TECHNICAL WHITE PAPER

A Dummy's Guide to Position Sensors

Author: **Mark Howard**, General Manager Zettlex UK Ltd

File ref: technical articles/Dummy's Guide Position Sensors

1. Introduction

This paper is for engineers, technicians and students – especially those who need to get up a learning curve quickly and gain a basic understanding of position sensing and position sensors. The paper's deliberately short and aims to provide an overview rather than an exhaustive treatise.

2. Terminology

Engineers love jargon – it helps us differentiate engineers from mere mortals. Unfortunately, jargon also makes it difficult for a competent engineer from one area of engineering to get to grips with another. Position sensing is no exception, so let's start with some clarifications on terminology.

Firstly, you'll come across various terms for 'sensor' – encoder, transmitter, detector, transducer even sender. There are some differences, but for most intents and purposes we can consider that they are all terms for the same thing. We'll use the universal term sensor.

Confusingly, some sensors – notably proximity sensors – are actually proximity switches since they determine the presence or absence of an object. What this means is that they produce a simple digital or on/off output rather than a continuous measurement of position. In this paper we will focus on true sensors rather than switches. In other words, sensors that produce a signal (usually electrical) proportional to position along a measurement path.

There's also a bunch of terms that refer to (linear and rotary) position – displacement, angle, angular position, rotary, rotation, linear – again, for the purposes of this paper we will use the universal term, 'position' to cover both linear and angular geometries.

Figure 1 – rotary and linear position sensors

Many, but not all, position sensors can also be considered as speed or velocity sensors. Since speed or velocity can be defined as the rate of change of position, then any position sensor whose position is updated frequently, is – in effect – also a speed sensor. Speed can be readily determined by typical, modern control systems by differentiating the sensor's output with respect to time or, more simply, by counting changes in position relative to time.

All position sensors can be classed as either absolute or incremental. The output from an incremental sensor only changes when position changes. Absolute sensors produce a signal which is proportional to true position whether stationary or moving. A good test to determine whether a sensor is absolute or incremental is to consider what happens at power up. If there's a true position signal without any motion, then it's an absolute sensor.

3. Position Measurement Basics

Maybe you were away from college on the day they did instrumentation theory: you know accuracy, resolution, repeatability and all that stuff. You are in good company – lots of engineers have either forgotten or have never really understood this topic. The terminology and fairly esoteric technical concepts applied to instrumentation are confusing. Nevertheless, they are important in selecting the right position sensor for your application. Get this selection wrong and you could end up paying way over the odds for your position sensors; get it wrong the other way and your product or control system may lack critical performance.

Firstly some more definitions:-

- A sensor's ***Accuracy*** is a measure of its output's veracity
- A sensor's ***Resolution*** is a measure of the smallest increment or decrement in position that it can measure
- A sensor's ***Precision*** is its degree of ***Repeatability***.
- A sensor's ***Linearity*** is the difference between a sensor's output to the actual position being measured

To understand the difference between accuracy and precision we can use the analogy of an arrow fired at a target. Accuracy describes how close the arrow is to the bullseye.

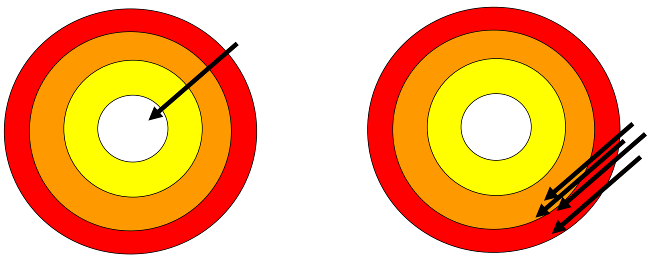

Figure 2 – an accurate shot (left) and precision shooting (right)

If lots of arrows were shot, precision equates to the size of the arrow cluster. If all arrows are grouped close together, the cluster is considered precise or, in other words, highly repeatable.

A perfectly linear position sensor is also perfectly accurate. For most applications linearity can be considered as tantamount to accuracy.

So, that's pretty straightforward then – just specify very accurate sensors every time and you'll be OK? Unfortunately, there are some big snags with such an approach. Firstly, high accuracy sensors are expensive. Secondly, high accuracy sensors may require careful installation and this may not be practical because of vibration, differential thermal expansion or, most likely, cost. Thirdly, some types of high accuracy sensor are also delicate and might suffer malfunction or failure in harsh environments.

