Sensors of Physical Parameters

So far we have confined our attention in terms of sensors to the problem of analogue-to-digital conversion of voltages but there are a large number of other parameters which we need to measure which are not, per se, voltages. Examples such as shaft rotation, temperature, pressure and humidity come to mind. We therefore need to look seriously at the question of transducers.

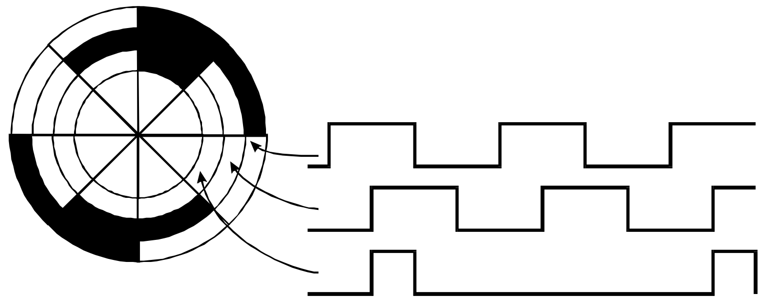
Before looking at any methods of conversion in any detail, consider one or two significant points: In order to be used in a digital system any physical parameter has to be converted into a digital form and, unless the conversion can be done directly, an intermediate form has to be used. Direct digital conversion is always preferable where possible because it reduces the potential for error in the analogue systems. By far the most prevalent intermediate forms are voltage and time (current is assumed the same as voltage here as they are readily interchangeable). Of these two forms voltage can be measured to a very finite limit of resolution and accuracy whereas time can be measured to any conceivable resolution and accuracy, provided you have enough patience. The order of preference for conversion is therefore:

- direct conversion ot digital form
- convert to time (or frequency)
- convert to voltage, current and then digitise

Shaft Encoders

Consider first the monitoring of rotation - the shaft encoder. The objective of this device is to sense the position or rotation of a shaft. There are two possibilities here and one should be aware of which one is required as it affects the cost of the operation. If we are trying to measure <u>rotation-rate</u>, i.e. rpm then we probably do not need to know exactly where the shaft is, whereas if we want to know where the shaft currently is we shall find we have a more complex problem. Thus the simplest form of rpm measurement consists of a flag vane and a sensor, which is almost always an optical device. The flag interrupts the light beam once per revolution and we measure (note a time measurement here) the frequency of the resulting waveform. However if the shaft rotates very slowly then we will spend an enormous amount of time measuring the rotation. So we can add vanes, say 100 and then the rotation rate is 1% of the encoder output frequency. This simple encoder is known as the incremental encoder.

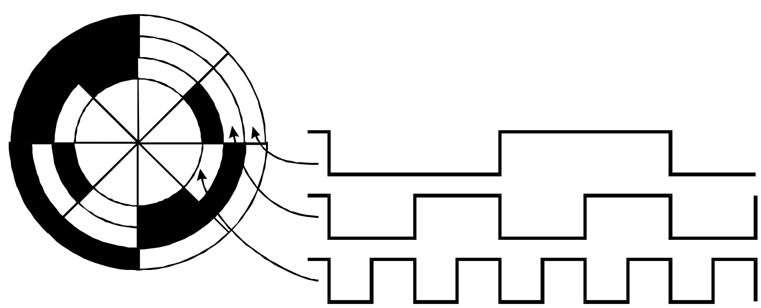
The incremental encoder can be modified into a simple form of absolute encoder by the addition of a second channel with only one vane per revolution. In this case by counting the pulses after the "zero vane" pulse we can measure where the shaft is - 50 pulses = 1/2 way round. However in a bidirectional problem we cannot work out which way the shaft is moving and therefore cannot see the difference between +1/4 and -1/4 in revolution. This again can be solved by a third channel which is the same as the incremental (100 pulse) channel but arranged 90 degrees out of phase with it. Thus the two outputs go as the top two waveforms of the diagram with the zero vane marking off the reference point. Therefore the count sequence from the two upper channels goes 00 01 11 10 00… and any state transition gives information about both the number of steps and the direction. With a computer looking on at a fast enough rate one can therefore measure shaft position accurately.



