Encoder Feedback Devices: Basics for Motion Control Engineers

From Battlebots to the Mars Rover, large construction vehicles to semiconductor manufacturing equipment, wherever there is motion, an encoder is sure to be found. Encoders may not be as glamorous as the products they go in, but they are critical in today's complex and sophisticated motion control systems. The number of applications that require encoders are very large and can be quite different from each other, and as a result, the types and styles of encoders available are just as numerous. It can be quite daunting for the design engineer to choose just the right encoder for their application. This paper looks at some of the features that can be specified for an encoder and how that will determine where it can best be used.

Mechanical Configuration

Encoders can be split into one of two mechanical configurations: linear or rotary. As the name suggests, the linear encoder measures linear motion or speed. Typical linear applications are the control of linear motors or X-Y tables such as those found on vertical CNC mills. They are also commonly found in measurement instruments such as metrology instruments or digital calipers. A rotary encoder, on the other hand, is used to measure rotation; some typical examples are on a motor shaft for speed control or on a PV panel for solar tracking. This paper will focus on the more common type: the rotary encoder.

Within rotary encoders, there are further basic differentiations on mechanical configurations, which are primarily a function of the end application of the encoder. There are two primary categories: industrial and motor feedback automation. Industrial encoders are used for rugged and extreme working conditions found in heavy manufacturing conditions. Extreme temperatures, significant vibration, dirt and debris are some of the challenges that these encoders have to survive on a daily basis. As a result, industrial encoders tend to be big and solidly built to withstand the tough environment.

US Digital EC35 Commutational Kit Encoder

Motor feedback encoders are typically presented in two further distinctions, encoders with bearings, and modular encoders which use an existing bearing set like is found on the tail shaft of a servo motor. The determining factor of which motor feedback encoder to utilize is a function of the stability of the shaft/bearings to which the encoder is attached. Motor feedback encoders with bearings are typically utilized when the application shaft has a significant amount of axial or radial run out. This style of encoder will incorporate some sort of flexible member, either a flexible shaft coupling or flexible body mount member, to allow mechanical compliance with the application shaft irregularities and run out.

Modular encoders rely on a mechanically stable application shaft as the shaft is responsible for holding a rotating code wheel in a precise location relative to the encoder's sensing element, as is the case with a tail shaft of a high quality servo motor. Motor manufacturers put a considerable amount of effort in designing a motor with a very stable shaft/bearing assembly which in turn allows them to provide a high performance motor. Modular encoders take advantage of the stable motor tail shaft assembly provided by manufacturers and are assembled from components supplied by the encoder manufacturer to the tail shaft and motor end bell.

The primary difference between the two motor feedback encoders is that the modular encoder requires the user to supply labor for assembly to the application shaft. However, the modular design does not add the expense of the mechanical components associated with an extra set of bearings in a motor feedback encoder. If the user is able to apply a modular encoder they will have the most cost effective motor feedback solution. In both cases, motor feedback encoders are generally less environmentally sealed when compared to the industrial style encoder.

Optical and Magnetic Encoders

Linear and rotary encoders may use optical or magnetic to sense movement. Optical encoders are more common and provide higher accuracy and resolution. Magnetic encoders have an advantage in that they do not need to be in as tightly controlled of an environment as optical encoders, and can be used in certain environments that have higher humidity, dust, et cetera. Magnetic encoders may also operate in various fluid environments.

Optical encoders use an optical sensor to detect light that is transmitted through or reflected from a disk (also known as a code wheel) whose pattern has both transparent and nontransparent lines. When the light is received by the sensor, the encoder puts out a high signal and conversely when the light is blocked by a line on the code wheel, the sensor puts out a low signal. So with a known pattern on the disk, the distance moved or speed of movement can be measured by the encoder.

Magnetic encoders use magnetic code wheels where magnetic poles are separated according to the resolution required. A magnetic sensor detects the change in the magnetic field and produces digital pulse trains from the encoder. Magnetic encoders use less power than their optical counterparts but struggle to provide the same resolution or positional accuracy of an optical encoder due to inherent nonlinearities in the magnetic field.

