TOPIC 1.0
INTRODUCTION TO MEASUREMENT AND INSTRUMENTATION SYSTEMS

1.1. INTRODUCTION

The ability to make accurate measurements is fundamental in enabling us to engage in Science and Engineering. Our daily performance is centred on our ability to quantify delivered or consumed amounts (e.g. electric energy), to verify desirable amounts (e.g. blood pressure) and safe amounts (e.g. traffic).

Over the last decade tremendous advances have been made in the fields of electrical/electronic engineering. Today virtually all measurements are made in the *'electric domain'*. Instrumentation devices and systems used in manufacturing/process industries are continually being refined to accurately monitor performance and provide up to date reports at all levels of production.

This course introduces the fundamentals of industrial and process instrumentation. The course seeks to provide students information that should enable them to: -

- Effectively operate existing industrial/process instrumentation devices and systems
- Guide the acquisition of new instrumentation devices when need arises
- Engage in research and development of new measurement systems.

1.1.1 PROCESS VARIABLES AND THEIR MEASUREMENT

All engineering operations depend on the measurement and control of process variables.

A process variable is defined as "*any condition or state of a process material or of its environment that is subject to change*." It can also be described as *"a quantity or condition associated with a process value which is subject to change with time."*

1.1.2 *Classification of process variables*

Three classes of process variables are introduced.

- Energy variables
- Quantity and rate variables
- Physical and chemical characteristics

a) Energy variables are those that are affected by the energy state of the material. They include temperature, pressure/vacuum, electricity, sound, radiation, etc.

b) Quantity and rate variables are affected by the quantity and flow rate relations of several materials in the process. Examples include fluid flow, liquid level, weight, thickness and speed.

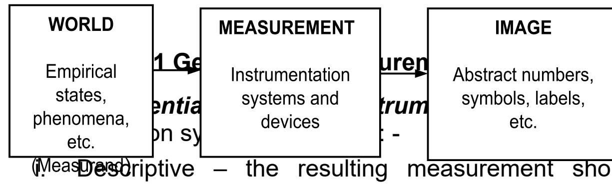
c) Physical and chemical characteristics are those that are dependent on the physical and chemical characteristics of the material. These include density, humidity, moisture content, viscosity, calorific value, colour, electrical and thermal conductivity, chemical absorption, refractive index, X-ray diffraction, polarity, pH, oxidation-reduction potential.

1.1.3 Definitions

Measurement: The acquisition of information that concerns characteristics, states or phenomena of the world around us with the aid of an instrumentation system.

Instrument: A device used in determining the magnitude of a variable.

Instrumentation: The field of engineering that covers the study (design and fabrication) of measurement devices.



i. Descriptive – the resulting measurement should relate to the phenomenon or state constituting the measurand.

ii. Selective – It should provide specific information about individual parameters that are measured in an environment consisting of several other states or phenomena.

iii. Objective – independent extracts of the same measurand should yield the same measurement outcome when subjected to the system.

1.2 CHARACTERISTICS OF INSTRUMENTATION SYSTEMS

Measurement systems are designed to accurately detect changes in parameters encountered in industrial/scientific/bio-medical processes such as pressure, fluid flow, motion resistance, voltage current and power. The information they generate facilitates manual or automated control. The following characteristics are used to describe the performance of instrumentation systems and devices.

1.2.1 Accuracy

This is *the closeness of the measured value to the true value of the measurand*. Accuracy is specified in terms of the absolute deviation of the image portrayed by a measurement system from the real world parameter. If the true value cannot be expressed with 100% certainty then an accepted standard value (or best estimate) is used.

Accuracy, a = $\pm$ | Measured value – true value |, or

= $\pm$ | Measured value – accepted standard |

Consider a system whose accuracy limits of its constituent elements are given as $\pm a_1$, $\pm a_2$, $\pm a_3$.
Define:

Overall accuracy, $A = \pm (a_1 + a_2 + a_3)$

Root mean square accuracy, $A_{RMS} = \sqrt{(a_1^2 + a_2^2 + a_3^2)}$

The overall accuracy of a device or system is usually specified as a percent of the entire range of measurement (% FS – **F**ull **S**cale). This specifies the ***maximum*** possible deviation of the measurement outcome from the true value.

1.2.2. *Precision*

Precision is *the closeness with which individual parameters are distributed about their mean value*.

Precision is also quoted in terms of deviation and it gives the degree of reproducibility among several measurements of the same true value under similar reference conditions. Expressed mathematically, precision is the mean of the scatter of individual measurements.

Precision = $\pm \sum$ | Measured value

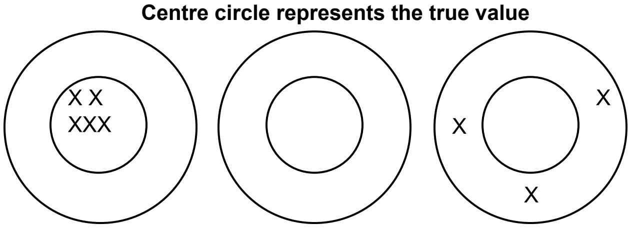
–mean value |

Total number measurements (Range)

NOTE: Precision does not necessarily guarantee the accuracy of a device or system.

Precision is a term that describes an instrument's degree of freedom from random errors. If a large number of readings are taken of the same quantity by a high-precision instrument, then the *spread of readings will be very small*. Precision is often, though incorrectly, confused with accuracy. High precision does not imply anything about measurement accuracy. A high-precision instrument may have a low accuracy. Low-accuracy measurements from a high-precision instrument are normally caused by a bias in the measurements, which is removable by recalibration.

X - measurement result
Centre circle represents the true value



(a) (c) (b)

High accuracy Low accuracy Low accuracy
High Precision High precision Low precision

Figure 1.2: The difference between accuracy and precision

1.2.3 Repeatability

Repeatability describes *the closeness of output readings when the same input is applied repetitively over a short period of time, with the same measurement conditions, same instrument and observer, same location and same conditions of use maintained throughout.*

1.2.4 Reproducibility

Reproducibility describes *the closeness of output readings for the same input when there are changes in the method of measurement, observer, measuring instrument, location, and conditions of use and time of measurement*. Both terms thus describe the spread of output readings for the same input. *This spread is referred to as repeatability if the measurement conditions are constant and as reproducibility if the measurement conditions vary.*

The degree of repeatability or reproducibility in measurements from an instrument is an alternative way of expressing its precision. Figure 1.2 illustrates this more clearly. The figure shows the results of tests on three industrial robots which were programmed to place components at a particular point on a table. The target point was at the centre of the concentric circles shown, and the black dots represent the points where each robot actually deposited components at each attempt. Both the accuracy and the precision of Robot C are shown to be low in this trial. Robot B consistently puts the component down at approximately the same place but this is the wrong point; therefore, it has high precision but low accuracy. Finally, Robot A has both high precision and high accuracy, because it consistently places the component at the correct target position.

1.2.5 Tolerance

Tolerance is a term which is closely related to accuracy and defines the minimum error which is to be expected in some value. Whilst it is not, strictly speaking, a static characteristic of measuring instruments, it is mentioned here because the accuracy of some instruments is sometimes quoted as a tolerance figure. Tolerance, when used correctly, describes *the maximum deviation of a manufactured component from some specified value*.

Crankshafts, for instance, may be machined with a diameter tolerance quoted as microns (10^{-6} m), and components in an electric circuit such as resistors have tolerances of perhaps 5%. One resistor chosen at random from a batch having a nominal value 1000Ω and tolerance 5% might have an actual value anywhere between 950Ω and 1050Ω.

1.2.6 Range / Span

The range or span of an instrument defines the minimum and maximum values of a quantity that the instrument is designed to measure.

1.2.7 Error

Error is similar in definition to accuracy as both give an indication of the deviation of the measurement outcome from the true value. Two categories of error feature in measurement systems:

(i) Mistakes (human errors) – These arise from the erratic human operator’s actions such as reading the wrong scale, incorrect device settings, overloading, improper instrument application and computational mistakes. These errors can easily be avoided by adhering to the stipulated measurement protocols.

(ii) Measurement errors – These are usually independent of the operator’s actions. They may be classified as systematic and random.

1.2.7.1 Systematic (Bias) errors

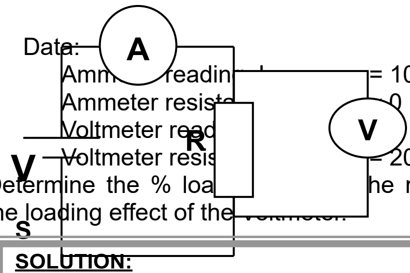
These are errors of a distinctively fixed magnitude that appear to occur every time that a measurement is made. These errors exist over the full range of measurement of an instrument. They include:

Instrument errors – result from friction in bearings and moving parts, irregular spring tension, loosely fitting parts, etc. They can be avoided through regular calibration of the measurement devices.

Environmental errors – These errors result from the presence of non-ideal electromagnetic, electrostatic, gravitational or thermal interactions tied to the environment where the measurement is made. This negative influence can be minimised through adjustments to the measurement system to provide for shielding (magnetic/electrostatic), temperature compensation and/or re-calibration of elements in the measurement path. Most instrument manufacturers specify ambient conditions under which measurements should be made. To avoid error, where possible, the test environment may be conditioned to conform to the instrument specifications.

Loading/mismatch errors – These are caused by energy losses occasioned by the effect of the instrumentation system’s load impedance on the object under test. The instrument tends to use some energy in acquiring the measurement input, thereby altering the true value of measurand. These errors cannot be eliminated completely. However various techniques can be applied to minimise their magnitude. The most common of these is called ***matching*** (energic or anenergic). This technique attempts to balance the load conditions between the instruments input terminals to match those of the test object (measurand).

<u>*Example 1.1*</u>
Consider the Voltmeter-Ammeter application used to determine the value of a resistor R as shown below.



Data:
Ammeter reading = 10 mA
Ammeter resistance 0
Voltmeter reading = 250 V
Voltmeter resistance = 200 kΩ

Determine the % loading error in the reading of R due to the loading effect of the Voltmeter.

SOLUTION:
Ideally V_S = V = I R_T
R_T = Total Circuit resistance
= V / I = 25 k

Since R_A 0, the apparent resistance of R (denoted as R_a) is approximately equal to the total circuit resistance R_T.

i.e., R_a $R_A + R = R_T$,
R_a 25 k

Taking into account the loading effect of the Voltmeter, the true magnitude of R is evaluated from the revised expression of the total circuit resistance as follows:
$R_T = \{R^{-1} + R_V^{-1}\}^{-1}$
= 28.56 k

Hence the loading error = $\frac{|R - R_a|}{R}$ x 100 12.46 %

1.2.7.2 Random errors
Random errors are those which vary unpredictably for every successive measurement of the same physical quantity. Statistical methods are used to determine the probabilities of their occurrence. For a set of measurements, prediction of error may be accomplished by determining the mean or average, standard deviation, probability distribution functions, etc.

1.2.8 ***Sensitivity***
Sensitivity (S) is the *ratio of the magnitude of an instrumentation system's output signal to that of its input*. Sensitivity is thus the ratio:

$$S = \frac{\text{Scale deflection}}{\text{Value of measurand causing deflection}} = \frac{Y}{X}$$

If, for example, a pressure of 2 bar produces a deflection of 10 degrees in a pressure transducer, the sensitivity of the instrument is 5 degrees/bar (assuming that the deflection is zero with zero pressure applied).

Define,
Differential Sensitivity, S_d, $= \left.\frac{dy}{dx}\right|_{X=X_0}$

With y = f(x)
For linear systems the differential sensitivity is equal to the static sensitivity, i.e. $S_d = S$

$$\textbf{Sensitivity factor,} S_x^y = \frac{dy}{dx} \cdot \frac{x}{y}$$

Also called ***cross sensitivity.*** Again y = f(x). Linear systems have a sensitivity factor of 1

<u>*Example 1.2*</u>
The following resistance values of a platinum resistance thermometer were measured at a range of temperatures. Determine the measurement sensitivity of the instrument in Ω/^{0} C.

Resistance (Ω)	Temperature (^{0}C)
307	200
314	230
321	260
328	290

SOLUTION:
If these values are plotted on a graph, the straight-line relationship between resistance change and temperature change is obvious.
For a change in temperature of 30 ^{0} C, the change in resistance is 7 . Hence the measurement sensitivity is equal to 7/30 = 0.23/^{0} C.

SENSITIVITY TO DISTURBANCE

All calibrations and specifications of an instrument are only valid under controlled conditions of temperature, pressure, etc. These standard ambient conditions are usually defined in the instrument specification. As variations occur in the ambient temperature, etc., certain static instrument characteristics change, and the sensitivity to disturbance is a measure of the magnitude of this change. Such environmental changes affect instruments in two main ways, known as zero drift and sensitivity drift.

1.2.8.1 Zero Drift

Zero drift describes the effect where the zero reading of an instrument is modified by a change in ambient conditions. In the case of a voltmeter affected by ambient temperature changes, typical units by which zero drift is measured are volts/ ^{0} C. This is often called the zero drift coefficient related to temperature changes. If the characteristic of an instrument is sensitive to several environmental parameters, then it will have several zero drift coefficients, one for each environmental parameter. The effect of zero drift is to impose a bias in the instrument output readings; this is normally removable by recalibration in the usual way.

1.2.8.2 Sensitivity Drift
Sensitivity drift (also known as scale factor drift) defines the amount by which an instrument's sensitivity of measurement varies as ambient conditions change. It is quantified by sensitivity drift coefficients which define how much drift there is for a unit change in each environmental parameter that the instrument characteristics are sensitive to.

Many components within an instrument are affected by environmental fluctuations such as temperature changes; for instance, the modulus of elasticity of a spring is temperature dependent. Taking the example of a pressure transducer where a dial gauge rotates by a certain angle for an applied pressure, temperature dependent sensitivity drift would be measured in units of [(angular degree/bar)/ °C].

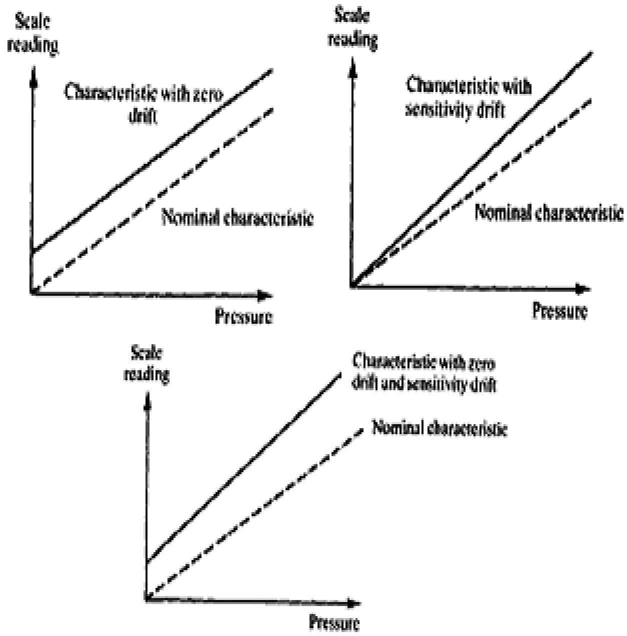


Figure 1.3: Instrument drift

Example 1.3

A spring balance is calibrated in an environment at a temperature of 20 ^{0}C and has the following deflection/load characteristic:

Load (kg)	0	1	2	3
Deflection (mm)	0	20	40	60

It is then used in an environment at a temperature of 30^{0}C and the following deflection/load characteristic is measured:

Load (kg)	0	1	2	3
Deflection (mm)	5	27	49	71

Determine the ***zero drift*** and ***sensitivity drift*** per ^{0}C change in ambient temperature.

SOLUTION

At 20 ^{0}C, deflection/load characteristic is a straight line.

Sensitivity = $\frac{\text{Deflection}}{\text{Load}}$ = 20 mm/kg.

At 30 ^{0}C, deflection/load characteristic is still a straight line.

Sensitivity = 22 mm/kg

Bias (zero drift) = 5 mm (the no-load deflection)

Sensitivity drift = (22-20) = 2mm/kg

Zero drift/ ^{0}C = (5-0)/(30-20) = 0.5 mm/ ° C

Sensitivity drift/^{0}C = 2/(30-20) = 0.2 (mm per kg)/ ° C

1.2.9 *Resolution And Scale Readability*

Resolution defines *the smallest change in the input variable (measurand) that is sufficient to cause a change in the output (display)*

The resolution of an instrumentation system indicates its ability to discriminate between two nearly equal input values.

Scale readability indicates the number of significant figures that can be recorded in output data. In analogue systems it depends on the scale marking and pointer thickness; while in digital systems the least significant bit determines it.

1.2.10 ***Linearity*** *(or Non-linearity)*

Linearity is the *maximum deviation, in the range of the measurement output, from the ideal response of the instrumentation system.*

It indicates the ability of a measurement system to reproduce the input signal symmetrically. In the illustration Fig 1.4, the best approximate of the input/output (i/o) characteristic is the mathematical relation; y = f(x) =ax. Here 'a' is an arbitrary constant obtained as the gradient.

For this system

$$\text{Linearity (\% F.S)} = \frac{|f(x) - ax|_{max}}{Ax}$$

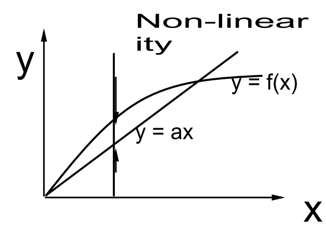
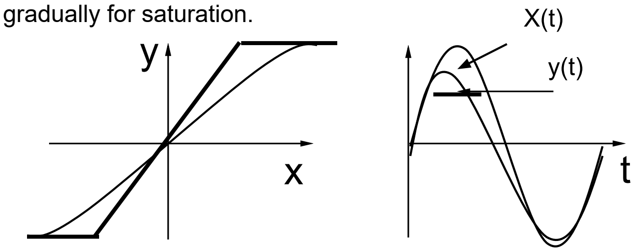


Figure 1.4: Static (or frequency independent) non-linearity

Forms of static non-linearity

1.2.10.1 Saturation and Clipping

This non-linearity is characterised by a *decreasing differential sensitivity with an increasing input*. The decrease occurs abruptly in the case of clipping and gradually for saturation.



Saturation

Clipping

(a)

(b)

Static i/o characteristic

Dynamic i/o characteristic

Figure 1.5: Saturation and clipping

1.2.10.2 ***Hysteresis***

Hysteresis is the *non-coincidence in the i/o characteristic observed when the input is increased steadily from a negative to a more positive value and then gradually* *decreased to the lower value.* It may result from play in a mechanical gear drive.

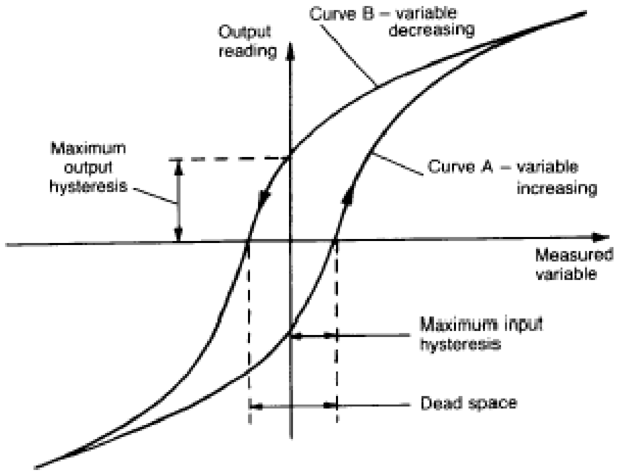
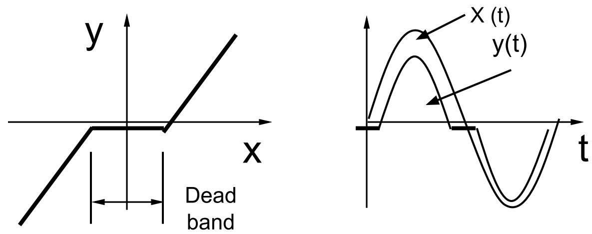


Figure 1.6: Hysteresis behaviour

The two quantities of maximum input hysteresis and maximum output hysteresis are defined as shown in Figure 1.6. They are normally expressed as a percentage of the full-scale input or output reading respectively.

1.2.10.3 Dead band/ Dead space/ Threshold

This is the *range of input values over which the measured variable changes without causing an output response*. It is usually caused by backlash in gears or static friction in movable parts of the instrument and results in the instrument i/o characteristic shown in Figure 1.7



Static i/o characteristic
Dynamic i/o characteristic

Figure 1.7: Dead band

1.3 MATHEMATICAL MODELLING OF DYNAMIC INSTRUMENTATION SYSTEMS

1.3.1 The Laplace Transform

The Laplace transform F(s) is defined mathematically as:

$$F(s) = L[f(t)]$$

$$= \int_0^\infty f(t)e^{-st}dt$$

where s – Laplace operator

The transfer (i/o) characteristic of an instrumentation system's dynamic response is determined by applying the Laplace transform to its characteristic equation.

1.3.2 Kinds of dynamic inputs

When the measurand changes there will be a time lag before a change in input of the instrument is observed. To characterise the dynamic (i.e. time dependent) response of the instrument we consider three cases in the way in which the measurand might change, i.e.:

- Step change
- Ramp change
- Sinusoidal (Harmonic) change

1.3.2.1 *Step change*

This is where there is an abrupt, discontinuous change in the measurand. It is fully described mathematically by the *step function*, defined as:

f(t) = 0 for t< 0
= A for t >0

Where A is a constant

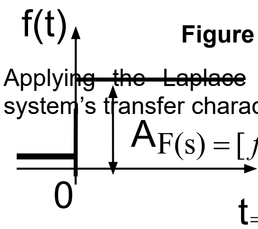


Figure 1.8: Step function

Applying the Laplace transform to the input gives the system's transfer characteristic:

$$F(s) = [f(t)] = L[A] \quad \text{for } t \geq 0$$

$$= \int_0^\infty Ae^{-st}dt = \frac{A}{s}$$

1.3.2.2 Ramp change

Here the change in the measurand is linearly proportional to the change in time. It is fully described mathematically by the *ramp function*, defined as:

f(t) = 0 for t< 0
= At for t >0

Where A is the slope's gradient

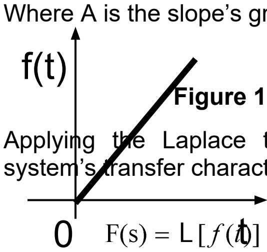


Figure 1.9: Ramp function

Applying the Laplace transform to the input gives the system's transfer characteristic:

$$F(s) = L[f(t)] = \int_0^\infty Ate^{-st}dt$$

$$= \left[At\frac{e^{-st}}{-s}\right]_0^\infty - \left[-\frac{1}{s}\int_0^\infty Ae^{-st}dt\right] = \frac{A}{s^2}$$

For the unit step function, A= 1

1.3.2.3 *Sinusoidal/ Harmonic change*

The harmonic change describes a sinusoidal change in the measurand. It is represented mathematically by the sine *function*, defined as:

f(t)= 0 for t< 0
= Asinωt for t ≥0

Where A is a constant

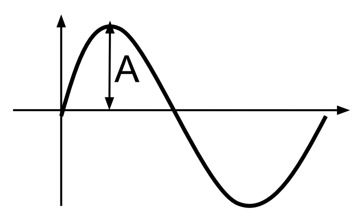


Figure 1.9: Harmonic function

The system's transfer characteristic:

$$F(s) = L[f(t)] = L[A\sin\omega t] \quad \text{for } t \geq 0$$

$$= \frac{A\omega}{s^2 + \omega^2}$$

1.3.3 Dynamic performance of systems

The underlying differential equations that describe the dynamic performance can be broadly categorised into 3 classes of instrument: zero, first and second order systems.

1.3.3.1 Zero Order Systems

Here the *output of the system is always linearly proportional to the input* no matter how fast it changes. If we take θ_o as the output from the instrument and θ_i as the input quantity, then we have:

$$\theta_o = K\theta_i$$

Where K is the system's static (i.e. time independent) sensitivity, as earlier defined. The expression given by the ratio θ_o / θ_i is described as the ***transfer operator*** which in this case is given by:

$$\frac{\theta_o}{\theta_i} = K$$

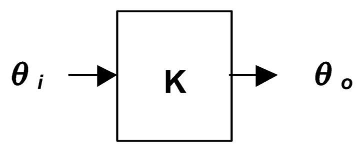


Figure 1.10: The transfer function for a Zero Order system

While complete instruments are rarely classed as zero order systems, components such as transducers can approach this behaviour e.g. a variable resistor (or potentiometer): changes in the position of the control contact wiper produce near instantaneous changes in output voltage.

1.3.3.1 First Order Systems

The dynamic performance of this system is described by a first order ODE.

$$a\frac{d\theta_o}{dt} + b\theta_o = c\theta_i$$

>(1)

Where,

$\theta_i = \theta_i(t) =$ input function

$\theta_o = \theta_o(t) =$ output function

$a, b, c =$ constants

The transfer operator is obtained by rearranging equation (1) to give:

$$\frac{a}{b}\frac{d\theta_o}{dt} + \theta_o = \frac{c}{b}\theta_i$$

> (2)

This gives the standard form of the general equation:

$$\tau\frac{d\theta_o}{dt} + \theta_o = K\theta_i$$

> (3)

Where

$\tau =$ system time constant $= a/b$

$K =$ static sensitivity $= c/b$

Introducing the ***D operator*** where:

$$D = \frac{d}{dt},\ D^2 = \frac{d}{dt^2}$$

The transfer operator from Eq. (3) can be re-written as:

$$\tau D\theta_o + \theta_o = K\theta_i$$

$$(\tau D + 1)\theta_o = K\theta_i$$

$$\therefore \frac{\theta_o}{\theta_i} = \frac{K}{(\tau D + 1)}$$

> (4)

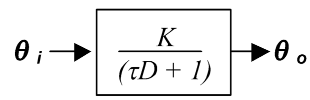


Figure 1.11: The transfer operator for a First Order system

Examples of 1st order systems include:

- thermal element / glass thermometer – heat conduction through the glass of the bulb is determined by 1st order ODE.
- build up of air pressure in a restrictor / bellows system.

<u>*Example 1.4*</u>

The differential equation describing a mercury-in-glass thermometer is:

$$4\frac{d\theta_o}{dt} + 2\theta_o = 2\times10^{-3}\theta_i$$

Where θ_o is the height of the mercury column in metres and θ_I is the input temperature in ^{0}C. Determine the time constant and static sensitivity of the thermometer.

<u>SOLUTION:</u>

Required: Time constant () and static sensitivity (K).

Comparing coefficients of the system equation with those of equation (1):

a = 4; b = 2; c = 2 x 10 $^{-3}$

EMBED Equation.3

lve

and

Complementary equation = τp + 1 ∴ p = -1 / τ

General solution $= Be^{pt} = Be^{(-t/\tau)}$

>(5)

$\theta_o/K\theta_i$ = particular solution + general solution >(6)

The particular solution is obtained when t → ∞, the output approaches the input ∴*$\theta_o/K\theta_i \approx 1$*

Hence:

$\theta_o/K\theta_i$ = particular solution + $\underbrace{Be^{(-t/\tau)}}_{0}$ ≈ 1

∴ Particular solution = 1

Substituting in eqn (6) yields

$\theta_o/K\theta_i = 1 + Be^{(-t/\tau)}$

>(7)

When t =0, both the output and input are at zero and hence B = -1. This yields:

$$\theta_o / K\theta_i = 1 - e^{(-t/\tau)}$$

or

$$\theta_o = K\theta_i \, [1 - e^{(-t/\tau)}] \quad >(8)$$

For $t \geq 0$, $\theta_i = A$,

$$\theta_o = KA \, [1 - e^{(-t/\tau)}] \quad >(9)$$

Two cases:

i.) When $t = \tau$, $\theta_o = KA \, [1 - e^{(-1)}] = 0.632$ KA, i.e. the output rises to 63.2% of the step input.

ii.) When $t \to \infty, \tau$, $\theta_o \approx KA$, *i.e.* the output approximates to the input.

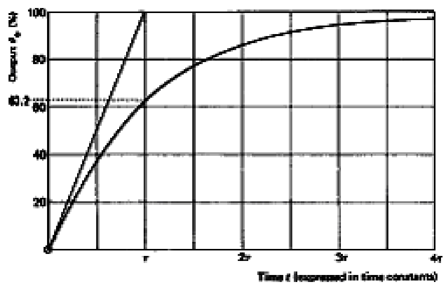


Fig 1.12 Step response of a First Order System

The difference between the *ideal output* and the *actual output* is referred to as the ***dynamic erro***r.

NOTE: The initial slope of the response curve at t =0 is **not** zero which is a characteristic of 1st order systems. 2nd order systems **do** have zero initial slope in their step response.

Frequency Response of 1st Order Systems

Frequency response describes the dynamic behaviour of an instrument system when the input signal is sinusoidal in nature. By varying the frequency of the harmonic wave we can derive the frequency response of the instrument transfer function.

In general we will see a *change in amplitude* and the *introduction of a phase lag* between the input and output. Generally, as we increase frequency we observe an increasing phase lag and a decrease in output signal amplitude.

When $\theta_i = \sin(\omega t)$ we can show that the output has the form:

$$\theta_o = \underbrace{A \exp\left(-t/\tau\right)}_{Transient} + \underbrace{\frac{K}{\sqrt{1+\omega^2\tau^2}} \sin(\omega t + \phi)}_{Steady\ state}$$

Where:

$$A = \frac{K}{\sqrt{1+\omega^2\tau^2}} \sin\phi \quad \text{and} \quad \phi = \tan^{-1}(-\omega t)$$

The Laplace transform and frequency response determination

The ***transfer function*** of a linear, time invariant, function is defined as the ratio of the Laplace transform of the output (response function) to that of the input (driving function), derived under the assumption that all initial conditions are zero.

Let $\theta_i = x(t) = x \; ; \theta_o = y(t) = y$

The 1st Order system equation is:

$$(\tau D + 1) y = Kx \quad >(10)$$

Transfer function,

$$G(s) = \frac{\mathrm{L}\,[y(t)]}{\mathrm{L}\,[x(t)]} = \frac{Y}{X} \quad >(11)$$

Applying the Laplace transform to both sides of equation (10) gives:

$$(\tau s + 1) Y = KX$$

∴For the 1st Order System

$$G(s) = \frac{Y}{X} = \frac{K}{\tau s + 1} \quad >(12)$$

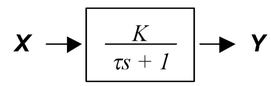


Figure 1.13: The transfer function for a First Order system

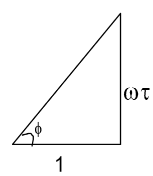
Complex sensitivity (***S****ˡ*), is obtained by setting s = *jω*, to yield:

$$S^1 = G(j\omega) = \frac{K}{j\omega\tau + 1} \quad >(13)$$

Amplitude ratio, the ratio of the amplitude of the output waveform to that of the input, is the modulus of the complex sensitivity.

$$\text{Amplitude ratio} = |G(j\omega)| = \frac{K}{\sqrt{1+\omega^2\tau^2}} \quad >(14)$$

Phase lag (Argument), φ,



$$\phi = -\tan^{-1}\left|\frac{\mathrm{Im}[G(j\omega)]}{\mathrm{Re}[G(j\omega)]}\right| = -\tan^{-1}(\omega t / 1)$$

$$\varphi = -\tan^{-1}(\omega t), \text{ as earlier stated.} \quad >(15)$$

The negative sign is used in equation (15) because the output waveform lags behind the input.

The frequency response of the transfer (i/0) function is therefore fully described as:

$$\frac{y}{x}(j\omega) = \frac{K}{\sqrt{1+\omega^2\tau^2}} \angle -\tan^{-1}(\omega\tau) \quad >(16)$$

If '*y' and 'x'* are of the same form, the static sensitivity K ≈1.

The characteristic form of the output amplitude ratio, equation (16) is shown in Figure 1.13. Shown is the case for $K = 1$ and the frequency axis is generalised in units of ωτ.

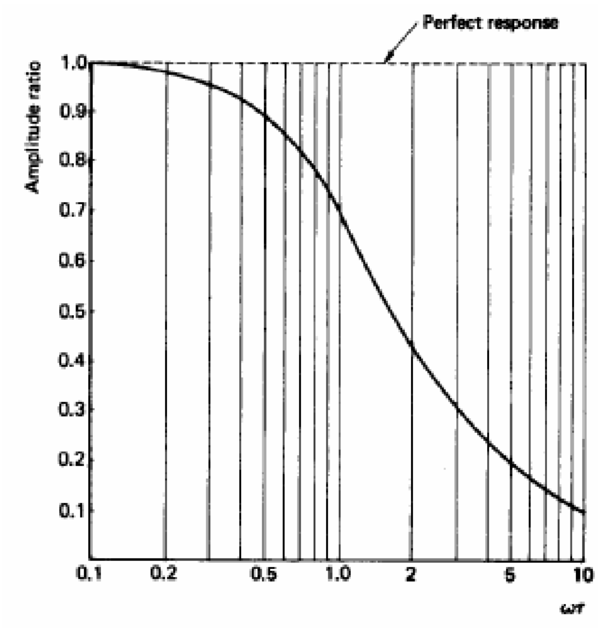


Fig1.14: Harmonic response of 1st Order Systems

Example 1.5

A first-order measuring system has a time constant of 0.01s. Determine the approximate range of input signal frequencies that the system could follow if the error is limited to within 10%.

SOLUTION:

For a10% error the limiting amplitude ratio (Y/X) = 0.9.
From Figure 1.13, this coincides with an upper frequency limit *=0.5 rad,*
Given *= 0.01s*

⇒ *= 2f =0.5/0.01*

SOLUTION:

Given *= 6ms=0.006s and = 1 rad*

⇒ = 2f *=0.5/0.006*

=166.7 rad/s

f =166.7/ 2 = 26.5 Hz

From Figure 1.13 at = *1 rad* the corresponding amplitude ratio is 0.7.

% dynamic error at this frequency =30%

Rearranging this expression

$$\frac{d^2y}{dt^2}+\frac{b}{a}\frac{dy}{dt}+\frac{c}{a}y=\frac{e}{a}x$$

Which can be rewritten as:

$$\frac{d^2y}{dt^2}+2\zeta\omega_n\frac{dy}{dt}+\omega_n^2y=K^*x$$

>(18)

Where

ω_n= undamped natural frequency (rad/s)

ω = frequency of the input variable

ζ = damping ratio

K^* = constant

Under steady state conditions the differential terms in eq (18), d^2/dt^2 and d/dt, are equal to zero, hence: $\omega^2y = K^*x$
When $\omega = \omega_n$

Static sensitivity, K = y/x = K^*/ω_n^2
Applying the D operator to eq(18) gives:

$$\left(D^2+2\zeta\omega_nD+\omega_n^2\right)y=K^*x$$

Dividing through by ω_n^2 yields:

$$\left(\frac{1}{\omega_n^2}D^2+\frac{2\zeta}{\omega_n}D+1\right)y=Kx$$

Where K= K^*/ω_n^2
The systems transfer operator

$$\frac{y}{x}=\frac{K}{\left(\frac{1}{\omega_n^2}\right)D^2+\left(\frac{2\zeta}{\omega_n}\right)D+1}$$

>(19)

$$x \rightarrow \boxed{\frac{K}{\left(\frac{1}{\omega_n^2}\right)D^2+\left(\frac{2\zeta}{\omega_n}\right)D+1}} \rightarrow y$$

Figure 1... ...ator for a Second Order ...n

Step resp... ...r systems

The undamped natural frequency (ω_n) measures the ability of a system to respond rapidly to rapid changes in the input. The higher the value of ω_n a system has, the faster it's *speed of response*. The damping ratio ζ is equal to the ratio of *actual damping* in the system to the *critical damping* condition when:

$\zeta = 1$ — **critical damping**

no oscillation or overshoots in the step response of the system

$\zeta < 1$ — **under damping**

implying oscillation occurs in the step response. For values $\zeta<0.707$ considerable overshoot occurs

$\zeta > 1$ — **over damping**

the system response time is slow and no overshoot or oscillation is seen

Examples: galvanometers, piezo-electric transducers, pen controls in *x-y* plotters.

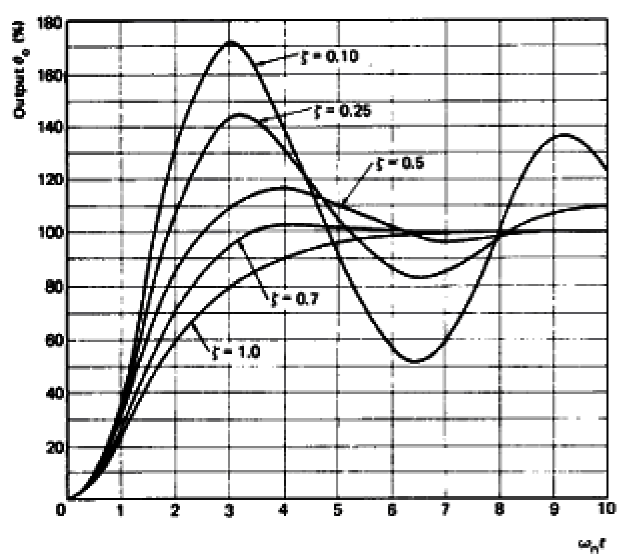


Figure 1.16: Step response of a 2nd order system

In Figure 1.17 the percentage overshoot (% first overshoot/steady-state response) plotted versus damping ratio. This may be used to estimate a value for ζ from the under-damped step response of a system. The quadratic nature of the output solution depends on which case being examined: under, over- or critical damping.

For the case where $\zeta < 1$ the output behaviour is of the form:

$$\theta_{out} = \exp(-\zeta\omega_n t).R \sin\left(\omega_n t\sqrt{1-\zeta^2} + \varepsilon\right) + K$$

where *R* and ε are arbitrary constants determined by boundary conditions.

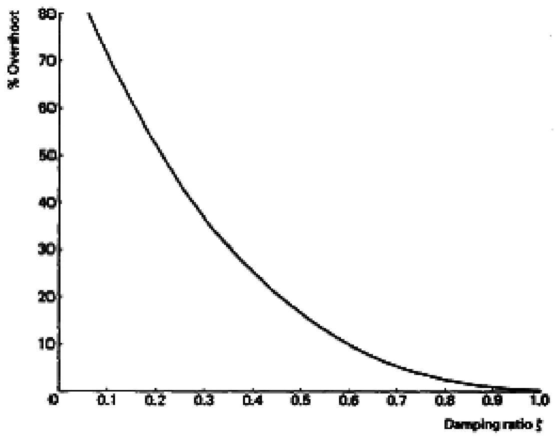


Figure 1.17: Overshoot as a function of the damping ratio

Example 1.7

A mass-spring-damper system has a first overshoot of approximately 40% of its final value when subjected to a step input force. Estimate the values of the damping ratio ζ and ω_n if the time taken to reach the first overshoot is 0.8s from the application of the step.

SOLUTION

From Figure 1.17, □ = 0.28 results in a 40% overshoot.

If an approximation of □ = 0.25 is made then Figure 1.16 can be used to determine □$_n$ as follows:

□$_n$ t = 3.2 units at the first overshoot and t=0.8s

□□$_n$ = 3.2/0.8 = 3.24 rad/s

case the amplitude ratio tends to infinity. As the damping increases the maximum resonance amplitude shifts away from ω_n to lower frequencies and as input frequency increases the output amplitude drops off. The dynamic error in the ideal response diminishes for ζ between 0.6 and 0.7.

For both step and frequency response the optimal value ζ lies in this range and 0.64 is often used as a best compromise.

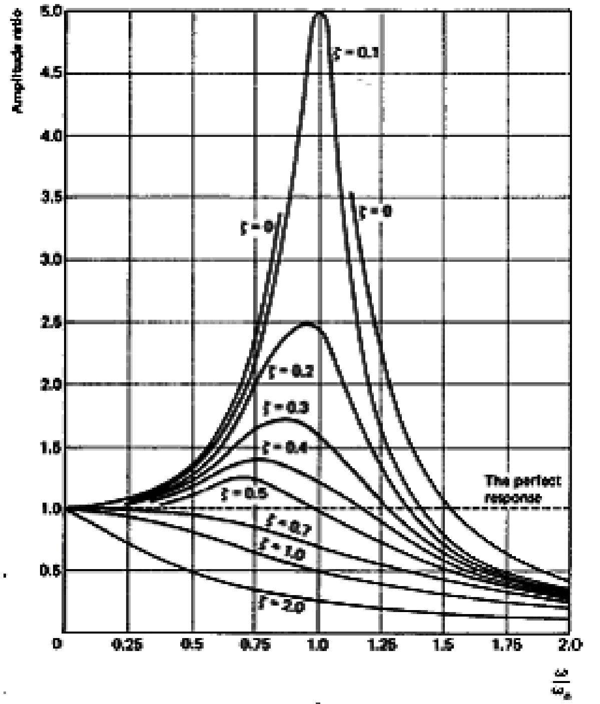


Figure 1.18: Frequency response of a 2nd order system

The frequency (harmonic) response is obtained by setting the input x(t)= sin (ωt).

