that can measure current and resistance as well as DC and AC voltage signals. Basically, the instrument consists of a moving coil analog meter with a switchable bridge rectifier to allow it to measure AC signals, as shown in Figure 9.6. A set of rotary switches allows the selection of various series and shunt resistors, which make the instrument capable of measuring both voltage and current over a number of ranges. An internal power source is also provided to allow it

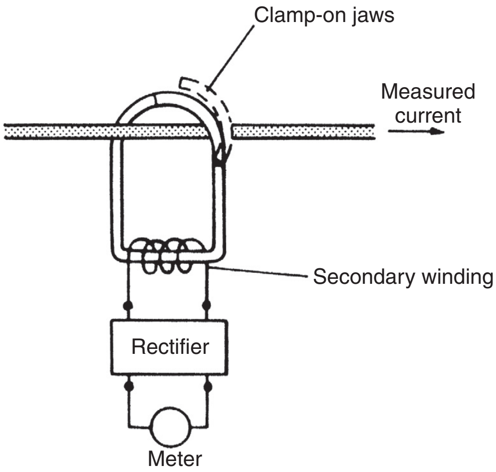

Figure 9.5
Schematic drawing of clamp-on meter.

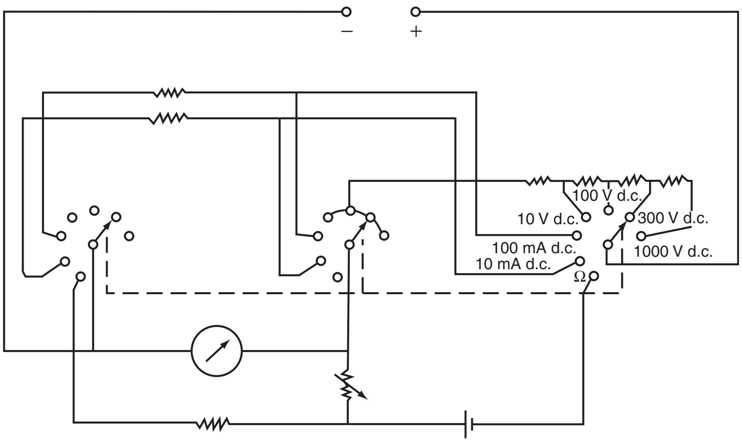

Figure 9.6
Circuitry of analog multimeter.

to measure resistances as well. While this instrument is very useful for giving an indication of voltage levels, the compromises in its design that enable it to measure so many different quantities necessarily mean that its accuracy is not as good as instruments that are purpose-designed to measure just one quantity over a single measuring range.

Measuring high-frequency signals with analog meters

One major limitation in using analog meters for AC voltage measurement is that the maximum frequency measurable directly is low, 2 kHz for the dynamometer voltmeter and only 100 Hz in the case of the moving iron instrument. A partial solution to this limitation is to rectify the voltage signal and then apply it to a moving coil meter, as shown in Figure 9.7. This extends the upper measurable frequency limit to 20 kHz. However, the

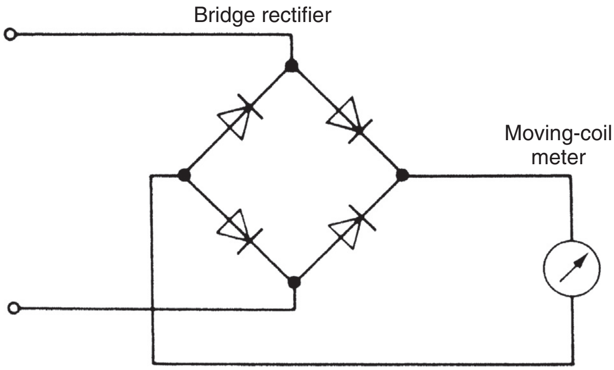

Figure 9.7
Measurement of high-frequency voltage signals.

inclusion of the bridge rectifier makes the measurement system particularly sensitive to environmental temperature changes, and nonlinearities significantly affect measurement accuracy for voltages that are small relative to the full-scale value.

Calculation of meter outputs for nonstandard waveforms

The two examples below provide an exercise in calculating the output reading from various types of analog voltmeter. These examples also serve as a useful reminder of the mode of operation of each type of meter and the form that the output takes.

Example 9.1

Calculate the reading that would be observed on a moving coil ammeter when it is measuring the current in the circuit shown in Figure 9.8.

Solution

A moving coil meter measures mean current.

$$I_{\text{mean}} = \frac{1}{2\pi}\left(\int_0^{\pi} \frac{5\omega t}{\pi} d\omega t + \int_{\pi}^{2\pi} 5\sin(\omega t) d\omega t\right) = \frac{1}{2\pi}\left(\left[\frac{5(\omega t)^2}{2\pi}\right]_0^{\pi} + 5[-\cos(\omega t)]_{\pi}^{2\pi}\right)$$

$$= \frac{1}{2\pi}\left(\frac{5\pi^2}{2\pi} - 0 - 5 - 5\right) = \frac{1}{2\pi}\left(\frac{5\pi}{2} - 10\right) = \frac{5}{2\pi}\left(\frac{\pi}{2} - 2\right) = -0.342 \text{ A}$$

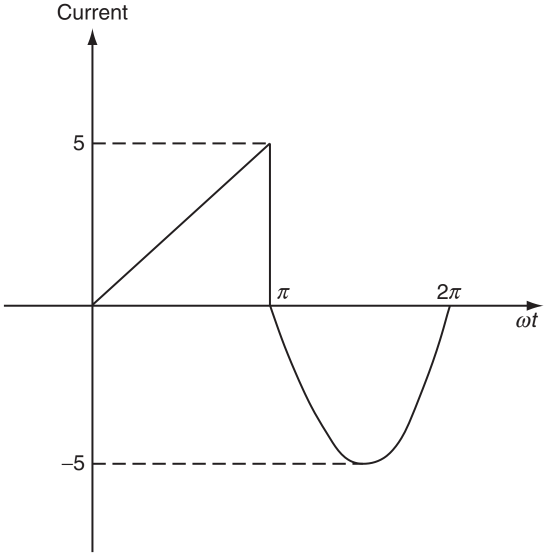

Figure 9.8
Circuit for Examples 9.1 and 9.2.

Example 9.2

Calculate the reading that would be observed on a moving iron ammeter when it is measuring the current in the circuit shown in Figure 9.8.

Solution

A moving iron meter measures rms current.

$$I_{\text{rms}}^2 = \frac{1}{2\pi}\left(\int_0^{\pi}\frac{25(\omega t)^2}{\pi^2}d\omega t + \int_{\pi}^{2\pi} 25\sin^2(\omega t)d\omega t\right)$$

$$= \frac{1}{2\pi}\left(\int_0^{\pi}\frac{25(\omega t)^2}{\pi^2}d\omega t + \int_{\pi}^{2\pi}\frac{25(1-\cos 2\,\omega t)}{2}d\omega t\right)$$

$$= \frac{25}{2\pi}\left(\left[\frac{(\omega t)^3}{3\pi^2}\right]_0^{\pi} + \left[\frac{\omega t}{2} - \frac{\sin 2\,\omega t}{4}\right]_{\pi}^{2\pi}\right) = \frac{25}{2\pi}\left(\frac{\pi}{3}+\frac{2\pi}{2}-\frac{\pi}{2}\right) = \frac{25}{2\pi}\left(\frac{\pi}{3}+\frac{\pi}{2}\right)$$

$$= \frac{25}{2}\left(\frac{1}{3}+\frac{1}{2}\right) = 10.416$$

Thus, $I_{\text{rms}} = \sqrt{(I_{\text{rms}}^2)} = 3.23$ A

9.2.3 Oscilloscopes

The oscilloscope is probably the most versatile and useful instrument available for signal measurement and display. While oscilloscopes still exist in both analog and digital forms, analog models tend to be found more often in the form of low specification, low cost instruments that are produced for educational use in schools, colleges, and universities. Most oscilloscopes used for professional work now tend to be digital models. These can be divided into digital storage oscilloscopes, digital phosphor oscilloscopes, and digital sampling oscilloscopes.

The basic function of an oscilloscope is to draw a graph of an electrical signal. In the most common mode of usage, the y-axis (vertical) of the display represents the voltage of a measured signal and the x-axis (horizontal) represents time. Thus, the basic output display is a graph of the variation of the magnitude of the measured voltage with time.

The oscilloscope is able to measure a very wide range of both AC and DC voltage signals, and is used particularly as an item of test equipment for circuit fault-finding. Besides measuring voltage levels, it can also measure other quantities such as the frequency and phase of a signal. It can also indicate the nature and magnitude of noise that may be corrupting the measurement signal. The most expensive models can measure signals at frequencies up to 25 GHz, while the very cheapest models can only measure signals up to 10 MHz. One particularly strong merit of the oscilloscope is its high input impedance, typically 1 MΩ, which means that the instrument has a negligible loading effect in most measurement situations. As a test instrument, it is often required to measure voltages whose frequency and magnitude are totally unknown. The set of rotary switches that alter its time base so easily, and the circuitry that protects it from damage when high voltages are applied to it on the wrong range, make it ideally suited for such applications. However, it is not a particularly accurate instrument and is best used where only an approximate measurement is required. In the best instruments, inaccuracy can be limited to $\pm 1\%$ of the reading but inaccuracy can approach $\pm 5\%$ in the cheapest instruments.

The most important aspects in the specification of an oscilloscope are its bandwidth, rise time, and accuracy. The bandwidth is defined as the range of frequencies over which the oscilloscope amplifier gain is within 3 dB[1] of its peak value, as illustrated in Figure 9.9. The -3 dB point is where the gain is 0.707 times its maximum value. In most oscilloscopes, the amplifier is direct-coupled, which means that it amplifies DC voltages by the same factor as low frequency AC ones. For such instruments, the minimum frequency measurable is zero and the bandwidth can be interpreted as the maximum

[1] The decibel, commonly written dB, is used to express the ratio between two quantities. For two voltage levels V_1 and V_2, the difference between the two levels is expressed in decibels as $20\log_{10}(V_1/V_2)$. It follows from this that $20\log_{10}(0.7071) = -3$ dB.

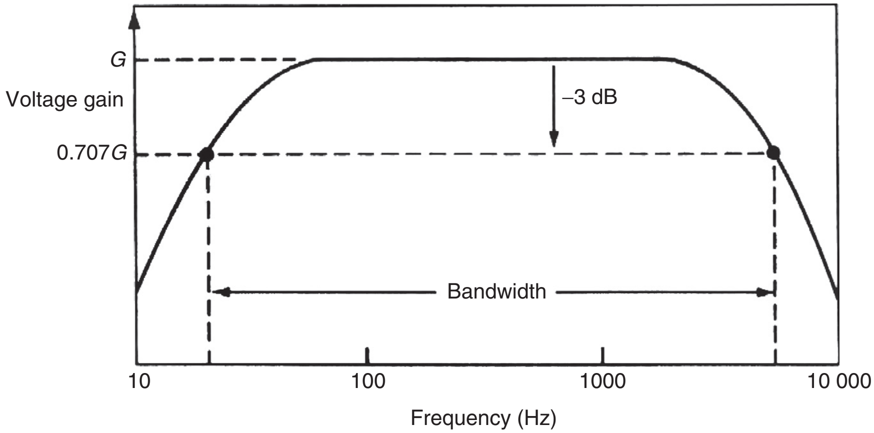

Figure 9.9
Bandwidth.

frequency where the sensitivity (deflection/volt) is within 3 dB of the peak value. In all measurement situations, the oscilloscope chosen for use must be such that the maximum frequency to be measured is well within the bandwidth. The −3 dB specification means that an oscilloscope with a specified inaccuracy of ±2% and bandwidth of 100 MHz will have an inaccuracy of ±5% when measuring 30 MHz signals, and this inaccuracy will increase still further at higher frequencies. Thus, when applied to signal-amplitude measurement, the oscilloscope is only usable at frequencies up to about 0.3 times its specified bandwidth.

