Chapter 4

Velocity and position transducers

Within a closed-loop control system, feedback is used to minimise the difference between the demanded and actual output. In a motion-control system, the controlled variable is either the velocity or the position. The overall performance of a motion-control system will depend, to a large extent, on the type and quality of the transducer which is used to generate the feedback signal. It should be noted that velocity- or position-measuring transducers need not be used; other process variables (for example, the temperature and the chemical composition) can be used to determine the speed or position of a drive within a manufacturing process. However, as this book is concerned with robotic and machine-tool applications, the primary concentration will be on velocity and position transducers. In order to appreciate the benefits and limitations of the available systems, the performance of measurement systems in general must be considered.

4.1 The performance of measurement systems

The performance of a measurement system is dependent on both the static and dynamic characteristics of the transducers selected. In the case of motion-control systems where the measured quantities are rapidly changing, the dynamic relationships between the input and the output of the measurement system have to be considered, particularly when discrete sampling is involved. In contrast, the measured parameter may change only slowly in some applications; hence the static performance only needs to be considered during the selection process. The key characteristics of a transducer are as follows.

- **Accuracy** is a measure of how the output of the transducer relates to the true value at the input. In any specification of accuracy, the value needs to be qualified by a statement of which errors are being considered and the conditions under which they occur.

- **Dead band** is the largest change in input to which the transducer will fail to respond; this is normally caused by mechanical effects such as friction,

backlash, or hysteresis.

- **Drift** is the variation in the transducer's output which is not caused by a change in the input; typically, it is caused by thermal effects on the transducer or on its conditioning system.

- **Linearity** is a measure of the consistency of the input/output ratio over the useful range of the transducer.

- **Repeatability** is a measure of the closeness with which a group of output values agree for a constant input, under a given set of environmental conditions.

- **Resolution** is the smallest change in the input that can be detected with certainty by the transducer.

- **Sensitivity** is the ratio of the change in the output to a given change in the input. This is sometimes referred to as the gain or the scale factor.

A clear understanding is required of the interaction between accuracy, repeatability and resolution as applied to a measurement system. It is possible to have measurement systems with either high or low accuracy and repeatability; the measurements compared to the target position are shown in Figure 4.1. A motor drive system needs to incorporate a position measurement system with both high accuracy and repeatability to ensure that the target point is measured. If the system has low resolution, Figure 4.2, the uncertainty regarding each measure point increases.

All measurement systems suffer from inherent inaccuracies; and estimation of the uncertainty requires knowledge of the form that the error takes. In general, an error can be classified either as a random or a systematic error. Random errors arise from chance or random causes, and they must be considered using statistical methods. Systematic errors are errors which shift all the readings in one direction; for example, a shift in the zero point will cause all the readings to acquire a constant displacement from the true value.

4.1.1   Random errors

If a large set of data is taken from a transducer under identical conditions, and if the errors generated by the measurement system are random, the distribution of values about the mean will be Gaussian, Figure 4.3. In this form of distribution, sixty eight per cent of the readings lie within ± 1 standard deviation of the mean and ninety five per cent lie within ± 2 standard deviations. In general, if a sample of n readings are taken with values x_1, x_2 ... x_n, the mean $\bar{x}$ is given by

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} n_i \qquad (4.1)$$

and the standard deviation, s, by

(a) Low repeatability and low accuracy.

(b) High repeatability and low accuracy.

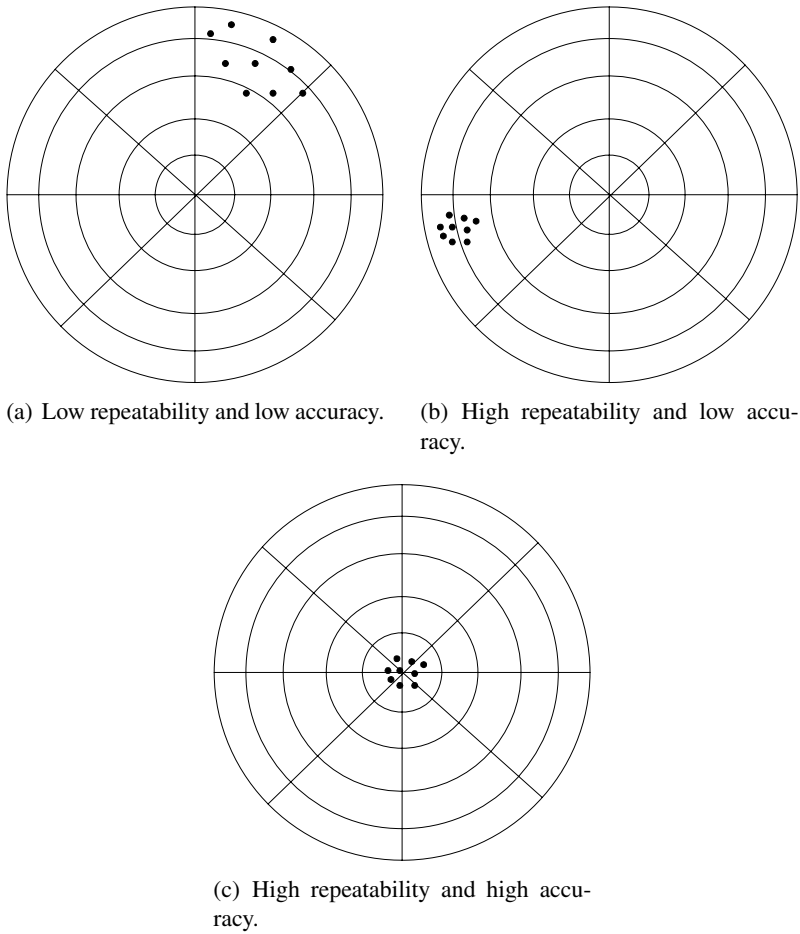(c) High repeatability and high accuracy.

Figure 4.1. Effect of accuracy and repeatability on the performance of a measurement system. The dots represent the individual measurements. Only when the system has both high accuracy and repeatability can the measurement error with respect to the target point be minimised.
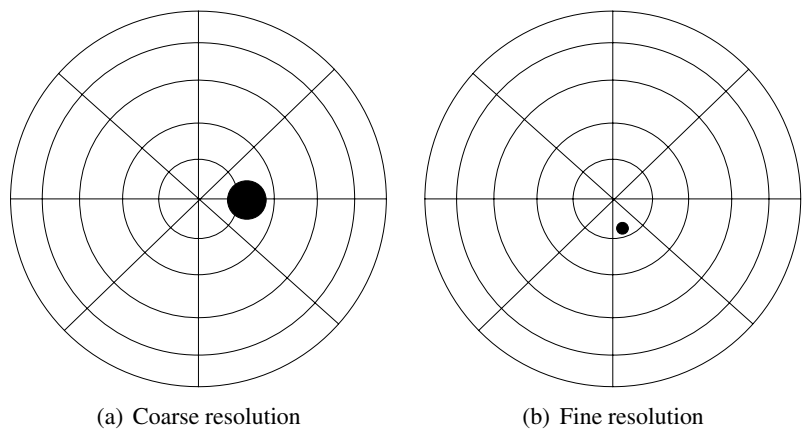
(a) Coarse resolution   (b) Fine resolution

Figure 4.2. Effect of resolution on the performance of a measurement system: the coarser the resolution (i.e. area of the dot), the more uncertainty there is in the measurement.
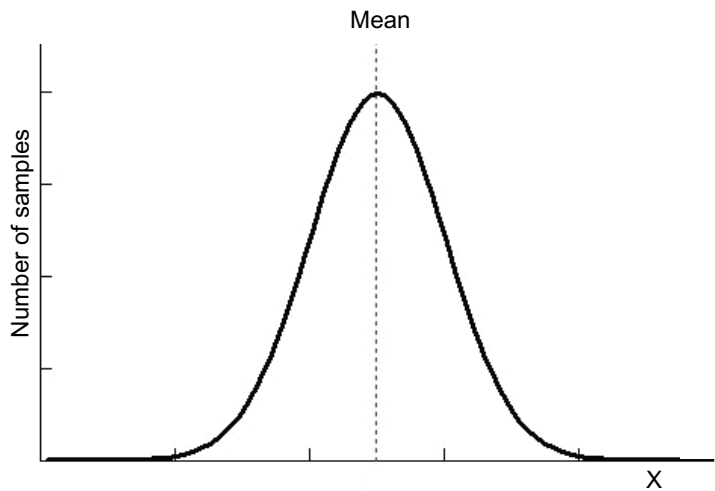
Figure 4.3. A Gaussian data distribution.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}} \tag{4.2}$$

The mean value which is obtained is dependent on the number of samples taken and on the spread; the true mean value can never be determined since this would require an infinite number of samples. However, by the use of the standard error of the mean, s_m, the probability of how close the mean of a set of data is to the true mean of the system can be evaluated. The standard error is given by

$$s_m = \frac{s}{\sqrt{n - 1}} \tag{4.3}$$

It is possible, using probability theory, to state that with a Gaussian distribution the probability of an individual reading, x_i, being within $\pm s_m$ of the true value is sixty eight per cent and that the probability of being within $\pm 2s_m$ is ninety five per cent.