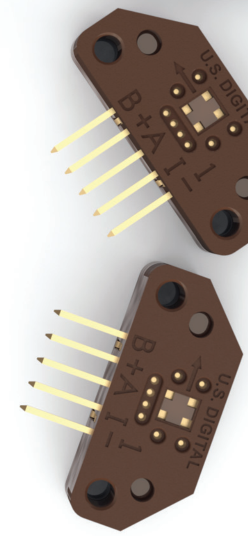
▲ US Digital S4 Miniature Optical Shaft Encoder (left) and E4P OEM Miniature Optical Kit Encoder (right)

An example of when a magnetic encoder is applied in place of an optical encoder is when there is a chance for an optical disk to become “fogged” as moisture condenses on the code wheel. Consider an application where the encoder is held at a very low temperature and then the ambient temperature quickly increases. In most cases this quick temperature change causes condensation on all surfaces of the encoder, including the optical code wheel. When the code wheel surface collects droplets of moisture, the light transfer of the code wheel image to the optical sensor becomes disrupted and a false or missing signal may occur on the output. With a magnetic encoder design, condensation of moisture is not an issue with the rotating magnet and magnetic sensor.

▼ US Digital HD25 Industrial Rugged Metal Optical Encoder

US Digital E5 Optical Kit Encoder (cover removed) ▼

▲ US Digital MA3 Miniature Absolute Magnetic Shaft Encoder (foreground)
US Digital MAE3 Absolute Magnetic Kit Encoder (background)



Absolute versus Incremental

One of the basic classifications used for encoders is whether their architecture is absolute or incremental in design. This refers to the type of output the encoder emits. An incremental encoder uses the lines on the code wheel to output a digital pulse train that corresponds to the sensor detecting light and dark regions. Typically incremental optical sensors use multiple sensing elements separated by various mechanical degrees and simple analog to digital electronics to ultimately produce two phase shifted output pulse trains (commonly referred to as channel A and channel B). The speed of rotation can be deduced from the frequency of these pulse trains while direction of rotation is derived from the phase difference between A and B. Many times an encoder offers a third output called the index. The code wheel has a second optical channel with a specific optical pattern designed to give a once per rotation pulse, index, which is often used as a known absolute position during the rotation of the encoder code wheel.

Incremental encoders are very useful for both position and speed control, as the pulses are very simple to process. The drawback to incremental encoders is that anytime power is lost, the true position of the mechanical system will be lost as well. In this case the mechanical system will need to be reset to known initial location and restarted. This process is commonly called *homing the system*. In some applications, this type of recalibration process is not possible, or is very expensive. An example would be semiconductor manufacturing equipment, where the machines have very tight positional tolerances. After a power loss event, the machines require the ability to continue from the exact at the moment power was lost, without performing a recalibration procedure. In such situations, absolute encoders are necessary.

An absolute encoder differs from an incremental encoder in that it outputs a position word that corresponds to a specific angle of shaft rotation, as opposed to the incremental encoder series of pulses that must be counted to define position. The absolute position word is defined as a number of bits (10 bits, 12 bits, et cetera) that will determine the resolution granularity of the encoder over its rotation. The code wheel of an optical absolute encoder consists of an optical pattern that has a specific pattern for each address or bit of resolution over the rotation of the code wheel. A multi-turn absolute encoder is a variant that also has the ability to provide not only positional information over one rotation of the encoder shaft, but also has the ability to count turns to give absolute positional information over multiple turns of the encoder shaft. Once again the absolute encoder turns are defined as a number of bits: 10 bits, 12 bits, et cetera. The total available resolution of a multi-turn absolute encoder is the sum of the number of single turn bits plus the turns counting bits. For instance, a 10 bit single turn resolution with 12 bit turns counting resolution absolute encoder will yield 22 total bits of resolution over 4096 turns of the shaft.

As mentioned previously, an absolute encoder communicates a word that defines its position commonly over a serial bus. There are a number of industrial serial communication protocols that are used today with absolute encoders such as Modbus, CANOpen, and Profibus. The design engineer will have to choose what is best for the application. As a result of the added complexities, absolute encoders are more expensive than incremental, but they are also much better suited for critical positional control situations.

