Chapter 7

Optical Encoders

7.1. INTRODUCTION

As referred to in this discussion, an encoder is an electromechanical device used to monitor the motion or position of an operating mechanism, and to translate that information into a useful output. The output can take the form of a simple system status indication, or it can provide feedback control information or in other ways interface with related devices.

There are several position sensing techniques available to the system designer. The most widely used types of sensors fall into one of the following types: capacitive, magnetic, contact or optical. This discussion is confined to the latter type: linear and rotary optical encoders. Since optical encoders offer significant improvements over earlier and less sophisticated methods of position sensing, it is hoped that the reader will find this information to be useful in applying optical encoding techniques to new equipment designs, as well as to the upgrading of existing equipment.

7.2 OPTICAL ENCODER COMPONENTS

Typically, an optical encoder consists of four components: (1) a light source, which can be an incandescent lamp or light-emitting diode (LED), depending on design considerations; (2) a pattern of alternating opaque and translucent segments appearing on an assembly (disc or scale) that stands between the light source and its associated light sensor, and which usually tracks the movement of the device being monitored; (3) the light sensor, often a phototransistor, but possibly a photovoltaic component (solar cell); and (4) conditioning circuitry such as may be required to convert sensor output to properly formatted information for the desired readout or interface. The interrelationship of these components is shown in Figure 7.2.1, which is a functional sketch of the elements of a simple rotary optical encoder.

![Figure 7.2.1. Components of an Optical Encoder](image)

For increased resolution, the light source is collimated, and an additional element — the mask — is added between the disc and sensor. In such applications, the mask and disc produce a shuttering effect so that only when the translucent segments of
both are in alignment light is permitted to pass through to the sensor. This configuration is shown, also using a rotary encoder example, in Figure 7.2.2.

Figure 7.2.2. High Resolution Optical Encoder

7.3. TYPES OF OPTICAL ENCODERS

Two basic types of optical encoders are described here: linear and rotary. The distinction is simple: with the linear encoder, direct digital information can be obtained regarding the position and/or velocity of an arm moving along a linear axis. Rotary encoders are those designed to sense the movement and/or position of devices which rotate about an axis. By far the more common of the two, rotary encoders are either incremental or absolute in encoding function, as described below; and are available both as self-contained discrete components and as modules for integration into related end products.

7.3.1. INCREMENTAL ROTARY ENCODERS

Incremental encoders provide efficient and economical means for obtaining position or velocity data in many applications. This results from the inherent simplicity of the incremental encoder, as compared to the absolute. With only one or two output channels, the incremental has fewer components and less complex code discs. Hence, the cost of a standard incremental encoder is typically 60 percent less than that of an absolute with equivalent capabilities.

Functionally, the incremental encoder generates a serial pulsetrain output by reading the number of increments traversed on the code disc track as the encoder shaft rotates. In order to know the exact amount of travel, it is necessary to count the output pulses with an external digital counter. The counter then maintains a memory of the position information for the encoder.

Should a power failure occur, the counter can reset and lose track of the position. An index pulse, once per revolution, can be used to recalibrate the counter after a power failure. For a fail-safe system, a battery backup system should be provided for the counter and encoder; or else an absolute encoder should be used. The determining factor on whether to use an absolute versus an incremental encoder lies in the cost trade-offs of the system.

Incremental encoders are used exclusively where velocity is the desired parameter. Velocity sensing functions include tachometers and velocity-feedback systems, as utilized in the tape transport industry. In these cases, a modular rotary encoder is normally mounted directly on the motor shaft. Incremental encoders are also used
widely in many position-determining applications. A typical position-sensing application is provided by the machine tool industry, where the position of the tool with respect to the work piece must be constantly observed. The rotary encoder gives a number of cycles of output for each inch of tool travel. A typical 2500 cycle encoder with 4X multiplication provides measurement tolerances in the order of .0001 inch per inch of travel.

There is a variety of standard incremental encoder configurations for applications requiring resolutions up to 5000 cycles per revolution. Standard encoders can be provided with internal electronics for interfacing with any logic type (CMOS, TTL, DTL, etc.). Certain configuration sizes also permit the inclusion of multiplication and direction-sensing electronics. Additionally, incremental encoders can be provided with either a single channel output or with any of the following output combinations:

a. Dual output from a single count, with each output phased by 90 degrees (phase quadrature).

b. Single output plus once per revolution pulse (zero reference or index).

c. Dual output plus index.

d. All of the above options taken from two different tracks...as with the English/Metric encoder, which provides outputs both in English and Metric (e.g. 1000/2540), either individually or simultaneously.