The output is once again the summation of a transient and steady state response. For the case of $\zeta < 1$ we have:

$$\theta_{out} = \underbrace{\exp(-\zeta\omega_n t)\left[A\sin\left(\omega_n t\sqrt{1-\zeta^2}\right) + B\cos\left(\omega_n t\sqrt{1-\zeta^2}\right)\right]}_{transient}$$

$$+ \underbrace{\frac{K\omega_n^2}{\sqrt{(\omega_n^2-\omega^2)^2 + (2\zeta\omega_n\omega)^2}}\cos(\omega t+\phi)}_{steady-state}$$

A and *B* are again arbitrary constants with the phase term given by:

$$\phi = \tan^{-1}\left[\frac{2\zeta\,\omega_n\omega}{\omega^2 - \omega_n^2}\right]$$

The Laplace transform determination of the frequency response of a 2nd order system

The transfer operator

$$\frac{y}{x} = \frac{K}{\left(\frac{1}{\omega_n^2}\right)D^2 + \left(\frac{2\zeta}{\omega_n}\right)D + 1}$$

>(20)

Transfer function

$$G(s) = \frac{L\,[y(t)]}{L\,[x(t)]} = \frac{Y}{X}$$

$$= \frac{K}{\left(\frac{1}{\omega_n^2}\right)s^2 + \left(\frac{2\zeta}{\omega_n}\right)s + 1}$$

>(21)

$$X \rightarrow \boxed{\dfrac{K}{\left(\dfrac{1}{\omega_n^2}\right)s^2+\left(\dfrac{2\zeta}{\omega_n}\right)s+1}} \rightarrow Y$$

Figure 1.19: The transfer function for a 2nd Order system

Set s =$j\omega$ in (21) yields the *Complex sensitivity*

$$S^1 = G(j\omega) = \frac{K}{\left(1-\left(\dfrac{\omega}{\omega_n}\right)^2\right)+j2\zeta\left(\dfrac{\omega}{\omega_n}\right)}$$

$$= \frac{K}{\left(1-r^2\right)+j2\zeta r}$$

>(22)

Where, r = ω_n / ω_n = frequency ratio

$$\text{Amplitude ratio} = \left|G(j\omega)\right| = \frac{K}{\sqrt{\left(1-r^2\right)^2+\left(2\zeta r\right)^2}}$$

>(23)

The phase lag (or argument) is:

$$\phi = -\tan^{-1}\left|\frac{\mathrm{Im}[G(j\omega)]}{\mathrm{Re}[G(j\omega)]}\right| = -\tan^{-1}\left(\frac{2\zeta r}{1-r^2}\right)$$

>(24)

∴The steady state frequency response is stated as:

$$\frac{y}{x}(j\omega) = \frac{K}{\sqrt{\left(1-r^2\right)^2+\left(2\zeta r\right)^2}} \angle -\tan^{-1}\left(\frac{2\zeta r}{1-r^2}\right)$$

>(25)

Example 1.8

In a frequency-response test on a second-order system, resonance occurred at frequency of 216Hz, giving a maximum amplitude ratio of 1.36. Estimate the values of ζ□and ω_n for the system.

SOLUTION

Examination of Figure 1.18 indicates that □□= 0.4 gives an amplitude ratio of 1.4, □□□□0.4

This resonance occurs at □□□$_n$ = 0.8

But □□ = □*f* = □□□x 216Hz

= 1357 rad/s

Example 1.9

A second order instrument is subjected to a sinusoidal input with an undamped natural frequency of 3Hz. Determine the frequency response for the system at 2Hz if the damping ratio is 0.5. Take K≈1.

SOLUTION

Given: *f =2Hz; f_n=3Hz; □=0.5; K* 1; *Required*: |G(*j*□)|

Frequency ratio, *r = f / f_n =2/3=0.67*

EMBED Equation.3

EMBED Equation.3

Similarly

EMBED Equation.3 = -50.6^0

Hence the frequency response is 1.153 -50.6^o [units]

Consider a high-order instrumentation system made up of three devices with individual frequency responses as shown in Fig 1.20 below.

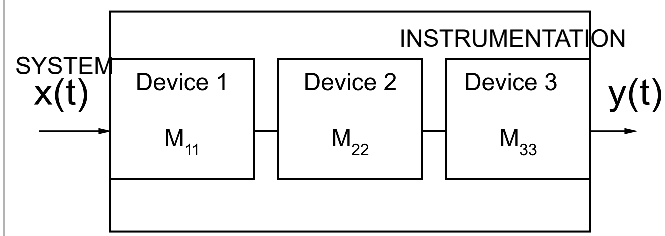


Figure 1.20: General configuration for high order instrumentation systems

For this system,

Overall frequency response = M ∠ φ

Where M = $M_1 \times M_2 \times M_3$

$\varphi = \angle(\varphi_1 + \varphi_2 + \varphi_3)$

Dynamic Specifications

When specifying the dynamic and characteristics of a system the following terms are often used:

Step Response

- *Response Time (*t_{res})

Time for the output to rise from 0% to 100% of its steady state for the first time. Obviously only applies to under-damped systems.

- *Rise Time (*tr)

Time to rise from 10 to 90% of steady state

- *Settling Time*

Time taken for the output to reach and **stay** within a certain percentage band of its steady-state. Typically 2% or 5%.

Frequency Response

Frequency response characteristics are specified depending on the type of device. There are 2 basic classes of devices: AC and DC.

1. *AC Devices*

Amplitude ratio is usually constant over a given range of frequencies but falls off at lower and higher frequencies (Figure 1.21). The ***bandwidth*** is the range and frequencies over which the amplitude is constant to with –3 dB (reduction of about 30% in amplitude)

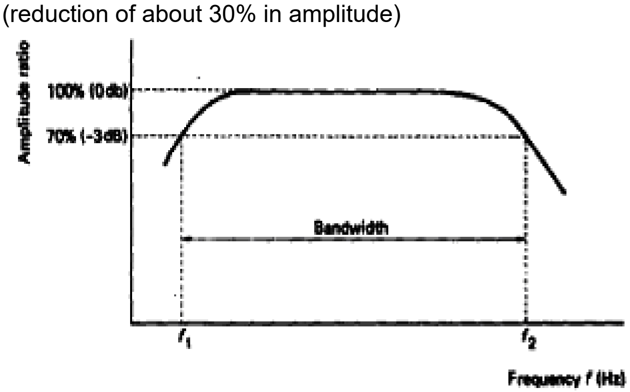


Figure 1.21: Bandwidth of an AC device

2. *DC Devices*

These devices actually respond to both DC and AC signals. Unlike AC devices there is *no lower limit* to frequency response but there remains an upper limit (see Figure 12). The upper limit may be specified as the frequency at which the amplitude ratio drops outside a certain tolerance range. Examples include ±3%, ±5%, -3dB or ±dB.

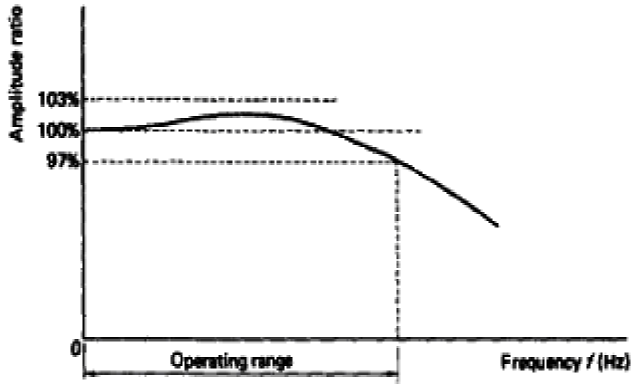


Figure 1.22: Operating range of a DC device

1.4 ELEMENTS OFAN INSTRUMENTATION SYSTEM

Any measuring system or instrument can be described as being made up of 3 basic elements:

1. **transducer** - sometimes called a sensor, this is the element which converts the quantity to be measured (the *measurand*) to another type of signal e.g. speed to voltage, temperature to voltage or flowrate to pressure
2. **signal conditioner** - processes the raw signal from the transducer into a signal compatible with the display/recorder system. Here is where we might amplify the signal or use signal processing techniques to reduce noise, filter or digitise the signal.
3. **recorder/display** - used to present the measurement to the 'end-user'. This can be an LED display, a needle gauge, data-logger/pc, etc.

Typical applications of instrumentation systems: Process control, general industry, medical and biological, environment and survey, military and aerospace, Research.

Review exercise 1

1. A diaphragm pressure gage using a Wheatstone bridge circuit of electrical resistance strain gages has the following specifications: -
- Pressure range = 200 kPa
- Bridge output = 2 mV/kPa
- Accuracy = ± 0.5 % of FS
- Linearity and hysteresis = ± 0.4 % of FS

Explain each of the terms used with respect to the instrument's static characteristics.

2. What are loading errors? A variable resistance potentiometric transducer having a resistance of 12 kΩ is connected to a dc voltage source of 60V. The voltage output of the transducer is measured by means of a voltmeter of internal impedance of 120 kΩ. Determine impedance-loading error at 25% position on the transducer and the actual voltage reading observed at this position.

3. (a) Show that the governing equation for a thermal measuring element is

$$\frac{mc}{ha}\frac{d\theta}{dt}+\theta=\theta_i$$

Where
θ_i = $\theta_i(t)$ – medium temperature (input)
θ = $\theta_o(t)$ – temperature of measuring element (output)
m = mass of thermal measuring element
c = specific heat of thermal measuring element
h = heat transfer coefficient between medium and the measuring element
a = exposed surface area of measuring element

(b) A thermocouple-based temperature measuring system is used to measure the temperature of a heating medium, which changes sinusoidally between 300 and 350 ºC with a periodic time of 20s. The thermocouple bead has an approximate shape of a sphere, 1.54mm diameter, with a density of 7360 kg/m^3 and specific heat of 100 J kg^{-1} K^{-1}. If the bead is exposed to a convection environment where the heat transfer coefficient is 300 kJh^{-1}m^{-2} K^{-1}, find the maximum and minimum values of temperature as indicated by the measuring system and the time lag between the output and input signals.

4. A thermal measuring element with a time constant of 0.4 s and a static sensitivity of 0.05 mV/ºC is used to measure the temperature of a medium that changes from 20 to 60^{0}C. Taking the output as zero at 20^{0}C, find (i) the time taken for the output voltage to reach 80% of the steady state value, if the temperature change occurs suddenly; (ii) the output voltage at the end of 5 s, if the temperature change from 20 to 60ºC, occurs at a constant rate in 5 s. [Hint: Ramp response for a First order system, $x_o(t) = Km(t - \tau + \tau e^{-t/\tau})$, where m is the gradient of the input ramp]

TOPIC 2.0
REVIEW OF ELECTRIC CIRCUITS

Many instruments increasingly use electrical or electronic technology. Some fundamental principles are reviewed below.

2.1 DEFINITIONS

Current *(unit Ampere, symbol A):* Essentially the rate at which electrons move through a circuit

1A ≈6.25 x 10^{18} electrons per second

where each electron carries a charge of -1.6 x 10^{-19} coulombs.

NOTE. Conventionally ***current flow*** direction is **opposite** to ***electron flow*** direction.

Voltage *(unit Volt, symbol V):* The voltage expresses a difference in ***potential*** between points in a circuit. Think of it as implying a difference in potential energy for the electrons of the circuit. The electrons will flow 'down' the potential to the +ve end of the potential (since electrons are negatively charged).

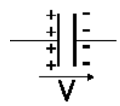
Resistance *(unit Ohm, symbol Ω):* When a voltage drives a current through a piece of material the electrons collide inelastically with the atoms of the material - losing energy. This energy is given up as heat. Different materials resist the current flow by differing amounts. The resistance is defined by the ratio of voltage/to current (V/I)

***Electrical Power** (unit Watt, symbol W):* The power lost due to resistance is easily calculated if any two of the following are known: the voltage, current or resistance

$$P = IV = I^2R = V^2/R$$

***Capacitance** (unit Farad, symbol F):* The capacity of a device (a capacitor) to store charge (i.e. electrons). The separation of +ve and -ve charge without current flow leads to a potential difference (i.e. a voltage).The current through a capacitor is given by:

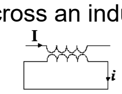
$$I = C\frac{dV}{dt}$$



Therefore if we supply constant current to a capacitor we will get a linear increase in voltage (a voltage ramp).
The value of a capacitor can be increased by the use of an insulating material (a dielectric) placed between the plates of the capacitor. The amount by which the capacitance is increased is a material property of the dielectric and is called its ***permitivity*** (symbol ε_r, and is dimensionless).

***Inductance** (unit Henry, symbol H):* The ability of a device to induce a current to flow in a secondary circuit by interaction of magnetic fields. A changing current through a coil will produce a magnetic field. Similarly a changing magnetic field will produce a current. Bringing two coils close enough together will induce current to flow in the 2nd, initially passive coil. Voltage across an inductor is given by:

$$V = L\frac{dI}{dt}$$



Therefore, applying a constant voltage gives a current ramp. We can increase the inductance of a coil by placing a ferro-magnetic material in the coil. The factor by which the inductance is increased is a property of the material used and is referred to as its ***permeabilit***y, (symbol μ_r, and is dimensionless).

***Decibel** (symbol dB):* When signals are amplified (or attenuated) we express the gain in terms of the ratio of output to input, on a logarithmic scale - the decibel scale.
Amplitude gain: If we consider the amplitude of a voltage signal with input/output given by V_{in} and V_{out} respectively, the amplitude gain in decibels is given by:

$$dB = 20\log_{10}\left(\frac{V_{out}}{V_{in}}\right)$$

Power gain: considering the power input/output

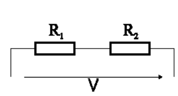
$$dB = 10\log_{10}\left(\frac{P_{out}}{P_{in}}\right)$$

(Recall that $P \propto V^2$or I^2).

2.2 *PASSIVE DEVICES*

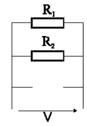
Resistors:

In ***serie***s: $R_{TOTAL} = R_1 + R_2 + \ldots + R_n$



In ***parallel***:

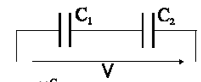
$$\frac{1}{R_{TOTAL}} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$$



Capacitors:

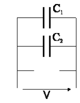
In ***series:***

$$\frac{1}{C_{TOTAL}} = \frac{1}{C_1} + \frac{1}{C_2} + \ldots + \frac{1}{C_n}$$



In ***parallel:***

$$C_{TOTAL} = C_1 + C_2 + \ldots + C_n$$



Energy stored in a capacitor when charged to a voltage V:

$$E_C = \frac{1}{2}CV^2 \text{ (Joules)}$$

Inductors:

Inductors are not generally combined in the same way as resistors and capacitors. However, they can store energy — when the current stops flowing in the primary coil, the current continues for a time in the secondary coil - inducing a secondary current in the primary.

The ***energy stored*** in an inductor when a current I is flowing in the primary is given by:

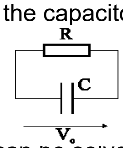
$$E_L = \frac{1}{2}LI^2 \text{ (Joules)}$$

2.3 *COMBINING COMPONENTS*

RC Circuits

If we have a fully charged capacitor with a voltage between the plates of V_o and discharge the capacitor through the resistor R as shown, we have:

$$I = C\frac{dV}{dt} = -\frac{V}{R}$$



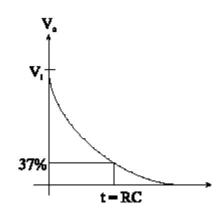
This differential equation in *V* can be solved:

$$V(t) = V_0 e^{-t/\tau}$$

Where τ =time constant (in seconds) =RC.

When t = t,

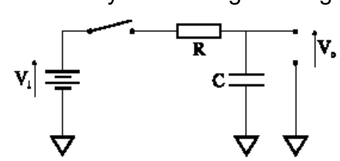
$$V(t) = \frac{V_0}{e} \approx 37\% V_o$$



Charging the capacitor from 0 volts by connecting a voltage across yields:

$$I = C\frac{dV}{dt} = \frac{V_i}{R} - \frac{V}{R}$$

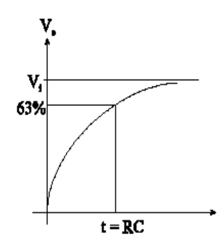
$$\Rightarrow V(t) = V_i + V_o e^{-t/\tau}$$



If at *t = 0, V(t) = 0* (initially the capacitor is discharged); $V_o = -V_i$; Hence,

$$V(t) = V_i(1 - e^{-t/\tau})$$

Now when $t = \tau = RC$ the voltage has reached 63% of its final value V_i

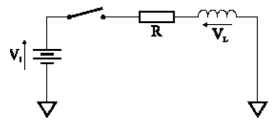


Note: The time constant (τ=*RC)* determines how quickly a circuit can react to a sudden change in input. This can be useful if the signal has high frequency noise - the time constant is fast enough to allow the signal through but not fast enough to allow the noise through. Resistance-capacitors can therefore be used to ***filter*** a signal.

LR Circuits

Consider the circuit with an inductor and resistor. We shall again consider the transient response as we close the switch S.

$V_S = V_L + V_R$



We have for V_L and V_R: $V_L = L\dfrac{dI}{dt}$

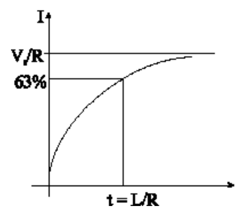
$VR = I\,R$

Therefore,

$$V_S = L\frac{dI}{dt} + IR \qquad \Rightarrow \frac{dI}{dt} = \frac{V_S - IR}{L}$$

Solving the differential equation using the boundary condition: $I = 0$ when $t = 0$; i.e. no current flows when the switch is open.

$$I(t) = \frac{V_S}{R}\left(1 - e^{-t/\tau}\right)$$



Here, $\tau = L / R$

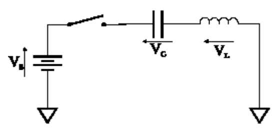
Again the transfer characteristic is appropriate for filter applications.

<u>LC Circuits (Resonance Circuits)</u>

Consider the circuit containing the inductor and capacitor shown below. If C is fully discharged at $t = 0$ then:

Total voltage = (voltage drop across inductor) + (voltage drop across capacitor)

$V_S = V_L + V_C$



Recall that $I = C\,(dV_C/dt)$

$$V_S = L\frac{dI}{dt} + \frac{1}{C}\int_o^t I\,dt$$

V_S is constant, hence differentiating it with respect to time and rearranging gives:

$$\frac{d^2I}{dt^2} = -\frac{1}{LC}I$$

This is similar to the equation for simple harmonic motion:

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

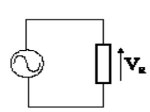
Thus the circuit will exhibit oscillation with a natural frequency:

$$\omega_n = \frac{1}{\sqrt{LC}}$$

2.4 *AC CIRCUITS*

Applying an alternating current (AC) to the 3 components seen above shows marked differences in their response.

<u>Resistors</u>

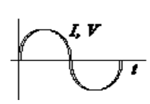


If we have the current varying sinusoidally:

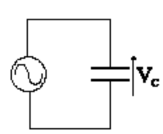
$I = I_o \sin wt$

From Ohms law: $V = IR;\ V = R\,I_o \sin \omega t$

Note that I and V are always in phase and the voltage is proportional to the current and the constant of proportionality is R.



<u>Capacitors</u>



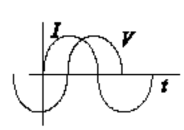
Again we assume: $I = I_o \sin \omega t$

We have the relationship between V and I as:

$$V = \frac{1}{C}.\int I dt$$

$$V = \frac{1}{C}\int I_o \sin \omega t\, dt$$

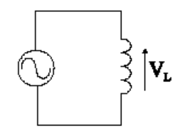
$$V = -\left(\frac{1}{\omega C}\right) I_o \cos \omega t$$



Note the term ($1/\omega C$) has units of resistance (Ohms) but depends on frequency. It is called the ***capacitive reactance*** (χ_c). Larger frequencies will see smaller “resistance” and so pass easily through the circuit while lower frequencies see more resistance - ***high pass*** behaviour.

Note also, the current is in terms of $\sin(\omega t)$ while the voltage is in terms of $-\cos(\omega t)$ which implies we have a 90° phase difference between the current and voltage.

<u>Inductors</u>



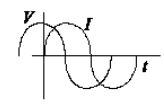
Again we assume: $I = I_o \sin \omega t$

We have the relationship between V and I as:

$$V = L\frac{dI}{dt}$$

$$V = L\frac{d}{dt}\left(I_o \sin \omega t\right)$$

$$V = (\omega L)\ I_o \cos \omega t$$



We similarly have an ***inductive reactance*** (χ_L) in Ohms:

$\chi_L = \omega_L$

Here lower frequencies see lower “resistance” while higher frequencies are attenuated (i.e. scaled down) - **low pass** behaviour. In the case where the current is in terms of $\sin(\omega t)$ while the voltage is in terms of $+\cos(\omega t)$: we again have a 90°phase difference between the current and voltage but in a different sense.

2.5 *COMPLEX NOTATION / PHASOR DIAGRAMS*

A harmonic wave has the form $A\cos(\omega t)$ where A is the amplitude of the wave and the argument of the cos function is its phase. From Euler’s equation we have the relationship between a complex quantity and harmonic waves:

$e^{i\omega t} = \cos(\omega t) + i\sin(\omega t)$

Often the complex conjugate ($i\sin(\omega t)$) is ignored and the harmonic wave is written in complex notation as:

$A\,e^{iwt} \equiv A\cos(\omega t)$

where ωt is our phase term and A is our amplitude.

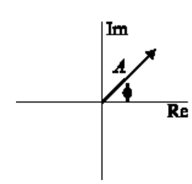
We can represent an expression in the form $A\,e^{i\varphi}$ as a vector in the ***complex plane***. This vector being known as a phasor since it contains information about the amplitude and phase of the wave it represents.

Our complex notation can be seen as the polar co-ordinate of the end-point of the phasor in the complex plane. Converting to rectangular co-ordinates we see our phasor as the sum of a real part (along the x-axis) and an imaginary part (along the y-axis):

$A\,e^{i\varphi} = x + iy;$

where A is the amplitude and φ is the phase.

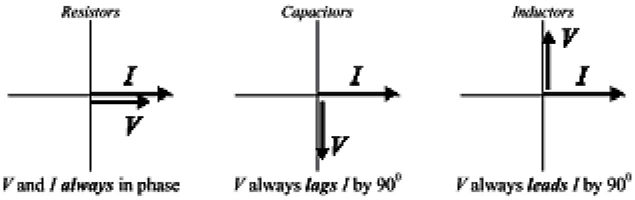
Plotting the amplitude as the magnitude of a vector in the complex plane and the phase as the angle of the vector to the real axis we have the



representation shown in the figure right.

Phasor diagrams are useful in illustrating the relative phase between different quantities.

R, C and L when subjected to an AC signal give the following relative behaviours, of the current and voltage through each of these circuits, as shown using phasor diagrams.

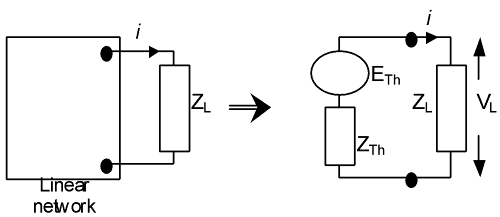


2.6 CIRCUIT ANALYSIS

In order to describe voltage and current behaviour at the connection of two elements there is need to represent each of the elements by equivalent circuits characterised by two terminals.

2.6.1 *Thevenin's equivalent circuit*

Thevenin's theorem states that *any network consisting of linear impedances and voltage sources can be replaced by an equivalent circuits consisting of a voltage source E_{Th} and series impedance Z_{Th}.* The source E_{Th} is equal to the open circuit voltage of the network across the output terminals, and Z_{Th} is the impedance looking back into these terminals with all voltage sources reduced to zero and replaced by their internal impedances.



Thevenin's equivalent circuit

Connecting a load Z_L across the network terminals is equivalent to connecting Z_L across the Thevenin circuit. The current *i* in Z_L is given by

$$i = \frac{E_{Th}}{Z_{Th} + Z_L}$$

and the voltage V_L across the load by

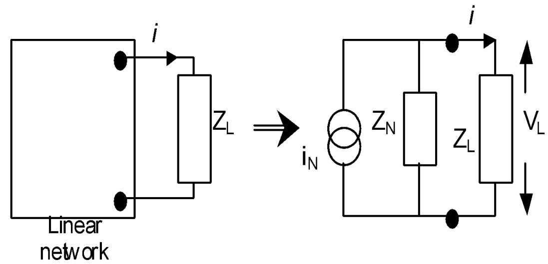
$$V_L = i\,Z_L = \frac{Z_L}{Z_{Th} + Z_L} E_{Th}$$

Maximum voltage transfer from the network occurs when $Z_L >> Z_{Th}$ (i.e. the load impedance is far greater than the network impedance) so that $V_L \rightarrow E_{Th}$.

Maximum power transfer from the network occurs when $Z_L = Z_{Th}$ (i.e. the load impedance is equal to the network impedance) so that $V_L = ½\, E_{Th}$.

2.6.2 Norton's equivalent circuit

Norton's theorem states that *any network consisting of linear impedances and voltage sources can be replaced by an equivalent circuit consisting of a current source i_N in parallel with an impedance Z_N.*



Norton's equivalent circuit

Z_N is the impedance looking back into the output terminals with all voltage sources reduced to zero and replaced by their internal impedances, and i_N is the current that flows when the terminals are short-circuited.

Connecting a load Z_L across the output terminals of the network is equivalent to connecting Z_L across the Norton circuit. The combined circuit impedance Z is given by $Z_N || Z_L$, i.e.

$$Z = \frac{Z_N.Z_L}{Z_N + Z_L}$$

The voltage V_L across the load by

$$V_L = i_N Z = i_N \frac{Z_N.Z_L}{Z_N + Z_L}$$

In order to develop the maximum current through the load $Z_L << Z_N$ (i.e. the load impedance is far smaller than the Norton impedance for the network) so that $V_L \rightarrow i_N Z_L$.

TOPIC 3 CLASSIFICATION AND SELECTION OF TRANSDUCERS

3.1. INTRODUCTION

A transducer is *a device that produces a usable output in response to a specific input measurand.*

Related terms: sensor, primary sensing element, detector, gage, pick-up, probe, converter.

A transducer consists of one or more sensing elements that perform the conversion of a physical (non-electric) stimulus, occurring at the input terminals, into a corresponding electric signal that is then made available at the output terminals.

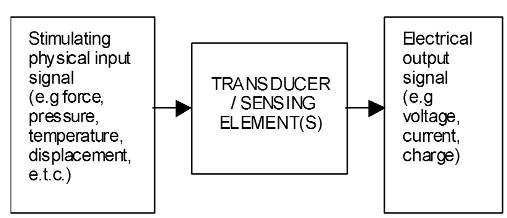


Fig 3.1 (a) The instrumentation transducer

3.2 ***CATEGORIES OF TRANSDUCERS***

Transducers can be categorised according to a broad range of considerations, which include:

- source of power (active or passive)
- number of stages in the conversion (simple and compound)
- frequency response characteristics (static and dynamic)
- nature of the output (analogue and digital)
- control actions (feedback/inverse)

3.2.1 *Active and passive transducers*

Active (self-generating) transducers: draw their energy required to activate an output response from the measurand. They are called self-generating because they do not require an external energising source to be able to operate.
Examples: piezo-electric vibration sensor, photovoltaic cell based units, thermoelectric and electromagnetic devices.

Passive transducers (or modulators): draw their activation energy from an external power source such as an AC or DC power supply. They are also called modulators because they operate under energy controlling principles, using the measurand to scale (i.e. modulate) the signal from the power supply to give a proportional output.
Examples: potentiometric displacement transducer, strain gage, variable capacitance level sensors.

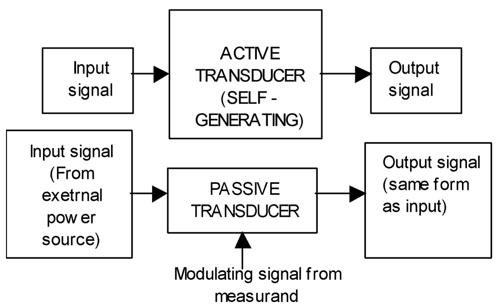


Fig 3.1(b) Active and passive transducers

3.2.2 *Simple and compound transducers*

Simple transducers are those where the conversion of the input from physical to electrical form (i.e. transduction) takes place in a single stage. They are single device transducers whose transfer characteristics match those of the constituent sensor.
Example: Piezo-electric crystal based force/pressure sensor.
Compound transducers are those having two or more stages of transduction. They are made up of a combination of sensors and the transfer (i/o) characteristic is the overall for the combined system.
Example: A pressure converter may incorporate a diaphragm as the primary sensing element coupled to a Linear Variable Displacement Transformer (LVDT) for conversion of the diaphragm movement into an electric output.

3.2.3 *Static and dynamic transducers*

Static (steady state) transducers are those whose transfer characteristics are time/frequency independent. A change in the input variable gives an almost immediate output response.
Examples: most displacement measuring transducers
Dynamic transducers are those whose transfer characteristics are frequency dependent. A time lag and phase difference exist between the input and output signals.
Examples: vibration/acceleration sensors, velocity sensors.

3.2.4 *Analogue and digital transducers*

Analogue transducers are dynamic devices whose output is continuous with respect to time.
Example: The X-Y plotter used in graphical recording, strain gages, thermocouples.

Digital transducers produce time –discontinuous (discrete) signals at their output terminals
Examples: analogue to digital converters.

3.2.5 *Feedback (inverse) transducers*

These transducers have primary sensing elements whose outputs are amplified and fed back into the input in polarity opposition. The device is located in the feedback path and its output signal is used in control as a measure of response to control action. These devices are common with process control instrumentation systems.
Example: speed governor used on automobiles.

3.3 *TECHNICAL CONSIDERATIONS IN TRANSDUCER SELECTION*

The following should be considered when selecting a transducer:

1. *General:* Manufacturer, model/series, media of measurand, measurement range, device applications and limitations.
2. *Electric characteristics:* Transduction principle, primary sensing element, sensitivity, transfer function, excitation, resolution (% FS).
3. *Output characteristics:* AC/DC, voltage range (V), power range (W), impedance (Ω), frequency range, linearity (%FS), hysteresis and creep (%FS), static error band, zero drift sensitivity (%FS/^{0}C), repeatability, threshold, overload, natural frequency, dynamic response, acceleration response (%FS/*g*), rise time, stability, temperature range, calibration, zero adjustment.
4. *Mechanical:* physical dimensions, weight, mounting specifications, construction material, life expectancy
5. *Connections:* inlet and outlet terminals, type of electric output, accessories, operational limits.
6. *Environment:* temperature, humidity, acceleration, shock, vibration, magnetic / electrostatic fields
7. *Cost*

3.4 *TRANSDUCTION PRINCIPLES*

Instrumentation transducers make use of one or more of following principles:

i. Variable *resistance* – e.g. the potentiometric resistance transducer, strain gauge
ii. Variable *inductance* –e.g. Linear Variable Displacement Transformer (LVDT)
iii. Variable *capacitance* – e.g. the variable air-gap capacitance transducer
iv. Piezo-electric principle – e.g. the piezo-electric transducer

3.4.1 *Resistance transducers*

These devices convert mechanical displacement into a change in resistance.

3.4.1.1 *Potentiometric Resistance Transducer*

The primary sensing element (PSE) is a variable resistance potentiometer consisting of a movable wiper contact that is coupled to an insulated plunger shaft mechanically linked to the point under measurement. Three types of potentiometric transducers are: wire-wound, carbon film and plastic film.

Construction: Fig. 3.2 gives the general construction of potentiometric resistance transducers. The ***wire-wound*** transducer has three functional construction elements:

a) *Resistor winding wire*: consists of precision drawn resistance wire of diameter 25 –50 microns, annealed in a reducing atmosphere to prevent surface oxidation. Wire resistivity: 0.4 – 1.3 μΩm; Temperature coefficient: 0.002 –0.01% per °C; Material: alloys of copper-nickel, nickel-chromium and silver-palladium
b) *Winding former:* consists of a cylindrical or flat mandrel of ceramic glass, anodised aluminium, steatite or moulded epoxies. Its construction should give good dimensional stability and surface insulation.
c) *Contact wiper:* consists of spring elements made from tempered phosphor bronze or other precious metals.

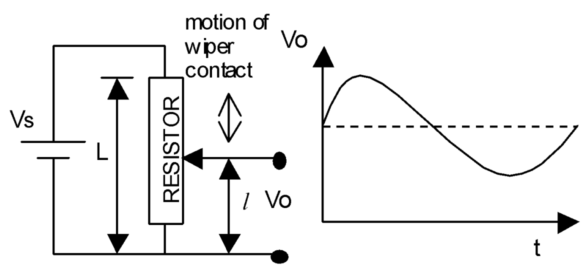


Fig 3.2 Potentiometric resistance transducers

Operation: The body whose motion is being measured is connected to the sliding element of the potentiometer, so that translational motion of the body causes a motion of equal magnitude of the slider along the resistance element and a corresponding change in the output voltage V_o. As the slider moves along the potentiometer track it makes contact with successive turns of the wire coil. This limits the ***resolution*** of the device to the distance from one coil to the next. Using a carbon or conducting plastic-film for the resistance element can achieve much better measurement resolution (up to 10^{-4}). Typical ***accuracy***, which is quoted for these devices, is $\pm$1% of the full-scale reading.
The life expectancy is quoted in terms of the number of reversals, i.e. the number of times the slider can be moved backwards and forwards along the track. Figures quoted for wire-wound, carbon-film, and plastic-film types are 1 million, 5 million and 30 million respectively. Wire –wound types have a lesser temperature coefficient compared to the rest.

Transfer characteristics: The following input/output relation can be deduced from Fig. 3.2

The output voltage (V_o) under ideal conditions:

$$V_O = \frac{\text{resitance at output terminals}}{\text{resistance at input terminals}} \times \text{input volage}$$

$$= \frac{l}{L} Vs \qquad > (3.1)$$

Advantages: Linear transfer characteristic; simple construction, low cost.

Disadvantages: Operational problems occur at the point of contact between the slider and resistance element; poor dynamic response

Example 3.1
The output voltage from a translational motion transducer of stroke length 0.1m is measured by a voltmeter with an internal resistance of 10kΩ. The maximum measurement error due to loading, which occurs when the slider is positioned two-thirds of the way along the resistor element (i.e. l / L =2/3), must not exceed 1% of the full-scale reading. If a family of potentiometers having a power rating of 1watt per 0.01m and resistances ranging from 100Ω to 1kΩ in 100Ω steps are available;

a) choose the most suitable potentiometer from the given range;
b) find the supply voltage for the highest possible measurement sensitivity
c) calculate the measurement sensitivity.

SOLUTION
Let resistance portion AC, R_{AC} = R_i and the total resistance of the element, R_{AB} = R_t. Also let the output voltage measured be V_m.
Circuit diagram
EMBED MSDraw.Drawing.8.2
Given data:
R_i / R_t =2/3 R_i = 2 R_t/3
Voltmeter resistance, R_V = 10 k = 10,000
Ideally EMBED Equation.3

Considering loading effects, the voltmeter resistance effectively reduces the resistance between AC to the sum of two resistances (R_i and R_v) in parallel.
EMBED Equation.3
The actual measured voltage, by the voltage divider rule becomes:
EMBED Equation.3 EMBED Equation.3

Maximum error = 1% FS = 0.01V_s, since at full scale the voltage drop across AB is equal to the supply voltage V_s.
EMBED Equation.3
EMBED Equation.3
Which simplifies into EMBED Equation.3
Substituting R_V =10,000 in this expression gives R_t ≈ 685

(a) The nearest value in the range of potentiometers available is 700. The 700 potentiometer gives a measurement error less than 1% and is therefore selected.

(b) The thermal rating of the potentiometers is quoted as 1W per 0.01m, i.e. 10 W for the total length of 0.1m.

By Ohm's law: $P = IV = I^2R = V^2/R$
V_{max} = (P_{max} x R) =(10 x 700)= 83.67 V

- Measurement sensitivity = output range/input range

3.4.1.2 ***Resistance strain gauge***
Strain gauges are the most commonly used transducers for the measurement of displacement. The resistance gauge consists of a fine wire which changes its resistance when mechanically strained due to physical effects. The gauge is usually mounted onto the measured surface so that it elongates or contracts with that surface. The resistance change from a strain gauge is usually coupled to a Wheatstone bridge for conversion to a voltage signal.

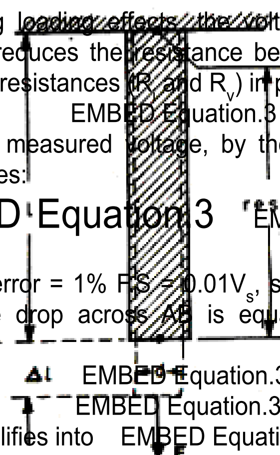


Fig. 3.3 Deformation of a resistance wire under strain

Theory of operation
Consider a conductor of length L [m], cross-sectional area A [m^2] and specific resistance ρ [Ω-m] as shown in Fig. 3.3. The resistance R [Ω] is given by:

$$R = \rho \frac{L}{A}$$

> (3.2)

Differentiating eq. (3.1) and dividing by the total resistance yields:

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dL}{L} - \frac{dA}{A}$$

or expressed in terms of fractional changes

$$\frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} + \frac{\Delta L}{L} - \frac{\Delta A}{A}$$

> (3.3)

The fractional change in the cross section area of the wire can be expressed as:

$$\frac{\Delta A}{A} \approx 2\frac{\Delta d}{d} = -2\nu \frac{\Delta L}{L}$$

> (3.4)

where v is the Poisson's ratio of the wire material.
Hence eq. 3.3 can be rewritten as:

$$\frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} + \frac{\Delta L}{L} + 2\nu \frac{\Delta L}{L}$$

or

$$\frac{\Delta R/R}{\Delta L/L} = 1 + 2\nu + \frac{\Delta \rho/\rho}{\Delta L/L} = G$$

> (3.5)

where G is the gauge factor which is the strain sensitivity of the wire defined as the resistance change per unit of initial resistance divided by the applied strain.

Experimental results show that for most metallic alloys $2 \le G \le 4$, and for pure metals $-12 \le G \le 6$. In both cases $0.25 \le \nu \le 0.35$ which implies that the change in specific resistance can be quite large for some metals (this has its origin in variations in the number of free electrons and their mobility with applied strain).

Gauge construction

There are two basic types of strain gauges:

- ✓ Bonded strain gauge
- ✓ Unbonded strain gauge

Bonded strain gauges (Fig.3.4) are usually bonded directly to the surface of the specimen being tested with a thin layer of adhesive cement which serves to transmit the strain from the specimen to the gauge and at the same time serves as an electrical insulator. Examples include the flat grid wire and the metal foil strain gauges and semiconductor strain gauges.

For good sensitivity and high transmission of strain to the gauge, it is essential that the gauge wire, used in wire grids, should have high resistivity and a large surface area. This is achieved by using a thin wire of long length. The normal gauge factor value is between 2.0 and 2.2, and materials having a higher gauge factor are usually not recommended because of their poor temperature characteristics.

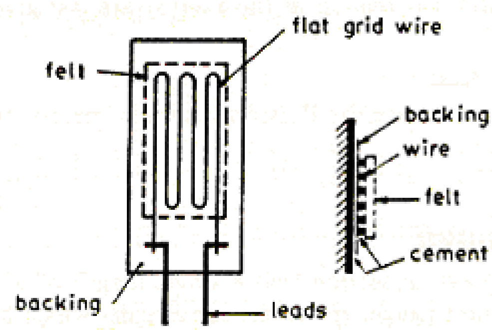


Fig 3.4 (A) Flat grid bonded strain gauge

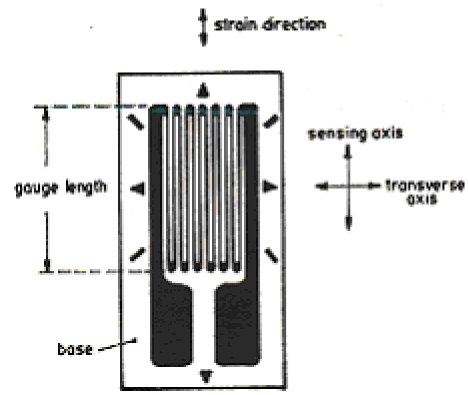


Fig 3.4 (B) Metal foil strain gauge

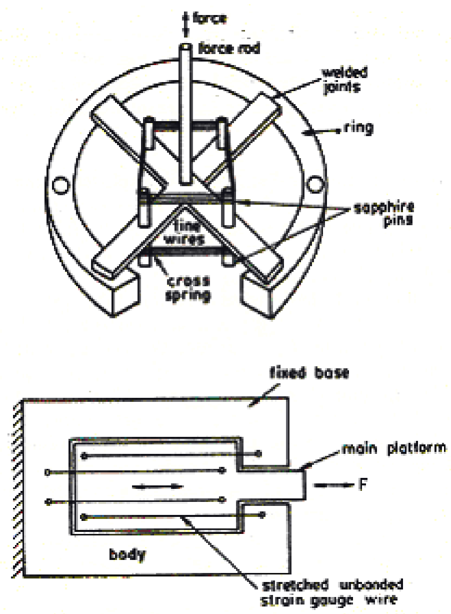


Fig 3.5 Unbonded strain gauge

Small foil gauges of varied patterns are produced by a versatile photo chemical etching process. Typical configurations include circular gauges, diaphragm gauges, orthogonal arrays for shear measurement, as well as Rosettes for two-dimensional stress analysis. In foil gauges there is no stress concentration at the terminals due to the absence of joints, thereby extending their durability.

An unbonded strain gauge device is a free filament sensing element where strain is transferred to the resistance wire directly without any backing (Fig 3.5). The main advantage of this device lies in its radial symmetry which assists in cancellation of spurious signals from transverse forces. The device also has very low hysteresis and creep.

3.4.2 *Variable inductance transducers*

Consist of variable reluctance sensors and the LVDT

3.4.2.1 *The Magnetic circuit*

A simple magnetic circuit (Fig 3.5) consists of a loop or core of ferromagnetic material on which is wound a coil of *n* turns carrying a current *i.* A *magnetomotive force*, m.m.f. (analogous to the electromotive force, e.m.f.) drives a flux φ through the magnetic circuit.

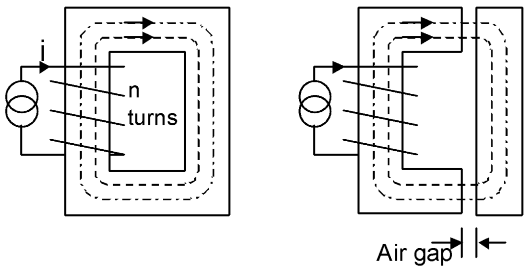


Fig 3.5 The magnetic circuit

For a magnetic circuit

m.m.f = flux x reluctance = $\varphi \times \Re = ni$

$\therefore$ the flux linked by a single coil is

$$\phi = \frac{ni}{\Re} \text{ webers} \qquad >(3.6)$$

The total flux N linked by the entire coil of n turns is:

$$N = n\,\phi = \frac{n^2 i}{\Re} \text{ webers} \qquad >(3.7)$$

The self-inductance of the coil L is defined as:

$$L = \frac{N}{i} = \frac{n^2}{\Re} \qquad >(3.8)$$

The reluctance $\Re$ of a magnetic circuit is given by

$$\Re = \frac{l}{\mu\mu_o A} \qquad >(3.9)$$

where l = total length of the flux path

μ = relative permeability of the circuit material

μ_o = permeability of free space $=4\pi \times 10^{-7}$ H / m

A = cross section area of the flux path

3.4.2.2 *Variable reluctance displacement sensor*

The primary sensing element consists of a magnetic circuit whose core is separated into two parts that are separated by an air gap of variable width.

