Renishaw plc
New Mills, Wotton-under-Edge,
Gloucestershire, GL12 8JR
United Kingdom

T +44 (0)1453 524524
F +44 (0)1453 524901
E uk@renishaw.com
www.renishaw.com

RENISHAW apply innovation™

White paper

The accuracy of angle encoders

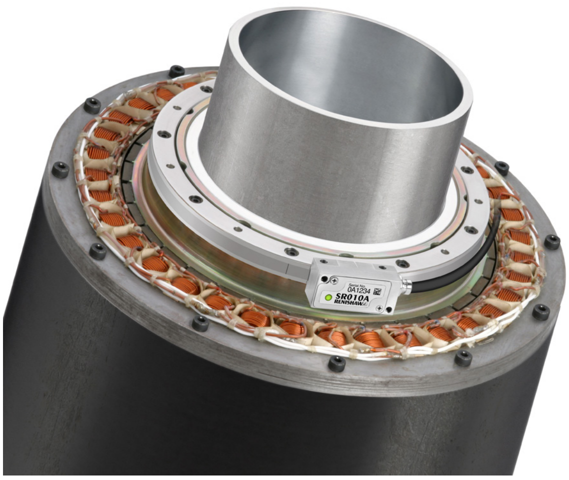
There's an encoder in nearly everything…

Many automation systems rely on precision rotary motion; rotary format computer-to-plate (CTP) pre-press machines, machine tool A, B & C axes, surface mount machines, form measurement, wafer handling and inspection and goniometers all use some form of rotary or angle encoder*.

Different applications demand different combinations of performance and features to optimise their function – some require accuracy, others repeatability, high resolution or low cyclic error for velocity loop control. Typically you would select the encoder that offers the optimum balance of specification and function, and there is a bewildering array to choose from. Yet few are capable of meeting all requirements.

Precision motion control cannot depend on accuracy alone, dynamic response of the system is equally important. Accurately measuring position is important, but without the ability to position accurately, the system is useless. Direct drive rotary motors (or torque motors) develop high torque and enable precision servo control over very small angles. Their dynamic response is excellent because the load is coupled directly to the drive, eliminating the need for transmission components that introduce backlash, hysteresis, gear tooth error or belt stretch. The frameless format of torque motors with large internal diameter offers no obvious coupling to fit a shaft encoder, but a ring encoder provides a convenient solution. Furthermore, like the load, the ring encoder is also coupled rigidly to the drive, eliminating unwanted 'play' in the system. In any measurement or control system, it is preferable for the encoder to be as close to the drive as practically possible - this helps to minimise potential resonances that influence servo performance - particularly as servo bandwidths increase.

Whatever the application, reliable direct position feedback is the key…



Direct drive rotary motor application

*Angle encoders generally have line counts of 10,000 or more with an accuracy better than ± 5 arc seconds. Strictly, the term 'rotary encoder' describes encoders that fall below these criteria, but it is often used as a generic term to describe all round' encoders.

The most obvious solution for providing precision position feedback is a rotary encoder. As with motor selection, choosing the correct rotary encoder depends on a realistic requirement specification, knowledge of the factors affecting encoder accuracy and a good understanding of how performance shortfalls may be overcome. This article outlines the basic factors affecting rotary encoder performance to help designers choose the 'right' encoder system.

When selecting a rotary or angle encoder, it is unwise to opt for the highest accuracy and resolution without considering data rates, system size, complexity and cost. Linear encoders are available with accuracy and resolution measured in tens of nanometres. Likewise, angle encoders can provide sub-arc second performance. A reminder of exactly what an arc second means is sometimes useful:

An arc second…

- Represents 1 micron at a radius of 206.25 mm.
- 30 m at the surface of the earth.
- Resolution gives a data rate of 1.3 MHz at 1 rev/sec.