7.3.2. ABSOLUTE ROTARY ENCODERS

Absolute or "whole word" encoders generate a unique digital parallel output for every shaft position. This absolute positional information is normally expressed in code formats such as binary, BCD, Gray, XS3, XS14, etc. Each bit in the digital word represents an independent track on the encoder disc, and an independent output channel with its associated electronics. Special codes for various purposes are routinely created.

Since it is not uncommon for an absolute encoder to have 13 or more output channels, as compared with the normal one or two in the incremental encoder, the higher cost of the absolute is apparent. For many applications, however, this higher cost is clearly justified. The positional information is read from the disc in an absolute encoder and from an external counter in an incremental encoder. Therefore, the information read from an absolute encoder is unaffected by power failure or noise.

Absolute encoders are normally used for generating positional feedback information. One example of usage is for monitoring the distribution of water through various canals and reservoirs in water conservation programs. Despite higher cost, absolute encoders are warranted for such applications, not only because of their inherent fail-safe features, but also because they afford a savings in power consumption. Absolute encoders need only be powered when information is required.
7.3.3. LINEAR ENCODERS

With a linear encoder you can discern the position or velocity of an arm moving along a linear axis. The encoder eliminates the need for a mechanical conversion, and thereby allows resolutions and accuracies that are unobtainable with a rack and pinion in the system.

Linear encoders operate on the same basic principle as the rotary encoders: the “transmissive optical” principle. A glass scale with light and dark increments is affixed to the arm in motion; the scale moves within a yoke-shaped read head that contains a source of illumination on one side of the scale and a sensing assembly on the other. Light passing through the scale is sensed by photodetectors, which generate the signal output.

The arm in motion can have any conceivable shape, configuration, length, depending upon the function being performed. There are as many different mechanical configurations of linear encoders as there are applications.

Linear scales can be designed for absolute as well as incremental encoding functions, in lengths ranging from less than an inch up to several feet. Resolution in linear encoders is expressed by the displacement of lines per inch, where one “line” equals an opaque segment on the scale and its adjacent clear segment. Scales have been designed for precision inspection equipment with resolutions of 10,000 lines per inch that are accurate to 50 millionths of an inch. A summary of the most common applications for linear encoders is given in Table 7.3.1 below. Figures 7.3.1 through 7.3.4 depict some of these applications.

<table>
<thead>
<tr>
<th>APPLICATION</th>
<th>USAGE</th>
<th>TYPICAL RESOLUTION AND ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disc Memory System</td>
<td>Position feedback information for locating the read heads over the respective tracks on the memory disc. Indicates the position of the tool with respect to the work piece. Indicates position and bypasses the lead screw.</td>
<td>Resolution: 100-200 lines per inch. Accuracy: 50 to 100 millionths.</td>
</tr>
<tr>
<td>Machine Tools</td>
<td></td>
<td>Resolution: 500-2500 lines per inch. Accuracy: 100 to 500 millionths.</td>
</tr>
<tr>
<td>Coordinatographs and Plotters</td>
<td></td>
<td>Resolution: 10,000 lines per inch. Accuracy: 50 to 100 millionths.</td>
</tr>
<tr>
<td>Inspection Equipment</td>
<td>Precision measurement with visual display as applied to height gauges, comparators, etc.</td>
<td>Resolution: 10,000 lines per inch. Accuracy: 50 to 100 millionths.</td>
</tr>
</tbody>
</table>

Table 7.3.1. LINEAR ENCODER APPLICATIONS
Figure 7.3.1. Disc memory type of linear head offers misaligned features, and provides 100 or 200 tracks per inch, and indications of BOT (beginning of travel) and EOT (end of travel).

Figure 7.3.2. This example of a linear head used for disc memory applications is similar to the head shown in Figure 7.3.1, offering 200 tracks per inch.

Figure 7.3.3. Another disc memory application shows a linear head and scale to be aligned by the user when the unit is integrated into an end product. This encoder offers 400 tracks per inch, plus BOT and EOT.