Construction: The displacement sensor consists of three elements: a ferromagnetic core in the shape or a *semitoroid* (semicircular ring), a variable air gap and a ferromagnetic plate or armature.

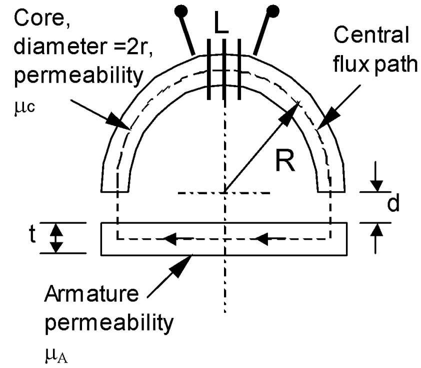


Fig 3.6 Variable reluctance displacement sensor

Operation: One of the elements (either the core or armature) is fixed while the other is coupled to the moving object in such a way as to vary the air gap d. A small variation in the air gap causes a measurable change in inductance which can be calibrated in terms of the input displacement.

Transfer characteristics: The i/o characteristic is derived as follows.

The total reluctance of the magnetic circuit is:

$$\Re_{total} = \Re_{core} + \Re_{gap} + \Re_{armature}$$

$$= \frac{R}{\mu_o \mu_c r^2} + \frac{2d}{\mu_o \pi r^2} + \frac{R}{\mu_o \mu_A r\, t}$$

$$= \Re_o + kd \qquad >(3.10)$$

where

$\Re_o$ = reluctance at zero air gap

$$= \frac{R}{\mu_o r}\left[\frac{1}{\mu_c r} + \frac{1}{\mu_A t}\right] \qquad >(3.11)$$

and

$$k = \frac{2}{\mu_o \pi r^2} \qquad >(3.12)$$

$$\therefore \quad L = \frac{n^2}{\Re} = \frac{n^2}{\Re_o + kd} \qquad >(3.13)$$

which can be rewritten as:

$$L = \frac{L_o}{1+\alpha d} \qquad >(3.14)$$

where $L_o = n^2 / \Re_o$, the inductance at zero air gap and $\alpha = k / \Re_o$. Both L and α depend on core geometry, permeability etc. Equations (3.13) and (3.14) apply for any variable displacement reluctance sensor

Example 3.2

If n=500 turns, R=2 cm, r=0.5 cm, t =0.5 cm, $\mu_c = \mu_A = 100$. Determine $\Re_o$, k and hence find L at d =1mm. Take μ_o= 1.257μ H/m

SOLUTION

Given data:

n=500 turns; R=2 cm =0.02m; r = t =0.5 cm =0.005m

$\mu_c = \mu_A = 100$

μ_o = 1.257μH/m

d =1mm =0.001m

Required: $\Re_o$, k and L

$\Re_o$ = EMBED Equation.3 $[H^{-1}]$

EMBED Equation.3

= $1.27 \times 10^{7}\ H^{-1}$

EMBED Equation.3 $[H^{-1}m^{-1}]$

= EMBED Equation.3 $2 \times 10^{10}\ [H^{-1}m^{-1}]$

L = EMBED Equation.3

= EMBED Equation.3

= 7.64×10^{-3} H or 7.64 mH

3.4.2.3 ***The differential (push-pull) displacement sensor***

The transfer between the input displacement d and the resulting change in inductance is non-linear. This problem is overcome by using a differential (push-pull) displacement sensor consisting of an armature between two cores separated by a fixed distance 2d, incorporated into an a.c. deflection bridge.

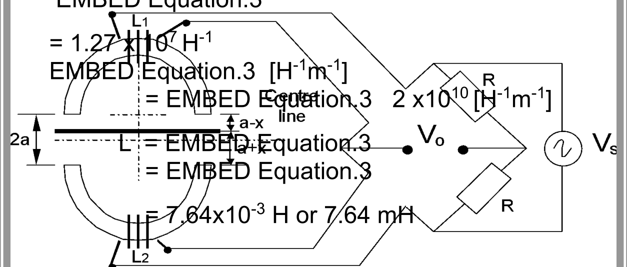


Fig 3.7 Differential (push-pull) displacement sensor

The coil inductances L_1 an L_2 are given by:

$$L_1 = \frac{L_o}{1+\alpha(a-x)}, \quad L_2 = \frac{L_o}{1+\alpha(a+x)}$$

Bridge parameters: $Z_1 = j\omega L_1$; $Z_2 = Z_3 = R$; $Z_4 = j\omega L_2$

$$V_o = V_s\left\{\frac{Z_1}{Z_1+Z_4} - \frac{Z_2}{Z_2+Z_3}\right\}$$

The output voltage V_o, for the inductive push -pull bridge is:

$$V_o = V_s\left\{\frac{L_1}{L_1+L_2} - \frac{1}{2}\right\}$$

From which

$$V_o = \frac{V_s\alpha}{2(1+\alpha a)}x \qquad >(3.15)$$

The transfer characteristic is linear and frequency independent (static).

3.4.2.4 ***Linear Variable Displacement Transformer (LVDT)***

This sensor is a transformer with a single primary winding and two identical secondary windings wound on a tubular ferromagnetic former (Fig. 3.8). The inductance as a function of the variable core displacement is achieved by a variation in mutual inductance of the coils.

Construction: Coils of enamelled copper wire are wound on phenolic or ceramic formers with good dimensional stability. The transformer is enclosed in a ferromagnetic case to provide full electromagnetic shielding. The moving core is made of ferromagnetic material of high permeability selected for optimum performance. It may be heat treated to enhance magnetic properties.

The object whose translational motion is to be measured is physically attached to the central iron core of the transformer, so that motions of the body are transferred to the core.

Operation: The primary coil is energised by an a.c. voltage of amplitude V_s and frequency f_s (Hz). Depending on the relative position of the ferromagnetic core, with respect to a central (null) position in the coil assembly, voltages V_1 and V_2 in the secondary coils. The two secondary coils are connected in series opposition so that the voltage that appears at the transformer's output terminals (V_o) is the difference of the voltages induced in the secondaries.

$$V_o = V_1 - V_2 = (M_1 - M_2)\frac{di_p}{dt} \qquad > (3.16)$$

Where M_1&M_2 are the mutual inductances of the secondary windings due to the relative position of the iron core and i_p is the current in the primary coil. The normal excitation voltage is 1.0 V at a carrier frequency of 2 kHz –10 kHz.

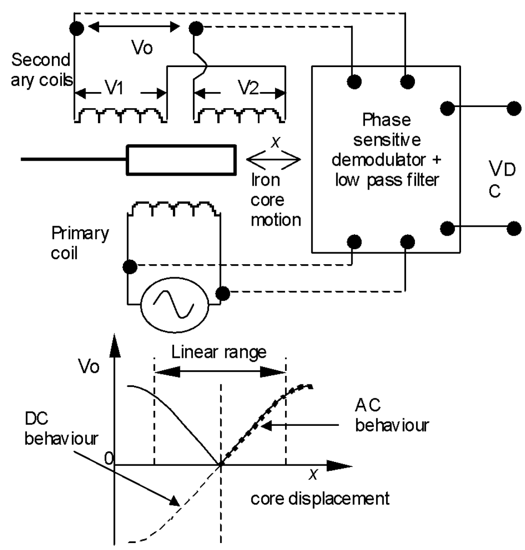


Fig. 3.8 The Linear Variable Differential Transformer

Transfer characteristics: The relationship between the magnitude of the output voltage and the core position is approximately linear over a reasonable range of movement of the core on either side of the null position. While the magnitude of the output voltages are the same on either side, the phase difference between output and input voltages changes by 180° when the core moves through the null position. The ***phase detector***, consisting of a ***phase sensitive demodulator*** and a ***low-pass filter***, senses this changeover and gives a negative or positive voltage output correspondingly. In relation to the other characteristics of the transformer, the i/o relation is stated as:

$$Vo = \frac{16\pi^3 f I_p n_p n_s}{10^9 \ln\left(\frac{r_o}{r_i}\right)} \frac{2bx}{3w}\left(1 - \frac{x^2}{2b^2}\right) \qquad > (3.17)$$

where f = excitation signal frequency; I_P = primary current, n_P =number of turns in primary; n_s = number of turns in secondary; b = width of primary coil; w = width of secondary

coil; x = core displacement, r_o = outer radius of coil r_i and inner radius of the coil.
The sensitivity is proportional to the frequency (f) and the primary current (I_p) and for best linearity x<<b.

Exercise 3.1
A LVDT type displacement transducer with a primary of 300 turns and secondary of 500 turns in each section is excited with a carrier wave signal of 1 kHz, 5 mA current. The width of the primary coil and each secondary coil is10 mm, the thickness of the windings is 4 mm and the radius of the outer coil is 8mm. Calculate the sensitivity of the device and the percentage output linearity at 10% and 20% of the primary coil width.

Advantages of the LVDT: Include -
1. Simple design: easy to fabricate and install
2. Wide displacement measuring range (±10 microns to 10 mm)
3. Frictionless movement of core and hence infinite resolution in output (the only limiting factors are the signal to noise ratio and input stability conditions)
4. Low core weight gives negligible operating force
5. Wide temperature operating range (-40^{0}C to100^{0}C)
6. Linear analogue output voltage (linearity better than 0.25%)
7. High sensitivity (2 mV/volt/10 microns at 4kHz excitation)
8. Low output impedance (100Ω)
9. Wide range of operating carrier frequencies (50 Hz to 20 kHz)
10. Very low cross sensitivity.

3.4.3 *Variable Capacitance Transducers*

These are displacement sensitive devices for which a variation in the configuration of the overlapping electrodes or the dielectric medium between them produces a change in the capacitance between the plates.
In general the governing equation for these devices is:

$$C = \frac{\varepsilon A}{\delta} = \frac{\varepsilon_0 \varepsilon_r A}{\delta} \quad > (3.18)$$

where C = capacitance [μF];
A= overlapping area of plates / electrodes [m^2],
δ = distance of separation between the plates [m];
ε = dielectric constant [F/m]
= $\varepsilon_0\varepsilon_r$; with
ε_0 =permitivity of free space ≈ 8.854 x 10^{-12}F/m
ε_r =relative permitivity of the medium between the plates. For air ε_r =1

Construction: The capacitance, as seen in eq. (3.18), varies in relation to δ, A and ε. The three practical configurations of variable capacitance transducers are:
- variable *plate separation, δ*
- variable overlap area, A
- variable dielectric, ε

The three are discussed below.

3.4.3.1 *Variable plate separation capacitance transducers*

The object whose displacement is to be measured is coupled to the capacitor in such a way that the perpendicular distance between the overlapping surfaces of the electrodes varies causing an inversely proportionate change in capacitance (Fig 3.9).

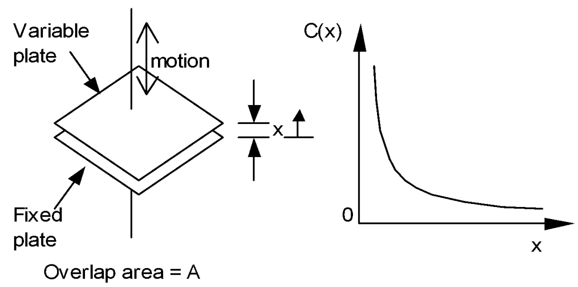


Fig. 3.9 Variable plate separation capacitance transducer construction and output response

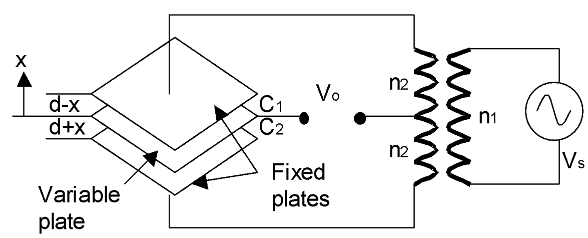
Transfer characteristic: Ignoring *fringe effects* that occur at the edges of the plates, the input/output relation is:

$$C(x) = \frac{\varepsilon_o \varepsilon_r A}{x}, \text{ for air } \varepsilon_r = 1,$$

hence

$$= \frac{\varepsilon_o A}{x} \quad > (3.19)$$

C(x) is inversely proportional to the displacement x. The transfer characteristic is non-linear (hyperbolic) therefore limiting the use of this transducer to measurements of small incremental displacements without making contact with the measurand. The output can be *linearised* by use of the *differential* (or balanced) capacitor configuration Fig. 3.10.



$$C_1 = \frac{\varepsilon_0 A}{d+x} \; ; C_2 = \frac{\varepsilon_0 A}{d-x}$$

Fig. 3.10 Balanced Variable plate separation capacitance transducer

The transfer equation of the transformer bridge is:

$$V_o = V_S \frac{n_2}{n_1} \frac{C_1 - C_2}{C_1 + C_2} \quad > (3.20)$$

Substituting for C_1 and C_2 in (5) yields:

$$V_o = V_S \frac{n_2}{n_1} \frac{x}{d}$$

$$= S_b x, \quad > (3.21)$$

Here S_b = bridge sensitivity (dV_o/dx)

$$= \frac{V_s}{d}\frac{n_2}{n_1} \qquad > (3.22)$$

where n_1 and n_2 are the number of turns in the primary and secondary transformer windings, respectively and d is the equal spacing between the plates at x=0 ($\Rightarrow C_1 = C_2$ and V_o =0).

3.4.3.2 ***Variable electrode-overlap capacitance transducers***

In this case the object whose displacement is to be measured is coupled to the capacitor in such a way that the overlapping surface area of the electrodes varies causing a directly proportional change in capacitance (Fig 3.11).

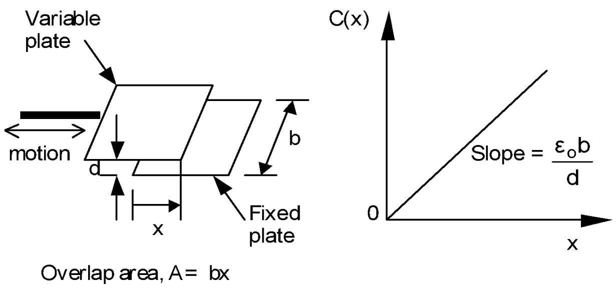


Fig. 3.11 Variable electrode-overlap capacitance transducer construction and output response

Transfer characteristic: Ignoring again the *fringe effects* that occur at the edges of the plates, the input/output transfer for this transducer is:

$C(x) = \frac{\varepsilon_o \varepsilon_r A}{d}$, for air ε_r =1and A =bx, hence

$$= \frac{\varepsilon_o b x}{d} \qquad > (3.23)$$

C(x) is directly proportional to the displacement x. The transfer characteristic is linear.

3.4.3.3 ***Variable dielectric capacitance transducers***

This device measures displacement of a dielectric whose *level* between the overlapping electrodes varies causing a directly proportional change in capacitance (Fig 3.12).

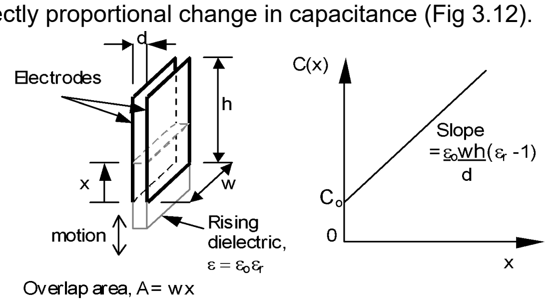


Fig 3.12 Variable dielectric capacitance transducer construction and output response

Transfer characteristic: Ignoring the *fringe effects*, the input/output transfer characteristic for the transducer (Fig 3.12) is derived as follows:

At any level x, the total capacitance:

$$C(x) = C_a + C_d \qquad > (3.24)$$

With C_a = capacitance of section with air = $\frac{\varepsilon_o w(h-x)}{d}$,

and, C_d = capacitance of rising dielectric = $\frac{\varepsilon_o \varepsilon_r w x}{d}$

Substituting C_a and C_d in (9) and simplifying gives:

$$C(x) = C_o\left[1 + \frac{x}{h}(\varepsilon_r - 1)\right] \qquad > (3.25)$$

Where $C_o = C(0) = \frac{\varepsilon_o wh}{d}$, the zero level capacitance.

The transfer characteristic is also linear in x. Typical applications incorporate the transducer in the form of two concentric cylinders measuring the level of fluid in a tank. In this case the fluid is the variable dielectric.

Example 3.3

A variable dielectric capacitive displacement sensor consists of two square metal plates, side 5mm, separated by a gap of 1mm. A sheet of dielectric 1mm thick and the same area as the plates, and having a dielectric constant of 4, can be slid between the plates as shown in Fig. 3.9. Calculate the capacitance of the sensor at x = 0.0, 2.5, 5.0 cm.

SOLUTION

Given data: w = h = 5 cm = 0.05m,

d = 1mm = 0.001m

ε_r =4 and for air $\varepsilon_r = \varepsilon_o$ =8.854 x 10^{-12} F/m

In general,

EMBED Equation.3

Hence: At x =0, EMBED Equation.3

= EMBED Equation.3

= 22.14 x 10^{-12} or 22.14 pF

At x =2.5cm, x/h = ½

EMBED Equation.3 [pF]

= 55.35 pF

At x = 5 cm, x/h =1

EMBED Equation.3 [pF]

=88.56 pF

Exercise 3.2

Show that the capacitance of two coaxial cylinders, radii b and a (b>a) and length (h), used in sensing the level (x) of a dielectric liquid is given by

$$C(x) = \frac{2\pi\varepsilon_o h}{\log_e\left(\frac{b}{a}\right)}\left[1 + \frac{x}{h}(\varepsilon_r - 1)\right] \qquad > (3.26)$$

COULOMB FORCE, F(x)

This is the *force that the measurement object must exert on a capacitive transducer in order to displace the electrodes*.

Taking F(x) to be positive for increasing values of x, performing an energy balance:

$$dE_f = dE_m + dE_e \qquad > (3.27)$$

where

dE_m = mechanical energy supplied by the object to cause a small displacement dx.

=F(x)dx

dE_e =electric energy provided by the power supply

= d(QV) =QdV +VdQ

dE_f =change in electric field energy

=½d(QV) =½ dE_e

Since the supply voltage across the capacitor electrodes is constant it follows that dV =0, hence taking Q = VC(x) and substituting appropriately gives:

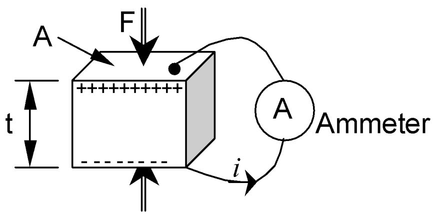
$$F(x) = \frac{S_c}{2} V^2 \quad \text{[newtons]} \qquad > (3.28)$$

where S_c = capacitor's displacement sensitivity

$$= \frac{dC(x)}{dx} \qquad > (3.29)$$

3.4.4 ***Piezoelectric transducers***

These transducers operate on the principle of the piezo-electric effect (Gr. Piezein = press), in which a *body of material exhibits electrical polarization (i.e. it develops a potential difference between it two opposite faces) when subjected to an external force.*



F= Applied force; A = cross section area; t =crystal thickness

Fig 3.13. Piezoelectric transducer

The phenomenon is characteristic of materials that have an asymmetric crystal lattice e.g. quartz. It may also occur in symmetric materials (e.g. ferroelectrics like barium titanate) by spontaneous (or random) polarisation if the centre of symmetry is disturbed by the presence of a strong magnetic field.

From solid mechanics, when a force F is applied to a crystal, the crystal atoms are displaced slightly from their normal positions in the lattice resulting in a displacement x in the direction of the force.

The ***steady state (or static) transfer*** between F and x is determined as:

$F = \sigma A = \varepsilon E A$

Where σ = stress; $\varepsilon = x / t$; E = Young's modulus (also called the modulus of elasticity, N/m^2)

Hence,

$$F = \frac{EA}{t} x = kx \qquad > (3.30)$$

with k = stiffness of the crystal

$$= \frac{EA}{t} \text{ [N/m]; typically 2GN/m} \qquad > (3.31)$$

The ***dynamic transfer*** between the input F and output x can be represented by the second order transfer function

$$G_1(s) = \frac{x}{f}(s) = \frac{1/k}{\frac{1}{\omega_n^2} s^2 + \frac{2\xi}{\omega_n} s + 1} \qquad > (3.32)$$

where $\omega_n = 2\pi f_n$. Typically f_n =10 to 100kHz and ξ=0.01

Charge Sensitivity, S_q

In a piezoelectric crystal deformation of the lattice is proportional to the charge developed, i.e.

q = Kx, where K [C/m] is a constant of proportionality.

But x =F/k, hence

$$q = \frac{K}{k} F = S_q F \qquad > (3.33)$$

Where S_q is the charge sensitivity (=K/k coulombs per metre[C/m] or coulombs per Newton [C/N]), i.e. the ratio of the amplitude of charge developed in the crystal to the amplitude of the applied force.

Voltage Sensitivity, S_v

The piezoelectric effect is reversible; eq. (3.33) represents the direct effect where an applied force produces an electric charge. In the inverse case, a voltage V applied across the crystal results in a proportional displacement x.

Here:

x =SV =F/k

or

$$V = \frac{F}{Sk} = S_V F \qquad > (3.34)$$

Where S_V = voltage sensitivity to the applied force [V/N], i.e. the ratio of the amplitude of the voltage applied to the amplitude of the resulting displacement / force and S_x is a constant of proportionality [m/V].

Crystal capacitance (C_{cr})

In order to measure the charge q, metal electrodes are deposited on opposite faces of the crystal, which are perpendicular to the displacement (Fig 3.13), to give a capacitor.

The capacitance of the parallel plate capacitor formed from the rectangular block of crystal of thickness t is called the crystal capacitance (C_{cr}) and is given as:

$$C_{cr} = \frac{\varepsilon A}{t} \qquad > (3.35)$$

where ε = dielectric constant [F/m]

$= \varepsilon_0 \varepsilon_r$

with ε_0 =permitivity of free space ≈ 8.854 x 10^{-12}F/m

ε_r =relative permitivity of the crystal material.

In practice, $q = C_{cr} V \qquad > (3.36)$

where V = the potential difference between the opposite crystal faces as a result of the applied force F.

Crystal resistance (R_{cr})

The resistance to flow of current across the terminals of the crystal is given as:

$$R_{cr} = \rho \frac{t}{A} \qquad > (3.37)$$

where ρ = resistivity of the crystal material, (Ω-m)

Relationship between S_q and S_v

From equations (15), (16) and (18) the relationship between the charge and voltage sensitivities is derived as:

$$q = S_q F = C_{cr} V$$

$$\Rightarrow \quad S_q = C_{cr}\frac{V}{F} = C_{cr} S_v$$

> (3.38)

A summary of the relevant properties of piezoelectric materials in common use is shown in Table 3.1

Table 3.1 Properties of commonly used piezoelectric materials.

	Material	Charge sensitivity Sq [pCN^{-1}]	Dielectric constant ε	Young's modulus E [GN/m^2]
Natural	Quartz	2.3	4.5	80
	Tourmaline	1.9, 2.4*	6.6	160
Ceramic	Lead Zirconate -titanate	265	1500	79
	Lead metaniobate	80	250	47

***Depends on direction of applied force**

3.5 ***ELASTIC SENSING ELEMENTS***

These devices convert an input force into a proportional output displacement by extension of the elastic-sensing element. They are commonly used to measure torque, pressure and acceleration related to the force by the following expressions:

Torque = force x distance [Nm]

> (3.39)

$$\text{Pressure} = \frac{\text{force}}{\text{area}} \ [\text{N/m}^2]$$

> (3.40)

$$\text{Acceleration} = \frac{\text{force}}{\text{mass}} \ [\text{m/s}^2]$$

> (3.41)

The output of these devices is converted to an electric signal by a secondary displacement sensor operating on the variable resistance/reluctance/capacitance or piezoelectric effects. The displacement may be translational or rotational.

<u>Transfer characteristics:</u> Elastic elements have associated mass (inertance) and damping (resistance) as well as spring characteristics. The general transfer equation for these elements is of the form:

$$G(s) = \frac{y(t)}{x(t)} = \frac{K}{\frac{1}{\omega_n^2}s^2 + \frac{2\xi}{\omega_n}s + 1}$$

> (3.42)

where K = steady state (i.e. static) sensitivity.

3.5.1 *Mass spring (linear) accelerometer*

The conceptual model consists of a casing containing a seismic mass which moves on frictionless guide rails against a spring.

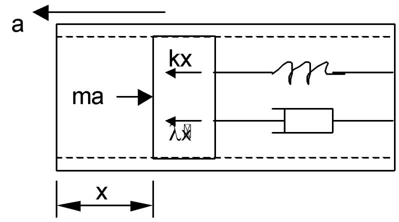


Fig 3.14 mass –spring linear accelerometer

Equation of motion: $m\ddot{x} + \lambda\dot{x} + kx = ma$

Transfer function:

$$G(s) = \frac{x}{a}(s) = \frac{\frac{1}{\omega_n^2}}{\frac{1}{\omega_n^2}s^2 + \frac{2\xi}{\omega_n}s + 1}$$

where $\omega_n = \sqrt{\frac{k}{m}}$, $\xi = \frac{\lambda}{2\sqrt{km}}$, K $=1/\omega_n^2$

3.5.2 *Bellows pressure sensor*

Consider the bellows construction as shown in Fig 7.15.

Equation of motion: $m\ddot{x} + \lambda\dot{x} + kx = AP$

Transfer function: $G(s) = \frac{x}{P}(s) = \frac{K}{\frac{1}{\omega_n^2}s^2 + \frac{2\xi}{\omega_n}s + 1}$

where $\omega_n = \sqrt{\frac{k}{m}}$, $\xi = \frac{\lambda}{2\sqrt{km}}$, K =A/k

3.5.3 *Angular accelerometer*

The construction of this device is as shown below:

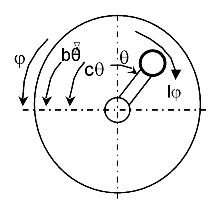


Fig 7.16 Rotating disc

Equation of motion:

$$I\ddot{\theta} + b\dot{\theta} + c\theta = I\varphi$$

Transfer function: $G(s) = \frac{\theta}{\varphi}(s) = \frac{\frac{1}{\omega_n^2}}{\frac{1}{\omega_n^2}s^2 + \frac{2\xi}{\omega_n}s + 1}$

where $\omega_n = \sqrt{\frac{c}{I}}$, $\xi = \frac{b}{2\sqrt{cI}}$, K $= 1/\omega_n^2$. For the torque sensor T =I ϕ.

3.6 ***DISPLACEMENT TRANSDUCER APPLICATIONS***

3.6.1 ***Measurement of force***

Force may be defined as a cause that produces resistance or obstruction to any moving body, or changes the motion of a body, or tends to produce these effects. Force is given as:

$$F = ma$$

where, F = force [N], m = mass, in kg and a - acceleration, in m/s^2

There are various methods of force measurement. But, there are several limitations to the methods of measuring force which are listed below:

- ✓ Force must either be reasonably constant in value or changing gradually and continuously in one direction. Otherwise the fluctuating forces of the masses of moving parts may result in a very low frequency and consequent lag in accurate indication,
- ✓ The force must act perpendicular to the platform of the scale, otherwise only the cosine component is measured
- ✓ The measurement may require correction for local variation in gravitational constant. Some of the methods of force measurement are described below

3.6.1.1 Cantilever load cell

In the cantilever force element (load cell) the applied force F causes the cantilever to bend stretching the top surface (tensile strain, $+e$) and compressing the bottom surface (compressive strain, $-e$). The force F is measured by means of four bonded strain gauges which are connected as shown in Fig 3.17 to enhance the signal output from the Wheatstone bridge and compensate for temperature variations in the cantilever beam.

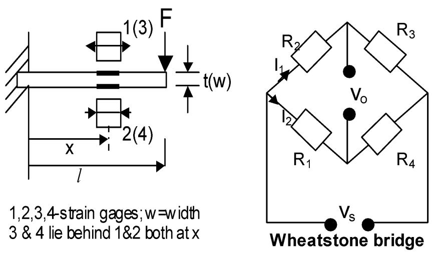


Fig 3.17 Cantilever load cell

The magnitude of strain *e* is given by:

$$e = \frac{6F(l - x)}{wt^2E} \qquad > (3.45)$$

Also for strain gauges, $\Delta R = GR_o e$

where G is the ***gauge sensitivity*** obtained as:

$$G = 1 + 2\nu + \frac{1}{e}\frac{\Delta\rho}{\rho}$$

For most metals v=0.3 and G ≈2.0

R_o is the unstrained resistance of the gauge before loading. Hence:

$$R_1 = R_3 = R_o + \Delta R = R_o(1 + Ge)$$

$$R_2 = R_4 = R_o - \Delta R = R_o(1 - Ge) \qquad > (3.46)$$

For the purely resistive Wheatstone bridge

$$V_o = V_s\left\{\frac{R_1}{R_1 + R_4} - \frac{R_2}{R_2 + R_3}\right\} \qquad > (3.47)$$

From eq. (3.46)

$$R_1 + R_4 = R_2 + R_3 = 2R_o$$

and also

$$R_1 - R_2 = 2R_oGe$$

Therefore

$$V_o = V_s\left\{\frac{R_1 - R_2}{R_1 + R_4}\right\} = V_sGe$$

$$= \frac{6V_sG(l - x)}{wt^2E}F \qquad > (3.48)$$

For example given the parameters for the cantilever beam and bridge circuit as being: G = 2.0, E = 200 GPa, V_s = 12 V, w = t = 0.005 m, l = 0.05 m, x = 0.01 m and F = 5 N then from eq. (3.45) e = 48 μm, V_o = 1.152 mV.

3.6.1.2 Pillar load cell

Here the applied force F causes a compressive stress –F/A, where A is the cross sectional area of the pillar this produces a longitudinal, compressive strain (e_L) which is accompanied by a transverse tensile strain (e_T). The strain gauge arrangement for temperature compensation is as shown in Fig 3.18 below:

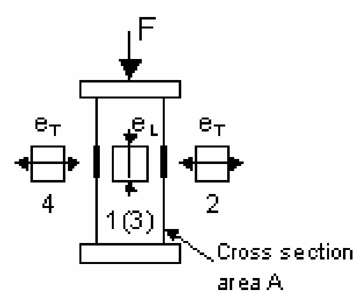


Fig 3.18 Cantilever load cell

The longitudinal and transverse strains are:

$$e_L = -\frac{F}{AE}; \qquad e_T = -\nu e_L = \frac{\nu F}{AE}$$

where ν and E are the Poisson ratio and Young's modulus respectively of the pillar material.

Here

$$R_1 = R_3 = R_o + \Delta R_o$$
$$= R_o(1 + Ge_L)$$

$$R_2 = R_4 = R_o + \Delta R_o$$
$$= R_o(1 + Ge_T) = R_o(1 - \nu Ge_L) \qquad > (3.49)$$

Observe again that

$$R_1 + R_4 = R_2 + R_3 = R_o(2 + Ge_L(1 - \nu)) \approx 2R_o$$

since $Ge_L(1 - \nu) << 2$

and also

$$R_1 - R_2 = Ge_LR_o(1 + \nu)$$

Therefore

$$V_o = V_s\left\{\frac{R_1 - R_2}{R_1 + R_4}\right\} = V_s G e_L (1+\nu)$$

$$= -\frac{V_s GF}{AE}$$

> (3.50)

The negative sign is a result of compression for which

$$F = -\frac{V_o AE}{V_s G}$$

> (3.51)

Example 3.4

A load cell is formed of a hollow steel cylinder loaded axially. The four strain gauges are so connected to enhance the signal and compensate for temperature variation. The load cell has a cross-sectional area of 20mm². Young's Modulus of steel is 207GPa and Poisson's ratio of 0.3. The unstrained gauge resistance is 1 kΩ and the gauge factor is 2.1. If the current in each strain gauge is limited to 25 mA, find the bridge supply voltage and the corresponding bridge voltage output when the load cell is subjected to a compressive force of 10^5 N.

SOLUTION

Given data:

$A = 20\text{ mm}^2 = 20 \times 10^{-6}\text{ m}^2$

$E = 207\text{ GPa} = 207 \times 10^9\text{ N/m}^2$

$\nu = 0.3$

$G = 2.1$, $R = 1k = 1000$

$F = -1 \times 10^5$ N (-ve sign denotes compression)

$I_1 = I_3 = 25$ mA

The supply voltage, Vs, for the bridge configuration Fig. 7.17 is given by:

EMBED Equation.3

EMBED Equation.3

But EMBED Equation.3

From eq. (7.50) the bridge output is:

EMBED Equation.3

3.6.2 *Measurement of Pressure*

Pressure is described as force per unit area and is analogous to stress. It is measured either as force per unit area or a head of a liquid column. It is, therefore, expressed as N/m², or in mm of water or mercury (1 Pascal, Pa = 1 N/m²). Some of the definitions of different pressure terms are illustrated in Fig. 3.19.

Absolute Pressure*:* Absolute pressure refers to the absolute value of force per unit area exerted by a fluid on the walls of its container. It is the fluid pressure measured above a perfect vacuum.

Atmospheric Pressure*:* The pressure exerted by the earth's atmosphere, as commonly measured by a barometer. At sea level, its value is close to 1.013 x 10^5 N/m² absolute, decreasing with altitude.

Gauge Pressure*:* Gauge pressure indicates the difference between the absolute pressure and local atmospheric pressure.

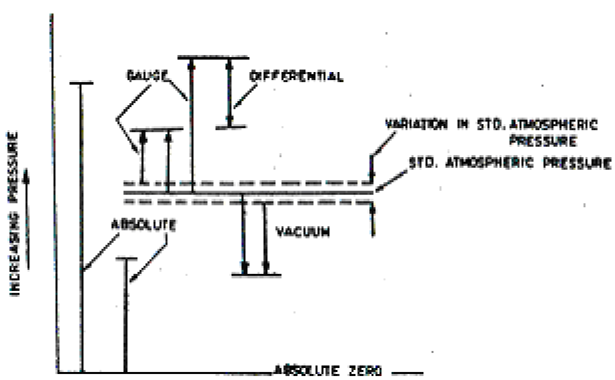


Fig 3.19 Different pressure terms

Differential Pressure*:* This is the difference between two measured pressures.

Vacuum: Vacuum represents the amount by which the atmosphere pressure exceeds the absolute pressure.

The pressure at a point in a fluid is equivalent to the weight of a certain column of the fluid acting on unit area at that point, i.e. $p = \rho gh$. This *h* is termed as the head.

3.6.2.1 The capacitive pressure sensor

The capacitive pressure sensor consists of a fixed metal disc that overlaps a flexible flat circular diaphragm.

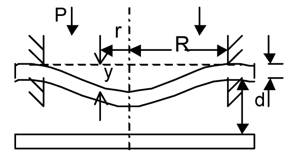


Fig 3.20 Capacitive pressure sensor

The elastic diaphragm is clamped around its circumference so that it bends as shown on application of a force P, effectively varying the electrode separation.

From Solid mechanics, the deflection y at any radius r can be shown to be:

$$y = \frac{3P}{16}\frac{(1-\nu^2)}{Et^3}(R^2 - r^2)^2$$

> (3.52)

where R = radius of diaphragm; t = thickness of diaphragm; E = Young's modulus and ν = Poisson's ratio.

A reduction in the separation of the electrodes results in an increase in capacitance given by: (Bentley, J. P, 1988)

$$\Delta C = C_0 \frac{(1-\nu^2)R^4}{16Edt^3}\Delta P$$

> (3.53)

with C_o = capacitance at zero pressure

$= \varepsilon_o \pi R^2/d$

d = initial separation of the electrodes

Example 3.5

A special purpose strain gauge circuit, consisting of a Wheatstone bridge, is bonded to a circular diaphragm of radius R_o and thickness t_d to create a pressure transducer. The device is used to measure pressure transients in a turbine where the pressure variation is $p(t) = 2500 + 500\sin 1000t$ kPa. The bridge output (V_o) is given by $V_o = 0.82V_s \frac{pR_o^2(1-\nu^2)}{t_d^2 E}$, where *p* is the pressure, *E* is the modulus of elasticity of the diaphragm material, *v* is the Poisson's ratio and V_s is the bridge supply voltage. The diaphragm is 30 mm in diameter and is fabricated from a stainless steel alloy with $v \approx 0.3$. For a linear output over the entire working range, the diaphragm thickness is chosen to be four times the maximum deflection at its centre, for which the bridge voltage ratio, V_o/V_s, must not exceed 0.3%. Evaluate: -

(i) the thickness of the diaphragm (t_d) and its modulus of elasticity (*E*)

(ii) the transducer's full scale voltage output, given that the bridge power supply is maintained at 230 V

SOLUTION

Given data: $R_o = 0.03/2 = 0.015$ m, $= 0.3$

(i) At $p = p_{max}$; $t_d = 4\,y_{max}$ or $y_{max} = 0.25\,t_d$

Hence EMBED Equation.3

EMBED Equation.3 > (1)

Also EMBED Equation.3

EMBED Equation.3 > (2)

Example 3.6

A compound transducer consisting of a bellows, coupled to a differential push-pull displacement sensor consisting of an armature between two identical semitoroid cores separated by a fixed distance *2d* and incorporated into an ac. deflection bridge, is used to measure both the density and the temperature of a hydraulic fluid in a reserve tank (see Fig 3.21).

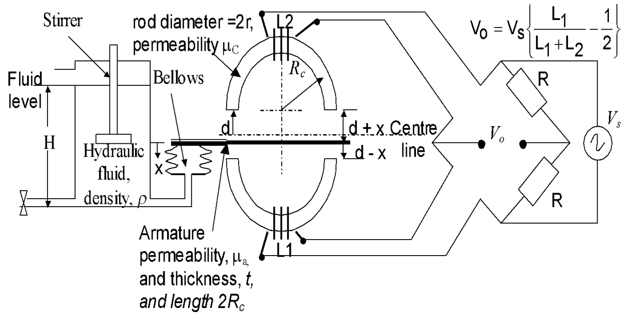


Fig 3.21 Density and temperature measurement

(a) Show that the transducer's bridge transfer can be expressed as $V_o(x) = \frac{V_s \alpha}{2(1+\alpha d)} x$ where $\alpha = \frac{2}{\pi r^2 \mu_o \Re_o}$.

(b) The density of the fluid varies with its temperature such that $\rho\,[kg/m^3] = 857 - 0.61114 \times T[^oC]$. The bellows has an effective area of 14 mm^2 with a spring rate of 14 N/m. The fluid level in the tank is maintained at 1m and the transducer is calibrated such that bridge output is zero volts when the hydraulic fluid is at 0 ºC. The number of turns on each coil is 20, the radius of curvature of the core is 2 cm and its diameter is 0.5 cm. The armature has a square section of sides 0.5 cm. Given that the bridge power supply is 230 V, find an expression for $V_o(T)$ and hence compute the fluid density and temperature when the transducer output is 25 V. (Take μ_o= 1.257 μH/m , d =1mm and $\mu_c = \mu_a$ =100)

SOLUTION

(a) The bridge transfer is, -

EMBED Equation.3

For the armature position fig 7.21

EMBED Equation.3

EMBED Equation.3

EMBED Equation.3

Therefore

EMBED Equation.3

(b) The bellows displacement *x* is given by: -

EMBED Equation.3

Where A is the effective area of the bellows and ρ_o is the threshold density of the hydraulic fluid at T = 0ºC = 857 kg/m^3

Given that **EMBED Equation.3** , then

EMBED Equation.3

Substituting for x in the transducer's transfer gives

EMBED Equation.3

Given data:

N = 20 turns, R_c = 0.02 m, 2r = t = 0.005 m, μ_o = 1.257 μH/m, $\mu_c = \mu_a$ = 100, d = 0.001m; A = 0.000014 m^2, H = 1 m; K = 14000 N/m and V_s = 230 V.

EMBED Equation.3 **EMBED Equation.3** EMBED Equation.3

EMBED Equation.3 **EMBED Equation.3** EMBED Equation.3

Given $V_o(T)$ = 25 V T = 25/(0.4707) 53 ºC

And hence

EMBED Equation.3

SOLUTION
(a) The bridge transfer is, -
For the armature position fig 7.21
EMBED Equation.3
Therefore
(b) The bellows displacement x is given by: -
Where A is the effective area of the bellows and ρo is the threshold density of the hydraulic fluid at T = 0ºC = 857 kg/m3
Given that EMBED Equation.3 , then
Substituting for x in the transducer's transfer gives
Given data:
N = 20 turns, Rc = 0.02 m, 2r = t = 0.005 m, μo = 1.257 μH/m, μc = μa = 100, d = 0.001m; A = 0.000014 m2, H = 1 m; K = 14000 N/m and Vs = 230 V.
Given Vo(T) = 25 V T = 25/(0.4707) 53 ºC
And hence
EMBED Equation.3

3.6.3 Measurement of torque and power

Torque is a turning moment and can be defined as the force which tends to change the linear motion or rotation of a body. It is often associated with the determination of mechanical power.

Consider a power transmission shaft of radius r that is rotating at a speed ω as shown in Fig. 3.22.

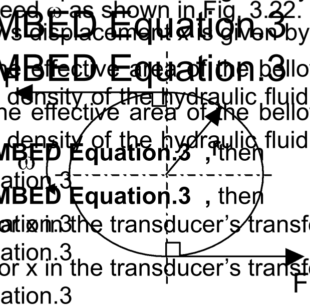


Fig 3.22 Rotating shaft

Torque (T) is given as the product of the force couple (F in newtons) and the perpendicular distance (D in metres), i.e.

