CHAPTER 10

ENCODERS

10.1 LINEAR ENCODERS

Linear displacement can be measured and communicated by a device called an encoder without using any form of analog-to-digital conversion because the basic output signal is already in a digital format. Although the term *encoder* has also been applied to devices based on laser interferometers, the term *linear encoder* will be used here in reference to standard industrial transducers based on geometric patterns applied along a linear scale and detected by any one of several methods. Linear encoders are available as incremental or absolute reading and encompass various detection techniques, including brush, optical, magnetic, and capacitive types. Besides selecting whether the output will be absolute or incremental, encoder designers and users make trade-offs among important product features, including ruggedness, resolution, and physical size.

10.2 HISTORY OF ENCODERS

The earliest type of linear encoder was the *brush* type, shown in Figure 10.1, in which mechanical contact fingers rub along a metal pattern printed onto an insulating base. The path of the brush moves over conducting and insulating segments. When the brush is in contact with a conducting segment, a contact closure occurs [23, p. 106]. The pattern is formed onto the base in the same

Linear Position Sensors: Theory and Application, by David S. Nyce
ISBN 0-471-23326-9

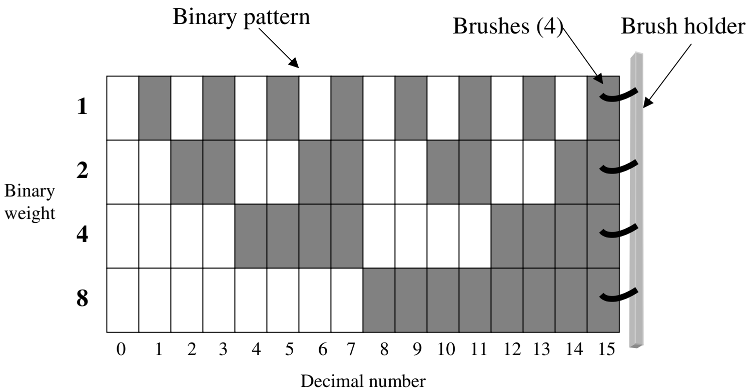

Figure 10.1 Brush type of linear position encoder with binary pattern. (See Section 10.8 to learn about other patterns.)

way as printed circuit boards are made for connecting electronic circuits. The major drawback of this contacting measurement technique was the wear of the metal pattern and fingers, sometimes resulting in errors, and eventually resulting in failure.

Noncontact encoders based on optical and magnetic techniques took over in the 1960s and 1970s because of their increased life and reliability. In the 1990s, the brush type largely became no longer acceptable for industrial measurement systems. For this reason, additional details of brush-type encoders are not included here, except where expedient to illustrate switching and coding theory.

10.3 CONSTRUCTION

A linear encoder can take many physical forms. Some have a housing similar to a potentiometric linear position transducer, comprising a housing, rod, bushing, wiper, scale, detector, and electronics. Others have a read head that rides in a track along the scale. Still others have a separate read head that must be mounted by the user so that it passes along parallel to the scale. An optical type is shown in Figure 10.2. The housing provides the base for combining the component parts, and includes means for mounting the encoder to the application.

In a *rod* type, the rod is usually made from hardened and polished steel and moves linearly into and out of the housing from one end. A bushing supports the rod, reducing wear and resisting the force due to side loading. One or more wipers clean particles away from the rod before it goes through the bushing. This reduces wear of the rod and bushing. The scale is an opaque member with slots (or reflective areas) for an optical transducer or a coded magnetic strip

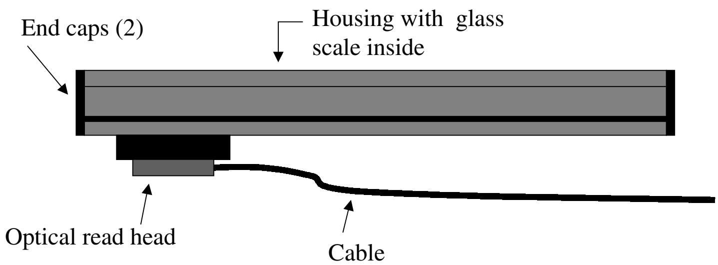

Figure 10.2 Optical type of linear position encoder.