The optimal strategy is to specify what is required – nothing more, nothing less. For example, in a position sensor for an industrial flow meter, linearity will not be a key requirement because it is likely that the fluid's flow characteristics will be highly non-linear. More likely, repeatability over varying environmental conditions is likely to be key. It should be noted that resolution and repeatability are often more important than linearity in many engineering applications.

In a CNC machine tool, for example, it is likely that both accuracy and precision will be a key requirement. Accordingly a position sensor with high accuracy (linearity), resolution and high repeatability even in dirty, wet environments over long periods without maintenance are likely crucial requirements.

4. Common types of position sensor

Position sensors are used in a wide variety of industrial and commercial applications, from high end military and defence applications through to low cost automotive and consumer appliances. In fact, after temperature measurement, position measurement is the second most common property that we need to measure in our lives.

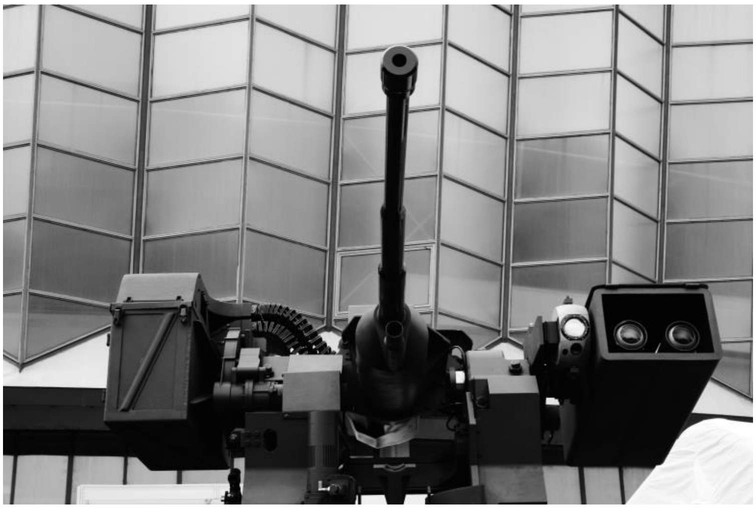

Figure 3 – position sensors are used in military and defence applications

There's a bewildering array of position sensors to choose from nowadays, so how do you choose the right one? This section outlines the main types of sensor and their respective strengths and weaknesses.

4.1 Potentiometers

Although there is a trend towards non-contact sensors, potentiometers (‘pots’) continue to be the most common position sensor. These sensors measure a voltage drop as electrical contact(s) slides along a resistive track, which means that position is proportional to voltage output. Pots are available in rotary, linear or curvilinear forms and are generally compact and lightweight. A simple device will cost pennies, whereas a high precision version may cost upwards of US$200. Linearities of less than 0.01% are possible by laser trimming the resistive tracks.

Potentiometers operate well in applications with modest duty cycles, benign environments and relaxed performance. Unfortunately, pots are susceptible to wear especially in high vibration environments and/or with foreign particles such as dust or sand which will abrade the resistive track. Higher quality devices quote long life in terms of the number of cycles, but this often ignores the effects of vibration.

It must also be noted that potentiometers often quote ‘infinite resolution’. Whilst theoretically true, many control systems require digital data and so actual resolution will be that of the analogue to digital converter (which needs to be included in any costings).

Figure 4 – a typical single turn potentiometer

Bizarrely, pots get classed as ‘simple devices’ in some safety related applications in aerospace, medical and petrochemical industries. This means that whilst they are subject to numerous failure modes they are not subject to the same rigorous design and selection scrutiny as electronic sensors by certifying bodies. It’s a daft but true situation that makes it difficult to replace unreliable pots in some applications.

Strengths: Low cost; simple; compact; lightweight; can be made accurate.
Weaknesses: Wear; vibration; foreign matter; extreme temperatures.

Further Reading:

- The True Cost of Potentiometers
- Why do engineers dislike potentiometers?

4.2 Optical

Optical sensors are usually referred to as encoders and they are a common form of position sensor, ranging from simple devices that cost a few dollars through to precision units that can cost upwards of US$10,000. Across all such devices the fundamental principle is the same: a light beam is shone through or onto a grating; the resulting light measured using a photo detector and a position signal generated.