T = F.D = 2r.F [N.m] > (3.54)

The power (P) is the product of the torque and the speed of angular rotation ω (in radians per second), i.e.

$$P = T\,\omega = 2\omega r.F \;[N.m.s^{-1} \text{ or Watts}] \qquad > (3.55)$$

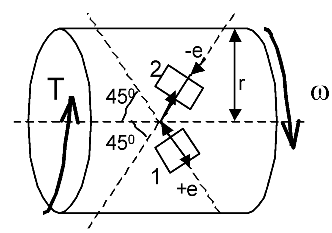
With $\omega\,[rad/s] = \frac{2\pi N}{60}$ where *N* is the speed of rotation in revolutions per minute (RPM).

Devices used to measure power are usually referred to as dynamometers.

3.6.3.1 Strain gauge torque sensor

This device consists of a cylindrical shaft that acts as the primary torque sensing element (Fig 3.23). The applied torque T produces a shear strain φ in the shaft and corresponding linear tensile and compressive strains on the shaft surface.

Gauge 1 is mounted with its active axis at +45º to the shaft axis where the tensile strain has a maximum value +e, and gauge 2 at -45º with a maximum compressive strain –e. Gauges 3 and 4 are mounted in similar angles on the other side of the shaft, and experience strain +e, -e respectively.



3.23 Strain gauge torque sensor

The maximum strain e is:

$$e = \frac{T}{\pi S r^3} \qquad > (3.56)$$

where S is the shear modulus of the shaft material and *r* is the shaft radius. Again for the wheatstone bridge arrangement as shown in Fig. 3.17

$$R_1 = R_3 = R_o + \Delta R = R_o(1 + Ge)$$
$$R_2 = R_4 = R_o - \Delta R = R_o(1 - Ge) \qquad > (3.57)$$

and

$$V_o = V_s\left\{\frac{R_1 - R_2}{R_1 + R_4}\right\} = V_s Ge$$
$$= \frac{V_s GT}{\pi S r^3} \qquad > (3.58)$$

3.6.4 *Measurement of vibration and shock*

Vibration and shock are two main physical characteristics that define motion in structures and machine components. Shock (or transient vibration) is defined as the transmission of kinetic energy to a system which takes place over a relatively short time compared to the natural period of the system and which is followed by a decay of the natural oscillatory motion developed in the system. Shock phenomena arise from explosions, impacts and other spasmodic releases of energy. Vibration, on the other hand, arises from imbalanced rotating parts, misalignments or external forces.

3.6.4.1 Relationship between the vibration parameters

Assuming that the vibration is simple harmonic motion, then the relationship between the different vibration parameters is:

Displacement, $x = A \sin \omega t$ > (3.59)

Velocity, $v = A\omega \cos \omega t$ > (3.60)

Acceleration, $a = -A\omega^2 \sin wt$ > (3.61)

where $\omega = 2\pi f$ [rad/s]

f = frequency of vibration [Hz]

Note that the frequencies are the same in each case, although there is a phase shift. The amplitudes of the above parameters are thus

Displacement amplitude = A

Velocity amplitude = $A\omega$

Acceleration amplitude = $A\omega^2$

In monitoring of machine vibration at least one of these parameters is measured, particularly at resonant frequency when the amplitudes are largest.

3.6.4.2 Piezoelectric accelerometer sensors

Piezoelectric elements are commonly used in measurements of acceleration and vibration. Tension cannot be applied to a crystal without using some kind of adhesive to make the tensile connection, and such a connection would be unreliable, so in the simplest form of piezoelectric accelerometer the crystal is kept permanently compressed by the seismic mass. Thus the effect of accelerations in alternate directions is an alternating increase and decrease in the compressive force on the crystal. The compressive pre-load is applied by screwing the seismic mass down on to the crystal to a given torque.

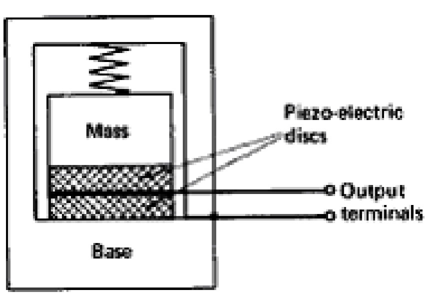


Fig. 3.24 Piezoelectric accelerometer (compression mode)

The electrical connections to the crystal are made by *metallising* (depositing a thin film of metal on) the end faces. This type of construction gives an accelerometer which is rugged, but because the casing is part of the 'spring' in the spring-mass system, it may be subject to spurious inputs. These include temperature change (causing expansion or contraction of the casing), acoustic noise, base bending (distorting the casing), cross-axis motion, and magnetic fields.

3.6.4.3 *Accelerometer charge sensitivity*

Consider a piezoelectric crystal used in conjunction with a seismic (vibrating) mass. For an accelerating mass,

$F = ma$

where F = actuating/driving force [N]

m = the seismic mass [kg]

a = the acceleration [ms^{-2}]

$= m\frac{d^2x}{dt}$, x =displacement [m]

Recall from eq. (3.33), the crystal acquires a charge proportional to the force, i.e. $q = S_q F = S_q ma$

$$\Rightarrow \left|\frac{q}{a}(t)\right| = S_q m \quad [\text{coulomb.m}^{-1}\text{s}^2]$$

$$= 9.81\, S'_q m \quad [\text{coulomb. g}^{-1}] \quad > (3.62)$$

where S'_q is the charge sensitivity expressed in coulomb. g^{-1} (S_q/g with g = 9.81m/s^2).

3.6.4.4 *Accelerometer voltage sensitivity*

If the accelerometer is used with the charge amplifier configuration as earlier discussed, the steady state voltage sensitivity can be expressed as:

$$\left|\frac{V_o}{a}(t)\right| = \frac{9.81 S'_q m}{C_F} \quad [\text{Volts g}^{-1}] \quad > (3.63)$$

The dynamic frequency response characteristics of piezoelectric accelerometers with or without charge amplifiers are as already described.

Displacement transducers may be used pick up the displacement which can be differentiated once to give velocity or twice to give acceleration. Piezoelectric transducers that pick up acceleration are however preferred (See Table 3.2).

Table 3.2 Piezoelectric versus displacement transducers in measurement of vibration

	Displacement transducer	Piezoelectric transducer
Sensitivity	Insufficient	Sufficient
Frequency response	Inadequate or mass too large for this application $\omega_n=\sqrt{(k/m)}$	Adequate frequency response and low mass
Noise immunity	Differentiation operations, to determine velocity and	Integration operations, to give velocity and displacement

	acceleration information, amplify unwanted high frequency noise	information, tends to attenuate (i.e. scale down) high frequency noise

3.7 TEMPERATURE MEASUREMENT

The temperature transducers which shall be discussed in some detail sense changes in temperature as:

- Changes in pressure (used in sealed-fluid thermometers, vapour-pressure thermometers).
- The change in contact potential between different metals (used in thermocouples).
- The change in resistance of a metal or a semiconductor (used in metallic resistance thermometers and thermistors).
- The change in the energy radiated by a hot object (used in radiation pyrometers).

Because all the types of device listed above produce an output signal, they can be used in environments where it is essential that the temperature indication be remote from the temperature transducer itself, for example in measuring the temperature of hot gases in a chemical process. The last-mentioned type of transducer - a radiation pyrometer - has the particular advantage of not needing to be in contact with the hot object whose temperature it is measuring; it uses the radiation given off by the hot object to measure temperature.

Table 3.3: The International Practical Temperature Scale (IPTS)

FIXED POINT	ASSIGNED TEMPERATURE	STANDARD INSTRUMENT
	> 1337.58 K	optical pyrometer
Freezing point of gold	1064.43 °C	Thermocouple (903.89 - 1337.58 K)
Freezing point of silver	961.93 °C	
Freezing point of zinc	419.58 °C	
Boiling point of water	100 °C	
Triple point of water	0.01 °C	
Boiling point of oxygen	-182.962 °C	Platinum resistance thermometer (13.81 - 903.89 K)
Triple point of oxygen	-218.789 °C	
Boiling point of neon 27	-246.048 °C	
Boiling point of equilibrium hydrogen	- 252.87 °C	
Equilibrium between the liquid and vapour phases of equilibrium hydrogen at 33 330.6 Pa	-256.108 °C	
Triple point of equilibrium hydrogen	-259.34 °C	

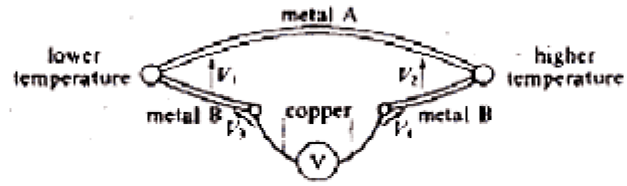
Not only are there several ways of measuring temperature, but there is more than one scale on which it is measured. Three scales are commonly used: Celsius (°C), Fahrenheit (°F) and kelvin (K).

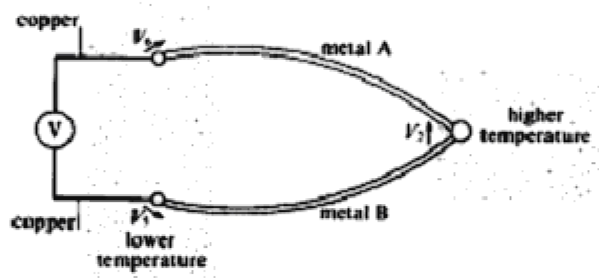
3.7.1 *Thermoelectric (thermocouple) temperature sensors*

Thermoelectric instruments are those that depend on a change in electrical characteristics of the sensing element to indicate changes in temperature.

3.7.1.1 *Operating principles of a thermocouple*

A thermocouple is a temperature transducer consisting of two wires of different metals joined at both ends. If the two junctions between the metals are at different temperatures an electric current will flow around the circuit. This phenomenon is called the ***Seebeck effect*** or sometimes the ***thermoelectric effect***.





Net voltage = V_2 - V_1 = V_5 - V_6

Fig 3.25 Thermoelectric effect

When a third metal is introduced into a thermocouple circuit, the net voltage is not affected, provided both junctions of` the third metal are at the same temperature.

This explains why a voltmeter can be inserted into the thermocouple circuit in either of the ways shown in Fig 3.25. It also tells us that thermocouple wires can be brazed or soldered, and that any new metals introduced in the process will not alter the net thermocouple voltage.

3.7.1.2 *Reference junctions and extension wires*

When a thermocouple is used to measure temperature, one of its two junctions, i.e. the hot junction, is placed in contact with the measurand so that it takes on the temperature of that material. However, the voltage across the output terminals of a thermocouple depends upon the temperature at both junctions. This means that the temperature at **the second junction** must be either known or controlled to at least the accuracy expected from the measurement. The second junction (i.e. the cold junction) is usually referred to as the ***reference junction***.

One strategy for establishing the temperature of the reference junction is to place it in some apparatus in which the temperature is accurately controlled. The most accurate method, used primarily in laboratory situations where great accuracy is required, is to use a triple-point apparatus, in which the temperature is maintained at the triple point of water. A more common method is to use an ice-water mixture to maintain a reference temperature of 0 °C, Such an ice bath can quite easily be constructed. There are also ice baths which are made commercially for use with thermocouples. Some of these contain built-in refrigeration units which eliminate the need to replenish the ice.

The reference temperature need not be 0°C. A temperature-controlled oven can be used to keep the temperature of the reference junction constant at any convenient temperature. One solution to the reference-junction problem is to use a ***cold-junction compensator***. Such a device monitors the temperature of the reference junction, which is allowed to take on the ambient temperature, and then automatically compensates for it in the thermocouple's voltage output. In fact, integrated circuits are now available which combine a *cold-junction compensator* and an instrumentation amplifier on one silicon chip. (Thermocouple outputs are fractions of a millivolt per ºC of temperature change, so in some applications an amplifier is essential.) When such a chip is used, the voltage signal from the thermocouple is supplied between two of the pins of the integrated circuit and the compensated, amplified voltage is taken between two other pins.

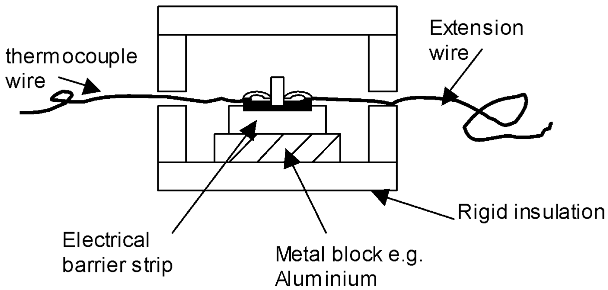


Fig 3.26 Construction of a zone box

Measuring instruments must often be located far from the point at which the measurements are to be made. With thermocouples this raises the problem of where to put the reference junction (which may well be located where the thermocouple wires join the wires leading to the meter). If it cannot be located near where the measurement is to be made, long lengths of thermocouple wire need to be used. Such wire is manufactured to very tight specifications and hence is expensive. The usual alternative is to use extension wires either made of the same material as the thermocouple wires, but of a smaller gauge and to less exacting standards, or made of some other metal which closely matches the temperature characteristics of the thermocouple wires. With extension wires connected, the reference junction is effectively moved to their ends. Connection of extension wires is done in a ***zone box*** which is a simple insulated box that maintains all the thermocouple junctions that it contains at a uniform temperature.

3.7.1.3 *Voltage to temperature conversion*

The thermocouple is a non-linear device i.e. the output temperature is not linearly proportional to the corresponding digital voltmeter reading at the terminals of the thermocouple. In general: -

$$\mathbf{T = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \ldots + a_n x^n} \quad > (3.64)$$

where

T = Temperature

x = Thermocouple EMF in Volts

a = Polynomial coefficients unique to each thermocouple

n = Maximum order of the polynomial

As n increases, the accuracy of the polynomial improves. A representative number is n = 9 for ± 1∘C accuracy. Lower order polynomials may be used over a narrow temperature range to obtain higher system speed.

The thermocouple types also differ greatly

(a) in their ability to withstand corrosive environments of different types and over different temperature range,

(b) in their ability to withstand thermal and mechanical stresses; and

(c) in their accuracy and reproducibility; and of course they also differ in price.

The particular combination of necessary attributes determines the choice in any given application.

Table 3.4 The thermocouple range

Type (Temp range, oC)	**1st wire material**	**2nd wire material**	**Environment of use**
S (0-1400)	Platinum	90%Platinum 10% Rhodium	Oxidising or inert.
R (0 - 1400)	Platinum	87%Platinum 13% Rhodium	Oxidising or inert.
J (-200-850)	Iron	Copper nickel	Reducing or inert
K (-200-1100)	Nickel chromium	Nickel aluminium	Oxidising or inert
T (-250-400)	Copper	Copper nickel	Most non corrosive
E (-200-850)	Nickel chromium	Copper nickel	Most non corrosive
B (0-1500)	70%Platinum 30% Rhodium	94%Platinum 6% Rhodium	Oxidising or inert

<u>*Example 3.4*</u>

A Copper-constantan thermocouple was found to have a near linear calibration between 0 - 400 ºC with an emf at maximum temperature (i.e. with reference junction temperature equal to 0 ºC) equal to 20.68 mV.

(a) Determine the correction to be made to the indicated emf if the cold junction temperature is 25 ºC.

(b) If the indicated emf in the thermocouple circuit is 8.92 mV, find the temperature of the hot junction.

<u>SOLUTION</u>

- Sensitivity of the thermocouple

= EMBED Equation.3 = 0.0517 mV/ºC

The thermocouple is calibrated for a reference junction temperature of 0 oC. When used at 25 oC: -

EMBED Equation.3

The correction to be made to the emf indicated

EMBED Equation.3

- The indicated emf = 8.92 mV. Therefore the corresponding □T from the I/O transfer curve is

EMBED Equation.3

Hence EMBED Equation.3 . T_{cold} = 25 ºC Hence T_{hot} = 172.53+25 = 194.53 ºC.

Alternatively

EMBED Equation.3

sheath also enables the thermocouple to be used in a system which must be sealed or must withstand high pressures.

Magnesium oxide insulation provides a high insulation resistance (of the order of MΩ) between the thermocouple wires and the metal sheath. At the same time it has a high thermal conductivity to enable the probe to respond rapidly to changes in the temperature to be measured. The response time of the thermocouple depends upon the rate of heat transfer from the material being measured and so varies not only from one thermocouple to another but also from one application to another.

Table 3.5 Thermocouple Sheath materials

Sheath material	Maximum operating Temp.	Affords protection against.
Metals		
Mild steel	800-900	Oxidation (but emits metal vapour above 500 oC)
Nickel-free stainless iron	800	Low-temperature molten metals (but emits metal vapour above 500 oC)
High chromium-nickel stainless steels	1000	Oxidation (but emits metal vapour above 500 oC)
Nickel-chromium alloy	1100	Oxidation (but emits metal vapour above 500 oC)
Refractories		
Fused silica	1000-1100	Oxidation and gas entry (but may emit silicon above 800 oC)
Porcelain	1400	Oxidation and gas entry
Recrystallised aluminium oxide	1700	Oxidation and gas entry

However, in general, the response time of a mineral-insulated probe increases as the probe diameter increases, so if a rapid response is a requirement for a particular application then a small probe diameter is essential. The best response times are obtained from bare wires.

3.7.2 ***Resistance temperature transducers***

Resistive temperature transducers can be divided into two major types: metals and semiconductors. Metallic resistance transducers are generally called ***resistance thermometers***. They include platinum, copper, tungsten, Balco (an alloy of nickel) and nickel.

The most commonly used semiconductor temperature transducers are made of oxides of chromium, manganese, iron, cobalt or nickel and are called ***thermistors***.

3.7.2.1 Resistance thermometers

The resistance of metallic resistance thermometers varies with temperature. The general relationship is approximately linear as shown in Fig 3.27.

Although the sensitivity of platinum is the smallest, it is the most stable and reproducible type of temperature transducer. It is thus used to define the *International Practical Temperature Scale* for temperatures in the range from 13.81K to 903.89K. Platinum is chemically inert and is resistant to contamination. This enables ***platinum resistance thermometers*** to be used in a wide range of environments. Unfortunately, platinum is relatively expensive.

The resistance-temperature relation is:

$$R_T = R_0(1 + \alpha T)$$

> (3.65)

where R_o = is the resistance at 0 °C [Ω]

α = is the temperature coefficient of resistance and T = is the temperature in °C.

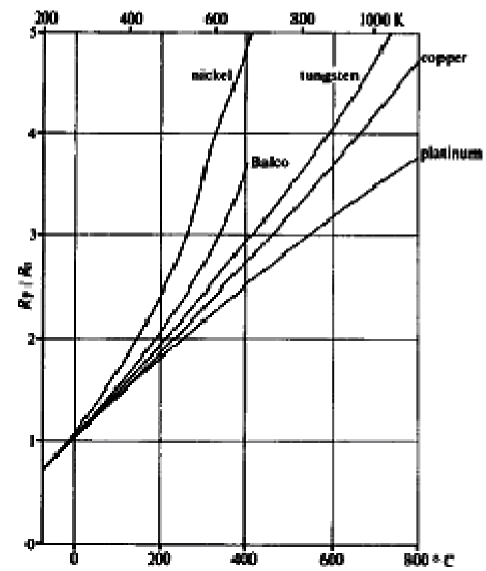


Fig. 3.27 Resistance vs. Temperature for materials for resistance thermometers

In Table 3.6 the values of α (to two significant figures) for some metals for the range from 0 °C to 100 °C are listed.

Table 3.6: Temperature coefficient of resistance for the range from 0 °C to 100 °C

Material	α / °C $^{-1}$
Copper	0.0043
Nickel	0.0068
Platinum	0.0039
Tungsten	0.0046

Typical specification: Comark Electronics Ltd, ***Platinum resistance immersion probe*** Model PRB4020/H2PC/DIN

Time constant 8 s in water

Temperature range -200 ^{0}C to 500 ^{0}C

Probe dimensions Length: 200 mm.
Diameter: 4mm.
Accuracy
To BS1904: 1984 Class B.
(Class A available on special order)
Connector Insulated PVC cable to standard DIN plug

Thermistors
Resistive temperature elements made from semiconductor materials are known as thermistors. Figure 3.28 shows a resistance against temperature curve for a thermistor. It is strikingly different from the curves for metallic resistance thermometers in two respects: it is much steeper and it decreases rather than increases with increasing temperature. Thermistors with a negative temperature coefficient of resistance like this are called ***negative-temperature-coefficient (NTC) thermistors***.

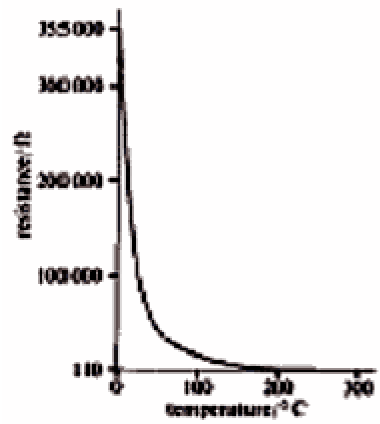


Fig 3.28 Resistance vs. temperature curve for a thermistor

This negative coefficient of resistance is much larger in magnitude than the coefficient of metals, at least over part of the range. The relationship between resistance and temperature θ for a thermistor is described by the equation:

$$R_\theta = A\exp\left(\frac{\beta}{\theta}\right) \qquad (3.66)$$

where R_θ is the resistance at temperature θ {K], A and β are constants for the thermistor. β is called the *characteristic temperature* of the thermistor. Its value is usually between 2000 K and 4000 K.

An alternative expression for Eq. (3.66) in terms of R_{ref} the resistance at reference temperature θ_{ref} is:

$$\frac{R_\theta}{R_{ref}} = \frac{A\exp\left(\frac{\beta}{\theta}\right)}{A\exp\left(\frac{\beta}{\theta_{ref}}\right)} = \exp\beta\left(\frac{1}{\theta} - \frac{1}{\theta_{ref}}\right) \qquad (3.67)$$

Hence

$$R_\theta = R_{ref}\exp\beta\left(\frac{1}{\theta} - \frac{1}{\theta_{ref}}\right) \qquad (3.68)$$

Note:
There are also thermistors with positive temperature coefficients of resistance. However, they are used primarily in applications such as overheating protection, rather than for temperature measurement.

3.7.3 Pyrometers

All the temperature-measuring methods discussed earlier require physical contact of thermometer with the body whose temperature is to be measured. At high temperatures (above 1400°C), the thermometer may melt due to direct physical contact. To solve these problems a non-contact method of temperature sensing is used. Also, for bodies that are moving, a non-contacting means of temperature sensing is most convenient.

Pyrometry is a technique for measuring temperature without physical contact. It depends upon the relationship between the temperature of a hot body and the electromagnetic radiation emitted by the body. When a body is heated, it emits thermal energy known as heat radiation. A black matt surface (or a black body) is a very good absorber of heat radiations and, also, a very good emitter of such radiations when heated. Pyrometry is a technique for determining a body's temperature by measuring its electromagnetic radiation. There are two types of pyrometers generally used in industries: (i) Radiation pyrometers (ii) Optical pyrometers

Radiation pyrometers measure the amount of radiant energy emitted by the hot body. Radiated heat from the hot body is focused onto the detector (e.g. the hot junction of a thermocouple) thus giving an indication of the temperature.

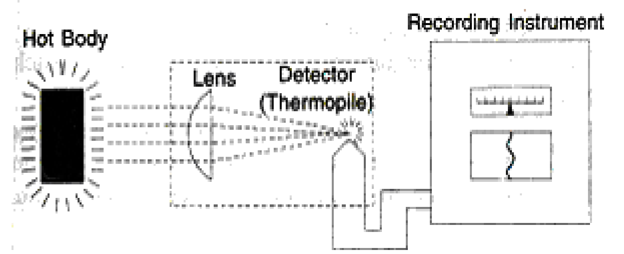
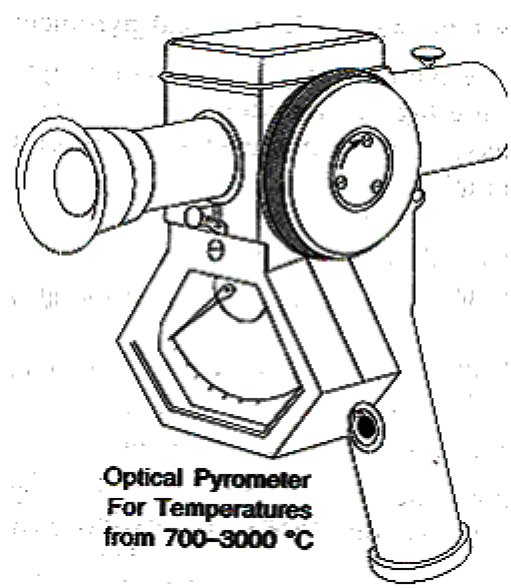


Fig 3.29 Radiation pyrometer



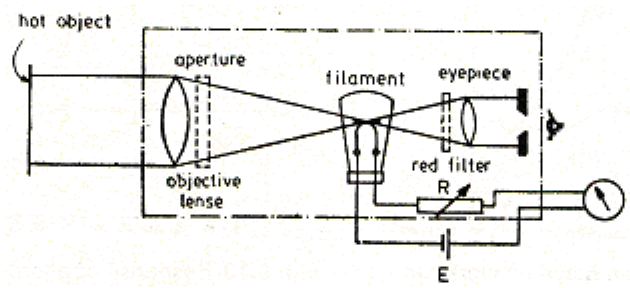


Fig 3.30 Schematic of the optical pyrometer

Radiation pyrometers have the following sources of errors: (i) They are quite sensitive to any obstructions in the line of sight between the pyrometer and the hot body. They should not be used when infrared-absorbing water vapour, dust, or other particles are in the air and (ii) they are also sensitive to emittance errors. To compensate for such errors, the device should be calibrated using an optical pyrometer which is less sensitive to emittance errors. Some of the advantages of radiation pyrometers are that they are able to measure high temperatures and need no contact with target of measurement. This makes it possible for them to indicate the temperature of moving targets. They also possess a fast response and have high output and moderate cost.

The disadvantages of radiation pyrometers include: a non-linear scale; errors due to presence of intervening gases or vapours that absorb radiating frequencies and the emissivity of the target material affects measurement.

Optical pyrometers make use of changes in colour of a hot body and interpret this phenomenon in terms of temperature. When a body is heated it initially becomes dark red, turns orange and then finally attains a white colour. The actual temperature measurement is based upon the determination of the variations in the intensity (brightness) of the radiation emitted by the hot body and comparing it with known values generated by a heated filament.

Optical pyrometers have the following advantages:
1. they possess flexibility, portability and convenience for use
2. have no need for contact and hence are useful for monitoring the temperature of moving and distant objects
3. Are useful for high temperature measurements
4. have good accuracy

Despite these advantages optical pyrometers are relatively expensive, are prone to human operational errors and are not useful for measuring the temperatures of clean burning gases that do not radiate visible energy.

Exercise 3.4

1. Thermocouple systems are some of the most widely used temperature measurement systems in use in Industry. Discuss the most significant sources of error in such systems.
2. A platinum resistance thermometer has a resistance of 145.6 Ω and 120.0 Ω at 100 °C and 0°C respectively. Given that the temperature coefficient of platinum is 0.0039 $°C^{-1}$, determine the temperature if the resistance is 350 Ω.
3. For a certain thermistor, β = 3200 K and the resistance at 20 °C is known to be 1200 Ω. The thermistor is used for temperature measurement in a cold room, determine the temperature if the thermistor shows a resistance of 2400 Ω.
4. The resistance of a thermistor is 800 ohms at 50 °C and 4 k Ω at the ice point. Calculate the characteristic constants (A, β) for the thermistor and the variation in resistance between 30 and 100 °C.
5. Distinguish between the operating principles of radiation and optical pyrometers. What advantages do pyrometers have over other temperature measuring devices?

3.8 FLOW SENSING ELEMENTS

Fluid flow is an important process variable, particularly in continuous mixing operations where inaccurate flow rates affect properties such as the viscosity, calorific value, specific gravity and corrosiveness of the end product. Inferential instruments for direct measurement of rate of flow are widely used in process industries.

3.8.1 Classification of Flow Measuring Devices

Bernoulli's theorem can be stated as: -

The total energy per unit weight of a continuous mass of fluid passing through an isolated system remains constant from point to point along its line of flow, without regard to what may be done with the fluid within the system.

i.e.

$$H = z + \frac{P}{w} + \frac{v^2}{2g} = \text{constant}$$

where H = Total pressure head in the system, m
z = Potential head, m
P/*w* = Pressure head
w = specific weight of fluid

An isolated system exchanges no energy with the environment (i.e. there is no addition or loss of heat from or to the environment respectively). The total mechanical energy of the fluid, neglecting friction, is assumed to be constant from point to point.

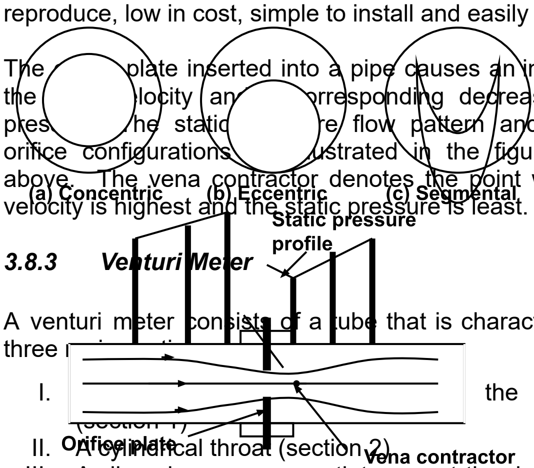
Three categories of flow measuring instruments based on Bernoulli's theorem are: -

1. Devices that change the velocity of the fluid momentarily as it passes through a constriction, and the resulting change in static head is taken as a measure of the velocity and hence the volume flow rate. Examples include orifices, flow nozzles and venture tubes.
2. Instruments in which the constriction area varies with the flow rate as the differential pressure is maintained at a constant value, e.g. rotameters.
3. Instruments that measure the velocity head directly without appreciably affecting the flow pattern of the fluid being measured. A blunt object placed in the path of flow measures the difference

between the impact pressure and the static pressure head, e.g. Pitot tube.

3.8.2 Orifice Plates

This is the most common type of head flow measuring device for medium and large pipe sizes. They are easy to reproduce, low in cost, simple to install and easily changed.

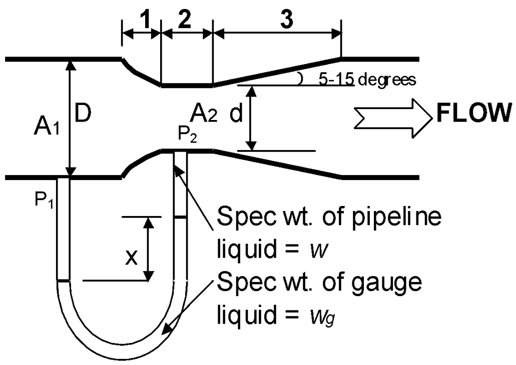


The orifice plate inserted into a pipe causes an increase in the flow velocity and a corresponding decrease in the pressure. The static pressure flow pattern and different orifice configurations are illustrated in the figure shown above. The vena contractor denotes the point where the velocity is highest and the static pressure is least.

3.8.3 Venturi Meter

A venturi meter consists of a tube that is characterised by three main sections:

I. A converging inlet cone at the upstream (section 1)
II. A cylindrical throat (section 2)
III. A diverging recovery outlet cone at the downstream (section 3).



Applying Bernoulli's equation to the upstream and throat sections (subscript 1 denotes 'upstream' and 2 denotes 'throat') and ignoring friction yields.

$$z_1 + \frac{v_1^2}{2g} + \frac{P_1}{w_1} = z_2 + \frac{v_2^2}{2g} + \frac{P_2}{w_2}$$

It can be shown that the actual discharge, Q [m3/s], is given by: -

$$Q = c_D A_1 \sqrt{\frac{2gH}{m^2 - 1}}$$

where

m = Area ratio = $\frac{A_1}{A_2}$

H = differential pressure between the two sections expressed as a column of the liquid in the meter, [m]

$$= \frac{P_1 - P_2}{w} = x\left(\frac{w_g}{w} - 1\right)$$

with w = specific weight of liquid in pipeline =ρg

w_g = specific weight of liquid in manometer gauge

x = u-tube manometer reading as indicated in the figure above.

Example 3.5

A horizontal venture meter measures flow of oil of specific gravity 0.9 in a 75 mm diameter pipeline. If the difference of pressure between the full bore and the throat tapings is 34.5kN/m^2 and the area ratio (m) is 4, calculate the flow rate assuming a coefficient of discharge of 0.97.

SOLUTION

w = 0.9 x 9.81 x 1000 = 8829 N/m3, m = 4, C_d = 0.97

EMBED Equation.3

EMBED Equation.3

Thus the actual discharge, EMBED Equation.3

EMBED Equation.3

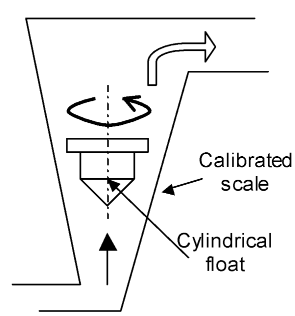
The discharge coefficient of a standard type of venture tube is about 0.99, valid for throat to pipe diameter ratios of 0.25 and 0.75. The approach to the throat is usually given a curved profile to maintain a constant discharge coefficient.

3.8.4 Area meters (Rotameters)

(Also called *constant pressure drop* or variable *aperture meters*)

The area flow meter operates on the principle that there is a different orifice area for each rate of flow and that consequently the differential pressure is constant.

A rotameter is an area flow meter that consists of a float that is supported in the flowing stream by the differential pressure as shown in the figure given below.



The construction consists of a conical tube made of glass, stainless steel, or monel.

The floats are made of brass, stainless steel, monel or special plastics.

Analysis of the variable area flow meter

The fundamental equation for incompressible flow through a tube is stated as: -

$$Q = c_D E A_2 \sqrt{\frac{2\Delta P}{\rho}}$$

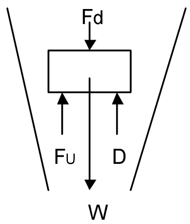
where

E = Approach factor

A_2 = Orifice area

ρ = fluid density

Consider the forces acting on a float in the vertical column of a fluid.



where

W = effective weight of the float

$= V_f(\rho_2 - \rho_1)$

with V_f = volume of float

ρ_2 = material density of float

ρ_1 = material density of fluid

F_d = Force acting downward on the upper surface of the float

$= P_2 A_f$

with P_2 = pressure per unit area on the upper surface of the float

A_f = effective surface area of the float.

F_U = Force acting upward on the lower surface of the float

$= P_1 A_f$

with P_1 = pressure per unit area on the lower surface of the float

D = Drag force that tends to pull the float upward in the direction of flow. This force depends on the float design and the conditions of flow.

Under equilibrium conditions,

$$F_U + D = W + F_d$$

Neglecting viscous forces, i.e. D = 0

$$F_U = W + F_d$$

$$P_1 A_f = V_f g(\rho_2 - \rho_1) + P_2 A_f$$

This gives

$$P_1 - P_2 = \frac{V_f}{A_f} g(\rho_2 - \rho_1) = \Delta P$$

Thus the flow rate Q is: -

$$Q = c_D E A_2 \sqrt{\frac{2\Delta P}{\rho_1}}$$

$$= c_D E A_2 \sqrt{2g \frac{V_f}{A_f}\left(\frac{\rho_2 - \rho_1}{\rho_1}\right)}$$

where A_2 = area between the float and the tube.

The flow rate is dependent on the relative densities of the material of the float and the fluid. Since the density of the fluid is temperature dependent, any changes in temperature during measurement could result in errors. Density compensation to minimise these errors can be achieved by selecting the material of the float such that its density is twice that of the fluid flowing in the meter, i.e.

$$\rho_2 \approx 2\rho_1$$

For liquids this condition is achieved by using a hollow float or by using plastic materials for the float.

<u>*Exercise*</u>

A rotameter uses a cylindrical float 0.02 m in height and 0.02 m in diameter, with proper density to give density compensation. If the discharge coefficient is 0.5 and the maximum inside diameter of the tube is 0.04 m, calculate the maximum flow rate. Take the approach factor to be approximately equal to 1. *(Hint: Set* $\rho_2 \approx 2\rho_1$ *, Answer = 0.00935 m3/s)*

3.9 Choosing Instrumentation for Ultrapure Corrosive Fluid Environments

Choosing flow, pressure, and level measurement instrumentation for the semiconductor industry can be particularly challenging. The primary issue is maintaining high purity in liquid environments that are often highly corrosive.

To meet this challenge, the following basic considerations should be followed when choosing measurement instrumentation for this environment: -

1. No moving parts
2. No fluid filled cavities
3. Metal free construction
4. Chemically resistant instrument exterior
5. Compact design
6. Reliability

NO MOVING PARTS

Instrumentation with moving parts that are used for prolonged periods inevitably risk becoming worn or degraded over time. When wearing or degradation occurs, devices with moving parts in the fluid stream can generate particles. Worn or partially degraded devices that generate particles will often function normally, making it difficult to identify the particle generation source. Avoiding the use of instruments with moving parts will minimize particle contamination.

NO FLUID FILLED CAVITIES

Fluid filled pressure transducers and gauge protectors typically have a metallic pressure sensor in contact with a stationary fill fluid. The fill fluid is in contact with a flexible seal (or isolation membrane) that separates it from the process fluid. The fill fluids from these instruments may leak, causing initial contamination from the fill fluid and continuous contamination after the process fluid fills the exposed cavity. While appropriate applications do exist for instruments utilizing fluid filled cavities, their use in ultrapure corrosive environments should be avoided.

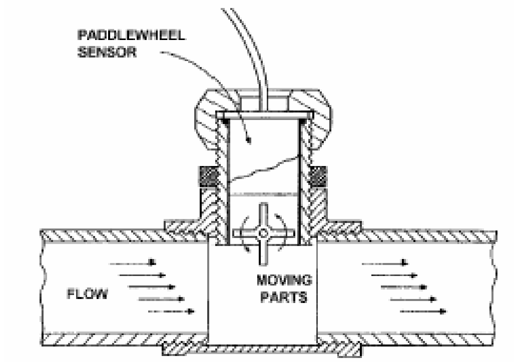


Figure 1: Example diagram of paddlewheel flowmeter with moving parts that can generate particles.

METAL FREE CONSTRUCTION

The chemicals commonly used in the semiconductor industry, such as hydrochloric acid, may chemically attack various metal parts of an instrument. The resultant attack upon the metallic surfaces can lead to instrument failure or contamination of the ultrapure process. The severity of this phenomenon is dependent on a number of factors, such as: the specific fluoropolymer or other materials used for wetted parts, the instrument design, chemical concentration, fluid temperature, and line pressure. In general, instruments in ultrapure or corrosive fluid environments should be constructed of metal free, high purity inert substances that are resistant to the chemicals being used.

COMPACT DESIGN

Flow, pressure, and level measurement devices come in all shapes and sizes, each with different installation requirements. Valve boxes, semiconductor processing tools, chemical distribution systems, and other equipment often have limited space considerations. Some instruments are not appropriate for installation into applications where footprint size is critical. These instruments may be too large, require specific installation or mounting considerations (i.e. vertical only), need straight lengths of tubing before or after the unit, are susceptible to electronic noise or vibration, or exhibit poor performance with fluids containing entrained solids or gas bubbles

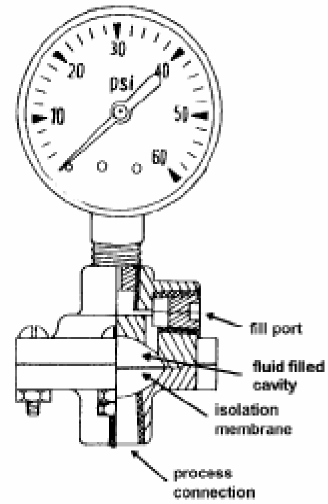


Figure 2: Example diagram of a fluid filled gauge protector that can leak fluid and cause process contamination.

CHEMICALLY RESISTANT EXTERIOR

Measurement instrumentation used in the semiconductor industry may be exposed to a variety of corrosive chemicals. These chemicals can attack the instrument exterior, typically from plumbing leaks or fumes. Electronics and wiring may corrode and will often lead to instrument failure. The product exterior and any electronics within the instrument must be protected from chemical exposure with appropriate design and materials of construction.

RELIABILITY AND EXPERIENCE

Critical applications require instruments with proven designs in order to provide years of problem free continuous duty in ultrapure and corrosive environments.

3.10 MEASUREMENT OF HYDROGEN CONCENTRATION (pH)

pH is a measure of the effective acidity or alkalinity of a solution. Some chemical processes that are dependent on the pH include precipitation, coagulation, neutralization, bleaching, settling, hydrolyzing, polymerizing, sizing, dyeing, etc.

Defined mathematically

$$pH = -\log\left[H^+\right]$$

The pH scale uses the negative exponent as a means of expressing the hydrogen ion concentration. For example the pH of a neutral solution = $-\log 10^{-7} = 7.0$. The practical range of the pH scale is between 0 and 14. A pH 0.0 corresponds to an acid solution of unit strength while a pH of 14 corresponds to a basic solution of unit strength.

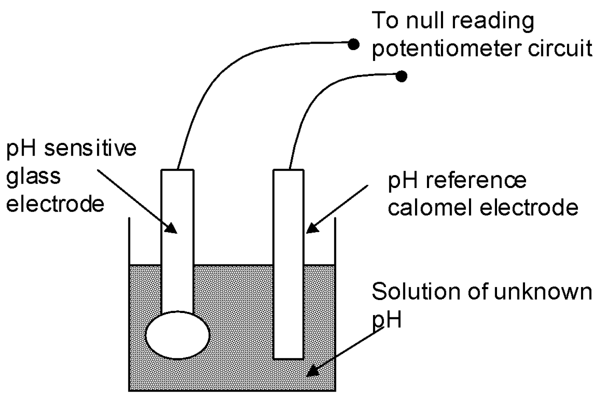
There are two methods of measurement of pH, namely: -

(i.) Derived calometric method based on the use of indicators and permanent colour standards. This technique is limited to visual comparison.

(ii.) Electrometric pH measurement which make use of transduction principles to give a proportionate electrical current/voltage signal.

3.10.1 Glass electrode pH sensor

This device consists of two electrodes: a measuring electrode and a reference electrode.



