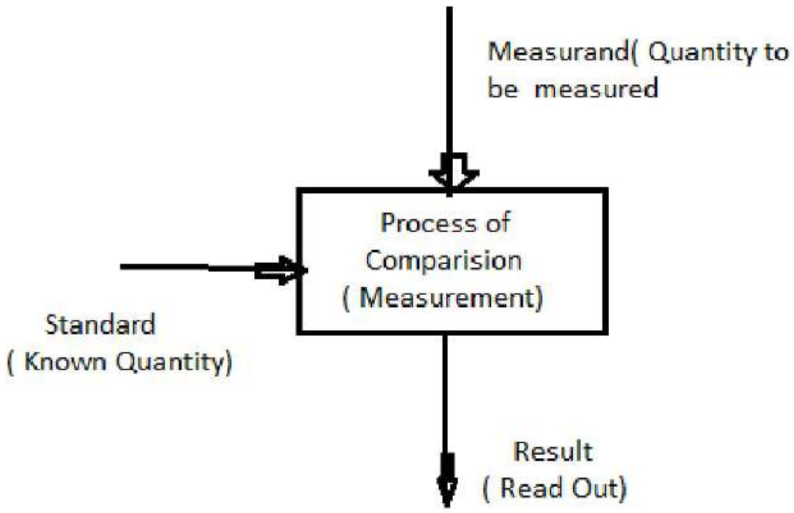
WHAT IS MEASUREMENT?

Measurement is the process of comparing unknown magnitude of certain parameter with the known predefined standard of that parameter. For instance, if we have to measure the temperature of the body, we measure it with the thermometer that has predefined scale indicating different values of the temperature. If we have to measure the length of the wall, we measure it with the measuring tape that has predefined markings on it. The measurement enables us obtaining magnitude of certain parameters whose value is not known by comparing them with the standards whose value is predefined.



PROCESS OF MEASUREMENT

INTRODUCTION & REQUIREMENTS OF MEASUREMENT-

For the measurements results to be accurate, two conditions should be met. Firstly, the standard which is used for comparison must be defined accurately and it should be universally accepted. For instance the weight cannot be just light or heavy. It should be light or heavy in comparison to some standard weight and should be measured accurately against it. The comparison of the unknown magnitude should be made with the recognized standard and it should produce some meaningful reading of the value.

The second important condition to be met for measurements is that the procedure applied for the measurements should be provable and there should be provable instruments for measurements. This means the methods for making the measurements and the instruments used for them should be reliable enough to make the correct measurements.

MEASUREMENT : A method to obtain information regarding the physical values of the variable.

- Definition —the quantitative comparison between a predefined standard and a measurand to produce a measured result

Measurement is the process of comparing unknown magnitude of certain parameter with the known predefined standard of that parameter.

GENERALIZED MEASUREMENT SYSTEM:-

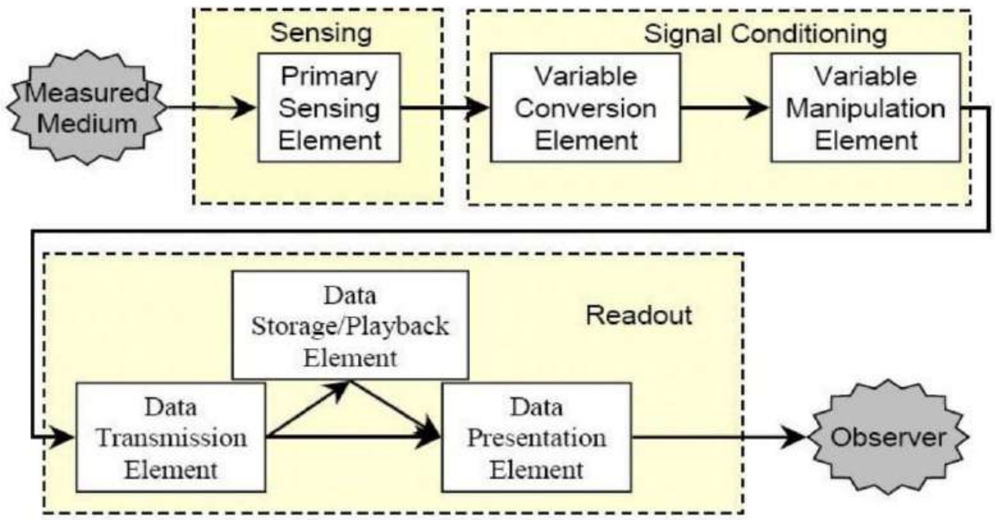
In the direct method of measurement the physical quantity like length or mass are measured directly by the measuring instruments. The indirect method of measurement comprises of various stages for the measurement of the physical quantity like temperature, pressure, force etc, since they cannot be measured by the direct instruments. In this method the transducer is used which is connected to a host of other instruments to convert one form of energy that cannot be measured into the other form that can be measured easily. The input and the output values are calibrated so that for all the value of output the value of the input can be calculated.

The General System of Measurement comprises of three stages (see the fig1), these are:

1. **I) First stage - the detector-transducer stage.**

2. **II) Second stage – Intermediate modifying stage**

3. **III) Final stage – Terminating stage comprising of: indicator, recorder,some controller as individuals or in combination.**



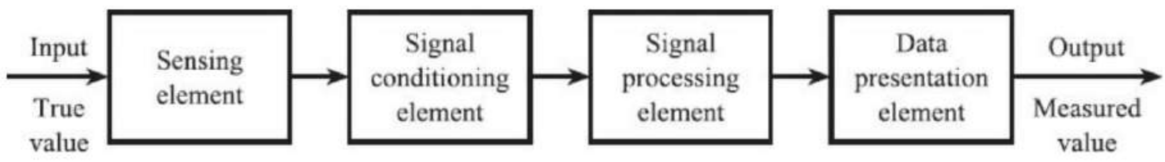


Fig.1 Generalized Measurement Systems

i) **Detector Unit: Primary Stage** : The main function of the first stage is to detect or sense the input signal. The input signal is usually the physical quantity that has to be measured. It can be temperature, pressure, force, velocity, or any other quantity that has to be measured. The devise most commonly used for detecting the input signal is the transducer or sensor. The transducer can detect the input quantity that has to be measured but cannot be measured directly by the instruments. It has the ability to convert these signals into analogous output that can be measured easily.

The detector, which is transducer, should be able to sense the input signals that are to be measured, but at the same time it should be insensitive to the other possible types of inputs.

For example if the pressure measuring transducer like bellow is being used, it should be sensitive to the pressure, but insensitive to the acceleration or other disturbances at the input. Similarly, if strain gauges are being used, they should be sensitive to temperature only. In actual practice it is very rare that the transducer is sensitive only to the signals that are to be measured.

II) Second Stage: Intermediate Modifying Stage

Let us continue with the second stage of general measurement system. The second stage or the intermediate modifying stage converts the input signal in the form that can be used easily by the final stage to indicate the value of the input physical quantity. The modifying stage may change the type of the input signal so that the output value can be measured easily. Or it may increase the amplitude and/or the power of the input signal to the level so as to drive the final terminating devices.

The intermediate modifying stage should be designed such that there is proper matching of characteristics between the first stage and second stage and the second stage and the final stage of measurement so that the output obtained is analogous to the input. The intermediate may also have to perform the functions of filtering the unnecessary input, and also integration, differentiation, telemetering etc, wherever required.

III) Final Stage: Terminating Stage (output stage)

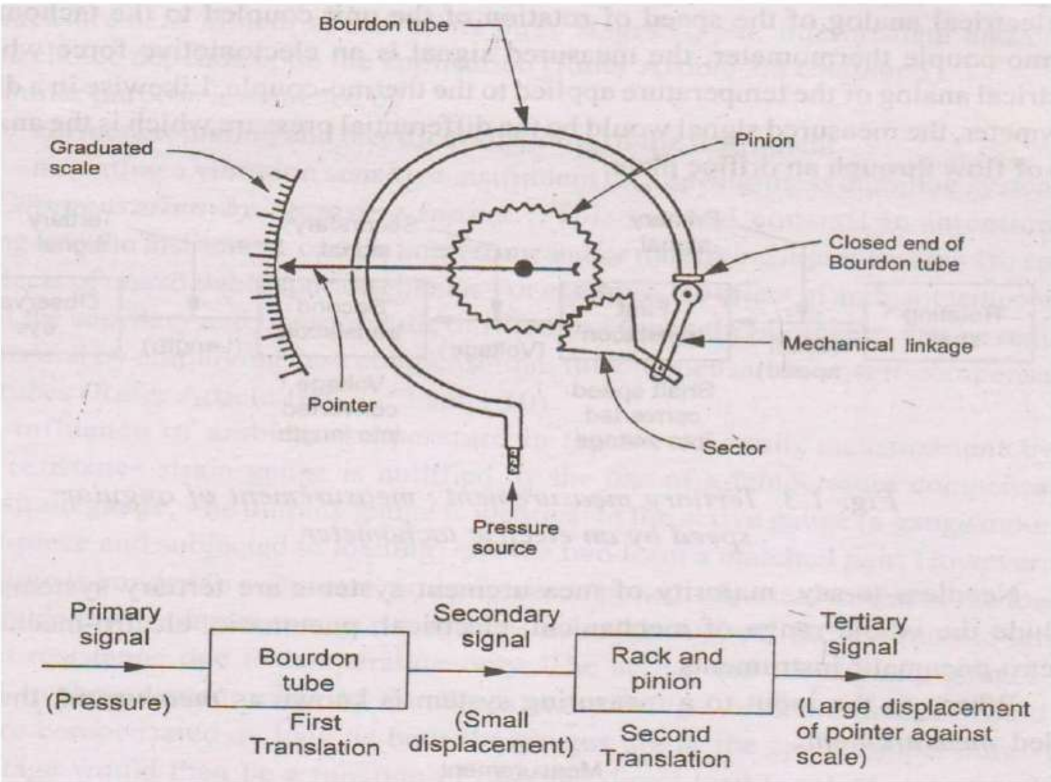
The final stage or the terminating stage provides the information about the input physical quantity in the form that can be easily read by the human beings or the controller. If the final output is intended to be recognized by the human beings it can be obtained in the form of the pointer movement on the predefined scale, in the digital format, by the graph etc. If the output is meant to perform some automatic function, it can be given to the controller that controls the operation of the next unit.

The output obtained at the final stage is calibrated against the input quantity so that the value shown by the pointer or the digital indicator is analogous to the input value. Thus the physical quantities like temperature, pressure, force etc which cannot be measured directly are measured with the help of the transducer in the form of pointer movement on the scale or the digital output.

EXAMPLE OF THE GENERAL MEASURMENT SYSTEM

Here the Bourdon tube senses the pressure, and it acts as the transducer that detects the quantity to be measured.

Let us come back to our previous example of measuring the pressure by using Bourdon tube. Here, the coiled oval shaped tube acts as the intermediate stage. When the pressure is applied at the inlet the oval tube tends to become circular, but inner and the outer diameters of the coil tend to remain the same. Due to this the coil tends to uncoil producing the angular motion of the tip of the coil, which is connected to the final stage of the pointer that indicates the value of the applied pressure.



The sealed tip of the coil is connected to the pointer via linkages and the gear arrangement. The pointer moves over the predefined scale that indicates the value of the pressure. When the pressure is applied the tip of the Bourdon tube uncoils, which moves the linkages and the gear that finally produce the rotary motion of the pointer on the scale indicating the actual value of the applied pressure.

APPLICATION OF MEASUREMENTS

Measurements are one of crucial parts of not only mechanical engineering but all types of engineering fields. Every branch of engineering involves two processes:

- **Monitoring and operation of process.**
- **Control of a process**.
- **Experimentation:**

- **Testing and performance operation**
- **Verification of properties or theory**
- **Information needed for analysis**

e.g. Checking or evaluation of:

Oil viscosity variation with temp.

Pump performance curve

piping head loss

Lift and drag of new airfoil shape…….etc.

Design, and operations and maintenance. The design may be machine design, building design, circuit design, transportation design, automobile design etc. The operations part involves operation of the machines, automobiles, various plants, circuits etc.

Both, the design, and operations and maintenance involve measurements. For instance while designing automobile we have to consider dimensions of various parts of the automobiles, the loads they can pick up etc. Similarly during the operations of the plant, say like industrial refrigeration plant, we have to measure parameters like pressure, temperature, etc.

In the power plant we have to measure

- Various quantities of the coal.
- The quantity of water in the boiler.
- The amount of steam produced along with its flow rate.
- Temperature and pressure.
- The amount of power produced.
- The outlet temperature of the steam from condenser etc.

Other field of application:

- Home
  - Thermometer
  - Barometer

    - Watch
- Road vehicles
    - speedometer
    - fuel gauge
- Industry
    - Automation
    - Process control
    - Boiler control

Types of Measurement

There are two methods of measurement:

1) Direct comparison with the standard

2) In-Direct comparison with the standard.

Both the methods are discussed below:

1.) Direct Comparison with the Standard

In the direct comparison method of measurement, we compare the quantity directly with the primary or secondary standard. Say for instance, if we have to measure the length of the bar, we will measure it with the help of the measuring tape or scale that acts as the secondary standard. Here we are comparing the quantity to be measured directly with the standard.

Even if you make the comparison directly with the secondary standard, it is not necessary for you to know the primary standard. The primary standards are the original standards made from certain standard values or formulas. The secondary standards are made from the primary standards, but most of the times we use secondary standards for comparison since it is not always feasible to use the primary standards from accuracy, reliability and cost point of view. There is no difference in the measured value of the quantity whether one is using the direct method by comparing with primary or secondary standard.

The direct comparison method of measurement is not always accurate. In above example of measuring the length, there is limited accuracy with which our eye can read the readings, which can be about 0.01 inch. Here the error does not occur because of the error in the standards, but because of the human limitations in noting the readings. Similarly, when we measure the mass of any body by comparing with some standard, it's very difficult to say that both the bodies are of exactly the same mass, for some difference between the two, no matter how small, is bound to occur. Thus, in direct method of measurement there is always some difference, however small, between the actual value of the quantity and the measured value of the quantity.

2) In-Direct Method of Measurement

There are number of quantities that cannot be measured directly by using some instrument. For instance we cannot measure the strain in the bar due to applied force directly. We may have to record the temperature and pressure in the deep depths of the ground or in some far off remote places. In such cases indirect methods of measurements are used.

In the indirect method of measurements some transducing devise, called transducer, is used, which is coupled to a chain of the connecting apparatus that forms the part of the measuring system. In this system the quantity which is to be measured (input) is converted into some other measurable quantity (output) by the transducer. The transducer used is such that the input and the output are proportional to each other. The readings obtained from the transducer are calibrated to as per the relations between the input and the output thus the reading obtained from the transducer is the actual value of the quantity to be measured. Such type of conversion is often necessary to make the desired information intelligible.

The indirect method of measurements comprises of the system that senses, converts, and finally presents an analogues output in the form of a displacement or chart. This analogues output can be in various forms and often it is necessary to amplify it to read it accurately and make the accurate reading of the quantity to be measured. The majority of the transducers convert mechanical input into analogues electrical output for processing, though there are transducers that convert mechanical input into analogues mechanical output that is measured easily.

STATIC & DYNAMIC CHARACTERISTICS OF INSTRUMENTS:

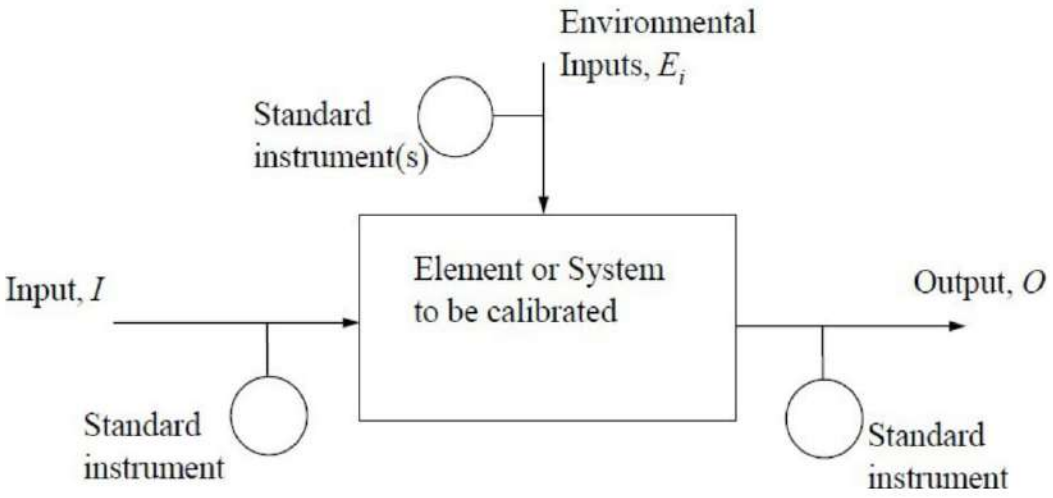
Characteristics of measurement systems

The system characteristics are to be known, to choose an instrument that most suited to a particular measurement application. The performance characteristics may be broadly divided into two groups, namely '**static**' and '**dynamic**' characteristics.

STATIC CHARACTERISTICS- the performance criteria for the measurement of quantities that remain constant, or vary only quite slowly.

DYNAMIC CHARACTERISTICS- the relationship between the system input and output when the measured quantity (measurand) is varying rapidly

calibration:



General diagram for calibration

STATIC AND DYNAMIC CHARACTERISTICS -.

- Instrument systems are usually built up from a serial linkage of distinguishable building blocks.
- The actual physical assembly may not appear to be so but it can be broken down into a representative diagram of connected blocks.
- The sensor is activated by an input physical parameter and provides an output signal to the next block that processes the signal into a more appropriate state.
- A fundamental characterization of a block is to develop a relationship between the input and output of the block.
- All signals have a time characteristic.
- It is essential to consider the behaviour of a block in terms of both the static and dynamic states.
- The behaviour of the static regime alone and the combined static and dynamic regime can be found through use of an appropriate mathematical model of each block.

CLASSIFICATION OF INSTRUMENTS

The classifications of instruments are given below

1. Automatic and manual instruments: -

The manual instruments require the services of an operator while the automatic types do not. For example, the temperature measurements by mercury-in-glass thermometer is automatic as the instrument indicates the temperature without requiring any manual assistance. However, the measurement of temperature by a resistance thermometer incorporating a wheatstone bridge in its circuit is manual in operation as it needs an operator for obtaining the null position.

Automatic instruments are proffered because of their fast dynamic response and low operational cost.

2. Self-generating and power-operated instruments:-

In a self-generating instrument, the output entirely or almost entirely by the input signal. The instrument Does not require any outside power in performing its function. For example, the motive power in a mercury-in-glass thermometer is supplied wholly by the thermal expansion of mercury and as such it constitutes a self-operated device. Likewise, the dial indicator is an active device since all the energy required to operate the instrument is furnished by the system whose displacement is being measured.

Some other common examples of active instruments are:-

exposure meter of a camera which is essentially a photo voltaic cell.

- bourdon gauge for the measurement of pressure.
- pitot-tube for the measurement of velocity.
- tacho-generator for rotational speed measurement.

Some instruments require some auxiliary source of power such as compressed air, electricity, hydraulic supply etc. For there operation and hence are called externally powers instruments. In these devices, the input signal supplies only an insignificant portion of the output power. For example, in the electro-mechanical measurement system an exciter has been incorporated to supply electric energy for the transducer. Likewise, the digital revolution counter is a passive device. The power to drive the solenoid comes from the a.c. power lines and not from the rotating shaft.

Some other common examples of passive instruments are:-

- · lvdt used in the measurement of displacement, force, pressure.
- · voltage-dividing potentiometer which converts rotation or displacement into potential difference.
- · photo-conductive transducer which translate light information to resistance information.
- · strain-gauge load cell using wheat stone bridge circuit.
- · resistance thermometers and thermisters.

3. Self-contained and remote indicating instruments: -

The different elements of a self-contained instrument are contained in one physical assembly. In a remote indicating instrument, the primary sensing element may be located at a sufficiently long distance from the secondary indicating element. In the modern instrumentation technology, there is a trend to install remote indicating instruments where the important indications can be displayed in the central control rooms.

4. Deflection and null output instruments:-

In null-type instruments, the physically effect caused by the quantity being measured in nullified by generating an equivalent opposing effect. The equivalent null causing effect then provides a measure of the unknown quantity.

A deflection type instrument is that in which the physical effect generated by the measuring quantity is noted and correlated to the measurand.

A distinction between the null and deflection mode of operation can be made by considering the working of a pan balance and a platform scale. In the pan balance the unknown weight is placed of one pan of the balance and weights of known value are placed in the other pan until a balance condition is indicated by zero or null position of the pointer.

In a platform scale, i.e. the deflection mode, the weight of an object placed on the platform of the scale is indicated by the relative displacement between the pointer and a dial.

The null-type devices are slow in operation, have poor dynamic response but are more accurate and sensitive, and do not interfere with the state of the quantity being measured. Deflection instruments are simple in construction and operation, have good dynamic response. However, they interfere with the state of measurand and as such do not determine its exact state/value/condition.

5. Analog and digital instruments: -

The signals of an analog unit vary in a continuous fashion and can take on infinite number values in a given range. Wrist watch, speedometer of an automobile, fuel gauge, ammeters and voltmeters are examples of analog instruments.

Signals varying in discrete steps and taking on a finite number of different values in a given range are digital signals and the corresponding instruments are of digital type. For example, the timers on a scoreboard, the calibrated balance of a platform scale, and odometer of an automobile are digital instruments.

The digital instruments convert a measured analog voltage into digital quantity which is displayed numerically, usually by neon indicator tubes. The output may either be a digit for every successive increment of the input or be a coded discrete signal representative of the numerical value of the input. The digital devices have the advantage of high accuracy high speed and the elimination of human operational errors. However, these instruments are unable to indicate the quantity which is a part of the step value of the instrument. The importance of the digital instrumentation is increasing very fast due to the applications of the digital computers for data handling, reduction and in automatic controls. Apparently it becomes necessary to have both analog-to-digital converters at input to the computers and digital-to-analog converters at the output of the computers.

STANDARDS & ITS TYPES:

Standards are the fundamental reference for a system of weights and measures, against which all other **measuring** devices are compared.**standards are subdivided into following four categories or grades:**

1. Primary Standards (Reference Standards).
2. Secondary Standards (Calibration Standards).
3. Tertiary Standards (Inspection Standards).
4. Working Standards (Workshop Measuring Standards).

1. Primary Standards (Master Standards):

The primary standard is also known as Master Standard, and is preserved under the most careful conditions. These standards are not commonly in use. They are used only after long internals. They solely used for comparing the secondary standards. Sometimes it is also called Reference Standards.

2. Secondary Standards (Calibration Standards):

The secondary standard is more or less similar to the primary standard. They are nearly close in accuracy with primary standards. The secondary standard is compared at regular intervals with primary Stands and records their deviation. These Standards are distributed to a number of places where they are kept under safe custody. They are used occasionally for comparing the territory standards.

3. Tertiary Standards (Inspection Standards):

The Tertiary standard is the first standard to be used for reference purpose in workshops and laboratories. They are used for comparing the working standards. These are not used as frequently and commonly as the working standards but more frequency than secondary standards. Tertiary standards should also be maintained as a reference for comparison at intervals for working standards.

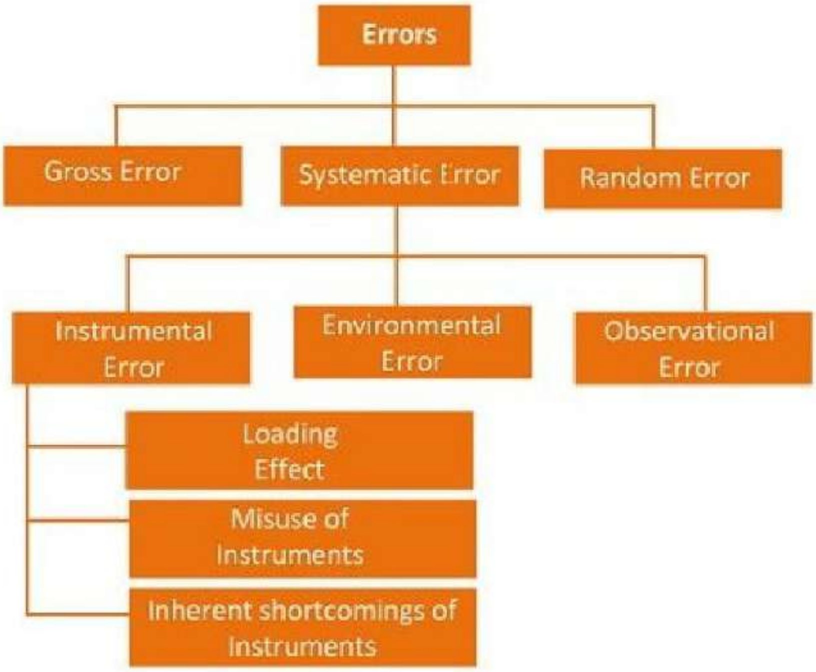
4. Working Standards (Workshop Measuring Standards):

The working standard is used for actual measurement in workshop or laboratories by the workers. These standards should also be as accurate as possible to the tertiary standard. But sometimes, lower grades of materials can be used for their manufacturing to reduce cost.

TYPES OF ERRORS IN MEASUREMENT

The error may arise from the different source and are usually classified into the following types. These types are

1. **Gross Errors**
2. **Systematic Errors**
3. **Random Errors**



Their types are explained below in details.

1. Gross Errors

The gross error occurs because of the human mistakes. For examples consider the person using the instruments takes the wrong reading, or they can record the incorrect data. Such type of error comes under the gross error. The gross error can only be avoided by taking the reading carefully.

For example – The experimenter reads the 31.5ºC reading while the actual reading is 21.5Cº. This happens because of the oversights. The experimenter takes the wrong reading and because of which the error occurs in the measurement.

Such type of error is very common in the measurement. The complete elimination of such type of error is not possible. Some of the gross error easily detected by the experimenter but some of them are difficult to find. Two methods can remove the gross error.

Two methods can remove the gross error. These methods are

- The reading should be taken very carefully.
- Two or more readings should be taken of the measurement quantity. The readings are taken by the different experimenter and at a different point for removing the error.

2. Systematic Errors

The systematic errors are mainly classified into three categories.

1. Instrumental Errors
2. Environmental Errors
3. Observational Errors

2 (i) Instrumental Errors

These errors mainly arise due to the three main reasons.

(a) Inherent Shortcomings of Instruments – Such types of errors are inbuilt in instruments because of their mechanical structure. They may be due to manufacturing, calibration or operation of the device. These errors may cause the error to read too low or too high.

For example – If the instrument uses the weak spring then it gives the high value of measuring quantity. The error occurs in the instrument because of the friction or hysteresis loss.

(b) Misuse of Instrument – The error occurs in the instrument because of the fault of the operator. A good instrument used in an unintelligent way may give an enormous result.

For example – the misuse of the instrument may cause the failure to adjust the zero of instruments, poor initial adjustment, using lead to too high resistance. These improper practices may not cause permanent damage to the instrument, but all the same, they cause errors.

(c) Loading Effect – It is the most common type of error which is caused by the instrument in measurement work. For example, when the voltmeter is connected to the high resistance circuit it gives a misleading reading, and when it is connected to the low resistance circuit, it gives the dependable reading. This means the voltmeter has a loading effect on the circuit.