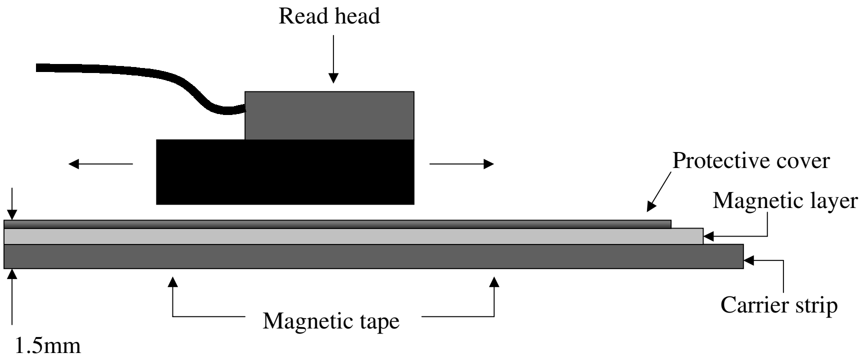

Figure 10.3 Incremental magnetic encoder with separate magnetic tape.

for a magnetic transducer. As the rod moves in or out, displacement or absolute position information is read from the scale via the detector.

An alternative construction is shown in Figure 10.3. In this magnetic encoder, the magnetic tape is supplied to the user as a separate component. The tape is attached to a stationary member, along the axis to be measured. The moving member is fitted with a magnetic pickup device. The combination of the coded tape and the pickup device comprises the linear encoder. The incremental pulse counting encoder circuit is zeroed at a starting position, often at one end of the tape. Then, as the pickup moves along the tape, the variations on the tape are counted and used to indicate the position as a distance from the starting position. As in all incremental encoders, this is actually a measurement of displacement rather than one of position.

10.4 ABSOLUTE VERSUS INCREMENTAL ENCODERS

It may be obvious that the information encoded onto the track of an incremental encoder can have a finer pitch than that of an absolute encoder. This is because only one or two bits must be toggled while the position is changing with an incremental encoder. With an absolute encoder, sufficient information must be read at each position to represent the position totally at that point in the transducer stroke with respect to the reference datum. So if the position count is 2046 counts, an absolute encoder must encode the number 2046 (requiring 11 binary bits) at that exact location, whereas with an incremental encoder, there is only one bit of information located at that point (or two, counting the quadrature bit at 90° separation from the first, as presented later). In addition to the A quad B tracks, an incremental encoder also may have an index track. The index track has a mark at one point that indicates the index or starting point. Then the count can be rezeroed automatically each time the index point is passed. So an advantage of an incremental encoder is that of higher resolution for a given spacing of the detected feature (e.g., optical slots). An absolute encoder, however, has the advantage of instant startup after a corruption of power or data, without needing to rezero or reset the count.

10.5 OPTICAL ENCODERS

An optical encoder uses a transmitter/receiver pair of optoelectronic devices. The transmitter part of the pair is a light source, usually a light-emitting diode (LED). The receiver part is usually a phototransistor. A track having opaque and transparent sections arranged in series is passed through the optoelectronic pair as the position changes. (Alternatively, reflective and nonreflective sections can be used; see Figure 10.4.)