Figure 7.3.4. The head and scale used on coordinateographs provide 1000 lines per inch. Such units can be butted together, thereby providing readout over several feet.
7.3.4. MODULAR ENCODERS

Modular encoders are offered to enable a manufacturer to build an encoder into his product, rather than couple an external encoder onto the shaft. Modular encoders are provided as a sub-assembly, with two basic elements: 1) code disc and hub, 2) housing, sensor, light source and electronics. The customer then assembles and integrates these elements directly onto the shaft within the casement of his end product. The modular elements are prealigned in order to minimize the time and expense required of the customer for the final integration.

Applications which lend themselves best to the modular concept generally fall into three basic categories: (1) where coupling one shaft to another might introduce torsional resonance, create errors or in other ways detract from system performance; (2) where lowest possible cost is essential; and (3) where encoders are an integral part of the motor housing or case.

Most modular encoders are incremental, but are also available as absolute.

7.4. SPECIAL DESIGNS & OPTIONS

Encoders for unique applications fit basically into three categories of design: 1) altitude reporting encoders, 2) multiturn encoders, 3) high-resolution encoders.

7.4.1. ALTITUDE REPORTING ENCODERS

These encoders are designed to be incorporated into the final assembly of altimeters. They derive motion through gears which are driven by the aneroid or by the servo. As the disc rotates, the differential pressure resulting from variations in altitude is converted into a digital output. The code is ICAO, which relates directly to altitude in increments of 100 feet. Encoders are supplied to the altimeter manufacturer as a modular subassembly. Figure 7.4.1 shows an example of this type.

Figure 7.4.1. An Altitude Reporting Encoder Disc

7.4.2. MULTITURN ENCODERS

Where high-bit capacity is required for defining long distances of motion, multiturn encoders make use of two or more discs mounted on a gear train. The gear ratio between the input and output discs dictates the total number of revolutions of the input shaft required to generate the full capacity output. For example, assume the input disc is a 100 count unit, and that the output disc is a 1000 count unit; also, the reduction gear ratio between input and output is 1000 to 1. The reduction ratio causes the output disc to advance one count every time the input shaft makes one complete revolution (100 counts). Therefore, 1000 revolutions of the input
are required for one revolution of the output. The total capacity is thus the product of both discs, which equals 100,000. By changing the disc resolutions and the gear ratios, multiturn encoders can provide a variety of count ranges.

7.4.3. HIGH RESOLUTION ENCODERS

For special requirements, high resolution encoders are available, such as Size 40 (four inch diameter) with resolutions up to 36999 in 8-4-2-1 BCD. This encoder can be modified to provide 16 bits in natural binary or Gray code.

7.5. ENCODER DISCS

The disc determines the resolution and accuracy of the encoder. The purpose of the disc is to operate as a rotating shutter, alternately blocking and transmitting light from its source to a light sensor. The disc is designed with an arrangement of opaque and transparent segments, and may be manufactured with any material that lends itself to this criterion. Figure 7.5.1 presents some examples of discs.

Discs are made from plastic, mylar, glass or metal. The superior properties of glass, in the areas of stability, rigidity, hardness and flatness, afford it as the most practical material for disc production, although the other materials have a higher shock and vibration resistance than glass. The opaque segments can be photo emulsion or metalized. The metalized glass has a harder surface than photo emulsion, and is more resistant to harsh environments.

Figure 7.5.2 illustrates two code discs: incremental and an absolute. Note that the incremental code (at left in the figure) is a single track. Provisions must be made to count the number of transitions to determine the position or velocity information. The absolute code represents a 4 bit natural binary. The 4 bits permit 16 discreet positions. Each position has a unique binary word; i.e., position 7 is represented by the word 0111.

![Figure 7.5.1. Typical Encoder Discs](image1)

![Figure 7.5.2. Incremental (left) and absolute code discs.](image2)

7-7
7.6. DIGITAL CODES

With the large number of applications requiring the interfacing of mechanical devices and electronic controllers, the need for converting mechanical position into an electronic signal is increasingly being solved by optical encoders. Depending on the applications, various methods of describing or coding an analog value into a digital pattern have been devised.

An absolute code makes use of one track on the disc for each bit; therefore, a 4 bit encoder in natural binary has 4 tracks, an 8 bit encoder disc has 8 tracks, and so on. The track having the greatest number of opaque and translucent signals is considered to be the LSD — least significant digit. The track having the least number of signals is known as the MSD — most significant digit. The tracks provide information which is weighted in an increasing importance from the LSD to the MSD. In the natural binary code pattern, when changing from the decimal number 7 to number 8 and from number 15 to number 0, all of the bits change value from logic 1 to logic 0 (or vice versa).