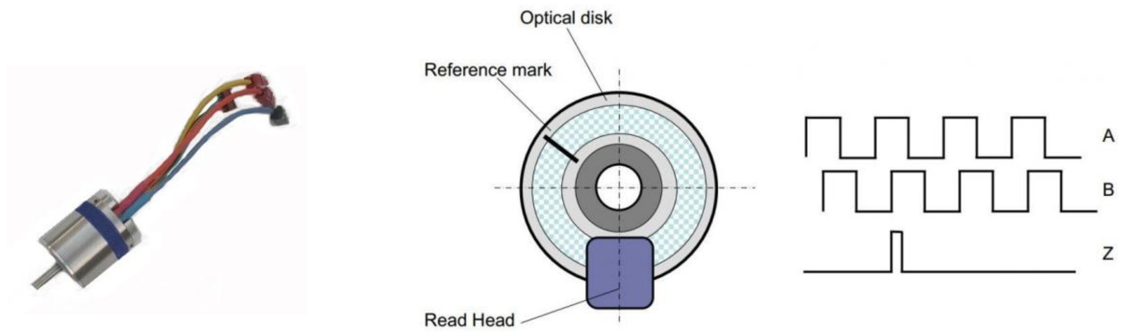

Figure 5 – optical sensors use an optical disk to measure angle

Packaged, rotary encoders are widely available, typically with 50-5,000 counts per revolution and they are proven to work well in benign applications. However, in more robust environments, if the lens or grating system becomes obscured by foreign particles such as dirt, swarf or water, then measurements will fail.

In selecting an optical sensor, it is important to note that if the sensor is quoted as having 1,000 counts per revolution, this does not mean that it is accurate to 1/1000th of a revolution. The data sheet for the sensor will need to be read carefully, particularly with encoder kits or ring encoders which require the user to mount the sensor extremely accurately and to ensure no contamination.

If the encoder has a glass disk, the unit will have limited shock resistance.

Strengths: High resolution; good accuracy if mounted precisely; wide availability.
Weaknesses: Foreign matter; catastrophic failure with no warning; shock; extreme temperatures.

Further reading:

- Optical Encoders Vs. Inductive Encoders

4.3 Magnetic

Magnetic sensors all use a similar measuring principle: as a magnet moves relative to a magnetic detector, the magnetic field changes in proportion to their relative displacement. A common form is the Hall Effect device which is available in chip form. They are often used in automotive and electric motor applications with modest measurement performance.

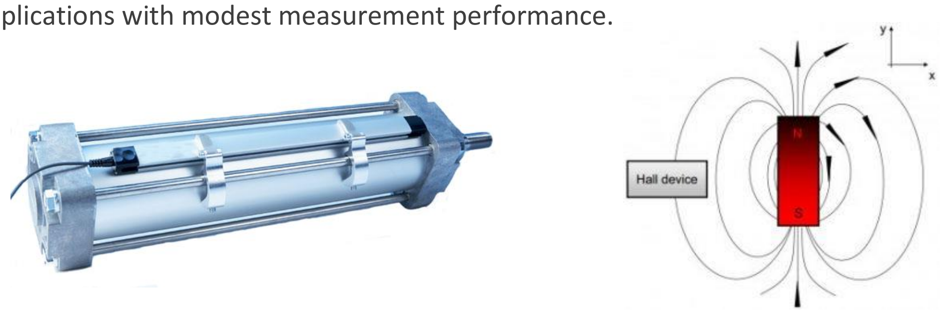

Figure 6 – Hall Effect sensors are the most common type of magnetic sensor

Magnetic sensors overcome many of the drawbacks associated with optical devices, as they are more tolerant to foreign matter. Nevertheless, these sensors are rarely used for high accuracy applications due to magnetic hysteresis and the need for precision mechanical engineering between the moving and stationary parts. Any magnetic sensor's data sheet should be studied carefully with respect to installation tolerances, temperature coefficient and operating temperature.

A further consideration is the proximity of magnetic materials or electrical cables. Magnets may attract some foreign particles and one source of failure is the build-up of swarf or particulates over time. Magnetic sensors are typically not chosen for applications with harsh impact or shock conditions since the modern NdFeB magnets are notoriously brittle.