The rise time is the transit time between the 10% and 90% levels of the response when a step input is applied to the oscilloscope. Oscilloscopes are normally designed such that:

$$\text{Bandwidth} \times \text{Rise time} = 0.35$$

Thus, for a bandwidth of 100 MHz, rise time = 0.35/100,000,000 = 3.5 ns.

All oscilloscopes are relatively complicated instruments that are constructed from a number of subsystems, and it is necessary to consider each of these in turn in order to understand how the complete instrument functions. To achieve this, it is useful to start with an explanation of an analog oscilloscope, since this was the original form in which oscilloscopes were made and many of the terms used to describe the function of oscilloscopes emanate from analog forms.

Analog oscilloscope (cathode ray oscilloscope)

Analog oscilloscopes were originally called cathode ray oscilloscopes because a fundamental component within them is a cathode ray tube. In recent times, digital oscilloscopes largely replaced analog versions in professional use. A particular attribute

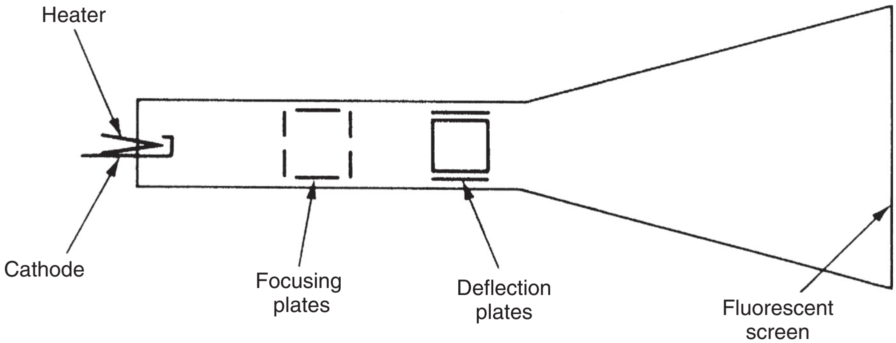

Figure 9.10
Cathode ray tube.

of analog oscilloscopes is that they are less expensive than digital ones. This means that they still find substantial educational usage in schools, colleges, and universities. However, the low cost of basic analog models is their only real merit, since their inclusion of a cathode ray tube makes them very fragile, and the technical performance of digital equivalents is greatly superior.

The *cathode ray tube* within an analog oscilloscope is shown schematically in Figure 9.10. The cathode consists of a barium and strontium oxide coated, thin, heated filament from which a stream of electrons is emitted. The stream of electrons is focused onto a well-defined spot on a fluorescent screen by an electrostatic focusing system that consists of a series of metal discs and cylinders charged at various potentials. Adjustment of this focusing mechanism is provided by a *focus* control on the front panel of an oscilloscope. An *intensity* control varies the cathode heater current and therefore the rate of emission of electrons, and thus adjusts the intensity of the display on the screen. These and other typical controls are shown in the illustration of the front panel of a simple oscilloscope given in Figure 9.11. It should be noted that the layout shown is only one example. Every model of oscilloscope has a different layout of control knobs, but the functions provided remain similar irrespective of the layout of the controls with respect to each other.

Application of potentials to two sets of deflector plates mounted at right angles to one another within the tube provide for deflection of the stream of electrons, such that the spot where the electrons are focused on the screen is moved. The two sets of deflector plates are normally known as the horizontal and vertical deflection plates, according to the respective motion caused to the spot on the screen. The magnitude of any signal applied to the deflector plates can be calculated by measuring the deflection of the spot against a cross-wires graticule etched on the screen.

Channel: One channel describes the basic subsystem of an electron source, focusing system and deflector plates. This subsystem is often duplicated one or more times within

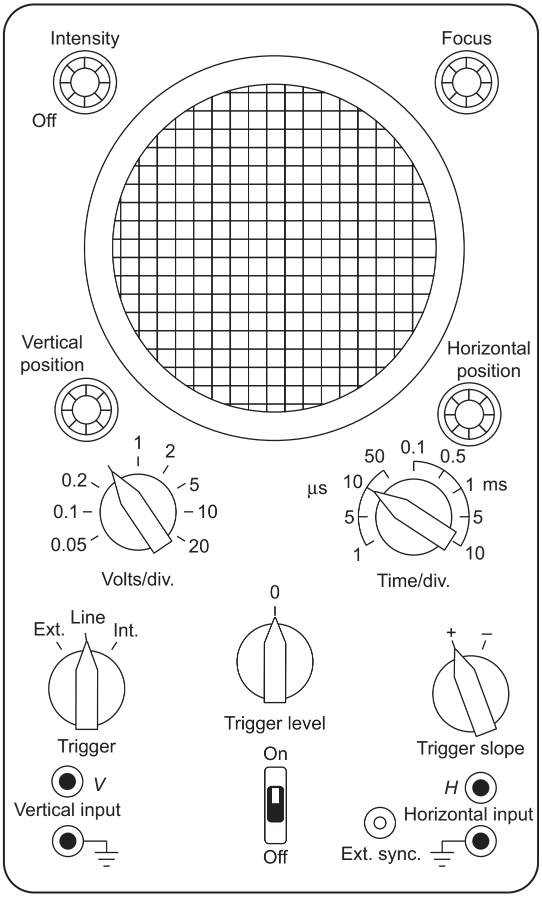

Figure 9.11
Controls of a simple oscilloscope.

the cathode ray tube to provide a capability of displaying two or more signals at the same time on the screen. The common oscilloscope configuration with two channels can therefore display two separate signals simultaneously.

Single-ended input: This type of input only has one input terminal plus a ground terminal per oscilloscope channel and, consequently, only allows signal voltages to be measured relative to ground. It is normally only used in simple oscilloscopes.

Differential input: This type of input is provided on more expensive oscilloscopes. Two input terminals plus a ground terminal are provided for each channel, which allows the potentials at two nongrounded points in a circuit to be compared. This type of input can also be used in single-ended mode to measure a signal relative to ground by using just one of the input terminals plus ground.

Time base circuit: The purpose of a time base is to apply a voltage to the horizontal deflector plates such that the horizontal position of the spot is proportional to time. This voltage, in the form of a ramp known as a sweep waveform, must be applied repetitively, such that the motion of the spot across the screen appears as a straight line when a DC level is applied to the input channel. Furthermore, this time base voltage must be synchronized with the input signal in the general case of a time-varying signal, such that a steady picture is obtained on the oscilloscope screen. The length of time taken for the spot to traverse the screen is controlled by a *time/div* switch, which sets the length of time taken by the spot to travel between two marked divisions on the screen, thereby allowing signals at a wide range of frequencies to be measured.

Each cycle of the sweep waveform is initiated by a pulse from a pulse generator. The input to the pulse generator is a sinusoidal signal known as a triggering signal, with a pulse being generated every time the triggering signal crosses a preselected slope and voltage level condition. This condition is defined by the *trigger level* and *trigger slope* switches. The former selects the voltage level on the trigger signal, commonly zero, at which a pulse is generated, while the latter selects whether pulsing occurs on a positive or negative going part of the triggering waveform.

Synchronization of the sweep waveform with the measured signal is most easily achieved by deriving the trigger signal from the measured signal, a procedure that is known as *internal triggering*. Alternatively, *external triggering* can be applied if the frequencies of the triggering signal and measured signals are related by an integer constant such that the display is stationary. External triggering is necessary when the amplitude of the measured signal is too small to drive the pulse generator, and it is also used in applications where there is a requirement to measure the phase difference between two sinusoidal signals of the same frequency. It is very convenient to use the 50 Hz line voltage for external triggering when measuring signals at mains frequency, and this is often given the name *line triggering*.

Vertical sensitivity control: This consists of a series of attenuators and preamplifiers at the input to the oscilloscope. These condition the measured signal to the optimum magnitude for input to the main amplifier and vertical deflection plates, thus enabling the instrument to measure a very wide range of different signal magnitudes. Selection of the appropriate input amplifier/attenuator is made by setting a *volts/div* control associated with each oscilloscope channel. This defines the magnitude of the input signal that will cause a deflection of one division on the screen.

Display position control: This allows the position at which a signal is displayed on the screen to be controlled in two ways. The horizontal position is adjusted by a *horizontal position* knob on the oscilloscope front panel and similarly a *vertical position* knob controls the vertical position. These controls adjust the position of the display by biasing the measured signal with DC voltage levels.

Digital storage oscilloscopes

Digital storage oscillosopes are the most basic form of digital oscilloscopes but even these usually have the ability to perform extensive waveform processing and provide permanent storage of measured signals. When first created, a digital storage oscilloscope consisted of a conventional analog cathode ray oscilloscope with the added facility that the measured analog signal could be converted to digital format and stored in computer memory within the instrument. This stored data could then be reconverted to analog form at the frequency necessary to refresh the analog display on the screen, producing a nonfading display of the signal on the screen.

While examples of such early digital oscilloscopes might still be found in some workplaces, modern digital storage oscillosopes no longer use cathode ray tubes and are entirely digital in construction and operation. The front panel of any digital oscilloscope has a similar basic layout to that shown for an analog oscilloscope in Figure 9.11, except that the controls for "focusing" and "intensity" are not needed in a digital instrument. The block diagram in Figure 9.12 shows the typical components used in the digital storage oscilloscope. The first component (as in an analog oscilloscope) is an amplifier/attentuator unit that allows adjustment of the magnitude of the input voltage signal to an appropriate level. This is followed by an analog-to-digital converter, which samples the input signal at discrete points in time. The sampled signal values are stored in the acquisition memory component before passing into a microprocessor. This carries out signal processing functions, manages the front panel control settings and prepares the output display. Following this, the output signal is stored in a display memory module before being output to the display itself. This electronic output display consists of either monochrome or multicolor liquid crystal elements (see Section 9.2.4). The signal displayed is actually a

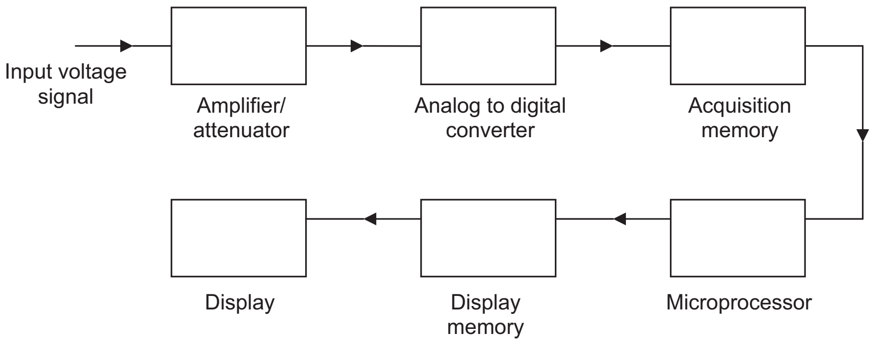

Figure 9.12
Components of a digital storage oscilloscope.

sequence of individual dots rather than a continuous line as displayed by an analog oscilloscope. However, as the density of dots increases, the display becomes closer and closer to a continuous line. The density of the dots is entirely dependent upon the sampling rate at which the analog signal is digitized and the rate at which the memory contents are read to reconstruct the original signal. As the speed of sampling and signal processing is a function of instrument cost, more expensive instruments give better performance in terms of dot density and the accuracy with which the analog signal is recorded and represented. Nevertheless, the cost of computing power is now sufficiently low to mean that all but the very cheapest instruments now have a display that look very much like a continuous trace.