4.1.2 Systematic errors

It can be seen from equation (4.3) that by taking a large number of samples, the random errors can be reduced to a very low value. However, when a systematic error occurs all the measurements are shifted in one direction by an equal amount. Figure 4.4 shows the spread of readings caused by both types of errors. The terms *accurate* and *precise* are used to cover both these situations; a measurement is accurate if the systematic error is small, and it is precise if the random error is small. A prime example of a systematic error is a zero offset, that is, when a instrument or a measured value does not return to zero when the parameter being measured is zero. This can be introduced by the transducer itself, or, more probably, by any conditioning electronics being used. Systematic errors are cumulative, so if a measurement, M, is a function of x, y, z, such that

$$M = f(x, y, z) \tag{4.4}$$

then the maximum value of the systematic error, ΔM, will be

$$\Delta M = \delta x^2 + \delta y^2 + \delta z^2 \tag{4.5}$$

where δx, δy and δz are the respective errors in x, y, and z. However, this approach can be considered to be rather pessimistic, because the systematic errors may not all operate in the same direction, and therefore they can either increase or decrease the reading. It is useful, therefore, to quote the systematic error in the form

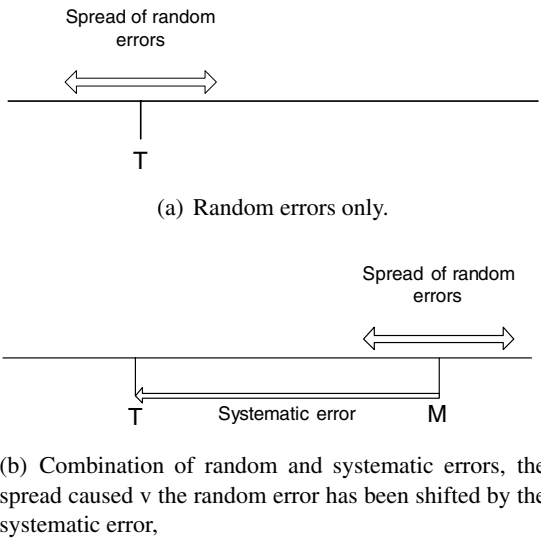$$\Delta M = \sqrt{\delta x^2 + \delta y^2 + \delta z^2} \tag{4.6}$$

(a) Random errors only.

(b) Combination of random and systematic errors, the
spread caused v the random error has been shifted by the
systematic error,

Figure 4.4. The effects of systematic and random errors on measurements where
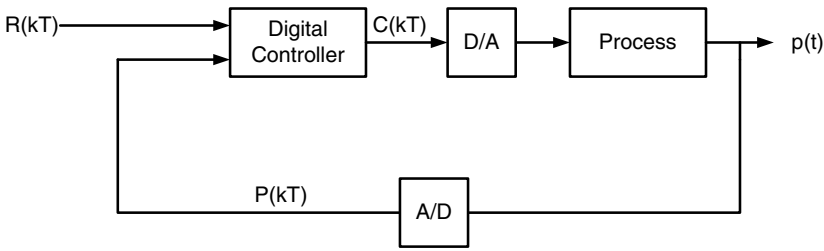T is the true value and M is the mean value of the data.

Figure 4.5. A block diagram of a digital-control system, showing the location of
the analogue to digital (A/D) and the digital to analogue (D/A) converters.

4.1.3   Digital-system errors

There is an increasing reliance on digital-control techniques in drive systems. Dig-
ital controllers require the transducer's output to be sampled and digitised. The
actual process of sampling will introduce a number of errors of its own. Consider
Figure 4.5, where a reference signal, $R(kT)$, a feedback signal, $P(kT)$, and the
resultant computed value, $C(kT)$, are discrete signals, in contrast to the output,
$p(t)$, which is a continuous function of time. If the sampling period, T, is small
compared with the system's time constant, the system can be considered to be
continuous; however, if the sampling time is close to the system's time constant,
the effects of digital sampling must be considered. A more detailed discussion of
digital controllers is to be found in Section 10.1.1.

A sampler can be considered to be a switch that closes for a period of time
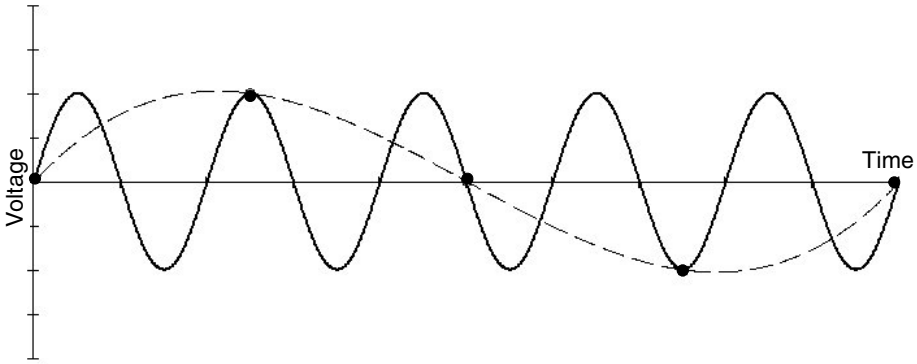every T seconds; with an ideal sampler for an input $p(t)$, the output will be

Figure 4.6. Aliasing caused by a sampling frequency. The sampling point are shown as dots, the sampling frequency is below frequency of the waveform being sampled. The reconstituted waveform is shown as the dotted line.

$$p^*(t) = p(nT)\delta(t - nT) \qquad (4.7)$$

where δ is the Dirac delta function (Nise, 1995). The input signal can be accurately followed if the sampling time is small compared to the rate of change of the signal; this ensures that the transients are not missed. In order to obtain an accurate picture of the signal being sampled, the sampling frequency must be selected with care. The sampling frequency is largely determined by the loop time of the control system; a high sample rate will place restrictions on the complexity of the algorithms that must be employed. If the highest frequency present in the signal to be sampled is f_p, then the minimum sampling rate is $2f_p$ as defined by Shannon's sampling theorem. The effect of a sampling frequency which is considerable less than the frequency of a signal is shown in Figure 4.6. It can be seen that the reconstituted signal is at a far lower frequency than the original waveform; this signal is referred to as the alias of the original signal. It is impossible to determine whether the sampled data is from the original signal or its alias. A frequently made mistake is the selection of a sampling rate at twice the frequency of interest, without considering the effect of noise, particularly interference from the mains supply. The solution, to this problem is to apply an anti-alias filter which blocks frequencies higher than those of interest.

4.1.4   Analogue-digital and digital-analogue conversion errors

Conversion of an analogue signal to a digital value involves a process of quantisation. In an analogue-to-digital (A/D) converter, the change from one state to the next will occur at a discrete point (the intermediate values are not considered, Figure 4.7). The difference between any two digital values is known as the quan-
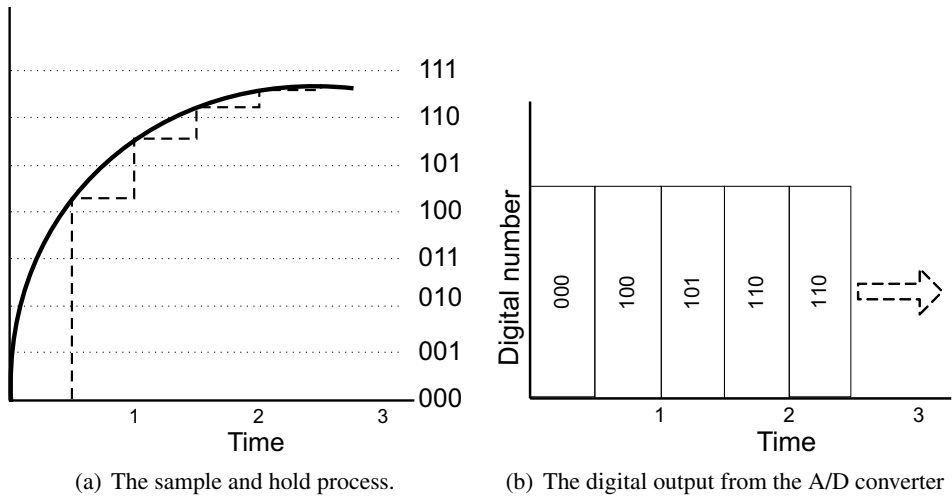
(a) The sample and hold process.   (b) The digital output from the A/D converter

Figure 4.7. The analogue to digital conversion process. The voltage being con-
verted is the solid line in (a), the input to the ADC is the dotted line, showing the
change of the sampled value.

tisation size, V_q, and it is commonly termed the resolution of the converter. For an
n-bit system the steps due to quantisation step V_q, and the subsequent error E_q are
equal to

$$V_q = \frac{\text{Full scale input}}{2^n} \tag{4.8}$$

$$E_q = \frac{1}{2}\frac{\text{Full scale input}}{2^n} = \frac{\text{Full scale input}}{2^{n+1}} \tag{4.9}$$

The resolution is equal to the input voltage, V_q, which will change the state of the
least-significant bit (LSB).