Strengths: Fairly robust; most liquids have no effect.
Weaknesses: Temperature; hysteresis; precision mechanical engineering; nearby steel/DC sources and poor impact/shock performance.

Further reading:

- Magnetic Vs. Inductive Position Sensors

4.4 Magnetostrictive

These sensors use an unusual phenomenon called 'magnetostriction', which is present in a few materials. When a magnet approaches the material it causes energy passing along the material to reflect. Position can be measured from the time it takes a pulse of energy to move along and back a strip of magnetostrictive material – usually a thin wire or strip.

Nearly all magnetostrictive sensors are linear because the delicate magnetostrictive strip must be carefully held in a housing such as an aluminium extrusion. The housing means that magnetostrictive devices suffer no wear or lifetime issues and they can be used in high pressure applications such as hydraulic rams.

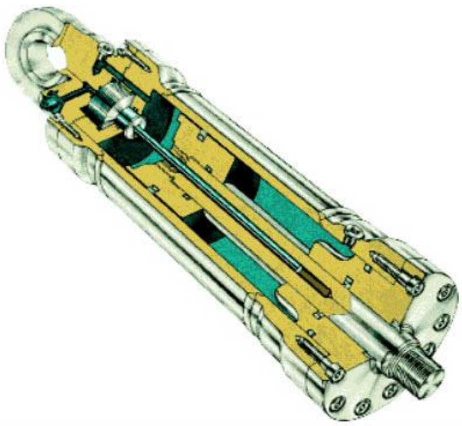

Figure 7 – magnetostrictive sensors are nearly always linear

Each sensor needs to be calibrated by the manufacturer and this, combined with the precision housing, makes magnetostrictive sensors relatively expensive. The technique is also sensitive to any other influences on the time of flight – most notably temperature. Magnetostrictive data sheets often quote accuracy at constant temperature, so design engineers will need to do their own calculations using the quoted temperature coefficient.

The tiny magnetostrictive is delicate and the mounts at either end of its length are critical. The net result is that magnetostrictive sensors should not be chosen for harsh shock or vibration environments.

Strengths: Robust; well suited to high pressures; % accuracy increases with length.
Weaknesses: Fairly expensive; shock; temp. effects; inaccurate over short distances (<100mm).

Further reading:

- Magnetic Vs. Inductive Position Sensors

4.5 Capacitive

A capacitor is an electrical device that accumulates charge. Typically, it has two conductive plates separated by an insulator. The amount of charge the capacitor can store varies according to the size of the plates, their percentage overlap, their separation and the permeability of the material between the plates. In its simplest form, a capacitive position sensor measures plate separation. Displacements are typically over ranges of less than 1mm for load, strain and pressure measurement.

Another form is used for rotary or linear position sensing in which a series of plates are cut or etched along the measurement axis. As another plate moves across them, the capacitance of the circuits along the axis varies indicating the relative position of the two parts. Capacitive position sensors are uncommon and rarely used in safety related applications. Unfortunately, as well as overlap of the plates etc., capacitance also varies with temperature, humidity, surrounding materials and foreign matter, which makes engineering a stable, high accuracy position sensor challenging.

I've been working in sensors, automation and electronics for almost 30 years. I'm yet to meet a design engineer who is happy with his selection of a capacitive position sensor. Capacitive sensors have a poor reputation with experienced engineers and are unlikely to be chosen for safety related applications. Some manufacturers have stopped referring to 'capacitive' and instead use alternative terms like charge storage, charge coupling or electric effect to obfuscate matters. Not good. There's so many things to go wrong they are best avoided unless you have a need for high accuracy measurements in highly stable and clinical applications.

Strengths: Compact; low power.
Weaknesses: Significant temperature and humidity coefficients; sensitive to foreign matter; tight installation tolerances.

Further reading:

- A Comparison of Capacitive and Inductive Position Sensors

4.6 Traditional Inductive

Traditional inductive position sensors work on inductive or transformer principles and have been used for more than 100 years. They have an excellent reputation for safe and reliable operation in tough conditions making them an almost automatic choice in many safety related applications.