Besides their ability to display the magnitude of voltage signals and other parameters such as signal phase and frequency, most digital oscilloscopes can also carry out analysis of the measured waveform and compute signal parameters such as maximum and minimum signal levels, peak—peak values, mean values, rms values, rise time, and fall time. These additional functions are controlled by extra knobs and push buttons on the front panel. They are also ideally suited to capturing transient signals when set to single-sweep mode. This avoids the problem of the very careful synchronization that is necessary to capture such signals on an analog oscilloscope. In addition, digital oscilloscopes often have facilities to output analog signals to devices like chart recorders and output digital signals in a form that is compatible with standard interfaces like IEEE488 and RS232.

The principal limitation of a digital storage oscilloscope is that the only signal information captured is the status of the signal at each sampling instant. Thereafter, no new signal information is captured during the time that the previous sample is being processed. This means that any signal changes occurring between sampling instants, such as fast transients, are not detected. This problem is overcome in the digital phosphor oscilloscope.

Digital phosphor oscilloscope

This newer type of oscilloscope, first introduced in 1998, uses a parallel-processing architecture rather than the serial-processing architecture found in digital storage oscilloscopes. The components of the instrument are shown schematically in Figure 9.13. The amplifier/attenuator and analog-to-digital converter are the same as in a digital storage oscilloscope. However, the signal processing mechanism is substantially different. The output from the analog-to-digital converter passes into a digital phosphor memory unit, which is in fact entirely electronic and not composed of chemical phosphor as its name might imply. Thereafter, data follow two parallel paths. Firstly, a microprocessor processes the data acquired at each sampling instant according to the settings on the control panel and sends the processed signal to the instrument display unit. In addition to this, a snapshot of the input signal is sent directly to the display unit at a rate of 30 images per second. This enhanced processing capability enables the instrument to have a higher

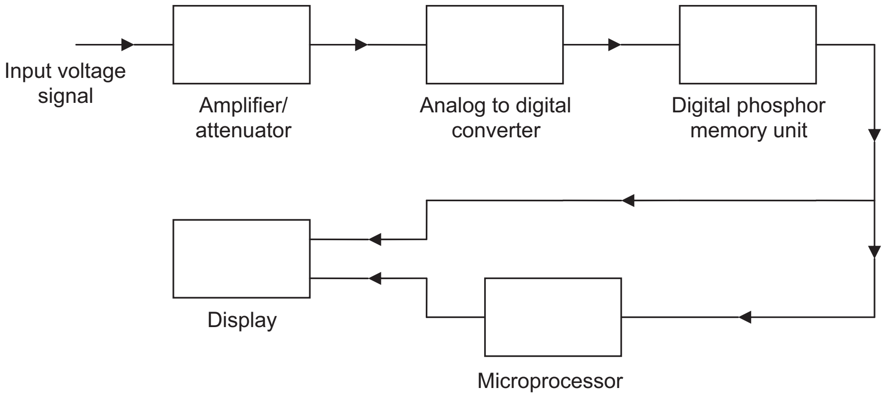

Figure 9.13
Components of a digital phosphor oscilloscope.

waveform capture rate and to detect very fast signal transients that are missed by digital storage oscilloscopes.

Digital sampling oscilloscope

The digital sampling oscilloscope has a bandwidth of up to 25 GHz, which is about 10 times better than that achieved by other types of oscilloscope. This increased bandwidth is achieved by reversing the positions of the analog-to-digital converter and the amplifier, as shown in the block diagram in Figure 9.14. This reversal means that the sampled signal applied to the amplifier has a much lower frequency than the original signal, allowing the use of a low bandwidth amplifier. However, the fact that the input signal is applied directly to the analog-to-digital converter without any scaling means that the instrument can only be used to measure signals whose peak magnitude is within a relatively small range of typically 1 V peak−peak. By contrast, both digital storage and digital phosphor oscilloscopes can typically deal with inputs up to 500 V.

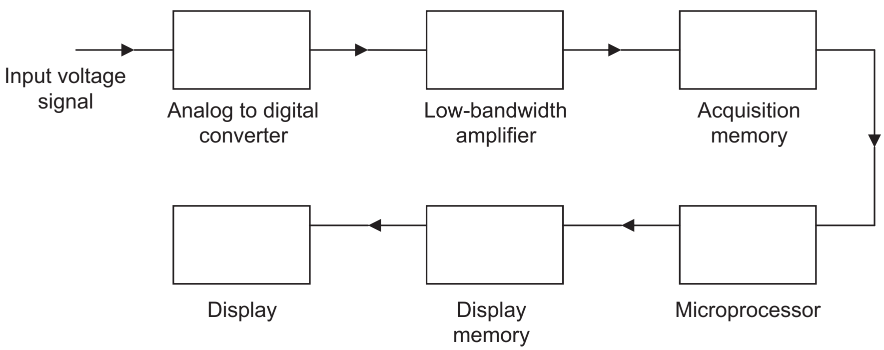

Figure 9.14
Components of a digital sampling oscilloscope.

PC-based oscilloscope

A PC-based oscilloscope consists of a hardware unit that connects to a standard PC via either a USB or a parallel port. The hardware unit provides the signal scaling, analog-to-digital conversion and buffer memory functions found in a conventional oscilloscope. More expensive PC-based oscilloscopes also provide some high-speed digital signal processing functions within the hardware unit. The host PC itself provides the control interface and display facilities.

The primary advantage of a PC-based oscilloscope over other types is one of cost, the cost saving being achieved because the use of the PC obviates the need for the display unit and the front control panel found in other forms of oscilloscope. The larger size of a PC display compared with a conventional oscilloscope often makes the output display easier to read. A further advantage is one of portability, since a laptop plus add-on hardware unit is usually smaller and lighter than a conventional oscilloscope. PC-based oscilloscopes also facilitate the transfer of output data into standard PC software such as spreadsheets and wordprocessors.

Although pc-based oscilloscopes have a number of advantages over conventional oscilloscopes, they also have disadvantages. Firstly, electromagnetic noise originating in PC circuits require the hardware unit to be well shielded in order to avoid corruption of the measured signal. Secondly, signal sampling rates can be limited by the mode of connection of the hardware unit into the PC.

9.2.4 Electronic Output Displays

Electronic displays enable a parameter value to be read immediately, thus allowing for any necessary response to be made immediately. The main requirement for displays is that they should be clear and unambiguous. Two common types of character format used in displays, seven-segment and 7×5 dot matrix, are shown in Figure 9.15. Both types of display have the advantage of being able to display alphabetic as well as numeric information, although the 7-segment format can only display a limited 9-letter subset of the full 26-letter alphabet. This allows added meaning to be given to the number displayed by including a word or letter code. It also allows a single display unit to send information about several parameter values, cycling through each in turn and including alphabetic information to indicate the nature of the variable currently displayed.

Electronic output units usually consist of a number of side-by-side cells, where each cell displays one character. Generally, these accept either serial or parallel digital input signals, and the input format can be either binary-coded decimal (BCD) or ACSII. Technologies used for the individual elements in the display are either light-emitting diodes (LEDs) or liquid-crystal elements.

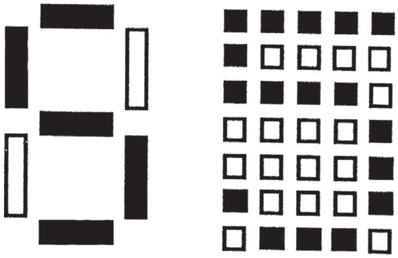

Figure 9.15
Character formats used in electronic displays: (a) Seven-segment; (b) 7 × 5 dot matrix.

9.2.5 Computer Monitor Displays

Now that computers are part of the furniture in most homes, the ability of computers to display information is widely understood and appreciated. Computers are now both cheap and highly reliable, and they provide an excellent mechanism for both displaying and storing information. As well as alphanumeric displays of industrial plant variable and status data, for which the plant operator can vary the size of font used to display the information at will, it is also relatively easy to display other information such as plant layout diagrams, process flow layouts etc. This allows not only the value of parameters that go outside control limits to be displayed, but also their location on a schematic map of the plant. Graphical displays of the behavior of a measured variable are also possible. However, this poses a difficulty when there is a requirement to display the variable's behavior over a long period of time since the length of the time axis is constrained by the size of the monitor's screen. To overcome this, the display resolution has to decrease as the time period of the display increases.

Touch screens have the ability to display the same sort of information as a conventional computer monitor, but they also provide a command-input facility in which the operator simply has to touch the screen at points where images of keys or boxes are displayed. A full "qwerty" keyboard is often provided as part of the display. The sensing elements behind the screen are protected by the glass and continue to function even if the glass gets scratched. Touch screens are usually totally sealed, and thus provide intrinsically safe operation in hazardous environments.

9.3 Recording of Measurement Data

As well as displaying the current values of measured parameters, there is often a need to make continuous recordings of measurements for later analysis. Such records are particularly useful when faults develop in systems, as analysis of the changes in measured parameters in the time before the fault is discovered can often quickly indicate the reason for the fault. The options for recording data include chart recorders, digital oscilloscopes, digital data recorders, and hard-copy devices such as inkjet and laser printers. The various types of recorder used are discussed below.

9.3.1 Chart Recorders

Chart recorders have particular advantages in providing a noncorruptible record that has the merit of instant "viewability." This means that all but paperless forms of chart recorder satisfy regulations set for many industries that require variables to be monitored and recorded continuously with hard-copy output. ISO9000 quality assurance procedures and ISO14000 environmental protection systems set similar requirements, and special regulations in the defense industry go even further by requiring hard-copy output to be kept for 10 years. Hence, while many people have been predicting the demise of chart recorders, the reality of the situation is that they are likely to be needed in many industries for many years to come.