Encoder Accuracy and Resolution

Ultimately, the encoder gives the motion controller information about the velocity and/or position of a rotating shaft or linear motion. The encoder must provide position information plus or minus a specific accuracy. Defining the accuracy of an encoder is the most commonly misunderstood aspect of an encoder. Often an encoder with higher resolution is specified thinking that more lines on a disk will provide more positional accuracy. This is incorrect, as accuracy and resolution of an encoder are not connected in any way.

The accuracy of an encoder is primarily defined by the precision at which the code information is placed on the disk and how concentric the disk pattern rotates with respect to the encoder's sensing element. Encoder accuracy is specified in units of arc-minutes or arc-seconds. For example: a low resolution incremental encoder of 100 lines per revolution will just as accurately report 180 degrees as an encoder with 10,000 lines per revolution if they both share the same positional accuracy specification. The high resolution encoder can just break the steps between 0 and 180 degrees into finer increments.

Another way to understand the difference between accuracy and resolution is to think of encoders as a form of analog to digital converter. They convert the analog value of mechanical shaft position to a digital form, and like their electronic A to D cousins, it isn't enough to simply specify resolution. The accuracy of a traditional electronic A to D is usually specified with two terms: integral nonlinearity, and differential nonlinearity.

Differential nonlinearity is the amount by which adjacent codes differ from each other and is directly tied to resolution; it has to be less than one bit, or else the resolution has been over-specified (a 12 bit encoder with a differential non-linearity of 2 bits is really an 11 bit encoder). With encoders, differential nonlinearity corresponds to the monotonicity of the encoder output codes. For an incremental encoder this is almost never a problem: if the encoder has 1000 lines on the disk you will get 1000 pulses out of the encoder.

Integral non-linearity is a measure of how much the actual response of the converter departs from the ideal response of a perfect converter; most electronic A to D's have a bow-shaped error curve relative to a line drawn from zero to maximum input value. However, integral nonlinearity is more elusive to pin down in the world of encoders. It won't be found on most data sheets, though it still exists, usually as a sine wave shaped error curve. The major cause of integral nonlinearity in encoders is the concentricity error between the code wheel and encoder sensing element. There often is a set screw that secures the code wheel hub to the shaft; this, combined with the clearance necessary to slip the hub onto the shaft, gives rise to a very slight eccentricity of motion relative to the encoder's disk pattern detection system. This eccentricity of motion means that the distance of the sensor from the axis of rotation is not constant: it is at a maximum at one position on the disk, and a minimum 180 degrees distant. Thus the amount of rotary motion necessary to produce a one-count change at one angular position of an eccentric code wheel may be different from the amount of rotary motion at another point, simply because the radii may be different. The typical result is that on half the disk (the half where the radius is small) the increments of motion corresponding to each count of change are too large, while on the other half of the disk (the half with the longer radius) the increments of motion are too small. This gives rise to the above-mentioned sine-shaped error curve.

Fortunately, there is an equation that describes the worst case integral non-linearity error in terms of optical radius and eccentricity:

$$INL\ (worst\ case) = 60\ sin^{-1}\left(\frac{TIR}{OR}\right)$$

Where INL is maximum integral nonlinearity in arc minutes, TIR is the total indicated run out of the pattern on the code wheel, and OR is the nominal optical radius (the radial distance at which the encoder senses the code wheel pattern). Notice that this calculation is completely independent of the chosen disk resolution; therefore, selecting a higher resolution disk simply does not increase accuracy. The only means to improve the accuracy error is to reduce TIR run out and/or increase the optical radius.

Conclusion

Encoders are available in many mechanical and functional versions. Understanding some of the basic considerations of an encoder will allow a system designer to properly select an encoder that will provide accurate information and performance over the life of their application.

About US Digital

US Digital designs and manufactures motion control components for OEM manufacturers as well as end users.

Motion control building blocks include absolute and incremental magnetic and optical encoders, inclinometers, interfaces, drives, and more.

With complete manufacturing capabilities in-house, standard products are delivered within 1-2 working days and custom orders that exceed industry expectations.

US Digital is located in Vancouver, Washington, and recently celebrated their 30th anniversary in the motion control industry.

For more info please contact:

US Digital Sales
sales@usdigital.com

US DIGITAL
usdigital.com
360.260.2468