The least number of channels or bits are required with the natural binary code. Each bit represents a power of 2, such as 1, 2, 4, 8, 16, etc. The analog value is represented by binary bits: "1" representing a transparent segment and "0" representing the opaque segment. When the values of each bit are added together, the value of the analog number is obtained. For example, as shown below, the analog number "178" is represented by:

\[
\begin{align*}
128 & \quad 64 & \quad 32 & \quad 16 & \quad 8 & \quad 4 & \quad 2 & \quad 1 & \quad (\text{bit value}) \\
1 & \quad 0 & \quad 1 & \quad 1 & \quad 0 & \quad 0 & \quad 0 & \quad 1 & \quad (\text{state of each bit}) \\
128 + 0 + 32 + 16 + 0 + 0 + 2 + 0 = 178
\end{align*}
\]

Since values or weights can be assigned to each binary digit or "bit", and then added to convert the analog value they represent, the natural binary code is a weighted code. And, since several bits may change state at a transition between two numbers, the natural binary code is also a polystrophic code. Some example analog numbers are shown here with their binary equivalents. 1 = 1, 2 = 10, 3 = 11, 4 = 100, 5 = 101, 6 = 110, 7 = 111, 8 = 1000 and, as above, 178 = 10110010.

Polystrophic codes are so known because two or more bits may change at a transition between adjacent words. Figure 8.6.1 shows a natural binary count of 16 code (zero through 15), and could be read with four in-line sensors. Due to mechanical tolerances, there will be a finite timing error, or ambiguity, among the four sensors. Techniques to compensate for this ambiguity include "V" scan and "U" scan sensor arrangements. In Figure 7.6.1 the sensors form a "U" shape; hence the term U scan.

![Figure 7.6.1. U Scan Arrangement](image)
Figure 7.6.2 represents the waveform from the U scan sensors. Note the advanced and retarded wave forms of the two sensors for track 2. When these two wave forms are gated by the least significant track, a waveform is generated whose transitions have zero time errors. The logic gating to generate this decoded track can be expressed at \((T \cdot 2A + 1 \cdot 2R)\) = track 2 decoded.

**WAVEFORMS**

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Waveform</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;1&quot;</td>
<td>RARARARARARARAR</td>
</tr>
<tr>
<td>Sensor 2A</td>
<td>RARARARARARARAR</td>
</tr>
<tr>
<td>Sensor 2R</td>
<td>RARARARARARARAR</td>
</tr>
<tr>
<td>Sensor 4A</td>
<td>RARARARARARARAR</td>
</tr>
<tr>
<td>Sensor 4R</td>
<td>RARARARARARARAR</td>
</tr>
<tr>
<td>Sensor 8A</td>
<td>RARARARARARARAR</td>
</tr>
<tr>
<td>Sensor 8R</td>
<td>RARARARARARARAR</td>
</tr>
</tbody>
</table>

"2" Decoded: RARARARARARARARARARAR
"4" Decoded: RARARARARARARARARARAR

Figure 7.6.2. U Scan Sensor Waveform

The same logic expression is used to decode tracks 4 and 8. For encoders with higher resolution, the sensors may be arranged in a "V" pattern. With this V Scan technique each decoded track is used to decode the next significant track. The V Scan technique simplifies mechanical alignment of the sensors.

A conversion of digital to decimal form can be simpler if a binary number is used to represent each decimal digit. This is known as Binary Coded Decimal system (BCD) or the 8421 code. The digits 0 - 9 are natural binary numbers. However, ten is represented as 0001-0000 rather than 1010 as above. Thus, "178" is represented as 0001-0111-1000 in BCD format. More bits or tracks are required if such a code is used to encode a shaft position; however, the output is in BCD.

Gray code, being monostrophic, has the advantage of requiring only a single bit to change from any one word to adjacent word. The conversion from Gray Code to natural Binary code is fairly simple, requiring one exclusive-or gate. Decimal, Gray Code and Binary equivalents are shown below:

<table>
<thead>
<tr>
<th>DECIMAL</th>
<th>GRAY CODE</th>
<th>BINARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0000</td>
<td>0000</td>
</tr>
<tr>
<td>1</td>
<td>0001</td>
<td>0001</td>
</tr>
<tr>
<td>2</td>
<td>0011</td>
<td>0010</td>
</tr>
<tr>
<td>3</td>
<td>0010</td>
<td>0011</td>
</tr>
<tr>
<td>4</td>
<td>0110</td>
<td>0100</td>
</tr>
<tr>
<td>5</td>
<td>0111</td>
<td>0101</td>
</tr>
<tr>
<td>6</td>
<td>0101</td>
<td>0110</td>
</tr>
<tr>
<td>7</td>
<td>0100</td>
<td>0111</td>
</tr>
<tr>
<td>8</td>
<td>1100</td>
<td>1000</td>
</tr>
<tr>
<td>9</td>
<td>1101</td>
<td>1001</td>
</tr>
<tr>
<td>10</td>
<td>1111</td>
<td>1010</td>
</tr>
<tr>
<td>11</td>
<td>1110</td>
<td>1011</td>
</tr>
</tbody>
</table>

Many tradeoffs must be made in deciding which of the codes (monostrophic or polystrophic) should be directly read from the code disc. In most cases, a monostrophic code pattern on the disc will be used and convert through digital logic to a polystrophic code if deemed necessary for the application.
The following codes are commonly used in building encoders.

**MONOSTROPHIC**
- Gray Code
- ICAO (altitude report code)
- XS-BCD (i.e. XS3, XS3/XS14, etc.)
- Biquinary (2 out of 5 code - Johnson code)

**POLOSTROPHIC**
- Natural Binary
- 8421 BCD

The Excess 3 (XS3) Code is sometimes used when simple arithmetic circuits are used to perform subtraction by the 9's compliment. If all bits are complimented or inverted in the Excess 3 Code, the 9's compliment is easily formed.

**DECIMAL/EXCESS 3**

<table>
<thead>
<tr>
<th>DECIMAL</th>
<th>EXCESS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0011</td>
</tr>
<tr>
<td>1</td>
<td>0100</td>
</tr>
<tr>
<td>2</td>
<td>0101</td>
</tr>
<tr>
<td>3</td>
<td>0110</td>
</tr>
<tr>
<td>4</td>
<td>0111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DECIMAL</th>
<th>EXCESS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0010</td>
</tr>
<tr>
<td>1</td>
<td>0010</td>
</tr>
<tr>
<td>2</td>
<td>0010</td>
</tr>
<tr>
<td>3</td>
<td>0010</td>
</tr>
<tr>
<td>4</td>
<td>0010</td>
</tr>
</tbody>
</table>

Some devices require the *direct decimal code*.

<table>
<thead>
<tr>
<th>DECIMAL</th>
<th>EXCESS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0011</td>
</tr>
<tr>
<td>1</td>
<td>0100</td>
</tr>
<tr>
<td>2</td>
<td>0101</td>
</tr>
<tr>
<td>3</td>
<td>0110</td>
</tr>
<tr>
<td>4</td>
<td>0111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DECIMAL</th>
<th>EXCESS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0010</td>
</tr>
<tr>
<td>1</td>
<td>0010</td>
</tr>
<tr>
<td>2</td>
<td>0010</td>
</tr>
<tr>
<td>3</td>
<td>0010</td>
</tr>
<tr>
<td>4</td>
<td>0010</td>
</tr>
</tbody>
</table>

Numerous other codes serve special purposes including error correcting (Hamming) codes, ICAO code used for automatic aircraft altitude reporting, and others.

Natural Binary and Gray Codes require the least number of tracks. The justification for the additional tracks needed for BCD is the ease of its application in interfacing with decimal display or BCD systems.

**7.7. MECHANICAL CONSIDERATIONS**

To the end user, the optical encoder is primarily a device designed to provide an electrical signal. The manufacturer views the optical encoder as an opto-mechanical assembly from which rotational or linear position, speed and/or acceleration can be determined. The optical encoder, as an opto-mechanical device, is only as good as the cumulative effects, or limits, of error that all the mechanical components of the assembly might exhibit. Such parameters as speed, slew rate, torque, inertia, size and weight, runout, shaft loading, eccentricity and end-play, are of major importance.
7.7.1 SIZE AND WEIGHT

Except for aerospace applications, encoder weight is rarely of any concern. However, size can significantly limit encoder resolution. Figure 7.7.1. depicts the relationship of counts/revolution to code pattern centerline diameter and window opening (each adjacent translucent and opaque section on the code provides one complete cycle.)

\[
\text{Resolution} \equiv N = \frac{\pi D}{2X}
\]

Where:
\[
D = \text{Code Pattern Centerline}
\]
\[
X = \text{Window Opening}
\]
\[
N = \text{Counts/Revolution}
\]

Figure 7.7.1 Counts per revolution vs. centerline diameter and window opening.