Transitions occur from one digital number to the next at integral multiples of
the LSB, giving a maximum uncertainty of one bit within the system. The resolu-
tion can only be decreased by increasing the number of bits within the converter.
A range of techniques are used for analogue to digital conversion, including high-
speed-flash (or parallel) converters, integrating, and successive-approximation con-
verters. It is not common to construct a discrete system; one of the commonly
available proprietary devices is usually used in the selection of a suitable device,
and consideration must be given to the device's conversion time, resolution, and
gain. A variant of the successive approximation converter is the tracking converter
that forms an integral part of a resolver's decoder; this is discussed later in this
chapter.

Digital-to-analogue (D/A) converters are used to provide analogue signals from
a digital systems. One of the problems with a D/A converter is that glitches occur
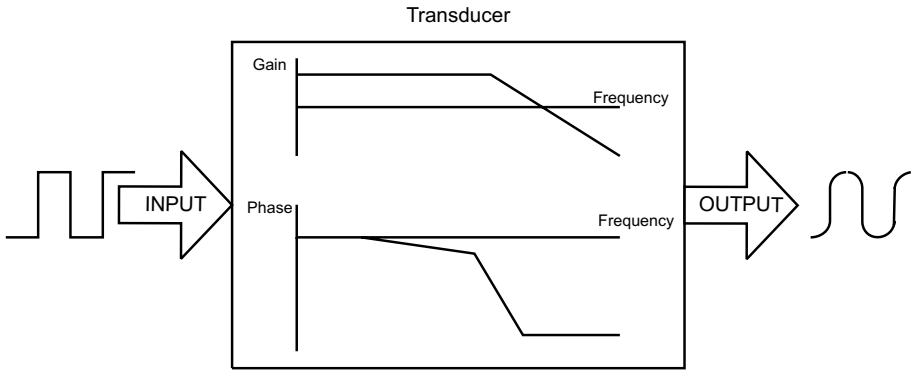as the digital signal (that is, the switches) change state, requiring a finite settling

Figure 4.8. The effects of a transducer's frequency-dependent gain and phase shift on an input signal.

time. As the code changes, the switches will not change state at the same instant; this is particularly acute when the code changes from, say 01111 to 10000, where the output for 11111 may transiently appear. It is possible to add a deglitching function to a D/A converter by increasing the transfer time of the converter.

4.1.5 Dynamic performance

Only the static characteristics of transducers have been considered up to this point. However, if the measured signal is rapidly changing, the dynamic performance of the measurement system has to be considered. A transducer with a linear characteristic will achieve a constant performance for all inputs; but this is not true in a practical system, since the input will have a non-linear distortion caused by the transducer's frequency-dependent gain and the phase shift, Figure 4.8. The formal analysis of these effects can be conducted, and represented, by a first-order, linear, differential equation. The dynamic performance needs to be considered in the selection of any transducer; even if the speed or position changes slowly, to ensure that any transient effects are considered. A limited bandwidth transducer will seriously limit the overall system bandwidth, and hence its ability to respond to transients (such as the application or removal of torques from the load).

4.2 Rotating velocity transducers

While the velocity can be determined from position measurement, a number of transducers are able to provide a dedicated output which is proportional to the velocity.
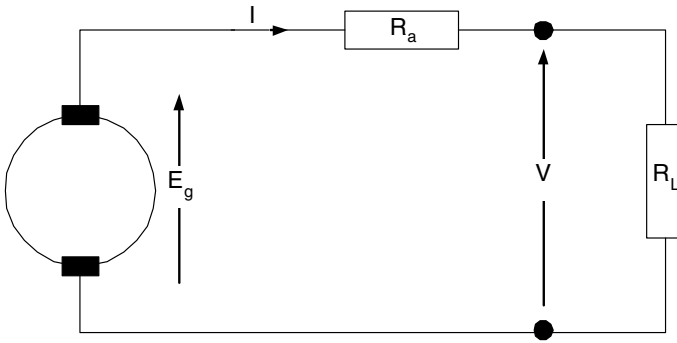
Figure 4.9. The equivalent circuit of a brushed tachogenerator.

4.2.1   Brushed d.c. tachogenerators

A brushed d.c. tachogenerator can be considered to be a precision d.c. generator, consisting of a permanent-magnet stator, with a wound armature. The output voltage, E_g, is related to the tachogenerator speed, N (rev min^{-1}), by the voltage constant, K_g (V rev^{-1} min)

$$E_g = K_g N \qquad (4.10)$$

In a tachogenerator with a conventional iron-copper armature, a ripple voltage will be superimposed on the d.c. output because of the relatively low number of commutator segments; the frequency and the magnitude of this ripple voltage will be dependent on the number of poles, armature segments, and brushes. A ripple-voltage component with a peak-to-peak value of five to six per cent of the output voltage is typical for brushed tachogenerators. The ripple voltage can be reduced by the use of a moving-coil configuration which has a high number of coils per pole; this minimises the ripple voltage to around two to three per cent. The armature consists of a cylindrical, hollow rotor, composed of wires held together by fibreglass and a polymer resin and has a low moment of inertia which ensures that the system performance is not compromised, similar to that of the ironless-rotor d.c. machine, discussed in Chapter 5. In addition to the low inertia and the low ripple content of the output, the axial magnets ensure that the motor length is small. In practice, this could add as little as 1 mm to the length of the overall package. A further refinement is the provision of frameless designs: this allows the system designer to mount the tacho directly on the shaft to be measured, thus removing any coupling errors.

The performance of a brushed tachogenerator depends on it being used within its specified operating capabilities; the linearity of the output will suffer if the load resistance, R_L, is allowed to fall below the manufacturer's recommended value. From Figure 4.9

$$E_g = R_a I + R_L I \qquad (4.11)$$

where R_a is the armature resistance; hence the terminal voltage, V, is given by

$$V = \frac{R_L K_g N}{R_a + R_L} \qquad (4.12)$$

The load resistance should be as large as possible to ensure that the terminal voltage is maximised; however, the current which is drawn should be sufficiently high to ensure that the commutator surface does not become contaminated.

4.2.2 Brushless d.c. tachogenerators

With the increasing use of brushless d.c. motors in servo systems, motor speeds are no longer limited by brushes; this leads to shaft speeds approaching 100 000 rev min^{-1} in some high-performance machine tools. The maximum speed of a brushed tachogenerator is limited to the speed at which aerodynamic lifting of the brushes occurs, and by increased armature-core losses which result in the output linearity deteriorating. Brushless tachogenerators have been developed as a response to these problems. The principle of their operation is identical to that of brushless motors (as discussed in Chapter 6), with the switching between phases being controlled by stator-mounted Hall-effect sensors. If the tachogenerator is integral to the motor, the Hall-effect sensors can be used for both motor and tachogenerator control. The maximum operational speed is only limited by the physical construction of the rotor assembly. There are no moving parts other than the rotor; this leads to a high reliability device, suitable for remote applications.

4.2.3 Incremental systems

An incremental-velocity measurement system is shown in Figure 4.10. A slotted disc, located on the shaft whose speed is to be measured, is placed between a light source and a detector. The source is usually a light-emitting diode; these diode have a longer life, and they are more rugged than filament bulbs, but are restricted to a temperature range of -10 to +75°C. The output of the photodetector needs to be conditioned prior to the measurement to ensure that the waveform presented has the correct voltage levels and switching speeds for the measurement system. The frequency of the signal, and hence the speed of the shaft, can be measured by one of two methods. Firstly, the frequency can be measured, in the conventional fashion, by counting the number of pulses within a set time period. This is satisfactory as long as the speed does not approach zero, when the timing period becomes excessive. To overcome this, an enveloping approach (shown in Figure 4.10) can be used. Each half-cycle of the encoder output is gated with a high-frequency clock; the number of cycles which are enveloped is determined, and this value is used to calculate the shaft speed. It should be noted that even this method will prove difficult to use at very low speeds, because the number of cycles per half-cycle becomes excessive.
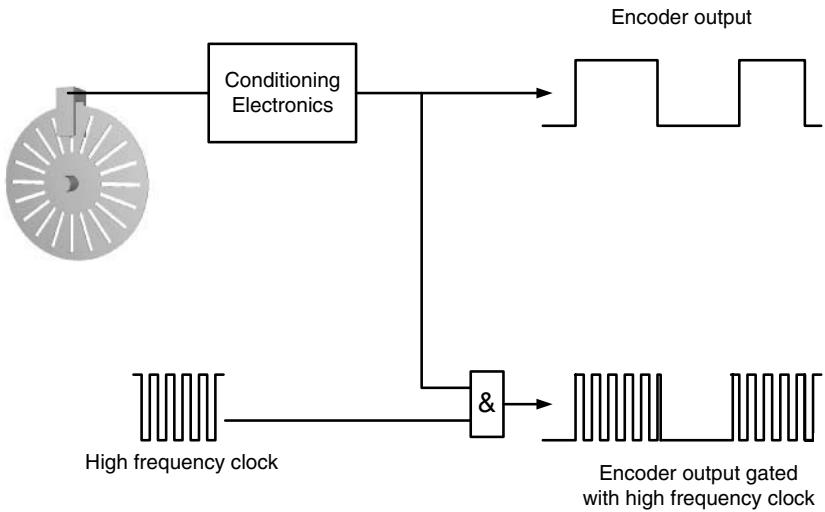
Figure 4.10. Speed-measurement using an incremental encoder. The output can be directly taken from the conditioning electronics, or to increase the resolution, the encoder waveform can be gated by a high frequency carrier.