Incremental Encoder with Directional Indication

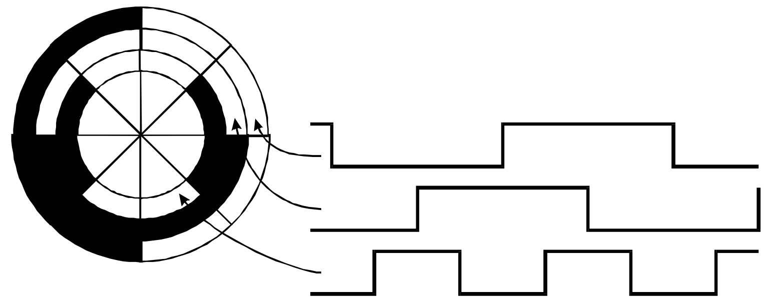
However there is one problem with the above scheme and that is that if we lose our place - we are lost for ever unless the zero vane appears again to recalibrate us. The scheme is therefore quite suitable for systems which rotate on a fairly consistent basis but not for systems which must rely on being accurate without recalibration. For these cases we use the absolute shaft encoder which is more complicated.

The absolute encoder consists of a number of channels producing a binary output for shaft rotation and the simplest scheme that one can think of for three channels is shown in the first diagram.



Absolute Shaft Encoder - Binary Version

However this suffers from a severe limitation in that the channels do not always change state simultaneously, similar to, but not the same as, the "ripple" problem in counters. We can solve this by the use of a "strobe" channel to tell us when readings can be taken but then we do not get a continuous readout, there are "blank spots" in the rotation pattern.



Absolute Shaft Encoder - Gray-Code Version

The commonest practical scheme for overcoming this problem is to use a "Gray code" encoding scheme. I think Gray was a person, by the way. In this scheme the binary numbers are still the same but the order is shuffled such that only one bit changes at any one time. A simple 2-bit scheme is 00 01 11 10 (as above). The three-bit scheme can be derived by writing the 2-bit scheme with the 3rd bit clear

```
000
001
011
010
```

and then writing the same sequence backwards with the 3rd bit set

```
110
111
101
100
```

The 4-bit scheme is derived the same way 3-bit with 4th clear, then 3-bit backwards with 4th bit set and so on. Because there is only one transition in one bit for every state change, the encoder does not suffer from transient states and the output is continuous and clean. There is a minor problem with sorting out the codes to get a simple binary sequence again but since there is a one-to-one correspondence between Gray code and straight binary a simple array look-up will suffice (although there are calculation schemes).

```
position := gray_array[port[IN]];
```

A simple calculation will suffice to show why hi-resolution shaft encoders are expensive. With 12-bit output we require 4096 states in rotation. If the encoder body is 7cm in diameter then the perimeter length of the encoder disc is 20cm and the length of each element, i.e. the pattern scale is 20/4096cm = 50 μm and things are already very precise.

There are some other schemes for shaft encoders, or rather for playing with encoder outputs, which merit a bit of attention. If the shaft is known to rotate at a constant (or nearly constant rate) then the output of a hi-resolution encoder is a clock-count sequence. If we therefore take a simple zero-vane and count a clock to the next occurrence and find that there are n pulses/revolution then we can say that the pulse count is exactly the same as the output from an encoder with n states and use the count instead. We need to monitor the value of n continuously to allow conversion from angle to counts, but we do not need the hi-resolution shaft encoder. Note however that if the rotation rate changes during a revolution we will get erroneous results.

Similar schemes allow us to interpolate encoders but be wary here as the same problems occur with encoders as with a/d converters, the $\pm 1/2$LSB problem, whereby the output resolution is only guaranteed generally to be of the same order as the size of the LSB.

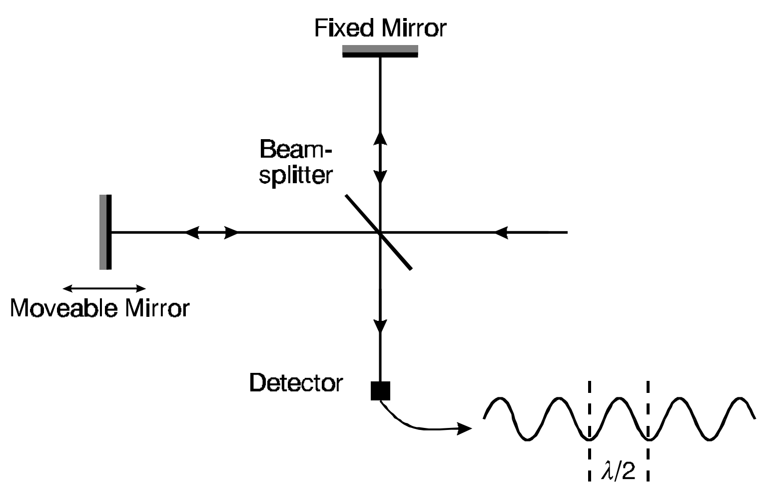
The accuracy of an encoder also follows a similar set of rules to DACs and ADCs - with the proviso that since a single revolution brings everthing back to the same place the system is self-calibrating on a single rotation basis. However the accuracy at an intermediate point in the revolution depends upon the accuracy of construction of the encoder, even if all the states are present (monotonicity) the width of the states might vary.