Originally, all chart recorders were electromechanical in operation and worked on the same principle as a galvanometric moving coil meter (see analog meters in Section 9.2.2) except that the moving coil to which the measured signal was applied carried a pen as shown in Figure 9.16 rather than carrying a pointer moving against a scale as it would do in a meter. The pen drew an ink trace on a strip of ruled chart paper that was moved past the pen at constant speed by an electrical motor. The resultant trace on the chart paper showed the variations with time in the magnitude of the measured signal. Even early

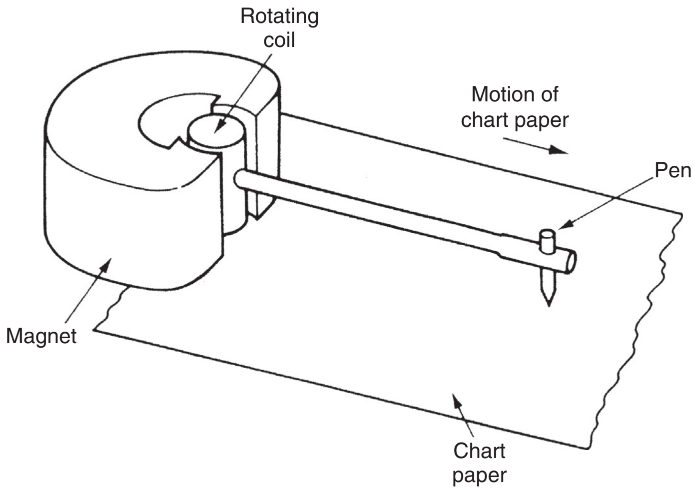

Figure 9.16
Original form of galvanometric chart recorder.

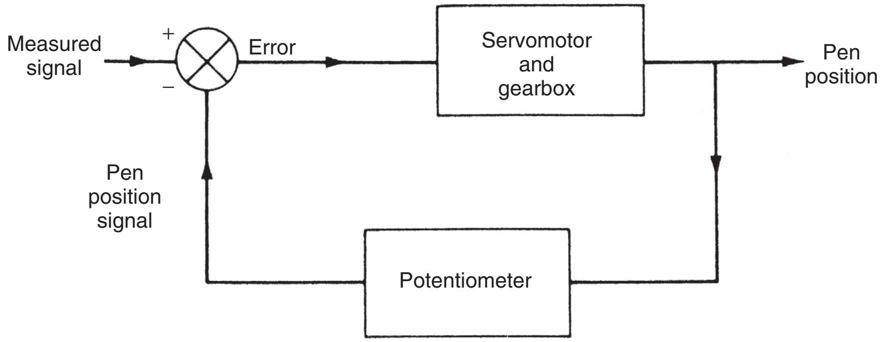

Figure 9.17
Servo system of potentiometric chart recorder.

recorders commonly had two or more pens of different colors so that several measured parameters could be recorded simultaneously.

The first improvement to this basic recording arrangement was to replace the galvanometric mechanism with a servo system, as shown in Figure 9.17, in which the pen is driven by a servomotor, and a sensor on the pen feeds back a signal proportional to pen position. In this form, the instrument is known as a *potentiometric recorder*. The servo system reduces the typical inaccuracy of the recorded signal to ±0.1%, compared to ±2% in a galvanometeric mechanism recorder. Typically, the measurement resolution is around 0.2% of the full-scale reading. Originally, the servo motor was a standard DC motor, but brushless servo motors are now invariably used to avoid the commutator problems that occur with DC motors. The position signal is measured by a potentiometer in cheaper models, but more expensive models achieve better performance and reliability by using a noncontacting ultrasonic sensor to provide feedback on pen position. The difference between the pen position and the measured signal is applied as an error signal that drives the motor. One consequence of this electromechanical balancing mechanism is that the instrument has a slow response time in the range of 0.2–2.0 s, which means that electromechanical potentiometric recorders are only suitable for measuring DC and slowly time-varying signals.

All current potentiometric chart recorders contain a microprocessor controller, where the functions vary according to the particular chart recorder. Common functions are selection of range and chart speed, and also specification of alarm modes and levels to detect when measured variables go outside acceptable limits. Basic recorders can record up to three different signals using three different colored pens. However, multipoint recorders can have 24 or more inputs and plot six or more different colored traces simultaneously. As an alternative to pens, which can run out of ink at inconvenient times, recorders using a heated stylus recording signals on heat-sensitive paper are available. Another variation is the circular chart recorder, in which the chart paper is circular in shape and is rotated rather than moving translationally. Finally, paperless forms of recorder exist where the output display is generated entirely electronically. These various forms are discussed in more detail below.

Pen strip chart recorder

The pen strip chart recorder refers to the basic form of electromechanical potentiometric chart recorder mentioned above. It is also called a *hybrid chart recorder* by some manufacturers. The word "hybrid" was originally used to differentiate chart recorders that had a microprocessor controller from those that did not. However, since all chart recorders now contain a microprocessor, the term hybrid has become superfluous.

Strip chart recorders typically have up to three inputs and up to three pens in different colors, allowing up to three different signals to be recorded. A typical commercially available model is shown in Figure 9.18. Chart paper comes in either roll or fan-fold form. The drive mechanism can be adjusted to move the chart paper at different speeds. The fastest speed is typically 6000 mm/h and the slowest is typically 1 mm/h.

As well as recording signals as a continuous trace, many models also allow for the printing of alphanumeric data on the chart to record date, time, and other process information. Some models also have a digital numeric display to provide information on the current values of recorded variables.

Figure 9.18
Honeywell DPR100 strip chart recorder. *Reproduced by kind permission of Honeywell International, Inc.*

Multipoint strip chart recorder

The multipoint strip chart recorder is a modification of the pen strip chart recorder that uses a dot matrix print head striking against an ink ribbon instead of pens. A typical model might allow up to 24 different signal inputs to be recorded simultaneously, using a six-color ink ribbon. Certain models of such recorders also have the same enhancements as pen strip chart recorders in terms of printing alphanumeric information on the chart and providing a digital numeric output display.

Heated stylus chart recorder

The heated stylus chart recorder is another variant that records the input signal by applying a heated stylus to heat-sensitive chart paper. The main purpose of this alternative printing mechanism is to avoid the problem experienced in other forms of paper-based chart recorder of pen cartridges or printer ribbons running out of ink at inconvenient times.

Circular chart recorder

The circular chart recorder consists of a servo-driven pen assembly which records the measured signal on a rotating circular paper chart, as shown in Figure 9.19. The rotational speed of the chart can be typically adjusted between one revolution in 1 h and one revolution in 31 days. Recorded charts are replaced and stored after each revolution, which means replacement intervals that vary between hourly and monthly according to the chart speed. The major advantage of the circular chart recorder over other forms is compactness. Some models have up to four different colored pen assemblies, allowing up to four different parameters to be recorded simultaneously.

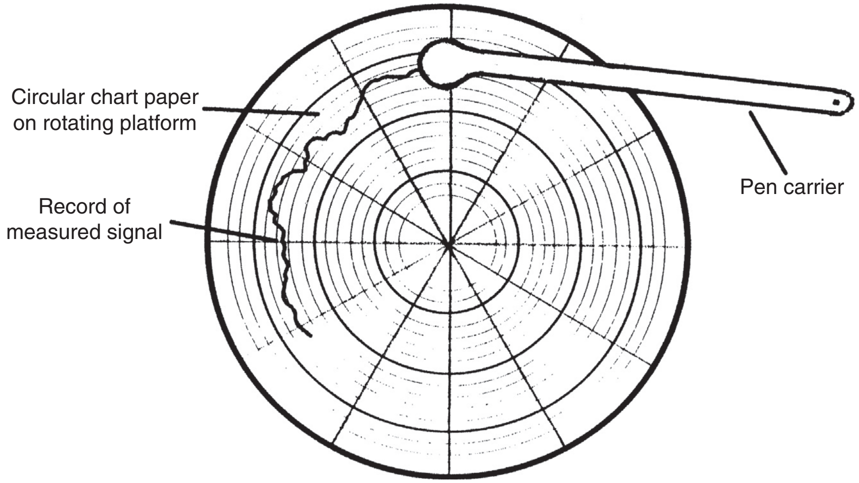

Figure 9.19
Circular chart recorder.

Paperless chart recorder

A paperless chart recorder, sometimes alternatively called a *virtual chart recorder* or a *digital chart recorder*, displays the time history of measured signals electronically, using a color-matrix liquid crystal display. This avoids the chore of periodically replacing chart paper and ink cartridges associated with other forms of chart recorder. Reliability is also enhanced compared with electromechanical recorders. As well as displaying the most recent time history of measured signals on its screen, the instrument also stores a much larger past history. This stored data can be recalled in batches and redisplayed on the screen as required. The only downside compared with other forms of chart recorder is this limitation of only displaying one screenful of information at a time. Of course, conventional recorders allow the whole past history of signals to be viewed at the same time on the hard-copy, paper recordings. Otherwise, specifications are very similar to other forms of chart recorder, with vertical motion of the screen display varying between 1 mm/h and 6000 mm/h, typical inaccuracy less than $\pm 0.1\%$ and capability of recording multiple signals simultaneously in different colors.

Videographic recorder

The videographic recorder provides exactly the same facilities as a paperless chart recorder but has additional display modes such as bar graphs (histograms) and digital numbers. However, it should be noted that the distinction is becoming blurred between the various forms of paperless recorder described earlier and videographic recorders as manufacturers enhance the facilities of their instruments. For historical reasons, many manufacturers retain the names that they have traditionally used for their recording instruments but there is now much overlap between their respective capabilities as the functions provided are extended.

9.3.2 Inkjet and Laser Printers

Standard computer output devices in the form of inkjet and laser printers are now widely used as an alternative means of storing measurement system output in paper form. Since a computer is a routine part of many data acquisition and processing operations, it often makes sense to output the data in a suitable form to a computer printer rather than a chart recorder. This saves the cost of a separate recorder and is facilitated by the ready availability of software that can output measurement data in a graphical format.

9.3.3 Other Recording Instruments

Many of the display devices mentioned earlier in this chapter also have facilities for storing measurement data digitally. These include data logging acquisition devices and digital storage oscilloscopes. This data can then be converted into hard-copy form as required by transferring it to either a chart recorder or a computer and printer.

9.3.4 Digital Data Recorders

Digital data recorders, also known as *data loggers*, have already been introduced in Chapter 6 in the context of data acquisition. They provide a further alternative way of recording measurement data in digital format. Data so recorded can then be transferred at a future time either to a computer for further analysis, to any of the forms of measurement display device discussed in Section 9.2, or to one of the hard-copy output devices described earlier in this section.

Features contained within a data recorder/data logger obviously vary according to the particular manufacturer/model under discussion. However, most recorders have facilities to handle measurements in the form of both analog and digital signals. Common analog input signals allowed include DC voltages, DC currents, AC voltages, and AC currents. Digital inputs can usually be either in the form of data from digital measuring instruments or discrete data representing events such as switch closures or relay operations. Some models also provide alarm facilities to alert operators to abnormal conditions during data recording operations.

Many data recorders provide special input facilities that are optimized for particular kinds of measurement sensor such as accelerometers, thermocouples, thermistors, resistance thermometers, strain gauges (including strain gauge bridges) linear variable differential transformers and rotational differential transformers. Some instruments also have special facilities for dealing with inputs from less common devices like encoders, counters, timers, tachometers, and clocks. A few recorders also incorporate integral sensors when they are designed to measure a particular type of physical variable.