The measuring electrode is sensitive to the presence of hydrogen ions. Four types of measuring electrodes used include: -

(a) Hydrogen electrode - this is a laboratory standard. It is prone to contamination and requires a continuous supply of pure hydrogen

(b) Quinhydrone electrode - use is limited by its small range

(c) Antimony electrode - is usually poisoned by solutions containing copper, silver, mercury or lead.

(d) Glass electrode- Used most widely. When a thin membrane of special glass separates two solutions of different pH a potential drop is created that is a function of the difference in pH between the two solutions. If the pH of one solution is kept constant, that of the other can be expressed in terms of the potential developed. A typical glass electrode consists of a glass tube about 0.5" in diameter with a bulbous glass tip that contains the solution with a constant pH.

The reference electrode is not affected by pH but is used to complete the circuit. The calomel electrode possesses a reasonably constant potential over a wide range of temperatures and is not easily poisoned by solutions. This electrode comprises mercury and mercurous chloride (calomel) with saturated potassium chloride as an electrolyte. A liquid junction between the electrode and the process solution is established by means of a small hole covered by a ground-glass sleeve. Diffusion of the electrolyte from the electrode to the solution is slow, but sufficient to provide excellent electrical conductivities.

pH Recording Equipment for Electrode Based Sensors.
Null balance potentiometers or ac- amplifier circuits are used for pH measurements as opposed to deflection type devices. This is because the recording device must not draw any appreciable current from the electrodes as this will cause polarization thereby affecting the accuracy of the measurement. Potentiometers available are in the ranges of 0 - 7, 3 - 10, 6 - 13 and 2 - 12 pH.

3.10.2 FIELD EFFECT TRANSISTORS (FETS) AS TRANSDUCERS IN ELECTROCHEMICAL SENSORS

Sensor technology has advanced with the progress in technology associated with semiconductors, microprocessors, fiber optics, optoelectronics, ferroelectrics, acoustics and material science

Chemical sensors are microdevices that connect the chemical and electrical domains (i.e. transduction of the chemical information into electric signal). The response of the sensors should be fast and selective for the analyte. Moreover, these devices should have a lifetime in the order of months. The construction of chemical sensors requires the integration of a sensing receptor and a transducing element into a defined chemical system. Field effect transistors (FETs) are very interesting because they can be made very small with current planar IC technology and have the advantage of a fast response time.

Advantages of integrated sensors for instrumentation:
These are summarized as:

1. A reduction in external leads
2. On board signal conditioning
3. Lower signal to noise ratio
4. Standardisation of output signals for computer compatibility
5. Small physical size
6. High reliability
7. Incorporation of diagnostic routines and alarm generation
8. Greater sophistication
9. Better performance.

The application of field effect transistors (FETs) as transducers in electrochemical sensors was firstly described in 1970 by Bergveld. These devices can transduce an amount of charge present on the surface of the gate insulator into a corresponding drain current. The fast expansion of these transducing elements was possible due to the introduction of IC-technology in their construction, which allowed mass fabrication.

This modern technology provides a possibility to design multi-ion sensors integrated with the reference cell (REFET). These sensors are very small, long living and use only very small amounts of ion-sensing compounds.

TOPIC 4
SIGNAL CONDITIONING

4.0 INTRODUCTION

Usually the electric signal available at the output terminals of the transducer must be subjected to some form of conditioning (or modification) to render it suitable for transmission and display/recording. Signal conditioning elements *convert the signal occurring at their input terminals into an ac. or dc. voltage or current form that is suitable for further processing.*

4.1 *DEFLECTION BRIDGES*

These devices convert the output of resistive, capacitive and inductive transducers into a voltage signal.

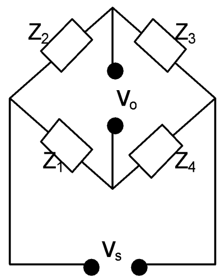


Fig. 4.1 Deflection bridge

Fig 4.1 gives the basic circuit for the deflection (or Wheatstone) bridge.

The bridge has for arms each with the impedances Z as indicated. A supply voltage V_s powers the circuit and the output V_o is obtained at the terminals shown.

The ***combined bridge impedance Z*** can be shown to be the parallel combination of Z_2 and Z_3 in series with the parallel combination of Z_1 and Z_4 i.e.

$$Z = \frac{Z_2 Z_3}{Z_2 + Z_3} + \frac{Z_1 Z_4}{Z_1 + Z_4}$$

>(4.1)

The ***bridge transfer*** is:

$$V_o = V_s \left\{ \frac{Z_1}{Z_1 + Z_4} - \frac{Z_2}{Z_2 + Z_3} \right\}$$

>(4.2)

4.1.1 *Design of resistive bridges*

In this bridge all impedances Z_1 to Z_4 are purely resistive. Hence the *bridge transfer* is:

$$V_o = V_s \left\{ \frac{R_1}{R_1 + R_4} - \frac{R_2}{R_2 + R_3} \right\}$$

>(4.3)

Let one of the resistance of one of the resistors in the bridge (R1) be that of the sensing element (R_{se}), i.e. $R_1 = R_{se}$; and R_2, R_3, R_4 be fixed resistors. Equation (4.3) can then be rewritten as:

$$V_o = V_s \left\{ \frac{1}{1 + R_4/R_{se}} - \frac{1}{1 + R_3/R_2} \right\}$$

>(4.4)

__Consideration #1__: Only V_s, R_4 and the ratio R_3/R_2 need to be specified in the design of the single element deflection bridge. These three are chosen considering the range of linearity of the output voltage and the electrical power limitations of the sensor.

The voltage output range limits corresponding to the minimum and maximum sensor resistances are:

$$V_{o\,(min)} = V_s \left\{ \frac{1}{1 + R_4/R_{se(min)}} - \frac{1}{1 + R_3/R_2} \right\}$$

>(4.5)

$$V_{o(max)} = V_s \left\{ \frac{1}{1 + R_4/R_{se(max)}} - \frac{1}{1 + R_3/R_2} \right\}$$

>(4.6)

__Consideration #2__: The bridge should be balanced initially, i.e. $V_{o(min)} = 0$.

Hence (5) reduces to:
$$\frac{R_4}{R_{se(min)}} = \frac{R_3}{R_2}$$

> (4.7)

or
$$R_4 = \frac{R_3}{R_2} R_{se(min)}$$

Substituting for R_4 in (4.4) and rearranging gives:

$$\frac{V_o}{V_s} = \left\{ \frac{1}{1 + \frac{R_3}{R_2} \Big/ \frac{R_{se}}{R_{se(min)}}} - \frac{1}{1 + R_3/R_2} \right\}$$

or
$$v(x) = \left\{ \frac{1}{1 + r/x} - \frac{1}{1 + r} \right\} = \left\{ \frac{x}{x + r} - \frac{1}{1 + r} \right\}$$

> (4.8)

where v(x) = is the variable bridge voltage transfer

= V_0/V_s

r =fixed resistor ratio =R_3/R_2, and

x = variable resistor ratio = $R_{se} / R_{se(min)}$

__*Note:*__ v(x) is always equal to zero at x =1(i.e. the balanced bridge condition) irrespective of r.

__*Exercise 4.1*__

Plot the graphs of v(x) for x =0.1 to 2.0 and for r =0.1, 1.0,10.0 and 100. Comment on the linearity of each of the graphs obtained.

__Consideration #3__: The resistance power (i^2R_{se}) of the sensor should be limited to a level which it is all dissipated as heat flow to the surroundings.

If W_{max} is the maximum power dissipation then the design condition is:

$$V_s^{\,2} \frac{R_1}{(R_1 + R_4)^2} \le W_{max}$$

>(4.9)

4.1.1.1 *Single strain gauge bridge voltage transfer*

The change in resistance $\Delta R = R_o Ge$ is small such that $x = R_{se} / R_{se(min)} \approx 1$. For maximum sensitivity r =1 (the graph that gives the greatest slope [$\partial v(x)/\partial x$] in the graph Exercise 4.1.). Hence R_3=R_2. From (7) $R_4 = R_{se(min)}$, hence the design condition for the resistors is R_2=R_3= R_4=$R_{se(min)}$. The supply voltage V_s is determined using (9).

Putting r = 1in (4.8) yields.

$$v(x) = \frac{x - 1}{2(x + 1)} \approx \frac{1}{4}(x - 1)$$

$$\frac{V_o}{V_s} = \frac{1}{4}\left[\frac{R_{se}}{Rse(min)} - \frac{R_{se(min)}}{Rse(min)} \right]$$

Hence
$$V_o = \frac{V_s}{4} \frac{\Delta R}{R_o} = \frac{1}{4} V_s Ge$$

>(4.10)

where $R_o = R_{se}$. The transfer is linear.

4.1.1.2 *Single element metallic resistance thermometer bridge voltage transfer*

Resistance temperature relation: $R_T = R_o(1+\alpha T)$. Typically $0 \le T \le 250$ ^{0}C and $R_0 = 100\ \Omega$ and $R_{250} = 200\ \Omega$. Therefore x varies between 1 and 2. The linearity of this device is small (less than 1%) hence the selection of r should give a linear bridge transfer. This implies choosing a large value of r e.g. 100. In (8) if r>>x then r + x ≈ r, therefore the equation can be reduced to:

$$v(x) \approx \frac{1}{r}(x-1)$$

⇒ if at T = 0^0C, $R_{se\,(min)} = R_0$ then

$$x = \frac{Rse}{Rse(min)} = \frac{R_T}{R_0} = 1 + \alpha T$$

Hence further substitution yields

$$V_o = V_s \frac{R_2}{R_3}\alpha T$$

>(4.11)

Example 4.1

Design a deflection bridge with two similar metal resistance thermometer elements to give an output that is approximately proportional to the temperature differences $T_1 - T_2$ if the first element is at T_1 and the second is at T_2 ^{0}C.

4.1.1.3 *Thermistor bridge voltage transfer*

$$R_\theta = A\exp\left(\frac{\beta}{\theta}\right)$$

Resistance temperature relationship. For a typical thermistor the resistance at 298 K (25 ^{0}C) i.e. R_{298} = 12 kΩ and R_{348} (at 75 ^{0}C) = 2 kΩ, therefore x varies from 1.0 i.e. 12/12) to 0.17 (i.e. 2/12) over this measurement range. Selecting r between 0.25 and 0.30, the bridge non-linearity partially compensates the thermistor non-linearity thereby realizing a fairly linear transfer over this range.

SOLUTION

For metallic resistance thermometers: $R = R_0(1 + \alpha T)$

Hence at T_1, $R_1 = R_0(1 + \alpha T_1)$ and at T_2

$R_2 = R_0(1 + \alpha T_2)$, $R_0 = R_{se\,(min)}$

The bridge output should be zero when $T_1 - T_2 = 0$ hence at $T_1 = T_2$, $R_1 = R_2 = R_0$

EMBED Equation.3 i.e. $R_4 = R_3$

The general bridge transfer is:

EMBED Equation.3

EMBED Equation.3

Substituting known values with $R_4 = R_3$:

EMBED Equation.3

EMBED Equation.3

If EMBED Equation.3 the denominator approximates to R_3/R_0 and the transfer function approximates to:

EMBED Equation.3 > (4.12)

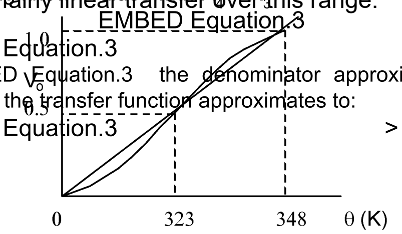


Fig 4.2. Thermistor bridge design conditions

If the output range is 0 to 1.0 V corresponding to an input range of 298 K to 348K generating three equations and solving for Vs, R_4 and r completes the design. Two of the equations are easily obtained using the required limits of the input and output. The third is obtained by equating corresponding mid-range values of the input and output ranges. (i.e. set V_o = (1.0-0)/2=0.5V at (348+298)/2 = 323K). The three equations to solve in this case are:

$$0.0 = V_s\left\{\frac{1}{1+R_4/R_{298}} - \frac{1}{1+R_3/R_2}\right\}$$

$$0.5 = V_s\left\{\frac{1}{1+R_4/R_{323}} - \frac{1}{1+R_3/R_2}\right\}$$

$$1.0 = V_s\left\{\frac{1}{1+R_4/R_{348}} - \frac{1}{1+R_3/R_2}\right\}$$

Exercise 4.2

1. Show that for the cantilever load cell with four active gauges $V_o = V_s G e$, where $e = \dfrac{6(l-x)}{wt^2E}$
2. Show that for the pillar load cell with four active gauges $V_o = \dfrac{V_s}{2}\dfrac{GF}{AE}(1+\nu)$

4.1.2 *Design of reactive deflection bridges*

In these bridges two arms are usually reactive impedances and the other two are purely resistive.

For the capacitive parallel plate level sensor it was shown that $C(x) = C_o\left[1 + \frac{x}{h}(\varepsilon_r - 1)\right]$ where $C_o = C(o)$. If $Z_1 = 1/(j\,\omega C_b)$; $Z_2 = R_2$; $Z_3 = R_3$; $Z_4 = 1/[j\,\omega C(x)]$. Hence

$$V_o = V_s\left\{\frac{1}{1+C_b/C(x)} - \frac{1}{1+R_3/R_2}\right\}.$$

For zero output at the minimum level of the sensor (i.e the balanced bridge condition) the bridge balancing capacitor $C_b = C(0)(R_3/R_2) = C_0(R_3/R_2)$.

$$V_o = V_s\left\{\frac{1}{1+\frac{C_o}{C(x)}\frac{R_3}{R_2}} - \frac{1}{1+\frac{R_3}{R_2}}\right\}$$

>(4.13)

If R_3/R_2 >>1the transfer is approximately linear, i.e.

$$V_o \approx V_s\frac{R_2}{R_3}\left(\frac{C(x)}{C_0} - 1\right)$$

>(4.14)

Example 4.2

A capacitive olive oil level sensor consists of two parallel plates of width 5 cm and height 10 cm. The separation of the plates is 2 mm. The bridge voltage supply is 230 V and $R_3 = 1000\ R_2$. Find C_b and determine the bridge output and displacement sensitivity (in volts per metre), when the liquid level is 50% of the plate height. (Take ε_0 = 8.85 pF/m, for olive oil $\varepsilon_r \approx 3$)

4.2 *MEASUREMENT AMPLIFIERS*

The low level output signal from a transducer is scaled up by an amplifier which increases its power while preserving the signal's information content. Amplifiers used in measurement have the following characteristics: -

- High accuracy
- Low noise distortion
- Accurately defined gain that may be adjustable.

The typical amplifier circuit is a compensation network that allows a comparison between the input V_i (i.e. the signal from transducer) and the device output V_o without loading the measurement object.

SOLUTION

Given data w = 0.05 m; h = 0.1 m; R_3/R_2 = 1000; x/h = 0.5, d = 2 mm; V_s = 230V

EMBED Equation.3

Bridge displacement sensitivity

EMBED Equation.3

The transfer function for the amplifier configuration with negative feedback Fig 4.4 (b) is derived as:

At the summing junction $V_f = V_i - V_{ref}$

Hence, $V_i = V_f + V_{ref}$ >(4.15)

Also $V_o = A_o V_f \Rightarrow V_f = \frac{V_o}{\beta}$ and $V_{ref} = \beta V_o$. Substituting V_f and V_{ref} in Eq.(4.15) gives:

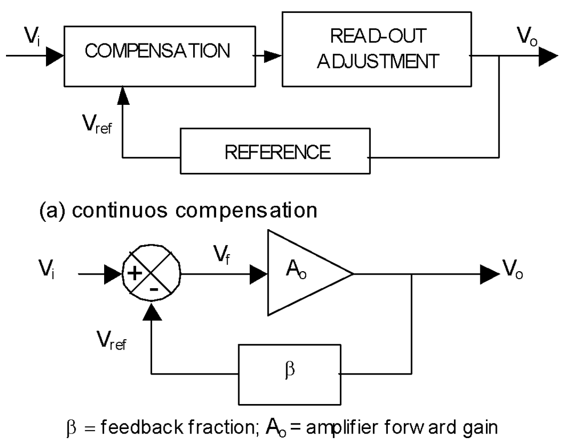
$$V_i = \frac{V_o}{A_o} + \beta V_o$$

and hence the transfer function is:

$$\frac{V_o}{V_i} = \frac{A_o}{1 + A_o\beta} = A$$

>(4.16)

Where A = overall gain (closed loop gain)



(a) continuos compensation

β = feedback fraction; A_o = amplifier forward gain

(b) continuos compensation with negative feedback

Fig 4.3 Block diagram representation of the measurement amplifier

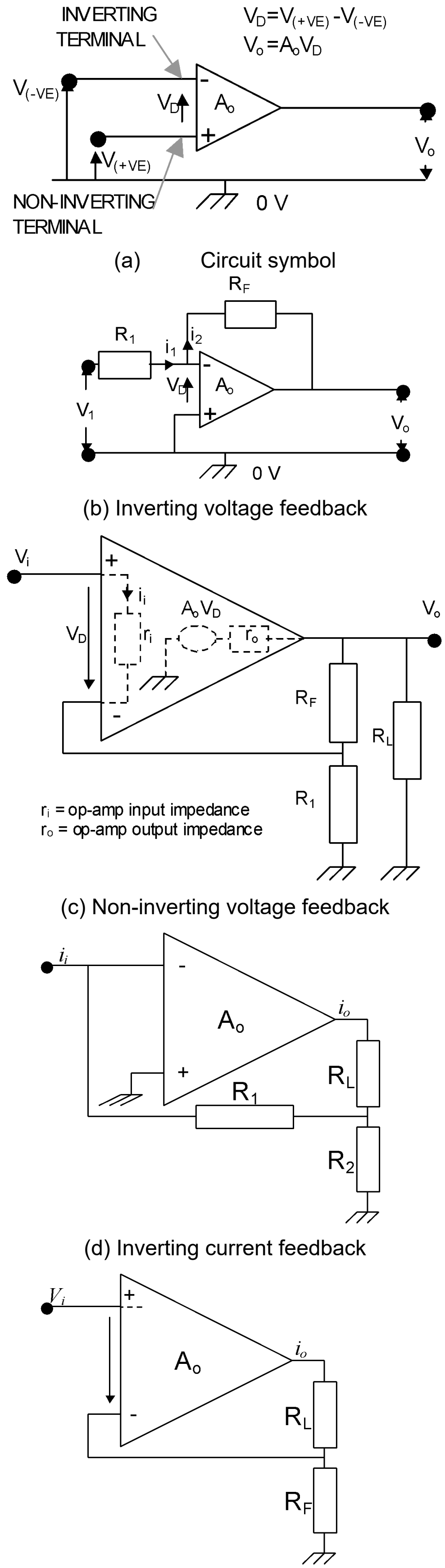
Amplifier Desensitivity (W)

This is the magnitude of the denominator in the amplifier's transfer function. W = $|1 + A_o\beta|$. It indicates how much the voltage gain is reduced by negative feedback. For a strong negative feedback the desensitivity is much greater than one ($1 << A_o\beta$) and the overall gain approximates to the inverse of the feedback fraction (i.e. $A = l / \beta$)

4.2.1 The Operational Amplifier (op-amp)

This amplifier derives its name from the ability to perform mathematical operations as addition, integration and differentiation in analogue computers. The typical op- amp consists of a single silicon integrated circuit and has the following characteristics: -

- high dc forward gain (ideally infinite)
- high input impedance (ideally infinite)
- low output resistance (ideally zero)



(a) Circuit symbol

(b) Inverting voltage feedback

(c) Non-inverting voltage feedback

(d) Inverting current feedback

(e) Non-Inverting current feedback

Fig.4.4 Different op-amp circuits

Assumptions made when evaluating op-amp parameters: -

The following assumptions based on the *ideal op-amp* are made when evaluating op-amp parameters: -

i.) the differential voltage across the input terminals is zero ($V_D = 0$)

ii.) the op-amp input current is zero

iii.) the output voltage change resulting from an input voltage change is such as maintain the difference voltage at the op-amp inputs at zero.

4.2.1.1 **Overall (Closed loop) Voltage Gain**

i.) *Inverting voltage feedback, Fig. 4.4. (b)*

Here V_D=0 and the overall input resistance and output resistances are R_1 and R_2 respectively.

From rule 2: $i_1 = i_2$, hence $V_1/R_1 = -V_o / R_2$. (The input is fed into the inverting terminal and therefore the output is oppositely charged, i.e. inverted)

⇒ Overall voltage gain, $A = V_o / V_i$

$$= -R_F/R_1 = -1/\beta \qquad > (4.17)$$

Where β = feedback fraction = R_1 / R_2

ii.) Non-inverting voltage feedback, Fig. 4.4. (c)

Here $V_D = V_{(+VE)} - V_{(-VE)}$,

>(4.18)

where $V_{(+VE)}$ = the voltage at the non-inverting terminal $= V_i$

$V_{(-VE)}$ = the voltage at the inverting terminal

$$= \frac{R_1}{R_1 + R_F} V_o \text{, by the voltage divider rule}$$

$$= \beta V_o \text{, with } \beta = \frac{R_1}{R_1 + R_F}$$

>(4.19)

$\therefore V_D = V_i - \beta V_o$

>(4.20)

Also, $V_o = A_o V_D$

$= A_o (V_i - \beta V_o)$

The overall gain, $$A = \frac{V_o}{V_i} = \frac{A_o}{1 + A_o\beta}$$

>(4.21)

Again if $1 \ll A_o\beta$, then $A \approx 1/\beta = 1 + R_F/R_1$

The overall gain is usually quoted in *decibels*.

$$A\,[dB] = 20 \log_{10} |A|$$

Closed loop input impedance, R_i,

As seen before for the inverting op-amp $R_i = R_1$. For the non-inverting op-amp, $R_i = V_i / i_i$.

From (4.20), $V_i = V_D + \beta V_o = (1 + A_o\beta)V_D$

But $V_D = i_i r_i$ hence,

$$R_i = (1 + A_o\beta) r_i = W\, r_i$$

>(4.22)

Closed loop output impedance, R_o,

It can similarly be shown that

$$R_o = \frac{r_o}{1 + A_o\beta} = \frac{r_o}{W}$$

>(4.23)

<u>*Example 4.3*</u>

Given the circuit configuration Fig.4.4(c), with Ao = 100,000, r_i = 2 MΩ, r_o = 75Ω, R_1=100Ω and R_F=100,000Ω, determine the: -

i.) feedback fraction

ii.) desensitivity

iii.) overall gain

iv.) closed loop input impedance

v.) closed loop output impedance.

<u>SOLUTION</u>

i.) Feedback fraction = EMBED Equa

Equation.3

0.001

ii.) Desensitivity, W = 1 + A_o

= 1 + 100,000 (0.001

iii.) Overall gain, A = A_o / W

101 990

iv.) Closed loop input impedance,

=202 M

v.) Closed loop output impedance, R_o = r

0.743

(iii) Inverting current feedback, Fig. 4.4. (d)

Here

$$\beta = \frac{R_2}{R_1 + R_2}$$

>(4.24)

$$A = \frac{i_o}{i_i} = \frac{A_o}{1 + A_o\beta} \approx \frac{1}{\beta}$$

>(4.25)

$$R_i = \frac{R_1}{1 + A_o\beta}$$

>(4.26)

$$R_o = (1 + A_o)R_2$$

>(4.27)

<u>*Example 4.4*</u>

Fig. 4.5 shows a sensitive dc ammeter using inverting current feedback. How much input current do we need to get full – scale deflection of the output meter? Find the closed loop input resistance and closed loop output resistance.

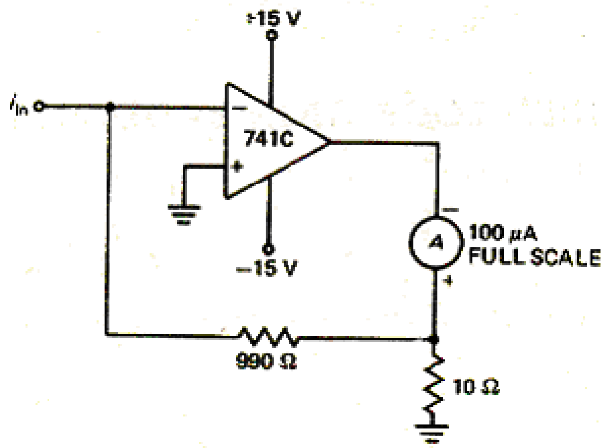


Fig 4.5 DC ammeter using inverting current feedback

SOLUTION

From Fig. 4.5, $R_1 \equiv 990$, $R_2 \equiv 10$, and for the 741C op-amp A_o 100000, hence:

Feedback fraction β = EMBED Equation.3 = EMBED Equation.3

Desensitivity, $W = 1 + A_o\beta$ = 1 + 100,000 (0.01) = 1001

The overall closed loop gain is

EMBED Equation.3

Closed loop input impedance is

$R_i = R_1 / W = 990/1001 \quad 1$

Closed loop output impedance is

$R_o = (1 + A_o)R_2 \quad A_oR_2 = (100000)(10) = 1$ M

(iv) Non-inverting current feedback, Fig. 4.4 (e)

A feedback amplifier with a non-inverting current feedback is a transconductance amplifier. The circuit Fig 4.4 (e) is also called a voltage to current converter because an input voltage controls an output current.

Therefore EMBED Equation.3 $g_m = \frac{i_o}{V_i} \cong \frac{1}{R_F}$

Here the transconductance

>(4.28)

Also $\beta = \frac{R_L}{R_L + R_F}$

>(4.29)

$$R_i = (1 + A_o\beta)r_i$$

>(4.30)

$$R_o = (1 + A_o)R_F \approx A_oR_F$$

>(4.31)

<u>*Exercise 4.3*</u>

A sensitive dc voltmeter using a non-inverting current feedback is shown below (Fig 4.6). For each switch position calculate the transconductance, the feedback fraction and the closed loop input and output resistances assuming that a full scale deflection of the ammeter is observed in each case.

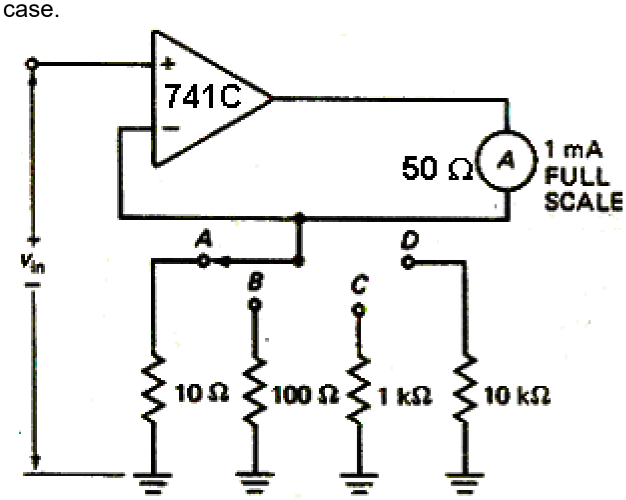


Fig. 4.6 DC voltmeter using inverting current feedback

4.2.1.2 The 741 op-amp characteristics

The 741 has become the industry standard and is often used in test circuits before upgrading to a better op-amp if design specifications are not met. The commercial grade

741C is least expensive and is the most widely used. Typical specifications for this op- amp are summarised in the table below: -

Table 4.1 Ideal and typical op-amp characteristics

Parameter	Ideal op-amp	741 op-amp
D.C. open loop gain	∞	100 dB≈ (10^5)
Input impedance, r_i	∞	2MΩ
Output impedance, r_o	0	75Ω
Input offset voltage, V_{os}	0	1 mV
Temp coeff. of V_{os}, γ	0	5 μV°C^{-1}
Input bias current, i_b	0	80nA
3 dB bandwidth 0 to f_B	0 to ∞	0 to 10 Hz
CMMR	∞	90 dB

(a) Slew rate (S_R) for the 741 op-amp

The slew rate indicates gives the maximum rate at which the input can be varied. It is defined mathematically in terms of TTL parameters as the ratio of the maximum value of tail current (I_T) to the charge capacitance (C_c),

i.e. $S_R = dV_c / dt = I_{T(max)} / I_C$ [V/μs].

>(24)

For the 741C, typically I_T = *15 μA,* C_c= 30 pF, hence

S_R =*15 μA* / *30* pF = 0.5 V / μs

⇒ the output voltage can change no faster than 0.5 V / μs. Beyond this value the output signal is distorted (*Slew rate distortion*).

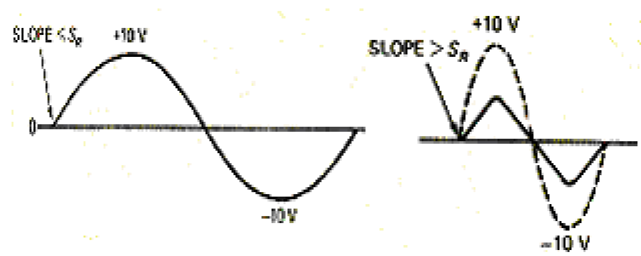


Fig. 4.7 Slew rate distortion

(b) Power Bandwidth

For a sinusoidal waveform input signal,

$f_{max} = S_R / 2\pi V_p$

>(4.25)

Where f_{max} = power bandwidth (i.e. the highest undistorted frequency)

V_p = peak of output sine wave. For the 741C, V_p = 10V, S_R= 0.5 V /μs

∴ f_{max} ≈8 kHz

Hence the l0 V power bandwidth of the 741C is approximately 8 kHz. The output voltage drops off at input frequencies higher than this.

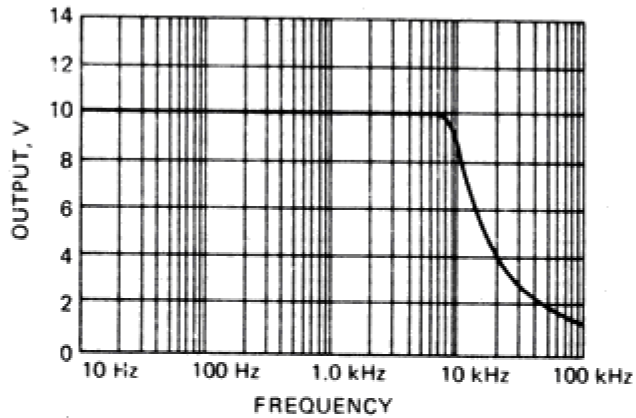


Fig. 4.8 Large signal response of 741 C

4.2.2 ***The Differential Amplifier***

Most transducers produce signals of only a few millivolts and, in many cases, the signal must be carried to the amplifier over relatively long lengths of wire. Noise signals picked up along the wire in the form of "unwanted voltages can easily swamp the signal.

This problem may be overcome by using a differential amplifier that is fed by two wires that are closely twisted together and caged in a common earth screen. Any noise signals that may be picked up are referred to as ***common mode*** signals. The amplifier is designed to amplify only ***differential mode*** signals.

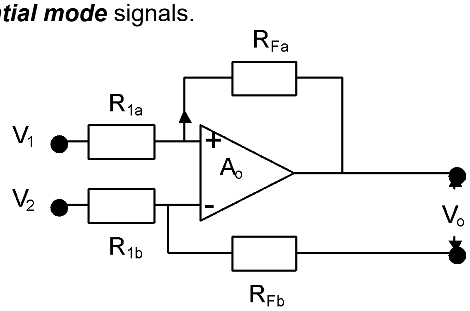


Fig 4.9 The Differential amplifier circuit

The transfer equation for the differential amplifier (Fig. 4.9.) can be derived as follows:

$$V_o = A_1V_1 + A_2V_2 \qquad (26)$$

Where A_1 = amplification along non-inverting path "a"

$A2$ = amplification along inverting path "b"

Rearranging (4.26) to a sum and difference form yields:

$$V_o = \frac{A_1 + A_2}{2}(V_1 - V_2) + \frac{A_1 - A_2}{2}(V_1 + V_2)$$

$$= A_dV_j - A_cV_n \qquad (4.27)$$

Where A_d = differential mode voltage gain

$$= \frac{A_1 + A_2}{2}$$

A_c = common mode voltage gain

$$= -\frac{A_1 - A_2}{2}$$

V_i= differential input voltage = $V_1 - V_2$

V_n = unwanted noise voltage = $V_1 + V_2$

COMMON MODE REJECTION RATIO, CMMR

This is the relative ability of a differential amplifier to maximise the differential mode gain at the expense of the common mode gain.

$$\text{CMMR} = \frac{A_d}{-A_c} \qquad (4.28)$$

Example 4.5

A pressure transducer develops an output voltage of 40 mV and is connected to a differential amplifier that has a differential mode gain of 46 dB and a common mode rejection ratio of 80 dB. Find the maximum common mode noise signal that can be tolerated if the component of output voltage caused by the noise signal is not to exceed 1% of the required signal.

SOLUTION

Given: V_i = 40 mV

A_d = 46 dB = $10^{46/20}$ = 200

CMMR = A_d/-A_c = 80 dB = 10,000

A_c = 200/10,000 = -0.02

Required: $V_{n, max}$ if $A_cV_n \leq$ 1% of A_dV_i

From (27) $V_o = A_dV_i - A_cV_n$

A_dV_i = 200 X 40 mV = 8 V

Set $A_cV_{n, max}$ = 0.01A_dV_i = 0.08 V

$V_{n, max}$ = 0.08/0.02 = 4 V

Hence the maximum common mode noise signal is 4 V.

4.2.3 **The Instrumentation Amplifier**

The differential voltage amplifier has a low input impedance that may load measurement object considerably. This problem is overcome by incorporating two extra voltage amplifiers to compensate the differential input voltage. The resulting configuration is known as an **instrumentation amplifier** (Fig. 4.10).

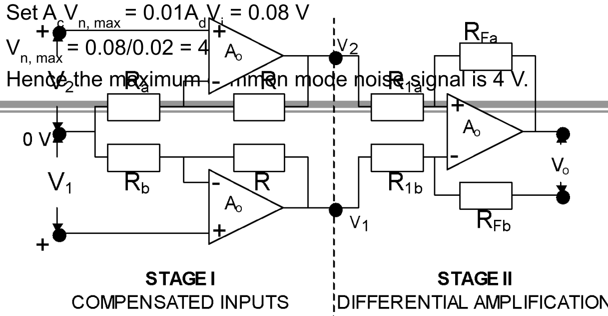


Fig. 4.10 Circuit configuration of the Instrumentation amplifier

The overall differential mode gain (A_d) for this amplifier is the combined gain from the two stages.

STAGE I

$$A_{d1} = \frac{v_2 - v_1}{V_d} = \frac{R_a + R_b}{R_a} \quad \text{where } V_d = V_2 - V_1$$

STAGE II

If the resistance ratios on each side of the amplifier are equal: -

$$A_{d2} = \frac{V_o}{v_2 - v_1} = \frac{R_{Fa}}{R_{1a}} = \frac{R_{Fb}}{R_{1b}}$$

$$\therefore A_d = \frac{V_o}{V_d} = A_{d1}A_{d2}.$$

$$= \frac{R_a + R_b}{R_a}\frac{R_{Fa}}{R_{1a}} =$$

$$\frac{R_a + R_b}{R_a}\,\frac{R_{Fb}}{R_{1b}} \qquad >(4.29)$$

The overall CMMR for the differential amplifier = A_d / A_c =

$$\frac{A_{d1}}{A_{c1}}\,\frac{A_{d2}}{A_{c2}} = \frac{A_d}{A_{c2}} \quad \text{since } A_{c1} = 1 \qquad >(4.30)$$

Exercise 4.4

Determine the CMMR of the instrumentation amplifier Fig 4.6. if the following parameters are specified: $R_a = R_{1a} = R_{1b} = 1\ k\Omega$ and $R_b = R_{Fa} = R_{Fb} = 100\ k\Omega$.

4.3 INSTRUMENTATION FILTERS

Filters are used in measurement systems to improve the quality of the output signals from a transducer so as to facilitate further processing or the extraction of information.

4.3.1 Filter Classification

Filters may broadly be classified as follows:-

i.) According to *construction*: Passive and Active filters

ii.) According to the *nature of the input and output signal*: Analogue and Digital

iii.) According to *filter magnitude*: Low -pass, High-pass, Band–pass and Band -reject

iv.) According to the *frequency response / transfer characteristics*: Butterworth, Chebychev, and Bessel

4.3.1.1 Passive and Active Filters

Passive filter networks make use of resistance (R), inductance (L), and capacitance (C), and hence are also referred to as RLC- networks. Drawbacks of these networks include difficulty in cascading of filters in spite of the design. At lower frequencies inductors are bulky and expensive.

Active filters make use of operational amplifiers that are suitable for low frequency applications and hence avoid the use of inductors. Active capacitance multiplication enables the use of capacitors with low practical values and giving fractional cut-off frequencies. However, these filter networks have a limited gain bandwidth product making them inappropriate for high frequency applications. Passive filters are preferred for signal frequencies above a few hundred kilohertz.

4.3.1.2 Analogue and Digital Filters

***Analogue** filters* process time-continuous signals received at the input terminals to give a time-continuous output signal. Laplace transforms are used in their analysis.

***Digital** filters* receive a sequence of input data values that are discrete (i.e. discontinuous with respect to time) and process them to give an output that has a corresponding sequence of digital values.

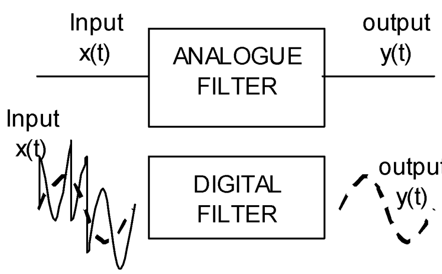


Fig. 4.11 Schematic presentation of analogue and digital filters

Advantages of digital over analogue filters include: -

i.) compatibility with computer hardware and software and hence the ease with which they are built and tested

ii.) unlike analogue filters that use differential operations, digital filters use arithmetic operations (addition, subtraction and multiplication) making them easier to understand and or modify.

4.3.1.3 Filter Magnitude Specifications

Where the main consideration is the *gain* or *attenuation* we have- low pass, high pass, band pass and band reject filter networks.

Low Pass (LP) Filter

This filter *passes low frequency signals occurring at the input terminals up to a desired cut-off above which higher frequency signals are attenuated (scaled down).* This filter is specified by the cut-off frequency (ω_c), Stop band (SB) frequency (ω_s), dc gain, Pass Band (PB) ripple. The filters' pass band frequency range is $0 \leq \omega \leq \omega_c$ and the stop band frequency range is $\omega > \omega_s$. This characteristic is illustrated in Fig.4.12 (a.)

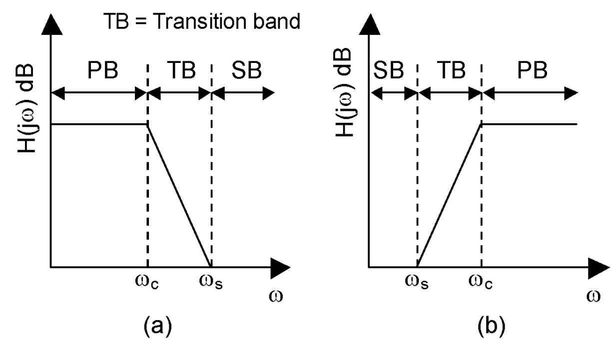


Fig. 4.12 (a.) Low Pass Filter Characteristic and (b.) High Pass Filter Characteristic

High Pass (HP) Filter

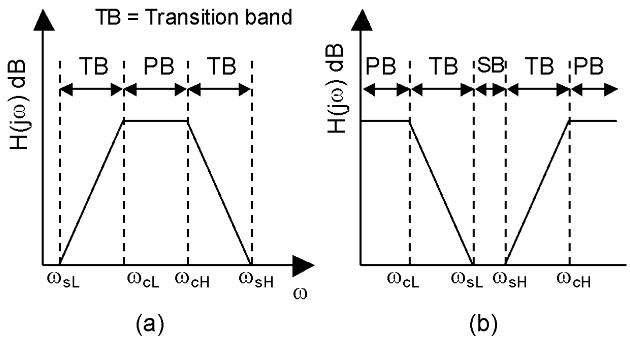
This filter passes high frequency signals occurring at the input terminals above a desired cut-off frequency below which lower frequency signals are attenuated (scaled down), Fig. 4.12 (b).

Band Pass (BP) Filter

This filter passes signals with a finite band of frequencies while attenuating both lower and higher frequency signals below or above the desired cut-off frequencies. They have an upper as well as a lower stop band. Fig. 4.13(a)

Band Reject (BR) Filter

This filter attenuates a finite band of frequencies while passing both lower and upper frequency pass bands. Fig.4.13 (b)



ω_{sL} = lower SB frequency; ω_{sH} = upper SB frequency

ω_{cL} = lower cut-off frequency; ω_{cH} = upper cut-off frequency

Fig. 4.13 (a.) Band Pass Filter Characteristic and (b.) Band Reject Filter Characteristic

4.3.1.4 Frequency Response

A final classification of filters is based upon their frequency response characteristics.

Butterworth Filters

This is a class of filters that are optimised for maximum flatness of their frequency response characteristic in the pass band. Their attenuation monotonically (i.e. step input-step output change) increases within the pass band to a value 3.01 dB at the cut-off frequency and further increases at the rate of (n x 6) dB per octave, ultimately at higher frequencies, for an n^{th} order filter.

Chebychev Filters

These filters provide a higher rate of attenuation near the cut-off frequency than that available from Butterworth filters. Their frequency response does not show monotonic attenuation in the pass band but the attenuation may increase or decrease depending on the order of the filter.

Bessel Filters

Butterworth and Chebychev filters are suitable for operation with continuous signals whose transient response shows considerable overshoot. Bessel filters are used when handling signals with negligible overshoot in the step response.

4.3.2 *Active Filters*

By using op-amps it is possible to build *active RC* filters that produce the sharp roll-off associated with *passive LC* filters.

4.3.2.1 Low-Pass Filter

In Fig. 4.14, a lag network is tacked onto the input side of a non-inverting voltage amplifier.