The maximum operation speed of an incremental system is limited by the high-frequency characteristics of the electronics, and particularly by the opto-electronics. The resolution of the disc will determine the maximum speed at with the encoder can be operated, as shown in Figure 4.11.

4.2.4   Electromechanical pulse encoders

Using the counting techniques discussed above, it is possible to replace an optical encoder with an electromechanical system. A steel or a soft-iron toothed wheel is fitted to the shaft, and a magnetic, inductive, or capacitative proximity sensor is used to detect the presence of the teeth. While such a system is not normally capable of producing highly accurate speed measurements it can provide a rugged system which can be used in high-reliability applications such as over-speed/underspeed detectors for motors or generators.

4.3   Position transducers

Position transducers are available in three main types: incremental, semi-absolute, and absolute. A typical incremental encoder is an encoder that produces a set number of pulses per revolution, which are counted to produce the positional information. If the power is lost, or the data is corrupted, rezeroing is required to obtain the true information. An incremental encoder can be improved by the addition of a once-per-revolution marker; this will correct against noise in the system, but complete rezeroing will still be required after a power loss, because the count-
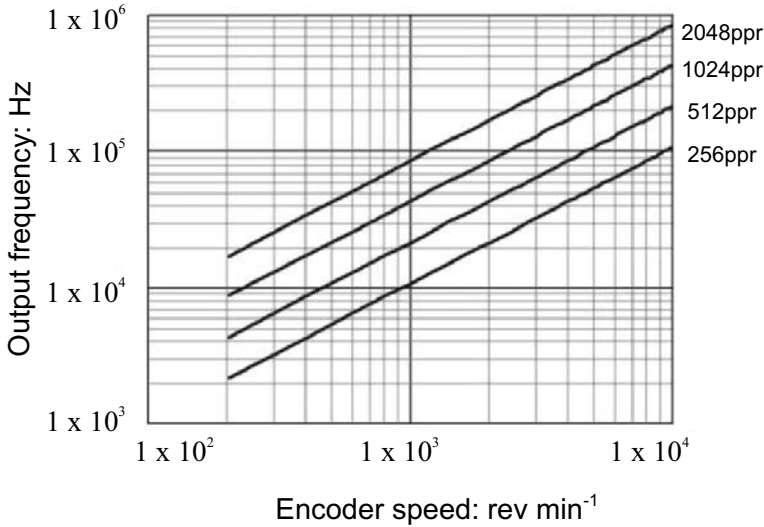
Figure 4.11. Encoder output frequencies as a function of speed for a range of incremental encoders, (ppr – pulses per revolution).

ing of the number of revolutions is also lost. An absolute transducer will maintain the zero and thus it will provide true information despite a loss of power for any length of time.

4.3.1 Brushed potentiometers

The principle of a potentiometer can be used in either a linear or a rotary absolute-position transducer in which the output voltage is a function of displacement. An excellent performance can be obtained if the drawbacks of the non-uniform track resistance and of the brush contact are considered to be acceptable. The accuracy of such a device will be dependent on regulation of the excitation voltage, which can be maximised by the use of a bridge circuit. A typical servo grade device will have a resolution of 0.05% of the full scale, with an accuracy of $\pm 0.1\%$. The maximum operating speed of a rotational version is typically limited to 500 rev min^{-1} by the brushes.

4.3.2 Linear variable differential transformers – LVDT

One of the most common methods of directly measuring a linear displacement to a high degree of accuracy uses a linear variable differential transformer (LVDT); the principal features of LVDTs are shown in Figure 4.12(a). The operation is based on a transformer in which the coupling between the primary and secondary coils (see Figure 4.12(b)), is determined by the position of a movable ferromagnetic core. The core is assembled using precision linear bearings to give low friction

and wear. The most widely used design has a secondary winding which is split into two, on either side of the primary. The secondary coils are wound in opposite directions and they are half the length of the moving core. In order to achieve high accuracies the windings have to be identical both in length and in inductance otherwise an unwanted quadrature signal will be produced, leading to non-linearities in the measurement; values of 0.5% for the accuracy are typical for LVDTs, increasing to 0.1% on selected devices. To operate an LVDT, the primary winding is energised with a sinusoidal excitation voltage, in the frequency range 2–10 kHz; the exact frequency depends on the type of device. With the secondary windings connected in series, the output voltage is

$$V_{out} = V_1 + V_2 \qquad\qquad (4.13)$$

When the core is in midposition, V_1 will equal V_2, and the output will be zero. As the core is displaced, the magnitude of the output rises linearly as shown in Figure4.12(c), with a $0°$ phase difference in one direction and a $180°$ phase difference in the opposite direction. Hence the magnitude of the output signal is proportional to the displacement of the central core, and the phase indicates the direction of travel. By the use of a suitable demodulator, a bipolar analogue voltage which is directly proportional to the displacement can be produced. Commercially available transducers can be obtained with displacements as small as 1 mm up to 600 mm in a variety of linearities and sensitivities. Because there is no physical contact between the core and the coils, the main mechanical components of the LVDT will not degrade with use. If precision bearings are used in the design, an almost infinite resolution, with zero hysteresis, is possible. The small core size and mass, and the lack of friction, mean that LVDTs have a high-response capability for dynamic measurements (for example, measurement of vibrations). Due to their rugged construction, it is possible to obtain LVDTs that are capable of operatinge in extreme environments, for example, ambient pressures up to 10^7 Pa and temperatures up to 700°C are not uncommon.

4.3.3   Resolvers

Resolvers are based on similar principles to LVDTs, but the primary winding moves relative to the two secondary windings rather than having a moving solid core, as shown in Figure 4.13(a). As the relative positions of the primary and secondary windings change, the output varies as the sine of the angle. By having two windings ninety electrical degrees apart and considering only the ratio of the outputs (Figure 4.13(b)), the variations due to the input voltage and the frequency changes become unimportant. The signals from the resolver are therefore relatively insensitive to an electrically noisy environment, and they can be transmitted over considerable distances with little loss in accuracy. In order to dispense with the need for sliprings, a separate rotary transformer is used to provide power to the rotating primary windings. The stator consists of the two output windings spaced
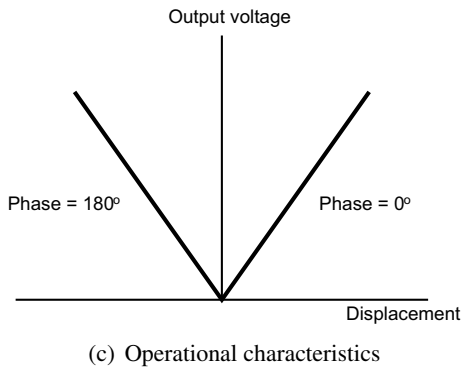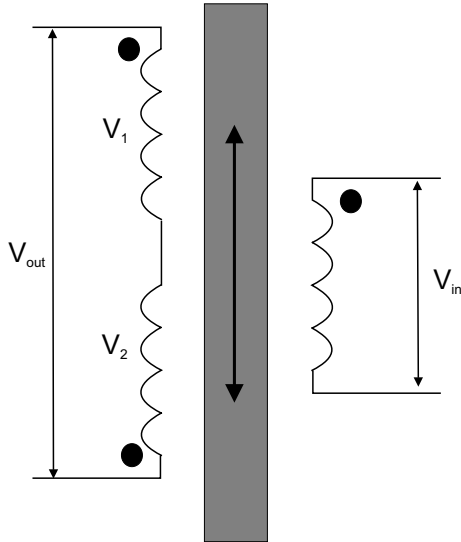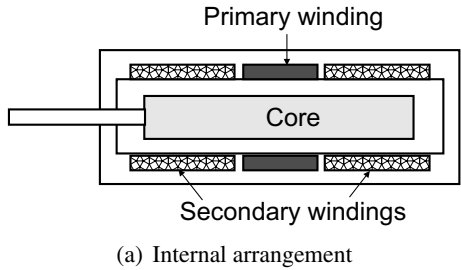
(a) Internal arrangement

(b) Electrical circuit, the dots signify the positive ending of the winding.