When determining the accuracy required, it is worth

separating precision, resolution, and repeatability:

- For applications demanding repeatability (e.g. a pick & place machine) the precise angle of each station is secondary to the system's ability to stop at the same encoder count time after time.
- For continuous smooth motion, the resolution and precision selected must not allow 'jitter' within the control servo bandwidth.
- For a slow moving device such as an astronomical telescope, precise angular measurement is more important than the maximum system data rate.
- For a helicopter camera mount, requiring accurate manual positioning, resolution is more important than repeatability or absolute precision, though the latter becomes more important if the same sensor provides target data for a weapon system.

- For high-speed systems, velocity verses positioning accuracy trade-offs may arise; coarser pitch (lower line count) systems suit high data rates, but finer pitch (higher line count) systems usually provide lower interpolation errors.

Once system accuracy requirements are understood, selection of the appropriate encoder is much easier. Despite some manufacturers' claims, accuracy in rotary measurement is rarely 'plug & play' - understanding the error budget is the key to optimising performance.

Building the error budget

Think of a child at school measuring the angle between two pencil lines on a piece of paper with a plastic protractor. He will lay the protractor directly on the paper so that the baseline of the protractor is directly over one of the lines and adjust its position until the origin coincides with the point where the two lines meet. He will then read off the angle between the lines on the graduated scale, interpolating if necessary to give the resolution he needs. The first few times, his reading may differ from his teacher's who'll need to stress the importance of getting the protractor accurately centred and aligned with respect to the lines. These alignment errors will have a bigger effect than any irregularity in the angular scale moulded into the plastic.

Our child has learned three prerequisites of accurate angular measurement:

1. To align the centre of the rotary scale as closely as possible with the vertex of the angle being measured.

2. To keep the measurement scale as close as possible to the item being measured.

3. To minimise relative (angular) motion between the measurement scale and the item being measured.

Three others that he may not have considered are:

4. The circumferential distance between the graduations should be consistent around the circle.

5. The radial distance between the centre of the radial scale and the edge of the scale where the measurement is made should be the same for all circumferential positions.

6. He should take the measurement viewing the line perpendicularly through the protractor to minimise parallax error.

These prerequisites apply equally to a plastic protractor on a page as to a rotary encoder in a machine. In Figure 1 the component of interest whose angular motion is to be measured/controlled rotates on a shaft mounted on two bearings. An angle encoder with integral bearing is coupled to this shaft and read by a readhead mounted on the non-rotating structure. Rephrasing the prerequisites above: for the encoder system output to reflect the actual rotary motion of the component, the following must apply:

DDR motor, Shinko Electric

1. Each part of the system must rotate in its bearings without radial runout (i.e. lateral motion) of its axis of rotation.

2. The shaft system connecting the component of interest to the encoder should be rigid in torsion.

3. The coupling should be designed such that the angular motion of the encoder rotating in its bearing is the same as that of the component of interest rotating in its own bearing system; i.e. a precise constant velocity joint is needed.

4. The spacing of the lines around the edge of the encoder scale should be uniform and the readhead should interpolate between them in a linear manner.

5. The encoder scale should be truly circular with the axis of rotation passing perpendicularly through its centre.

6. The readhead should read the scale without parallax or other geometric error and be rigidly mounted to the non-rotating frame of reference.

If any of the above are not true there will be discrepancies between the angular position of the component of interest and that reported by the encoder system. By investigating each of these potential sources of error, it is possible to determine their individual contributions and thus the total error budget for the whole system.

Effect of bearing wander

The term 'Bearing Wander' is used to describe a variety of system attributes that result in a radial run-out (or lateral translation) of the component's and/or the encoder's axis of rotation, the majority of which can be ascribed to deficiencies in the bearing system. It includes play and higher harmonics (e.g. ball/roller/race imperfections) but, conceivably, not eccentricity.

The magnitude of the radial run-out of a spindle running in roller/ball bearings is affected by the design and adjustment of the bearing system, but is typically not less than ±1 µm.

As the encoder system can read the circumferential position of its rotary scale to at least a 10th of this value, it can be seen that the errors caused by bearing wander can swamp those caused by the remainder of a well designed system. The error contribution of bearing wander is given by:

Angular measurement error (arc seconds) = bearing wander (μm) x 412.5/D

where D is the diameter of the encoder scale in mm.