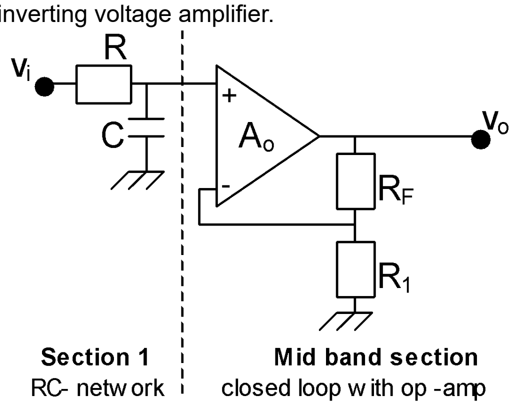


Fig. 4.14 First order (One-pole) low-pass filter

The transfer function is the product of the gains from the two sections. For the RC part of the circuit the transfer is

$$G_1(s) = \frac{1}{1+\tau s} \qquad >(31)$$

Where τ = time constant [s] = RC

Setting s = jω in (31) with $\omega = 2\pi f$ gives

$$G_1(j\omega) = \frac{1}{1+j\omega\tau},$$

$$= \frac{1}{1+j\,2\pi RC\,f} = \frac{1}{1+j\dfrac{f}{f_c}}$$

where f = frequency of the input signal [Hz]

f_c = - 3 dB cut-off frequency (also called corner frequency).

$= 1/2\pi RC$ [Hz]

As seen earlier the midband gain is

$$G_2(j\omega) = A \approx 1/\beta$$

Where β = feedback fraction

$$= \frac{R_1}{R_1+R_F}$$

$$\therefore \quad A \approx 1 + R_F/R_1$$

Hence the overall transfer function:

$$G(s) = v_o/v_i$$

$$= G_1G_2(s) = \frac{A}{1+j\dfrac{f}{f_c}} \qquad >(32)$$

The transfer function has one *j* factor in the denominator (due to one *lag network* in the filter circuit) and hence the term *<u>first order</u>* or *<u>one-pole</u>* filters. Above the cut-off frequency, the voltage gain rolls off at 6 dB per octave or 20 dB per decade.

4.3.2.2 Second Order (Two Pole) Low-Pass Filter

Fig. 4.15 is a two-pole low pass filter because it has two lag networks. The capacitor of the first lag network receives a feedback voltage that modifies the cut-off frequency and the response of the active filter.

Here the transfer function is:

$$\frac{v_o}{v_i} = \frac{A}{s^2R^2C^2 + (3-A)sRC + 1}$$

The damping factor ξ = (3 - A)/2. If A = (3 - √2) or 1.586, ξ = 1 / √2 and a maximally flat (or also called ***Butterworth***) response is obtained.

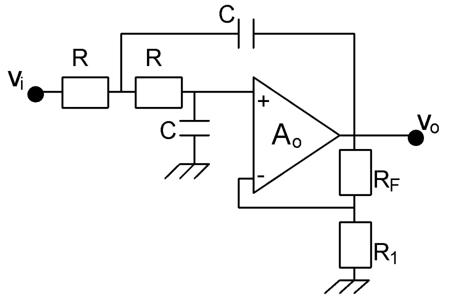


Fig. 4.15 Second order (two-pole) low-pass filter

For this filter a closed loop non-inverting gain of 1.586 is a critical value. If,

- A < 1.586 -the filter response approaches a linear phase shift with frequency (Bessel response)
- A > 1.586- ripples in the mid-band (Chebychev response)
- A = 1.586- maximally flat (Butterworth response).

For the Butterworth response,

$$1.586 = 1 + R_F/R_1$$

$$\Rightarrow R_F = 0.586\, R_1$$

If R_1 = 1 kΩ then R_F= 586 Ω ≈ 560 Ω (nearest standard value). Again f_c = - 3 dB cut-off frequency

$$= 1/\,2\pi RC \text{ [Hz]}$$

Above this frequency the voltage gain decreases by 12 dB per octave (i.e. n x 6, with n = 2 the number of poles / lag networks) or 40dB per decade.

4.3.2.3 High-Pass Filter

A lead network instead of a lag network is tacked onto the input side of the circuit configuration Fig. 4.14 to realise the one-pole high pass filter (Fig. 4.16).

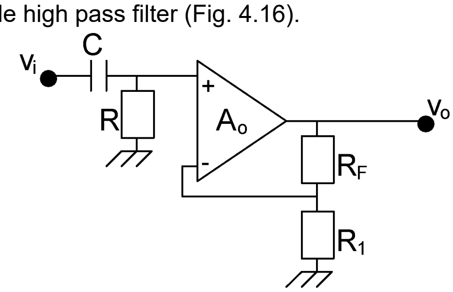


Fig. 4.16 First order (One-pole) high-pass filter

As before the transfer function is,

$$v_o/v_i = \frac{A}{1 + j\,\frac{f}{f_c}}$$

4.3.2.4 Higher Order Filters

Filters with more than two poles are realised by cascading (i.e. connecting in series) two or more first or second order filter networks. The gain in each section is selected so as to realise the Butterworth condition (i.e. maximally flat response in the pass band).

Table 4.2. Gains for Butterworth Filters

No. of Poles	1st section (1or 2 poles)	2nd section (2 poles)	3rd section (2 poles)
1	Optional		
2	1.586		
3	Optional	2	
4	1.152	2.235	
5	Optional	1.382	2.382
6	1.068	1.568	2.482

Example 4.4

For the filter circuit configuration shown in Fig. 4.17: -

(i) show that the filter response approaches the Butterworth response: -

(ii) determine the cut-off frequency and

(iii) the rate at which the voltage gain drops below cut-off.

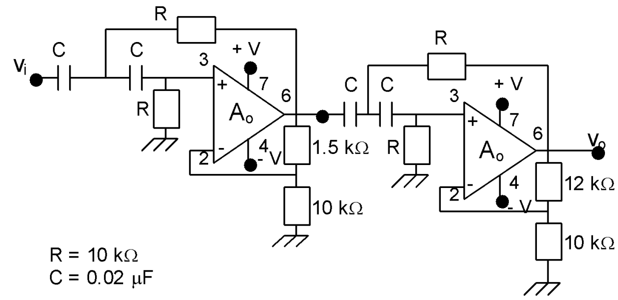


Fig. 4.17 4-pole active filter network

SOLUTION

The circuit has four lag networks and is therefore a four-pole or fourth order high pass filter,

(i) From Table 4.2 the closed loop gain for the first and second sections should ideally be 1.152 and 2.235 respectively.

First section: R_F = 1.5 k, R_1 = 10 k

A = 1 + R_F / R_1

= 1.15 (close to the ideal; 1.152)

Second section: R_F = 12 k, R2 = 10 k

Hence A = 2.2 (Again this is close to the ideal, 2.235)

The filter response in both sections approximates to the Butterworth response.

(ii), cut-off frequency, fc = 1/2πRC

For the circuit configuration given, R = 10 k and C = 0.02 μF hence *fc* = 796 Hz

(iii) n = 4 the voltage gain drops off at a rate of 24 (n x 6) dB per octave or 80 dB per decade.

Exercise 4.5

1. With the aid of appropriate sketches differentiate between a charge amplifier and an instrumentation amplifier.
2. The voltage follower circuit configuration (Fig. 4.18), uses a 741C op-amp with a forward gain of 100 dB, a differential input resistance of 2 MΩ and an output resistance of 75 Ω. Determine the:
   i.) desensitivity factor
   ii.) closed-loop voltage gain
   iii.) closed-loop input impedance
   iv.) closed-loop output impedance

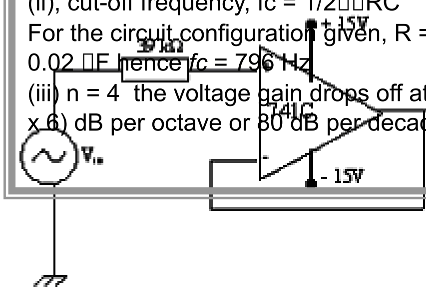


Fig 4.18 Voltage follower

3. Figure 4.19 shows an electronic thermometer. At 0°C, the thermistor has a resistance of 20 kΩ and a characteristic temperature (β) of 2000 K. What does the ammeter read at 0°C? At 25°C? At 50°C?

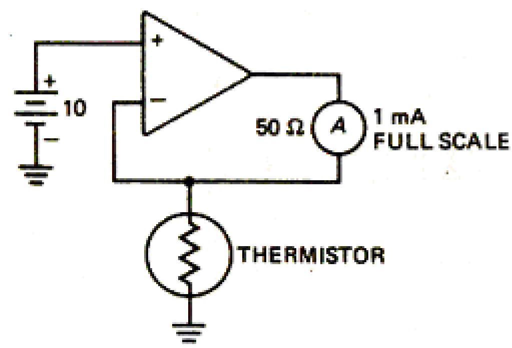


Fig 4.19 Electronic thermometer

3. Describe, using appropriate sketches, the output frequency response of the following categories of filters (i) low pass (ii) high pass (iii) Band reject

4. For the filter circuit configuration shown in Fig. 4.20, find the lower cut-off frequency, upper cut-off frequency and the rate at which the voltage gain decreases outside the passband of the filter [15.9 Hz, 1.59 kH]

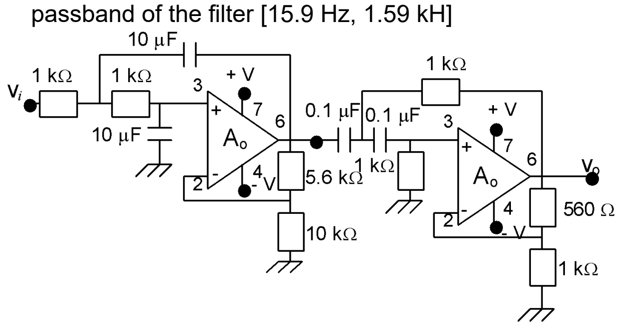


Fig 4.20 Passband filter

5. Design an active six-pole low-pass Butterworth filter having a cut-off frequency of 20 kHz.

TOPIC 5

DATA ACQUISITION AND CONVERSION SYSTEMS

5.0 INTRODUCTION

Typically a data acquisition system (DAS) consists of transducers, signal conditioning elements, multiplexers, data conversion, data processing, transmission and storage or display. Analogue data is acquired and converted to digital form for the purposes of processing, transmission, display and storage. A schematic block diagram of a generalised data acquisition system is given in Fig. 5.1.

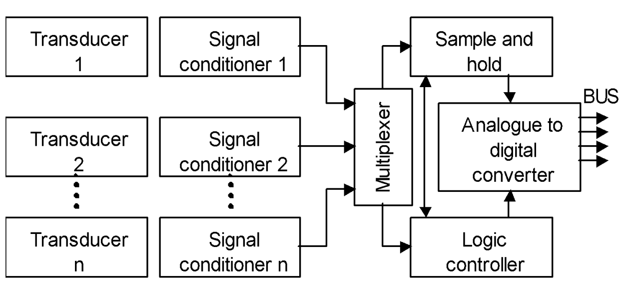


Fig. 5.1 Generalised data acquisition system

Factors determining the choice of a data acquisition configuration system

i. Resolution and accuracy requirements
ii. Number of analogue channels
iii. Sampling rate per channel
iv. Throughput
v. Signal conditioning requirements
vi. Cost factors

5.1 DATA ACQUISITION CONFIGURATIONS

Several possibilities exist under single and Multi-channel configurations. These include:

5.1.1 Single Channel Configurations

A single channel DAS consists of a transducer, signal conditioning device and an analogue to digital converter *(A/D)* that performs repetitive conversions at a pre-determined rate. Single channel configurations include: -

(i.) *Direct conversion*- does not use a pre-amplifier and a pre-conditioner

(ii.) *Pre-amplification and direct conversion*- makes use of an instrumentation amplifier to raise the signal level of the inputs to match the converter input requirements.

(iii.) *Sample and hold (S/H) and conversion*

(iv.) *Pre-amplification, S/H and conversion*

(v.) *Pre-amplification, signal conditioning* and any of the above

5.1.2 Multi-channel Configurations

Consist of a combination of more than one single channel DAS. *Multiplexers* are used to facilitate the sharing of the A/D between the outputs of the individual channels. When a large number of channels are to be monitored synchronously (i.e. at the same time) the outputs of the S/H devices in individual channels are multiplexed and linked to one A/D. Where several A/Ds are available each can be assigned to an individual channel and then their outputs multiplexed. This configuration provides additional advantage in industrial DASs where many strain gages, thermocouples and LVDTs are distributed over a large plant area. Since analogue signals are digitised at the source, the digital transmission of the data to the centre provides enhanced immunity against line frequency and other ground-loop interferences.

5.2 *ANALOGUE INPUT/OUTPUT FUNCTIONALITY*

Transducers generate electric analogue voltage or current signals that can be converted to digital form using an A/D in data transmission and digital signal processing. These digital outputs can be processed back into analogue form by a Digital to Analogue (D/ A) converter.

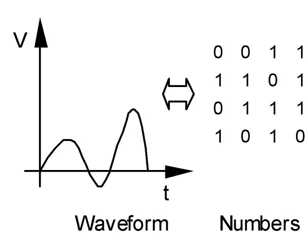


Figure 5.2. Analog-to-Digital Interface

Today, digital computers and other microprocessor - based devices have replaced analogue recording and display technologies in all but the simplest data acquisition applications. And while computers have had an undeniably positive impact on the practice of data acquisition, they speak only a binary language of ones and zeroes. Manufacturing processes and natural phenomena, however, are still by their very nature analogue. That is, natural processes tend to vary smoothly over time, not discontinuously changing states from black to white, from on to off.

To be meaningfully recorded or manipulated by a computer then, analogue measurements such as pressure, temperature, flow rate, and position must be translated into digital representations. Inherently digital events, too, such as the tripping of a motor or a pulse generated by a positive displacement flowmeter, must be made interpretable as a transistor-to-transistor logic (TTL) level changes in voltage, hence, the origination and ongoing development of input/output (I/O) systems for converting analogue and digital information about real-world processes and events into the language of computers.

Most sensors for measuring temperature, pressure, and other continuous variables provide a continuously varying electrical output to represent the magnitude of the variable in question. To make this signal interpretable by a microprocessor, it must be converted from a smooth continuous value to a discrete, digital number (Fig.5.2).

This analogue-to-digital (A/D) conversion process involves three key operations, namely:

- ✓ Sampling
- ✓ Quantisation
- ✓ Encoding

The first operation is performed by a sample and hold device; the second and third are combined in an analogue to digital converter.

5.2.1 Sampling and aliasing

A continuous signal $y(t)$ can be represented by a set of samples y_i, i = 1, 2, …, N, taken at discrete intervals of time. The ***Nyquist theorem*** defines the necessary relationship between the highest frequency contained in a signal and the minimum required sampling speed. Nyquist stated as follows:

> *A continuous signal can be represented by, and reconstituted from, a set of sample values providing that the number of samples per second is at least twice the highest frequency present in the signal.*

Stated mathematically: $f_s \geq 2f_{max}$, i.e. the sample frequency (f_s) must be at least twice the highest frequency component (f_{max}) contained within the input signal. So, to sample a 1Hz sine wave, the sample rate should be at least 2 Hz (but a rate of 8 -16 Hz would be much better for resolving the true shape of the wave).

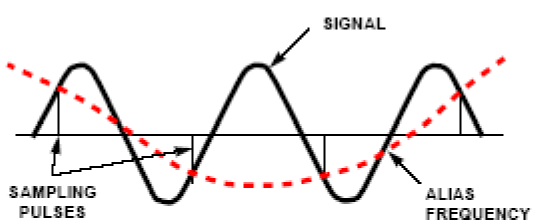


Fig. 5.3 Aliasing due to a slow sample rate

The primary implications of ignoring the Nyquist criterion include not only missing high frequency information but of introducing aliasing; if the sample rate is not fast enough, the presence of totally nonexistent frequencies may be indicated in the alias waveform (Figure 5.3). Here ***aliasing*** involves *reconstruction of the harmonic wave from the sampled values*. It is aliasing that makes a helicopter's rotors or a car's wheels appear to turn slowly backwards when seen in a movie. Low-pass, or anti-aliasing filters can be used to limit the measured waveform's frequency spectrum so that no detectable component equals or exceeds half of the sampling rate.

The sample and hold (S/H) device

Sample and holds are used in data acquisition systems to 'freeze' fast moving signals so that sampled level at a chosen instant can be converted to digital form.

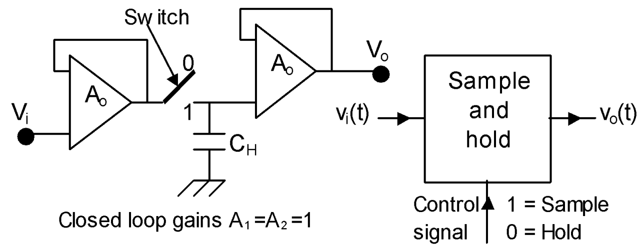


Fig 5.4 Realisation of a sample and hold circuit using op –amps.

Switch closed: When the switch is closed (position '1') the output voltage, V_o, follows (or tracks) the input voltage, V_i, and the capacitor is charged to the value V_i. This is called the *tracking* or *sampling* mode. $V_o = A_I A_2 V_i = V_i$

Switch open: The switch is in position' 0' .At the instant the switch opens the capacitor holds the last value of V_i. This is called the *hold* mode.

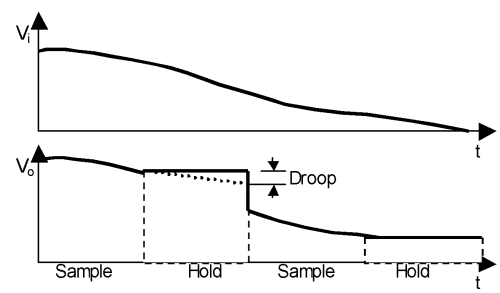


Fig 5.5 Input/output voltage transfer of a Sample and Hold circuit

The sample and hold waveform shown is ideal. In practice errors can occur due to the finite time for transition between the sample and hold states (called *aperture time*) and the reduction in the hold signal (*droop*) as the capacitor discharges over time.

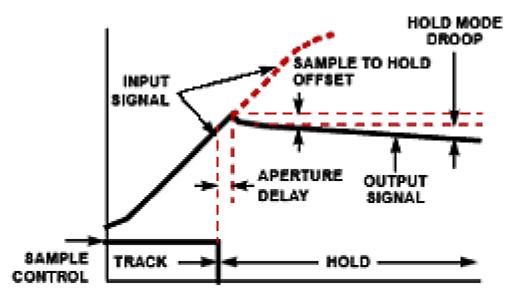


Fig. 5.6. Sinusoidal input parameters

Acquisition and aperture time

The ***acquisition time*** is the time taken by a S/H circuit to acquire the input signal within a stated accuracy.

The ***aperture time*** (t_a) is the time delay between the transition of the command logic from 'sample' to 'hold' and the time that the output voltage stabilises in the 'hold' position.

Consider a sinusoidal analogue input signal (V_i) of amplitude E occurring at the input terminals of the S/H device (Fig. 5.6).

$V_i = E \sin \omega t$

$$= \frac{V_{FS}}{2}\sin\omega t$$, where V_{FS} is the range.

The quantisation interval

ΔV = maximum conversion rate x aperture time

$$= \frac{d}{dt}\left[\frac{V_{FS}}{2}\sin\omega t\right]_{t\to 0} \times t_a$$

$$= \frac{V_{FS}\omega t_a}{2}$$

>(5.1)

Making t_a the subject, with $\omega = 2\pi f$

$$t_a = \frac{\Delta E}{V_{FS}}\frac{1}{\pi f} = \frac{1}{R\pi f}$$

>(5.2)

where R = resolution of the converter

$$= \frac{V_{FS}}{\Delta E}$$

>(5.3)

Droop rate

The *droop rate* is the rate of change of the output voltage ($V_o = V_h$) when the S/H device is in the 'hold' mode. It is a function of the leakage/droop current (I_d) and the capacitance of hold capacitor (C_h)

$$\text{Droop rate} = \frac{dV_h}{dt} = \frac{I_d}{C_h} \quad [\text{V}/\mu\text{s}]$$

>(5.4)

where V_h = voltage developed across the hold capacitor

$$= \frac{1}{C_h}\int i_c\, dt$$

>(5.5)

with i_c = amplifier charging current.

5.2.2 Quantisation

Quantisation is the process of transforming a continuous analog signal into a set of discrete output states. Measurement transducers or transmitters typically provide continuously varying signals between 0-10 V dc, ±5 V dc, 0-100 mV dc, or 4-20 mA dc. Thermocouples and resistance temperature devices (RTDs) are other common low voltage inputs. When this analogue value is represented as a digital number, however, this essentially continuous resolution is limited to *discrete steps*. This resolution of an A/D conversion often is stated in terms of bits-***the more bits the finer the resolution***. The number of bits determines the number of divisions into which a full-scale input range can be divided (i.e. *quantised*) to approximate an analog input voltage.

In quantisation sample voltages are rounded either up or down to one of the quantisation levels V_q, where q = 0, 1, 2, …, Q -1, (Fig. 5.7(a)). These quantum levels correspond to Q digital numbers 0, 1, 2,…, Q-1. The spacing width or ***quantisation interval* ΔV** is determined as:

$$\Delta V = \frac{V_{FS}}{2^n}$$

>(5.6)

where V_{FS} = Full scale voltage range = $V_{max} - V_{min}$

The ***resolution R*** of the converter is therefore:

$$R = \frac{V_{FS}}{\Delta V} = 2^n$$

>(5.7)

For example, the nonlinear transfer function shown in Fig 5.7(b) is that of an ideal quantiser with 8 output states; with output code words assigned, it is also that of a 3-bit A/D converter. The 8 output states are assigned the sequence of binary numbers from 000 through 111. The analog input range for this quantiser is 0 to +10V. There are several important points concerning the transfer function of Fig 5.7(b). First, the resolution of the quantiser is defined as the number of output states expressed in bits; in this case it is a 3-bit quantiser. The number of output states for a binary coded quantiser is 2^n, where n is the number of bits. Thus, an 8-bit quantiser has 256 output states and a 12-bit quantiser has 4096 output states.

As shown in the diagram, there are 2^n-1 analogue decision points (or threshold levels) in the transfer function. These points are at voltages of +0.625, +1.875, +3.125, +4.375, +5.625, +6.875, and +8.125. The decision points must be precisely set in a quantiser in order to divide the analogue voltage range into the correct quantised values. The voltages +1.25, +2.50, +3.75, +5.00, +6.25, +7.50, and +8.75 are the centre points of each output code word. The analogue decision point voltages are precisely halfway between the code word centre points. The quantiser staircase function is the best approximation which can be made to a straight line drawn through the origin and full scale point; notice that the line passes through all of the code word centre points.

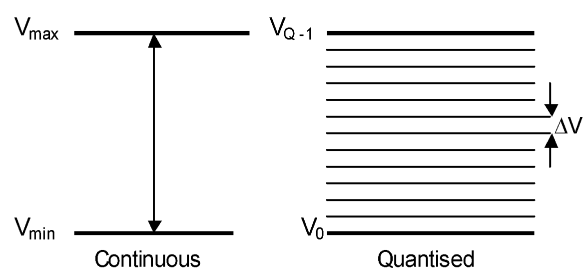


Fig. 5.7(a): Quantisation

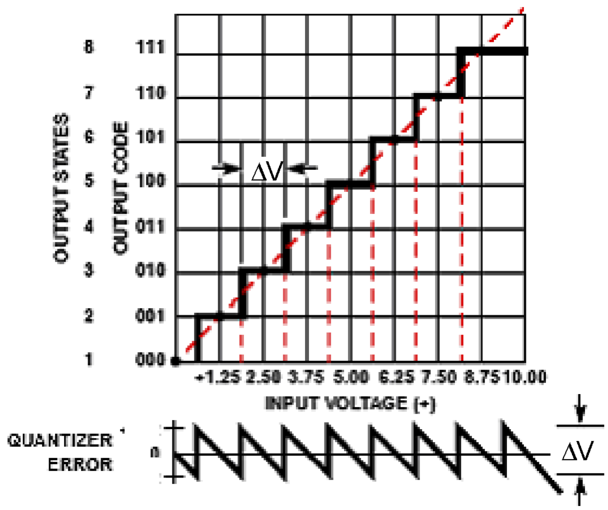


Fig. 5.7(b) Transfer function of an ideal 3-bit quantiser

The maximum ***quantisation error, $e_{q,\ max}$*** is $\pm$ ΔV/2. Expressed as a percentage of the range

$$e_{q,\ max} = \pm\frac{\Delta V}{2} = \frac{V_{FS}}{2^{n+1}} \qquad (5.8)$$

where R is the resolution of the converter. From (5.8) it can be deduced that the quantisation error reduces with increasing resolution. Designing or specifying a device for A/D conversion consists of a series of trade-offs. More resolution (more bits) means more accurate conversion but more expensive hardware. Similarly, slower sample rates mean cheaper A/D conversion, but the Nyquist criterion must still be satisfied.

5.2.3 Encoding

An encoder converts the quantisation values V_q into a parallel digital signal corresponding to a binary coded version of the digital numbers 0, 1, 2,…, Q-1.
A/D and D/A converters interface with digital systems by means of an appropriate digital code. While there are many possible codes to select, a few standard ones are almost exclusively used with data converters.

5.2.3.1 Natural binary code

The most popular code is natural binary, or straight binary, which is used in its fractional form to represent a number

Most significant bit (MSB) — Least significant bit (MSB)

$$b_n 2^n + b_{n-1} 2^{n-1} + \ldots + b_i 2^i + \ldots + b_1 2^1 + b_0 2^0 \qquad (5.9)$$

where b_0, b_1,…, b_n are called ***bits*** or ***digits***. A bit can take either of the values 0 or 1 corresponding to the logic levels LOW or HIGH, FALSE or TRUE, ON or OFF.

For example to convert the 8-bit binary number 10110111 to decimal proceed as follows:

$b_0 = b_1 = b_2 = b_4 = b_5 = b_7 = 1$;
$b_3 = b_6 = 0$

$\Rightarrow$ 10110111_{bin}
$=(1)2^7+(0)2^6+(1)2^5+(1)2^4+(0)2^3+(1)2^2+(1)2^1+(1)2^0$
$= 183$

The number of binary digits **n** required to encode Q decimal numbers is given by:

$$Q = 2^n \qquad (5.10)$$

Fractional designation of the natural binary code

A binary fraction is normally written as 0.110101, but with data converter codes the decimal point is omitted and the code word is written 110101. This code word represents a fraction of the full scale value of the converter and has no other numerical significance. The binary code word 110101 therefore represents the decimal fraction

$$1(2^{-1})+1(2^{-2})+0(2^{-3})+1(2^{-4})+0(2^{-5})+1(2^{-6}) = 0.828125$$

or 82.8125% of full scale for the converter. If full scale is +10V, then the code word represents +8.28125V. The natural binary code belongs to a class of codes known as positive weighted codes since each coefficient has a specific weight, none of which is negative. The leftmost bit has the most weight, 0.5 of full scale, and is called the most significant bit, or MSB; the rightmost bit has the least weight, 2^{-n} of full scale, and is therefore called the least significant bit, or LSB. The bits in a code word are numbered from left to right from 1 to n. The LSB has the same analog equivalent value as ΔV discussed previously,

namely; LSB (analogue value) = $\dfrac{V_{FS}}{2^n}$

The dynamic range of a data converter in dB is given by:

$$DR = 20 \log 2^n = 6.02n$$

where DR is a dynamic range, n is the number of bits, and 2^n the number of states of the converter. Since 6.02dB corresponds to a factor of two, it is simply necessary to multiply the resolution of a converter in bits by 6.02. A 12-bit converter, for example, has a dynamic range of 72.2dB.

An important point to notice is that the maximum value of the digital code, namely all 1's does not correspond with analog full scale, but rather with one LSB less than full scale, or $F_S = (1-2^{-n})$. Therefore a 12-bit converter with a 0 to 10V analog range has a maximum code of 1111 1111 1111 and a maximum analog value of $+10V(1-2^{-12}) = +9.99756V$. In other words, the maximum analog value of the converter, corresponding to all one's in the code, never quite reaches the point defined as analog full scale.

5.2.3.2 Other binary codes

Several other binary codes are used with A/D and D/A converters in addition to straight binary. These codes are offset binary, two's complement, binary coded decimal (BCD), and their complemented versions. Each code has a specific advantage in certain applications. BCD coding for example is used where digital displays must be interfaced such as in digital panel meters and digital multimeters. Two's complement coding is used for computer arithmetic logic operations, and offset binary coding is used with bipolar analog measurements.

Not only are the digital codes standardised with data converters, but so are the analog voltage ranges. Most converters use unipolar voltage ranges of 0 to +5V and 0 to +10V although some devices use the negative ranges 0 to -5V and 0 to -10V. The standard bipolar voltage ranges are ±2.5V, ±5V and ±10V. Many converters today are pin programmable between these various ranges.

Table 5.1 Binary coding for 8-bit unipolar converters

FRACTION OF f_S	+ 10V f_S	STRAIGHT BINARY	COMPLEMENTARY BINARY
$+f_S$-1 LSB	+9.961	1111 1111	0000 0000
$+3/4f_S$	+7.500	1100 0000	0011 1111
$+1/2f_S$	+5.000	1000 0000	0111 1111
$+1/4f_S$	+2.500	0100 0000	1011 1111
$+1/8f_S$	+1.250	0010 0000	1101 1111
+1 LSB	+0.039	0000 0001	1111 1110
0	0.000	0000 0000	1111 1111

Table 5.1 shows straight binary and complementary binary codes for unipolar 8-bit converter with a 0 to +10V analog full scale range. The maximum analog value of the converter is +9.961V, or one LSB less than +10V. Note that the LSB size is 0.039V as shown near the bottom of the table. The complementary binary coding used in some converters is simply the logic complement of straight binary. When A/D and D/A converters are used in bipolar operation, the analog range is offset by half scale, or by the MSB value. The result is an analog shift of the converter transfer function as shown in Fig. 5.8.

Notice for this 3-bit A/D converter transfer function that the code 000 corresponds with -5V, 100 with 0V, and 111 with +3.75V. Since the output coding is the same as before the analog shift, it is now appropriately called offset binary coding.

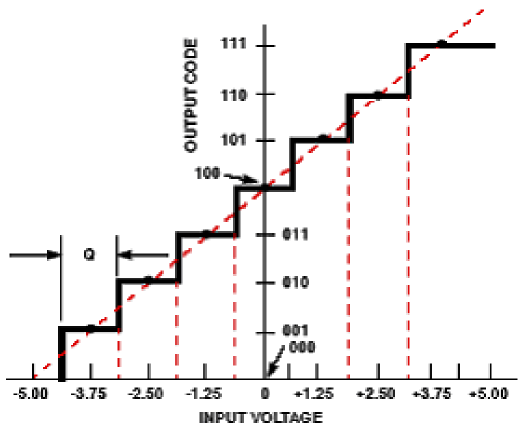


Fig 5.8 Transfer function of a bipolar 3-bit A/D converter

Table 5.2 shows the offset binary code together with complementary offset binary, two's complement, and sign magnitude binary codes. These are the most popular codes employed in bipolar data converters.

The two's complement code has the characteristic that the sum of the positive and negative codes for the same analog magnitude always produces all zero's and a carry. This characteristic makes the two's complement code useful in arithmetic computations. Notice that the only difference between two's complement and offset binary is the complementing of the MSB. In bipolar coding, the MSB becomes the sign bit. The sign-magnitude binary code, infrequently used, has identical code words for equal magnitude analog values except that the sign bit is different. As shown in Table 3 this code has two possible code words for zero: 1000 0000 or 0000 0000. The two are usually distinguished as 0+ and 0-, respectively. Because of this characteristic, the code has maximum analog values of $\pm$ (FS-1 LSB) and reaches neither analog +FS or -FS.

Binary coded decimal

Table 5.3 shows BCD and complementary BCD coding for a 3 decimal digit data converter. These are the codes used with integrating type A/D converters employed in digital panel meters, digital multimeters, and other decimal display applications.

Here four bits are used to represent each decimal digit.

BCD is a positive weighted code but is relatively inefficient since in each group of four bits, only 10 out of a possible 16 states are utilised.

Table 5.3 BCD and complementary BCD coding

FRACTION OF f_S	+10V f_S	BINARY CODED DECIMAL	COMPLEMENTARY BCD
$+f_S$-1 LSB	+9.99	1001 1001 1001	0110 0110 0110
$+3/4f_S$	+7.50	0111 0101 0000	1000 1010 1111
$+1/2f_S$	+5.00	0101 0000 0000	1010 1111 1111
$+1/4f_S$	+2.50	0010 0101 0000	1101 1010 1111
$+1/8f_S$	+1.25	0001 0010 0101	1110 1101 1010
+1 LSB	+0.01	0000 0000 0001	1111 1111 1110
0	0.00	0000 0000 0000	1111 1111 1111

Each decade of the decimal number is separately coded into decimal. Four binary digits DCBA are used to encode the 10 numbers 0 to 9 in each decade. In 8:4:2:1 BCD; A $=2^0$, B$=2^1$, C$= 2^2$ and D$=2^3$.

The decimal number 75 is denoted in straight BCD as:

$75_{dec} = 0111\ 0101_{bcd}$

The number of digits **n** in BCD required to encode Q decimal numbers is:

$n = 4 n_d$

$= 4\log_{10}Q$

where n_d = number of decades in the signal.

$= \log_{10}Q$

The signal to alphanumeric displays is usually in BCD form.

The LSB analog equivalent value in this case is:

$$\text{LSB (analogue value)} = \frac{V_{FS}}{10^d}$$

Where d is the number of decimal digits. For example if there are 3 digits and the full scale range is 10V, the LSB value is $10/10^3$ = 10 mV.

BCD coding is frequently used with an additional overrange bit which has a weight equal to full scale and produces a 100% increase in range for the A/D converter. Thus for a converter with a decimal full scale of 999, an overrange bit provides a new full scale of 1999, twice that of the previous one. In this case, the maximum output code is 1 1001 1001 1001. The additional range is commonly referred to as ½ digit, and the resolution of the A/D converter in this case is 3½ digits. Likewise, if this range is again expanded by 100%, a new full scale of 3999 results and is called 3¾ digits resolution. Here two overrange bits have been added and the full scale output code is 11 1001 1001 1001. When BCD coding is used for bipolar measurements another bit, a sign bit, is added to the code and result is sign-magnitude BCD coding.

5.3 DIGITAL TO ANALOGUE CONVERTERS (DACs)

A *D/* A converter changes a digital signal into an analogue signal. Analog outputs commonly are used to operate valves and motors in industrial environments and to generate inputs for electronic devices under test. As with most analog input configurations, a common D/A converter often is shared among multiplexed output signals. Standard analog output ranges are essentially the same as analog inputs: ±5 V dc, ±10 V dc,0-10 V dc, and 4-20 mA dc.

The input signal is represented by digital word consisting of **n** parallel bits, which are defined at any time t, by D, where:

$$D = (b_{n-1}b_{n-2}\ldots b_1b_0)$$

The parallel bits drop in (or drop out, depending on whether the bit is 1 or 0) a series of resistors from a circuit driven by a reference voltage. This ladder of resistors can be made of either

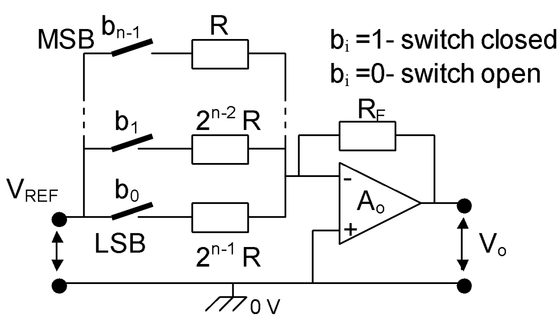
- weighted value resistors, Fig 5.9 (a) or
- a R-2R network, Figure 5.9 (c).

The **resolution** in both cases = Q – 1, with Q = 2^n.

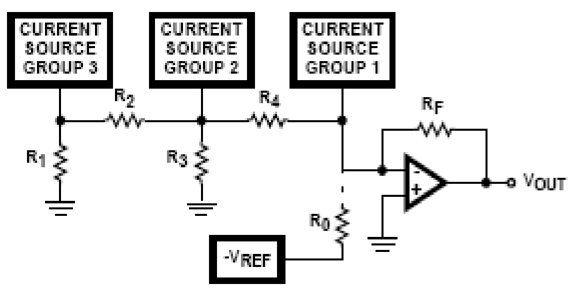
Table 5.2 Popular bipolar codes used with data converters

FRACTION OF f_S	±5V f_S	OFFSET BINARY	COMP OFF BINARY	TWO'S COMPLEMENT	SIGN-MAG BINARY
$+f_S$-1 LSB	+4.9978	1111 1111	0000 0000	0111 1111	1111 1111
+1/4f_S	+3.7500	1110 0000	0001 1111	0110 0000	1110 0000
+1/2f_S	+2.5000	1100 0000	0011 1111	0100 0000	1100 0000
+3/4f_S	+1.2500	1010 0000	0101 1111	0010 0000	1010 0000
0	0.0000	1000 0000	0111 1111	0000 0000	1000 0000
-1/4f_S	-1.2500	0110 0000	1001 1111	1110 0000	0010 0000
-1/2f_S	-2.5000	0100 0000	1011 1111	1100 0000	0100 0000
-3/4f_S	-3.7500	0010 0000	1101 1111	1010 0000	0110 0000
$-f_S$+1 LSB	-4.9978	0000 0001	1111 1110	1000 0001	0111 1111
$-f_S$	-5.0000	0000 0000	1111 1111	1000 0000	--

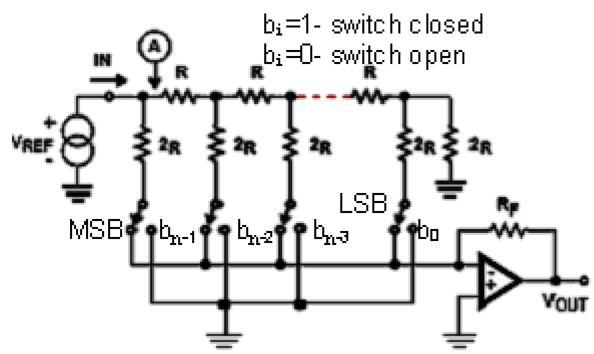
NOTE: Sign Magnitude Binary has two code words for zero as shown here.



(a) Weighted resistors



(b) current dividing the outputs of weighted current source groups



(c) R –2R ladder network

Figure 5.9: Digital to analogue conversion

5.3.1 Weighted Resistor/Current Source D/A Converter

The most popular D/A converter design in use today is the weighted current source circuit illustrated in Figure 5.9(a). An array of switched transistor current sources is used with binary weighted currents. The binary weighting is achieved by using resistors with binary related values of R, 2R, 4R, 8R, ...,2^nR. The resulting collector currents are then added together at the current summing line. The current sources are switched on or off from standard TTL inputs by means of the control diodes connected to each resistor.

When the TTL input is high the current source is on; when the input is low it is off, with the current flowing through the control diode. Fast switching speed is achieved because there is direct control of the transistor current, and the current sources never go into saturation. The summed output currents from all current sources that are on go to an operational amplifier summing junction; the amplifier converts this output current into an output voltage.

Transfer characteristics

For the weighted resistor network Fig 5.9. (a)

$$V_o = -V_{REF} \frac{R_F}{R} \sum_{i=0}^{n-1} b_i 2^i$$

>(5.11)

The weighted current source design has the advantages of simplicity and high speed. Both PNP and NPN transistor current sources can be used with this technique although the TTL interfacing is more difficult with NPN sources. This technique is used in most monolithic, hybrid, and modular D/A converters in use today.

A difficulty in implementing higher resolution D/A converters designs is that a wide range of emitter resistors are required, and very high value resistors cause problems with both temperature stability and switching speed. To overcome these problems, weighted current sources are used in identical groups, with the output of each group divided down by a resistor divider as shown in Fig. 5.9 (b).

The resistor network, R_1 through R_4, divides the output of Group 3 down by a factor of 256 and the output of Group 2 down by a factor of 16 with respect to the output of Group 1. Each group is identical, with four current sources of the type shown in Fig. 5.9(a), having binary current weights of 1, 2, 4, 8. Fig. 5.9(b) also illustrates the method of achieving a bipolar output by deriving an offset current from the reference circuit which is then subtracted from the output current line through resistor R_0. This current is set to exactly one half the full scale output current.

5.3.2 R – 2R ladder network D/A Converter

The operation of the R-2R ladder network is based on the binary division of current as it flows down the ladder. Examination of the ladder configuration reveals that at point A looking to the right, one measures a resistance of 2R; therefore the reference input to the ladder has a resistance of R. At the reference input the current splits into two equal parts since it sees equal resistances in either direction. Likewise, the current flowing down the ladder to the right continues to divide into two equal parts at each resistor junction. The result is binary weighted currents flowing down each shunt resistor in the ladder. The digitally controlled switches direct the currents to either the summing line or ground.

The R-2R scheme is more practical. Because only one resistor value need be used, it is easier to match the temperature coefficients of an R-2R ladder than a weighted network, resulting in more accurate outputs. Plus, for high-resolution outputs, very high resistor values are needed in the weighted-resistor approach.

Transfer characteristics

For the R-2R ladder network Fig 5.9. (c)

$$V_o = -i_o R = -\frac{V_{REF}}{2} \sum_{k=0}^{n-1} \frac{b_{[(n-1)-k]}}{2^k}$$

>(5.12)

Key specifications of an analog output include:

***Settling time**:* Period required for a D/A converter to respond to a full-scale setpoint change.

Linearity: This refers to the device's ability to accurately divide the reference voltage into evenly sized increments.

Range: The reference voltage sets the limit on the output voltage achievable.

Because most unconditioned analog outputs are limited to 5 mA of current, amplifiers and signal conditioners often are needed to drive a final control element. A low-pass filter may also be used to smooth out the discrete steps in output.

5.3.3 Multiplying D/A converters

The R-2R ladder method is specifically used for multiplying type D/A converters. With these converters, the reference voltage can be varied over the full range of $\pm V_{MAX}$ with the output the product of the reference voltage and the digital input word. Multiplication can be performed in 1, 2, or 4 algebraic quadrants. If the reference voltage is unipolar, the circuit is a one-quadrant multiplying DAC; if it is bipolar, the circuit is a two-quadrant multiplying DAC. For four-quadrant operation the two current summing lines shown in Fig. 5.9 (b) must be subtracted from each other by operational amplifiers. In multiplying D/A converters, the electronic switches are usually implemented with CMOS (i.e. complementary metal oxide semiconductor) devices. Multiplying DACs are commonly used in automatic gain controls, CRT character generation, complex function generators, digital attenuators, and divider circuits.