The light source has a means for focusing the light onto the detector. This may include a focusing lens, collimator, and/or a slit or pinhole. If the encoder is incremental, only two receivers are needed. The second receiver is spaced

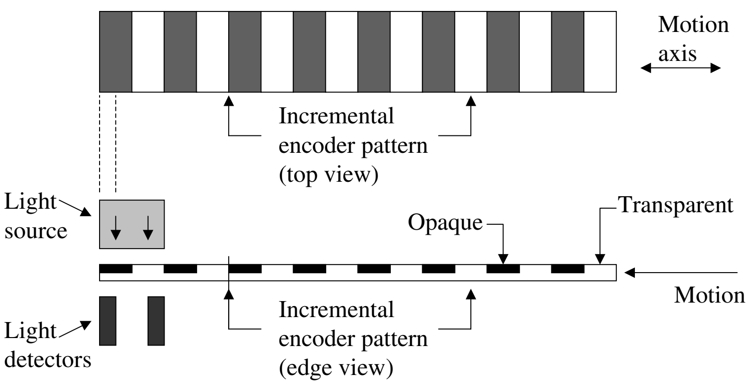

Figure 10.4 Incremental optical encoder with LED and phototransistor arranged for quadrature output.

from the first by a set distance to yield a 90° shift between the two outputs, resulting in a quadrature output (see Section 10.7). In this case, the alternating dark/light areas are counted to obtain the present position. A closer spacing between adjacent dark or light areas will yield a higher resolution. A third receiver may be used to detect an index mark.

In an absolute encoder, a sufficient number of light receivers are needed to represent the maximum number of bits to be indicated. The reciprocal of this number of bits is the resolution. For example, a 12-bit encoder represents 4096 counts and has a theoretical resolution of 0.024% or 1/4096. In this example there will be 4096 sets of position data along the full-range stroke of the sensor. Twelve parallel data bits will be read by 12 phototransistors. It is possible to use one long light bar to drive all of the phototransistors, or multiple LEDs can be used.

10.6 MAGNETIC ENCODERS

A magnetic encoder uses a magnetic tape onto which the position information is recorded, together with one or more magnetic sensors to read the data. The magnetic sensors in modern encoders are usually Hall effect or magnetoresistive types. Information on the Hall effect and on magnetoresistance is included here in Chapters 7 and 8, respectively. An incremental magnetic encoder is shown in Figure 10.5. Within the track is a magnetic tape having reversals of polarization along its length. These magnetic field variations are detected by the magnetoresistive pickups. If the encoder is incremental, only two pickups are needed. The second pair is spaced from the first pair by a set distance to yield a 90° shift between the two outputs, resulting in a quadrature output (see Section 10.7). In this case the pulses from the pickups are counted to obtain the present position. The highest resolution that is possible is limited by the smallest size of two adjacent field variations that can be

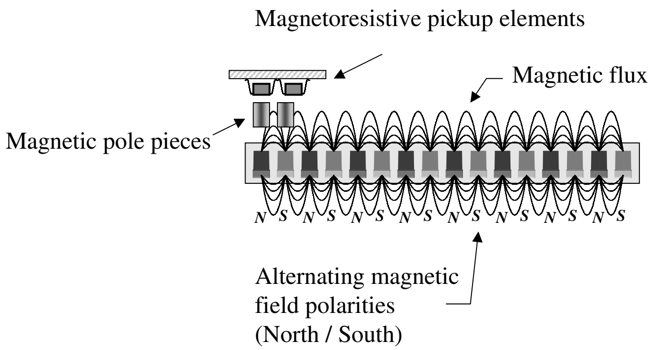

Figure 10.5 Incremental magnetic encoder with magnetoresistive pickup.

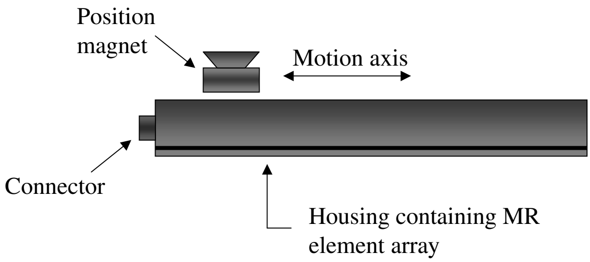

Figure 10.6 Absolute magnetic encoder.