The error caused by the loading effect can be overcome by using the meters intelligently. For example, when measuring a low resistance by the ammeter-voltmeter method, a voltmeter having a very high value of resistance should be used.

2 (ii) Environmental Errors

These errors are due to the external condition of the measuring devices. Such types of errors mainly occur due to the effect of temperature, pressure, humidity, dust, vibration or because of the magnetic or electrostatic field. The corrective measures employed to eliminate or to reduce these undesirable effects are

- The arrangement should be made to keep the conditions as constant as possible.
- Using the equipment which is free from these effects.
- By using the techniques which eliminate the effect of these disturbances.
- By applying the computed corrections.

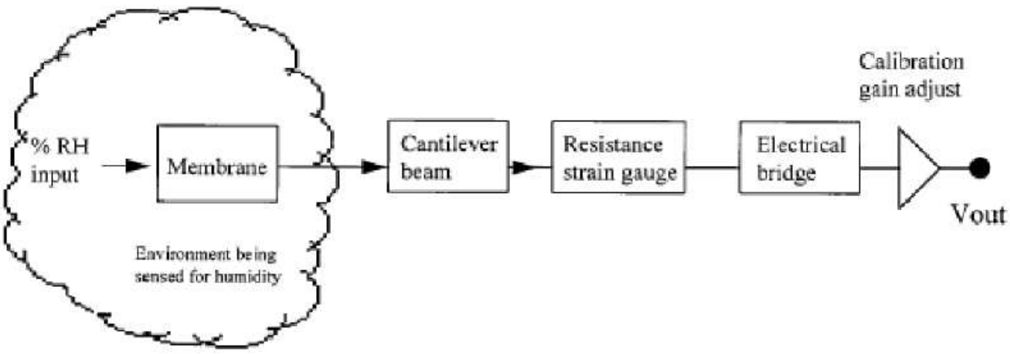
2 (iii) Observational Errors

Such types of errors are due to the wrong observation of the reading. There are many sources of observational error. For example, the pointer of a voltmeter resets slightly above the surface of the scale. Thus an error **occurs** (because of parallax) unless the line of vision of the observer is exactly above the pointer. To minimise the parallax error highly accurate meters are provided with mirrored scales.

3. Random Errors

The error which is caused by the sudden change in the atmospheric condition, such type of error is called random error. These types of error remain even after the removal of the systematic error. Hence such type of error is also called residual error.

STATIC PERFORMANCE OF INSTRUMENT-



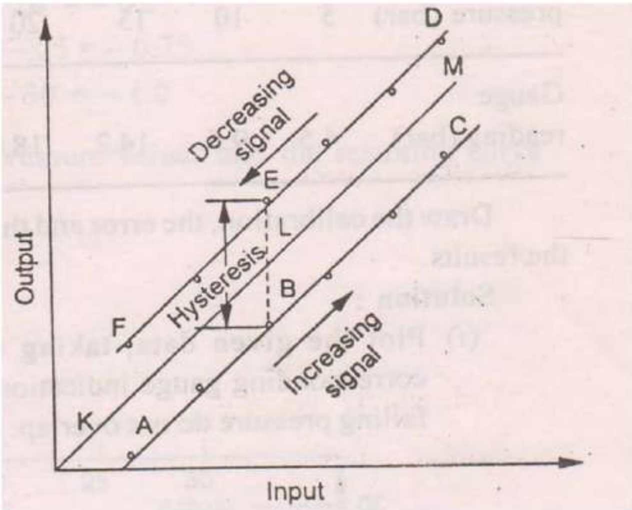
The **static characteristics** of instruments are related with steady state response.The relationship between the output and the input when the input does not change, or the input is changing at a slow rate.

CALIBRATION-Parameters

- **Range & Span**
- **Accuracy & Precision**
- **Threshold & Resolution**
- **Sensitivity**
- **Linearity**
- **Drift**
- **Hysteresis & Dead Zone**
- **Reproducibility**
- **Uncertainty**
- **Traceability**
- **Readability**

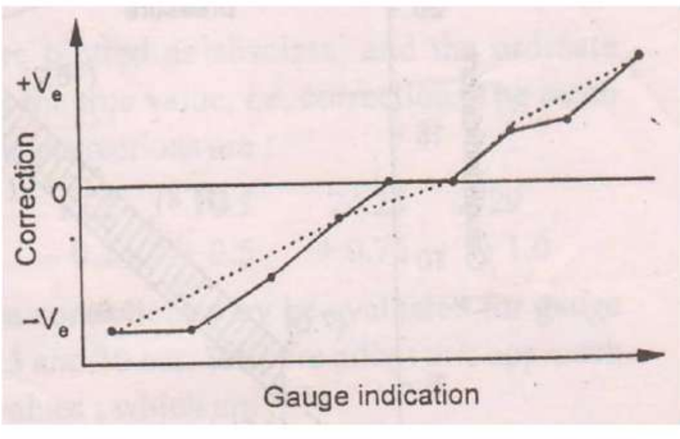
Calibration:

Calibration consists of comparing the output of the instrument or sensor (under test) against the output of an instrument of known accuracy (higher accuracy) when the same input (the measured quantity is applied to both instrument)



Calibration Curve

The procedure is carried out for a range of inputs covering the whole measurement range of the instrument or sensor Ensures that the measuring accuracy of all instruments and sensors used in a measurement system is known over the whole measurement range, provided that the calibrated instruments and sensors are used in environmental conditions that are the same as those under which they were calibrated as shown in figure.



Correction curve

Range: The region between the limits within which an instrument is designed to operate for measuring.

For a standard thermometer this is 0 to 100°C. This is the same as the full scale.

Span: The algebraic difference between higher calibration values to lower calibration values.

For above e.g. span is 100-0 = 100°C.

Accuracy - is how close the measurement to the true value is. It is more easily quantified by percentage error where:

Percentage error = (indicated value – true value) / true value x 100%

It can also be related to the percentage error of the full measuring range:

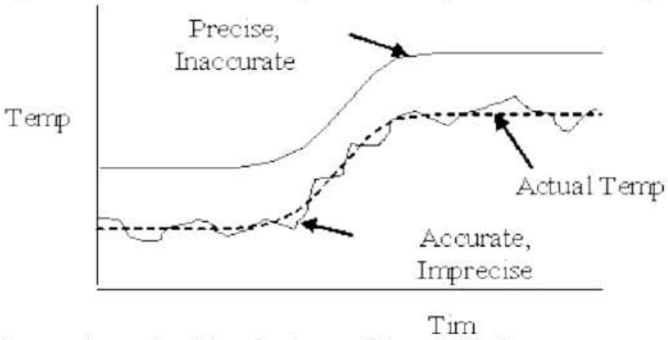
= (indicated value – true value) / Maximum scale value x 100%

Therefore, if a thermometer is defined as having an accuracy of +-2.5% with a range of 0 to 50°C, then the maximum error is 1.25°C. This error may not be a problem when measuring in the 40 to 50°C range but at the lower end, 0 to 10°C, an error of +-1.25°C is quite large if defined using the true value. It is important to select an instrument so that it is not normally operating at the extremes of its measuring range.

Precision – is the capability of an instrument to reproduce a reading.

Precision = ($X_{maximum}$ – $X_{average}$) x 100 / $X_{average}$

Precision may be defined as the reproducibility with which repeated measurements of the same



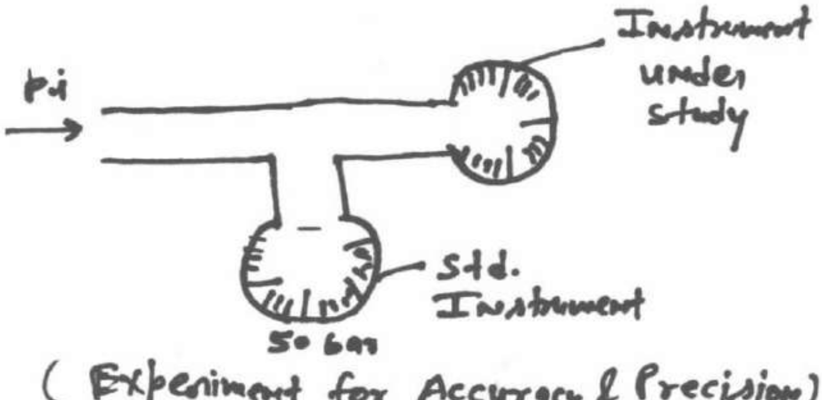
variable can be made under identical conditions. An instrument can be precise but inaccurate and, likewise, it is possible to have an accurate but imprecise instrument. See below:

In this figure pressure is applied at inlet & we measure the pressure at outlet. There are two instruments which records the pressure.

1. Standard Instrument (at 50 bar)
2. Instrument under study (shows various reading)

Let us make a graph between input & output signal & we find out as shown

For understanding the accuracy & precision we conduct an experiment on a pressure gauge as shown in figure



(Experiment for Accuracy & Precision)

Calibration values of a pressure gauge.

From reading accuracy is –

Accuracy = (indicated $value_{max}$ – true value) / true value x 100%

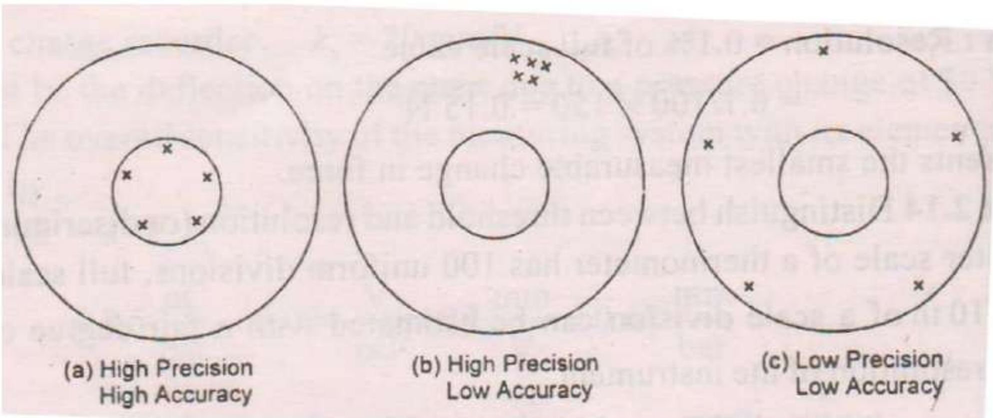
i.e (53-50) /50 x 100 % = +6%

From reading Precision is –

Precision = ($X_{maximum}$ – $X_{average}$) x 100 / $X_{average}$

i.e 53-52/52 x 100 % = +2%

Example: (game of darts)



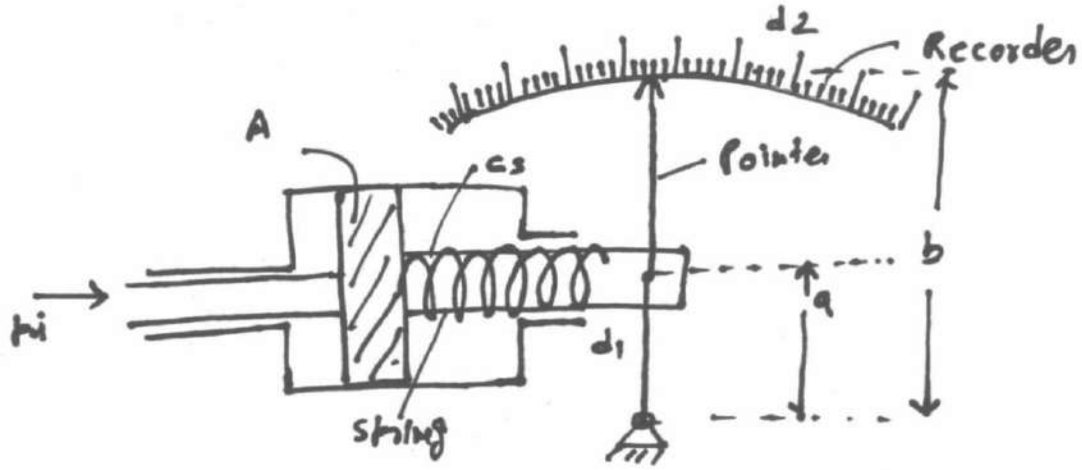
so from data accuracy i.e inaccuracy in a system is always more than precision(i.e imprecision).

Because the factor which are responsible for imprecision are lies within the instrument & factor which are lies inside & outside the instrument (like atmospheric conditions , nearby power lines , Mounting Conditions etc) . So factor responsible are more in case of inaccuracy.

Threshold & Resolution :

Threshold maybe defined as the minimum change in input signal to make the pointer move from zero value.

For understanding let us take the e.g of piston type pressure gauge as shown in figure



(Piston cylinder Pressure gauge)

where p_i = Input pressure ; $\frac{b}{a}$, leverage

$$d_2 = \frac{A}{c_s}\left(\frac{b}{a}\right) p_i$$ – Signal Relation

From figure as the pressure is increase (i.e Δp_i) from any zero value the pointer is likely starts to move with displacement d2.

Hence Δp_i is known as the threshold in system.

Resolution: Resolution maybe defined as the minimum change in input signal to make the pointer move from any Non- zero value.

Note: (Threshold value at non-zero point is called resolution & Resolution at zero point is called threshold.)

Sensitivity: Sensitivity of an instrumentation system is the ratio of magnitude of response (i.e. incremental change in output to input signal)

Sensitivity = Δ Xo/ Δ Xi

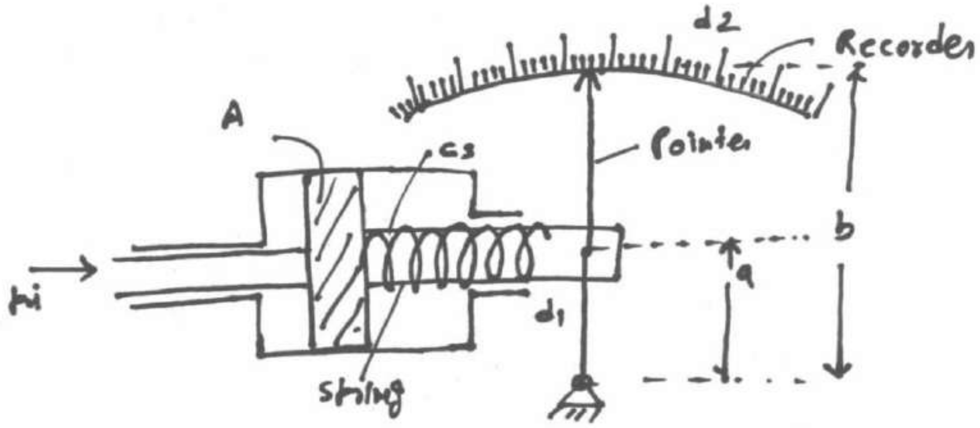
Where Δ Xo - incremental change in output
Δ Xi - incremental change in input

Note sensitivity of an instrument not a dimensionless quantity. For
Units = $\Omega/^{O}C$ (in case of Resistance Thermometer)
Units = mm/bar (In case of Pressure gauge piston type)

Overall sensitivity – a measurement system consists of a number of devices. If the sensitivities of these devices are K1, K2, K3, etc., then the overall sensitivity is the product of the individual sensitivities, K1xK2xK3xetc

For a pressure gauge as shown the sensitivity relation is given

From above figure :It is clear that sensitivity depends upon so many factors i.e



(Piston cylinder Pressure gauge)

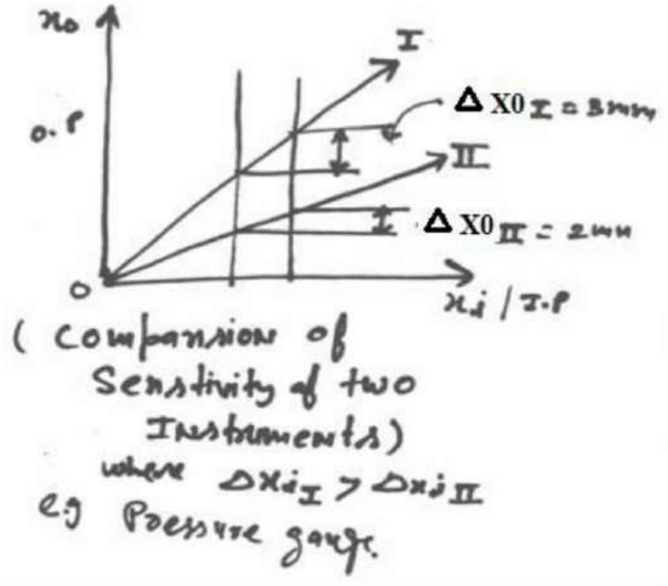
where p_i = Input pressure ; $\frac{b}{a}$, leverage

$$d_2 = \frac{A}{C_s}\left(\frac{b}{a}\right) p_i$$ – Signal Relation

1. Area of piston(A) 2. Spring Constant (Cs) 3. Leverage(b/a)

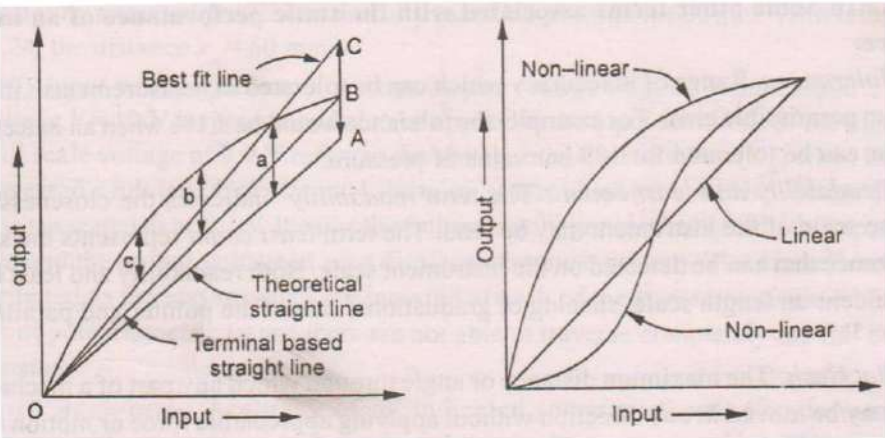
Comparison of Sensitivity of two instruments:

Let us take two instruments having sensitivity relation S1 & s2 respectively & at the same input the output change in instrument 1 is more as compared to instrument 2 as shown by graph.

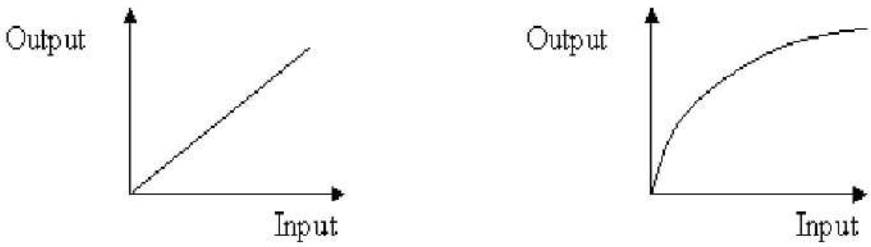


From figure it is clear that the instrument having more sensitivity shows the larger value of tangent (tanø).

Linearity: This is usually referred to as non-linearity. It is the difference between actual and ideal straight line behavior. One way to define non linearity is to divide the maximum non linearity value by the full scale deflection.



Linear **Nonlinear**



The linearity of an instrument may be defined as deviation of the reading from the linear characteristics.

Reasons of Non-Linerity :

- Non –Linear elements of instruments
- Mechanical Hysteresis
- Viscous flow or creep

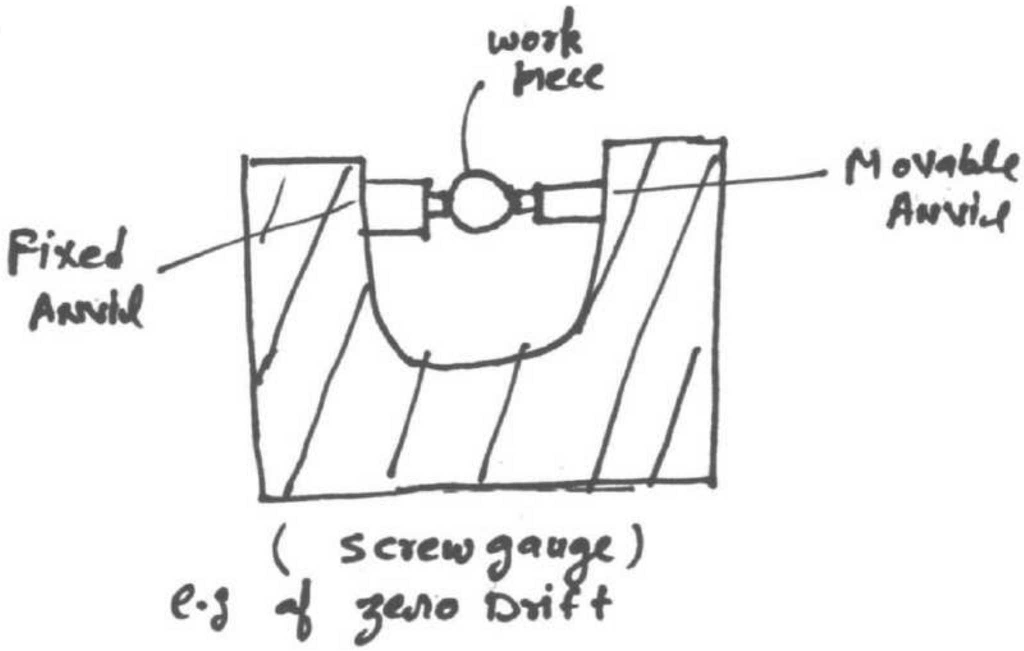
Drift: a measurement must be made with respect to a known datum or base line. It is very common and convenient to adjust the output of the instrument to zero at the datum. For example, a thermometer is set up to display zero at the freezing point of water. A pressure gauge is adjusted to read zero when open to atmosphere.

A common problem with instruments is that the output at the datum drifts and introduces an error to the measurement. All sensors are affected by drift to some extent whether it be short term or long term. Short term drift is usually associated with changes in temperature or electronics stablising. Long term drift is usually associated with aging of the transducer.

Types of Drift:

- Zero Drift
- Sensitivity drift

Zero Drift-defind as the drift from zero value during usage of an instrument Example screw gauge as shown in figure:



Sensitivity drift: occurs due to some atmospheric change is called sensitivity drift. E.g. pressure gauge:

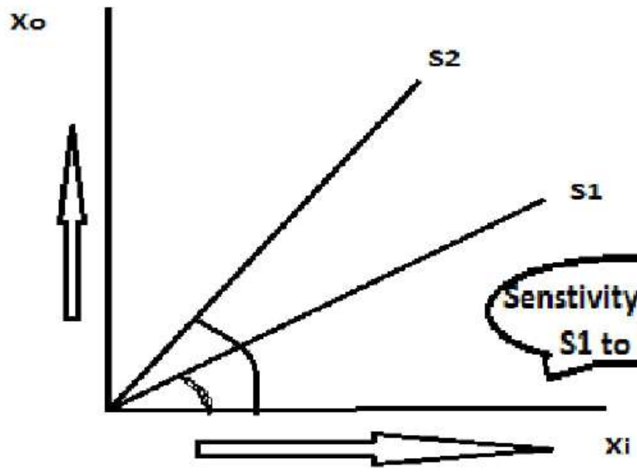
In Pressure gauge sensitivity drift calculated as :

$$\text{Sensitivity Drift} = \Delta X_o/\Delta X_i = (A/C_s) \times b/a$$

As temperature is increased the

- Cs reduces (i.e. spring becomes softer) so deflection is more.
- Friction b/w piston increase.

Hence sensitivity is shifted as shown in figure by S1 to S2

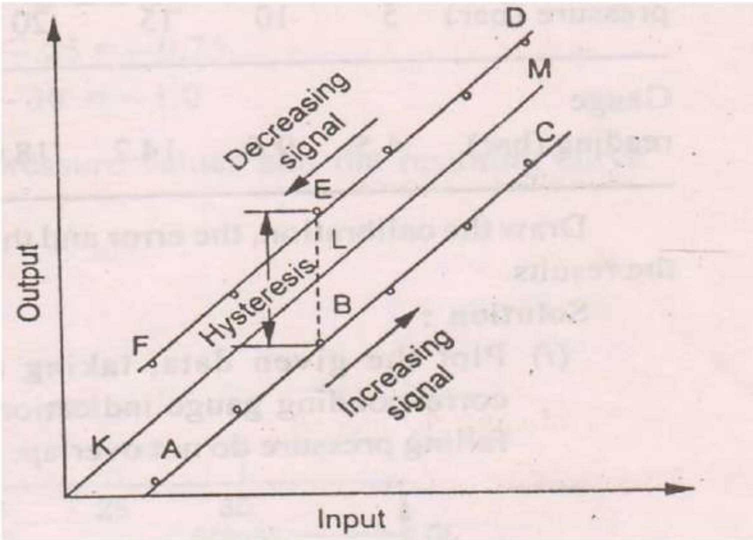


- Sensitivity drift effect the error.

Hysteresis.:

Hysteresis is the maximum difference between the same measured quantity (input signal) between the upscale & downscale readings during a full range transverse in each direction.

This refers to the situation where different readings (outputs) are sometimes observed for the same input because the input was approached from different directions. For example a thermometer exposed to an increasing temperature input (i.e. going from 0 to 100°C) may show a slightly different profile to that for the decreasing input (i.e. decreasing from 100 to 0 °C).



Hysteresis Loss

Hysteresis is often noticeable in mechanical systems where degradation of parts due to wear create slightly different results when the direction of the input is reversed. Imagine two cogs that have small gaps between the teeth due to wear. This will create hysteresis.

Dead Zone: is the largest zone through which an input signal can be varied without initiating from the indicating instrument.

- Friction & play is the direct cause of Dead band & Zone.

Backlash: The maximum distance through which one part of the instrument is moved without disturbing the other part.

Reproducibility: It is the consistency of pattern of variation in measurement. When individual measurements are carried out the closeness of the agreement between the results of measurements of the same quantity.

Uncertainty: The range about the measured value within the true value of the measured quantity is likely to lie at the stated level of confidence.

Traceability: It is nothing establishing a calibration by step by step comparison with better standards.

THE DYNAMIC CHARACTERISTICS OF ANY MEASUREMENT SYSTEM ARE:

(i) Speed of response and Response time

(ii) Lag

(iii) Fidelity

(iv) Dynamic error

(v) Overshoot

Out of the above four characteristics the Speed of Response and the Fidelity are desirable in a dynamic system, while Lag and Dynamic error are undesirable.

(i) Speed of Response and Response Time

Speed of Response is defined as the rapidity with which an instrument or measurement system responds to changes in measured quantity.

Response Time is the time required by instrument or system to settle to its final steady position after the application of the input. For a step input function, the response time may be defined as the time taken by the instrument to settle to a specified percentage of the quantity being measured, after the application of the input. This percentage may be 90 to 99 percent depending upon the instrument. For portable instruments it is the time taken by the pointer to come to rest within 0.3 percent of final scale length and for switch board (panel) type of instruments it is the time taken by the pointer to come to rest within 1 percent of its final scale length.