The relationship between radial run-out, repeatability and angular position will depend on

- The relative dimensions of the inner and outer bearing races
- The number and diameter of its balls/rollers and
- The wear/adjustment of the bearing system.

Despite displaying cyclic components, to all intents and purposes the relationship should be considered to be random, as any error map used to compensate would have to map many spindle revolutions.

For high accuracy systems the use of well designed air bearings is preferred as radial run-out can be reduced to sub-micron levels by correct selection of the bearing radial stiffness. When using air bearings, the effect of out-of-balance forces must be considered. At low speed the spindle will rotate about its geometric centreline, but at high speed, when the effect of out-of-balance centrifugal force exceeds the radial stiffness of the bearing and its mounting, the spindle will rotate about its centre of mass. Although this transition typically occurs at high speeds, it may introduce a discrepancy of several microns between the static and dynamic centre lines. That said, this radial run-out is entirely predictable at one cycle per spindle revolution.

Whatever the bearing used, the following should be noted:

- For the system illustrated in Figure 1, only the bearings supporting the encoder will contribute to bearing wander. Any advantage this brings may, however, be reduced by additional errors introduced by the coupling.
- Although techniques exist to eliminate the effect of bearing wander (notably the use of 2 or more readheads on the same encoder scale) the purpose of the angular measurement must be considered.

▶ If the measurement is required to determine the angular position of a point remote from the axis of rotation (as with an astronomical telescope), even significant bearing wander compensated for in this manner will not affect the accuracy of the results obtained.

▶ If, however, position feedback is required to locate a given point on the rotating component using polar co-ordinates (as with a wafer inspection machine), unless 3 or more readheads are used on the same encoder ring, any bearing wander supporting the component of interest will affect positioning accuracy.

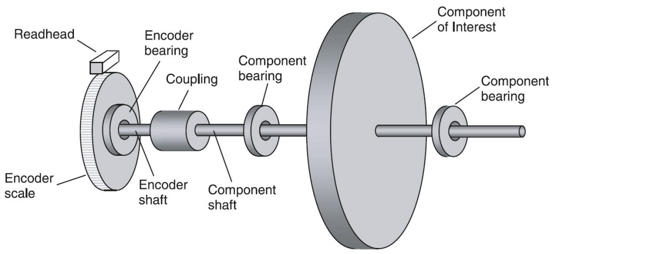


Figure 1. Generic machine system

Where larger amounts of bearing wander are unavoidable, consideration must be given to the selection of an appropriate scale pitch. As a rule of thumb, where the incremental signal is the average of 2 or more readheads with the reference mark derived from one, the scale pitch should exceed the bearing wander by a factor of 3 to 4 - any less than this and problems with reference mark repeatability may become significant, unless compensatory techniques such as Renishaw's propoZ™ technology are used.

Effect of coupling errors

Figure 1. shows a system with a self-contained angle encoder with its own bearings connected to the component of interest via a coupling. This can be a benefit because only the wander in the encoder bearing will impact angular measurement accuracy. However, this 'benefit' must be considered with caution because wander in the main bearings will affect positioning accuracy if the system is designed to return the polar co-ordinates of a point on the component of interest rather than the angular bearing of a distant object.

The design of the coupling itself may have a significant impact on system accuracy. It is beyond the scope of this article to fully document the deficiencies of different coupling designs but the problems are:

Backlash

Any backlash (e.g. from a worn drive-plate in an Oldhams style coupling) will introduce differences in reported angular position with direction of rotation and have the most significant impact on system repeatability.

Torsional stiffness

The coupling may not be as stiff as the shafts it connects and may therefore be affected by vibration/resonance and shaft wind-up, which if used in a feedback loop could significantly affect transient performance, settling time, permissible closed loop gain and bandwidth.