The quality of the data recorded by a digital recorder is a function of the cost of the instrument. Paying more usually means getting more memory to provide a greater data storage capacity, greater resolution in the analog to digital converter to give better recording accuracy, and faster data processing to allow a greater data sampling frequency.

9.4 Presentation of Data

The two formats available for presenting data on paper are tabular and graphical ones. The relative merits of these are compared below. In some circumstances, it is clearly best to use only one or other of these two alternatives alone. However, in many data collection exercises, part of the measurements and calculations are expressed in tabular form and part graphically, so making best use of the merits of each technique. Very similar arguments apply to the relative merits of graphical and tabular presentations if a computer screen is used for the presentation instead of paper.

9.4.1 Tabular Data Presentation

A tabular presentation allows data values to be recorded in a precise way that exactly maintains the accuracy to which the data values were measured. In other words, the data values are written down exactly as measured. Besides recording the raw data values as measured, tables often also contain further values calculated from the raw data. An example of a tabular data presentation is given in Table 9.1. This records the results of an experiment to determine the strain induced in a bar of material that is subjected to a range of stresses. Data were obtained by applying a sequence of forces to the end of the bar and using an extensometer to measure the change in length. Values of the stress and strain in the bar are calculated from these measurements and are also included in the table. The final row, which is of crucial importance in any tabular presentation, is the estimate of possible error in each calculated result.

Table 9.1: Sample tabular presentation of data

Table of Measured Applied Forces and Extensometer Readings and Calculations of Stress and Strain				
	Force Applied (kN)	Extensometer Reading (Divisions)	Stress (N/m^2)	Strain
	0	0	0	0
	2	4.0	15.5	19.8×10^{-5}
	4	5.8	31.0	28.6×10^{-5}
	6	7.4	46.5	36.6×10^{-5}
	8	9.0	62.0	44.4×10^{-5}
	10	10.6	77.5	52.4×10^{-5}
	12	12.2	93.0	60.2×10^{-5}
	14	13.7	108.5	67.6×10^{-5}
Possible error in measurements (%)	±0.2	±0.2	±1.5	$\pm 1.0 \times 10^{-5}$

A table of measurements and calculations should conform to several rules as illustrated in Table 9.1.

1. The table should have a title that explains what data are being presented within the table.
2. Each column of figures in the table should refer to the measurements or calculations associated with one quantity only.
3. Each column of figures should be headed by a title that identifies the data values contained in the column.
4. The units in which quantities in each column are measured should be stated at the top of the column.
5. All headings and columns should be separated by bold horizontal (and sometimes vertical) lines.

6. The errors associated with each data value quoted in the table should be given. The form shown in Table 9.1 is a suitable way to do this when the error level is the same for all data values in a particular column. However, if error levels vary, then it is preferable to write the error boundaries alongside each entry in the table.

9.4.2 Graphical Presentation of Data

Presentation of data in graphical form involves some compromise in the accuracy to which the data are recorded, as the exact values of measurements are lost. However, graphical presentation has important advantages over tabular presentation.

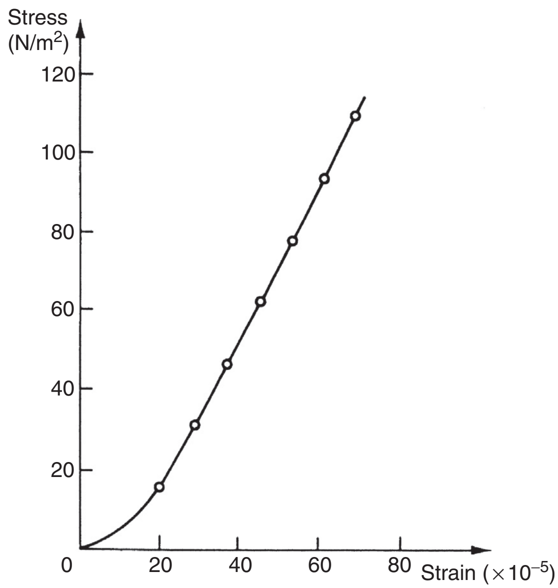

Figure 9.20
Sample graphical presentation of data: graph of stress against strain.

- Graphs provide a pictorial representation of results that is more readily comprehended than a set of tabular results.
- Graphs are particularly useful for expressing the quantitative significance of results and showing whether a linear relationship exists between two variables. Figure 9.20 shows a graph drawn from the stress and strain values given in the Table 9.1. Construction of the graph involves first of all marking the points corresponding to the stress and strain values. The next step is to draw some line through these data points that best represents the relationship between the two variables. This line will normally be either a straight one or a smooth curve. The data points will not usually lie exactly on this line but instead will lie on either side of it. The magnitude of the excursions of the data points

from the line drawn will depend on the magnitude of the random measurement errors associated with the data.

- Graphs can sometimes show up a data point that is clearly outside the straight line or curve that seems to fit the rest of the data points. Such a data point is probably due either to a human mistake in reading an instrument or else to a momentary malfunction in the measuring instrument itself. If the graph shows such a data point where a human mistake or instrument malfunction is suspected, the proper course of action is to repeat that particular measurement and then discard the original data point if the mistake or malfunction is confirmed.

Like tables, the proper representation of data in graphical form has to conform to certain rules:

1. The graph should have a title or caption that explains what data are being presented in the graph.
2. Both axes of the graph should be labeled to express clearly what variable is associated with each axis and to define the units in which the variables are expressed.
3. The number of points marked along each axis should be kept reasonably small—about five divisions is often a suitable number.
4. No attempt should be made to draw the graph outside the boundaries corresponding to the maximum and minimum data values measured, i.e., in Figure 9.20, the graph stops at a point corresponding to the highest measured stress value of 108.5.

Fitting curves to data points on a graph

The procedure of drawing a straight line or smooth curve as appropriate that passes close to all data points on a graph, rather than joining the data points by a jagged line that passes through each data point, is justified on account of the random errors that are known to affect measurements. Any line between the data points is mathematically acceptable as a graphical representation of the data if the maximum deviation of any data point from the line is within the boundaries of the identified level of possible measurement errors. However, within the range of possible lines that could be drawn, only one will be the optimum one. This optimum line is where the sum of negative errors in data points on one side of the line is balanced by the sum of positive errors in data points on the other side of the line. The nature of the data points is often such that a perfectly acceptable approximation to the optimum can be obtained by drawing a straight line through the data points by eye. In other cases, however, it is necessary to fit a line mathematically, using regression techniques.

Regression techniques

Regression techniques consist of finding a mathematical relationship between measurements of two variables y and x, such that the value of variable y can be predicted from a measurement of the other variable x. However, regression techniques should not be regarded as a magic formula that can fit a good relationship to measurement data in all circumstances, as the characteristics of the data must satisfy certain conditions. In determining the suitability of measurement data for the application of regression techniques, it is a recommended practice to draw an approximate graph of the measured data points, as this is often the best means of detecting aspects of the data that make it unsuitable for regression analysis. Drawing a graph of the data will indicate, for example, whether there are any data points that appear to be erroneous. This may indicate that human mistakes or instrument malfunctions have affected the erroneous data points, and it is assumed that any such data points will be checked for correctness.

Regression techniques cannot be successfully applied if the deviation of any particular data point from the line to be fitted is greater than the maximum possible error that is calculated for the measured variable (i.e., the predicted sum of all systematic and random errors). The nature of some measurement data sets is such that this criterion cannot be satisfied, and any attempt to apply regression techniques is doomed to failure. In that event, the only valid course of action is to express the measurements in tabular form. This can then be used as an x–y look-up table, from which values of the variable y corresponding to particular values of x can be read off. In many cases, this problem of large errors in some data points only becomes apparent during the process of attempting to fit a relationship by regression.

A further check that must be made before attempting to fit a line or curve to measurements of two variables x and y is to examine the data and look for any evidence that both variables are subject to random errors. It is a clear condition for the validity of regression techniques that only one of the measured variables is subject to random errors, with no error in the other variable. If random errors do exist in both measured variables, regression techniques cannot be applied and recourse must be made instead to correlation analysis (covered later in this chapter). A simple example of a situation where both variables in a measurement data set are subject to random errors are measurements of human height and weight, and no attempt should be made to fit a relationship between them by regression.

Having determined that the technique is valid, the regression procedure is simplest if a straight-line relationship exists between the variables, which allows a relationship of the form $y = a + bx$ to be estimated by linear least squares regression. Unfortunately, in many cases, a straight-line relationship between the points does not exist, which is readily shown by plotting the raw data points on a graph. However, knowledge of physical laws governing the data can often suggest a suitable alternative form of relationship between the two sets of variable measurements, such as a quadratic relationship or a higher order polynomial relationship. Also, in some cases, the measured variables can be transformed into a form where a linear relationship exists. For example, suppose that two variables y and x are related according to: $y = ax^c$. A linear relationship from this can be derived, using a logarithmic transformation, as: $\log(y) = \log(a) + c\log(x)$. Thus, if a graph is constructed of $\log(y)$ plotted against $\log(x)$, the parameters of a straight-line relationship can be estimated by linear least squares regression.

In some cases, the background theory describing the relationship between two sets of measured variables is not known. The way to proceed in this case is to fit two relationships to the data, one using linear least squares regression and one using quadratic (second-order) least squares regression. A confidence test is then applied (see later section headed *confidence tests in curve-fitting*) to compare the two relationships fitted. If this shows that the linear relationship is best, the curve-fitting process ends. However, if the confidence test shows that a quadratic relationship is better than a straight line one, it is necessary to fit a third line to the data using a third-order relationship. A confidence test is then applied again, this time comparing the quadratic relationship to the third-order relationship. This continues until the next higher relationship tried fits the data less well than the previous relationship.

All quadratic and higher order relationships relating one variable y to another variable x can be represented by a power series of the form:

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_px^p$$

Estimation of the parameters $a_0 \cdots a_p$ is very difficult if p has a large value. Fortunately, a relationship where p only has a small value can be fitted to most data sets. Quadratic least squares regression is used to estimate parameters where p has a value of two, and for larger values of p, polynomial least squares regression is used for parameter estimation.

Where the appropriate form of relationship between variables in measurement data sets is not obvious either from visual inspection or from consideration of physical laws, a method that is effectively a trial and error one has to be applied. This consists of estimating the parameters of successively higher order relationships between y and x until a curve is found that fits the data sufficiently closely. What level of closeness is acceptable is considered in the later section on confidence tests.

Linear least squares regression

Linear least squares regression estimates the optimal values for the constant coefficients a and b of a linear relationship $y = a + bx$ for a set of n measurements $y_1 \cdots y_n$, $x_1 \cdots x_n$, such that the relationship gives the best fit to the measurement data. Linear regression is applicable whenever the relationship between two sets of measurement variables is either known to be linear or is suspected to be so. For example, the deflection of a spring is linearly proportional to the force applied according to the known background theory for a spring (as long as the magnitude of the force applied is not sufficient to cause the elastic limit for the spring to be exceeded).

Typical data for a set of output measurements $y_1 \cdots y_n$ for a corresponding set of inputs $x_1 \cdots x_n$ subject to random errors is shown in the graph in Figure 9.21. In the absence of errors, all data points would fall exactly on a straight line. However, because of measurement errors, the actual data points fall above and below the straight line expected according to the theoretical relationship between the data points. The deviation of each point (x_i, y_i) from the line can be expressed as d_i, where $d_i = y_i - (a + bx_i)$.