7.7.2. SLEW SPEED AND SLEW RATE

Angular speed of the encoder disc, the code pattern diameter and segment size are related parameters defining the frequency of a generated output signal from an optical encoder. The linear speed of a point on a disc equals the angular velocity \(\omega\) times the radius \(R\) of the point in question. It is this linear speed, divided by the segment cycle, that generates the output frequency of the encoder, as shown in Figure 7.7.2.

\[
\text{Frequency} = \frac{\text{WN}}{2\pi} = \frac{n \text{ (RPM's)}}{60}
\]

Thus, the frequency output of the optical encoder is generated by the angular speed of the disc.

Slew rate is the maximum velocity at which an encoder will be required to perform. A typical example is tape rewind in a tape transport system. Slew speed is an indicator of how fast the encoder may be rotated without introducing conditions detrimental to the bearings or other components of the mechanical assembly. As slew speed becomes an important factor, it can be enhanced by reducing the preload on the bearings and by using a high grade type of bearing in the assembly.
7.7.3. TORQUE, INERTIA
AND ACCELERATION

Torque, inertia and angular acceleration are equated in rotational motion by \( T = J\alpha \) where \( T = \) torque, \( J = \) inertia and \( \alpha = \) acceleration. This represents the interdependency of available torque of the driving motor, the desired system acceleration (or deceleration) and the added system inertia of the optical encoder. Acceleration is usually not a major concern in optical encoders. However, in extremely high acceleration applications, bonding strength of disc to hub and material stresses must be adequate to withstand the higher torques.

Friction torque must be considered as either starting torque or running torque, which are analogous to static friction and dynamic friction. Starting and running torques are greatly affected by the encoder requirements, and are contributed by not only the code disc, but also by the quality of bearings, bearing preload and bearing seals. Where extremely small, low inertia DC motors are employed in servo systems, encoder torques can be greatly reduced.

However, this torque reduction has trade-off penalties. Reducing bearing preload and/or bearing sealing requirements will reduce torque, but will also increase radial and axial endplay, and will permit greater particle contamination in the bearing.

Disc inertia may be of paramount importance in some systems. In high acceleration or velocity servos where fast access, over-shoot or settling time are significant, low inertia discs are mandatory. As resolution requirements also increase, larger discs may be necessary. And, since inertia increases by the 4th power of the disc radius, many trade-offs have to be considered. Inertia varies in direct proportion to the mass of the disc. Therefore, certain disc material trade-offs must be considered. Mylar and photoplast discs, while lighter in weight than glass, may present other mechanical difficulties. the best results have been obtained with very thin glass discs.

7.8. MANUFACTURING AND
HANDLING

Manufacturing techniques and assembly sequence will vary considerably, depending on the application the encoder is designed to fulfill. The variables are dictated by the options required to meet the expected encoder performance, environmental conditions, and reliability.

7.9. ENCODER RELIABILITY AND
DESIGN CONSIDERATIONS

Encoder failure can be divided into four areas: (1) light source failure, (2) bearing failure, (3) electronics failure and (4) environmentally caused failure.

The lifetime of an encoder’s light source depends on the type used, vibration level, power supply, and ambient temperature. Incandescent lamps can withstand higher temperatures, while LED’s withstand
higher vibration levels and have a longer average life. Power supply regulation has a large effect on incandescent lamp life, as filament life is inversely proportional to the twelfth power of the applied voltage.

\[ t_2 = t_1 \left( \frac{1}{v + \Delta v} \right)^{12} \]

- \( t_1 \) = Lamp life at voltage \( v \)
- \( t_2 \) = Lamp life at voltage \( (v + \Delta v) \)

Thus, a one percent increase in filament voltage will shorten lamp life by eleven percent. When incandescent lamps are used in encoders they are derated and burned in for approximately 300 hours, to maximize life and stabilize output. Thus, average filament lifetimes of 50,000 to 100,000 hours are possible.

To minimize error caused by axial and radial play in the encoder shaft, the bearings are usually class 7, and are preloaded near the recommended limit. However, excessive preload is the usual cause of early bearing failure. Thus, the bearing housing used in encoders must be carefully designed to obtain maximum life without excessive wear.

The environment has a large impact on encoder design. Potential humidity, temperature or dust contamination may require special consideration in the specification of encoders.