(c) Operational characteristics

Figure 4.12. The operation of the LVDT
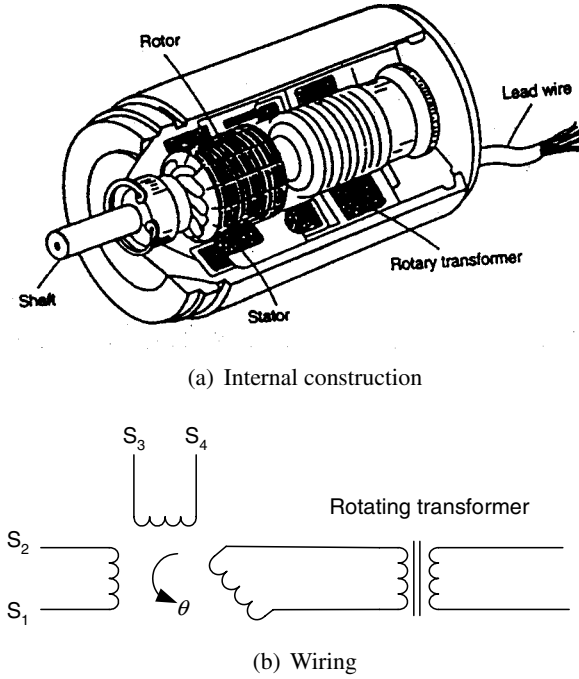
(a) Internal construction

(b) Wiring

Figure 4.13. Resolver construction and wiring

90 electrical degrees apart and the primary of a rotary transformer. The rotor also carries the secondary of the rotary transformer that is used to excite the rotor of the resolver. In the construction of resolvers, considerable care is taken to ensure that the cores, windings, and the air gap are constructed to an accuracy which ensures that non-linearity does not occur. In practice, errors can be caused by a number of factors including: a difference in the primary/secondary transformation ratio, an electrical phase shift, or a zero shift error between the two secondary windings and unequal loading of the windings by the external decoder. If the input to the resolver is

$$V = A \sin \omega t \tag{4.14}$$

the two outputs signals will be

$$V_{out1} = A k_1 \sin \theta \sin(\omega t + \alpha) \tag{4.15a}$$

$$V_{out2} = A k_2 \sin \theta \sin(\omega t + \alpha) \tag{4.15b}$$

where A is the amplitude of the excitation voltage, and k_1 and k_2 are the transformation ratios between the primary and the two secondary windings (which ideally should be equal), ω equals $2\pi f$ where f is the carrier frequency, and α is the rotor/stator phase shift (including any zeroing error).
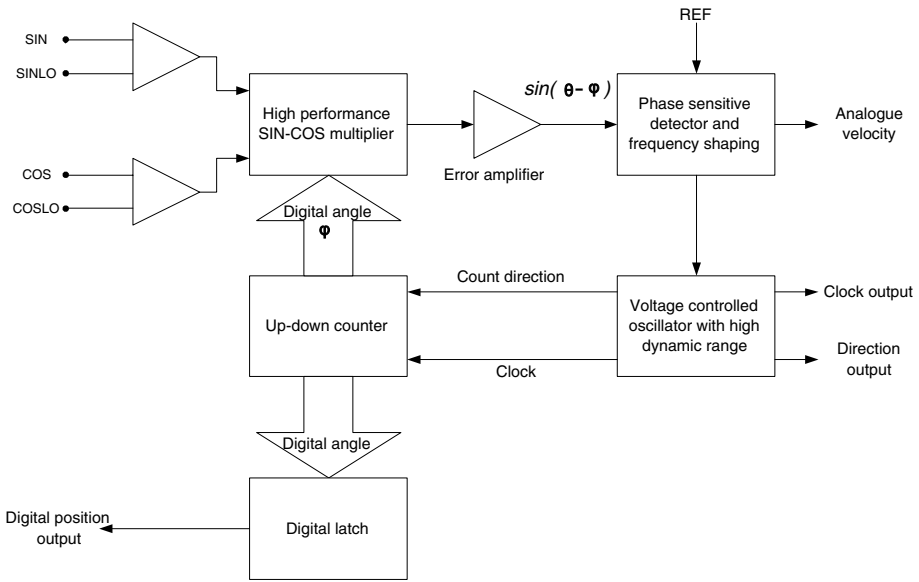
Figure 4.14. Internal function block diagram of a resolver-to-converter. The two stationary windings of the resolver are connected to SIN and SINLO, and COS and COSLO respectively. The resolver is powered by an external oscillator, which also provides the REF signal.

The output from the resolver can be used either directly as an analogue signal or after conversion to a digital signal. The advent of resolve-to-digital converters (RDC) has allowed digital data to be easily produced from resolvers. A modern RDC uses a ratiometric method; therefore, the system is not affected by changes to the absolute values of the signal to and from the resolver. This is of considerable importance if the transmission distance between the resolver and the RDC is large. It is current practice to provide a complete RDC as integrated circuits or as hybrid packages. This ensures that the best possible performance is obtained, with the packages' components optimised for temperature drift and other external sources of inaccuracy. A number of manufacturers provide devices that determine the resolver's velocity and position in a number of different formats: as a bipolar analogue signal or as a digital clock proportional to the speed, together with a logical direction signal. It is not uncommon for a device to haves 12-bit resolution up to 375 rev s^{-1}.

In this type of tracking converter, the two inputs (assuming a perfect resolver where $\alpha = 0$ and $k = k_1 = k_2$) are multiplied by the value held in a counter; if the output of the counter is assumed to be equivalent to an angle φ, then

$$V_1' = Ak \sin \theta \cos \varphi \sin \omega t \qquad (4.16a)$$

$$V_2' = Ak \cos \theta \sin \varphi \sin \omega t \qquad (4.16b)$$

Table 4.1. Resolution over 360°

Number of bits	Angle in radians	Angle in degrees
1	3.1415	180.00
2	1.5707	90.00
4	0.3927	22.5
8	0.02545	1.4063
10	0.00614	0.3516
12	0.001534	0.08789
16	0.000096	0.00549

and the difference from the error amplifier is

$$
\begin{aligned}
V_1' - V_2' &= Ak \sin \omega t (\cos \varphi \sin \theta - \cos \theta \sin \varphi) \\
&= Ak \sin \omega t \sin(\theta - \varphi)
\end{aligned}
\tag{4.17}
$$

A phase-sensitive detector, a voltage-controlled oscillator, and a counter form a closed-loop control system that attempts to minimise $\sin(\theta - \varphi)$. At the zero point, θ will equal φ, and the output of the counter will equal the angle of the resolver. In the selection of a tracking RDC, two major parameters need to be considered: the resolution (see Table 4.1) and the accuracy, both static and dynamic. The dynamic accuracy depends on how fast the voltage controlled oscillator (VCO) input tracks the error signal, which is dependent on the excitation frequency of the resolver that is used as part of the phase sensitive detector. One of the most significant forms of error is the lag in the tracking converter as the system accelerates; these errors may need to be considered in very-high-performance systems.

While a single resolver is only absolute over one revolution, applications often require absolute measurements over a number of revolutions. One possible solution is to couple two resolvers by a gear system (see Figure 4.15) so that the second resolver will rotate once for n turns of the input shaft. While this solution is perfectly acceptable, accuracies can be compromised by the backlash and tooth wear in the gearing. If anti–backlash gears are used, these effects will be very small; but they could be significant if the full 16-bits accuracy is required. In an anti-backlash gear, two independent gears are mounted on the same hub with a spring between the two providing a constant full-tooth engagement with the mating spur gear, thereby eliminating backlash in the mesh.

While the mechanical approach is satisfactory, it is more convenient to use a multipole resolver, where up to 32 cycles of stator voltage can be produced within 360 mechanical degrees. To provide absolute angular information, a second, coarse (one speed), winding is provided. By cascading a number of resolver-to-digital converters together, very-high-resolution systems can be constructed.
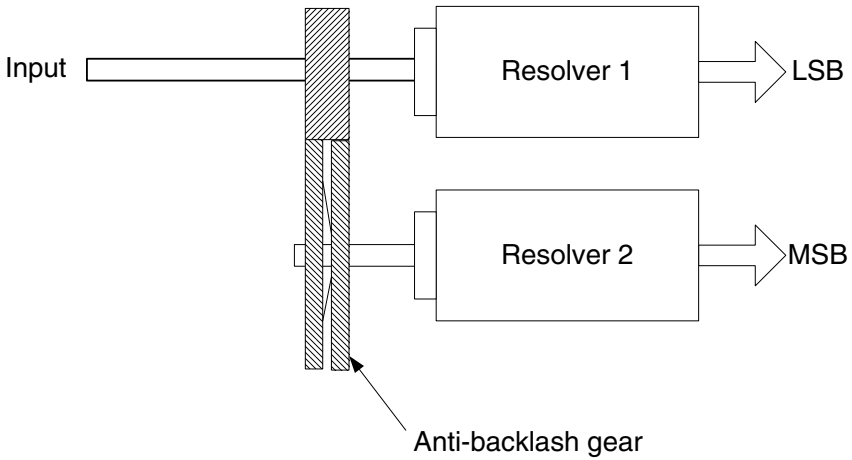
Figure 4.15. The use of anti-backlash gearing to increase the range of resolvers.