Linear Motion Sensing

There are several methods for measuring linear motion which "revolve" around the concept of the optical sensor.

Moiré systems can be used which give large light variations for small amount of movement and straightforward sensing of graticules are used for less critical applications.

A method which is rapidly



Fourier Transform Spectrometer

gaining favour is to use a laser beam and count the interference fringes as the motion proceeds. This is exemplified in the fourier transform spectrometer where the laser beam is used to measure the linear displacement of the mirror. Such schemes are of course incremental in nature and require the use of a calibration point but the interferometer produces its own zero-point by the existence of the ZPD (Zero Path-Difference) in the scan. Just to get things in perspective a He-Ne laser has a wavelength of about 600nm and an interference system produces 2 counts/wavelength so that the resolution is already 300nm before we use fractional fringe techniques for increasing the resolution.

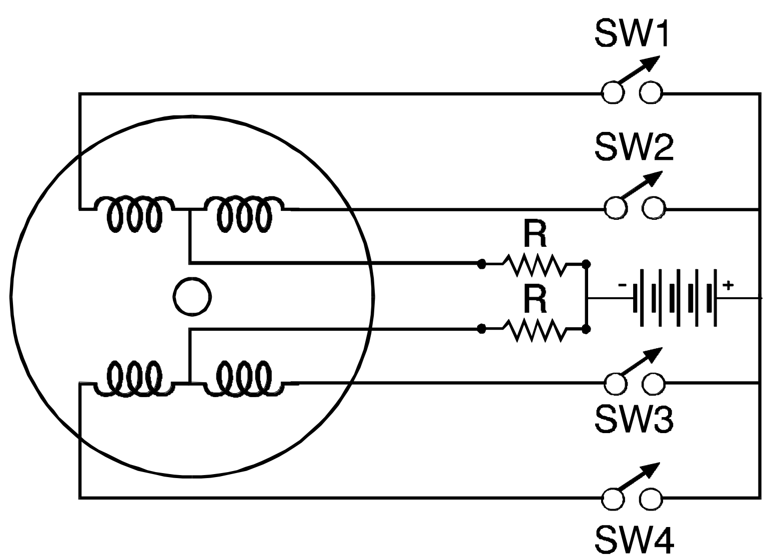
Motors

A word about motors to go with the sensors. The degree of control that you can get with a motor depends upon the type and the suitability of the type depends upon the application. There are essentially three ways to control rotation:

Start/Stop motors (with optional speed control) are used in many pieces of equipment and are suitable for applications where almost continuous rotation is required most of the time or, with some form of sensor for intermittent movement between points with non-critical positioning, e.g. automatic door openers with limit switches.

Clutch Systems give a fast response but a very jerky take-up unless clutch control is arranged to be smooth (fluid flywheel or similar). This technique has great applications where the source of power is non-electric, e.g. a gas engine.

Stepper Motors are precision positioners but tend to be slow. A stepper motor may be moved a number of steps in either direction by the application of a suitable sequence of voltages. A stepper is therefore a well-defined system which, if perfect, does not need a sensor to tell you how far you have got so far. It is also capable of very precise rotation rates. They tend



Stepper Motor Windings and Control

to be slow but can move large loads. A further advantage with some is that they have strong "cog" points where the load tends to stay and therefore "1/2" steps are semi-impossible. This same "cogging" however means that they are difficult to use where an extremely smooth rotation is required without some very fancy driving techniques.

The standard drive sequence is either a "four-step" sequence which has a high retention torque at each step point and a coarser resolution (which will of course depend on the motor construction) and an "eight-step" sequence which has higher resolution but poorer retention.