(ii) Measuring Lag

As discussed earlier, an instrument does not react to a change in input immediately. The delay in the response of an instrument to a change in the measured quantity is known as *measuring lag*. Thus it is the retardation delay in the response of a measurement system to changes in the measured quantity. This lag is usually quite small, but this small lag becomes highly important when high speed measurements are required. In the high speed measurement systems, as in dynamic measurements, it becomes essential that the time lag be reduced to minimum.

Measuring lag is of two types

i) **Retardation type:** In this type of measuring lag the response begins immediately after a change in measured quantity has occurred.

ii) **Time delay:** In this type of measuring lag the response of the measurement system begins after a dead zone after the application of the input.

(iii) Fidelity

Fidelity of a system is defined as the ability of the system to reproduce the output in the same form as the input. It is the degree to which a measurement system indicates changes in the measured quantity without any dynamic error. Supposing if a linearly varying quantity is applied to a system and if the output is also a linearly varying quantity the system is said to have 100 percent fidelity. Ideally a system should have 100 percent fidelity and the output should appear in the same form as that of input and there is no distortion produced in the signal by the system. In the definition of fidelity any time lag or phase difference between output and input is not included.

(iv) Dynamic Error

The dynamic error is the difference between the true value of the quantity changing with time and the value indicated by the instrument if no static error is assumed.

However, the total dynamic error of the instrument is the combination of its fidelity and the time lag or phase difference between input and output of the system.

(v) Overshoot

Moving parts of instruments have mass and thus possess inertia. When an input is applied to instruments, the pointer does not immediately come to rest at its steady state (or final deflected) position but goes beyond it or in other words overshoots its steady position.

The overshoot is evaluated as the maximum amount by which moving system moves beyond the steady state position. In many instruments, especially galvanometers it is desirable to have a little overshoot but an excessive overshoot is undesirable.

Overshoot response graph

A typical overshoot response graph can be shown as the response time stated in terms of rise time, peak percentage overshoot and settling time. Such an under damped graph in control system technology of a measuring instrument is shown in Fig. .

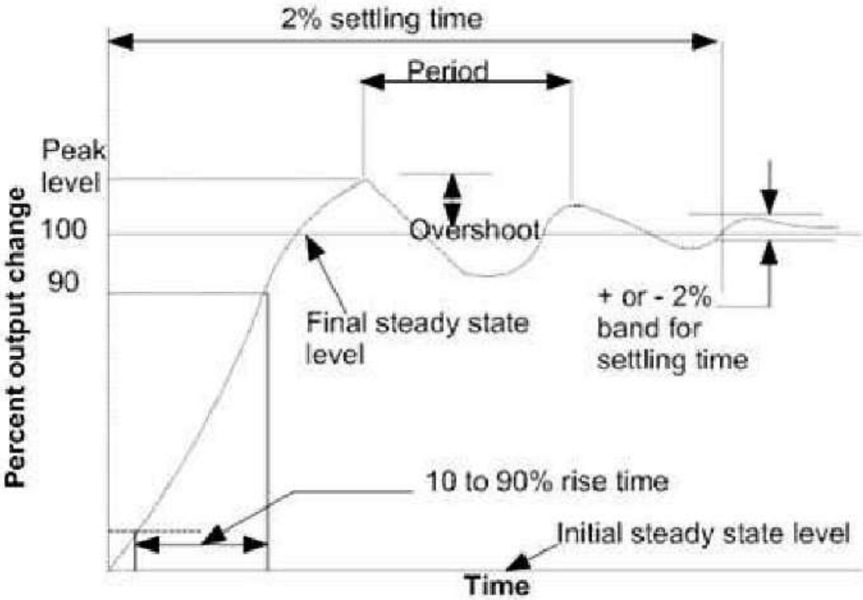


Fig. Overshoot response graph

2.4.2. Dynamic Characteristics of a Measurement System

The *dynamic characteristics* of a measurement system are:

1. Speed of response ... *Desirable*
2. Measuring lag ... *Undesirable*
3. Fidelity ... *Desirable*
4. Dynamic error ... *Undesirable*

1. **Speed of response.** The *"speed of response"* or *"responsiveness"* is defined as the *rapidity with which a measurement system responds to changes in the measured quantity.*

2. **Measuring lag.** It refers to *retardation or delay in the response of a measurement system to changes in measured quantity.*

   The lag *is caused by conditions such as capacitance, inertia, or resistance.*

   The measuring lags are of the following two types:

   (*i*) ***Retardation type lag.*** In this type of measuring lag the response of the measurement system *begins immediately after a change in measured quantity has occurred.*

   (*ii*) ***Time delay type lag.*** In this case the response of the measurement system *begins after a dead time after the application of the input.* If the measured quantity varies at a fast rate, the dead time has a severe adverse effect on the performance of the system. The measurement lags of this type are *very small* and are of the order of a fraction of a second and hence can be *ignored.*

3. **Fidelity.** It is defined as the *degree to which a measurement system indicates changes in the measured quantity without any dynamic error.* It refers to the ability of the system to reproduce the output in the same form as the input.

4. **Dynamic error.** The dynamic error, also called **"measurement error"**, is the *difference between the true value of the quantity changing with time and the value indicated by the measurement system if no static error is assumed* (see Fig. 2.5.)

   - *The maximum amount by which the pointer moves beyond the steady state is called* ***overshoot****.*

Example 2.16

On the application of a step input of 80 bar to a pressure gauge, the pointer swings to pressure of 82.5 bar and finally comes to rest at 81.3 bar. Determine:

(*i*) *The overshoot of the gauge reading and express it as a percentage of the final reading.*

(*ii*) *The percentage error of the gauge.*

Solution. (*i*) *Overshoot of the gauge reading* = 82.5 – 81.3 = **1.2 bar** **(Ans.)**

$$\text{Percentage overshoot} = \frac{1.2}{81.3} \times 100 = \mathbf{1.476\%} \quad \textbf{(Ans.)}$$

(*ii*)

$$\text{Percentage error} = \frac{81.3 - 80}{80} \times 100 = \mathbf{1.625} \quad \textbf{(Ans.)}$$

2.4.3. 'Dynamic' Analysis of Measurement Systems

The dynamic behaviour of measurement systems is studied in the following two domains:

1. Time domain analysis.
2. Frequency domain analysis.

1. Time domain analysis:

In this the input signal is applied to the measurement system and the behaviour of the system is studied as a *function of time*. The dynamic response of the system to different types of inputs, which are a function of time is analysed at different intervals of time after the application of the input signal. In most cases, the actual input signals vary in random fashion with respect to time and therefore cannot be mathematically defined. Consequently the performance of a system can be analysed (in the time domain analysis) by using the following standard *test signals/inputs*:

(*i*) Step input;

(*ii*) Ramp input;

(*iii*) Parabolic input;

(*iv*) Impulse input.

2. Frequency domain analysis:

This type of analysis of a system *pertains to the steady state response of the system to a sinusoidal input.* Here, the system is subjected to a sinusoidal input and the system response is studied with *frequency as the independent variable.*

- *Frequency response.* It is the maximum frequency of the measured variable that an instrument is capable of following without error. The usual requirement is that the frequency of measurand should not exceed 60 per cent of the natural frequency of the measuring instrument.

Standard test signals/inputs:

The most common standard inputs used for dynamic analysis are discussed below:

1. Step function:

Refer to Fig. 2.9(*a*). It is a *sudden change from one steady value to another.*

It is mathematically represented by the relationship:

$$x = 0 \quad \text{at} \quad t < 0$$

$$x = x_c \quad \text{at} \quad t \geq 0$$

where x_c is a constant value of the input signal x_i.

- The *"transient response" indicates the capacity of the system to cope with changes in the input signal.*

2. Ramp or linear function:

In this case, the *input varies linearly with time.*

This input is mathematically represented as:

$$x = 0 \quad \text{at} \quad t < 0$$

$$x = \psi \quad \text{at} \quad t \geq 0$$

where, ψ is the slope of the input versus time relationship.

- *The ramp-response becomes indicative of the steady state error in following the changes in the input signal.*

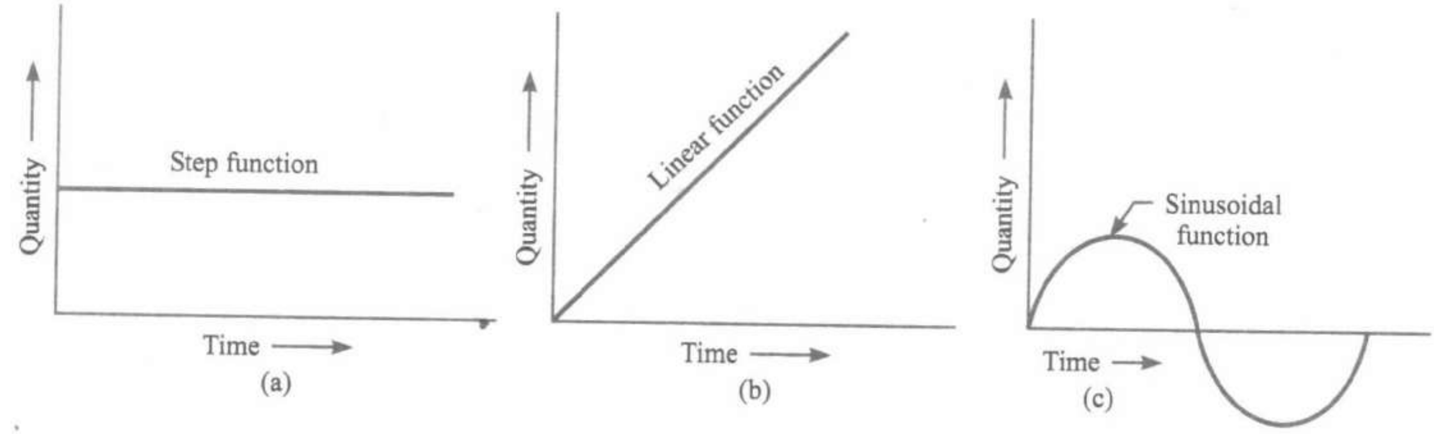


Fig. 2.9. Standard input functions.

3. Sinusoidal function:

In this case the *input varies sinusoidally with a constant maximum amplitude.*

It is represented mathematically as follows:

$$I_i = A \sin \omega t$$

where, A = Amplitude, and

ω = Frequency in rad/s.

- *The frequency or harmonic response is a measure of the capability of the system to respond to inputs of cyclic nature.*

A **general measurement system** can be mathematically described by the following **differential equation**:

$$(A_n D^n + A_{n-1} D^{n-1} + ... + A_1 D + A_0) I_o = (B_m D^m + B_{m-1} D^{m-1} + ... + B_1 D + B_0) I_i \quad ...(2.11)$$

where, A's and B's = Constants, depending upon the physical parameters of the system,

D^k = Operative derivative of the order k,

I_o = The information out of the measurement system, and

I_i = The input information.

The order of the measurement system is generally classified by the value of the power of n.

- *Zero-order system:* $n = 0$ and $A_1, A_2, A_3 ..... A_n = 0$
- *First-order system:* $n = 1$ and $A_2, A_3, A_4 ..... A_n = 0$
- *Second-order system:* $n = 2$ and $A_3, A_4, A_5 ..... A_n = 0$

The above method of classification is used for most of the instruments and systems.

Although general equation can be solved by various methods, we shall be using method of D-operator for getting its solution.

2.4.4. Zero, First and Second Order Systems

2.4.4.1. Zero-order systems

Fig. 2.10 shows the block diagram of a 'Zero-order system'. In this case the output of the measuring system (ideal) is directly proportional to input, no matter how the input varies. The *output is faithful reproduction of input without any distortion or time lag.*

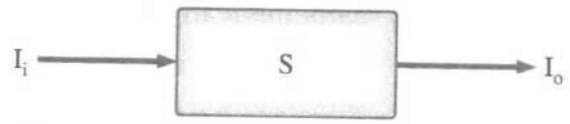


Fig. 2.10. Block diagram for Zero-order system.

The behaviour of the zero-order system is represented by the following mathematical solution:

$$I_o = S\, I_i \quad ...(2.12)$$

where, I_o = Information out of the measuring system,

S = Sensitivity of the system, and

I_i = Input information

This equation is obtained by putting $n = 0$ in the general equation (2.11),

i.e.,
$$A_o I_o = B_o I_i$$

or,
$$I_o = \frac{B_o}{A_o} I_i = S I_i \quad ...(2.13)$$

The zero-order system is characterised only by the static sensitivity (parameter), the value of which is obtained through the process of static calibration.

***Examples of zero-order system*:**

- Mechanical levers;
- Amplifiers;
- Potentiometer (It gives an output voltage which is proportional to wiper's displacement) etc.

2.4.4.2. First-order systems

Fig. 2.11 shows the block diagram of a 'First-order system'.

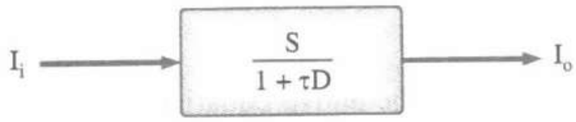


Fig. 2.11. Block diagram for First-order system.

The behaviour of a first-order system is given by following first-order differential equation:

$$A_1 \frac{dI_o}{dt} + A_o I_o = B_o I_i \quad ...(2.14)$$

(This equation is obtained by inserting $n = 1$ in the general equation).

Eqn. (2.14) may be written in standard form as follows:

$$\frac{A_1}{A_o}\frac{dI_o}{dt} + I_o = \frac{B_o}{A_o} I_i \quad ...(2.15)$$

or,
$$\tau \frac{dI_o}{dt} + I_o = S\, I_i \quad ...(2.16)$$

where, $\tau = \frac{A_1}{A_0}$ = Time constant, and

$S = \frac{B_o}{A_o}$ = Sensitivity.

Using D-operator, we get:

$$\left[\text{where, } D = \frac{d}{dt}, \text{ and } D^2 = \frac{d^2}{dt^2}\right]$$

$$\tau D I_o + I_o = S I_i$$

or,
$$I_o(\tau D + 1) = S I_i$$

or, $$\frac{I_o}{I_i} = \frac{S}{1+\tau D} \qquad \text{...(2.17)}$$

Equation (2.17) gives the standard form of *transfer operator* for first-order system.

Examples of first-order system :

- Vetocity of a true falling mass;
- Air pressure build-up in bellows;
- Measurement of temperature by mercury-in-glass thermometers;
- Thermistors and thermocouples;
- Resistance-capacitance network.

2.4.4.3. Second-order systems

Fig. 2.12 shows the block diagram of 'Second-order system':

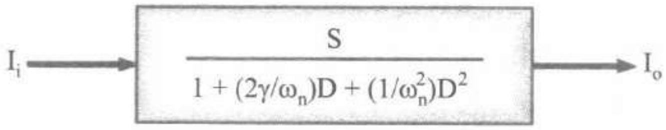


Fig. 2.12. Block diagram for Second-order system.

The behaviour of a second-order system is given by the following differential equation (obtained by putting $n = 2$ in the general equation);

$$A_2 \frac{d^2 I_o}{dt^2} + A_1 \frac{dI_o}{dt} + A_o I_o = B_o I_i \qquad \text{...(2.18)}$$

Dividing the above equation by A_o, we have:

$$\frac{A_2}{A_o}\frac{d^2 I_o}{dt^2} + \frac{A_1}{A_0}\frac{dI_o}{dt} + I_o = \frac{B_o}{A_o} I_i \qquad \text{...(2.19 a)}$$

Let, $$\omega_n = \sqrt{\frac{A_o}{A_2}} = \text{Undamped natural frequency, rad/s,}$$

$$\gamma = \frac{A_1}{2\sqrt{A_0\ A_2}} = \text{Damping ratio, dimensionless, and}$$

$$S = \frac{B_o}{A_o} = \text{Static sensitivity or steady-state gain.}$$

Then, by substituting these values in eqn. 2.19 (*a*), we get:

$$\frac{1}{\omega_n^2}\cdot\frac{d^2 I_o}{dt^2} + \frac{2\gamma}{\omega_n}\cdot\frac{dI_o}{dt} + I_o = SI_i \qquad \text{...(2.20)}$$

or, in terms of D-operator, we have:

$$\left(\frac{D^2}{\omega_n^2} + \frac{2\gamma}{\omega_n}D + 1\right)I_o = SI_i$$

or, $$\frac{I_o}{I_i} = \frac{S}{\frac{1}{\omega_n^2}D^2 + \frac{2\gamma}{\omega_n}D + 1} \qquad \text{...(2.21)}$$

- Piezoelectric pick-up;
- Spring-mass system (used for acceleration and force measurements);
- Pen control system on X-Y plotters;
- U.V. galvanometer, etc.

Damping ratio :

In the design of instruments a term which is very frequently used is the *"damping ratio"* (γ) defined as the *ratio of the actual value of coefficient of viscous friction in movement and the value required to produce critical damping.*

i.e.,
$$\gamma = \frac{A_1}{2\sqrt{A_o A_2}}$$

This dimensionless term is very useful because to determine its value, it is not necessary that the values of A_1, A_0 and A_2 may be known. In practice it is not easy to determine accurately the values of A_1 and A_2. Further, even if these values are known, they do not in themselves specify whether the instrument is under, over or critically damped, since a numerical calculation has to be performed with them first. Therefore, designers find "*damping ratio*" as a very convenient measure of the amount of damping present in the movement.

The terms *damping ratio* (γ) and *underdamped natural frequency* (ω_n) immediately conjure up a physical picture of the response of an instrument and both of the quantities are very easy to measure. Thus γ and ω_n easily do away with quantities A_2, A_1 and A_0.

2.4.5. First-order System Responses

The complete solution of an equation which describes the dynamical behaviour of a system consists of the following *two parts*:

(*i*) **Complementary function.** *It corresponds to the short time or transient response.*

(*ii*) **Particular integral.** *It refers to the long time steady state response.*

The transfer operator form of the first-order system is given by :

$$\frac{I_o}{I_i} = \frac{S}{1 + \tau D}$$

When S (static sensitivity or steady state gain) equals *unity*, we get:

$$(1 + \tau D)\, I_o = I_i \qquad \text{...(2.22)}$$

Now we shall obtain the solution of this equation for different standard inputs (The solutions are not mathematically rigorous, but are practical).

Transient response (Complementary function):

The transient response from the auxiliary equation is obtained by putting input I_i equal to zero;

i.e.,
$$(1 + \tau D) I_{o,t} = 0 \qquad \text{...(2.23)}$$

(subscript t refers to the transient value)

Let the solution be of the form :

$$I_{o,t} = A\,e^{mt}$$

where, m is an algebraic variable)

or,
$$(1 + \tau D)\,A\,e^{mt} = 0$$

or,
$$A\,e^{mt} + \tau \cdot \frac{d}{dt}(A\,e^{mt}) = 0$$

or,
$$A\,e^{mt} + \tau \cdot A\,m\,e^{mt} = 0$$
$$A\,e^{mt}(1 + \tau \cdot m) = 0$$

$\therefore$
$$m = -\frac{1}{\tau}$$

Then,
$$I_{o,t} = A\,e^{mt} = Ae^{-t/\tau} \quad ...(2.24)$$

The transient response of a first-order system is same for different standard inputs.

Steady state response (Particular integral) :

The steady state response is given by :

$$(1 + \tau D)\,I_{o,s} = I_i \quad ...(2.25)$$

(Subscript s refers to the steady state value)

or,
$$I_{o,s} = (1 + \tau D)^{-1}\,I_i$$
$$= (1 - \tau D + \text{terms in } D^2 \text{ and higher})\,I_i \quad ...(2.26)$$

1. Step input :

Since the input I_i is a step of constant magnitude; its differential equals zero, and subsequently, we get:

$$I_{o,s} = (1 - \tau D)\,I_i = I_i \quad ...(2.27)$$

Total response = Transient response + steady state response

or,
$$I_o = Ae^{-t/\tau} + SI_i \quad ...(2.28)$$

The constant A is evaluated from the initial conditions as follow :

At,
$$t = 0, \qquad I_o = 0$$

$\therefore$
$$0 = A + SI_i \quad \text{or,} \quad A = -SI_i$$

$\therefore$
$$I_o = \underbrace{-I_i\,e^{-t/\tau}}_{\text{Transient}} + \underbrace{I_i}_{\text{steady state}}$$

or,
$$I_o = I_i(1 - e^{-t/\tau}) \quad ...(2.29)$$

or,
$$\frac{I_o}{I_i} = (1 - e^{-t/\tau}) \quad ...(2.30)$$

.... in *non-dimensional form.*

Salient features (with step input):

Following are the salient features of first-order system with step input:

(*i*) The transient response of the first-order system is *time dependent*; as the time passes, grows its value decreases (Refer to eqn. 2.30) and after a very long time the value becomes zero approximately. Thus magnitude of output (I_o) will be same as input (I_i) when the time is very large.

(*ii*) The speed of response relates to the time constant τ. A large τ indicates that response of the system is slow, whereas a small τ represents a fast system response. Thus in order to get good fidelty (*i.e.*, for accurate dynamic measurements) efforts should be made to minimise the value of τ.

(*iii*) Refer to Fig. 2.13, which shows the time response of a first-order system to a step-input when $t = \tau; \frac{I_o}{I_i} = (1 - e^{-1}) = 0.632$. Thus, the *time constant* (τ), *for a rising exponential function, is defined as the time to reach* 63.2% of its steady state value. The time constant, for a *decaying function* would correspond to the time taken to fall to 36.8% *of its initial value*.

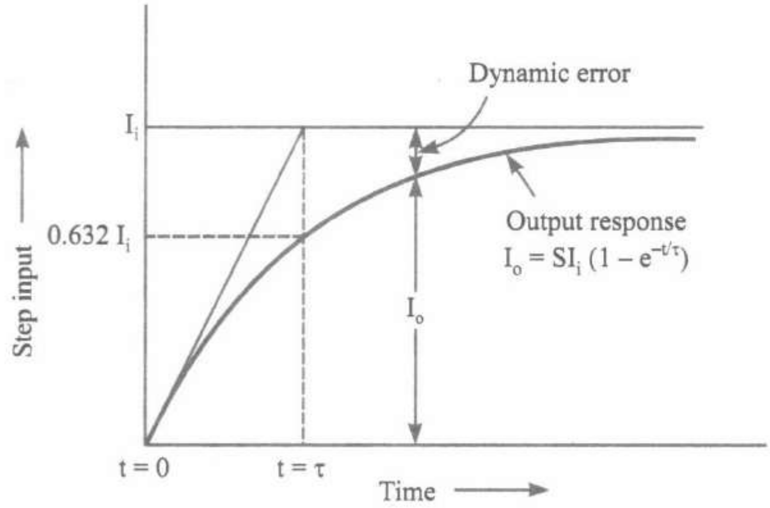


Fig. 2.13. Time response of a first-order system to step input.

(*iv*) Dynamic error (*i.e.*, vertical difference between the input and output response curve),

$$E_{dy.} = I_i - I_o$$
$$= I_i - I_i(1 - e^{-t/\tau})$$
$$= -I_i e^{-t/\tau} \quad ...(2.31)$$

or, $$\frac{E_{dy.}}{I_i} = -e^{-t/\tau} \quad ...(2.32)$$

.... (*in dimensionless form*)

(*v*) In case the measurand has an initial value of I_{initial} at $t = 0$, then the output. I_o at any instant t is given by:

$$I_o = I_i[1 - e^{-t/\tau}] + I_{\text{initial}}\, e^{-t/\tau}$$

or, $$I_o = I_i + (I_{\text{initial}} - I_i)\, e^{-t/\tau} \quad ...(2.33)$$

(*vi*) The speed response of a system is defined in terms of *settling time* (it is time taken by the system to reach and remain within a certain percentage tolerance band of the final steady state value). Smaller the setling time, faster is the response. Typical value of tolerance band are 2% and 5% settling times.

For unit step input, the output reaches the values given in Table 2.1 at various intervals of time. This table also shows the per unit dynamic error at different intervals of time.

TABLE 2.1 Per unit output and dynamic error for a step input for a first-order system.

t/τ	Per unit output (I_o / I_i)	Per unit dynamic error ($E_{dy.}/I_i$)
0	0	1.000
1	0.632	0.368
2	0.865	0.135
3	0.950	0.050
4	0.982	0.018
5	0.993	0.007
∞	1.000	0

Thus, 5% settling time means that the system has reached its specified value after a time which is thrice the time constant (Table 2.1).

2. Ramp Input

Consider that the input signal varies linearly with time so that ramp input I_i is prescribed by the relation $I_i = \psi t$, where ψ is constant. Then the governing differential equation is given by:

$$(1 + \tau D) I_o = \psi t \quad \text{...(2.34)}$$

Now, *Transient response* (complimentary function) is given as:

$$I_{o,t} = A\, e^{-t/\tau} \quad \text{... as before}$$

and, *steady state response* (particular integral) will be:

$$I_{o,s} = (1 - \tau D + \text{terms in } D^2 \text{ and higher})\, \psi t = \psi t - \tau \frac{d}{dt}(\psi \tau)$$

or,
$$I_{o,s} = \psi t - \psi \tau \quad \text{...(2.35)}$$

∴ Complete response = Transient response + steady state response

$$I_o = A\, e^{-t/\tau} + (\psi t - \psi \tau) \quad \text{...(2.36)}$$

The value of constant A can be evaluated by applying the initial condition,

At, $t = 0 \qquad I_0 = 0$

or, $0 = A - \psi\tau \quad \therefore A = \psi\tau$

$$\therefore \quad I_o = \psi t - \psi\tau + \psi\tau \times e^{-t/\tau} = \psi(t - \tau) + \psi\tau\, e^{-t/\tau} \quad \text{...(2.37)}$$

$$\text{or,} \quad I_o = \psi[t - \tau(1 - e^{-t/\tau})] \quad \text{...(2.37a)}$$

Fig. 2.14 shows the time response of a first-order system to a ramp input.

The dynamic error,

$$E_{dy.} = I_i - I_o$$
$$= \psi t - [\psi t - \psi\tau + \psi\tau\, e^{-t/\tau}]$$
$$= \underbrace{\psi\tau}_{\text{steady}} - \underbrace{\psi\tau\, e^{-t/\tau}}_{\text{Transisent}} \quad \text{...(2.38)}$$

$$\text{or,} \quad \frac{E_{dy.}}{\psi\tau} = 1 - e^{-t/\tau} \quad \text{...(2.39)}$$

... (in dimensionless form)

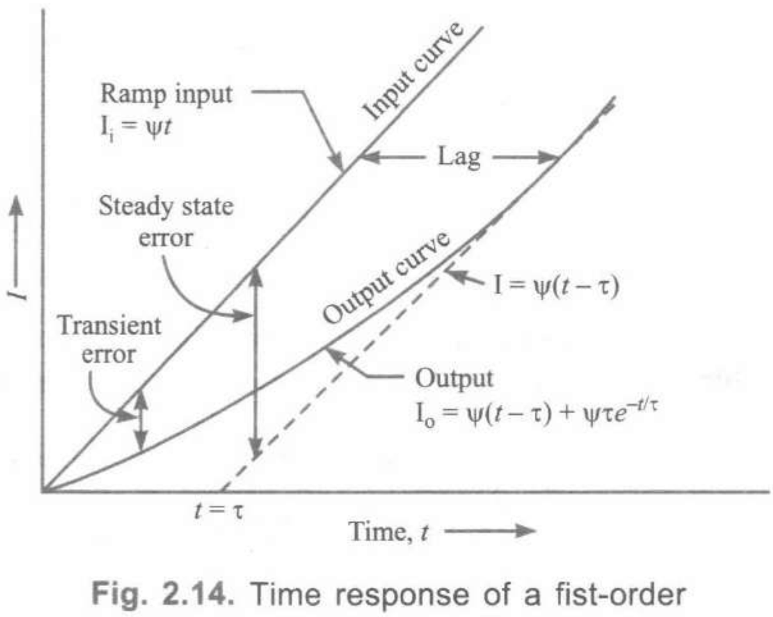


Fig. 2.14. Time response of a fist-order system to a ramp input.