The best-fit line is obtained when the sum of the squared deviations, S, is a minimum,

$$\text{i.e., when } S = \sum_{i=1}^{n} \left(d_i^2\right) = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \text{ is a minimum.}$$

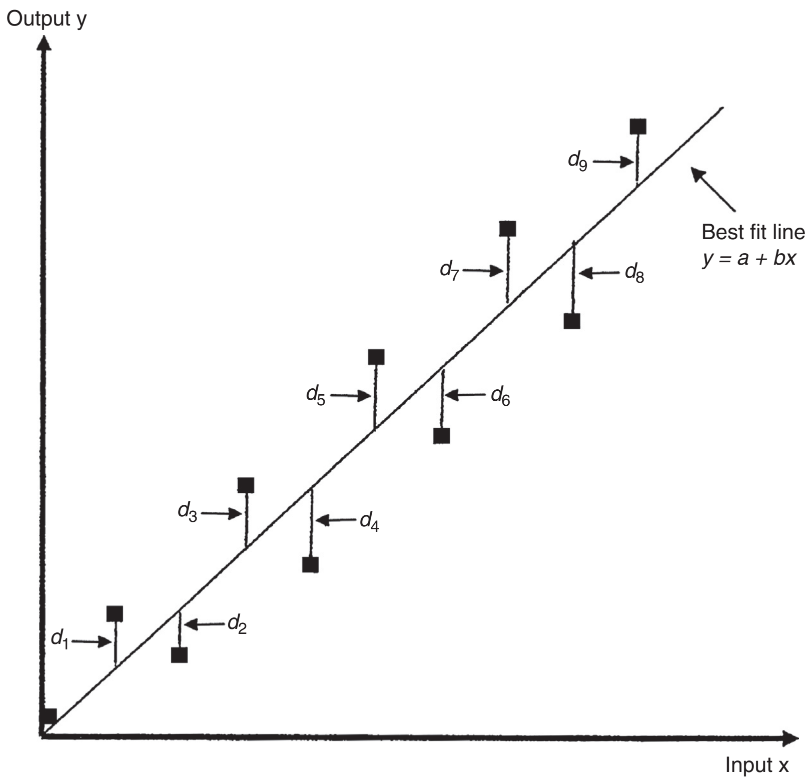

Figure 9.21
Deviations of data points away from best-fit straight line.

The minimum can be found by setting the partial derivatives $\partial S/\partial a$ and $\partial S/\partial b$ to zero and solving the resulting two simultaneous (normal) equations:

$$\partial S/\partial a = \sum 2\left(y_i - a - bx_i\right)(-1) = 0 \tag{9.5}$$

$$\partial S/\partial b = \sum 2\left(y_i - a - bx_i\right)(-x_i) = 0 \tag{9.6}$$

The values of the coefficients a and b at the minimum point can be represented by $\hat{a}$ and $\hat{b}$, which are known as the least squares estimates of a and b. These can be calculated as follows:

From (9.5),

$$\sum y_i = \sum \hat{a} + \hat{b} \sum x_i = n\hat{a} + \hat{b} \sum x_i$$

and thus,

$$\hat{a} = \frac{\sum y_i - \hat{b} \sum x_i}{n} \tag{9.7}$$

From (9.6),

$$\sum \left(x_i y_i\right) = \hat{a} \sum x_i + \hat{b} \sum x_i^2 \tag{9.8}$$

Now substitute for $\hat{a}$ in (9.8) using (9.7):

$$\sum \left(x_i y_i\right) = \frac{\left(\sum y_i - \hat{b} \sum x_i\right)}{n} \sum x_i + \hat{b} \sum x_i^2$$

Collecting terms in $\hat{b}$,

$$\hat{b}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right] = \sum \left(x_i y_i\right) - \frac{\sum x_i \sum y_i}{n}$$

Rearranging gives:

$$\hat{b}\left[\sum x_i^2 - n\left\{\left(\sum x_i/n\right)\right\}^2\right] = \sum \left(x_i y_i\right) - n \sum (x_i/n) \sum \left(y_i/n\right).$$

This can be expressed as:

$$\hat{b}\left[\sum x_i^2 - n x_m^2\right] = \sum \left(x_i y_i\right) - n x_m y_m,$$

where x_m and y_m are the mean values of x and y.

Thus:

$$\hat{b} = \frac{\sum(x_i y_i) - n x_m y_m}{\sum x_i^2 - n x_m^2} \tag{9.9}$$

And, from (9.7):

$$\hat{a} = \frac{\sum y_i - \hat{b}\sum x_i}{n} = y_m - \hat{b} x_m \tag{9.10}$$

Example 9.3

In an experiment to determine the characteristics of a displacement sensor with a voltage output, the following output voltage values were recorded when a set of standard displacements was measured:

Displacement (cm)	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Voltage (V)	2.1	4.3	6.2	8.5	10.7	12.6	14.5	16.3	18.3	21.2

Fit a straight line to this set of data using least squares regression and estimate the output voltage when a displacement of 4.5 cm is measured.

Solution

Let y represent the output voltage and x represent the displacement. Then a suitable straight line is given by $y = a + bx$. We can now proceed to calculate estimates for the coefficients a and b using Eqns (9.9) and (9.10) above. The first step is to calculate the mean values of x and y. These are found to be $x_m = 5.5$ and $y_m = 11.47$. Next, we need to tabulate $x_i\, y_i$ and x_i^2 for each pair of data values:

x_i	y_i	$x_i\, y_i$	x_i^2
1.0	2.1	2.1	1
2.0	4.3	8.6	4
3.0	6.2	18.6	9
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
10.0	21.2	212.0	100

Now calculate the values needed from this table:

$n = 10$; $\sum(x_i y_i) = 801.0$; $\sum(x_i^2) = 385$ and enter these values into (9.9) and (9.10)

$$\hat{b} = \frac{801.0 - (10 \times 5.5 \times 11.47)}{385 - (10 \times 5.5^2)} = 2.067; \quad \hat{a} = 11.47 - (2.067 \times 5.5) = 0.1033;$$

i.e., $y = 0.1033 + 2.067x$

Hence, for $x = 4.5$, $y = 0.1033 + (2.067 \times 4.5) = 9.40$ V. Note that in this solution, we have only specified the answer to an accuracy of three figures, which is the same accuracy as the measurements. Any greater number of figures in the answer would be meaningless.

■

Least squares regression is often appropriate for situations where a straight line relationship is not immediately obvious, for example, where $y \propto x^2$ or $y \propto \exp(x)$.

■ Example 9.4

From theoretical considerations, it is known that the voltage (V) across a charged capacitor decays with time (t) according to the relationship: $V = K \exp(-t/\tau)$. Estimate values for K and τ if the following values of V and t are measured.

V	8.67	6.55	4.53	3.29	2.56	1.95	1.43	1.04	0.76
t	0	1	2	3	4	5	6	7	8

■

■ Solution

If $V = K \exp(-T/\tau)$ then, $\log_e(V) = \log_e(K) - t/\tau$. Now let $y = \log_e(V)$, $a = \log(K)$, $b = -1/\tau$ and $x = t$. Hence, $y = a + bx$, which is the equation of a straight line whose coefficients can be estimated by applying Eqns (9.9) and (9.10). Therefore, proceed in the same way as Example 9.3 and tabulate the values required:

V	$\log_e(V)$ (y_i)	t (x_i)	$(x_i y_i)$	(x_i^2)
8.67	2.16	0	0	0
6.55	1.88	1	1.88	1
4.53	1.51	2	3.02	4
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
0.76	−0.27	8	−2.16	64

Now calculate the values needed from this table:
$n = 9$; $\sum(x_i y_i) = 15.86$; $\sum(x_i^2) = 204$; $x_m = 4.0$; $y_m = 0.9422$, and enter these values into (9.9) and (9.10).

$$\hat{b} = \frac{15.86 - (9 \times 4.0 \times 0.9422)}{204 - (9 \times 4.0^2)} = -0.301; \;\; \hat{a} = 0.9422 + (0.301 \times 4.0) = 2.15$$

$$K = \exp(a) = \exp(2.15) = 8.58; \;\; \tau = -1/b = -1/(-0.301) = 3.32$$

■

Quadratic least squares regression

Quadratic least squares regression is used to estimate the parameters of a relationship $y = a + bx + cx^2$ between two sets of measurements $y_1 \cdots y_n$, $x_1 \cdots x_n$. It is applied when theory suggests that this is the appropriate form of relationship for the data.

The deviation of each point (x_i, y_i) from the line can be expressed as d_i, where $d_i = y_i - (a + bx_i + cx_i^2)$.

The best-fit line is obtained when the sum of the squared deviations, S, is a minimum,

$$\text{i.e., when } S = \sum_{i=1}^{n} \left(d_i^2\right) = \sum_{i=1}^{n} \left(y_i - a - bx_i + cx_i^2\right)^2 \text{ is a minimum.}$$

The minimum can be found by setting the partial derivatives $\partial S/\partial a$, $\partial S/\partial b$, and $\partial S/\partial c$ to zero and solving the resulting simultaneous equations, as for the linear least-squares regression case above. Standard computer packages to estimate the parameters a, b, and c by numerical methods are widely available (for example, Labview and Matlab) and therefore a detailed solution is not presented here.

Polynomial least squares regression

Polynomial least squares regression is used to estimate the parameters of the pth order relationship $y = a_0 + a_1 x + a_2 x^2 + \cdots + a_p x^p$ between two sets of measurements $y_1 \cdots y_n$, $x_1 \cdots x_n$.

The deviation of each point (x_i, y_i) from the line can be expressed as d_i, where:

$$d_i = y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p\right)$$

The best-fit line is obtained when the sum of the squared deviations given by $S = \sum_{i=1}^{n}(d_i^2)$ is a minimum.

The minimum can be found as before by setting the p partial derivatives $\partial S/\partial a_0 \cdots \partial S/\partial a_p$ to zero and solving the resulting simultaneous equations. Again, as for the quadratic least squares regression case, standard computer programs to estimate the parameters $a_0 \cdots a_p$ by numerical methods are widely available (for example, Labview and Matlab) and therefore a detailed solution is not presented here.

Confidence tests in curve fitting by least squares regression

Once data have been collected and a mathematical relationship that fits the data points have been determined by regression, the level of confidence that the mathematical relationship fitted is correct must be expressed in some way. The first check that must be made is whether the fundamental requirement for the validity of regression techniques is satisfied, i.e., whether the deviations of data points from the fitted line are all less than the maximum error level predicted for the measured variable. If this condition is violated by any data point that a line or curve has been fitted to, then use of the fitted relationship is unsafe and recourse must be made to tabular data presentation, as described earlier.

The second check concerns whether or not random errors affect both measured variables. If attempts are made to fit relationships by regression to data where both measured variables contain random errors, any relationship fitted will only be approximate and it is likely that one or more data points will have a deviation from the fitted line or curve that is greater than the maximum error level predicted for the measured variable. This will show up when the appropriate checks are made.