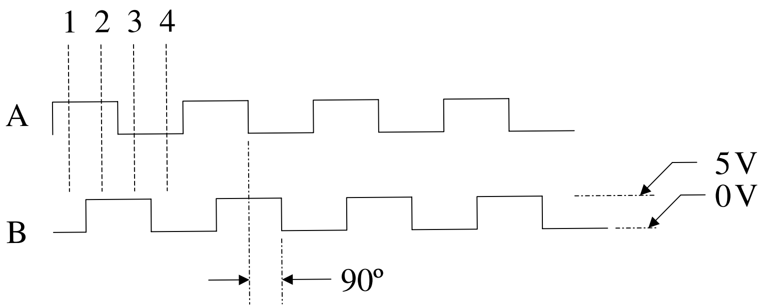

Figure 10.7 A quad B outputs are separated by 90°. States 1 through 4 occur during each count cycle.

differentiated by the pickup device. The magnetic encoder can be incremental or absolute (see Figure 10.6), following the same theory as presented for the optical type.

10.7 QUADRATURE

Incremental encoders have two outputs, called A and B. These are arranged in quadrature, which means that they are separated in phase by 90°, as shown in Figure 10.7. This arrangement is also called *A quad B*. It is called *quadrature* because "quad" means four, and a transition of one or the other data line (A or B) occurs four times per count cycle. A complete count cycle in most situations is generally considered to comprise 360° (the 360° is a count cycle, not the angle of rotation of a rotary sensor). So with four transitions per count cycle, one phase is delayed when compared to the other by 90°, or one-fourth of 360°. The four possible states are (1) A high, B low; (2) A high, B high; (3) A low, B high; and (4) A low, B low.

If the pulses from either A or B are counted, the change in position is indicated by the number of counts multiplied by the distance per count (e.g., 1000 counts with a resolution of 10 μm would be 1000 counts × 10 μm = 10 mm). The reason for having the other output (B or A) is to find the direction of motion, either incrementing or decrementing the count, so the current count represents the actual position. In Figure 10.7, for example, if A goes from low to high while B remains low, it is an increment. Conversely, if B is high at that time, it is a decrement. This basic explanation can be used to easily understand how to obtain a count and the count direction. It is also possible to obtain four counts per cycle by looking at all of the transitions as count inputs and looking at the relationship between the A and B inputs, at the times of the counted transitions, to determine the direction of the change.

Modern circuits that read the A quad B signals from a position transducer do not actually wait for a transition to occur and then count that transition. Instead, it is more common to monitor the states of A and B continuously at a higher sampling rate than the transitions are expected to occur. With this (state, rather than transition) information and, usually, a microcontroller, smoother operation can be obtained with less likelihood of error.

10.8 BINARY VERSUS GRAY CODE

In an absolute encoder, the output is typically either in binary or Gray code, but binary-coded decimal (BCD) is also available. BCD is similar to binary, except that the data bits are arranged into sets of 4 bits each. Four bits of BCD data equal one character and can have a value of zero through nine. For reference, hexadecimal is also shown because the reader may be familiar with this binary number representation. In hexadecimal, an 8-bit binary word is viewed as two 4-bit nibbles. The 4-bit nibble can have 16 possible values. These are represented as the decimal numbers 0 through nine, followed by letters A through F (for a total of 16 characters).

Natural binary or BCD is easy to interface directly to standard digital circuits but has the disadvantage that a change of only one increment involves the simultaneous change of more than one bit. This means that a large error can be indicated if all the bits do not switch at exactly the same time. For example, when the count changes from 7 to 8, it is a change from 0111 to 1000 in binary or BCD. But if the most significant bit (MSB) changes a few milliseconds before the rest of the bits change, there will be a reading of 15 (1111 in binary, F in hexadecimal, or an error in BCD) for those few milliseconds. This can represent a large error and cause stability problems in a servo control system. A system called the Gray code was developed to solve this problem.