Salient features (with ramp input) :

(*i*) the term $\psi\tau$ being independent of time continues to exist and so it is called the steady state error. The term $\psi\tau\ e^{-t/t}$ gradually decreases with time and hence is called the *transient error*.

- Since the steady state error is directly proportional to τ (time constant), therefore, the larger the value of t the larger will be the magnitude of the error.
- When τ is made small the transient error decreases rapidly; this implies, that the system attains the steady state at a faster pace.

(*ii*) The output response curve always lags behind the input curve by a constant amount known as **lag**.

3. Sinusoidal (Harmonic) input :

The frequency analysis of a system pertains to the steady state response of the system to a sinusoidal input. In this analysis, the system is subjected to a sinusoidal input and the system response studied with frequency as the independent variable. The sinusoid is a unique input signal, and the resulting output signal for a linear system is sinusoidal in the steady state. However, the output signal differs from the input waveform in amplitude and phase.

In order to determine the frequency response of sinusoidal input to a first-order system, let us replace the transfer operator D by a factor $j\omega$ in the input/output relationship; then we get:

$$\frac{I_o}{I_i} = \frac{1}{1 + D\tau} = \frac{1}{1 + j\omega\tau} \quad ...(2.40)$$

where, ω = Input frequency, rad/s, and

$j = \sqrt{(-1)}$

In a frequency response the following two quantities are of interest : Refer to Fig. 2.15.

(*i*) *Amplitude ratio or modulus* $\left(\frac{I_o}{I_i}\right)$. It prescribes the size of the output amplitude relative to the input amplitude.

(*ii*) *Phase shift* of output relative to input.

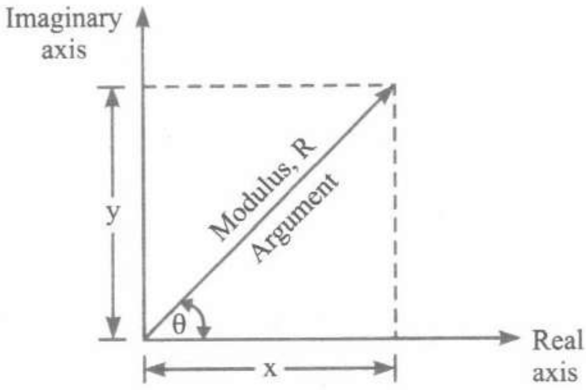


Fig. 2.15.

For the first-order system represented by the equation (2.40),

$$\text{Modulus} = \sqrt{1 + (\omega\tau)^2}$$

$$\text{Argument/Phase angle} = \tan^{-1}(\omega\tau) \quad ...(2.41)$$

Salient features (*with sinusoidal input*) : Refer to Fig. 2.16.

(*i*) When a system is subjected to a sinusoidal input with frequency ω, its output will also be sinusoidal, but the magnitude of the output amplitude necessarily may not be the same (as the input one). *The ratio of the amplitude* (often called *attenuation*) is given as :

$$\frac{I_o}{I_i} = \frac{1}{\sqrt{1 + (\omega\tau)^2}} \quad ...(2.42)$$

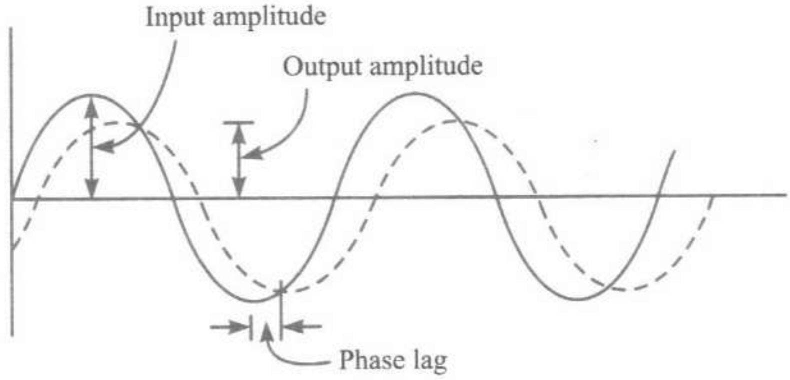


Fig. 2.16. Relationship between an input frequency and corresponding output frequency.

Thus, with the increase in input frequency, the amplitude ratio decreases.

(*ii*) The output from the system may not necessarily be in phase with the input; and the phase difference is given by:

$$\phi \text{ (phase angle) } = -\tan^{-1}(\omega\tau) \qquad \text{...(2.43)}$$

– ve indicates that output *lags* behind the input. When $\omega = \frac{1}{\tau}$ the phase lag is $\frac{\pi}{4}$ or 45°.

As the accuracy of an instrument measuring dynamic input depends upon the time constant, therefore, smaller the time constant, greater the accuracy; for phase shift to be small, the time period τ should be small.

(*iii*) When the input and output signals are given by the relations :

$$I_i = A \sin \omega t, \text{ and } I_o = B \sin(\omega t + \phi) = zA \sin(\omega t + \phi),$$

Then the amplitude ratio may be represented as follows:

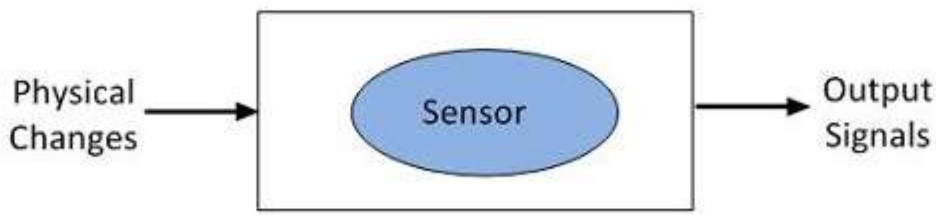
$$K = \left|\frac{I_o}{I_i}\right| = \frac{z}{\sqrt{1 + (\omega\tau)^2}} \qquad \text{...(2.44)}$$

In order to produce amplitude of sinewave without any attenuation ($K = 1$) we must use an instrument whose time constant , $\tau = \frac{\sqrt{z^2 - 1}}{\omega}$.

DEFINITION OF SENSOR

The sensor is a device that measures the physical quantity (i.e. Heat, light, sound, etc.) into an easily readable signal (voltage, current etc.). It gives accurate readings after calibration.

Examples – The mercury used in the thermometer converts the measurand temperature into an expansion and contraction of the liquid which is easily measured with the help of a calibrated glass tube. The thermocouple also converts the temperature to an output voltage which is measured by the thermometer.



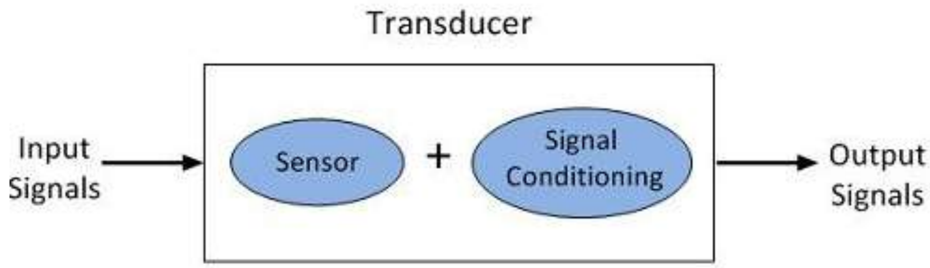
Sensor

The sensors have many applications in the electronics equipment. The few of them are explained below.

1. The motion sensors are used in the home security system and the automation door system.
2. The photo sensor senses the infrared or ultraviolet light.
3. The accelerometer sensor use in the mobile for detecting the screen rotations.

DEFINITION OF TRANSDUCER

The transducer is a device that changes the physical attributes of the non-electrical signal into an electrical signal which is easily measurable. The process of energy conversion in the transducer is known as the transduction. The transduction is completed into two steps. First by sensing the signal and then strengthening it for further processing.



.

The transducer has three major components; they are the input device, signal conditioning or processing device and an output device.

The input devices receive the measurand quantity and transfer the proportional analogue signal to the conditioning device. The conditioning device modified, filtered, or attenuates the signal which is easily acceptable by the output devices.

Key Differences Between Sensor and Transducer

The following are the key differences between the sensor and transducer.

1. The sensor senses the physical change across the surrounding whereas the transducer transforms the one form of energy into another.
2. The sensor itself is the major component of the sensor, whereas the sensor and the signal conditioning are the major elements of the sensor.
3. The primary function of the sensor is to sense the physical changes, whereas the transducer converts the physical quantities into an electrical signal.
4. The accelerometer, barometer, gyroscope are the examples of the sensors whereas the thermistor, and thermocouple is the examples of the transducer

Classifications:-

1.Based upon transduction principle used

2. Primary & Secondary transducers

3.Active & Passive transducers

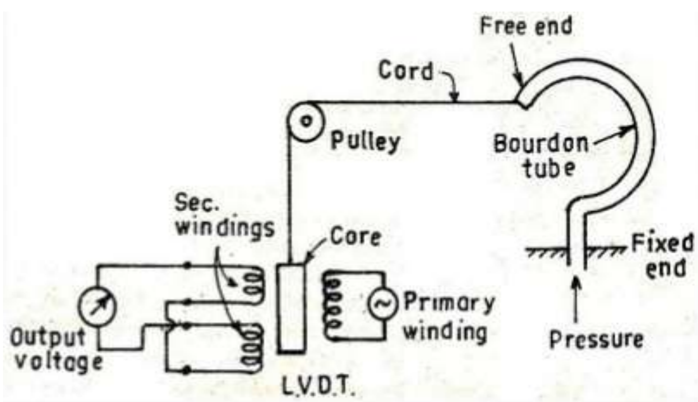
4. Analoge & Digital transducers

5.Transducers & Inverse transducers

1.Based upon the transduction principle:-

- **Resistance -** Potentimeter devices, Resistance strain gauge, Pirani guage or hot wire meter, Resistance thermometer, Thermister, Resistance hygrometer, Photoconductive cell.
- **Capacitance -** Variable capacitance pressure gauge, Capacitor microphone, Dielectric gauge.
- **Inductance -** Magnetic circuit transducer, Reluctance pick-up, Differential transformer, Eddy current guage, Magneto striction gauge.
- **Voltage & Current -** Hall effect transducer, Ionisation chamber, Photoemissive cell, Photomultiplier tube.
- **Self generating transducers -** Thermocouple, Thermopile, Moving coil generator, Piezoelectric transducer, Photovoltaic.

2. Primary & Secondary transducers:-



Look at the above picture, it is easy to understand the concept with this picture. Here the measuring variable pressure is passed through bourdon tube. Bourdon converts the pressure into displacement by the movement of free end. This displacement is transferred to iron core of LVDT. LVDT converts

displacement to voltage.
Here bourdon tube acts as primary transducer and LVDT acts as secondary transducer.

3. Active & Passive:-

Transducers those which don't require an auxilary power source to produce their output are known as 'Active transducers' or self generating type.

eg: Moving coil, Piezoelectric crystal, Thermocouple, Photovoltaic cell.
On the other hand transducer that can't work on the absence of external power supply are called 'passive transducers'.
eg: Resistive, Capacitive, Inductive.

4. Analoge & Digital :-

Analoge transducers converts input quantity into an analog output which is continous function of time. eg: strain gauge, LVDT, Thermocouple. Thermister.

Digital transducer converts input quantity into an electrical output is in the form of pulses eg: Glass scale, Metallic scale.

5. Transducers & Inverse Transducers

We have already looked into transducers, inverse transducers are just the opposite of transducers in function. I Inverse transducers converts electrical quantity in non electrical quantity. Just think about an current coil moving in magnetic field which is an inverse transducer.
eg: Peizoelectrical crystal.

SELECTION OF A TRANSDUCER

Transducers are often employed at the boundaries of automation, measurement, and control systems. Before choosing a transducer we have to ensure that the transducer is suitable for our need. These are factors which may influence the selection.

Operating principle:

There are operating principles such as resistive, inductive, capacitive, piezoelectric, photo-voltaic, ionization etc.,

.

Sensitivity:

Transducer must be sensitive enough not to give an output, but to give an detectable output.

Operating range:

A transducer should have good resolution over it's entire range of operation. We cannot choose a temperature sensing transducer that work in range 0-100 degree Celsius for a use in boiler where temperature is up to 1000 degree Celsius, that's what it means.

Accuracy:

High degree of accuracy is needed and small value of repeatablity.

Cross-sensitivity:

Cross-sensitivity must taken into account while measuring mechanical quantity. Cross-sensitivity must reduced. Because when transducer subjected to measure variations in one plain, while actual quantity to be measured is in another plain.

Error:

A transducer must avoid error. For which a transducer should maintain expected input output relationship by it's transfer function.

Transient & Frequency response:

Transducer should meet desired time domain specifications such as peak over shoot, raise time, setting time & small dynamic error, it should flat frequency response curve with higher cut-off frequency at high limit in-order to have a high value.

Loading effect:

Transducer should have high input impedance and low output impedance to avoid loading effect.

Environment compatibility:

Under specified environment conditions the transducer maintain input-output relationship and doesn't breakdown. Transducer should be able to withstand temperature, pressure shock etc., when subjected to it, if it is application require so.

Insensitivity to unwanted signals:

Noise should be avoided in measuring, so that a good transducer should insensitive to unwanted signals. But high sensitive to desired signals.

Usage and ruggedness:

Ruggedness both of mechanical and electrical intensities of transducer VS it's size and weight must be considered while selection.

STRAIN GAUGE MEASURMENTS

Introduction:

- When a metal conductor is stretched or compressed, its resistance changes an account of the fact that both length and diameter of conductor change.
- The value of resistivity of the conductor also changes. When it is strained its property is called piezo- resistance.
- Therefore, resistance strain gauges are also known as piezo-resistive gauges.

Strain gauge:

- A strain gauge is a device which is used to measure dimensional change on the surface of a structural member under test.
- Strain gauges give indication of strain at only one point.

Types of strain gauge:

Four types of strain gauges are:

- Wire-wound strain gauge
- Foil-type strain gauges.
- Semiconductor strain gauges

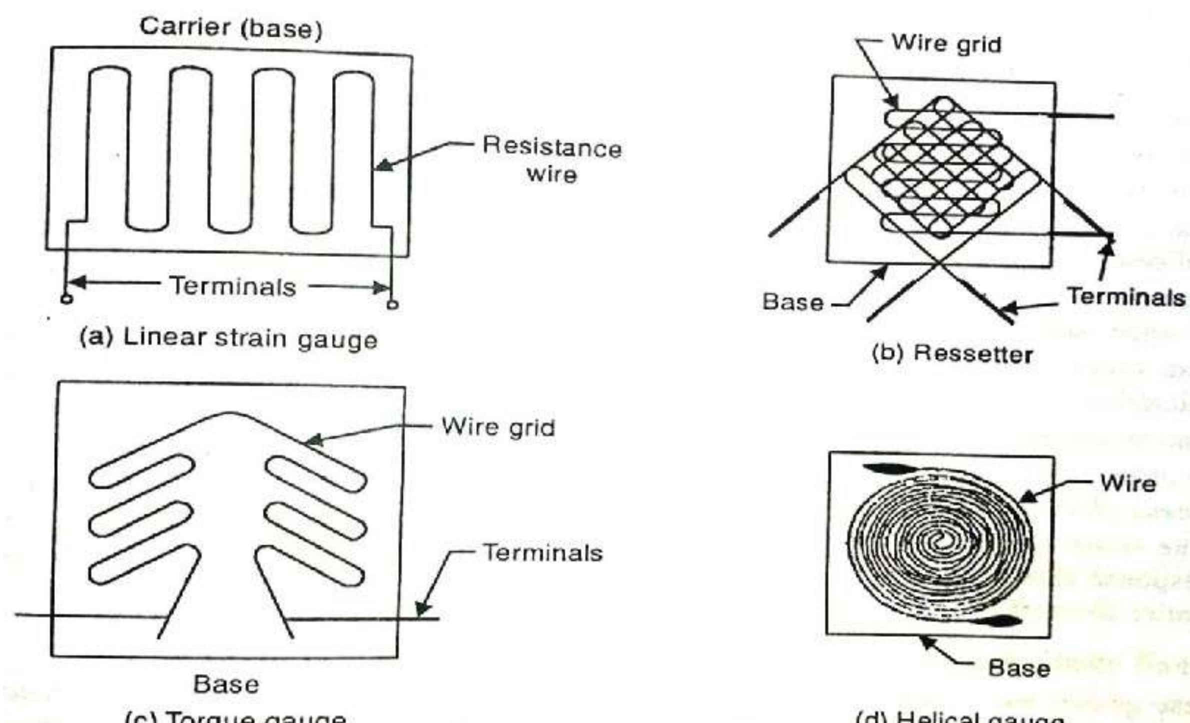
1. Wire wound strain gauges:

These are two main classes of wire-wounded strain gauges:

- Bonded strain gauge.
- Unbounded strain gauge.

Bounded strain gauge:

- It is composed of fine wire, wound and cemented on a resilient insulating support, usually a wafer unit.



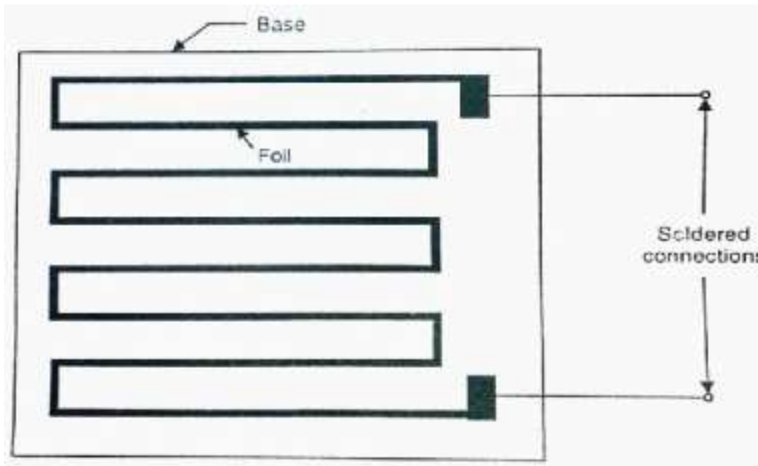
Resistance wire strain gauges.

- Units may be mounted upon or incorporated In mechanical elements whose deformations under stress are to be determined.

While there are no limits to the basic values which may be selected for strain-gauge resistance

2. Foil strain gauges:

In these gauges the strain is sensed with help of metal foil. Foil gauges have a much greater dissipation capacity as compared with wire wounded gauges on account of their greater surface area for the same volume. Due to this reason they can be employed for high operating temperature range.



Foil gauge

In these gauges, the bonding is better due to large surface area of the foil. The bonded foil gauges find a wider field of action.

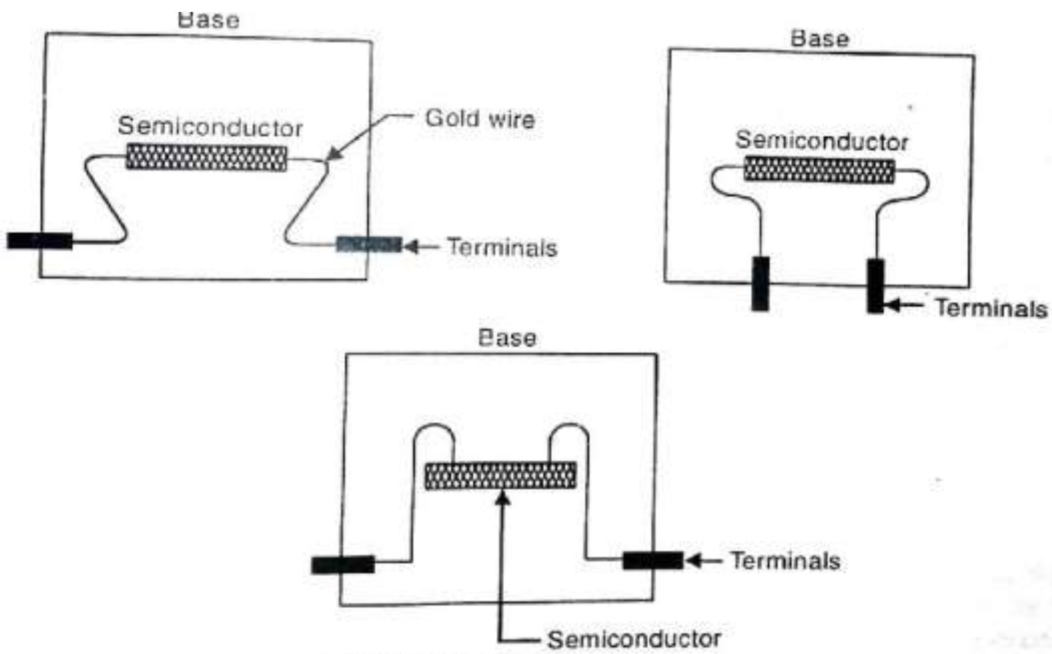
The characteristics of foil type's strain gauges are similar to those of wire wounded strain gauges and their gauge factors are typically the same as that of wire wounded strain gauges.

3. Semiconductor strain gauges:

- Semiconductor strain gauges depend for their action upon piezo-resistive effect i.e., the change in value of the resistance due to change in resistivity.
- These gauges are used where a very high gauge factor and small envelope are required.
- For semiconductor strain gauges semiconducting materials such as silicon and germanuium are used.
- A typical strain gauge consists of a strain sensitive crystal material and leads that are sandwiched in a protective matrix.

The production of these gauges employs conventional semiconductor technology using semiconducting filaments which have a thickness of 0.05 mm and bonding them on suitable insulating substances, such as Teflon

The resistance value of foil gauges which are commercially available is between 50 and 1000 Ω.



Semiconductor strain gauges.

- Gold, leads are generally applied for making the contacts.

Advantages:

- These gauges have high gauge factor.
- Excellent hysteresis characteristics.

Disadvantages:

- Linearity of these gauges is poor.
- Manometers are used up to 200 kN/m^2; above this pressure dead weight testers are used.

CHARACTERISTICS OF STRAIN GAUGES

The characteristics of strain gauges are as follows:

1. They are highly precise and don't get influenced due to temperature changes. However, if they do get affected by temperature changes, a thermistor is available for temperature corrections.

2. They are ideal for long distance communication as the output is an electrical signal.

3. Strain Gauges require easy maintenance and have a long operating life.

4. The production of strain gauges is easy because of the simple operating principle and a small number of components.

5. The strain gauges are suitable for long-term installation. However, they require certain precautions while installing.

6. All the strain gauges produced by Encardio-Rite are hermetically sealed and made up of stainless steel thus, waterproof.

7. They are fully encapsulated for protection against handling and installation damage.

8. The remote digital readout for strain gauges is also possible.

GAUGE FACTOR:

The gauge factor is defined as the unit change in resistance per unit change in length.

It is denoted as K or S. It is also called sensitivity of the strain gauge.

$$S = \frac{\Delta R/R}{\Delta l/l}$$

S = Gauge factor or sensitivity

R = Gauge wire resistance

ΔR = Change in wire resistance

l = Length of the gauge wire in unstressed condition

Δl = Change in length in stressed condition.

Derivation: Consider that the resistance wire is under tensile stress and it is deformed by ~l as shown in the Fig.

When uniform stress (J is applied to this wire along the length, the resistance R

Let ρ = Specific resistance of wire material in Ω-m

l = Length of the wire in m

A = Cross-section of the wire in m^2

changes to R + ~R because of change in length and cross-sectional area.

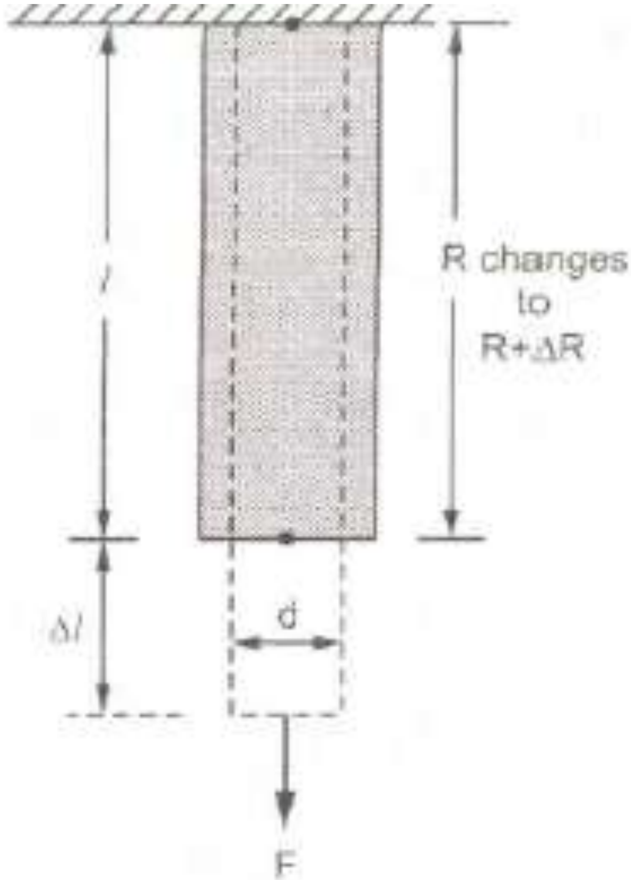
$$\sigma = \text{Stress} = \frac{\Delta l}{l}$$

$\Delta l/l$ = Per unit change in length

$\Delta A/A$ = Per unit change in area

$\Delta \rho/\rho$ = Per unit change in resistivity (specific resistance)

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



$$\therefore \qquad \frac{dR}{d\sigma} = \frac{d\left(\frac{\rho l}{A}\right)}{d\sigma} = \frac{\rho}{A}\frac{\partial l}{\partial \sigma} - \frac{\rho l}{A^2}\frac{\partial A}{\partial \sigma} + \frac{l}{A}\frac{\partial \rho}{\partial \sigma}$$

Note that $\qquad \frac{\partial}{\partial \sigma}\left(\frac{1}{A}\right) = -\frac{1}{A^2}\frac{\partial A}{\partial \sigma}$

Multiply both sides by $\frac{1}{R}$,

$$\frac{1}{R}\frac{dR}{d\sigma} = \frac{\rho}{RA}\frac{\partial l}{\partial \sigma} - \frac{1}{R}\frac{\rho l}{A^2}\frac{\partial A}{\partial \sigma} + \frac{l}{RA}\frac{\partial \rho}{\partial \sigma}$$

Using $\qquad R = \frac{\rho l}{A}$ on right hand side,

$$\frac{1}{R}\frac{dR}{d\sigma} = \frac{1}{l}\frac{\partial l}{\partial \sigma} - \frac{1}{A}\frac{\partial A}{\partial \sigma} + \frac{1}{\rho}\frac{\partial \rho}{\partial \sigma}$$

Canceling $\partial\sigma$ from both sides,

$$\frac{dR}{R} = \frac{dl}{l} - \frac{dA}{A} + \frac{\partial\rho}{\rho}$$

i.e.