Where seals are necessary, encoders employ an O-ring seal between the outer cover and bearing assembly. Condensation or liquid splashing usually requires the use of sealed bearings. By incorporating O-ring shaft seals, low rpm encoders can be made submersible. Where both humidity and temperature must be tolerated, metalized discs and masks are usually used instead of photographic emulsions.

To achieve high resolution and accuracy, the mask and disc surfaces are usually spaced .001 to .006 inches apart. Any emulsion swelling due to high temperature, 120°F or above, or high humidity with possible condensation, or dust contamination may limit the life of the encoder.

The electronic circuits used in most encoders are generally very reliable and normally outlast the light source and bearings in encoders. Two exceptions are: In very high temperature environments above 125°C, and with complex encoders using several integrated circuits. Early failure of integrated circuits or "infant mortality" can be as high as .5 or 1% during the first one thousand hours if commercial non-stress-tested devices are used. Thus, if an encoder design uses eight IC's, as many as 4% of the encoders may fail during the first 1,000 hours due to "infant mortality." Simple burn-in or artificial aging procedures can reduce this rate from .5% to .1%.

Design considerations as related to encoder life are summarized in Table 7.9.1, following.
<table>
<thead>
<tr>
<th>ENCODER COMPONENT OPTION</th>
<th>ADVANTAGE</th>
<th>DISADVANTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incandescent Lamp</td>
<td>High light output. Resistant to high temperatures.</td>
<td>Filament sensitive to vibration. Requires lens for collimation.</td>
</tr>
<tr>
<td>L.E.D.</td>
<td>Long life. Simple to install. Requires less current</td>
<td>Less light output. Usually infrared cannot be visually observed. Limited to about 80°C. Higher cost than incandescent.</td>
</tr>
<tr>
<td>Silicon Photo Voltaic Cell</td>
<td>Less expensive. Resistant to temperatures to 150°C.</td>
<td>Low cost prevents hermetic sealing.</td>
</tr>
<tr>
<td>Photo Transistor</td>
<td>Higher sensitivity; can be hermetically sealed. Higher output.</td>
<td>Limited to upper temperature limit. Higher cost.</td>
</tr>
<tr>
<td>Photographic Disc Emulsion</td>
<td>Less expensive</td>
<td>Only moderately hard. Not resistant to high humidity. Limited to 120°C.</td>
</tr>
<tr>
<td>Metal on Glass</td>
<td>Resistant to high temperature, humidity &amp; scratching.</td>
<td>Expensive.</td>
</tr>
<tr>
<td>Mylar Disc</td>
<td>Low inertia.</td>
<td>Limited disc size.</td>
</tr>
<tr>
<td>Photoplastic Disc</td>
<td>Low inertia. High resistance to shock.</td>
<td>Less rigid disc.</td>
</tr>
</tbody>
</table>
### 7.10. ACCURACY AND ERRORS

The total accuracy of an encoder—incremental or absolute—is affected by the combined error potential of the encoder/system configuration. This error can be summarized as follows:

\[
\text{Total Error} = \text{encoder error} + \text{system interaction errors}
\]

#### 7.10.1 ENCODER ERROR

Encoder error is the aggregate of quantization error, instrument error and signal processing errors, including generation and transmission errors.

**7.10.1.1. Quantization Error** — Some quantization error is present in any digital device. In the case of optical encoders, it is determined by the intricacy of the pattern printed or deposited on the disc or scale, which in turn is dictated by desired accuracy, output requirements and other interfacing provisions. As shown in the figure below, quantization error will be a maximum of \( \frac{1}{2N} \), where \( N \) is the number of lines in one disc revolution (or for linear types, the number of lines in one scale length). Thus, the accuracy of position readout is limited by this type of error.

\[
\text{Quantization error} = \frac{1}{2N} \quad \text{maximum}
\]

#### 7.10.1.2. Instrument Error — is caused by the following defects; the total is usually of the same order of magnitude as the quantization error.

1. Assembly error and manufacturing mechanical tolerance buildup
2. Mask or disc misalignment
3. Mechanical vibration
4. Temperature variation
5. Electrical noise
6. Faulty power supply regulation

The limiting errors for rotary encoders are usually eccentricity, disc pattern variation and runout on the disc face. These mechanical variations will produce “jitter” or “flutter”. These are partly interdependent. Jitter will always be increased by disc eccentricity, but is also caused by inaccuracies in the photo reduction process.