The step sequences are:

STEP	SW1	SW2	SW3	SW4
1	ON	OFF	ON	OFF
2	ON	OFF	OFF	ON
3	OFF	ON	OFF	ON
4	OFF	ON	ON	OFF
1	ON	OFF	ON	OFF

FOUR-STEP SEQUENCE

STEP	SW1	SW2	SW3	SW4
1	ON	OFF	ON	OFF
2	ON	OFF	OFF	OFF
3	ON	OFF	OFF	ON
4	OFF	OFF	OFF	ON
5	OFF	ON	OFF	ON
6	OFF	ON	OFF	OFF
7	OFF	ON	ON	OFF
8	OFF	OFF	ON	OFF

1	ON	OFF	ON	OFF

EIGHT-STEP SEQUENCE

Sensing Temperature

Let us now consider another major area of sensing - temperature. Virtually all measures of temperature use a form of a/d conversion of a voltage which is related to the temperature in some way and there are three major forms of sensor:

Thermocouples are the primary calibration source for most temperature measurements. If one junction of the thermocouple is kept at 273K (well strictly speaking at the triple point of water - but the melting point of ice is more common) and the materials of the thermocouple are known then the voltage out as a function of temperature is known to a very high accuracy. However there are several problems with thermocouple systems which make them awkward to use:

1) They require a reference temperature of 273K provided by an ice-bath. There are ways of getting around the problem but these render the primary calibration property a bit suspicious.

2) The output voltage is small (and how!). The maximum voltage that can be got out of a thermocouple is of order mV and in order to measure fractional degrees resolution to µV is required. Thus considerable analogue processing is required before the output can be digitised.

3) They require great care with materials in complex apparatus. This means keeping similar metal contacts at all joints, ensuring an absence of thermal gradients in connectors, etc., etc.

4) They are also slightly non-linear but that is just a matter of using an appropriate correction curve and is a slight inconvenience.

Overall thermocouples are good primary standards and not too good for general use.

Resistance thermometers share some of the disadvantages of thermocouples but they do lose the requirement for a reference point. They have the advantage of being reproducible standards but have two disadvantages:

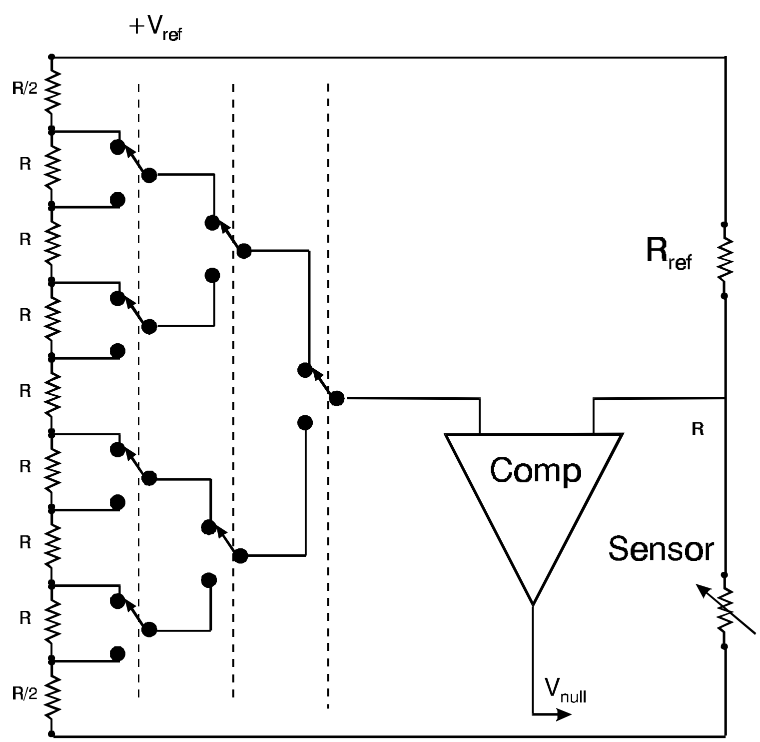
1) They require the measurement of a small change of a large resistance. This is in contrast to a thermocouple where the requirement is to measure something which is small, period. They require both amplification and offsetting or an enormous dynamic range. Therefore they require considerable analogue processing.