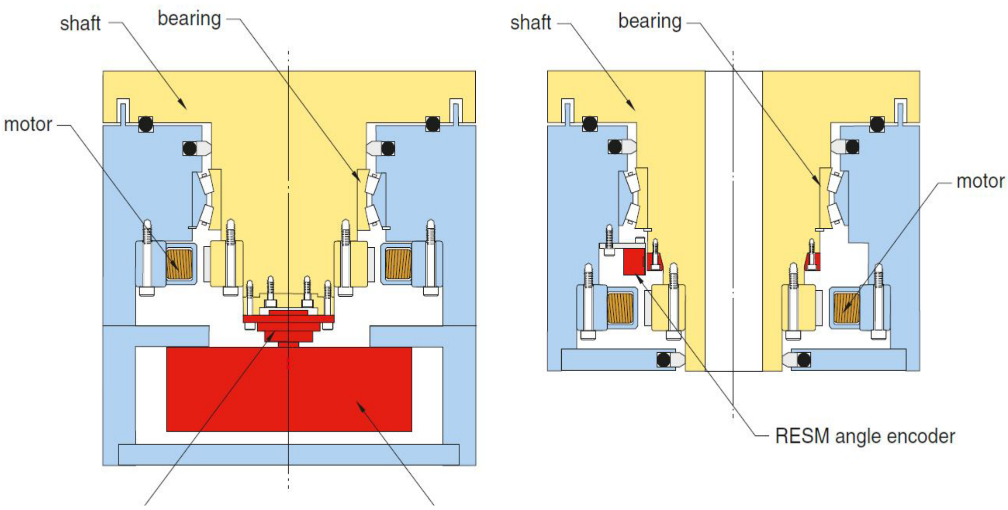


Figure 2. Sealed encoder v open ring encoder

Angular error

Most couplings can, under certain conditions of alignment, introduce angular error between the driving shaft and the driven (e.g. an Oldhams coupling gives a 4-per-rev error if the axes of the 2 shafts are not parallel).

For high accuracy systems the angle encoder should be rigidly mounted on the same shaft as the component of interest and rotate in the same bearings.

Effect of shaft torsion

In the same manner as discussed with the coupling above, lack of torsional rigidity in the shaft(s) between the component of interest and the angle encoder scale will induce dynamic errors which will degrade system performance. To minimise this effect, a non-contact encoder mounted as close as possible to or 'on' the component of interest is recommended (See figure 2).

Effect of scale eccentricity and distortion

Whilst an accurate scale can be made using a combination of non-uniform graduation and varying radius, if accurate angular measurement is to be achieved, a scale of uniform linear graduation must be positioned at a consistent distance from the axis of rotation. Variations in radius, as caused by eccentricity of a perfectly round rotary scale, can generate errors that vary once per revolution. These combine with other errors varying 2 or more times per revolution which are the result of scale distortion.

Consider a perfectly round scale of radius r_0 . Instead of mounting it to rotate about its nominal centreline, it could be mounted to rotate about another point at a distance a_1 from the nominal at a phase angle Φ_1 (see Figure 3). At an arbitrary azimuth angle, θ_1, the distance from the centre of rotation to the scale surface, R_θ, will be given by:

$R_\theta = r_0 - a_1\cos(\theta - \Phi_1)$

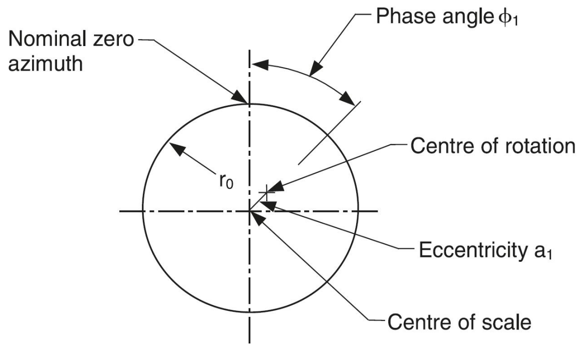


Figure 3. Eccentricity

The true radius will therefore vary sinusoidally once per revolution with an amplitude equal to the eccentricity.