4.3.4   Rotary and linear Inductosyn

Inductosyn is the trademark of a position transducer manufactured by Inductosyn International of Valhalla, New York; the most the widely used version is based on an inductive principle. Linear Inductosyns can be fabricated in lengths of up to 40 m, or they can be manufactured in a rotary form up to 0.5 m in diameter (see Figure 4.16(a)). As Inductosyns are inductively coupled, non-contact transducers, they are very tolerant to changes in the local dielectric constant; therefore, their operation will not be affected by dust, oil, or pressure changes in a hostile environment. Inductosyns have applications in machine-tool, subsea, and aerospace areas, where very high resolution and accuracy are required.
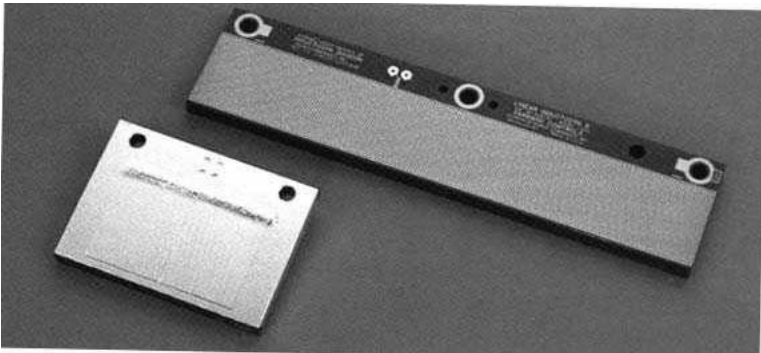
Inductosyns can be considered to be planar resolvers; the rotor and stator elements consist of a high precision hairpin element printed as a track over the complete length of the device (see Figure 4.16(b)). The length of one complete cycle of the pattern is the pitch P. An alternating current in the primary will induce a signal in the secondary. The amplitude is dependent on the relative positions of the primary and secondary windings, giving

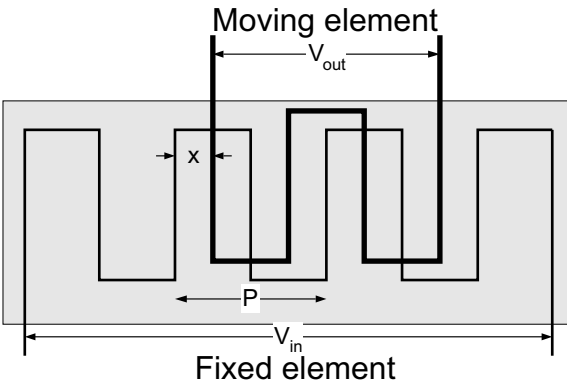$$V_{out1} = kV \cos \left[\frac{2\pi x}{P} \right] \qquad (4.18)$$

where V is the excitation voltage ($V = V_{pk} \sin \omega t$), k is the transformation ratio, and x is the displacement. If a second output winding is displaced by $\pi/2$ electrical degrees from the first winding, its output voltage will be:

$$V_{out1} = kV \sin \left[\frac{2\pi x}{P} \right] \qquad (4.19)$$

As these output voltages have the same form as those of a resolver, an identical converter can be used to determine the displacement, x. In practice, the number

(a) A commercial linear system, photograph reproduce by permission of Inductosyn International, Farrand Controls, Valhalla, NY.

(b) The relationship between the fixed and moving elements in an Inductosyn

Figure 4.16. The linear Inductosyn.

of complete pitches are counted to determine the total distance move. The pitch of a metric linear Inductosyn is is such that a resolution of 5×10^{-6} m to be achieved. Rotary Inductosyns are supplied with pitch counts in the range 32–2048 per revolution, with achievable accuracies to $\pm 0.5''$.
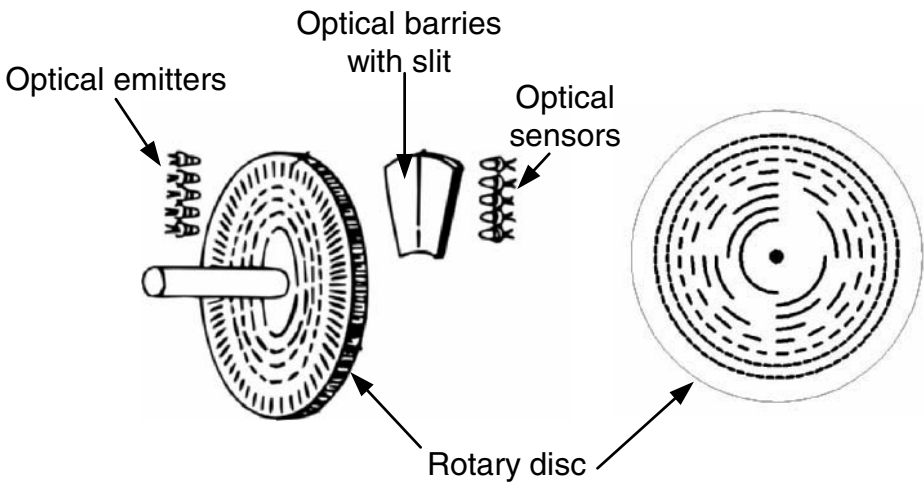
4.3.5   Optical position sensors

Optically based encoders are widely used for position measurements in robots and machine tools. They can take one of three forms: absolute, semi-absolute, and incremental. Each of these types of encoder consists of three elements: an optical receiver, a light source, and a code wheel. The receiver is normally a phototransistor or diode which responds to the light intensity which is received. As discussed earlier, the light source can either be a solid-state light-emitting diode or a filament bulb. The difference between the types of encoders is characterised by the information contained on the code wheel and by how it is interpreted by an external control system.
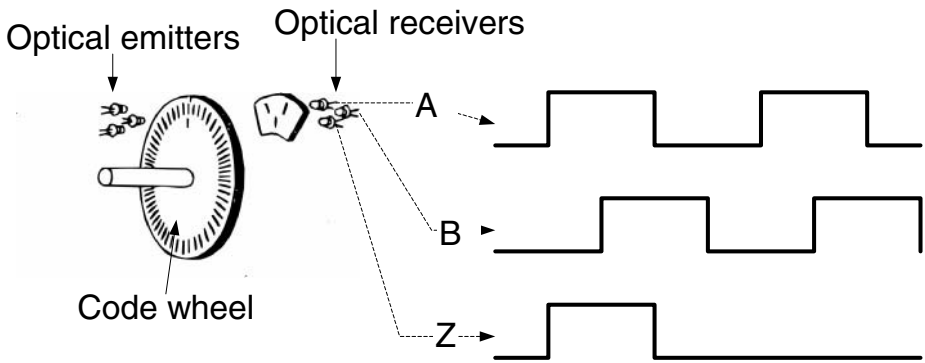
Absolute optical encoders   incorporate a code wheel that is encoded in binary, either in pure binary or in grey code, with one bit per track. The latter is preferred because only one bit changes between any two states. This prevents errors, since there is no way of guaranteeing that all the bits will change simultaneously at the boundary between two states, due to inherent manufacturing problems with the code wheel. For example, if pure binary is used (see Table 4.2), it would be possible to generate an output of 15 during the transition from 7 to 8. Code wheels are normally produced on glass substrates by photographic methods. This is costly for high resolutions; as will be readily appreciated, as the resolution of an absolute encoder increases, so does the size and complexity of the code wheel (see Figure 4.17(a)).

Semi-absolute and incremental optical encoders   are identical in most respects, and they can thus be considered together. The construction of an incremental encoder is based on a code wheel which has a single track of equal-sized, opaque and translucent slots; and, as the wheel is rotated, an alternating signal is produced with a frequency which is proportional to the speed of rotation (see Figure4.17(b)). Semi-absolute encoders are incremental encoders with an additional output giving one pulse per revolution. As the output of these detectors is typically a distorted sine wave, the output needs to be suitably conditioned to produce a clean square wave for other electronic systems. This circuitry can be mounted in the encoder or it can form part of the external system. As the resolution of the encoder increases, the use of physical slots in the code wheel will become unreliable; hence, use is made of gratings. As the code wheel is moved, the whole field observed by the optical receiver goes dark as the lines move in and out of phase.