2) The materials problem with an additional twist. Because we are measuring resistance we need to be wary of extra resistance in the circuit and that necessitates the use of a 4-wire system as shown here.

File Contains Data for PostScript Printers Only

4-Wire Temperature Measurement

At this point it might be worth pointing out that a specialised form of a/d converter can work as a bridge circuit for measuring resistance directly. Notice that a reference voltage is not required for this circuit but a somewhat specialised form of comparator and ladder may be. This technique is sometimes used for resistance thermometry because of the very small changes in resistance being detected (temperature coefficient of resistance 0.0038/K).

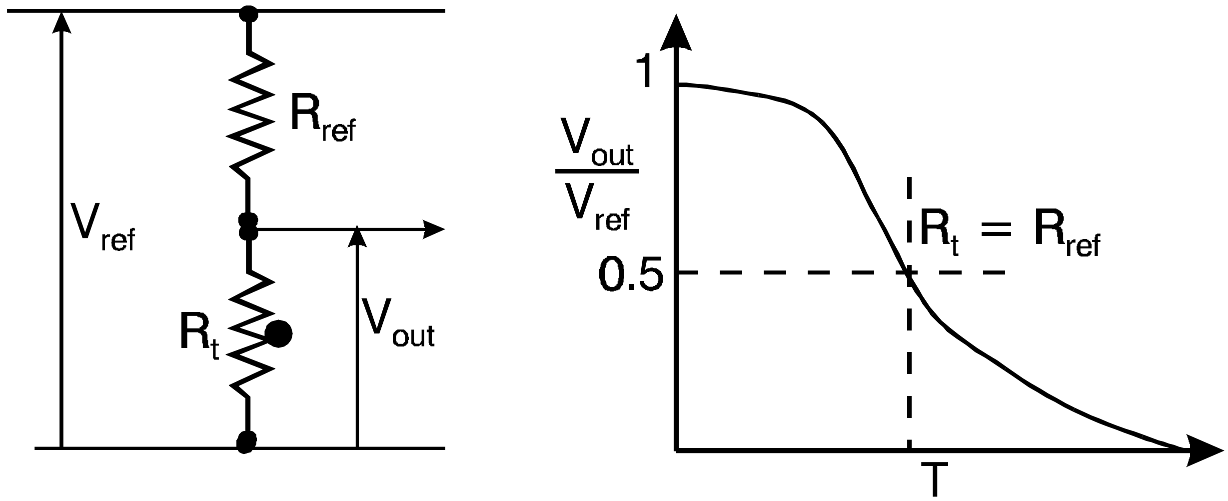


Bridge-Nulling A/D Converter

The last category of temperature sensors is the *thermistor* type which have a large temperature-sensitive part of their resistance. Their advantage is that the signal is large, no reference is required but their disadvantage is their inaccuracy. They are also, in general, not interchangeable. The resistance vs temperature relation for a particular type of thermistor is

$$\frac{1}{T} = A + B\ln(R) + C(\ln(R))^3$$

which is non-linear - and how! If we put the thermistor into a circuit thus:



Thermistor Circuit and Output

We shall have good resolution where $R_t \approx R$ but poorer at the ends. Therefore we should always set up the circuit so that the operating range is in the centre. Calibration is tedious, using either a lot of measurements or a complex relationship, but once done then a table look-up or other "unkinker" will give a temperature readout. Note that in general thermistors are not interchangeable and therefore this must be done for each one. However some firms (for a price) do make interchangeable versions which make life easier.

.

Linearisation

Many sensors as we have seen are non-linear and it is very useful to be able to linearise the output so that subsequent processing can be simplified. There are many techniques for linearisation and I'll just put forward a few of the most useful.

Analytic equation If the non-linear equation is known, then it can always be applied to the input to get the linear quantity back again. A simple example is the generation of the power in a heater by squaring the voltage drop across it.

Look-up table Good and fast if you can afford the storage. It has the advantage that it can cope with any non-linearity including non-analytic ones. However it can get tedious initially entering the values and the tables can get large if good resolution is required.

Piece-wise Linearisation A very acceptable technique for small non-linearities. Essentially the range is broken up into small enough sub-ranges so that a straight line in the sub-range is an adequate approximation. We've already seen a piece-wise approximation in the companding a/d converter. The advantages are that the precision can be good at a modest storage requirement and the speed is intermediate between the analytic solution and the look-up table.