To add the effects of scale distortion, the overall shape of the ring can be considered to be the summation of a series of sine waves of differing frequency n, phase Φ_n and amplitude a_n such that the scale radius at azimuth angle θ is given by:

$R_\theta = r_0 - a_1\cos(\theta - \Phi_1) - a_2\cos(2\theta - \Phi_2) - a_3\cos(3\theta - \Phi_3) - \ldots\ldots - a_n\cos(n\theta - \Phi_n)$

It can be shown that the maximum circumferential error E_n induced by a sinusoidally varying distortion of amplitude a_n (mean to peak), cycling n times per revolution is given by

$\pm E. = a_n/n$

As a sanity check, an eccentricity (i.e. n = 1) of 1 µm will induce a ±1 µm sinusoidal linear error at the circumference.

From this it can be seen that the higher order scale distortions, which will be of increasingly smaller amplitude, will have a progressively smaller effect on scale accuracy. It also shows that the lower harmonics will have significant impact; eccentricity being on a par with the effect of bearing wander.

The relatively flexible nature of Renishaw's ring encoders will raise suspicion that metrology will be affected by eccentricity and distortion, in particular the effects of a potential 'lobing' induced by the multiple securing bolts and the taper mount. On a standard installation of a 200 mm ring secured using 12 bolts tightened to the correct torque settings, no appreciable errors are introduced; the error 'noise' at 12 cycles per revolution being approximately ± 0.05 µm. On a typical installation, eccentricity is responsible for 60% or more of the error (due to that installation) with the lower harmonics (predominantly 2nd to 4th) taking a progressively smaller proportion.

Fortunately, errors induced by eccentricity and distortion respond to compensation techniques, the most powerful of which is the use of multiple readheads. Use of two readheads will remove errors due to eccentricity and all other odd harmonics. Four readheads have been employed in some installations to good effect, but more give progressively lower return on investment; careful selection of the ring cross-section offers a more powerful way of limiting the higher distortion harmonics.

The patented taper mount used on the ring encoders effectively converts a potentially eccentric and distorted ring into one with a small degree of swash, which has a significantly lower impact on accuracy. The taper mount turns 1 µm of eccentricity on a 200 mm ring into a concentric ring with 0.002° of swash. The accuracy achieved without using multiple heads is therefore increased.

Effect of scale swash

Swash refers to the condition where the angle encoder scale is mounted concentrically with the component of interest, but has its geometric axis inclined to the axis of rotation (see Figure 4). Viewed from the side, i.e. radially, this would have the effect of imparting a once-per-revolution sinusoidal axial motion to the periphery of the angle encoder scale.

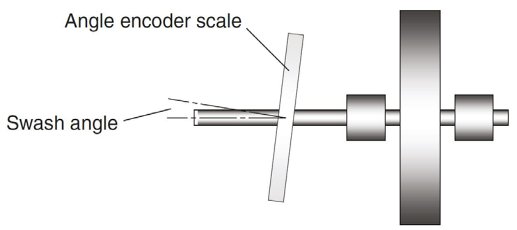


Figure 4. 'Swash'

Swash imparts two distinct error mechanisms, neither of which are obvious. Let us assume an angle encoder scale (with axial graduation) of 200 mm diameter mounted with 0.1° of swash. During installation, the scale has been adjusted to be concentric using a dial gauge (DTI) running on the scale surface. During one revolution, not only will the scale move axially by ± 0.175 mm relative to the readhead, but the yaw angle of the scale graduations will also change through ± 0.1° either side of their nominal value. If the readhead has been positioned on the same part of the scale as the DTI, then the error produced will be second order. However should the readhead be displaced axially from that point by 1 mm, the combination of axial motion, change in yaw and readhead position will induce an error of approximately ± 1.74 µm at the circumference (± 3.6 arc seconds) varying once per revolution.

The second mechanism is as follows: A round coin viewed face on, appears circular. If that coin is then swashed with respect to the eye, it will appear elliptical. The impact of swashing a rotary scale is similar and has the effect of imparting a twice-per-revolution error, the magnitude of which is inversely proportional to the cosine of the swash angle. This is a second order effect, and in the example above would give errors in the order of ± 0.16 arc second – for most applications it can be ignored.