As previously discussed, an encoder with a single track will allow the magnitude of the speed to be measured; the direction of rotation can be determined by the

(a) Absolute encoder

(b) Semi–absolute encoder

Figure 4.17. Rotary optical encoders.

addition of a second track or an additional sensor to produce a quadrature signal. The two signals A and B shown in Figure 4.17(b) are displaced by 90 electrical degrees. As a result, if the encoder moves forward, channel A will lead channel B, and vice versa when the motion is reversed. A number of techniques can be used to detect the direction of motion; one possible technique is shown in Figure 4.18. Figure 4.19 shows the waveforms used to discriminate direction.

The encoder signal is used to generate a pulse from the monostable, which can be inhibited by the other channel; the resulting pulse is used to latch a flip-flop, whose output indicates the direction of motion. The speed and position are measured by pulse-counting techniques, the resolution being determined by the size of the counter and the encoder. An encoder is specified by the number of lines per rotation; however, since channels A and B are shifted by 90 electrical degrees it is possible to divide each encoder cycle in four, hence the resolution

Table 4.2. Pure binary and grey codes as used in an absolute rotary position encoder.

State	Pure binary	Grey code
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100
9	1001	1101
10	1010	1111
11	1011	1110
12	1100	1010
13	1101	1011
14	1110	1001
15	1111	1000

of a 360 pulses per revolution (p.p.r.) encoder can be increased to 1440 counts per revolution by the addition of an electronic system. Since this increases the effective system resolution at a cost which is significantly lower than for encoders with four times the resolution, this can almost be considered to be a standard feature of position systems. Commercial systems are also available that will increase the encoder resolution by 8 and 12 times.

Linear optical encoders operate in an identical fashion to rotary incremental encoders, where a optical sensing head moves over the stationary grating, which is either a glass scale or a reflective steel strip. When the scale with a grating moves relative to another grating with an identical pattern – the index grating – the lines and gaps alternately align. The light-dark modulation produced is detected with optical sensors, a typical system is shown in Figure 4.20. The arrangement of the optical sensors and the signal processing required is a function of the design and the encoder resolution. It is possible to obtain reflective linear optical encoders in lengths of up to 50m, the performance depends on the care of the installation, particularly the alignment between the encoder track and the moving sensor head.

It is possible to purchase absolute linear encoders which have up to seven tracks, the information from which is combined to provide the absolute position. Due to the complexity of the process, these encoders are limited to lengths of 3 m or less, with a resolution of up to 0.1 μm.

As with rotary encoders a reference mark is provided on a second track, parallel to the incremental track, which are scanned, and used to locate the datum position,
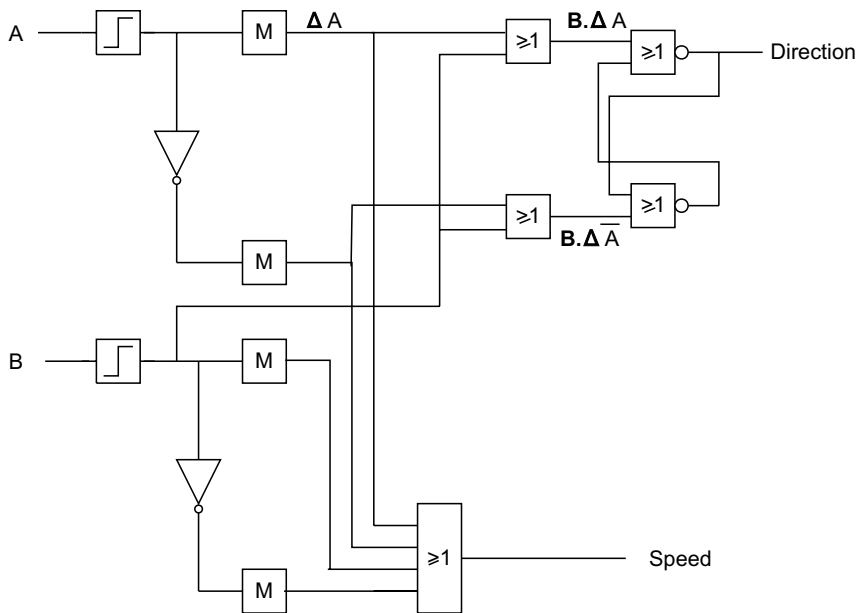
Figure 4.18. Position decoding for an incremental encoder, the blocks marked M are single-shot monostables, operating on the rising edge, the waveforms are shown in Figure 4.19.
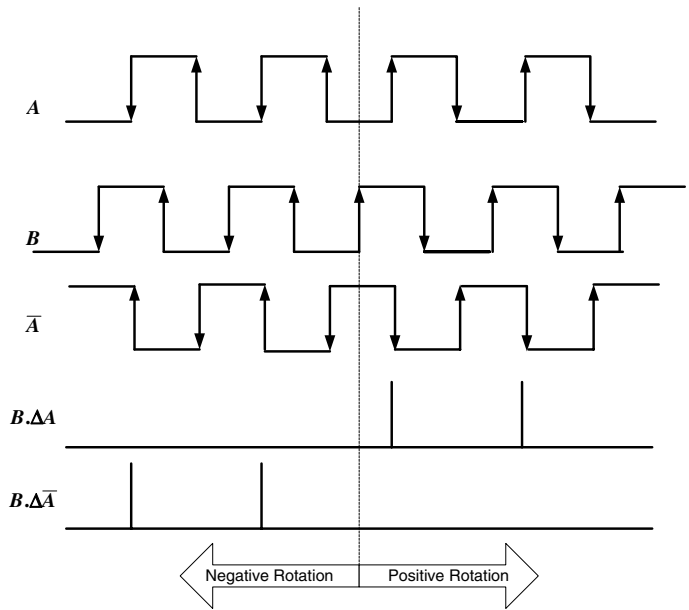

Figure 4.19. The discrimination of direction using an incremental encoder.
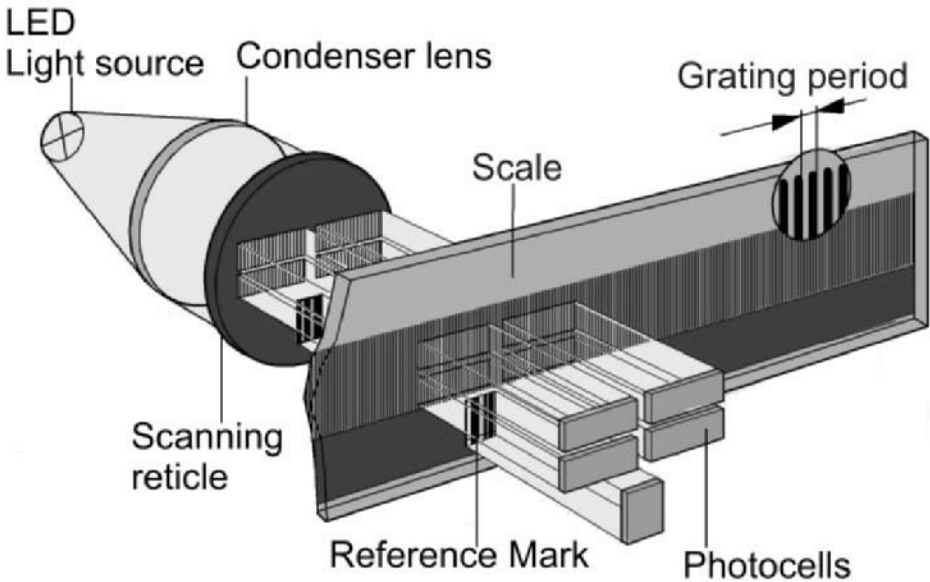
Figure 4.20. A linear optical encoder showing the grating arrangement. Image reproduced by permission of Dr. Johannes Heidenhain GmbH, Traunreut, Germany.

as discussed in section 4.4.3.

4.4 Application of position and velocity transducers

The correct installation of an encoder or transducer is critical to its satisfactory operation. During installation, particular consideration must be given to the mechanical aspects and to the connection to the system's measurement electronics.

4.4.1 Mechanical installation

The previous sections have described the operation of a range of velocity and position transducers. In practice, units are supplied either complete or as a set of components in a frameless design. Frameless transducers are supplied to allow direct integration into the mechanical structure of a system, therefore reducing, or eliminating, errors caused by windup in couplings or shafts and eliminating backlash in gears. A range of common sizes has been developed for resolvers and optical encoders; the more significant sizes are listed in Table 4.3. These standard sizes permit easy interchangeability between manufacturers' products. It should be noted that the shafts can either be solid or hollow, giving designers a number of integrations options for the design of a mechanical systems.