Another important D/A converter design takes advantage of the best features of both the weighted current source technique and the R-2R ladder technique. This circuit, shown in Fig. 5.10, uses equal value switched current sources to drive the junctions of the R-2R ladder network. The advantage of the equal value current sources is obvious since all emitter resistors are identical and switching speeds are also identical. This technique is used in many ultra-high speed D/A converters.

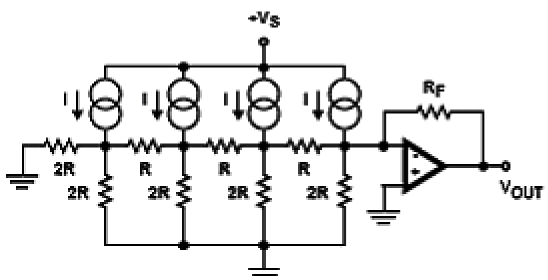


Fig 5.10 D/A converter employing R-2R ladder with equal value switched current sources

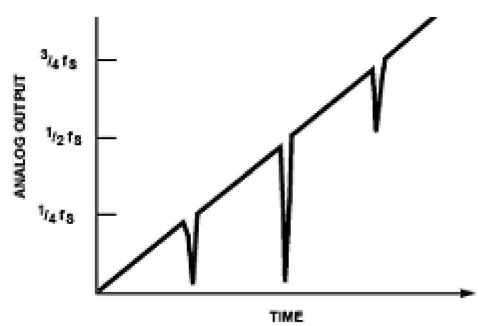


Fig 5.11 (a) Output glitches

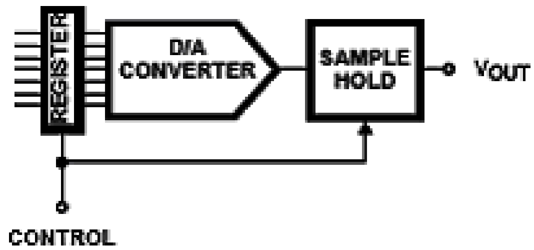


Fig 5.11 (b) Deglitched D/A converter

5.3.4 Deglitched D/A Converters

Glitches are undesired transient spikes that are observed in the output of DACs. They are caused by small time differences between some current sources turning off and others turning on. Take, for example, the major code transition at half scale from 0111 ...1111 to 1000...0000. Here the MSB current source turns on while all other current sources turn off. The small differences in switching times results in a narrow half scale glitch. Such a glitch produces distorted characters on CRT displays.
Glitches can be virtually eliminated by the circuit shown in Figure 5.11(b). The digital input to a D/A converter is controlled by an input register while the converter output goes to a specially designed sample-hold circuit. When the digital input is updated by the register, the sample-hold is switched into the hold mode. After the D/A has changed to its new output value and all glitches have settled out, the sample-hold is then switched back into the tracking mode. When this happens the output changes smoothly from its previous value to the new value with no glitches present.

5.4 ANALOGUE TO DIGITAL CONVERTERS (ADCs)

Continuous electrical signals are converted to the digital language of computers using analog-to-digital (A/D) converters. An A/D converter may be housed on a PC board with associated circuitry or in a variety of remote or networked configurations. In addition to the converter itself, sample-and-hold circuits, an amplifier, a multiplexer, timing and synchronization circuits, and signal conditioning elements also may be on board (Figure 5.1). The logic circuits necessary to control the transfer of data to computer memory or to an internal register also are needed.

When determining what type of A/D converter should be used in a given application, performance should be closely matched to the requirements of the analog input transducer(s) in question. Accuracy, signal frequency content, maximum signal level, and dynamic range all should be considered.

Central to the performance of an A/D converter is its resolution, often expressed in bits. An A/D converter essentially divides the analog input range into 2^n bins, where **n** is the number of bits. In other words, resolution is a measure of the number of levels used to represent the analog input range and determines the converter's sensitivity to a change in analog input. This is not to be confused with its absolute accuracy! Amplification of the signal, or input gain, can be used to increase the apparent sensitivity if the signal's expected maximum range is less than the input range of the A/D converter. Because higher resolution A/D converters cost more, it is especially important to not buy more resolution than you need-if you have 1% accurate (1 in 100) temperature transducers, a 16-bit (1 in 65,536) A/D converter is probably more resolution than you need.

Absolute accuracy of the A/D conversion is a function of the reference voltage stability (the known voltage to which the unknown voltage is compared) as well as the comparator performance. Overall, it is of limited use to know the accuracy of the A/D converter itself. Accuracy of the system, together with associated multiplexer, amplifier, and other circuitry is typically more meaningful.
The other primary A/D converter performance parameter that must be considered is speed-throughput for a multi-channel device.

Overall, *system speed* depends on the *conversion* time, *acquisition* time, *transfer* time, and the number of channels being served by the system: Acquisition is the time needed by the front-end analog circuitry to acquire a signal. Also called *aperture time*, it is the time for which the converter must see the analog voltage in order to complete a conversion. Conversion is the time needed to produce a digital value corresponding to the analog value. Transfer is the time needed to send the digital value to the host computer's memory. Throughput, then, equals the number of channels being served divided by the time required to do all three functions.

A/D Converter Options

While all analog-to-digital converters are classified by their resolution or number of bits, how the A/D circuitry achieves this resolution varies from device to device. There are four primary types of A/D converters used for industrial and laboratory applications

- successive approximation
- flash/parallel
- integrating, and
- ramp/counting.

Some are optimized for speed, others for economy, and others for a compromise among competing priorities (Table 5.2.). Industrial and lab data acquisition tasks typically require 12 to 16 bits -12 is the most common. As a rule, increasing resolution results in higher costs and slower conversion speed.

Table 5.2: Alternative A/D Converter Designs

DESIGN	SPEED	RESOLUTION	NOISE IMMUNITY	COST
Successive approximation	Medium	10-16 bits	Poor	Low
Integrating	Slow	12-18 bits	Good	Low
Ramp/counting	Slow	14-24 bits	Good	Medium
Flash/parallel	Fast	4-8 bits	None	High

5.4.1 *Successive approximation ADC*

By far, the most popular A/D conversion technique in general use for moderate to high speed applications is the successive-approximation type A/D. This method falls into a class of techniques known as feedback type A/D converters, to which the counter type also belongs. In both cases a D/A converter is in the feedback loop of a digital control circuit which changes its output until it equals the analog input. In the case of the successive-approximation converter, the DAC is controlled in an optimum manner to complete a conversion in just n-steps, where n is the resolution of the converter in bits.

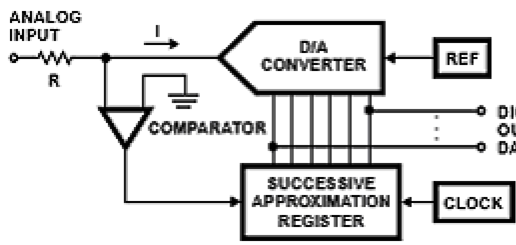


Fig 5.12 Successive Approximation ADC

The operation of this converter is analogous to weighing an unknown on a laboratory balance scale using standard weights in a binary sequence such as 1, 1/2, 1/4, 1/8 ...1/n kilograms. The correct procedure is to begin with the largest standard weight and proceed in order down to the smallest one. The largest weight is placed on the balance pan first; if it does not tip, the weight is left on and the next largest weight is added. If the balance does tip, the weight is removed and the next one added. The same procedure is used for the next largest weight and so on down to the smallest. After the n^{th} standard weight has been tried and a decision made, the weighing is finished. The total of the standard weights remaining on the balance is the closest possible approximation to the unknown.

In the successive-approximation A/D converter illustrated in Fig.5.12, a successive-approximation register (SAR) controls the D/A converter by implementing the weighing logic just described. The SAR first turns on the MSB of the DAC and the comparator tests this output against the analog input. A decision is made by the comparator to leave the bit on or turn it off after which bit 2 is turned on and a second comparison made. After n-comparisons the digital output of the SAR indicates all those bits which remain on and produces the desired digital code. The clock circuit controls the timing of the SAR.

The conversion efficiency of this technique means that high resolution conversions can be made in very short times. For example, it is possible to perform a 10-bit conversion in 1μs or less and a 12-bit conversion in 2μs or less. Of course the speed of the internal circuitry, in particular the D/A and comparator, are critical for high speed performance.

5.4.2 *Flash/parallel ADC*

When higher speed operation is required, parallel, or flash-type A/D conversion is called for. This design uses multiple comparators in parallel to process samples at more than 100 MHz with 8 to 12-bit resolution. Conversion is accomplished by a string of comparators with appropriate references operating in parallel (Figure 5.13).

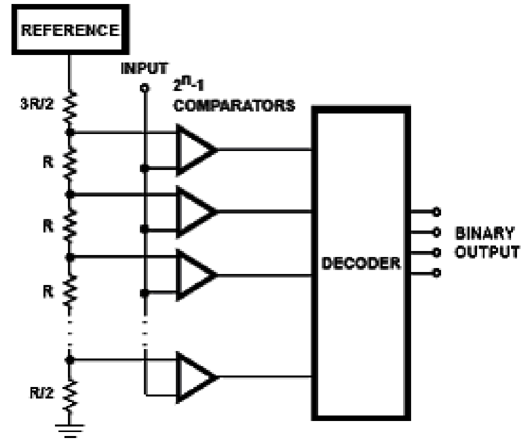


Fig. 5.13 Flash /parallel ADC

This circuit employs 2^{n-1} analog comparators to directly implement the quantiser transfer function of an A/D converter. The comparator trip-points are spaced 1 LSB apart by the series resistor chain and voltage reference. For a given analog input voltage all comparators biased below the voltage turn on and all those biased above it remain off. Since all comparators change state simultaneously, the quantisation process is a one-step operation. A second step is required, however, since the logic output of the comparators is not in binary form.

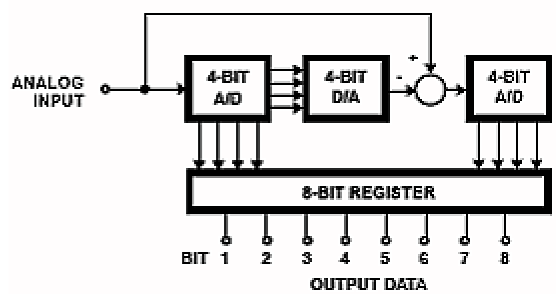


Fig. 5.14 Two-stage parallel 8-bit a/d converter

Therefore an ultra-fast decoder circuit is employed to make the logic conversion to binary. The parallel technique reaches the ultimate in high speed because only two sequential operations are required to make the conversion. The limitation of the method, however, is in the large umber of comparators required for even moderate resolutions. A 4 - bit converter, for example, requires only 15 comparators, but an 8-bit converter needs 255. For this reason it is common practice to implement an 8-bit A/D with two 4-bit stages as shown in Figure 5.14.

The result of the first 4-bit conversion is converted back to analog by means of an ultra-fast 4-bit D/A and then subtracted from the analog input. The resulting residue is then converted by the second 4-bit A/D, and the two sets of data are accumulated in the 8-bit output register. Converters of this type achieve 8-bit conversions at rates of 20 MHz and higher, while single stage 4-bit conversions can reach 50 to 100 MHz rates.

5.4.3 *Integrating ADC*

Another class of A/D converters known as integrating type operates by an indirect conversion method. The unknown input voltage is converted into a time period which is then measured by a clock and counter. A number of variations exist on the basic principle such as single-slope, dual-slope, and triple-slope methods. In addition there is another technique - completely different - which is known as the charge-balancing or quantized feedback method.

The most popular of these methods are dual-slope and charge-balancing; although both are slow, they have excellent linearity characteristics with the capability of rejecting input noise. Because of these characteristics, integrating type A/D converters are almost exclusively used in digital panel meters, digital multimeters, and other slow measurement applications.

5.4.3.1 Dual-Slope ADC

The dual-slope technique, shown in Fig. 5.15, is perhaps best known. Conversion begins when the unknown input voltage is switched to the integrator input; at the same time the counter begins to count clock pulses and counts up to overflow. At this point the control circuit switches the integrator to the negative reference voltage which is integrated until the output is back to zero. Clock pulses are counted during this time until the comparator detects the zero crossing and turns them off. The counter output is then the converted digital word.

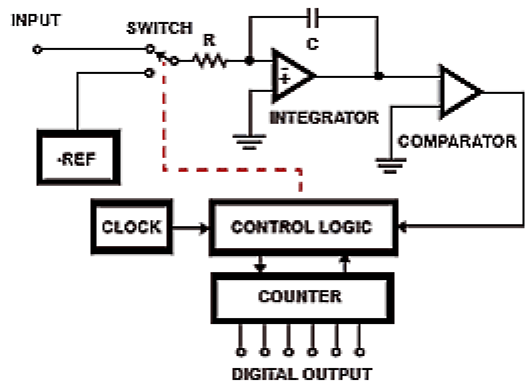


Fig. 5.15 Dual slope ADC

Fig.5.16 shows the integrator output waveform where T_1 is a fixed time and T_2 is a time proportional to the input voltage. The times are related as follows:

$$T_2 = T_1 \frac{E_{IN}}{\mathrm{V}_{\mathrm{REF}}}$$

> (5.13)

The digital output word therefore represents the ratio of the input voltage to the reference.

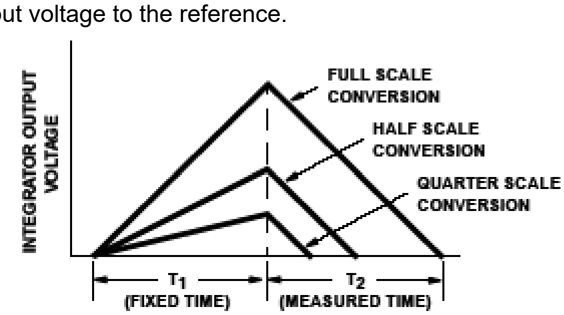


Fig. 5.16 Integrator output waveform for dual slope ADC

Dual-slope conversion has several important features. First, conversion accuracy is independent of the stability of the clock and integrating capacitor so long as they are constant during the conversion period. Accuracy depends only on the reference accuracy and the integrator circuit linearity. Second, the noise rejection of the converter can be infinite if T_1 is set to equal the period of the noise. To reject 60 Hz power noise therefore requires that T_1 be 16.667 ms.

5.4.3.2 Charge-Balancing A/D Conversion

The charge-balancing or quantized feedback, method of conversion is based on the principle of generating a pulse train with frequency proportional to the input voltage and then counting the pulses for a fixed period of time. This circuit is shown in Figure 5.17. Except for the counter and timer, the circuit is a voltage-to-frequency (V/F) converter which generates an output pulse rate proportional to input voltage.

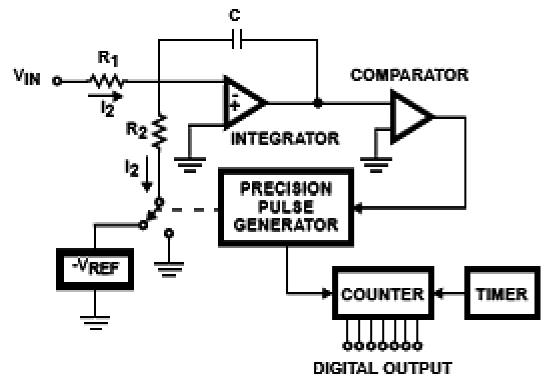


Fig. 5.17 Charge balancing ADC

The circuit operates as follows. A positive input voltage causes a current to flow into the operational integrator through R_1. This current is integrated, producing a negative going ramp at the output. Each time the ramp crosses zero the comparator output triggers a precision pulse generator which puts out a constant width pulse. The pulse output controls switch S_1 which connects R_2 to the negative reference for the duration of the pulse. During this time a pulse of current flows out of the integrator summing junction, producing a fast, positive ramp at the integrator output. This process is repeated, generating a train of current pulses which exactly balances the input current - hence the name charge balancing. This balance has the following relationship:

$$f = \frac{1}{\tau} \frac{V_{IN}}{V_{REF}} \frac{R_2}{R_1}$$

> (5.14)

where τ is the pulse width and f the frequency. A higher input voltage therefore causes the integrator to ramp up and down faster, producing higher frequency output pulses. The timer circuit sets a fixed time period for counting. Like the dual-slope converter, the circuit also integrates input noise, and if the timer is synchronized with the noise frequency, infinite rejection results. Figure 5.18 shows the noise rejection characteristic of all integrating type A/D converters with rejection plotted against the ratio of integration period to noise period.

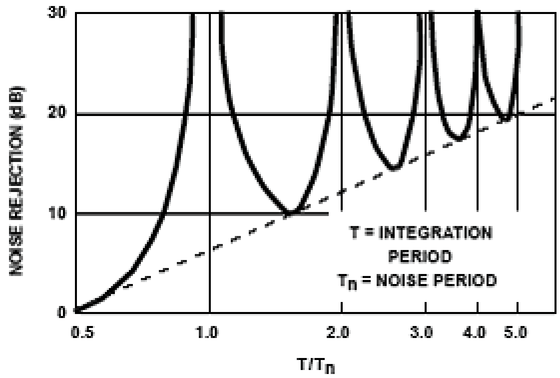


Fig. 5.18 Noise rejection for integrating type ADCs

5.4.4 *Counter and servo type ADCs*

As in the case of successive approximation designs, counting or ramp-type A/D converters, Fig 5.19 (a), use one comparator circuit and a D/A converter. This circuit employs a digital counter to control the input of a D/A converter. Clock pulses are applied to the counter and the output of the D/A is stepped up one LSB at a time. A comparator compares the DAC output with the analog input and stops the clock pulses when they are equal. The counter output is then the converted digital word. While this converter is simple, it is also relatively slow.

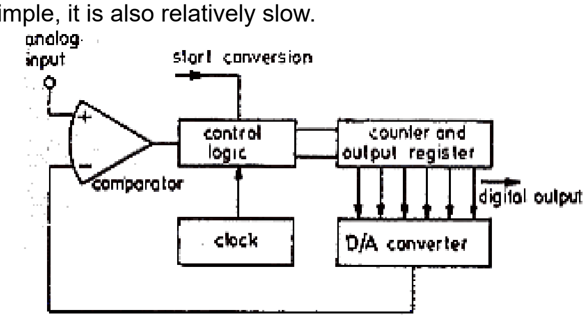


Figure 5.19 (a) Counter type ADC

A variation of this technique is the tracking / servo type ADC Fig. 5.19 (b), wherein, using an up-down counter controlling a D/A converter, the D/A converter output is forced to track the input signal. When the input signal is less than the D/A converter output, the counter is in the "count up" mode, and when the input signal exceeds the D/A converter output, the counter reverts to the "count down" mode. If the analog input is constant, the D/A converter "hunts" between two adjacent bit values. An advantage is that small changes are rapidly followed by the converter, though full-scale step changes require the full count time before the output becomes valid. A buffer storage register may often be necessary to store the results of a previous conversion, while the counter attempts to "servo" onto the next value. If the "up" or "down" count of the counter is disabled, the counter can also act as a "valley" or "peak" follower.

The obvious advantage of the tracking ADC is that it can continuously follow the input signal and give updated digital output data if the signal does not change too rapidly. Also, for small input changes, the conversion can be quite fast. The converter can be operated in either the track or hold modes by a digital input control.

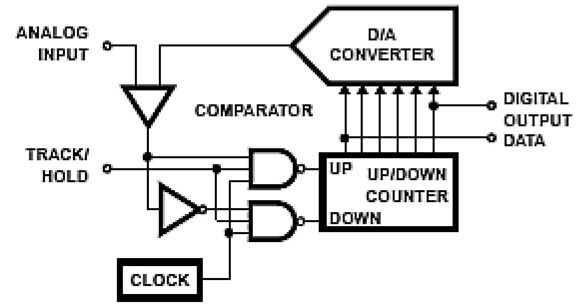


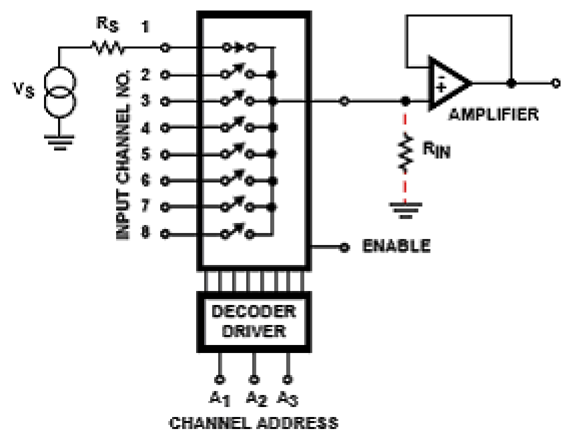
Figure 5.19 (b) Tracking /servo type ADC

5.5 ANALOGUE MULTIPLEXERS

5.5.1 Analogue multiplexer operation

Analogue multiplexers are the circuits that time-share an A/D converter among a number of different analogue channels. Since the A/D converter in many cases is the most expensive component in a data acquisition system, multiplexing analogue inputs to the A/D is an economical approach. Usually the analogue multiplexer operates into a sample-hold circuit which holds the required analogue voltage long enough for A/D conversion.

As shown in Fig. 5.20, an analogue multiplexer consists of an array of parallel electronic switches connected to a common output line. Only one switch is turned on at a time. Popular switch configurations include 4, 8, and 16 channels which are connected in single (single-ended) or dual (differential) configurations.



A_1	A_2	A_3	EN	ON-CHAN
X	X	X	0	NONE
0	0	0	1	1
0	0	1	1	2
0	1	0	1	3
0	1	1	1	4
1	0	0	1	5
1	0	1	1	6
1	1	0	1	7
1	1	1	1	8

Fig. 5.20 Analogue multiplexer circuit

The multiplexer also contains a decoder-driver circuit which decodes a binary input word and turns on the appropriate switch. This circuit interfaces with standard TTL inputs and drives the multiplexer switches with the proper control voltages. For the 8-channel analogue multiplexer shown, a one-of-eight decoder circuit is used.

Most analogue multiplexers today employ the CMOS switch circuit shown in Figure 5.21. A CMOS driver controls the gates of parallel-connected P-Channel and N-Channel MOSFETs. Both switches turn on together with the parallel connection giving relativity uniform on-resistance over the required analogue input voltage range. The resulting on -resistance may vary from about 50Ω to 2kΩ depending on the multiplexer; this resistance increases with temperature.

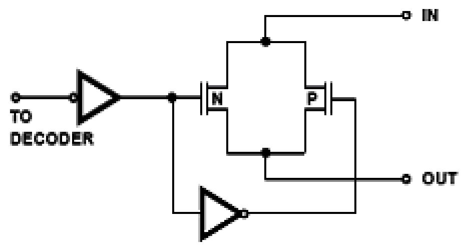


Fig. 5.21 CMOS analogue switch

5.5.2 *Analog Multiplexer Characteristics*

Because of the series resistance, it is common practice to operate an analogue multiplexer into a very high load resistance such as the input of a unity gain buffer amplifier shown in the diagram. The load impedance must be large compared with the switch on-resistance and any series source resistance in order to maintain high transfer accuracy. Transfer error is the input to output error of the multiplexer with the source and load connected; error is expressed as a percent of input voltage. Transfer errors of 0.1% to 0.01% or less are required in most data acquisition systems. This is readily achieved by using operational amplifier buffers with typical input impedances from 10^8 to 10^{12} Ω. Many sample-hold circuits also have very high input impedances.

Another important characteristic of analogue multiplexers is break-before-make switching. There is a small time delay between disconnection from the previous channel and connection to the next channel which assures that two adjacent input channels are never instantaneously connected together.

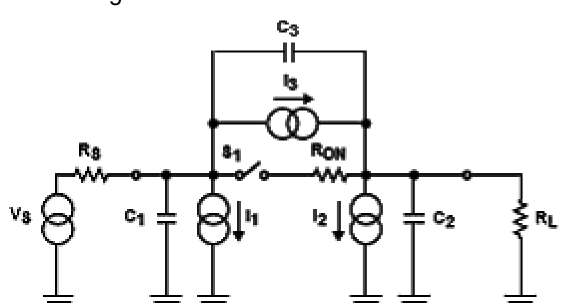


Fig 5.22 Equivalent circuit of analogue multiplexer switch

Settling time is another important specification for analogue multiplexers; it is the same definition previously given for amplifiers except that it is measured from the time the channel is switched on. Throughput rate is the highest rate at which a multiplexer can switch from channel to channel with the output settling to its specified accuracy. ***Crosstalk*** is the ratio of output voltage to input voltage with all channels connected in parallel and off; it is generally expressed as an input to output attenuation ratio in dB.

As shown in the representative equivalent circuit of Fig.5.22, analogue multiplexer switches have a number of leakage currents and capacitances associated with their operation. These parameters are specified on data sheets and must be considered in the operation of the devices. Leakage currents, generally in picoamperes at room temperature, become troublesome only at high temperatures. Capacitances affect crosstalk and settling time of the multiplexer.

5.5.3 Analogue multiplexer applications

Analogue multiplexers are employed in two basic types of operation: high-level and low-level.

In ***high-level multiplexing***; the most popular type, the analogue signal is amplified to the 1V to 10V range ahead of the multiplexer. This has the advantage of reducing the effects of noise on the signal during the remaining analogue processing.

In ***low-level multiplexing*** the signal is amplified after multiplexing, therefore great care must be exercised in handling the low-level signal up to the multiplexer. Low-level multiplexers generally use two-wire differential switches in order to minimize noise pick-up. Reed relays, because of essentially zero series resistance and absence of switching spikes, are frequently employed in low-level multiplexing systems. They are also useful for high common-mode voltages.

A useful specialized analogue multiplexer is the flying-capacitor type. This circuit, shown as a single channel in Fig. 5.23 has differential inputs and is particularly useful with high common-mode voltages. The capacitor connects first to the differential analogue input, charging up to the input voltage, and is then switched to the differential output which goes to a high input impedance instrumentation amplifier. The differential signal is therefore transferred to the amplifier input without the common mode voltage and is then further processed up to A/D conversion.

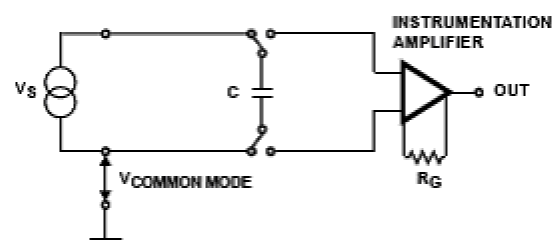


Fig 5.23 Flying capacitor multiplexer switch

In order to realize large numbers of multiplexed channels analogue multiplexers can be connected in parallel using the enable input to control each device. This is called ***single - level multiplexing***. The output of several multiplexers can also be connected to the inputs of another to expand the number of channels; this method is ***double -level multiplexing.***

5.6 SPECIFICATION OF DATA CONVERTERS

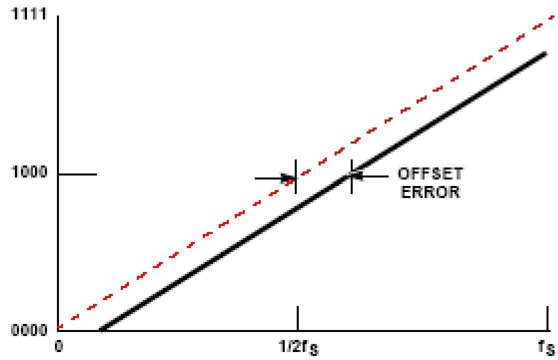
Real A/D and D/A converters do not have the ideal transfer functions discussed earlier. There are three basic departures from the ideal:

- ✓ offset
- ✓ gain, and
- ✓ linearity errors.

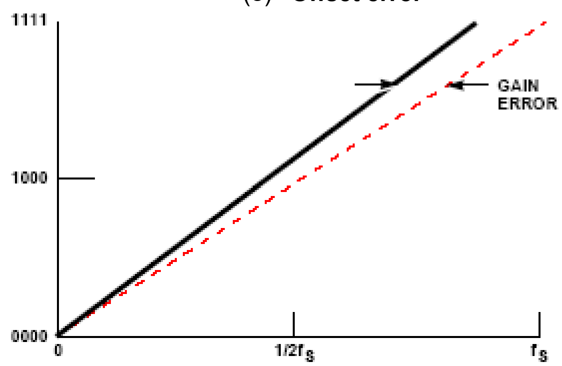
These errors are all present at the same time in a converter; in addition they change with both time and temperature. Fig 5.24 shows A/D converter transfer functions which illustrate the three error types.

5.6.1 Offset and gain errors

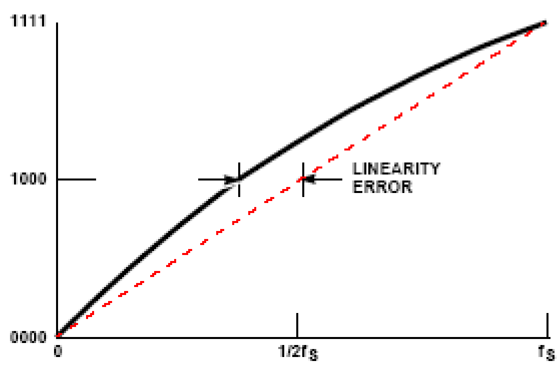
Fig. 5.24 (a) shows offset error, the analogue error by which the transfer function fails to pass through zero. Next, in Fig. 5.24 (b) is gain error, also called scale factor error; it is the difference in slope between the actual transfer function and the ideal, expressed as a percent of analogue magnitude. Most A/D and D/A converters available today have provision for external trimming of offset and gain errors. By careful adjustment these two errors can be reduced to zero, at least at ambient temperature.



(e) **Offset error**



(f) **Gain error**



(c) Integral linearity error

Fig. 5.24 Common A/D & D/A converter errors

5.6.2 Linearity errors

Basically there are only two ways to reduce linearity error in a given application. First, a better quality higher cost converter with smaller linearity error can be procured. Second, a computer or microprocessor can be programmed to perform error correction on the converter. Both alternatives may be expensive in terms of hardware or software cost. Four types of linearity errors associated with data converters are discussed below.

5.6.2.1 Integral linearity

Fig. 5.24 (c) shows the integral linearity error, or nonlinearity. This is defined as the maximum deviation of the actual transfer function from the ideal straight line at any point along the function. It is expressed as a percent of full scale or in LSB size, such as ±1/2 LSB, and assumes that offset and gain errors have been adjusted to zero.

5.6.2.2 Differential linearity

This is defined as the maximum amount of deviation of any quantum (or LSB change) in the entire transfer function from its ideal size of FSR/2^n. Fig. 5.25 shows that the actual quantum size may be larger or smaller than the ideal; for example, a converter with a maximum differential linearity error of ±1/2 LSB can have a quantum size between 1/2 LSB and 1½ LSB anywhere in its transfer function. In other words, any given analogue step size is (1± ½) LSB. Integral and differential linearities can be thought of as macro- and micro-linearities, respectively.

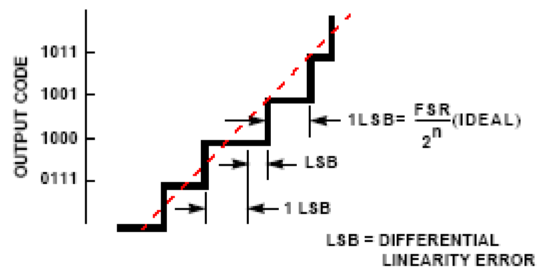


Fig. 5.25 Differential linearity error

5.6.2.3 Monotonicity

This is the characteristic whereby the output of a circuit is a continuously increasing function of the input. Fig. 5.26 shows a non-monotonic D/A converter output where, at one point, the output decreases as the input increases. A D/A converter may go non-monotonic if its differential linearity error exceeds 1 LSB; if it is always less than 1 LSB, it assures that the device will be monotonic.

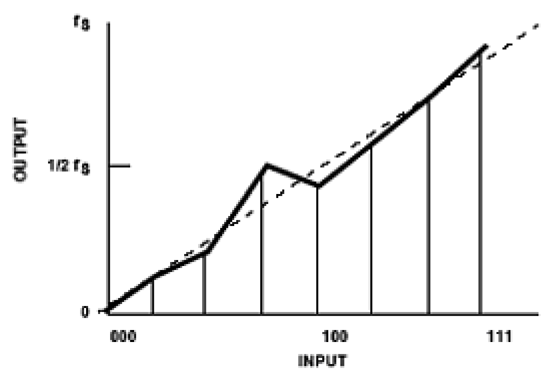


Fig. 5.26 Non-monotonic DAC

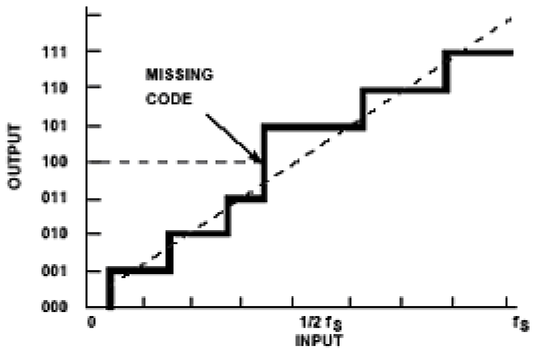


Fig. 5.27 ADC with missing code

5.6.2.4 Missing code error

The term missing code, or skipped code, applies to A/D converters. If the differential linearity error of an A/D converter exceeds 1 LSB, its output can miss a code as shown in Fig. 5.27. On the other hand, if the differential linearity error is always less than 1 LSB, this assures that the converter will not miss any codes. Missing codes are the result of the A/D converter's internal D/A converter becoming non-monotonic.

<u>*Exercise 5.1*</u>

1. Differentiate between a multi-channel Data Acquisition System (DAS) using digital multiplexing before transmission and that employing multiplexing of outputs of sample/holds. Give examples of instrumentation systems that require these two different schemes of data acquisition.
2. With the aid of a circuit diagram explain the operation of a 4-bit digital to analogue converter (DAC) making use of a network of precision-weighted resistors connected to the inverting input terminal of an op-amp. If the reference dc supply voltage for the converter is 10 V (negative), and the MSB weighting resistor is 15 kΩ. Determine: -
   i.) the magnitude of the LSB weighting resistor
   ii.) the output voltage for the following digital codes 1001, 1011.
3. i.) Using appropriate sketches explain the operation of a counter type A/D Converter.
   ii.) Explain the basic drawbacks of this type of converter.
   iii.) A 12 bit Analog to digital Converter is to be used and be driven by a 2 MHz clock. Calculate the maximum time it will take to perform a conversion and hence determine the maximum conversion rate.
4. (a) With the aid of appropriate sketches describe how the R-2R ladder type Digital to Analog Converter operates.

   (b) For the following binary inputs, determine the analog output voltage for an 8 bit R-2R ladder network if the reference voltage is 10 V.
   (i) 11001100, (ii) 00110011, (iii) 10101001
5. (a) With the aid of appropriate sketches describe the transfer function of a ***sample and hold*** (S/H) device. What is its relevance in Data Acquisition systems?
   (b) Explain the meanings of the following terms that are associated with S/H devices:
   - ✓ Droop
   - ✓ Aperture time
   - ✓ Acquisition time

   (b) A particular kind of S/H circuit has a maximum charge current of 40 mA and a droop current of 150 pA. Calculate the droop rate if the acquisition time for a 5V step is 6 microseconds.
6. What are multiplexers? Distinguish between the high and low-level modes of multiplexer operation.
7. With the aid of a sketch describe the working of the flying capacitor multiplexer switch.
8. Describe for types of linearity errors that characterize real data converters.

TOPIC 6.0 ANALOGUE & DIGITAL SIGNAL TRANSMISSION

6.1. *INTRODUCTION*

Although the microprocessor and digital network technologies have fundamentally reinvented the ways in which today's data acquisition systems handle data, much laboratory and manufacturing information is still communicated the "old" way, via analog electrical signals and a fundamental understanding of how analog signal transmission works must first begin with a discussion of electrical basics.

To understand the ways in which an analog signal is transmitted over a circuit, it is first important to understand the relationships that make analog signal transmission possible. It is the fundamental relationship between voltage, current, and electrical resistance (Figure 6.1) that allow either a continuously varying current or voltage to represent a continuous process variable.

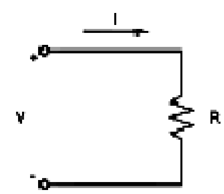


Figure 6.1. A Basic Electric Circuit

While charge flow is electric current, voltage is the work done in moving a unit of charge (1 coulomb) from one point to another. The unit of voltage is often called the potential difference, or the volt (V). The International System of Units (SI) unit for electrical flow is the ampere (A), defined as one coulomb per second (c/s).

A signal source of voltage, V, will cause a current, I, to flow through a resistor of resistance, R. Ohm's law, which was formulated by the German physicist George Simon Ohm (1787-1854), defines the relation: V=IR . While most single-channel analog signal transmissions use direct current (dc) variations in current or voltage to represent a data value, frequency variations of an alternating current (ac) also can be used to communicate information. In the early 19th century, Jean Baptiste Joseph **Fourier**, a French mathematician and physicist, discovered that ac signals could be defined in terms of sine waves. A sine wave is described by three quantities: amplitude, period, and frequency. The amplitude is the peak value of the wave in either the positive or negative direction, the period is the time it takes to complete one cycle of the wave, and the frequency is the number of complete cycles per unit of time (the reciprocal of the period).

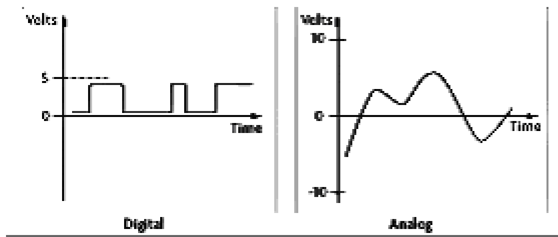


Figure 6.2:Digital and Analog Signal Representations

6.2. *ANALOG SIGNAL TYPES*

Most data acquisition signals can be described as: -

- analogue
- digital, or
- pulse.

While analog signals typically vary smoothly and continuously over time, digital signals are present at discrete points in time (Figure 6.2). In most control applications, analog signals range continuously over a specified current or voltage range, such as 4-20 mA dc or 0 to 5 V dc. While digital signals are essentially on/off (the pump is on or off, the bottle is there or isn't), analog signals represent continuously variable entities such as temperatures, pressures, or flow rates. Because computer-based controllers and systems understand only discrete on/off information, conversion of analog signals to digital representations is necessary.

Transduction is the process of changing energy from one form into another. Hence, a transducer is a device that converts physical energy into an electrical voltage or current signal for transmission. There are many different forms of analog electrical transducers. Common transducers include load cells for measuring strain via resistance, and thermocouples and resistance temperature detectors (RTDs) for measuring temperature via voltage and resistance measurement, respectively. Transmission channels may be ***wires*** or ***coaxial cables***.

For *noise-resistant transmission* over significant distances, the raw transducer signal is often converted to a 4-20 mA signal by a two-wire, loop-powered transmitter. The bottom value of a process variable's range, for example, a temperature, is typically designated as 4 mA, making it easy to distinguish transmitter failure (0 mA) from a valid signal. If the current source is of good quality, current loops tend to be less sensitive to noise pickup by electromagnetic interference than voltage-based signals.

6.3. ***NOISE*** *(also called* ***Interference)***

In transmitting analog signals across a process plant or factory floor, one of the most critical requirements is the protection of data integrity. However, when a data acquisition system is transmitting low-level analog signals over wires, some signal degradation is unavoidable and will occur due to noise and electrical interference. ***Noise*** and ***signal degradation*** are two basic problems in analogue signal transmission.

Noise is defined as *any unwanted electrical or magnetic phenomena that corrupt a message signal*. Noise can be categorised into two broad categories based on the source:

- internal noise and
- external noise.

While internal noise is generated by components associated with the signal itself, external noise results when natural or man-made electrical or magnetic phenomena influence the signal as it is being transmitted. Noise limits the ability to correctly identify the sent message and therefore limits information transfer.

6.3.1. *Sources of noise in instrumentation systems*

Some of the sources of internal and external noise include:

1. Electromagnetic interference (EMI);
2. Radio-frequency interference (RFI);
3. Leakage paths at the input terminals;
4. Turbulent signals from other instruments;
5. Electrical charge pickup from power sources;
6. Switching of high-current loads in nearby wiring;
7. Self-heating due to resistance changes;
8. Arcs;
9. Lightning bolts;
10. Electrical motors;
11. High-frequency transients and pulses passing into the equipment;
12. Improper wiring and installation;
13. Signal conversion error; and
14. Uncontrollable process disturbances.

6.3.2. **Common mode and normal mode noise**

Signal leads can pick up two types of external noise-*common mode* and *normal mode*. Normal mode noise enters the signal path as a differential voltage and cannot be distinguished from the transducer signal. Noise picked up on both leads from ground is referred to as common-mode interference.

6.3.3. *Signal to Noise Ratio (SNR)*

Whether the noise is detrimental to the proper performance of the system depends on the ratio of the total signal power to the total noise level. This is referred to as the signal-to-noise ratio. If the signal power is large in comparison to the noise signal, the noise can often times be ignored. However, with long-distance signals operating with limited signal power, the noise may disrupt the signal completely.

6.3.4. *Current transmitters*

Current-driven devices have been most widely accepted in processing plants, with a common current range of 4-20 mA. Low-level current signals are not only safe, but are not as susceptible to noise as voltage signals. If a current is magnetically coupled into the connecting wires in the transmission of the signal from a current source, no significant change in the signal current will result. If the transducer is a voltage-driven device, the error adds directly to the signal. After current transmission, voltage signals can be easily rederived.

6.4. *AVOIDANCE OF NOISE IN INSTRUMENTATION SYSTEMS*

Even though internal and external interference can be problematic in sending analog signals, analog signal transmission is widely and successfully used in industry. The effects of noise can be reduced by: -

1. careful engineering design,
2. proper mechanical arrangements & installation,
3. routing techniques of wires and cables,
4. separation of frequencies,
5. analogue to digital conversion
6. amplitude modulation using ac carrier systems,
7. screening / shielding or guarding, and
8. grounding.