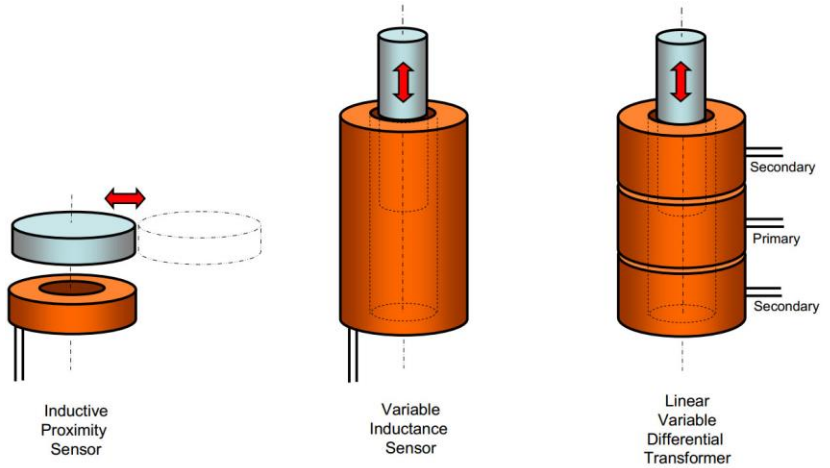

Figure 9 – Traditional inductive sensors have an excellent reputation for safe and reliable operation.

Linear inductive position sensors are commonly referred to as variable reluctance or linearly variable differential transformers (LVDTs). Rotary forms are known as synchros, resolvers or RVDTs. LVDTs use a transformer construction with at least three wire spools: a primary and two secondaries. As the rod moves, it varies the electromagnetic coupling between the primary and secondary spools. The ratio of the induced signals indicates the position of the rod relative to the spools. This ratiometric technique is key to the LVDT’s high stability and measurement performance.

Whereas optical and magnetic sensors require electronic circuitry adjacent to the sensing point, inductive sensors can displace the electronics away from the sensing area, which enables the sensor to be located in harsh environments with the electronics in more benign locations.

However, because of the wound transformer construction they tend to be big, bulky and expensive.

Strengths: High accuracy; reliable; robust; extreme environments; widely available.
Weaknesses: Expensive; bulky; heavy.

Further reading:

- Traditional Inductive Vs. New Generation Inductive Sensors

4.7 New Generation Inductive or Incoders

New generation inductive sensors – often referred to as incoders - use the same principles as traditional inductive sensors, so they offer good, non-contact measurement performance in tough environments. However, rather than use bulky spools of wire, these sensors use printed circuits on flexible or rigid substrates.

The transition to printed windings brings other specific advantages:

- A large reduction in production cost, size and weight
- Greater flexibility in form factor
- Eradication of sources of inaccuracy from the winding process
- Complex measurement geometries such as curvilinear, 2D & 3D position sensing
- Multiple sensors can be located in the same space by using multi-layer circuit boards (e.g. redundant sensors in safety-related applications).

Figure 10 – example of new generation Incoders

EMC performance is generally as good as that of resolvers or LVDTs. This is evidenced by the increasing selection of new generation inductive devices for aerospace and military applications.

Strengths: High accuracy; reliable; robust; multiple geometries; compact; lightweight.
Weaknesses: More expensive than potentiometers

Further reading:

- IncOder Inductive Encoders

5. Common pitfalls

The following is a list of the most common mistakes that engineers make with regards to position sensors:-

Not calculating the cost of sensor failure. All engineers want to select a low cost solution. This is not the same as simply selecting the lowest cost sensor. As a general rule, the cost of sensor failure in the field is going to be more trouble and many times more expensive than the cost of a position sensor. In other words, it is usually the best and least overall cost solution to select a sensor that will not fail in the field. Further, there's also the nature of the failure to be considered. A sensor that malfunctions and stops working is usually far less problematic and costly than a sensor that fails and produces a credible but wrong reading. The consequences of a wrong sensor reading in terms of cost and safety can be even higher than a sensor that simply stops working or produces an error warning.

Not understanding the difference between repeatability, resolution and accuracy. Have a look back at section 3 and make sure you understand these basics. You should avoid the mistake (often propagated by the position sensors industry) of confusing resolution and accuracy. Just because an optical encoder produces a million counts per rev, does not means that it is accurate to one millionth of a rev – far from it. Conversely, repeatability is often the key requirement in many engineering applications and high accuracy (and hence high cost) sensors need not be specified.