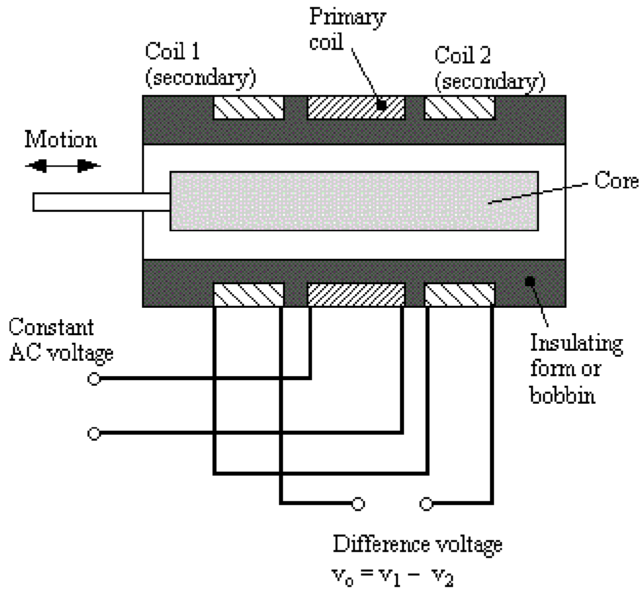
$$\frac{\Delta R}{R} = \frac{\Delta l}{l} - \frac{\Delta A}{A} + \frac{\Delta\rho}{\rho}$$

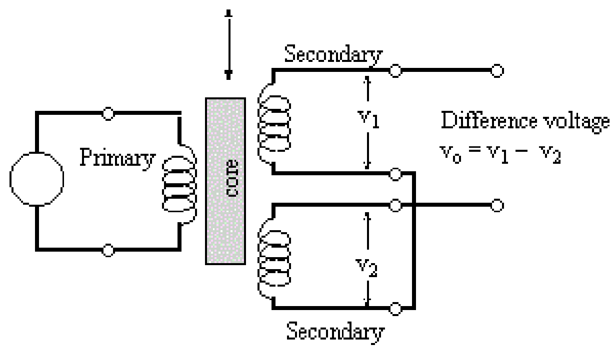
LINEAR VARIABLE DIFFERENTIAL TRANSFORMER (LVDT)

Linear Variable Differential Transformer, LVDT is the most used inductive transducer for translating linear motion into electrical signal. This transducer converts a mechanical displacement proportionally into electrical signal.

Main Features of Construction

- The transformer consists of a primary winding P and two secondary windings S_1 and S_2 wound on a cylindrical former (which is hollow in nature and contains the core).
- Both the secondary windings have an equal number of turns, and we place them on either side of primary winding
- The primary winding is connected to an AC source which produces a flux in the air gap and voltages are induced in secondary windings.
- A movable soft iron core is placed inside the former and displacement to be measured is connected to the iron core.
- The iron core is generally of high permeability which helps in reducing harmonics and high sensitivity of LVDT.
- The LVDT is placed inside a stainless steel housing because it will provide electrostatic and electromagnetic shielding.
- The both the secondary windings are connected in such a way that resulted output is the difference between the voltages of two windings.



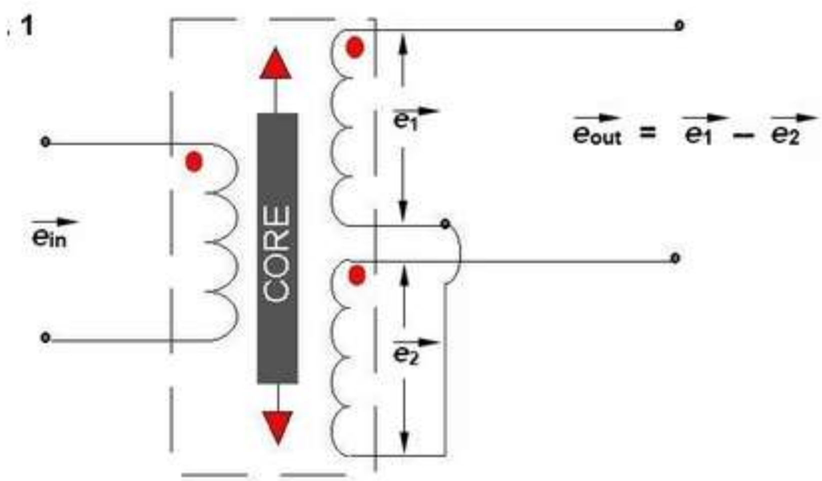


Principle of Operation and Working

As the primary is connected to an AC source so alternating current and voltages are produced in the secondary of the LVDT. The output in secondary S_1 is e_1 and in the secondary S_2 is e_2. So the differential output is,

$$e_{out} = e_1 - e_2$$

This equation explains the **principle of Operation of LVDT**.



Now three cases arise according to the locations of core which explains the working of LVDT are discussed

- CASE I When the core is at null position (for no displacement)
  When the core is at null position then the flux linking with both the secondary windings is equal so the induced emf is equal in both the windings. So for no displacement the value of output e_{out} is zero as e_1 and e_2 both are equal. So it shows that no displacement took place.

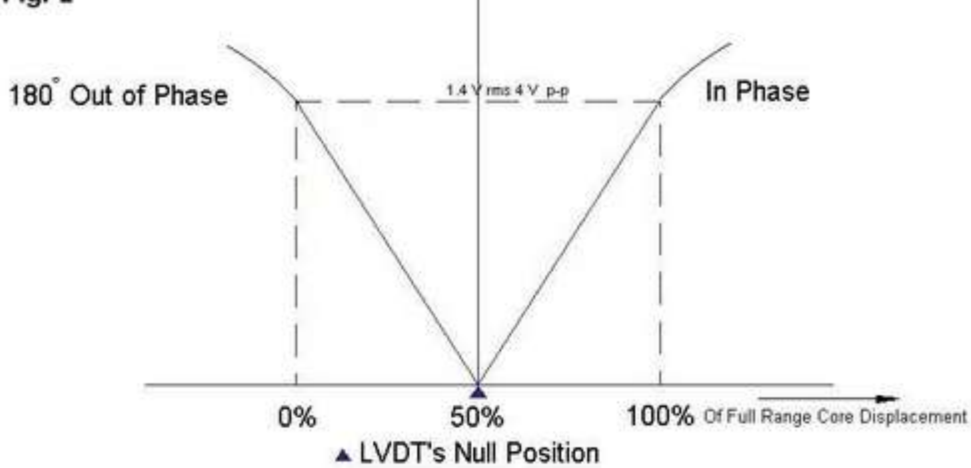
- CASE II When the core is moved to upward of null position (For displacement to the upward of reference point)
  In the this case the flux linking with secondary winding S_1 is more as compared to flux linking with S_2. Due to this e_1 will be more as that of e_2. Due to this output voltage e_{out} is positive.

- CASE III When the core is moved to downward of Null position (for displacement to the downward of the reference point). In this case magnitude of e_2 will be more as that of e_1. Due to this output e_{out} will be negative and shows the output to downward of the reference point.

Output V_S Core Displacement A linear curve shows that output voltage varies linearly with displacement of core.

LVDT'S AC Output Magnitude

Fig. 2



AC Output of Conventional LVDT Versus Core Displacement

Some important points about magnitude and sign of voltage induced in LVDT

- The amount of change in voltage either negative or positive is proportional to the amount of movement of core and indicates amount of linear motion.
- By noting the output voltage increasing or decreasing the direction of motion can be determined
- The output voltage of an LVDT is linear function of core displacement .

Advantages of LVDT

- High Range – The LVDTs have a very high range for measurement of displacement.they can used for measurement of displacements ranging from 1.25 mm to 250 mm
- No Frictional Losses – As the core moves inside a hollow former so there is no loss of displacement input as frictional loss so it makes LVDT as very accurate device.
- High Input and High Sensitivity – The output of LVDT is so high that it doesn't need any amplification. The transducer posseses a high sensitivity which is typically about 40V/mm.
- Low Hysteresis – LVDTs show a low hysteresis and hence repeatability is excellent under all conditions
- Low Power Consumption – The power is about 1W which is very as compared to other transducers.
- Direct Conversion to Electrical Signals – They convert the linear displacement to electrical voltage which are easy to process

Disadvantages of LVDT

- LVDT is sensitive to stray magnetic fields so it always requires a setup to protect them from stray magnetic fields.
- LVDT gets affected by vibrations and temperature.

It is concluded that they are advantageous as compared than any other inductive transducer.

Applications of LVDT

1. We use LVDT in the applications where displacements to be measured are ranging from a fraction of mm to few cms. The **LVDT** acting as a primary transducer converts the displacement to electrical signal directly.
2. The LVDT can also act as a secondary transducer. E.g. the Bourbon tube which acts as a primary transducer and it converts pressure into linear displacement and then LVDT coverts this displacement into an electrical signal which after calibration gives the readings of the pressure of fluid.

ROTARY VARIABLE DIFFERENTIAL TRANSFORMER (RVDT)

Definition: The transformer which **senses** the **angular displacement** of the conductor is **known** as the **Rotary Variable Differential Transformer** or RVDT. It is the type of electromechanical **transducer which** gives **the** linear output proportional **to the** input angular displacement**.**

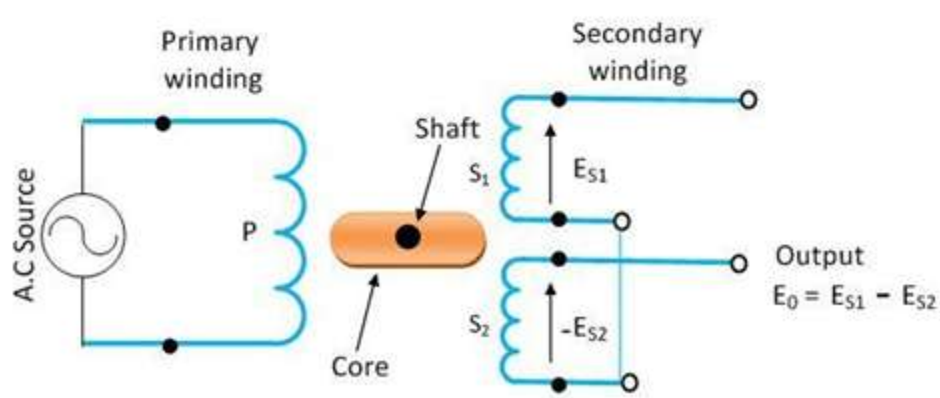
RVDT CONSTRUCTION AND ITS WORKING

RVDT transducer has two windings similar to a normal transformer such as primary winding and two secondary windings shown in the following **RVDT diagram**. The two windings of the transformer wounded, where the two secondary windings have an equivalent number of windings. These are located on both sides of the primary winding of the transformer. A cam formed a magnetic core which is made with a soft iron is coupled to a shaft. Thus, this core can be twisted among the windings. The construction of both the RVDT and LVDT are similar but the main difference is the shape of the core in transformer windings. This core will turn between the two windings of the transformer due to the shaft.

RVDT Construction

The typical RVDTs are linear over a +40 or -40 degrees, Sensitivity is about 2mV to 3mV per degree of rotation and the input voltage range is 3V RMS at frequency ranges from 400Hz to 20kHz. Based on the movement of the shaft in the transformer, the three conditions will be produced such as

- When the Core is at Null Position
- When the Core Rotates in Clockwise Direction
- When the Core Rotates in Anticlockwise Direction



When the Core is at Null Position

In the first condition, when the shaft is placed at the null position then the induced e.m.f in the secondary windings are similar although reverse in phase. Thus, the differential o/p potential will be zero, and the condition will be E1 = E2, where E0 = E1-E2 =0

When the Core Rotates in Clockwise Direction

In the second condition, when the shaft rotates in the direction of clockwise; more section of the core will enter across the primary winding. Therefore, the induced e.m.f across the primary winding is higher than secondary winding. Hence, the differential o/p potential is positive, and the condition will be E1 > E2, where E0 = E1-E2 = positive.

When the Core Rotates in Anticlockwise Direction

In the third condition, when the shaft rotates in the direction of anticlockwise, more section of the core will be entered across the secondary winding. Thus, the induced e.m.f across the secondary coil is higher than the primary coil. Hence, the differential o/p potential is negative that means 1800 phase shift, and the condition will be E1 < E2, where E0 = E1-E2 = negative.

RVDT Advantages and Disadvantages

The advantages of RVDT include the following.

- The consistency of RVDT is high
- The exactness of RVDT is high
- The lifespan is long
- The performance is repeatable
- The construction is compact and strong
- Durability
- Low cost
- Easy to handle electronic components

The disadvantages of RVDT mainly include the following

- The contact among the measuring exterior as well as the nozzle is not possible for all time.
- The output of the RVDT is linear (about +40 or -40 degrees), so it restricts the usability.

RVDT Applications

The applications of RVDT include the following.

- Fuel Valves as well as Hydraulic
- Modern machine tools
- Controls Cockpit
- Controls Fuel
- Brake with cable systems
- Engines bleed air-systems
- Robotics

PIEZO-ELECTRIC TRANSDUCER

Definition: The Piezoelectric transducer is an **electroacoustic transducer** use for **conversion** of **pressure** or mechanical stress into an alternating **electrical force.** It is used for measuring the physical quantity like force, pressure, stress, etc., which is directly not possible to measure.

The word piezoelectric means the electricity produces by the pressure. The Quartz is the examples of the natural piezoelectric crystals, whereas the Rochelle salts, ammonium dehydration, phosphate, lithium sulphate, dipotassium tartrate are the examples of the man made crystals. The ceramic material is also used for piezoelectric transducer.

The ceramic material does not have the piezoelectric property. The property is developed on it by special polarizing treatment. The ceramic material has several advantages. It is available in different shapes and sizes. The material has the capability of working at low voltages, and also it can operate at the temperature more than 3000ºC

Piezoelectric Effect

The EMF develops because of the displacement of the charges. The effect is changeable, i.e. if the varying potential applies to a piezoelectric transducer, it will change the dimension of the material or deform it. This effect is known as the piezoelectric effect.

The pressure is applied to the crystals with the help of the force summing devices for examples the stress is applied through mechanical pressure gauges and pressure sensors, etc. The deformation induces the EMF which determines the value of applied pressure.

Theory of Piezo-Electric Transducer

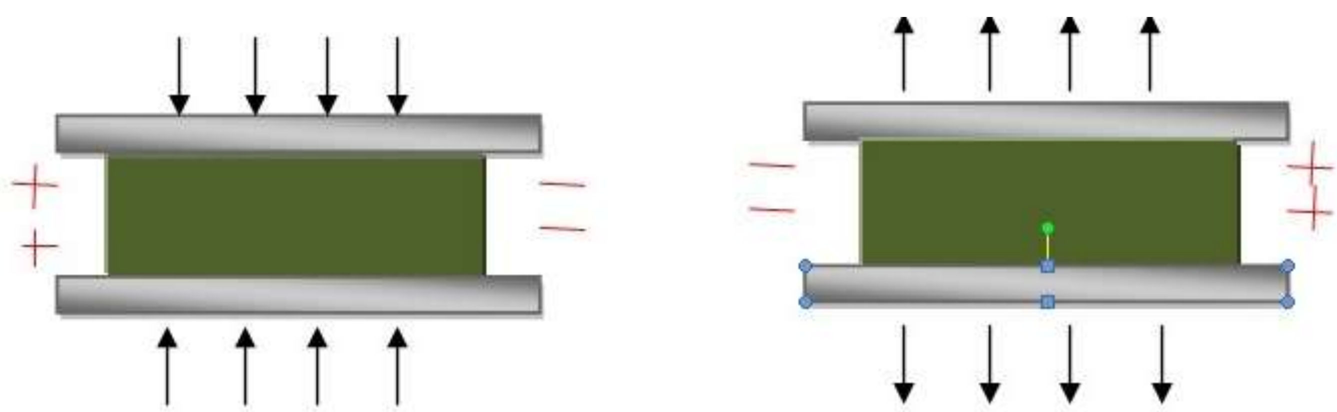
Piezoelectric Transducer Working

Piezoelectric Transducer works with the principle of piezoelectricity. The faces of piezoelectric material, usual quartz, is coated with a thin layer of conducting material such as silver. When stress has applied the ions in the material move towards one of the conducting surface while moving away from the other. This results in the generation of charge. This charge is used for calibration of stress. The polarity of the produced charge depends upon the direction of the applied stress. Stress can be applied in two forms as C**ompressive stress** and **Tensile stress** as shown below.

Working of a Piezoelectric Transducer

Piezoelectric Transducer Formula

The orientation of the crystal also effects the amount of voltage generated. Crystal in a transducer can be arranged in **longitudinal position** or **transverse position**.



Piezoelectric Transducer Formula

Longitudinal and Transverse Effect

In the longitudinal effect, the charge generated is given by

$$Q = F * d$$

Where F is the applied force,d is the piezoelectric coefficient of the crystal.

Piezoelectric coefficient d of quartz crystal is around $2.3 * 10^{-12}$ C/N.
In the transverse effect, the charge generated is given by

$$Q = F * d * (b/a)$$

When the ratio b/a is greater than 1 the charge produced by transverse arrangement will be greater than the amount generated by longitudinal arrangement.

Properties of Piezo Electric-Crystal

The following are the properties of the Piezoelectric Crystals.

1. The piezoelectric material has high stability.
2. It is available in various shapes and sizes.
3. The piezoelectric material has output insensitive to temperature and humidity.

Uses of Piezoelectric Crystal

The following are the uses of the Piezoelectric transducers.

1. The piezoelectric material has high stability and hence it is used for stabilizing the electronic oscillator.
2. The ultrasonic generators use the piezoelectric material. This generator is used in SONAR for underwater detection and in industrials apparatus for cleaning.
3. It is used in microphones and speakers for converting the electric signal into sound.
4. The piezoelectric material is used in electric lighter.

The transducer has low output, and hence external circuit is associated with it.

WHAT IS PHOTOELECTRIC TRANSDUCER?

The photoelectric transducer can be defined as, a transducer which changes the energy from the light to electrical. It can be designed with the semiconductor material.

The photoelectric transducer converts the light energy into electrical energy. It is made of semiconductor material. The photoelectric transducer uses a photosensitive element, which ejects the electrons when the beam of light absorbs through it.

This transducer utilizes an element like photosensitive which can be used for ejecting the electrons as the light beam soak ups through it. The electron discharges can change the photosensitive element's property. Therefore the flowing current stimulates within the devices. The flow of the current's magnitude can be equivalent to the whole light absorbed with the photosensitive element.

The diagram of the photoelectric transducer is shown below. This transducer soak ups the light radiation which drops over the semiconductor material. The light absorption can boost the electrons in the material, & therefore the electrons begin to move. The electron mobility can generate three effects like

- The material resistance will be changed.
- The semiconductor's o/p current will be changed.
- The semiconductor's o/p voltage will be changed.

Photoelectric Transducer Classification

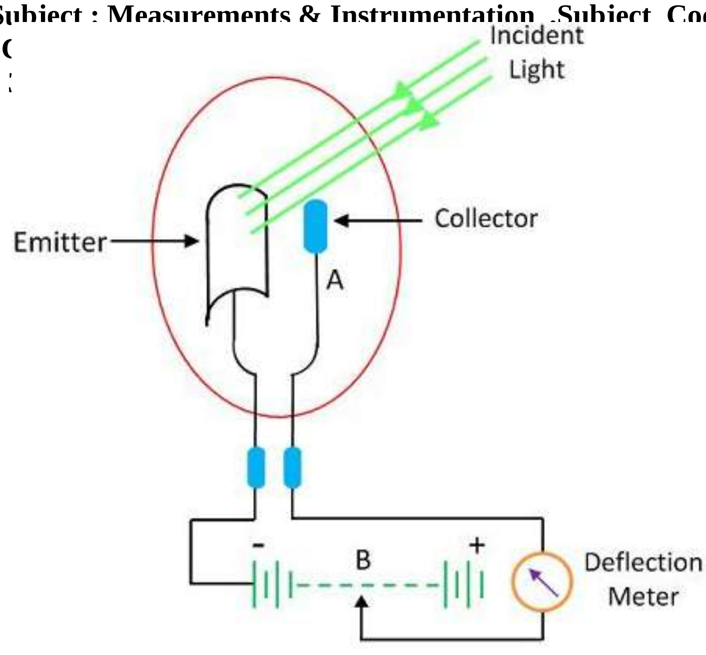
- Photo emissive Cell
- Photoconductive Cell
- Photo-voltaic cell

Classification of Photoelectric Transducers

The photoelectric transducers are classified into following ways.

1. PHOTOEMISSIVE CELL

The Photoemissive cell converts the photons into electric energy. It consists the anode rode and the cathode plate. The anode and cathode are coated with a Photoemissive material called caesium antimony.



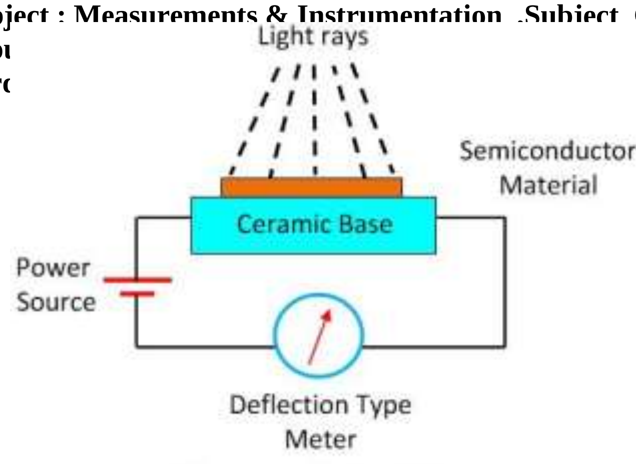
Photoemissive Cell

When the radiation of light fall on cathode plates the electrons starts flowing from anode to cathode. Both the anode and the cathode are sealed in a closed, opaque evacuated tube. When the radiation of light fall on the sealed tube, the electrons starts emitting from the cathode and moves towards the anode.

The anode is kept to the positive potential. Thus, the photoelectric current starts flowing through the anode. **The magnitude of the current is directly proportional to the intensity of light passes through it.**

2. PHOTOCONDUCTIVE CELL

The photoconductive cell converts the light energy into an electric current. It uses the semiconductor material like cadmium selenide, Ge, Se, as a photo sensing element.

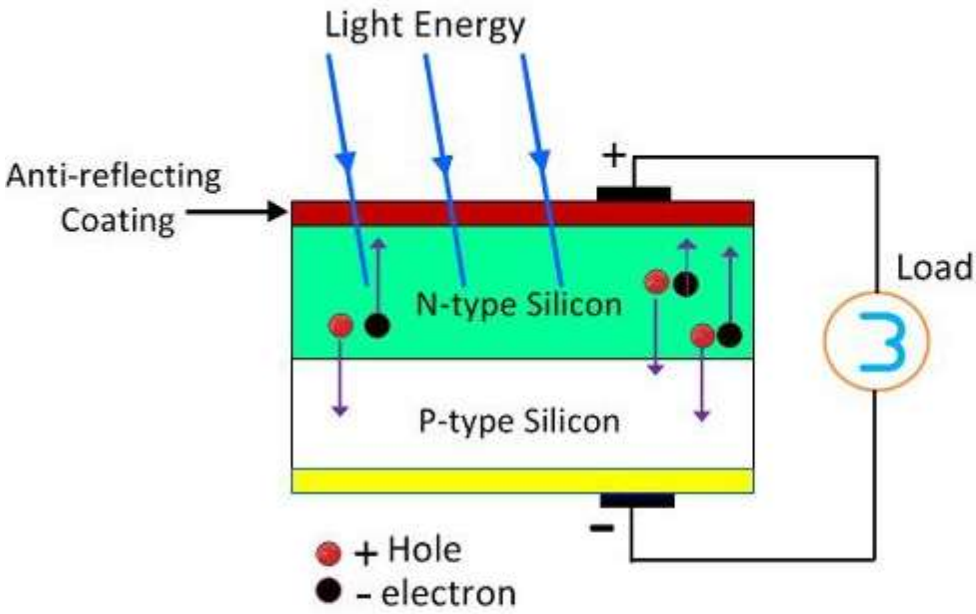


Photoconductive Cell

When the beam of light falls on the semiconductor material, their conductivity increases and the material works like a closed switch. The current starts flowing into the material and deflects the pointer of the meter.

3. PHOTO-VOLTAIC CELL

The photovoltaic cell is the type of active transducer. The current starts flowing into the photovoltaic cell when the load is connected to it. The silicon and selenium are used as a semiconductor material. When the semiconductor material absorbs heat, the free electrons of the material starts moving. This phenomenon is known as the photovoltaic effect.



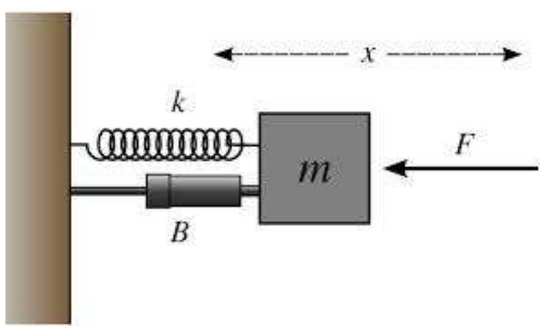
Photovoltaic Cell

The movements of electrons develop the current in the cell, and the current is known as the photoelectric current.

MECHANICAL, HYDRAULIC & PNUMATICS AMLIFYING ELEMENTS

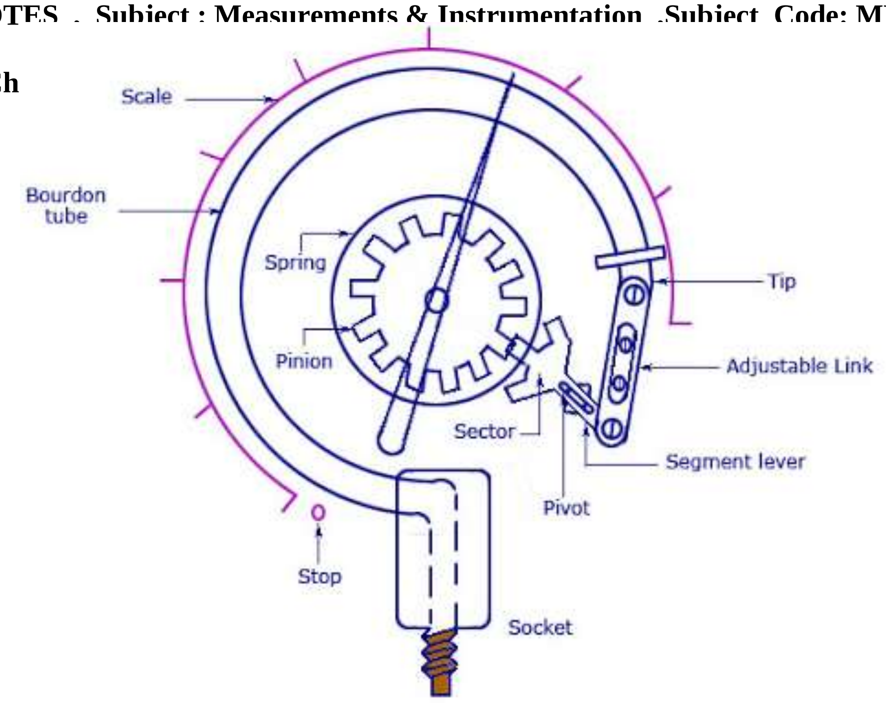
1.MECHANICAL AMPLIFYING ELEMENTS

A **mechanical** amplifier, or a **mechanical amplifying element**, is a linkage mechanism that **amplifies** the magnitude of **mechanical** quantities such as force, displacement, velocity, acceleration and torque in linear and rotational systems.