Having carried out the above checks to show that there are no aspects of the data, which suggest that regression analysis is not appropriate, the next step is to apply least squares regression to estimate the parameters of the chosen relationship (linear, quadratic, etc.). After this, some form of follow-up procedure is clearly required to assess how well the estimated relationship fits the data points. A simple curve-fitting confidence test is to calculate the sum of squared deviations S for the chosen y/x relationship and compare it with the value of S calculated for the next higher order regression curve that could be fitted to the data. Thus if a straight-line relationship is chosen, the value of S calculated should be of a similar magnitude or less than that obtained by fitting a quadratic relationship. If the value of S were substantially lower for a quadratic relationship, this would indicate that a quadratic relationship was a better fit to the data than a straight line one and further tests would be needed to examine whether a cubic or higher order relationship was a better fit still.

The stages of this simple confidence test can be summarized as follows:

1. Calculate $S = \sum (d_i)^2$ for the relationship fitted (for example, a linear relationship)
2. Calculate $S' = \sum (d_i)^2$ for the next higher order relationship, for example, a quadratic relationship
   If S' is significantly less than S, then the first relationship fitted (in this example the linear relationship) is incorrect.
3. Repeat steps (1) and (2) until S' is greater than or equal to S.

Curve fitting programs within packages such as Labview and Matlab are able to assist this procedure by calculating the value of S for each order of regression fit tried.

Other more sophisticated confidence tests exist such as the *F-ratio test*. However, these are outside the scope of this book.

Correlation tests

Where both variables in a measurement data set are subject to random fluctuations, correlation analysis is applied to determine the degree of association between the variables. For example, in the case already quoted of a data set containing measurements of human height and weight, we certainly expect some relationship between the variables of height and weight because a tall person is heavier *on average* than a short person. Correlation tests determine the strength of the relationship (or interdependence) between the measured variables, which is expressed in the form of a correlation coefficient.

For two sets of measurements $y_1 \cdots y_n$, $x_1 \cdots x_n$ with means x_m and y_m, the correlation coefficient Φ is given by:

$$\Phi = \frac{\sum (x_i - x_m)(y_i - y_m)}{\sqrt{\left[\sum (x_i - x_m)^2\right]\left[\sum (y_i - y_m)^2\right]}} \tag{9.11}$$

The value of $|\Phi|$ always lies between 0 and 1, with 0 representing the case where the variables are completely independent of one another and 1 being the case where they are totally related to one another. Because of the presence of measurement errors in almost all situations, it is very unlikely that value of $|\Phi|$ will ever equal 1, which would indicate that the data fitted a straight line exactly. Hence, how close to 1 does $|\Phi|$ have to be for there to be reasonable confidence that the data fits a straight line relationship? The usual criterion applied is to regard any value given by $\pm 0.9 \le |\Phi| \le 1.0$ as indicating that a straight line exists between the data.

For $0 < |\Phi| < 1$, linear least squares regression can be applied to find a relationship between the variables, which allows x to be predicted from a measurement of y, and y to be predicted from a measurement of x. This involves finding two separate regression lines of the form:

$$y = a + bx \text{ and } x = c + dy$$

These two lines are not normally coincident as shown in Figure 9.22. Both lines pass through the centroid of the data points but their slopes are different.

As $|\Phi| \to 1$, the lines tend to coincidence, representing the case where the two variables are totally dependent upon one another.

As $|\Phi| \to 0$, the lines tend to orthogonal ones parallel to the x and y axes. In this case, the two sets of variables are totally independent. The best estimate of x given any measurement of y is x_m and the best estimate of y given any measurement of x is y_m.

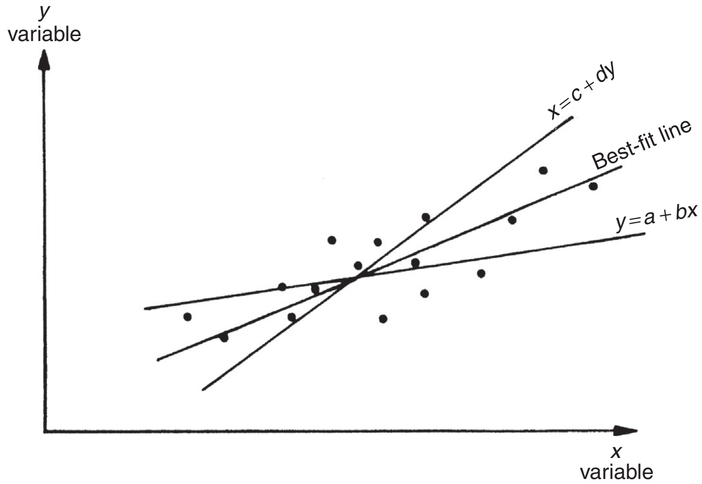

Figure 9.22
Relationship between two variables with random fluctuations.

For the general case, the best fit to the data is the line that bisects the angle between the lines on Figure 9.22.

Example 9.5

The height and weight of 33 students is measured and recorded as follows:

y (weight in kg)	70.3	75.2	88.1	75.0	72.8	80.7	81.3	76.6	73.8	85.2	81.2
x (height in meters)	1.69	1.68	1.82	1.69	1.81	1.76	1.84	1.74	1.70	1.81	1.75
y (weight in kg)	77.3	77.9	75.3	84.7	68.5	70.7	74.4	73.0	69.4	78.8	83.5
x (height in meters)	1.80	1.71	1.74	1.77	1.72	1.67	1.75	1.66	1.75	1.78	1.80
y (weight in kg)	75.5	64.2	77.2	76.1	90.6	66.6	79.5	64.1	65.8	69.4	73.5
x (height in meters)	1.76	1.68	1.75	1.79	1.85	1.70	1.83	1.76	1.64	1.82	1.73

Use linear least squares regression to find the optimal values for a, b, c, d in relationships between the variables of the form $y = a + bx$ and $x = c + dy$. Draw both of these relationships on a graph. Using these two plotted relationships as a guide, draw a single line that represents the best straight-line fit to the height/weight data given and calculate the gradient of this line. Also, calculate the correlation coefficient Φ.

Solution

Equations (9.9) and (9.10) provide expressions for estimating the optimal values of a and b for a relationship $y = a + bx$ between the data.

$$\hat{b} = \frac{\sum(x_i y_i) - n x_m y_m}{\sum x_i^2 - n x_m^2} \quad \text{and} \quad \hat{a} = \frac{\sum y_i - \hat{b}\sum x_i}{n} = y_m - \hat{b}x_m$$

Similar expressions for estimating the optimal values of c and d for a relationship $x = c + dy$ between the data can be found by an identical procedure of minimizing the sum of squared deviations as used earlier for calculating a and b (however, these expressions are found most simply by just interchanging x and y in Eqns (9.9) and (9.10)):

$$\hat{d} = \frac{\sum(x_i y_i) - n x_m y_m}{\sum y_i^2 - n y_m^2} \quad \text{and} \quad \hat{c} = \frac{\sum x_i - \hat{d}\sum x_i}{n} = x_m - \hat{d}y_m$$

It would be extremely tedious to calculate a, b, c, and d by hand for 33 data pairs, and so it is usual to write a computer program or use a spreadsheet package to evaluate the expressions for a, b, c, and d. The author used Microsoft Excel and obtained the following values:

$$\hat{a} = -53.23; \quad \hat{b} = 73.64; \quad \hat{c} = 1.3433; \quad \hat{d} = 0.005377$$

Thus, the best-fit lines are

$$y = -53.23 + 73.64x \tag{9.12}$$

and

$$x = 1.3433 + 0.005377y \tag{9.13}$$

Data points for these two best-fit lines can be calculated as follows:

Using (9.12), for $x = 1.67$, $y = 69.8$ and for $x = 1.85$, $y = 83.0$.

Using (9.13), for $y = 64$, $x = 1.687$ and for $y = 92$, $x = 1.838$.

These two relationships have been plotted in Figure 9.23.

A line that bisects these two relationships can then be estimated by eye as shown by the dashed line in Figure 9.23. This represents the best straight line fit to the height/weight data given.The gradient of this line can be calculated by selecting two data points on it.

At $y = 67.0$, $x = 1.680$; at $y = 86.0$, $x = 1.834$.

Hence, the gradient of the line is. $y/x = (86.0 - 67.0)/(1.834 - 1.680) = 123.4$ kg/m.

The equation for calculating Φ is given in Eqn (9.11). Using a computer program or a spreadsheet package to evaluate this gives $\Phi = 0.629$.

■

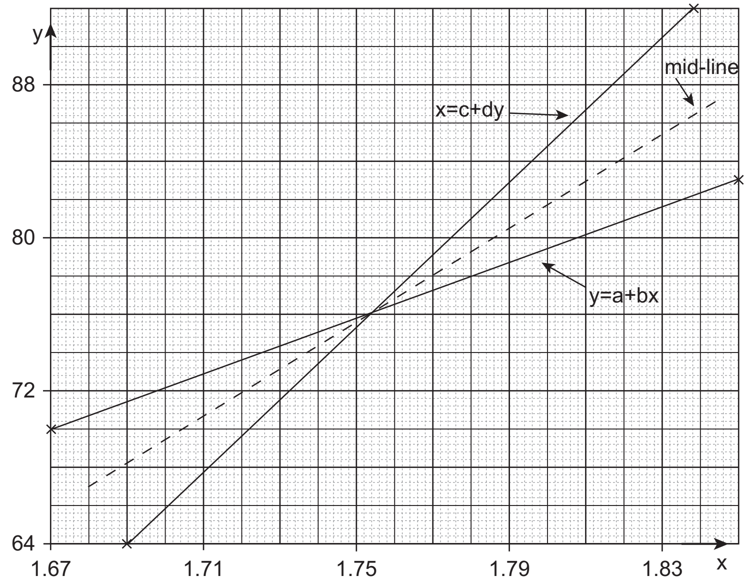

Figure 9.23
Figure showing solution of worked Example 9.5.

The noncoincidence of the lines for $y = a + bx$ and $x = c + dy$ calculated in Example 9.5 is explained by the value of 0.629 calculated for Φ. We have previously commented that a value of at least 0.9 is expected for Φ if a linear relationship exists between the data.

9.5 Summary

This chapter has covered what is essentially the final stage in a measurement system that of using the data collected. Such use might mean immediate use by displaying the data on a display device. Alternatively, the requirement may be to store the data for future use and analysis.

The chapter started with a discussion of the techniques available to display measurement data for immediate use. Within this discussion, we looked at the merits and usage of various types of electrical meter and also of oscilloscopes. We then went on to look at two newer ways of displaying measurement data, using an electronic display device and using a computer monitor.

Following this, we looked at the various means available to record and store measurement data for future use. We learned that the options available for recording data are numerous and include chart recorders, digital oscilloscopes, digital data recorders, and hard-copy devices such as inkjet and laser printers. We gave consideration to each of these and indicated some of the circumstances in which each alternative recording device might be used.

The chapter then ended with a discussion about standards of good practice in presenting data. This included coverage of both graphical and tabular forms of presentation, using either paper or a computer monitor screen as the display medium. We then went on to consider the best way of fitting lines through data points on a graph.This led us to a discussion of mathematical regression techniques and the associated confidence tests that are necessary to assess the correctness of the line fitted using regression. Finally, correlation tests were described that determine the degree of association between two sets of data when they are both subject to random fluctuations.