The Gray code is arranged so that an increment of 1 bit always changes the output by only 1 bit [37, pp. 6–126]. The differences among BCD, hexadecimal, binary, and a Gray code are shown in Table 10.1. Figure 10.8 is a corresponding Gray code pattern. One can see that a change between any two adjacent numbers requires only the change of 1 bit in the Gray code. When the transducer operates using the Gray code, the controller or other device to

TABLE 10.1 Decimal Equivalents of Hexadecimal, Binary-Coded Decimal (BCD), Natural Binary, and Gray Code

Decimal	Hexadecimal	BCD		Natural Binary	Gray
0	0	0000	0000	0000	0000
1	1	0000	0001	0001	0001
2	2	0000	0010	0010	0011
3	3	0000	0011	0011	0010
4	4	0000	0100	0100	0110
5	5	0000	0101	0101	0111
6	6	0000	0110	0110	0101
7	7	0000	0111	0111	0100
8	8	0000	1000	1000	1100
9	9	0000	1001	1001	1101
10	A	0001	0000	1010	1111
11	B	0001	0001	1011	1110
12	C	0001	0010	1100	1010
13	D	0001	0011	1101	1011
14	E	0001	0100	1110	1001
15	F	0001	0101	1111	1000

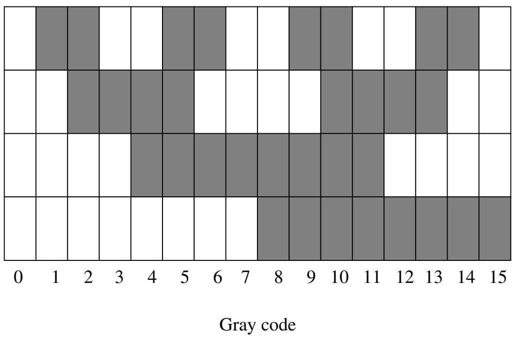

Figure 10.8 Gray code pattern corresponding to Table 10.1.

which it is connected will usually incorporate a Gray-to-binary conversion. A Gray-to-binary converter can be built in hardware with gates as shown in Section 10.9, or it can be stored in a lookup table in memory. Then the controller can use the binary numbers for operation as usual after they are converted from the Gray code.

10.9 ELECTRONICS

The light source of an optical encoder is usually one or more LEDs driven by a simple reference voltage with current limiting resistor, or by a constant-

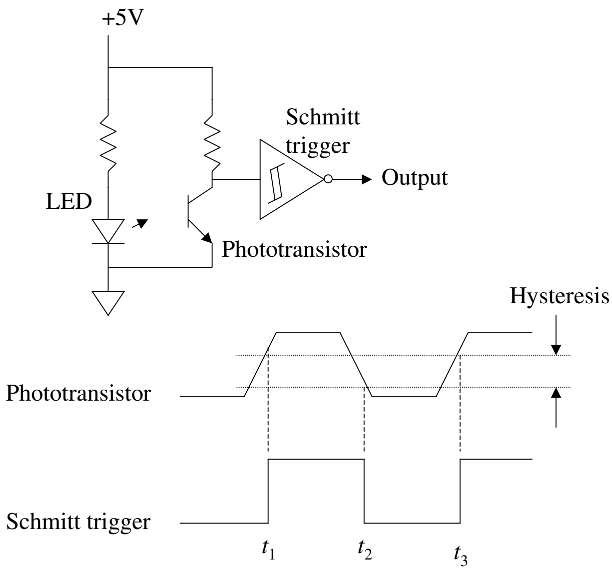

Figure 10.9 LED and phototransistor connection circuit, with Schmitt trigger.

current source. Temperature compensation capability may be added to maintain a constant light output with temperature variations. The LED and phototransistor are typically connected as shown in Figure 10.9. The signal directly from the phototransistor will not be a waveform with sharp transitions. The sharp transitions needed for the A quad B output are generated by a Schmitt trigger circuit. A Schmitt trigger transitions from one state to the other when the input crosses a voltage threshold. It has hysteresis built in so that the opposite transition will not take place until the input moves substantially past that same threshold. So there is a positive-going threshold and a negative-going threshold. In the figure, the positive-going threshold is approximately 4 V (with a 5-V power supply voltage). The negative-going threshold is approximately 1 V. This is a hysteresis of 3 V.