6.4.1 *Amplitude modulation*

In amplitude modulation the amplitude of the carrier signal (frequency f_s) is changed according to the instantaneous value of a modulating signal (frequency f_m) effectively shifting the spectrum of the measurement signal from lower to higher frequencies.

Consider a sinusoidal measurement signal

$$V_s = A_s \sin \omega_s t = A_s \sin 2\pi f_s t,$$

that is to be modulated by a modulating signal described as: $V_m = A_m \sin 2\pi f_m t$

with $A_s < A_m$ and $f_s < f_m$

where A_s = Amplitude of the measurement signal

A_m = Amplitude of the modulating signal

The amplitude modulated signal

$$V_i = V_s V_m = A_i \sin 2\pi f_s \sin 2\pi f_m, \text{ with } A_i = A_1 A_2$$

Using the identity

$$\sin A \sin B = \frac{1}{2}\cos(A - B) - \frac{1}{2}\cos(A + B)$$

, V_i can be expressed as

$$V_i = \frac{A_i}{2}\cos 2\pi(f_s - f_m) - \frac{A_i}{2}\cos 2\pi(f_s + f_m)$$

> (1)

The A.M. signal is the sum of two cosinusoidal signals with frequencies f_s - f_m and f_s + f_m. A pass band filter is used to discriminate against noise by choosing the upper and lower cut-off frequencies such that they include the measurement signal between f_s - f_m and f_s + f_m. Outside the pass band all other drift and interference are attenuated

6.4.2 *Grounding and shielding*

Proper grounding also is essential for effective operation of any measurement system. Improper grounding can lead to potentially dangerous ***ground loops*** and susceptibility to interference. To understand the principles involved in shielding and grounding, some terms must first be understood.

A **ground** is a conducting flow path for current between an electric circuit and the earth. Ground wires are typically made with materials that have very low resistance. Because current takes the path of least resistance, the ground wires connected from the system provide a stable reference for making voltage measurements. Ground wires also safeguard against unwanted common-mode signals and prevent accidental contact with dangerous voltages. Return lines carry power or signal currents (Figure 6.4). A ground loop is a potentially dangerous loop formed when two or more points in an electrical system are grounded to different potentials.

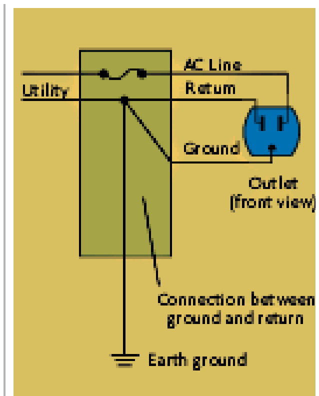


Figure 6.3. A Ground Conductor

There are many different grounding techniques designed to not only protect the data being transmitted, but to protect employees and equipment. There are two ways in which all systems should be grounded. First of all, all of the measuring equipment and recording systems should be grounded so that measurements can be taken with respect to a zero voltage potential. This not only ensures that potential is not being introduced at the measuring device, but ensures that enclosures or cabinets around equipment do not carry a voltage. To ground an enclosure or cabinet, one or more heavy copper conductors are run from the device to a stable ground rod or a designated ground grid. This system ground provides a base for rejecting common-mode noise signals. It is very important that this ground is kept stable.

The second ground is for the signal ground. This ground is necessary to provide a solid reference for the measurement of all low-level signals. It is very important that this ground is grounded separate and isolated from the system ground. If a signal return line is grounded at the signal source and at the system ground, a difference in potential between the two grounds may cause a circulating current (Figure 6.4). In this case, the circulating current will be in series with the signal leads and will add directly to the signal from the measuring instrument. These ground loops are capable of creating noise signals 100 times the size of the original signal. This current can also be potentially dangerous. In a single-point ground configuration, minimal current can flow in the ground reference. Figure 6.5 (**Single point grounding**) shows that by grounding the wire at the signal end only, the current has no path, eliminating the ground loop.

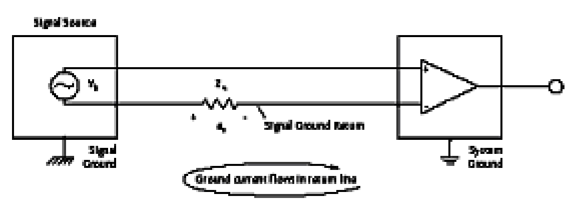


Figure 6.4. Incorrect Grounding of Signal Circuit

For off-ground measurements, the shield or the ground lead is stabilised with respect to either the low-level of the signal or at a point between the two. Because the shield is at a potential above the zero-reference ground, it is necessary to have proper insulation.

6.4.3 *Wire & Cable Options*

Another important aspect to consider in analog signal transmission is a proper wiring system, which can effectively reduce noise interference. Analog signal transmission typically consists of two-wire signal leads or three-wire signal leads. In systems that require high precision and accuracy, the third signal lead, or shield, is necessary. In the three-wire configuration, the shield is grounded at the signal source to reduce common-mode noise. However, this does not eliminate all of the possibilities of the introduction of noise. It is crucial to prevent the noise pickup by protecting the signal lines, for example, in the case where the noise and signal frequency are the same. In this scenario, the signal cannot be isolated / filtered from the noise at the receiving device.

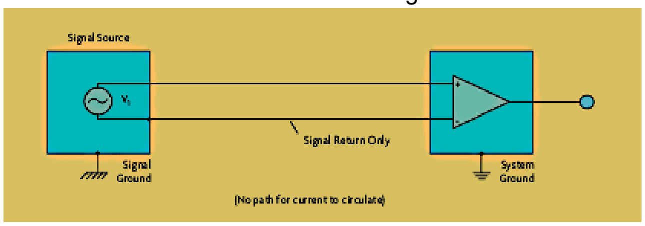


Figure 6-5: Single-point Grounding of Signal Circuit

Generally, *two-wire transmission* mediums are used to carry an analog signal to or from the field area. A wire carrying an alternating current and voltage may induce noise in a pair of nearby signal leads. A differential voltage/noise will be created since the two wires may be at different distances from the disturbing signal. There are many different wiring options that are available to reduce unwanted noise pickup from entering the line.

Types of wire in data acquisition

Four types of wires are fundamental in data acquisition:

- plain pair,
- shielded pair,
- twisted pair, and
- coaxial cable.

While plain wire can be used, it is generally not very reliable in screening out noise and is not suggested.

A shielded pair is a pair of wires surrounded by a conductor that does not carry current. The shield blocks the interfering current and directs it to the ground. When using shielded pair, it is very important to follow the rules in grounding. Again, the shield must only be grounded at one source, eliminating the possibility of ground-loop currents.

Twisted-pairs help in elimination of noise due to electromagnetic fields by twisting the two signal leads at regular intervals. Any induced disturbance in the wire will have the same magnitude and result in error cancellation.

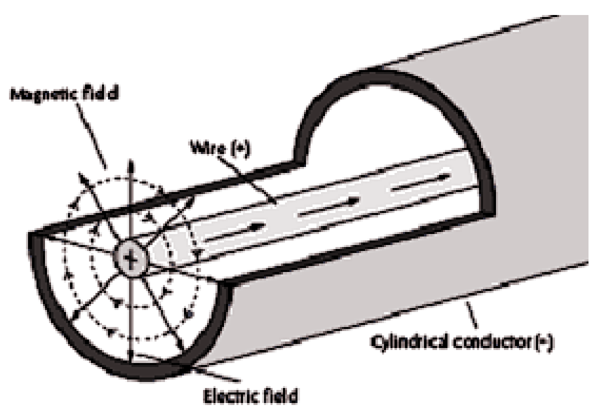


Figure 6.7 Coaxial Cable Construction

A coaxial cable is another alternative for protecting data from noise. A coaxial cable consists of a central conducting wire separated from an outer conducting cylinder by an insulator. The central conductor is positive with respect to the outer conductor and carries a current (Figure 6.6). Coaxial cables do not produce external electric and magnetic fields and are not affected by them. This makes them ideally suited, although more expensive, for transmitting signals.

Although noise and interference cannot be completely removed in the transmission of an analog signal, with good engineering and proper installation, many of the effects of noise and interference can be substantially reduced.

6.5 ***DIGITAL DATA TRANSMISSION***

The primary objective of data transmission is the transfer of information from one location to another. In contrast to the analog signal that has an infinite number of voltage levels, the digital signal information is "binary" or two-valued. A binary symbol is called a binary digit or bit, and a group of bits representing an information unit is called a byte. A selective arrangement of seven bits provide 2^7 (or 128) distinct character combinations or 128 bytes. The *American Standard Code for Information Interchange* ***(ASCII)*** is an excellent example of such an arrangement used in the transmission of numbers, alphanumeric characters, and miscellaneous information.

The major points that must be agreed upon by the sender and receiver, to accomplish uniform flow of data are:

(i) The nominal rate of transmission, or how many bits are to be emitted per second by the sender

(ii) The specified information code providing a one-to-one mapping ratio of the information to bit-pattern and vice versa

(iii) A particular scheme by which each bit can be positioned properly, within a byte by the receiver of the data (in the case of bit-serial transmission)

(iv) A protocol (handshaking) sequence necessary to ensure an orderly flow of information

(v) The electrical stress representing the logic values of each bit and the particular pulse code to be used.

The items (ii), (iii), and (iv) are essentially "software" type decisions and do not affect the characteristics of the actual signal sent along the transmission line or any other media used for transmission of information. Items (i) and (v) are directly influenced by the characteristics of the line-drivers, the line-receivers, and the transmission-lines used for data transmission

6.5.1 Data transmission system

A typical data-transmission system contains the "*information source"* from which data has to be transmitted through a suitable medium to the destination termed as the *"information sink".* The information source can be a computer terminal, a digitized transducer output, or any other device generating a stream of bits at the rate of one bit in every *ta* seconds. The "information rate" of the system is then $1/t_B$ bits per second. As shown in Fig. 6.8, the information source feeds to a "source encoder" which performs logic operations not only on the data but also on the associated clock, and perhaps, the past data bits as well. Thus the source encoder produces a data stream controlling the line driver. The line driver interfaces the internal logic levels of the source (viz., TTL, MOS, etc.) with the transmission line. The transmission line in turn carries the signal produced by the line driver to the line receiver.

The line receiver makes a decision on the signal logic state by comparing the received signal to a decision threshold level, and the "sink-decoder" performs logic operations on the binary bit stream recovered by the line receiver. For example, the sink decoder may contract the clock rate from the data or perhaps detect and correct errors in the data. From the optional sink decoder, the recovered binary data passes to the "information sink", which is the destination for the information-source data.

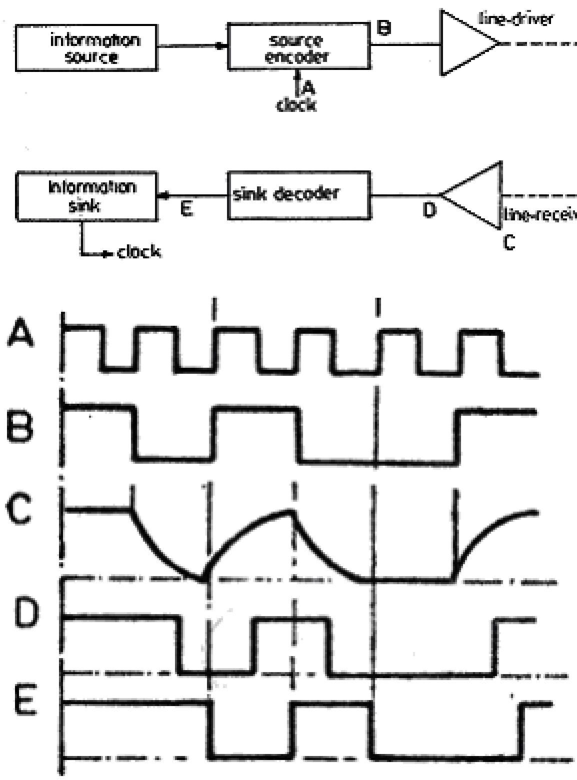


Fig 6.8 Block schematic for a digital data transmission system

6.5.2 Pulse code formats

Classification pulse code formats is based on three criteria, i.e.

- ✓ the form of information transmission
- ✓ the relation to zero, and
- ✓ the direction

Based on the form of information – transmission, the format used can be:

(i) Full binary transmission where both "0" and "1" bits are part of the formats.

(ii) Half binary transmission, where only the "1" are transmitted, recognising the "0" by the absence of a pulse at the time clock transition.

(iii) Multiple level transmission, where ternary and quadratic codes are used for each transmission pulse.

Based on the relationship with the zero level the transmission format can be either:

- ✓ Return to zero (RZ)
- ✓ Non – return to zero (NRZ)

In the RZ format there is a return to the zero level after the transmission of each bit of information while in the NRZ format a change in the voltage level occurs only if consecutive bits being transmitted are different e.g. from "0" to "1" or "1" to "0". When they are the same e.g. from "0" to "0" or "1" to "1" the voltage level is sustained.

Based on the direction the code format used cab be either unipolar, where the pulses are in a single direction or bipolar, where the pulses are in different directions. An illustration of these pulse code formats is given in Fig. 6.9.

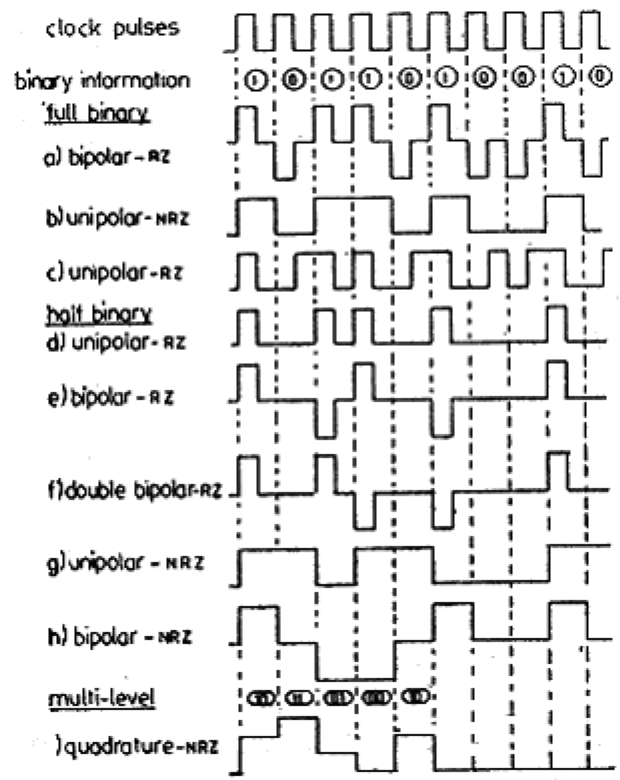


Fig. 6.9 Pulse code formats used in digital data transmission

6.5.3 Modulation techniques for digital data transmission

Three types:

- ✓ Amplitude modulation
- ✓ Frequency modulation
- ✓ Phase modulation

In each of these techniques a sinewave carrier is employed to convey data by means of a change in its amplitude, frequency or phase. The three modulation techniques follow the block diagram Fig. 6.10.

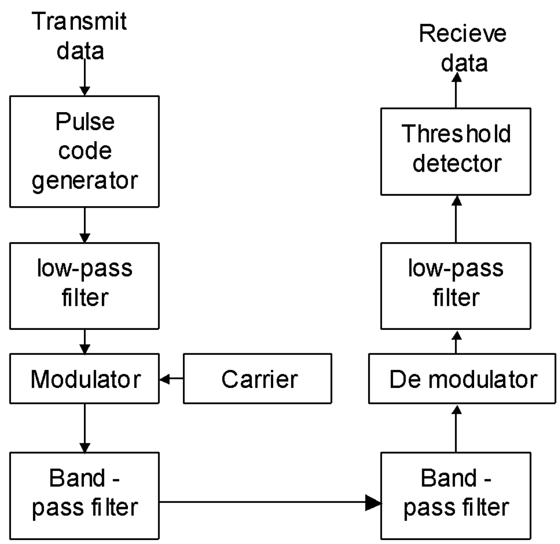


Fig 6.10 Block diagram of modulation/demodulation

Modulators –demodulators (***modems***) are used to interface the terminal equipment of computers and business machines with the telephone network. These devices convert the electrical waveforms in the terminals into a modulated form of stable transmission over voice –telephone facilities.

Telemetry systems use transducers to collect data from remote locations. The data is scanned and multiplexed before being transmitted to a centralised location through a suitable link. A receiver located at the collection centre reconstructs the data for distribution to individual display and recording systems.

TOPIC 7.0
DISPLAY & RECORDING DEVICES

7.1. *INTRODUCTION*

Once data has been acquired, there is often a need to store it for current and future reference. Today, alternative methods of data storage embrace both digital computer memory and that old traditional standby-paper. Covering in any detail the scores of instruments on the market today for recording, logging, printing, and storage of data is far beyond the scope of this section. The aim here is to identify and define major categories and offer some practical guidelines to a selection best suited for the application.

There are two principal areas where recorders or data loggers are used:

1. ***Process measurements*** for such variables as temperature, pressure, flow, pH, humidity;
2. ***Scientific and engineering applications*** such as high-speed testing (e.g., stress/strain), statistical analyses, and other laboratory or off-line uses where a graphic or digital record of selected variables is desired.

Instrumentation for each of these fields can have many operating principles and features in common. Each area, however, may place more emphasis on the importance of such factors as *cost*, *method of use* (e.g., benchtop vs. panel mounting), *portability*, degree of *accuracy*, *speed of response*, and *number of points (channels)* to be measured.

7.2 MODERN TRENDS IN DATA DISPLAY AND RECORDING

For some time now, digital computer systems have had the ability to provide useful trend curves on CRT displays-displays that could be analyzed and printed out. Today, a similar approach permits digital inputs to recorders and data loggers to be stored electronically on-board or sent to a remote PC, distributed control system (DCS) workstation, or full-colour printer.

There is now a plethora of data acquisition hardware and software that can work with PCs-making it quite plausible to classify them as recorders or, alternatively, data loggers. These are referred to as PC-based solutions.

There remains a continuing role for conventional standalone recorders that still use paper and pen or printer, but here, too, such products have gained greatly in capabilities through advances in electronics and new technologies such as inkjet and laser printing.

New technology has also made paperless, or videographic, recorders a practical reality, with attendant savings in materials and maintenance. This mushrooming class of instruments mimics conventional trend recorder displays; it also has the capabilities of a data logger and can export data to a PC or DCS workstation, in real-time or periodically for later analysis.

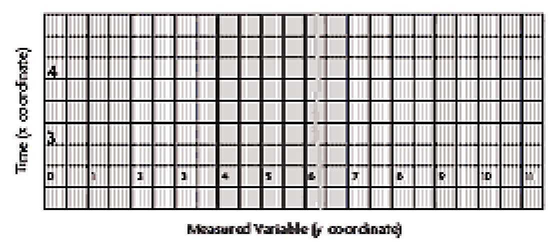


Figure 7.1:Typical Strip-Chart recorder Trace

The traditional data logger, whose original basic function was to print out digital values of a measured variable versus time, also has benefited from new technology that has added disk or card memory storage and faster logging capabilities. Some can provide *real-time recording* via network connection to a host system.

In recent years, conventional analog recorders have had the functions of a data logger added to their capabilities. Called "hybrid recorders", they combine analog trend representations with digital information on the same chart paper.

Scores of specialised software packages are offered not only to implement the PC-based functions mentioned above but also to perform complex mathematical operations or signal conditioning on input and output data.

For high-speed recording, the traditional analog oscilloscope has also been enhanced by a digital storage version (DSO), which finds an important application in test and analysis of today's many electronic products-both consumer and industrial.

There are therefore no longer sharp distinctions between many of the products that serve the "front end" of a data acquisition system. In fact, the breakneck evolution of computer and networking technologies will continue to meld alternative offerings for process monitoring, data acquisition, and test and measurement applications.

7.3 DEFINITIONS & CLASSIFICATIONS

Acknowledging that sharp lines no longer exist, consider the fundamental ways in which each type of recording instrument functions to store, analyze, and record data, the following definitions/classifications are possible: -

7.3.1 *Recorder vs. data logger*

A traditional ***recorder*** accepts an input and compares it to a full-scale value. The recording pen is then moved to a point on the scale that represents a useful value of the input-say, 500^0F with a chart range of 0 to 1000 ^{0}F-at a time that is given by the calibrations on the recorder's moving chart (Figure 7.1).

A **data logger** accepts an input which is fed into an analog-to-digital converter and stored digitally, or printed out as a series of time-stamped values. Today, many such instruments have some form of on-board intelligence which provides the user with diverse capabilities for signal conditioning as well as data storage and analysis.

As an alternative to a recorder, a data logger can have the ability to accept a greater number of inputs (referred to as channels), with better resolution and accuracy. Only a recorder, however, can provide a truly continuous trend display of the variable's change with time.

Fig.7.2 Paper-based options for recording devices

7.3.2 Graphic recorders

The term graphic recorder more aptly describes the majority of recorders used in data acquisition systems. It is defined as an instrument which, in some automatic fashion, draws a graph. Such a graph may relate two or more variables to time, or to each other (the latter, typified by ***an X-Y recorder***). "Graphic" is also used widely today in reference to ***CRT displays***, many of which do not qualify as graphs. Graphic coordinates hearken back to basic geometry, where curves are plotted on a rectangular coordinate matrix. The ***strip-chart recorder***, so common in a great variety of forms today, had its origin in this basic concept (Figure 7.1). Note that the vertical coordinate (time) is the X-axis of a graph and the horizontal coordinate (measured variable) is the Y- axis. The heavy vertical line in the center of the chart is the curve plotted by the recorder (often called the *chart record* or trace).

Chart speed is a term that expresses the rate at which the recording paper in a strip-chart recorder moves. The relationship between the period or the recorded signal and the chart speed is given as: -

$$\text{Period} = \frac{\text{Time}}{\text{Cycle}} = \frac{\text{Time base}}{\text{chart speed}}$$

>(1)

The frequency of the recorded signal is the inverse of the period (i.e. period^{-1}). For example if the chart speed is 25 mm/cycle and one cycle of the signal being recorded extends over 5mm (i.e. the time base is 5 mm/cycle) then the frequency is 5 cycles/second.

Recorders can have adjustable chart speeds-fixed, once configured-to suit the application. Thus, at 4 (o'clock) in Figure 7.1, an operator can see that the temperature was 625¡F. He or she can also note that the temperature has remained rather steady. The trend is often as important, or more important, than the precise value. Thus, this type of instrument is often called a **trend recorder**.

With an **X-Y recorder**, the time coordinate X is replaced by some other variable, such as stress, plotted versus strain. Other versions of this type may have two different variables, Y and Y' plotted vs. X or have time (t) also as one of the variables.

Circular-chart coordinates (Figure 7.3) were developed early in recorder history to utilise the convenience of a spring-wound clock and synchronous motor movements. Depending on application needs, the circular chart speed can be one rotation in 24 hours, one rotation in a week, or another arbitrary time interval. Note the advantage of a circular chart in giving at a glance a complete history of one or several variables over a period of time-albeit not with a high-accuracy readings.

Chart sizes are referred to by total diameter, such as 10 inch or 12 inch, although full-scale pen movement is often much less. Because many plant operators are used to reading circular chart recorders, suppliers still find a strong market for them.

From a commercial cataloguing standpoint, graphic recorders are generally classified by their common usages or physical appearance, such as strip chart, circular chart, or special purpose (X-Y, event, or oscillographic). A further broad distinction is whether the instrument is ***flatbed*** for ***benchtop*** mounting or ***vertical***, for *panel or rack* mounting.

Data loggers, too, can be classified as benchtop units; others are described as panel-mount printers and still others are "vertical" benchtop units. Smaller, handheld units also are available.

As with any method of classification, such names leave gray areas. For example, some flatbed types utilise strip charts and some are vertical. Many data loggers are bench-top, but some are vertical and panel-mounted. And videographic, or paperless recorders can provide screen displays that look like a strip chart or a circular chart.

7.4 TREND RECORDERS

Vertical strip-chart recorders head this category, embracing many types that are commonly designed for permanent installation and are normally mounted vertically-usually in panel cut-outs for process measurements but also on relay racks as for laboratory measurements. Size is one distinction of this class, with chart widths ranging from 100 mm to 10 inches (250 mm) or more.

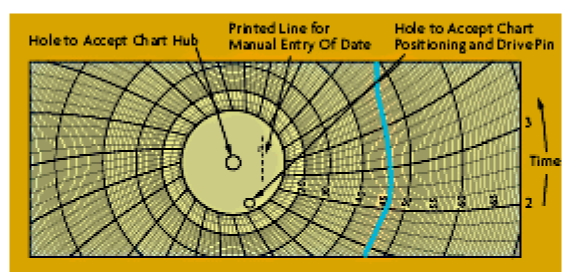


Figure 7.3 Typical Circular Chart Recorder Trace

Other terms used to describe vertical strip-chart recorders include electronic, hybrid, and multi-point. Certain models can record up to 32 channels, although common designs for process measurements normally do not exceed 12. Instruments in this class are designed to plot variables over long periods of time, utilizing paper take-up rolls or trays for Z-fold paper. Chart speeds, programmable in most cases, normally range from 1 cm/min to 30 cm/min.

This category contains many variations on the basic principle of strip-chart recording shown in Figure 7.1. One extensive catalog reference covers 29 different models; a majority of these are for the measurement of process variables, but there are also numerous designs customised for scientific and laboratory applications.

The impact of the microprocessor is obvious in many strip-chart recorder models, with digital storage of data points a common advanced feature. Another result of new technology is the ability to handle a variety of variables, such as flow, along with temperature and pressure.

Another result of microprocessor advances is the hybrid recorder, which provides high-speed recording of up to 30 recording channels as well as digital measurement data. On these units, the digital information is sometimes printed on the left margin of the chart in up to six colours. Typically, some form of keyboard on the front panel (or accessible with the door open) is used for programming of chart speed, alarm setpoints, full-scale range, and other features. Some of the features of strip-chart process recorders can include:

- 1 to 4-pen continuous writing and up to 24-point dot printing;
- Memory card for recorder setup and data storage; and,
- Digital display, horizontal bar graph, and digital totaliser for each channel.

In large case, 250-mm recorders, additional features may include:

- Interactive, menu-driven programming;
- Trend printing with user-assignable colours; and,
- Ability to print up to 60 analog input channels.

Flatbed and portable versions also are available, more suited for laboratory or test uses. Some include signal input modules which plug into compartments on the recorder so that, for example, they can accommodate the direct input from a thermocouple. Some models have as many as six such plug-in input modules and are referred to as modular recorders.

One recorder with a capability to handle 1,600 samples/second employs a thermo-array recording system, with 200 dots per inch (dpi) resolution on a 200-mm wide thermally sensitive chart. Referred to as therographic recording, this method uses a linear array of fixed thermal-printing elements (styli) across the chart width. A heated stylus affects the heat sensitive paper to produce the chart record. The chart grid system is generated by the printer.

A portable, benchtop model that handles 10 or 20 channels uses a printing technique described as raster scan. With this printing method, the recorder can handle high-speed scanning of 20 points/second and high-speed recording of 50 points/second.

7.4. Trend recorder with a long term paper record.

A highly-specialised "transient" recorder with a 170-mm wide thermally sensitive paper chart is offered to record signals for which standard recorders are too slow. It offers digital sampling and storage up to 1 MHz. It has a modular design that allows expansion to 30 channels. Its 16 K memory per channel can store up to 16,000 samples and its standard analog output can directly interface to an oscilloscope or strip-chart recorder.

7.5 FLATBED RECORDERS

For testing and other laboratory work, flatbed recorders are offered with a variety of switch-selectable, full-scale ranges to accommodate different inputs. Designed to meet various application needs, they typically offer a range of chart speeds from centimetres/minute to centimetres/hour. A typical slow speed is 1 cm/hr; a fast speed generally is 60 cm/min.

Most flatbed recorders, including X-Y types, record with ink, using fiber-tipped, disposable cartridges. Chart width is typically 100 mm; larger models range from 200 to 250 mm. Notable is the instrument's speed-typically 0.4 seconds full-scale response for this class.

Offerings in various flatbed models include:

1. Ability to accept a variety of inputs, such as from thermocouples and RTDs, ac/dc voltage/current, and relative humidity-with available inputs depending on the model;
2. Two and three pens with pen offset compensation. (One model has four channels with a choice of display modes: measured data, bar graph, or range data);
3. Widely adjustable chart speeds. One portable, battery-operated model for recording temperature and relative humidity comes with chart paper pre-printed for chart capacity of 1, 7, and 32 days;

Figure 7.5.
Portable flatbed recorders

4. Built-in 3.5-in. digital (LCD) display (on a one-channel model) indicating true value of input;
5. Protocol printing on chart to permanently record measurement range, chart speed, etc.;
6. Built-in remote control of paper advance, pen lift (automatically lifts the pen after 30 seconds if chart is not advancing), and event marker; and,
7. One model type, called a "***function recorder***," is offered with up to three channels, each of which can have its own plug-in input module to equip the recorder for specific inputs types.

More sophisticated units add RS-232 serial or IEEE-488 parallel computer interface to handle data output and remote configuration, 3.5-in. disk drive for data logging, and digital data print-out.

7.6 *XY* RECORDERS

XY recorders, though usually flatbed in design, are really a class by themselves. Most employ a potentiometric pen-positioning system, but some newer models feature a digital servo system. The instrument's basic function is to plot the relationship of two input variables with no regard for time. Input signals to the recorder can be analog or digital, with suitable conditioning as required. Common sizes for the chart paper are 8.5 by 11 and 11 by 17 inches.

In addition to features as described above for flatbed recorders, options available in certain X-Y recorders include:

Ability to also plot either the X or Y inputs versus time. Such a capability is expressed as X-Y, Y-t and X-t recording. The mode to be used is selectable on the instrument's front adjustment panel.

An additional *servo-actuated measuring element* and recording pen yields an XX'Y or XYY' recorder that can plot the relationship among three variables.

To distinguish different chart records plotted on the same chart, one design with a digital plotter can use eight different colour pens (disposable felt tipped). Pen change is selected manually or via RS-232C interface.

Recorder responsiveness is expressed in a "***slewing***" speed along each axis. The speed may be, say, 0.6 meter/second for the X-axis and 0.8 m/s for the Y-axis. Some thermal-array recorders also include an X-Y recording mode. Such instruments offer much higher real-time frequency response and the ability to print many Y channels against a single X channel. Some thermal-array recorders include a memory X-Y mode, so that the user can capture high-speed transients for later X-Y replay.

7.7 CIRCULAR CHART RECORDERS

7.6. Circular chart recorders with display, recording, and control functionality.

Basic characteristics of this type of recorder were covered with reference to Fig. 7.6. In this class, too, the microprocessor has had its beneficial effects. Recent developments include circular chart recorders that use blank circular chart paper, printing the scales and traces for up to four channels, as recording progresses.

One of the newer types uses heat-sensitive paper on which a single stylus printhead produces up to four analog traces and also prints alphanumeric chart data and time lines. The instrument is user configurable via English language prompts. Another unit can be programmed to print the equivalent of 10-inch, 11-inch, or 12-inch diameter charts. It is fully configurable from a PC or a keypad and prints straight time lines (not curved, as conventional recorders have used for many years).

Other capabilities that the microprocessor has brought to the circular-chart recorder include:

- Ability to handle a variety of inputs such as all common types of thermocouples, RTDs, dc voltages and currents;
- Control outputs (including PID) and adjustable alarms settings;
- Flow totalisation, including data logging of totals; and
- Retransmission options.

One model, custom-designed for measuring and recording temperature and relative humidity, uses a double-sided, 200-mm (8-inch) chart. Available chart capacities are 1, 7, or 32 days. Here again, an embedded microprocessor affords capabilities such as digital display with user-selectable read-out of temperature in ^{0}F or ^{0}C, relative humidity, and alarm setpoints.

7.8 *VIDEOGRAPHIC RECORDERS*

Evidence of the many benefits of videographic, or paperless, recorders are the large number of suppliers that now offer these types of instruments. Some are relative newcomers to the recorder field, utilizing their expertise in electronic technology to meet user needs. One example is the recent offering of a circular chart videographic recorder with a 10-inch screen display-equivalent in size to the common paper chart versions.

Display technology has been getting better and better in sharpness of detail, with a wide range of bright colours, similar to the proven quality of a newer PC. Some models offer the flexibility of plotting records horizontally or vertically. A common screen size on such recorders is small-on the order of 76 mm wide by 38 mm high (3 in. x 1.5 in.).

Some suppliers offer a 5-inch wide display area or larger. Most are designed for panel mounting and come in a size as small as 1/4 DIN. In addition to displaying trend records or bar graphs, the instruments can be configured at will to provide large digital displays of the measurements in engineering units or small digital readings along with the trend curves.

These recorders have some form of storage media-either a 3-1/2 inch floppy disk or a PCMCIA card-which continuously stores all the measurement inputs as well as the recorder configuration. From such stored data, an operator can use a touchscreen or membrane keypad to provide such display functions as: a split screen (two trend windows), zooming, faceplate displays, alarm summaries, and selected trends (such as six out of up to 24 inputs).

Figure 7.7. Videographic recorder with magnetic storage media (PCMCIA card or floppy disk)

The recorder's disk or card can be removed and used to download all its accumulated data to a PC that has the necessary software. An engineer can bring up any of the stored data for review and analysis, zooming in on a desired time, say, of a process upset (as well as just before and just after the event).

Videographic recorders may include a serial port that permits data to be sent, for example, over an RS-422/485 communication network directly to a PC or DCS with suitable software. This can be two-way communication. Thus, an operator at a PC in a control room could monitor the recorder readings in real-time, call for specific data, and even reconfigure the recorder.

7.9 RECORDER SPECIFICATIONS

The following are general features and specifications that should be considered when examining a recorder:

1. The type of writing instrument desired such as pen and ink, heated stylus, pressure stylus or electric discharge.
2. The type of recording chart desired, such as paper or wax-coated paper; or photographic film which is self-developing or requires darkroom processing.
3. Maximum amplification of the signal to be recorded
4. Frequency response expressed in cycles per second (Hertz)
5. Recording speed expressed in seconds for full-scale deflection.
6. Input signal's voltage or current.
7. Input impedance.
8. Chart speed expressed in inches or centimetres per second.
9. Input circuit. The input terminals should be *floating* (i.e. isolated from the ground.
10. Number of recordings (available from single channel to approximately 40 channels)
11. Channel width expressed in inches, centimetres or millimetres.
12. Drift expressed in inches or millimetres per hour.
13. Power requirements
14. Mounting requirements –fixed, vertical or horizontal
15. Weight and physical dimensions.

7.10 *COMPUTER AIDED INSTRUMENTATION*

7.10.1 *The General purpose computer*

A Computer is made of number of electronic and electromechanical devices. It consists of four basic parts (or subsystems).

- ✓ Central Processing Unit (CPU)
- ✓ Storage device
- ✓ Input/output (I/O) devices
- ✓ Bus interface

Fig. 7.8 shows the block diagram of a general-purpose computer system. All the physical components, as shown in the figure, are termed as the *hardware*. Set of instructions is required to guide these physical devices to perform specified functions, which are called *program* or *software.* The software instructs the hardware components to work. Without software, hardware simply cannot work.

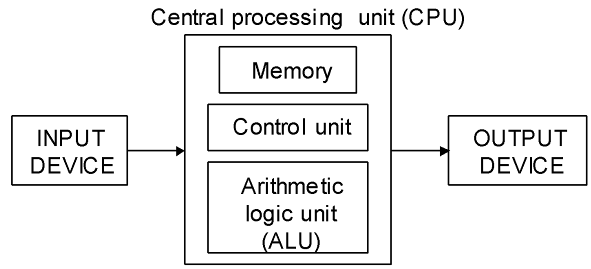


Fig. 7.8 Block diagram of general purpose computer

The *central processing unit (CPU),* also called the heart of computer, consists of control unit, arithmetic logic unit (ALU), temporary memory (primary storage) and general purpose registers. The control unit continually supervises and controls the operations/activities within the CPU. It fetches program instructions from main memory, decodes the instructions and sets up the necessary data paths and timing cycle for the execution of the instructions.

The *arithmetic logic unit (ALU)* contains the circuits necessary to carry out arithmetic and logical operations such as addition, subtraction, comparison, multiplication, division, etc. The main memory stores data and information, programs, etc. The general-purpose registers are used for storing data temporarily while it is being processed. Today, computers have CPUs with several general-purpose registers and hence, frequent access to main memory is not required thereby giving faster processing speed.

Computer fetches data from *primary (main) memory* under the command of control unit. It places these in one of the registers, performs arithmetic operations using ALU based on application program requirements, saves them temporarily in accumulator (a type of register), and transfers it finally to primary (main) memory storage after calculations/operations are over. The processed data is further transferred to Input/Output (I/O) devices as per the requirements of application program.

7.10.2 ***The computer based data logger***

As an example of the application of microprocessors in instrumentation, the use of a microprocessor in data logging can be examined. A typical data-logging system can use varied types of signal conditioners, amplifiers, and output devices. A printer is most commonly employed in the logging of slowly-varying data. A certain amount of processing of data may be involved in certain cases, such as linearization of the thermocouple output, computation of the average value of a certain set of data inputs, or conversion to engineering units. In addition, the comparison of the measured data with preset minimum and maximum limits, and actuation of out-of-limit indicators or initiation of alarms can also be carried out. Eventually, the system can also be made to sequentially shut down energisation systems, when maximum tolerable limits of working systems are reached. An exemplifying requirement of a data-logging system which is to collect data periodically from 100 locations can be the following:

(a) The provision to collect load information from 60 locations. Transducer used: load cell. The load information from each individual station is to be measured at least once in every five seconds. The measured data with the time at which measurement is made should be printed out as well as stored for analysis later.
(b) The provision to monitor temperature at twenty points. Temperature range: 0 to 1000°C. Transducer used: K-type thermocouple. Output data are to be linearised to provide direct display of temperature to within ±0.1% ±1°C. Visual indicators are to be actuated when the set temperature limits (settable for each station) are exceeded. Reading rate: once in every 5 minutes.
(c) The provision to monitor twenty control point voltages. Aural alarms are to be set when any of them falls out of the range of ±5 V. Simultaneous visual indication of the particular station where the voltage is out of range is to be displayed.

A data logger of this nature can be efficiently handled by a microprocessor-based system. The system can be designed in such a way that a built-in clock is used to function as a "real time" clock so that each measurement is logged along with the time at which it is done. Invariably, the channel number identifying the channel on which measurement is made can also be logged. In the case of the load-cell measurement, balancing of the individual strain gauges can also be done by the processor. The processor can do this by calling sub-routines at the start of logging by noting the outputs of the load cells when no load is applied and storing these for later subtraction from individual measurements (auto zero). The rate of reading (e.g., the necessity to take readings of the load in every five seconds) can be tailored to the requirement. The processor can command low-level or high-level multiplexers at appropriate times to access data from a desired channel, convert the data into digital values, store these in RAMs within the system, and proceed to log data from other locations. It can efficiently utilize the time available to it for doing the various functions by commanding the RAM data to be stored in cassettes in a serial form after adding parity and synchronization bits.

For the monitoring and display of temperature, the thermocouple output will be amplified and linearised by look-up tables stored within the system in read only memories. The linearised output can be transferred to output ports that control the displayed digits. Constant comparison can be made of the temperature with the limit values set in the operating console or fed to the microprocessor system at the start of a process. Visual annunciators can be actuated based on this comparison, when the limits are exceeded. The monitoring of the other twenty voltage levels corresponding to the control functions can also be carried out in much the same way.

This data-logging system can obviously have incorporated within it programmable data amplifiers and multiplexers to handle the inputs. It can also have a fast ADC for conversion of the analog inputs. Its output device can be a teleprinter (TTY) or a line printer that will provide the print-out of the data. It can have an operating console with keyboards for the entry of necessary data and commands, a panel display that displays the temperature of the various stations monitored by it, and visual and other alarms, It can also have a cassette interface for recording the "load cell" data from the 60 stations. The block schematic of the system is shown in Fig. 7.9. While the microprocessor has to necessarily operate on instructions sequentially, the fast speed at which it operates enables it to access, convert, compute, display and record information from various stations. It is thus ideally suited for logging of many parameters and to take intelligent decisions based on computations carried out on these measured data.

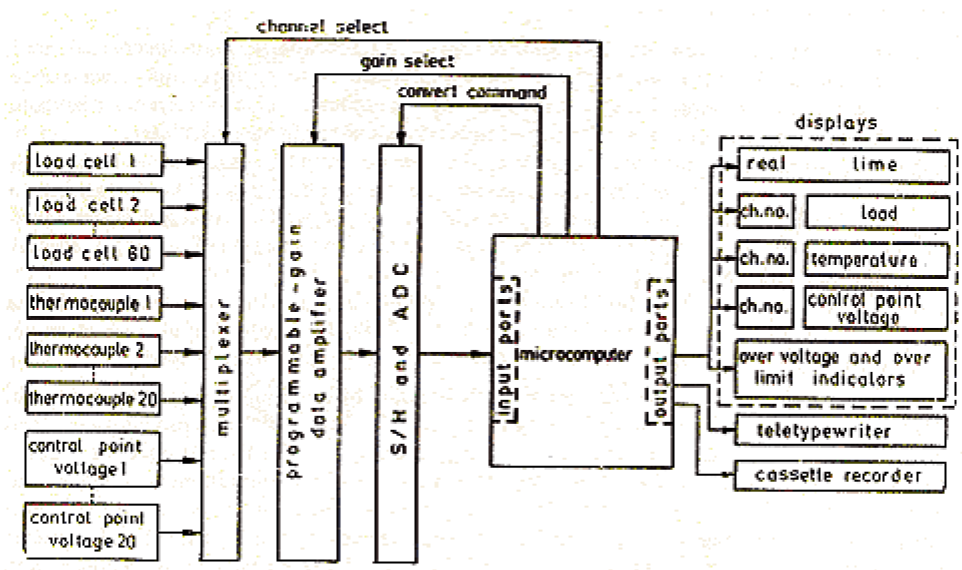


Fig 7.9 Computer- based data logger