9.6 Problems

9.1 Summarize the advantages of digital meters over their analog counterparts.

9.2 Explain the four main alternative mechanisms used for effecting analog-to-digital conversion in a digital voltmeter and discuss their relative advantages.

9.3 What sort of applications are analog meters still commonly found in?

9.4 (a) Explain the mode of operation of a moving coil meter.
(b) Calculate the reading that would be observed on a moving coil ammeter when the current for the waveform shown in Figure 9.24 is being measured. (The waveform between $\omega t = \pi$ and $\omega t = 2\pi$ is the positive half of a sine wave with a peak current of +20 A.)

9.5. (a) Explain the mode of operation of a moving iron meter.
(b) Calculate the reading that would be observed on a moving iron ammeter when the current for the waveform shown in Figure 9.24 is being measured. (The waveform between $\omega t = \pi$ and $\omega t = 2\pi$ is the positive half of a sine wave with a peak current of +20 A.)

9.6 Calculate the reading that would be observed on (a) a moving coil ammeter and (b) a moving iron ammeter when the current for the waveform shown in Figure 9.25 is

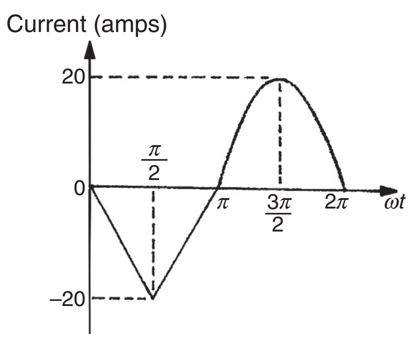

Figure 9.24
Waveform for problems 9.4 and 9.5.

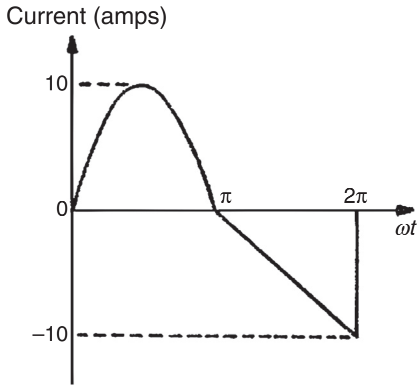

Figure 9.25
Waveform for problem 9.6.

being measured. (The waveform between $\omega t = 0$ and $\omega t = \pi$ is the positive half of a sine wave with a peak current of +10 A.)

9.7 Explain the following terms: (a) bandwidth, (b) rise time. In designing oscilloscopes, what relationship is sort between bandwidth and rise time?

9.8 Explain the following terms in relation to an oscilloscope: (a) channel, (b) single-ended input, (c) differential input, (d) vertical sensitivity, and (e) display position control.

9.9 Sketch a block diagram showing the main components in a digital storage oscilloscope and explain the mode of operation of the instrument.

9.10 Draw a block diagram showing the main components in a digital phosphor oscilloscope and explain the mode of operation of the instrument. What advantages does a digital phosphor oscilloscope have over a digital storage one?

9.11 Illustrate the main components in a digital sampling oscilloscope by sketching a block diagram of them and explain the mode of operation of the instrument. What advantages and disadvantages does a digital sampling oscilloscope compared with a digital storage one?

9.12 What is a PC-based oscilloscope? Discuss its advantages and disadvantages compared with digital oscilloscopes.

9.13 What are the main differences between a digital storage oscilloscope, a digital phosphor oscilloscope, and a digital sampling oscilloscope? How do these differences affect their performance and typical usage?

9.14 What are the main ways available for displaying parameter values to human operators responsible for controlling industrial manufacturing systems? (Discussion on electronic displays and computer monitors is expected.)

9.15 Discuss the range of instruments and techniques available for recording measurement signals, mentioning particularly the frequency response characteristics of each instrument or technique and the upper frequency limit for signals in each case.

9.16 Discuss the features of the main types of chart recorder that are available for recording measurement signals.

9.17 What is a digital data recorder and what recording features does it typically provide?

9.18 Discuss the relative merits of tabular and graphical methods of recording measurement data.

9.19 What would you regard as good practice in recording measurement data in graphical form?

9.20 What would you regard as good practice in recording measurement data in tabular form?

9.21 Explain the technique of linear least squares regression for finding a relationship between two sets of measurement data.

9.22 (a) Show that the optimal coefficients $\hat{a}$ and $\hat{b}$ for a linear relationship $y = a + bx$ between a set of n measurements $y_1 \cdots y_n$ and $x_1 \cdots x_n$ are given by:

$$\hat{b} = \frac{\sum(x_i y_i) - n x_m y_m}{\sum x_i^2 - n x_m^2} \quad \text{and} \quad \hat{a} = y_m - \hat{b} x_m$$

where x_m and y_m are the mean values of x and y.

(b) The following set of measurements is obtained for values of an output variable y and an input variable x that are believed to be related by a linear expression of the form: $y = a + bx$.

i	1	2	3	4	5	6	7	8	9	10
x_i	40	45	50	55	60	65	70	75	80	85
y_i	13.8	22.7	33.6	42.9	53.5	63.3	75.0	82.8	94.1	103.4

Apply linear least squares regression to find the values of a and b that produce the best-fit straight line.

9.23 (a) What is linear least squares regression and in what situations is it useful?

(b) If a set of n measurements $y_1 \cdots y_n$ and $x_1 \cdots x_n$ are believed to be related by a linear relationship given by $y = a + bx$, show that the optimal coefficients $\hat{a}$ and $\hat{b}$ between are given by:

$$\hat{b} = \frac{\sum(x_i y_i) - n x_m y_m}{\sum x_i^2 - n x_m^2} \quad \text{and} \quad \hat{a} = y_m - \hat{b} x_m$$

where x_m and y_m are the mean values of x and y.

(c) An experiment is carried out to determine the characteristic of a spring. A set of weights are added to the spring and the corresponding deflections are measured. The following set of measurements is obtained for the values of the output deflection y and the input weight x. It is known from the background theory for a spring that the deflection y is related to the weight x according to the linear expression: $y = a + bx$.

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i (gram)	5	10	15	20	25	30	35	40	45	50	55	60
y_i (mm)	11.4	21.9	32.3	44.8	54.2	66.7	75.7	89.1	98.5	111.0	120.6	133.4

Apply linear least squares regression to find the values of a and b that produce the best-fit straight line.

9.24 Explain the techniques of (a) quadratic least squares regression and (b) polynomial least squares regression. How would you determine whether either quadratic or polynomial least squares regression provides a better fit to a set of measurement data than linear least squares regression?

9.25 During calibration of a platinum resistance thermometer, the following temperature and resistance values were measured:

Resistance (Ω)	212.8	218.6	225.3	233.6	240.8	246.6
Temperature (°C)	300	320	340	360	380	400

The temperature measurements were made using a primary reference standard instrument for which the measurement errors can be assumed to be zero. The resistance measurements were subject to random errors but it can be assumed that are no systematic errors in them.

(a) Determine the sensitivity of measurement in Ω/°C in as accurate a manner as possible.

(b) Write down the temperature range that this sensitivity value is valid for.

(c) Explain the steps that you would take to test the validity of the type of mathematical relationship that you have used for the data.

9.26 Theoretical considerations show that quantities x and y are related in a linear fashion such that:

$y = ax + b$. Show that the best estimate of the constants a and b are given by:

$$\hat{a} = \frac{\sum(x_i y_i) - n x_m y_m}{\sum x_i^2 - n x_m^2}; \quad \hat{b} = y_m - \hat{a} x_m$$

Explain carefully the meaning of all the terms in the above two equations.

9.27 The characteristics of a chromel–constantan thermocouple is known to be approximately linear over the temperature range of 300 °C–800 °C. The output emf was measured practically at a range of temperatures and the following table of results obtained. Using least-squares regression, calculate the coefficients a and b for the relationship $T = a + bE$ that best describes the temperature–emf characteristic.

Temp (°C)	300	325	350	375	400	425	450	475	500	525	550
emf (mV)	21.0	23.2	25.0	26.9	28.6	31.3	32.8	35.0	37.2	38.5	40.7

Temp (°C)	575	600	625	650	675	700	725	750	775	800
emf (mV)	43.0	45.2	47.6	49.5	51.1	53.0	55.5	57.2	59.0	61.0

9.28 Measurements of the current (I) flowing through a resistor and the corresponding voltage drop (V) are shown below:

I (amps)	1	2	3	4	5
V (volts)	10.8	20.4	30.7	40.5	50.0

The instruments used to measure voltage and current were accurate in all respects except that they each had a zero error that the observer failed to take account of or to correct at the time of measurement. Determine the value of the resistor from the data measured.

9.29 A measured quantity y is known from theoretical considerations to depend on a variable according to the relationship: $y = a + bx^2$. For the following set of measurements of x and y, use linear least squares regression to determine the estimates of the parameters a and b that fit the data best.

x	0	1	2	3	4	5
y	0.9	9.2	33.4	72.5	130.1	200.8

9.30 The mean-time-to-failure ($MTTF$) of an integrated circuit is known to obey a law of the following form: $MTTF = C\exp T_0/T$, where T is the operating temperature and C and T_o are constants.
The following values of $MTTF$ at various temperatures were obtained from accelerated-life tests.

$MTTF$ (h)	54	105	206	411	941	2145
Temperature (°K)	600	580	560	540	520	500

(a) Estimate the values of C and T_o. (Hint:$\log_e(MTTF) = \log_e(C) + T_0/T$. This equation is now a straight-line relationship between log($MTTF$) and $1/T$, where log(C) and T_o are constants.)
(b) For a $MTTF$of 10 years, calculate the maximum allowable temperature.

9.31 The height and weight of 15 men is measured and recorded as follows, where y is the weight in kg and x is the height in meters of each man:

y	70.3	76.1	73.5	88.1	72.8	66.5	75.5	80.7	88.7	76.6	79.7	75.3	68.5	69.4	78.8
x	1.69	1.79	1.73	1.82	1.81	1.68	1.76	1.76	1.84	1.74	1.71	1.74	1.72	1.75	1.78

Use linear least squares regression to find the optimal values for a, b, c, d in relationships between the variables of the form $y = a + bx$ and $x = c + dy$. Draw both of these relationships on a graph. Using these two plotted relationships as a guide, draw a single line that represents the best straight-line fit to the height/weight data given and calculate the gradient of this line. Also, calculate the correlation coefficient Φ.

9.32 The height and weight of 15 male students is measured and recorded as follows, where y is the weight in lbs and x is the height in inches of each student:

y	68.0	69.5	68.5	72.5	69.0	70.5	67.5	70.0	68.0	71.5	66.0	70.0	71.5	66.5	69.5
x	151	166	172	198	153	168	180	174	162	181	147	178	161	155	176

Use linear least squares regression to find the optimal values for a, b, c, d in relationships between the variables of the form $y = a + bx$ and $x = c + dy$. Draw both of these relationships on a graph. Using these two plotted relationships as a guide, draw a single line that represents the best straight-line fit to the height/weight data given and calculate the gradient of this line. Also, calculate the correlation coefficient Φ.