Spring Mass System

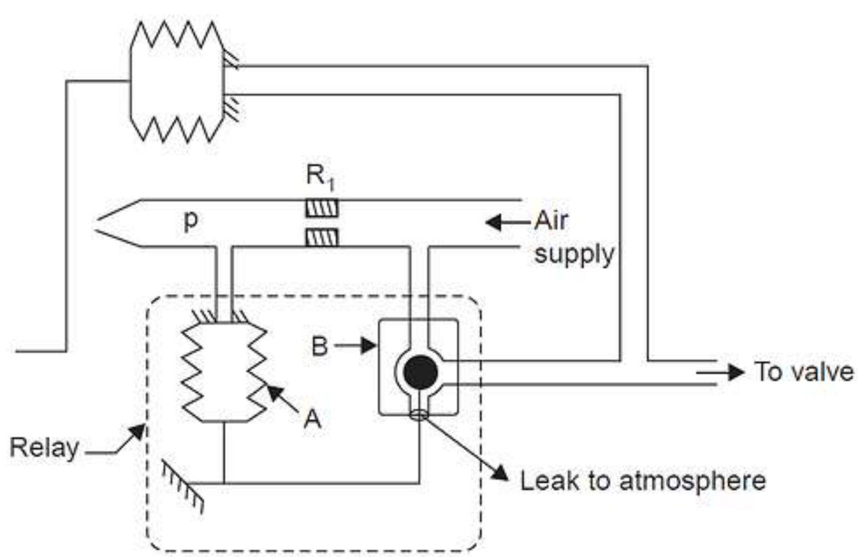
Ch



Bourdon Tube Pressure Gauge

In Bourdon tube pressure gauge Pinion , sector , C-clamp, Pointer, Link act as Mechanical amplifying Elements.

(2) PNEUMATIC AMPLIFYING ELEMENTS: A simple pneumatic relay being used as an amplifying element is shown below. These elements are used where abundant supply of compressed air is available, e.g. in factories and workshop



Pneumatic Relay working:

A power amplifier is needed to act on a large control valve. A pneumatic relay is a power amplifier. The relay is composed of small volume A bellows and a specially designed valve. In the nozzle pressure is transmitted to bellows A and, as the pressure increases, the bellows expand the ball down the valve B cavity, decreasing the leakage to the atmosphere and increasing the pressure Pv.

The relay is made in such a way that the output pressure Pv is proportional to the signal pressure p. The airflow of the relay will be much greater than the flow rate through the nozzle, since the resistance to the air supply may be too small compared to the resistance Rr in the nozzle. A relay valve can be designed as direct acting, in which case the outlet pressure is directly proportional to the inlet pressure, or it can be reverse acting, in which case the outlet pressure is inversely proportional to the inlet pressure.

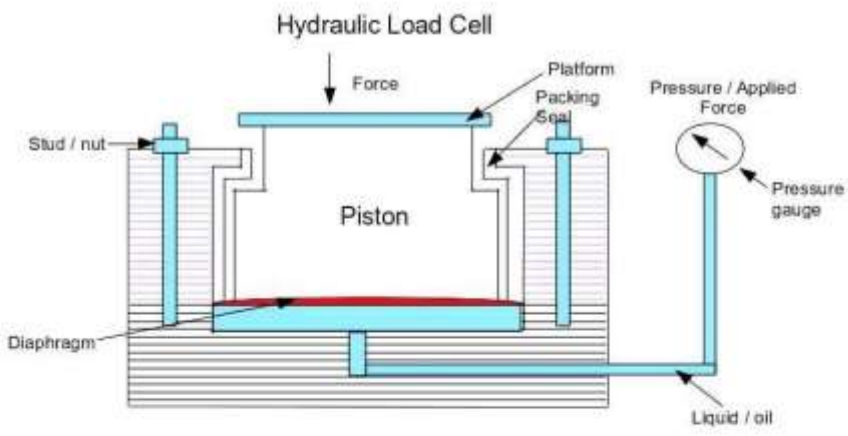
3. HYDRAULIC AMPLIFYING ELEMENTS

Hydraulic load cells are force-balance devices, measuring weight as a change in pressure of the internal filling fluid. In **hydraulic load cell**, a **load** or force acting on a loading head is transferred to a piston that in turn compresses a filling fluid confined within an elastomeric diaphragm chamber.

Principle of Hydraulic Load cell

When a force is applied on a liquid medium contained in a confined space, the pressure of the liquid increases. This increase in pressure of the liquid is proportional to the applied force. Hence a measure of the increase in pressure of the liquid becomes a measure of the applied force when calibrated.

Hydraulic Load Cell



The main parts of a hydraulic load cell are as follows

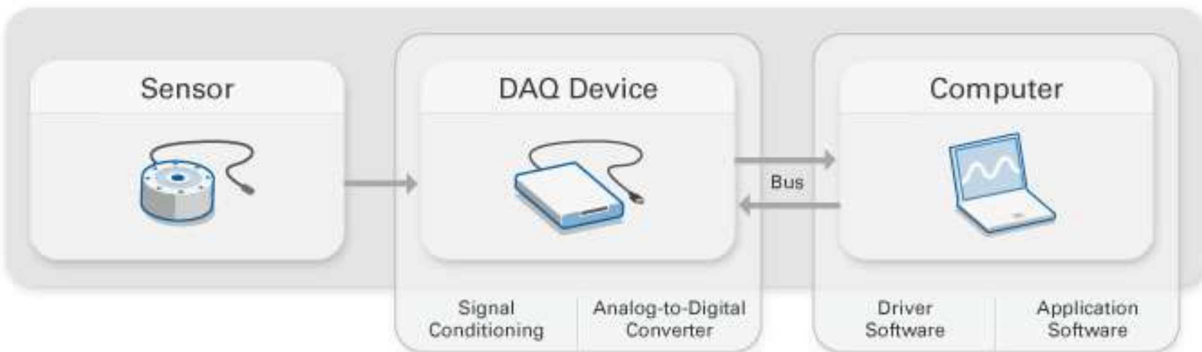
- A dirphragm
- A piston with a loading platform (as shown in figure) placed on top of the diaphragm.
- A liquid medium which is under a pre-loaded pressure is on the other side of the diaphragm.
- A pressure gauge (bourdon tube type) connected to the liquid medium.

Operation of Hydraulic Load Cell

- The force to be measured is applied to the piston.
- The appilied force moves the piston downwards and deflects the diaphragm and this deflection of the diaphragm increases the pressure in the liquid medium (oil).
- This increase in pressure of the liquid medium is proportional to the applied force. The increase in pressure is measured by the pressure gauge which is connected to the liquid medium.
- The pressure is calibrated in force units and hence the indication in the pressure gauge becomes a measure of the force applied on the piston.

DATA ACQUISITION SYSTEM

Data acquisition (DAQ) is the process of measuring an electrical or physical phenomenon such as voltage, current, temperature, pressure, or sound with a computer. A DAQ system consists of sensors, DAQ measurement hardware, and a computer with programmable software. Compared to traditional measurement systems, PC-based DAQ systems exploit the processing power, productivity, display, and connectivity capabilities of industry-standard computers providing a more powerful, flexible, and cost-effective measurement solution.



Data acquisition system (DAS)

Data acquisition system (DAS) is a computerized system that collects data from the real world, converts it into the form of electrical signals and do required processing on it for storage, and presentation on computers.

The complete system is controlled and operated by a software application. This software application is developed by using general-purpose high-level programming languages like C, C^{++}, java, etc.

These systems are used in industrial and commercial fields. They are used for collecting, storing and processing of data.

The data acquisition system can be divided into two types:

- Analog data acquisition system
- Digital data acquisition system

The analog data acquisition system gives an analog output whereas the digital data acquisition system gives a digital output. Analog DAS is used when wide frequency width is required or when lower accuracies can be tolerated.

Digital DAS is used when physical quantity being monitored has a narrow bandwidth (i.e. when the quantity varies slowly). Also, high accuracy and low per channel cost are required. These are more complex than analog DAS.

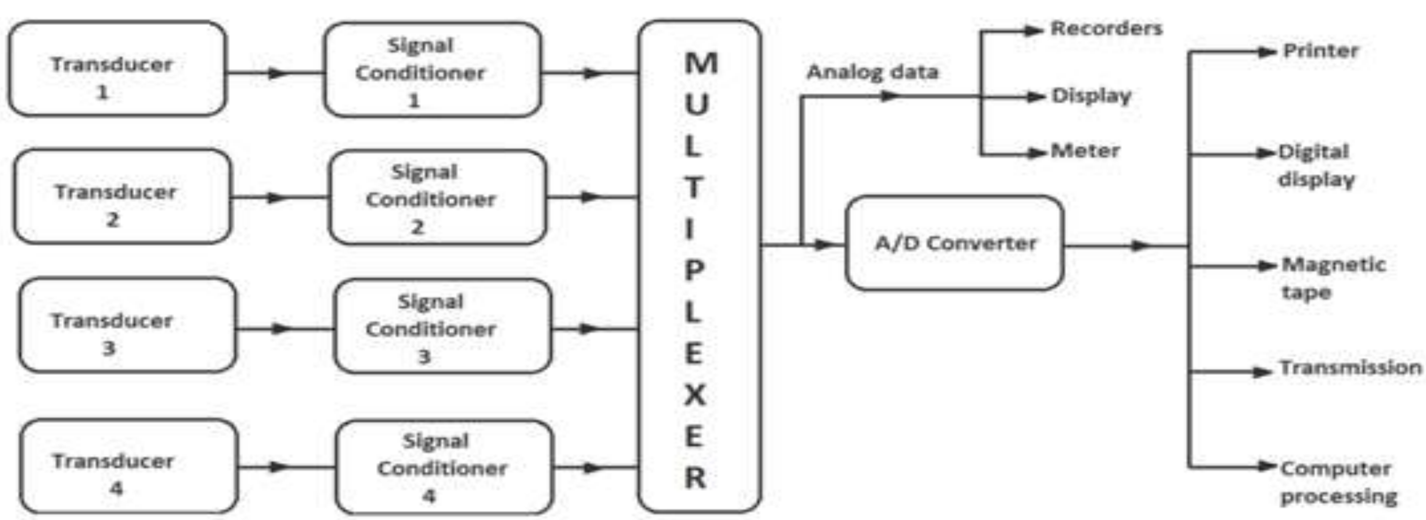
The digital data have more advantages over analog data. Some of those are:

- easy and fast processing,
- easy and fast transmission,
- easy display,
- less storage space is required,
- more accurate.

Due to these advantages, mostly the digital data acquisition system is preferred.

Data Acquisition System Block Diagram

A generalized **data acquisition system block diagram** is shown in Figure.



Generalised Data Acquisition System

The function of each block is as under:

Transducers: They are converting physical quantities (such as temperature, pressure, etc.) into electrical quantities, or measuring electrical quantities directly. They collect data from the physical world.

The most commonly used transducers are:

- RTDs, thermocouples, and thermistors for temperature measurements.
- Photosensors for light measurements.
- Strain gages, piezoelectric transducers for force and pressure measurements.
- Microphone for sound measurements.
- Potentiometer, LVDT, optical encoder for position and displacement measurements.

Signal Conditioning Unit: The signal produced by the transducers may or may not be very suitable for our system to work properly. It may be very weak, very strong or may have some noise.

To convert this signal into the most suitable form, amplification, and filtration is done respectively by signal conditioning unit. So the signal conditioning unit converts electrical signals in the most suitable form.

Multiplexer: The multiplexer receives multiple analog inputs and provides a single output signal according to the requirements. If a separate channel is used for each quantity, the cost of installation, maintenance, and periodic replacement becomes high. Therefore, a single channel is used which is shared by various quantities.

Analog to Digital (A/D) Converters: The data is converted into digital form by A/D converters.

After the conversion of data into digital form, it is displayed with the help of oscilloscopes, numerical displays, panel meters to monitor the complete system.

Also, the data can be either permanently or temporarily stored or recorded according to the requirement. The data is recorded on optical, ultraviolet, stylus or ink recorders for future use.

Objectives of Data Acquisition System

- It must collect the necessary data at the correct speed.
- It must use all the data efficiently to inform the operator about the state of the system.
- It must monitor the complete system operation to maintain on-line optimum and safe operations.
- It must be able to summarize and store data for the diagnosis of operation and record purpose.
- It must be flexible for future requirements.
- It must be reliable and not have a downtime of more than 0.1%.
- It must provide an effective communication system.

Applications of Data Acquisition System

The data acquisition system is used in industrial and scientific fields like aerospace, biomedical and telemetry industries.

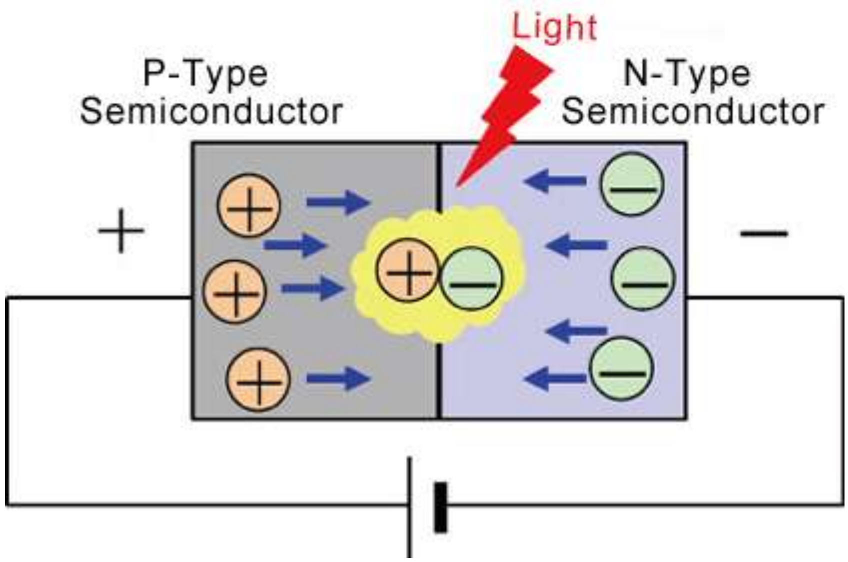
DATA DISPLAY & STORAGE DEVICE

A **display device** is an output device for presentation of information in visual[1] or tactile form (

1. **LIGHT EMITTING DIODE-** The "**Light Emitting Diode**" or LED as it is more commonly called, is basically just a specialised type of diode as they have very similar electrical characteristics to a PN junction diode. This means that an LED will pass current in its forward direction but block the flow of current in the reverse direction.

Light emitting diodes are made from a very thin layer of fairly heavily doped semiconductor material and depending on the semiconductor material used and the amount of doping, when forward biased an LED will emit a coloured light at a particular spectral wavelength.

When the diode is forward biased, electrons from the semiconductors conduction band recombine with holes from the valence band releasing sufficient energy to produce photons which emit a monochromatic (single colour) of light. Because of this thin layer a reasonable number of these photons can leave the junction and radiate away producing a coloured light output.



2. LIQUID CRYSTAL DIODE (LCD)

The LCD is defined as the **diode** that uses **small cells** and the **ionised gases** for the **production of images**. The LCD works on the **modulating property of light.** The light modulation is t**he technique of sending and receiving the signal** through the **light**. The **liquid crystal consumes** a small amount of energy because they are the **reflector and the transmitter of light.** It is normally used for **seven segmental display.**

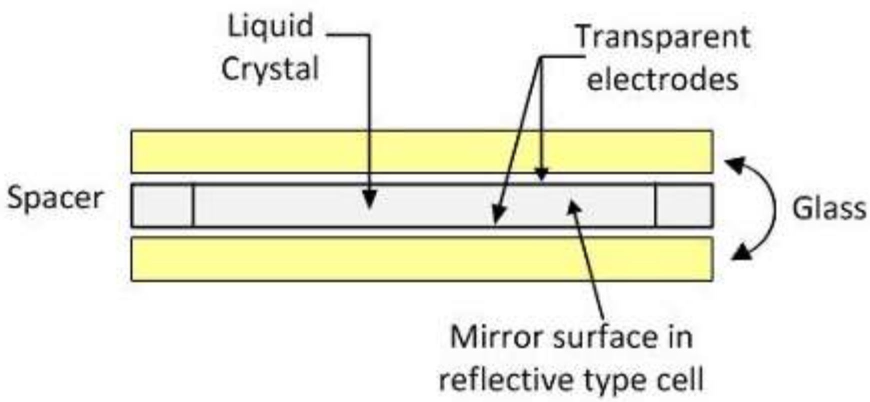
CONSTRUCTION OF LCD

The liquid crystals are the organic compound which is in liquid form and shows the property of optical crystals. The layer of liquid crystals is deposited on the inner surface of glass electrodes for the scattering of light. The liquid crystal cell is of two types; they are Transmittive Type and the Reflective Type.

Transmittive Type – In transmitter cell both the glass sheets are transparent so that the light is scattered in the forward direction when the cell becomes active.

Reflective Type – The reflective type cell consists the reflecting surface of the glass sheet on one end. The light incident on the front surface of the cell is scattered by the activated cell.

Both the reflective and transmittive type cells appear brights, even under small ambient light conditions.

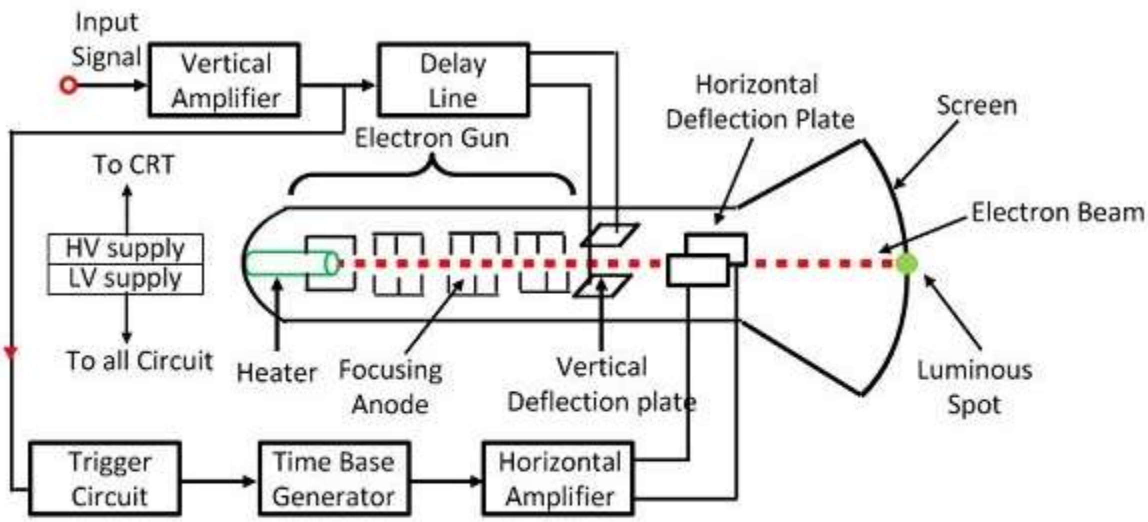


3. CATHODE RAY OSCILLOSCOPE (CRO)

The cathode ray oscilloscope (CRO) is a type of electrical instrument which is used for showing the measurement and analysis of waveforms and others electronic and electrical phenomenon. It is a very fast X-Y plotter shows the input signal versus another signal or versus time. The CROs are used to analyse the waveforms, transient, phenomena, and other time-varying quantities from a very low-frequency range to the radio frequencies.

Working of Cathode Ray Oscilloscope

When the electron is injected through the electron gun, it passes through the control grid. The control grid controls the intensity of electron in the vacuum tube. If the control grid has high negative potential, then it allows only a few electrons to pass through it. Thus, the dim spot is produced on the lightning screen. If the negative potential on the control grid is low, then the bright spot is produced. Hence the intensity of light depends on the negative potential of the control grid.



After moving the control grid the electron beam passing through the focusing and accelerating anodes. The accelerating anodes are at a high positive potential and hence they converge the beam at a point on the screen.

STOARGE:

A **recorder** records electrical and non-electrical quantities as a **function** of time. Currents and voltages can be recorded directly while the non-electrical quantities are recorded indirectly by first converting them to equivalent currents or voltages with the help of sensors or transducers

- A recorder records electrical and non electrical quantities as a function of time.
- Currents and voltages can be recorded directly while the non electrical quantities are recorded indirectly by first converting them to equivalent currents or voltages with the help of sensors or transducers.

UNIT-III (PART-I)

SUB: INSTRUMENTATION & CONTROL (ME 210C); SEM-4th B.tech

Concept of control system

A **control system** manages commands, directs or regulates the behavior of other devices or systems using control loops. It can range from a single home heating controller using a thermostat controlling a domestic boiler to large Industrial control systems which are used for controlling processes or machines. A control system is a system, which provides the desired response by controlling the output. The following figure shows the simple block diagram of a control system.

Examples − Traffic lights control system, washing machine

Traffic lights control system is an example of control system. Here, a sequence of input signal is applied to this control system and the output is one of the three lights that will be on for some duration of time. During this time, the other two lights will be off. Based on the traffic study at a particular junction, the on and off times of the lights can be determined. Accordingly, the input signal controls the output. So, the traffic lights control system operates on time basis.

Classification of Control Systems

Based on some parameters, we can classify the control systems into the following ways.

Continuous time and Discrete-time Control Systems

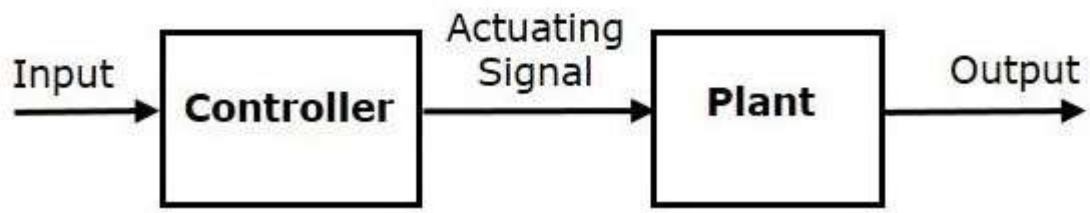
- Control Systems can be classified as continuous time control systems and discrete time control systems based on the **type of the signal** used.
- In **continuous time** control systems, all the signals are continuous in time. But, in **discrete time** control systems, there exists one or more discrete time signals.

Open Loop and Closed Loop Control Systems

Control Systems can be classified as open loop control systems and closed loop control systems based on the **feedback path**.

In **open loop control systems**, output is not fed-back to the input. So, the control action is independent of the desired output.

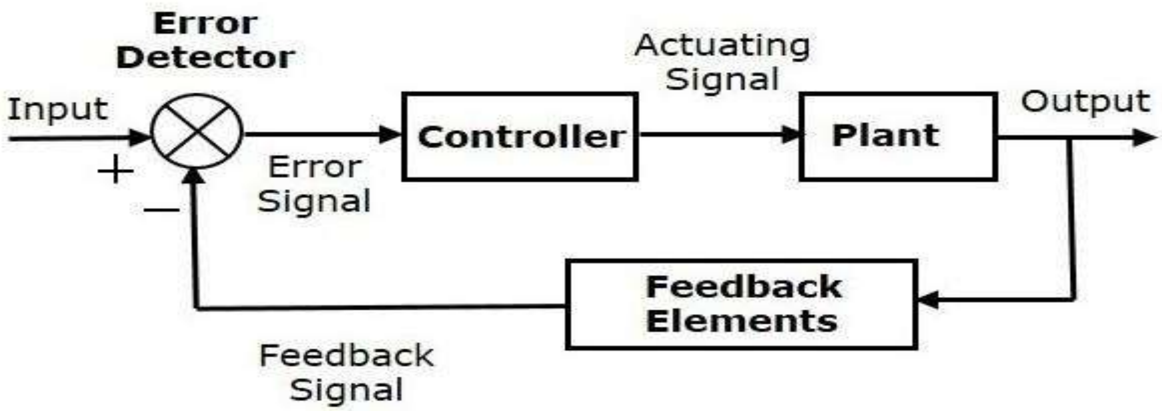
The following figure shows the block diagram of the open loop control system.



Here, an input is applied to a controller and it produces an actuating signal or controlling signal. This signal is given as an input to a plant or process which is to be controlled. So, the plant produces an output, which is controlled. The traffic lights control system which we discussed earlier is an example of an open loop control system.

In **closed loop control systems**, output is fed back to the input. So, the control action is dependent on the desired output.

The following figure shows the block diagram of negative feedback closed loop control system.



The error detector produces an error signal, which is the difference between the input and the feedback signal. This feedback signal is obtained from the block (feedback elements) by considering the output of the overall system as an input to this block. Instead of the direct input, the error signal is applied as an input to a controller.

So, the controller produces an actuating signal which controls the plant. In this combination, the output of the control system is adjusted automatically till we get the desired response. Hence, the closed loop control systems are also called the automatic control systems. Traffic lights control system having sensor at the input is an example of a closed loop control system.

The differences between the open loop and the closed loop control systems are mentioned in the following table.

Open Loop Control Systems	Closed Loop Control Systems
Control action is independent of the desired output.	Control action is dependent of the desired output.
Feedback path is not present.	Feedback path is present.
These are also called as **non-feedback control systems**.	These are also called as **feedback control systems**.
Easy to design.	Difficult to design.
These are economical.	These are costlier.
Inaccurate.	Accurate.

If either the output or some part of the output is returned to the input side and utilized as part of the system input, then it is known as **feedback**. Feedback plays an important role in order to improve the performance of the control systems. In this chapter, let us discuss the types of feedback & effects of feedback.