**Eccentricity** is caused by radial bearing play, inaccurate dynamic centering of the disc pattern on the hub and by excessive clearance between the disc hub and shaft. Eccentricity produces several errors in the signal output, noted as follows.

**Amplitude Modulation** — if the encoder output is an analog or quasi sine wave, eccentricity will produce an amplitude variation such as shown below.

\[
\frac{E_{\text{max}} - E_{\text{min}}}{E_{\text{max}}} = \frac{\Delta R}{R + \Delta R}
\]

\( R \) = track radius

\( \Delta R \) = Eccentricity error

7-15
Frequency Modulation — usually appearing as jitter on the oscilloscope, is one component of observed jitter, depicted below.

\[
\frac{F_2 - F_1}{F_1} \left( \frac{\Delta R}{R} \right) \times 100\% = \% \text{ Flutter}
\]

Inter-channel Jitter — if the optical sensors for the two channels are separated by angle \( \theta \), eccentricity will produce additional jitter of an amount equal to:

\[
\% \text{ Inter-channel Jitter} = \frac{\Delta R}{R} \frac{N}{2\pi} \frac{\theta}{180^\circ} \times 100\%
\]

Where

- \( N \) = number of lines on disc
- \( \theta \) = angular separation in degrees of sensors
- \( \frac{\Delta R}{R} \) = \% eccentricity

Oscilloscope traces of the output of Tracks A and B depicted above, would appear as shown:

Angular Error — is the difference between the angle indicated by the encoder and the actual shaft angle, due to imperfect centering of the disc pattern with respect to the axis of rotation.

\[
\text{Angular error} = \arcsin \left( \frac{R_{\text{max}} - R_{\text{min}}}{R_{\text{min}}} \right)
\]

Where \( R_{\text{max}} \) = maximum distance from the code pattern to the axis of rotation

\( R_{\text{min}} \) = the minimum distance

For example, encoder discs are usually centered by either optical or electronic means. Typical centering accuracy is \( \pm 5\% \) of one line width. Thus, for \( R = 1 \) inch, \( N = 1000 \), eccentricity error will be approximately \( 3 \times 10^{-4} \) inches.
If this is added to the eccentricity caused by shaft runout of .0002 in., and a hub/shaft clearance of .0002 in., total worst case eccentricity is .0007 in. Thus, a 1000-count, 2-inch diameter encoder could have .07% amplitude modulation, which would seldom be significant. Assuming .07% frequency modulation, and angular error of .04°; and if angular separation between tracks A and B is 20°, then Jitter = 1.2%.

7.10.1.3. Signal Processing Errors

Signal Generation — for most encoder applications, the electronic circuitry for signal generation senses the point at which the voltage output crosses through zero. Thus, any shift in the zero crossover can be an error. Besides eccentricity and disc runout discussed earlier, gap variation, excessive mask-disc gap, and temperature drift can cause crossover error. Zero crossover error appears as a constant, as a random variation or as a rapid periodic error (jitter).

As shown in the figures below, a single light sensor cannot produce a zero voltage output without an offset bias due to non-collimated light, scattering, finite mask-disc gaps, and mask misalignment.

Although a bias voltage can be used to cancel the DC offset of a single sensor, a small gap variation or temperature change would cause the DC offset to change; therefore, causing a change in the zero crossovers.

The use of two sensors per track output will nearly cancel this drift if the sensors are operated back-to-back or "push-pull". Thus, two masks and two sensors per track
are used with the masks arranged to produce signals with 180° phase difference from the sensors.

Thus, the modulations and drift in the zero crossover are minimized by the use of two sensors per track output. This error cancellation requires closely matched sensors for output, temperature tracking, and high frequency rolloff. Other errors reduced by "back-to-back" sensors are changes in light excitation and errors due to poor voltage regulation.

Signal Transmission — problems encountered in transmitting the encoder signal to the receiving electronics are usually those of electrical noise or signal distortion. These could result in gain or loss of counts, and hence incorrect data on the performance of the device being monitored.

Electrical noise can be minimized by the following techniques: (1) using shielded cable, (2) installing a low value pull-up resistor, typically 100 ohms, (3) use of a line driver and (4) use of a differential line driver. In the latter case, a differential line driver is used where long interconnections exist. A shielded twisted pair cable and mated line receiver are used.

7.10.2 SYSTEM INTERACTION ERRORS

System interaction errors are the result of the effects of gears, racks and pinions and drive belts used to couple the encoder with the system. Also, in the case of linear encoders, misalignment can cause significant cosine errors.