In coupling the motor or the load to a rotary transducer, care must be taken to ensure that the respective shafts are correctly aligned in all axes; if they are not
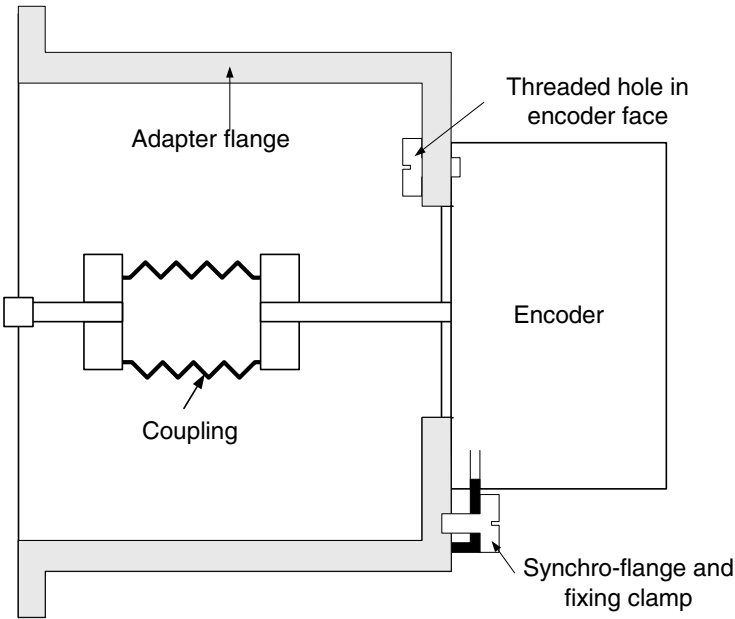
Figure 4.21. Connecting an encoder to a shaft using a bellow coupling and adapter flange. Threaded holes can be provided on the face of the encoder, or if the encoder is fitted with a synchro-flange, fixing clamps can be used.

Table 4.3. Standard encoder and resolver sizes, dimensions in mm

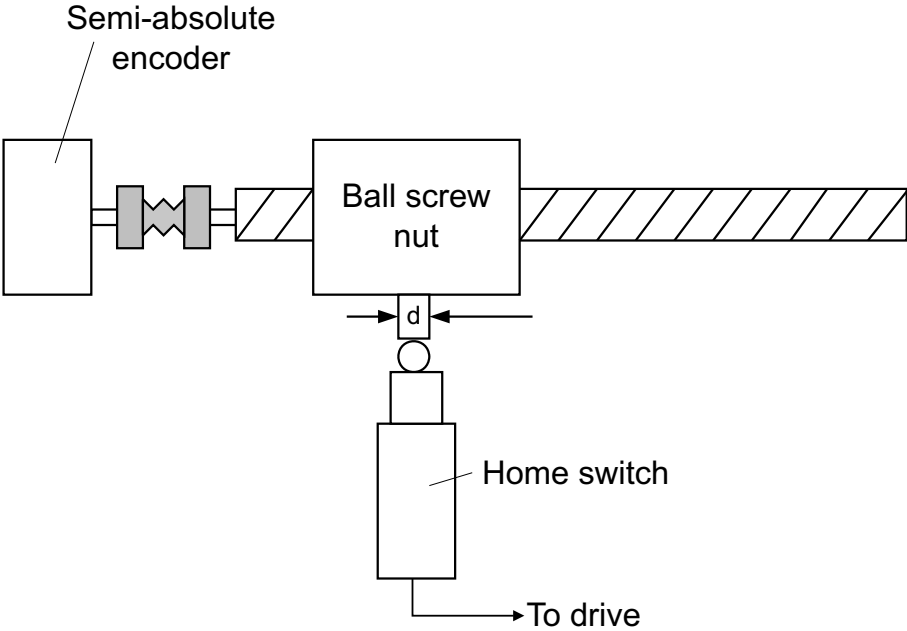| Type	Size	Diameter	Length	Shaft diameter
Frameless	15	36.83	25.4	6.35
Frameless	21	52.37	31.37	12.7
Housed	8	19.13	31.5	
Housed	11	27.05	40.39	

correctly aligned, a considerable load will be placed on the transducer bearings, leading to premature failure. Two methods of fixing are shown in Figure 4.21, during assembly is is normal practice to tighten the flange mounting screw last, to ensure that the input shaft and encoder shaft are in alignment. Since transducers can be supplied either with a conventional solid shaft or with a hollow shaft, the coupling which is used depends on the type of encoder and on the application. The use of bellows couplings will allow very small alignment errors to be eliminated, while still retaining a solid coupling between the motor and the transducer. If a hollow shaft encoder is used, the shaft can be coupled directly to the transducer, while its transducer itself is fitted to the system using a compliant mount. If a frameless transducer is used, its installation will be very specific to the unit and the application, and the manufacturer will normally supply details of the installation design where necessary.

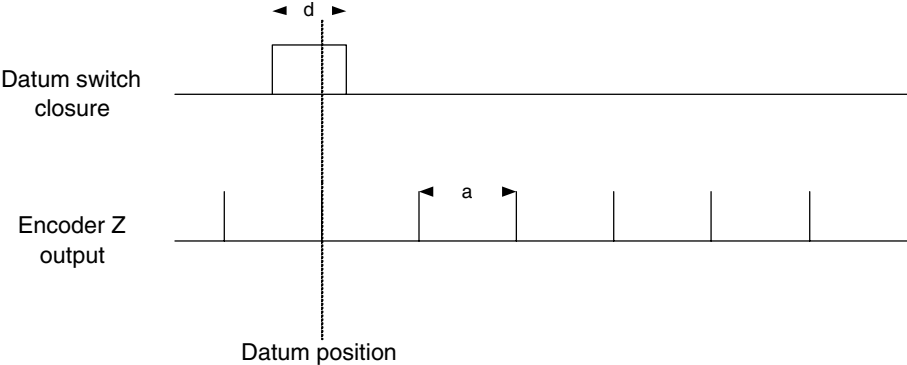In installing both linear and rotary transducers a number of additional requirements need to be considered:

- The operating temperature range specified by the manufacturer defines the limits within which the values given in the specifications for the encoders are maintained. Operating out side the limits will results in a degradation of performance.

- All encoders are subject to various types of acceleration during operation and mounting, again these will be detailed in the manufacturers' specifications. Due to presence of a glass code wheel or linear scale, encoders are considered fragile, particularly during the assembly of a system.

- Both linear and rotary encoders have internal friction, particularly if the design includes a seal, this is normally specified as a torque or force in the specifications.

- All types of encoders should not be subjected to vibration during operation: this can be more significant for a long linear encoder. To function properly, the more solid the mounting surface the better. It is recommended that linear encoders should not be mounted on hollow parts.

- As discussed in section 2.5.5 is a major factor in selecting drives and their associated components. For example linear encoders are normally protected IP 53 (see Table 2.3) provided that they are mounted with the linear seal is facing away from possible sources of contamination. If the encoder will be exposed to heavy concentrations of coolant and lubricant mist, the scale housing can be fed with compressed air to raise the IP rating to IP 64 effectively preventing the ingress of contamination.

4.4.2 Electrical interconnection

The wiring and connectors between the transducer and the processing electronics are critical to the operation of a system. If they are not satisfactory, in the case of a digital encoder, any electrical noise which is introduced will probably result in additional pulses being counted and hence in an increasing positional error. In analogue systems, electrical noise resulting from poor connections will result in a poor signal-to-noise ratio and hence in a degraded performance, These problems can be reduced by the use of twisted screened cables and high-quality connectors throughout the system. As shown Figure 4.11, at high speeds the encoder output frequency can exceed 100 kHz, and therefore the wiring and associated electronics must be designed to accommodate signals of this frequency; in particular stray capacitance must be minimised.

(a) Typical mechanical arrangement for a home or datum switch

(b) The relationship between the switch closure (d) and the distance between the output on the encoder's z-track, for reliability $a > d$.

Figure 4.22. The use of a home switch to define the datum position of a linear axis.

4.4.3   Determination of datum position

When a rotary incremental encoder is used in linear applications one of the design requirements is to accurately and repeatably determine the datum position. This is the point to where all measurements on a particular axis are referenced. As shown in Figure 4.17(b), it is common practice for an incremental to have track that provides one pulse per revolution. This is used to provide the exact datum point. To ensure that the datum point is selected within the correct encoder revolution, it is normal practice to provide a *home switch*, as shown in Figure 4.22. The distance moved once the switch is mode must be less than that which results in one encoder revolution, or the possibility exists that the datum position will have an error of plus/minus one encoder revolution.

4.5   Summary

This chapter has reviewed the range of velocity and position encoders that are currently available. In order to make a satisfactory selection, designers have to address a number of key question, including:

- What resolution and accuracy is required by the application, particularly as increasing both parameters directly affects the overall system cost?

- What are the environmental constraints at the location where the measurement system will be installed?

- How will the derived information be integrated into the system? The answer to this question will depend on the controller selected, and this question may need to be addressed at the completion the selection procedure.

The encoder or transducers which are selected will have a major effect on overall system performance; for if the wrong measurement is made, the system will never be able to produce the required result.