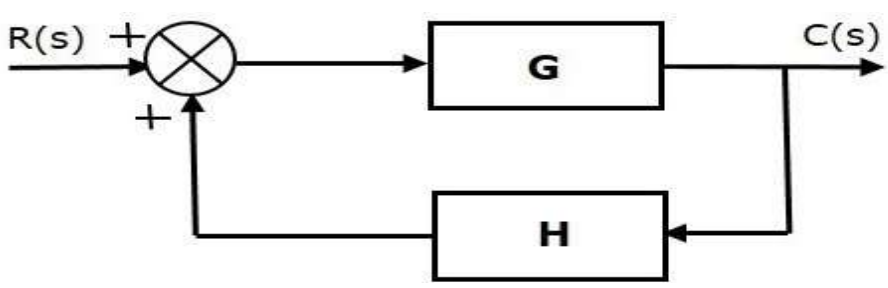
Types of Feedback

There are two types of feedback −

- Positive feedback
- Negative feedback

Positive Feedback

The positive feedback adds the reference input, R(s)R(s) and feedback output. The following figure shows the block diagram of **positive feedback control system**



he concept of transfer function will be discussed in later chapters. For the time being, consider the transfer function of positive feedback control system is,

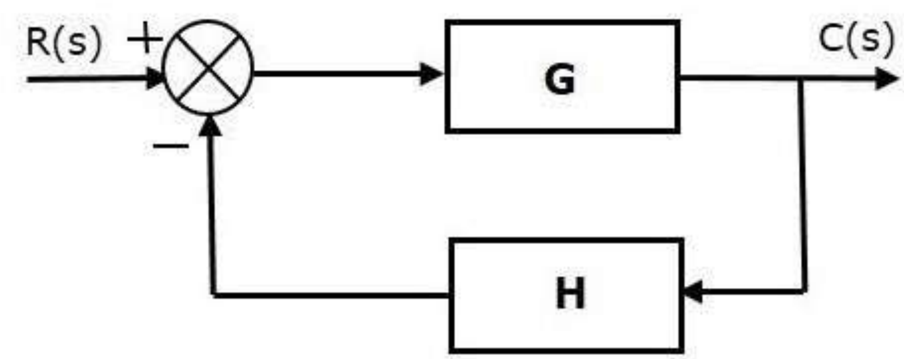
$$T = \frac{G}{1-GH} \qquad \text{(Equation 1)}$$

Where,

- **T** is the transfer function or overall gain of positive feedback control system.
- **G** is the open loop gain, which is function of frequency.
- **H** is the gain of feedback path, which is function of frequency.

Negative Feedback

Negative feedback reduces the error between the reference input, R(s)R(s) and system output. The following figure shows the block diagram of the **negative feedback control system**.



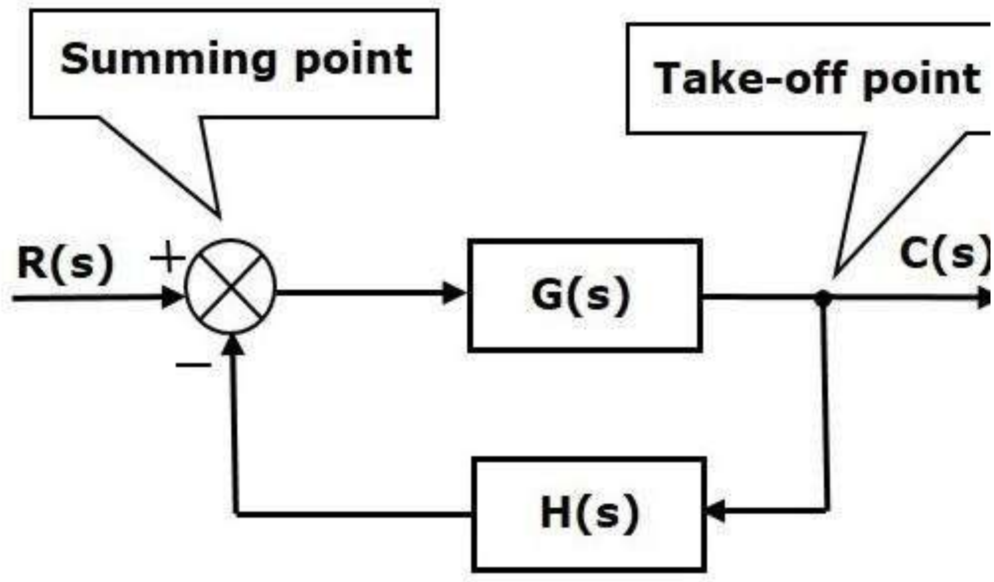
TRANSFER FUNCTION REPRESENTATION

Block Diagrams

Block diagrams consist of a single block or a combination of blocks. These are used to represent the control systems in pictorial form.

Basic Elements of Block Diagram

The basic elements of a block diagram are a block, the summing point and the take-off point. Let us consider the block diagram of a closed loop control system as shown in the following figure to identify these elements.



The above block diagram consists of two blocks having transfer functions G(s) and H(s). It is also having one summing point and one take-off point. Arrows indicate the direction of the flow of signals. Let us now discuss these elements one by one.

Block

The transfer function of a component is represented by a block. Block has single input and single output.

The following figure shows a block having input X(s), output Y(s) and the transfer function G(s).

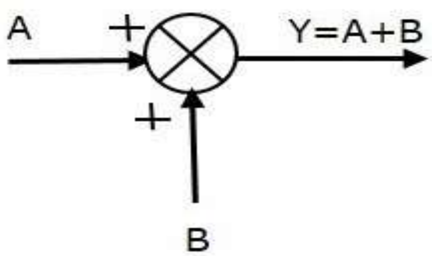
Transfer Function, $$G(s) = \frac{Y(s)}{X(s)}$$

$$\Rightarrow Y(s) = G(s)X(s)$$

Summing Point

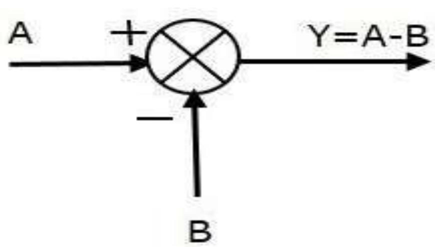
The summing point is represented with a circle having cross (X) inside it. It has two or more inputs and single output. It produces the algebraic sum of the inputs. It also performs the summation or subtraction or combination of summation and subtraction of the inputs based on the polarity of the inputs. Let us see these three operations one by one.

The following figure shows the summing point with two inputs (A, B) and one output (Y). Here, the inputs A and B have a positive sign. So, the summing point produces the output, Y as **sum of A and B** i.e. = A + B.



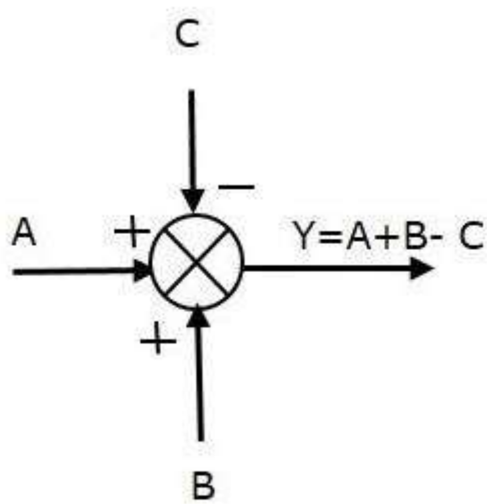
The following figure shows the summing point with two inputs (A, B) and one output (Y). Here, the inputs A and B are having opposite signs, i.e., A is having positive sign and B is having negative sign. So, the summing point produces the output **Y** as the **difference of A and B** i.e

Y = A + (-B) = A - B.



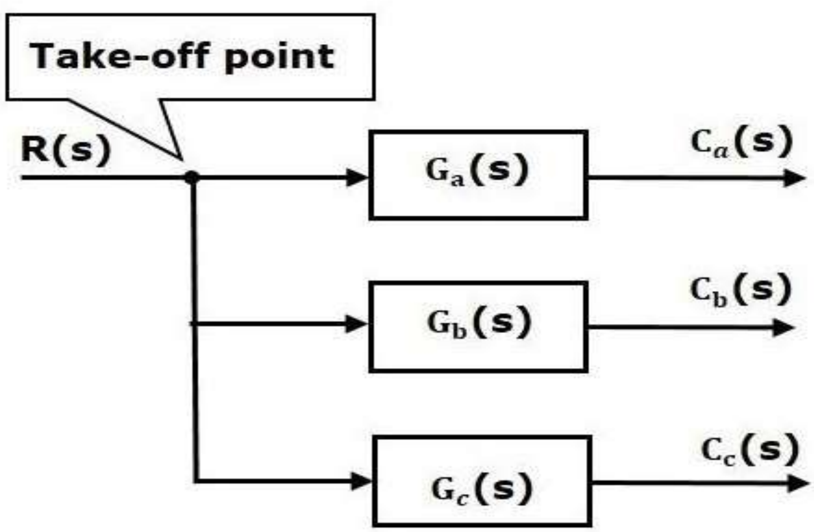
The following figure shows the summing point with three inputs (A, B, C) and one output (Y). Here, the inputs A and B are having positive signs and C is having a negative sign. So, the summing point produces the output **Y** as

Y = A + B + (−C) = A + B − C.

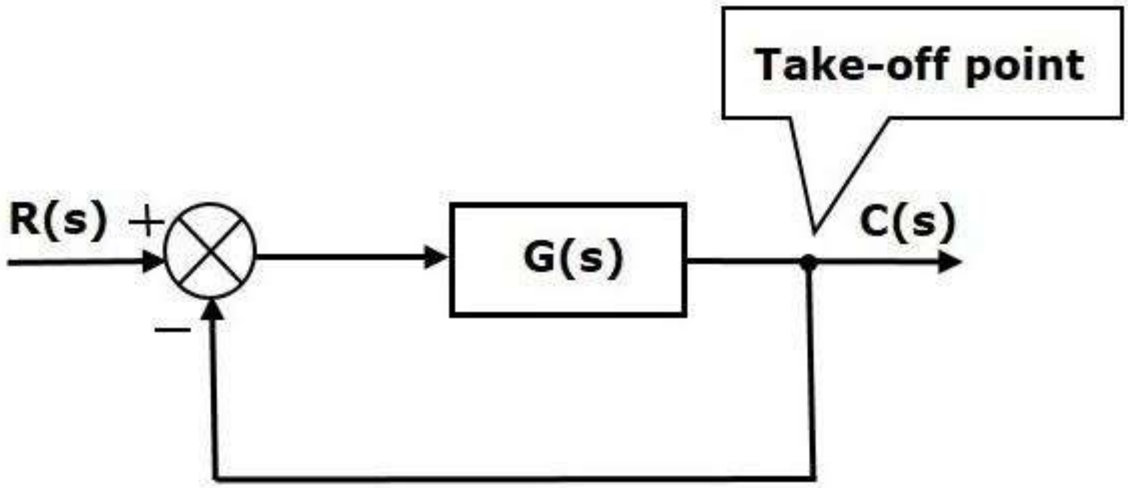


Take-off Point

The take-off point is a point from which the same input signal can be passed through more than one branch. That means with the help of take-off point, we can apply the same input to one or more blocks, summing points.In the following figure, the take-off point is used to connect the same input, R(s) to two more blocks.



In the following figure, the take-off point is used to connect the output C(s), as one of the inputs to the summing point.



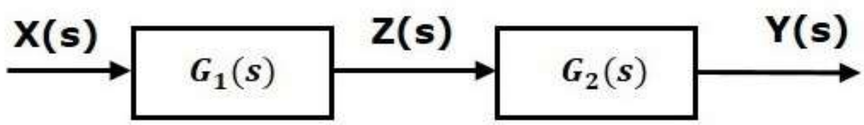
Block diagram algebra is nothing but the algebra involved with the basic elements of the block diagram. This algebra deals with the pictorial representation of algebraic equations.

Basic Connections for Blocks

There are three basic types of connections between two blocks.

Series Connection

Series connection is also called **cascade connection**. In the following figure, two blocks having transfer functions G1(s)G1(s) and G2(s)G2(s) are connected in series.



For this combination, we will get the output $Y(s)$ as

$$Y(s) = G_2(s)Z(s)$$

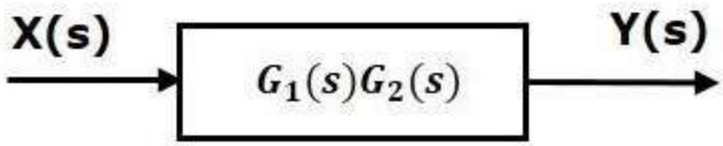
Where, $Z(s) = G_1(s)X(s)$

$$\Rightarrow Y(s) = G_2(s)[G_1(s)X(s)] = G_1(s)G_2(s)X(s)$$

$$\Rightarrow Y(s) = \{G_1(s)G_2(s)\}X(s)$$

Compare this equation with the standard form of the output equation, $Y(s) = G(s)X(s)$. Where, $G(s) = G_1(s)G_2(s)$.

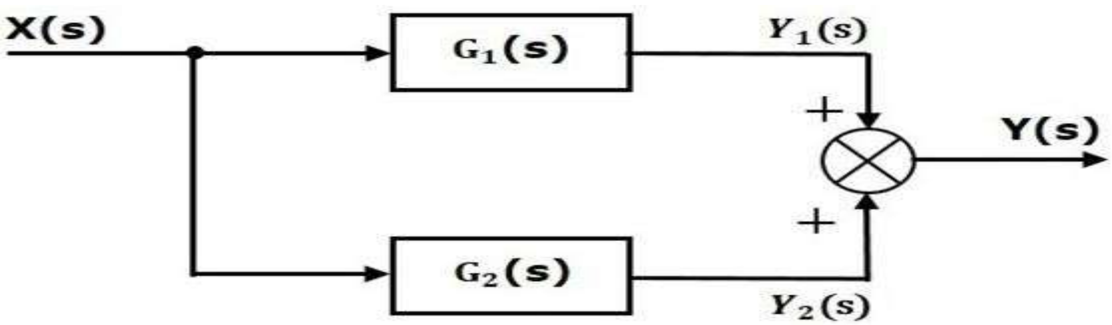
That means we can represent the **series connection** of two blocks with a single block. The transfer function of this single block is the **product of the transfer functions** of those two blocks. The equivalent block diagram is shown below.



Similarly, you can represent series connection of ‘n’ blocks with a single block. The transfer function of this single block is the product of the transfer functions of all those ‘n’ blocks.

Parallel Connection

The blocks which are connected in **parallel** will have the **same input**. In the following figure, two blocks having transfer functions G1(s)G1(s) and G2(s)G2(s) are connected in parallel. The outputs of these two blocks are connected to the summing point.



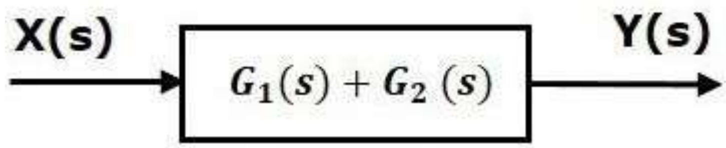
$$Y(s) = Y_1(s) + Y_2(s)$$

$Y_1(s) = G_1(s)X(s)$ and $Y_2(s) = G_2(s)X(s)$

$$\Rightarrow Y(s) = G_1(s)X(s) + G_2(s)X(s) = \{G_1(s) + G_2(s)\}X(s)$$

$$G(s) = G_1(s) + G_2(s).$$

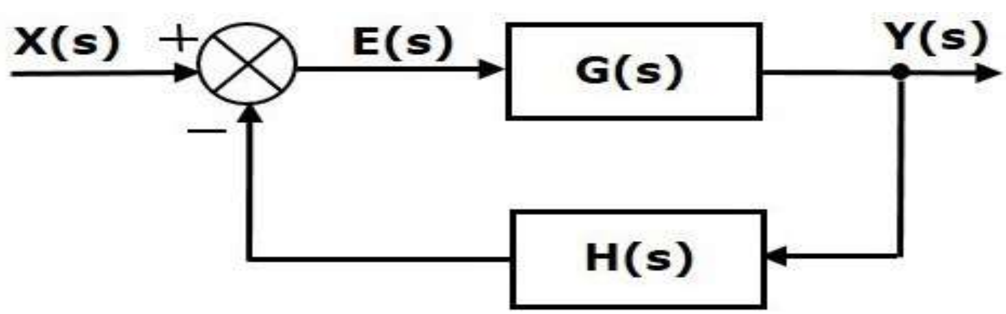
That means we can represent the **parallel connection** of two blocks with a single block. The transfer function of this single block is the **sum of the transfer functions** of those two blocks. The equivalent block diagram is shown below.



Similarly, you can represent parallel connection of 'n' blocks with a single block. The transfer function of this single block is the algebraic sum of the transfer functions of all those 'n' blocks.

Feedback Connection

As we discussed in previous chapters, there are two types of **feedback** — positive feedback and negative feedback. The following figure shows negative feedback control system. Here, two blocks having transfer functions G(s)G(s) and H(s)H(s) form a closed loop.



The output of the summing point is -

$$E(s) = X(s) - H(s)Y(s)$$

The output $Y(s)$ is -

$$Y(s) = E(s)G(s)$$

Substitute $E(s)$ value in the above equation.

$$Y(s) = \{X(s) - H(s)Y(s)\}G(s)\}$$

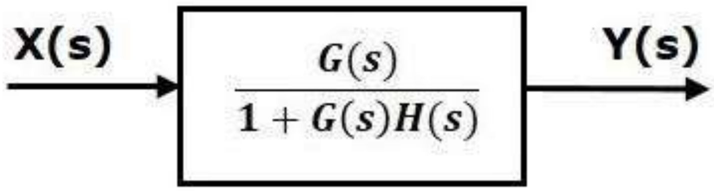
$$Y(s)\{1 + G(s)H(s)\} = X(s)G(s)\}$$

$$\Rightarrow \frac{Y(s)}{X(s)} = \frac{G(s)}{1 + G(s)H(s)}$$

Therefore, the negative feedback closed loop transfer function is :

$$\frac{G(s)}{1+G(s)H(s)}$$

This means we can represent the negative feedback connection of two blocks with a single block. The transfer function of this single block is the closed loop transfer function of the negative feedback. The equivalent block diagram is shown below.



UNIT-III (PART-II) TYPES OF CONTROLLERS:

Proportional Controller

The proportional controller produces an output, which is proportional to error signal.

$$u(t) \propto e(t)$$

$$\Rightarrow u(t) = K_P e(t)$$

Apply Laplace transform on both the sides -

$$U(s) = K_P E(s)$$

$$\frac{U(s)}{E(s)} = K_P$$

Therefore, the transfer function of the proportional controller is KPKP.

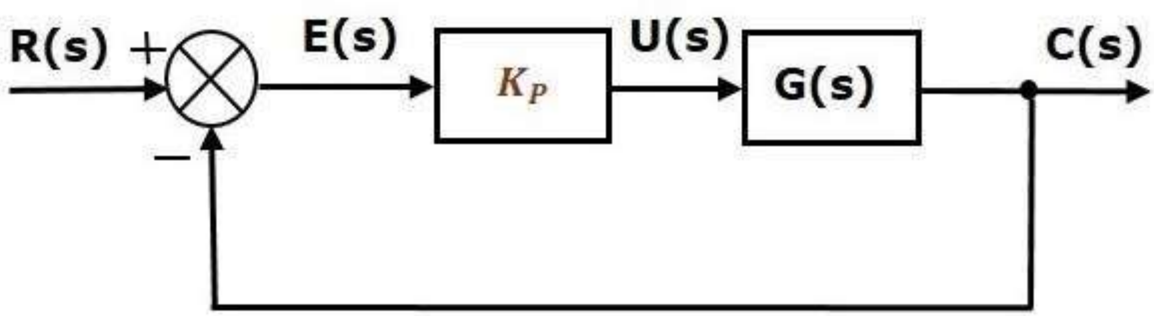
Where,

U(s) is the Laplace transform of the actuating signal u(t)

E(s) is the Laplace transform of the error signal e(t)

KP is the proportionality constant

The block diagram of the unity negative feedback closed loop control system along with the proportional controller is shown in the following figure.



Derivative Controller

The derivative controller produces an output, which is derivative of the error signal.

$$u(t) = K_D \frac{\mathrm{d}e(t)}{\mathrm{d}t}$$

Apply Laplace transform on both sides.

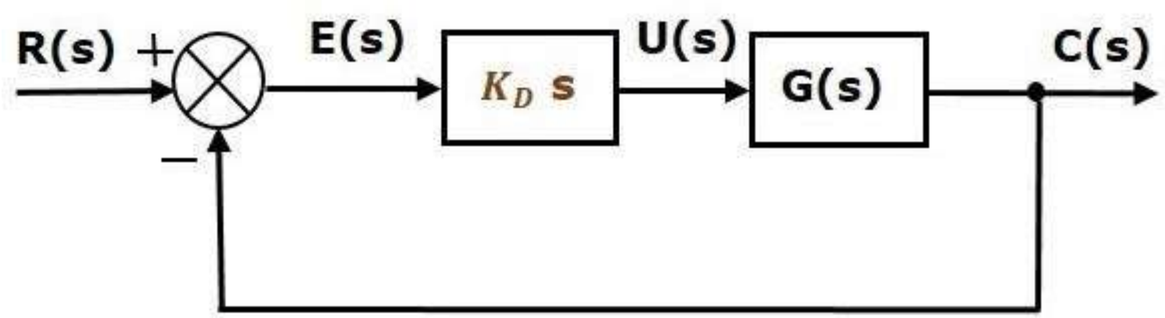
$$U(s) = K_D s E(s)$$

$$\frac{U(s)}{E(s)} = K_D s$$

Therefore, the transfer function of the derivative controller is $K_D s$.

Where, KD is the derivative constant.

The block diagram of the unity negative feedback closed loop control system along with the derivative controller is shown in the following figure.



The derivative controller is used to make the unstable control system into a stable one.

Integral Controller

The integral controller produces an output, which is integral of the error signal.

$$u(t) = K_I \int e(t)dt$$

Apply Laplace transform on both the sides -

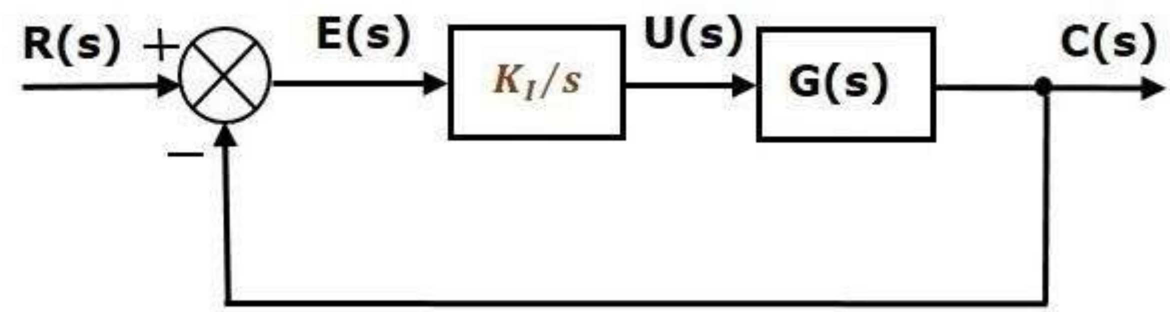
$$U(s) = \frac{K_I E(s)}{s}$$

$$\frac{U(s)}{E(s)} = \frac{K_I}{s}$$

Therefore, the transfer function of the integral controller is $\frac{K_I}{s}$.

Where, KIKI is the integral constant.

The block diagram of the unity negative feedback closed loop control system along with the integral controller is shown in the following figure.



The integral controller is used to decrease the steady state error.

Let us now discuss about the combination of basic controllers.

Proportional Derivative (PD) Controller

The proportional derivative controller produces an output, which is the combination of the outputs of proportional and derivative controllers.

$$u(t) = K_P e(t) + K_D \frac{\mathrm{d}e(t)}{\mathrm{d}t}$$

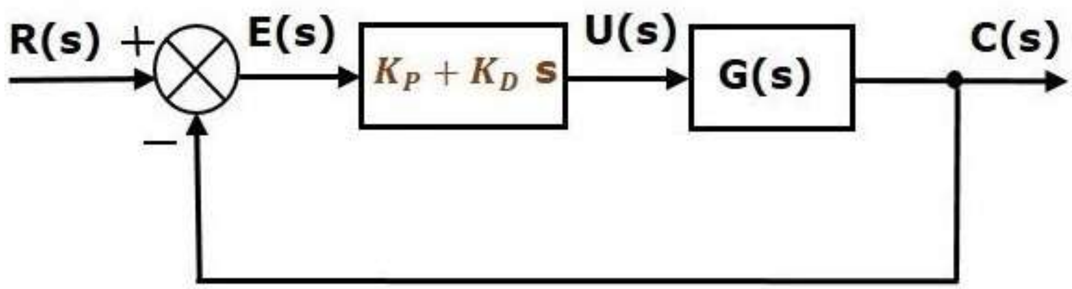
Apply Laplace transform on both sides -

$$U(s) = (K_P + K_D s)E(s)$$

$$\frac{U(s)}{E(s)} = K_P + K_D s$$

Therefore, the transfer function of the proportional derivative controller is K_P+K_Ds.

The block diagram of the unity negative feedback closed loop control system along with the proportional derivative controller is shown in the following figure.

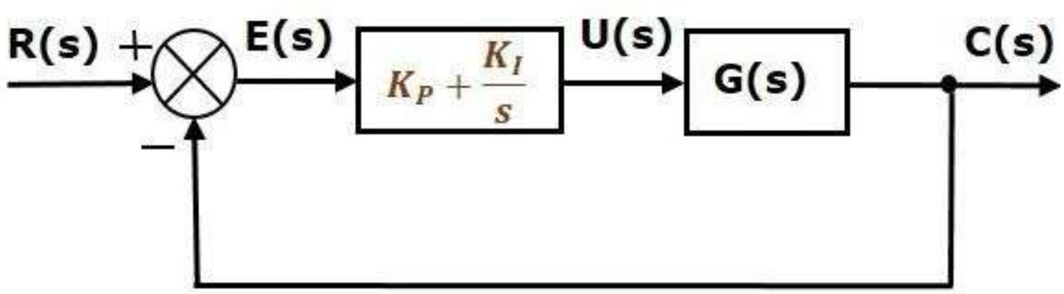


Proportional Integral (PI) Controller

The proportional integral controller produces an output, which is the combination of outputs of the proportional and integral controllers.

The proportional derivative controller is used to improve the stability of control system without affecting the steady state error.

The block diagram of the unity negative feedback closed loop control system along with the proportional integral controller is shown in the following figure.

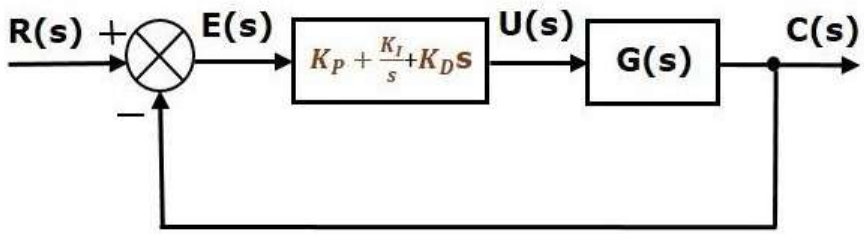


The proportional integral controller is used to decrease the steady state error without affecting the stability of the control system.

Proportional Integral Derivative (PID) Controller

The proportional integral derivative controller produces an output, which is the combination of the outputs of proportional, integral and derivative controllers.