Mismatching sensor type and environment. Man has devised ways of harnessing most of the basic physical phenomena to measure position by using optical, magnetic, capacitive, resistive and inductive techniques. Each technique has its own strengths and weaknesses. As a general rule don’t select

- Resistive (potentiometric), optical or capacitive sensors for dirty or wet environments. Condensation and surface ice in outdoor equipment is a common cause of failure.
- Optical, magnetic or capacitive sensors with applications with extreme operating temperatures (most will not operate above 125C)
- Magnetic sensors where high measurement performance is required, unless it’s also possible to eradicate magnetic fields and arrange precision mechanical sensor mounting
- Potentiometers in applications with harsh or prolonged vibration. This because their sliding electrical contacts are subject to wear and failure from lots of vibration induced microscopic movements.

Inferring a measurement rather than measuring directly. A good design rule for position sensors is to measure the position of the object that you’re interested in. In other words, measure its position *directly*. Try not to *infer* or calculate a component’s position by measuring the position of another component such as a gear at the end of transmission line or the position of a drive motor. There is likely to be backlash, clearances, part-to-part variability, mechanical failure, differential thermal expansion/contraction etc. that will inevitably degrade measurement performance and reliability.

Forgetting cables & connectors. Cables and connectors are a primary cause of sensor failure. Ensure that they are accounted for in any design and in particular the cables are strain relieved in any applications that experience motion, shock or vibration.

Not reading the small-print of a sensor’s datasheet. The position sensor industry is a competitive one. Unfortunately, this has led to some manufacturers being a bit too commercially sharp with specification data. Often they get away with it because the industry also knows that many engineers won’t have read a paper like this. The consequence is that sensors will be publicised with, for example, a resolution 10,000 counts per rev – but no mention of accuracy. Another example, is sensors with impressively high resolution but much less repeatability – in other words lots of resolution but also plenty of noise on the sensor’s output. The trick is not to be misled by head-line figures of a datasheet – read the small print.

6. How to specify a Position Sensor

The first and most important step in choosing a position sensor for your project is to be absolutely clear about what is needed, particularly with respect to sensor resolution, repeatability and linearity. Over-specifying any of these attributes will cause unnecessary expense. The trick is to find a sensor that is fit-for-purpose at minimum overall cost – remembering to include an allowance for field failure in your analysis.

You can use the following as a check list to ensure you’ve considered all the important stuff in your specification. Providing this to a position sensor supplier together with a mechanical drawing of the envelope will also provide a solid basis for your discussions:

A. **Geometry** – for example, linear or rotary or curvi-linear or 2D or 3D
B. **Space envelope**– mechanical fixing points, cable routings and space envelope
C. **Measurement type**– incremental or absolute
D. **Full-scale**– for example, 360 degrees or 600mm
E. **Resolution** – in other words, the smallest change that must be measured – for example 0,1 degrees or 0,2mm
F. **Repeatability** – in other words, the stability of the measurement in terms of going back to the same point – for example repeatability = +/-0,025mm
G. **Linearity** – the maximum allowable deviation from a perfectly accurate reading. You might want to think carefully about this since we often find that what is most important for many applications is actually repeatability.
H. **Operating and store temperature range** – -40 + 85Celsius is most typical
I. **Electrical supply** – for example, 5V, 12V or 24V
J. **Electrical output** – for example Serial Data, A/B pulses, 0-10V, 4-20mA
K. **Unusual stuff** – such as – “we want to keep power consumption as low as possible” or “it’s for submersion in hot sulphuric acid” or “we’re using a capacitive device and we have reliability problems”

Further Information / Contact

Zettlex designs and builds sensors for precisely measuring position or speed in tough environments. The position sensors use a unique non-contact technology to deliver high accuracy, high reliability measurements in harsh conditions. Our inductive position sensors are used for servo controls, motor encoders and user interfaces in the medical, defence, aerospace, industrial, marine, motor sports & petrochemical sectors.

UK Head Office

Zettlex UK Ltd
Newton Court, Newton, Cambridge, CB22 7ZE, United Kingdom
Contacts: Mark Howard or Josef de Pfeiffer
Email: info@zettlex.com
Telephone: +44 1223 874444
Web: www.zettlex.com