If an absolute encoder uses a Gray code pattern, the resulting output must be converted to natural binary before the data can be handled by a microcontroller. This conversion can be done by using a mathematical formula in the microcontroller, a lookup table, or in hardware. A schematic is shown in Figure 10.10 for converting the Gray code to natural binary.

10.10 ADVANTAGES

With resolution as fine as 0.01 μm (millionths of a meter), linear encoders are often the preferred method of position sensing in the precision environment

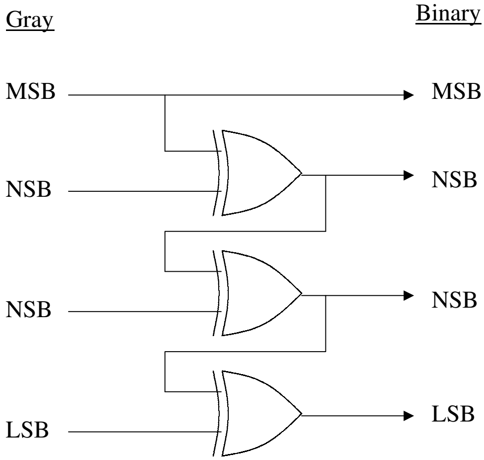

Figure 10.10 Schematic for converting Gray code to natural binary.

of machine tools. The highest-resolution transducers tend to be the incremental ones. A disadvantage to the incremental mode is that data can be corrupted due to electromagnetic noise or power fluctuations.

Magnetic linear encoders, with a magnetic tape and a sensor head, can have a long measuring range of up to 40 m (such as the Lincoder by Stegmann). The user makes the installation by attaching the magnetic tape to a fixed surface, along which the sensor head will travel. They retain the advantage of noncontact measuring, and so have a virtually unlimited life.

Shorter linear encoders can be magnetically or optically coupled and are self-contained with the magnetic tape or optical scale contained within a housing. Typical industrial models can have a resolution of better than 0.1 μm. The bushings and wipers, used with models having an actuator rod, have a finite lifetime. End of life, however, does not mean that the measurement accuracy is affected. It means that some mechanical drag is encountered as the rod rubs on worn bushings or bearings, sometimes accompanied by audible noise.

10.11 TYPICAL PERFORMANCE SPECIFICATION AND APPLICATIONS

Figure 10.2 showed an optical linear encoder. Representative specifications are as follows:

Full-scale range:	100 mm to 3 m
Resolution:	0.1 to 10 μm
Accuracy at 20°C:	0.1 to 10 μm
Hysteresis:	0.5 μm

Operating temperature:	0 to 50°C
Maximum speed:	1 m/s
Input power:	5 V dc at 180 mA maximum

Coding patterns can be photographically reproduced onto the measuring medium. Any nonuniformity of the pattern on the measuring medium is a source of error. In a linear encoder, these include the width and spacing of the optical, magnetic, or conductor tracks that represent the individual bits. Incremental encoders generally have finer resolution, in a given technology, than do absolute versions. Magnetic scales are somewhat more rugged than optical scales, because high-resolution optical scales are made from glass. Typical encoder applications include process machinery feedback and control, robotics, and measuring equipment. Equipment using an incremental encoder will often have a zeroing function when the machine is first turned on, and additional rezeroing cycles at opportune times during operation. This is to avoid, as much as possible, prolonged error of the position count if it becomes corrupted by erratic motions or externally induced noise.