Effect of readhead induced errors

So far, we have concentrated on the encoder scale and its error mechanisms, but the scale is only part of the encoder system; the readhead also contributes to the overall error budget. The most significant readhead induced errors are as follows:

Sub-divisional error

A rotary encoder with 3600 graduations will have one graduation every 0.1° or 360 arc seconds. If the required resolution is finer than that, and it probably will be, then the readhead will be required to interpolate. Any non-linearity in the interpolation will result in cyclic error, also known as sub-divisional error (SDE).

Using a Renishaw readhead for illustrative purposes, the scale and readhead index grating produce optical fringes which move laterally across the readhead photo-detector with movement of the scale. These fringes are sinusoidal in intensity and are decoded by the readhead into 2 sinusoidal voltages 90° out of phase to each other. If these two voltages are plotted against each other on an oscilloscope, a circular Lissajous is generated which rotates once per scale pitch of movement. If this Lissajous is perfectly circular and centred on the origin, rotates at a velocity truly uniform to scale motion and the means of interpolation has truly uniform angular discrimination, then the readhead interpolation will be perfect; if not, SDE will occur.

SDE is affected by the alignment, adjustment and cleanliness of the scale so good housekeeping and careful readhead alignment are important. However, to a greater extent, levels of SDE are determined by the optical design of the readhead. For Renishaw's 20 µm pitch TONiC systems, SDE is typically ± 30 nm (± 0.06 arc seconds on a 200 mm ring). Because of its high frequency, mapping does little to eliminate the effects of SDE, but averaging over small distances can be effective for certain applications.

Parallax

Should the distance between the scale and readhead change (due to e.g. ring eccentricity, temperature change etc.), then errors will be induced unless the readhead is correctly aligned with respect to the scale rotational centreline. If the readhead is pitched, changes in ride height will induce errors proportional to the sine of the pitch angle.

Mounting stability

This may seem obvious but the rigid and secure mounting of the readhead and reference mark is vital to accurate and repeatable angular measurement. The system should be designed such that the readhead does not move with respect to the scale axis of rotation with changes in attitude, loading, temperature, vibration etc. Should the system give unexpectedly high unrepeatable errors, it is worth checking that the bolts securing the readhead and associated brackets & mountings have not loosened off over time.

Effect of scale graduation accuracy

Consider the manufacturing process of an angle encoder where the graduations are marked directly onto the substrate as opposed to being marked on a linear scale which is subsequently secured to the circumference of a disk or ring. The manufacturer may secure the nascent scale to a mandrel which he then rotates to position each graduation. On completion of the graduation process, but before its removal from the mandrel, the measured scale accuracy (discrepancy between the actual and intended position of the graduations) is termed 'Graduation Error'. Were this measurement to be repeated, but this time using a correctly adjusted readhead, the error would, in addition to the graduation error, include components due to the readhead (significantly SDE); this is termed 'System Error'.

If the angle encoder is now removed and remounted on the same or a different mandrel and its accuracy checked with a readhead, once again, the error recorded would be different. The difference would correspond to the error caused by the change in eccentricity and higher order out-of-roundness of the encoder scale between its initial installation for graduation and its reinstallation for use. The overall error measured in this instance is logically termed 'Installed Error' and is the error definition that most closely reflects the performance achieved by the user in the field.

To summarise:

Graduation error = Error in spacing of graduations during manufacture.

System error = Graduation error + SDE

Installed error = System error + Effects of installation differences.

It is worth looking at the relative proportions of these errors; the results taken from testing a significant number of 200 mm diameter Renishaw ring encoders are given in Table 1, 1 µm corresponding to 2.06 arc seconds at this diameter.

Table 1:

Error type	**Typical error on Ø 200 ring**	
	µm	arcseconds
Graduation error	0.5	1.0
System error	0.53	1.1
Typical installed error (1 readhead)	2.5	5.2
Typical installed error (2 readheads)	1.0	2.1

Levels of graduation error and system error are in the hands of the encoder manufacturer, but responsibility for the additional ± 2 µm installed error lies with both the manufacturer and the customer. Even if the customer were able to mount the encoder perfectly concentric and circular, there would still be a difference between the system and installed errors - unless of course the customer mounted the ring in exactly the same position the manufacturer did to graduate it.

The causes of Graduation Error depend on the manufacturing technique:

1. For angle encoders made by the scribing/etching of individual axial lines round the edge of the disc/ring, graduation error is caused by errors in the dividing process.

2. For radial glass encoders made using mask & etch techniques, graduation error is caused by both errors in the accuracy of the mask and errors in the placement of the mask during etching.

3. For an angle encoder system that secures linear scale around the circumference of a prepared shaft; graduation error is caused by the accuracy of the linear scale manufacture, changes in both the thickness of the linear scale and radius of the prepared shaft and differences in tension in the linear scale as it is secured to the shaft.

This last effect is one of the few which can vary after preparation as changes in temperature coupled with differences in coefficient of thermal expansion between the scale and the shaft may cause the scale to creep relative to the shaft surface. Relaxation of glue, if used to secure the scale, may exacerbate this effect.

Considering the Renishaw encoder ring and others where it does not vary with time, graduation error will be predictable from revolution to revolution and, to the user, be indistinguishable from the effects of installation accuracy; it can therefore be reduced using the same techniques.

One, less obvious, manifestation of graduation error occurs if the axial graduations on the edge of an angle encoder are yawed or imperfectly aligned with its axis of rotation. This will have no effect during pure rotary motion, but any axial motion of the encoder scale relative to the readhead (due to end-float in the bearings) will give rise to an incorrect indication of scale rotation. This mechanism is similar to the 3rd swash mechanism described above. The effect of an incorrectly pitched readhead (i.e. with parallax) used with radial scale is similar.

Error compensation techniques

Once the effect of all the error sources of the draft design have been determined, a comparison can be made between the accuracy required to achieve the device specification and the performance that could be expected from the uncompensated rotary encoder system. Should the former exceed the latter, a choice must be made between a different encoder system of higher specification, if one can be found to meet the space envelope, delivery schedule and budget, and applying error compensation techniques to satisfy the shortfall. The two most powerful compensation techniques are the use of multiple readheads and error maps.

Multiple readheads This has already been discussed at some length in the appropriate sections on the error contributors above. Fitting two diametrically opposing readheads will remove the effects of eccentricity, and the higher odd harmonics of repeatable error. It will also remove the effects of bearing wander from angular measurement, but four readheads are normally required to combat bearing wander for accurate polar positioning. Increasing the number of readheads employed will further reduce the repeatable error but it is generally considered that the advantages of fitting more than 4 are outweighed by the complexity and cost. The beauty of this technique is that it does not require elaborate calibration to be effective; a great benefit both in terms of time and testing system design. There is nothing new in this technique; the best mechanical surveying instruments of old used up to 7 points of measurement around an arc to determine angles to sub-arc second accuracies.

Error maps An error map can be used either with or instead of multiple readheads to reduce repeatable errors if the control system of choice is configured to use one. For this technique to be effective, the rotary encoder system must be calibrated by the OEM using an interferometer or another recognised reference after final assembly of the device. He cannot rely on any calibration certificate provided by the encoder manufacturer, because, any errors introduced during the installation process would be ignored, rendering the error map worthless. It is worth optimising the number of points in the error map. For a sinusoidally varying cyclic error, seven points per cycle will remove approximately 90% of the error at that frequency. A hundred point error map will therefore compensate for most of the errors in the first fourteen harmonics but, it should be noted, can potentially increase the errors caused by the higher harmonics remaining. It goes almost without saying that this technique has no impact on the effects of bearing wander, shaft torsion or other time-dependent error sources.

Summary

This article has briefly examined some of the trade-offs that must be made to determine a realistic specification for an angle encoder system. It has also looked at some of the more significant factors that can limit achievable accuracy, detailing a number of techniques available for reducing any shortfall. For further information on this subject, refer to ISO230-7 DIS Axes of rotation.

For further product information, please visit **www.renishaw.com/encoder**.