The block diagram of the unity negative feedback closed loop control system along with the proportional integral derivative controller is shown in the figure



$$u(t) = K_P e(t) + K_I \int e(t)dt + K_D \frac{de(t)}{dt}$$

Apply Laplace transform on both sides -

$$U(s) = \left(K_P + \frac{K_I}{s} + K_D s\right) E(s)$$

$$\frac{U(s)}{E(s)} = K_P + \frac{K_I}{s} + K_D s$$

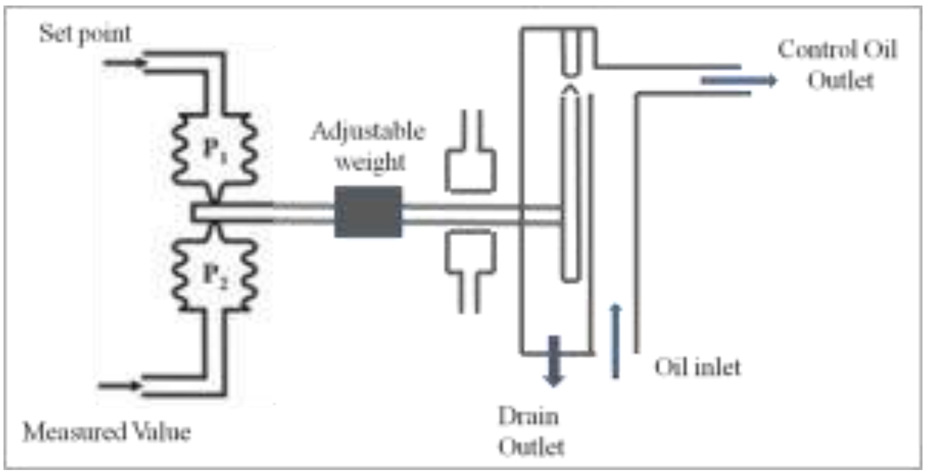
Therefore, the transfer function of the proportional integral derivative controller is $K_P + \frac{K_I}{s} + K_D s$.

TYPES OF PID CONTROL SYSTEM

As per the realization of the controllers these are divided as Pneumatic, Electronic and Digital controllers.

Pneumatic/hydraulic controller:

In earlier days all the process controllers (P, PI, PID) are pneumatic type where there is no so much advancement or progress in electronic components. The advantage of pneumatic controllers is its ruggedness, and major limitation is the slow response. This controllers are designed using mechanical components which operates according to air/liquid pressure. Due to this mechanical components these controllers are strong and insensitive to the temperatures in plant. However the response of mechanical components are slow compared to electronic components so these controller slower in response. These pneumatic controllers will acts on the difference between the air/liquid pressure of measured signal and set point signal. The best example of pneumatic/hydraulic controller is Hydraulic Governor which is placed to control the steam turbine parameters.



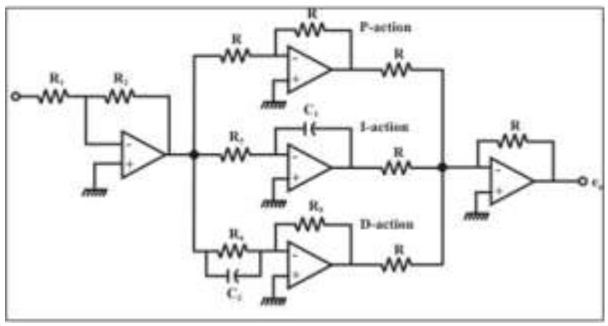
Hydraulic Controller

The simple scheme of implementation of Proportional controller is shown in figure. Here the set point and measured values are converted as the hydraulic pressure signals and applied to one end of the force beam. The force beam moves up/down according to the differential pressure of measured and set values. If the set point is high then the left end of the force beam moves down and the right end of the force beam moves up. the movement ratio of left and right end of beam can be controlled by the adjustable weight which can be called as proportional gain. Due to this movement of right end of the beam the flapper which is connected to it also moves upward towards fixed flapper. this reduces the nozzle gap between the fixed and moving flappers and hence reduces the oil flow through the nozzle. This in-turn increases the output control oil pressure and flow which is used drive the control valve actuator.

Electronic or Hardwired controller:

The later advancements in electronic components prompted the use of electronic components in control systems. The introduction of this electronic control systems enhanced the performance of the control much faster compared to mechanical systems. Due to the small size of electronic components the overall size of the controller became very small. However the system response is faster in this control system but it also has a disadvantage that it is very sensitive to temperature or any internal faults in components. There will be some internal and external noise because of electronic components and care has to be taken to filter the noise which otherwise affects overall response of the controller.

Nowadays the electronic controllers with advanced technology which are offering with very less noise with low sensitivity for the internal faults. A simple schematic of electronic PID controller is shown in below figure.



Electronic PID Controller

The PID controller is designed using basic electronic components like resistor, capacitor and op-amp comparator. The PID controller is a combination of Proportional gain, Integral component (Low-pass filter) and differential component (High Pass filter).

Output = Proportional gain * error + Low pass filter (error) + high pass filter (error).

The input comparator calculates the error between input set point and actual measured value. This error is parallelly applied to the gain amplifier, low pass filter and high pass filter. The parallel output is combined at the output node and given to the actuator to drive the actuator.

The biggest disadvantage in hardwired controller is that it is not flexible for changing the logic of the controller. That is if the PI controller is designed to control the liquid level in the tank, suppose with same controller should be used to control the temperature with PD design then the whole circuit should be redesigned for this operation. Hence it is not suitable in design and testing environment.

UNIT-IV FREQUENCY RESPONSE ANALYSIS (PART-I)

SUB: INSTRUMENTATION & CONTROL (ME 210C); SEM-4th B.tech

What is Frequency Response?

The response of a system can be partitioned into both the transient response and the steady state response. We can find the transient response by using Fourier integrals. The steady state response of a system for an input sinusoidal signal is known as the **frequency response**. In this chapter, we will focus only on the steady state response.

If a sinusoidal signal is applied as an input to a Linear Time-Invariant (LTI) system, then it produces the steady state output, which is also a sinusoidal signal. The input and output sinusoidal signals have the same frequency, but different amplitudes and phase angles. Let the input signal be

$$r(t) = A\sin(\omega_0 t)$$

The open loop transfer function will be −

$$G(s) = G(j\omega)$$

We can represent $G(j\omega)$ in terms of magnitude and phase as shown below.

$$G(j\omega) = |G(j\omega)|\angle G(j\omega)$$

Substitute, $\omega = \omega_0$ in the above equation.

$$G(j\omega_0) = |G(j\omega_0)|\angle G(j\omega_0)$$

The output signal is

$$c(t) = A|G(j\omega_0)|\sin(\omega_0 t + \angle G(j\omega_0))$$

- The **amplitude** of the output sinusoidal signal is obtained by multiplying the amplitude of the input sinusoidal signal and the magnitude of $G(j\omega)$ at $\omega = \omega_0$.
- The **phase** of the output sinusoidal signal is obtained by adding the phase of the input sinusoidal signal and the phase of $G(j\omega)$ at $\omega = \omega_0$.

Where, **A** is the amplitude of the input sinusoidal signal.

- **ω0** is angular frequency of the input sinusoidal

$\omega_0 = 2\pi f_0$

Here, f_0 is the frequency of the input sinusoidal signal. Similarly, you can follow the same procedure for closed loop control system.

Frequency Domain Specifications

The frequency domain specifications are

- **Resonant peak**
- **Resonant frequency**
- **Bandwidth**.

Consider the transfer function of the second order closed control system as

$$T(s) = \frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\delta\omega_n s + \omega_n^2}$$

Substitute, $s = j\omega$ in the above equation.

$$T(j\omega) = \frac{\omega_n^2}{(j\omega)^2 + 2\delta\omega_n(j\omega) + \omega_n^2}$$

$$\Rightarrow T(j\omega) = \frac{\omega_n^2}{-\omega^2 + 2j\delta\omega\omega_n + \omega_n^2} = \frac{\omega_n^2}{\omega_n^2\left(1 - \frac{\omega^2}{\omega_n^2} + \frac{2j\delta\omega}{\omega_n}\right)}$$

$$\Rightarrow T(j\omega) = \frac{1}{\left(1 - \frac{\omega^2}{\omega_n^2}\right) + j\left(\frac{2\delta\omega}{\omega_n}\right)}$$

Let, $\frac{\omega}{\omega_n} = u$ Substitute this value in the above equation.

$$T(j\omega) = \frac{1}{(1 - u^2) + j(2\delta u)}$$

Magnitude of $T(j\omega)$ is -

$$M = |T(j\omega)| = \frac{1}{\sqrt{(1 - u^2)^2 + (2\delta u)^2}}$$

Phase of $T(j\omega)$ is -

$$\angle T(j\omega) = -tan^{-1}\left(\frac{2\delta u}{1-u^2}\right)$$

Resonant Frequency

It is the frequency at which the magnitude of the frequency response has peak value for the first time. It is denoted by ω_r. At $\omega = \omega_r$, the first derivate of the magnitude of $T(j\omega)$ is zero.

Differentiate M with respect to u.

$$\frac{\mathrm{d}M}{\mathrm{d}u} = -\frac{1}{2}\left[(1-u^2)^2 + (2\delta u)^2\right]^{\frac{-3}{2}}\left[2(1-u^2)(-2u) + 2(2\delta u)(2\delta)\right]$$

$$\Rightarrow \frac{\mathrm{d}M}{\mathrm{d}u} = -\frac{1}{2}\left[(1-u^2)^2 + (2\delta u)^2\right]^{\frac{-3}{2}}\left[4u(u^2 - 1 + 2\delta^2)\right]$$

Substitute, $u = u_r$ and $\frac{\mathrm{d}M}{\mathrm{d}u} == 0$ in the above equation.

$$0 = -\frac{1}{2}\left[(1-u_r^2)^2 + (2\delta u_r)^2\right]^{-\frac{3}{2}}\left[4u_r(u_r^2 - 1 + 2\delta^2)\right]$$

$$\Rightarrow 4u_r(u_r^2 - 1 + 2\delta^2) = 0$$

$$\Rightarrow u_r^2 - 1 + 2\delta^2 = 0$$

$$\Rightarrow u_r^2 = 1 - 2\delta^2$$

$$\Rightarrow u_r = \sqrt{1-2\delta^2}$$

Substitute, $u_r = \frac{\omega_r}{\omega_n}$ in the above equation.

$$\frac{\omega_r}{\omega_n} = \sqrt{1-2\delta^2}$$

$$\Rightarrow \omega_r = \omega_n\sqrt{1-2\delta^2}$$

Resonant Peak

It is the peak (maximum) value of the magnitude of T(jω). It is denoted by M_r. At $u=u_r$, the Magnitude of T(jω) is -

$$M_r = \frac{1}{\sqrt{(1-u_r^2)^2+(2\delta u_r)^2}}$$

Substitute, $u_r = \sqrt{1-2\delta^2}$ and $1-u_r^2 = 2\delta^2$ in the above equation.

$$M_r = \frac{1}{\sqrt{(2\delta^2)^2+(2\delta\sqrt{1-2\delta^2})^2}}$$

$$\Rightarrow M_r = \frac{1}{2\delta\sqrt{1-\delta^2}}$$

Resonant peak in frequency response corresponds to the peak overshoot in the time domain transient response for certain values of damping ratio δδ. So, the resonant peak and peak overshoot are correlated to each other.

Bandwidth

It is the range of frequencies over which, the magnitude of T(jω) drops to 70.7% from its zero frequency value.

At ω=0, the value of u will be zero.

Substitute, u=0 in M.

$$M = \frac{1}{\sqrt{(1-0^2)^2+(2\delta(0))^2}} = 1$$

Therefore, the magnitude of T(jω) is one at ω=0

At 3-dB frequency, the magnitude of T(jω) will be 70.7% of magnitude of T(jω)) at ω=0 i.e., at $\omega = \omega_B, M = 0.707(1) = \frac{1}{\sqrt{2}}$

$$\Rightarrow M = \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{(1-u_b^2)^2 + (2\delta u_b)^2}}$$

$$\Rightarrow 2 = (1-u_b^2)^2 + (2\delta)^2 u_b^2$$

Let, $u_b^2 = x$

$$\Rightarrow 2 = (1-x)^2 + (2\delta)^2 x$$

$$\Rightarrow x^2 + (4\delta^2 - 2)x - 1 = 0$$

$$\Rightarrow x = \frac{-(4\delta^2-2) \pm \sqrt{(4\delta^2-2)^2 + 4}}{2}$$

Consider only the positive value of x.

$$x = 1 - 2\delta^2 + \sqrt{(2\delta^2 - 1)^2 + 1}$$

$$\Rightarrow x = 1 - 2\delta^2 + \sqrt{(2 - 4\delta^2 + 4\delta^4)}$$

Substitute, $x = u_b^2 = \frac{\omega_b^2}{\omega_n^2}$

$$\frac{\omega_b^2}{\omega_n^2} = 1 - 2\delta^2 + \sqrt{(2 - 4\delta^2 + 4\delta^4)}$$

$$\Rightarrow \omega_b = \omega_n \sqrt{1 - 2\delta^2 + \sqrt{(2 - 4\delta^2 + 4\delta^4)}}$$

Bandwidth ωb in the frequency response is inversely proportional to the rise time tr in the time domain transient response.

Bode plots

The Bode plot or the Bode diagram consists of two plots −

- Magnitude plot
- Phase plot

In both the plots, x-axis represents angular frequency (logarithmic scale). Whereas, yaxis represents the magnitude (linear scale) of open loop transfer function in the magnitude plot and the phase angle (linear scale) of the open loop transfer function in the phase plot.

The **magnitude** of the open loop transfer function in dB is -

$$M = 20\ \log|G(j\omega)H(j\omega)|$$

The **phase angle** of the open loop transfer function in degrees is -

$$\phi = \angle G(j\omega)H(j\omega)$$

Basic of Bode Plots

The following table shows the slope, magnitude and the phase angle values of the terms present in the open loop transfer function. This data is useful while drawing the Bode plots.

Type of term	G(jω)H(jω)	Slope(dB/dec)	Magnitude (dB)	Phase angle(degrees)
Constant	K	0	$20\log K$	0
Zero at origin	$j\omega$	20	$20\log\omega$	90
'n' zeros at origin	$(j\omega)^n$	$20\ n$	$20\ n\log\omega$	$90\ n$
Pole at origin	$\frac{1}{j\omega}$	-20	$-20\log\omega$	$-90\ or\ 270$
'n' poles at origin	$\frac{1}{(j\omega)^n}$	$-20\ n$	$-20\ n\log\omega$	$-90\ n\ or\ 270\ n$
Simple zero	$1+j\omega r$	20	$0\ for\ \omega < \frac{1}{r}$ $20\ \log\omega r\ for\ \omega > \frac{1}{r}$	$0\ for\ \omega < \frac{1}{r}$ $90\ for\ \omega > \frac{1}{r}$

Simple pole	$\frac{1}{1+j\omega r}$	-20	$0\ for\ \omega < \frac{1}{r}$ $-20\ \log \omega r\ for\ \omega > \frac{1}{r}$	$0\ for\ \omega < \frac{1}{r}$ $-90\ or\ 270\ for\ \omega > \frac{1}{r}$
Second order derivative term	$\omega_n^2\left(1-\frac{\omega^2}{\omega_n^2}+\frac{2j\delta\omega}{\omega_n}\right)$	40	$40\ \log\ \omega_n\ for\ \omega < \omega_n$ $20\ \log\ (2\delta\omega_n^2)\ for\ \omega = \omega_n$ $40\ \log\ \omega\ for\ \omega > \omega_n$	$0\ for\ \omega < \omega_n$ $90\ for\ \omega = \omega_n$ $180\ for\ \omega > \omega_n$
Second order integral term	$\frac{1}{\omega_n^2\left(1-\frac{\omega^2}{\omega_n^2}+\frac{2j\delta\omega}{\omega_n}\right)}$	-40	$-40\ \log\ \omega_n\ for\ \omega < \omega_n$ $-20\ \log\ (2\delta\omega_n^2)\ for\ \omega = \omega_n$ $-40\ \log\ \omega\ for\ \omega > \omega_n$	$-0\ for\ \omega < \omega_n$ $-90\ for\ \omega = \omega_n$ $-180\ for\ \omega > \omega_n$

Consider the open loop transfer function $G(s)H(s) = K$.

Magnitude $M = 20\ \log K$ dB

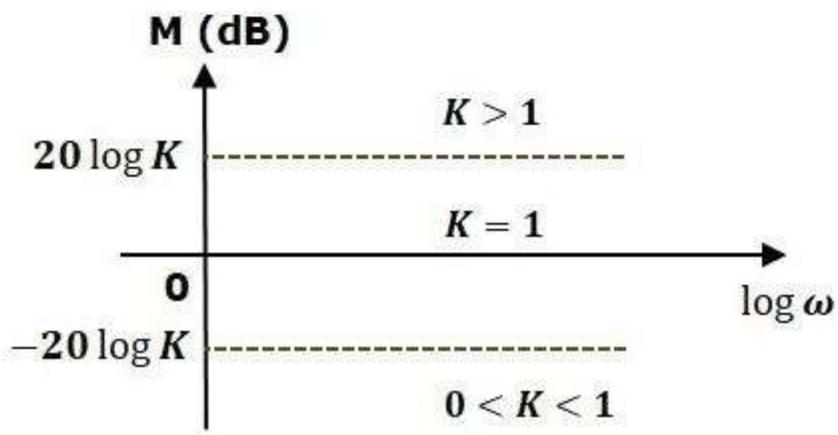
Phase angle $\phi = 0$ degrees

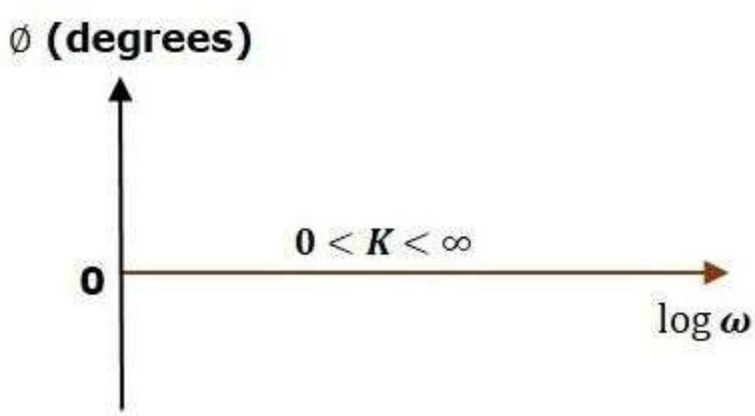
If $K = 1$, then magnitude is 0 dB.

If $K > 1$, then magnitude will be positive.

If $K < 1$, then magnitude will be negative.

The following figure shows the corresponding Bode plot.





The magnitude plot is a horizontal line, which is independent of frequency. The 0 dB line itself is the magnitude plot when the value of K is one. For the positive values of K, the horizontal line will shift 20logK dB above the 0 dB line. For the negative values of K, the horizontal line

will shift 20logK dB below the 0 dB line. The Zero degrees line itself is the phase plot for all the positive values of K.

Consider the open loop transfer function G(s)H(s)=s

Magnitude M=20logω dB

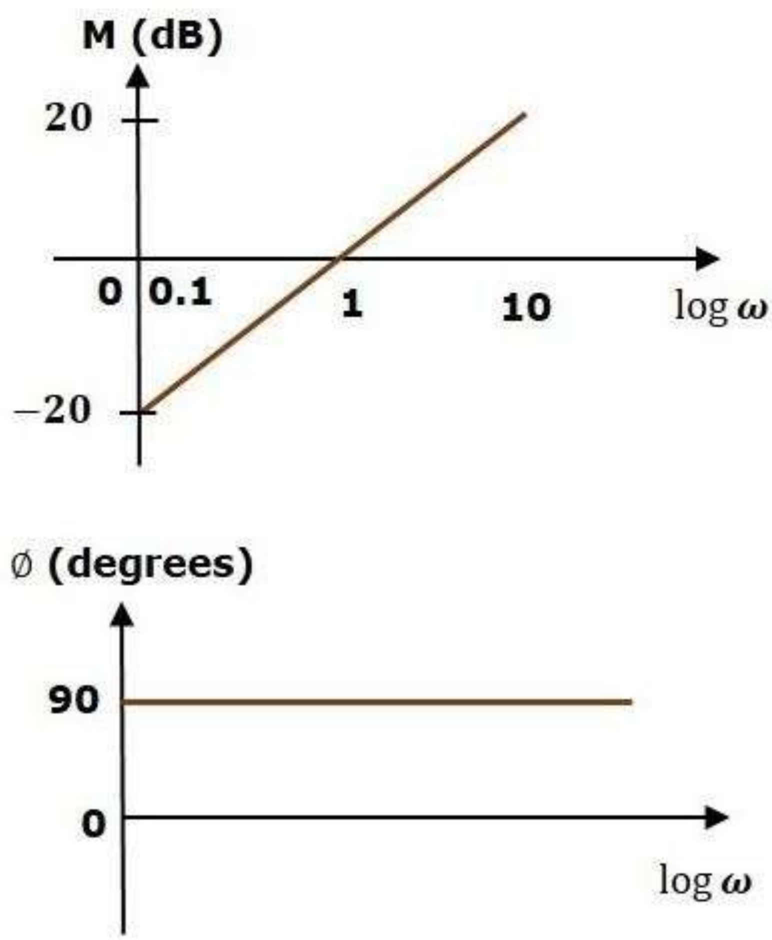
Phase angle $\phi=90^0$

At ω=0.1rad/sec, the magnitude is -20 dB.

At ω=1rad/sec, the magnitude is 0 dB.

At ω=10 rad/sec, the magnitude is 20 dB.

The following figure shows the corresponding Bode plot.



The magnitude plot is a line, which is having a slope of 20 dB/dec. This line started at ω=0.1rad/sec having a magnitude of -20 dB and it continues on the same slope. It is touching 0 dB line at ω=1 rad/sec. In this case, the phase plot is 90^0 line.

Consider the open loop transfer function

$$M = 20\ log\sqrt{1+\omega^2\tau^2}\ \text{dB}$$

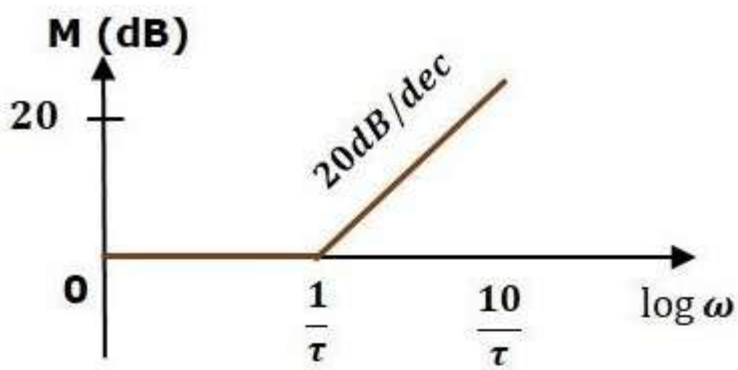
G(s)H(s)=1+sτ. Magnitude

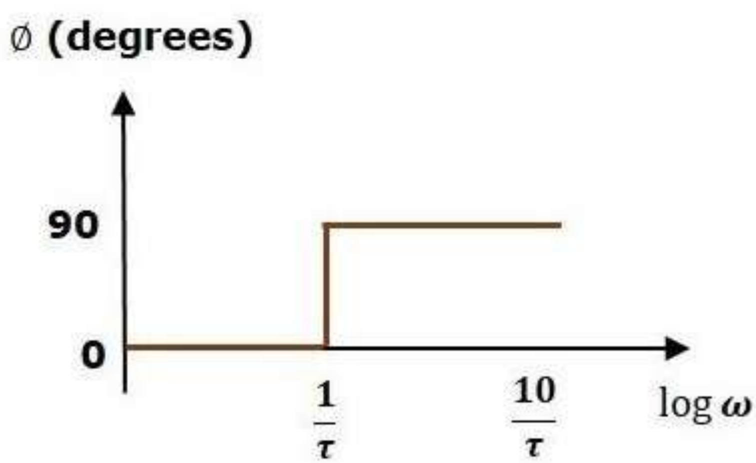
$$\phi = \tan^{-1}\omega\tau \text{ degrees}$$

Phase angle

For $\omega < \frac{1}{\tau}$, the magnitude is 0 dB and phase angle is 0 degrees.

For $\omega > \frac{1}{\tau}$, the magnitude is 20logωτ dB and phase angle is





90^0. The following figure shows the corresponding Bode plot

The magnitude plot is having magnitude of 0 dB upto ω=1τω=1τ rad/sec. From ω=1τ rad/sec, it is having a slope of 20 dB/dec. In this case, the phase plot is having phase angle of 0 degrees up to ω=1τ rad/sec and from here, it is having phase angle of 90^0. This Bode plot is called the **asymptotic Bode plot**.

As the magnitude and the phase plots are represented with straight lines, the Exact Bode plots resemble the asymptotic Bode plots. The only difference is that the Exact Bode plots will have simple curves instead of straight lines.

Similarly, you can draw the Bode plots for other terms of the open loop transfer function which are given in the table.

Rules for Construction of Bode Plots

Follow these rules while constructing a Bode plot.

- Represent the open loop transfer function in the standard time constant form.
- Substitute, s=jωs=jω in the above equation.
- Find the corner frequencies and arrange them in ascending order.
- Consider the starting frequency of the Bode plot as $1/10^{th}$ of the minimum corner frequency or 0.1 rad/sec whichever is smaller value and draw the Bode plot upto 10 times maximum corner frequency.
- Draw the magnitude plots for each term and combine these plots properly.
- Draw the phase plots for each term and combine these plots properly.

Note − The corner frequency is the frequency at which there is a change in the slope of the magnitude plot.

Example

Consider the open loop transfer function of a closed loop control syste

$$G(s)H(s) = \frac{10s}{(s+2)(s+5)}$$

Let us convert this open loop transfer function into standard time constant form.

$$G(s)H(s) = \frac{10s}{2\left(\frac{s}{2}+1\right)5\left(\frac{s}{5}+1\right)}$$

$$\Rightarrow G(s)H(s) = \frac{s}{\left(1+\frac{s}{2}\right)\left(1+\frac{s}{5}\right)}$$

So, we can draw the Bode plot in semi log sheet using the rules mentioned earlier.

Stability Analysis using Bode Plots

From the Bode plots, we can say whether the control system is stable, marginally stable or unstable based on the values of these parameters.

- Gain cross over frequency and phase cross over frequency
- Gain margin and phase margin

Phase Cross over Frequency

The frequency at which the phase plot is having the phase of -180^0 is known as **phase cross over frequency**. It is denoted by ωpc. The unit of phase cross over frequency is **rad/sec**.

Gain Cross over Frequency

The frequency at which the magnitude plot is having the magnitude of zero dB is known as **gain cross over frequency**. It is denoted by ωgc. The unit of gain cross over frequency is **rad/sec**.

The stability of the control system based on the relation between the phase cross over frequency and the gain cross over frequency is listed below.

- If the phase cross over frequency ωpc is greater than the gain cross over frequency ωgc, then the control system is **stable**.
- If the phase cross over frequency ωpc is equal to the gain cross over frequency ωgc, then the control system is **marginally stable**.
- If the phase cross over frequency ωpcis less than the gain cross over frequency ωgc, then the control system is **unstable**.

Gain Margin

Gain margin GMGM is equal to negative of the magnitude in dB at phase cross over frequency.

$$GM=20\log(1Mpc)=20\log Mpc$$

Where, MpcMpc is the magnitude at phase cross over frequency. The unit of gain margin (GM) is **dB**.

Phase Margin

The formula for phase margin PMPM is

$$PM=180^0+\phi gc$$

Where, ϕgc is the phase angle at gain cross over frequency. The unit of phase margin is **degrees**.

- NOTE:

The stability of the control system based on the relation between gain margin and phase margin is listed below.

- If both the gain margin GM and the phase margin PM are positive, then the control system is **stable**.
- If both the gain margin GM and the phase margin PM are equal to zero, then the control system is **marginally stable**.

If the gain margin GM and / or the phase margin PM are/is negative, then the control system is **unstable**.

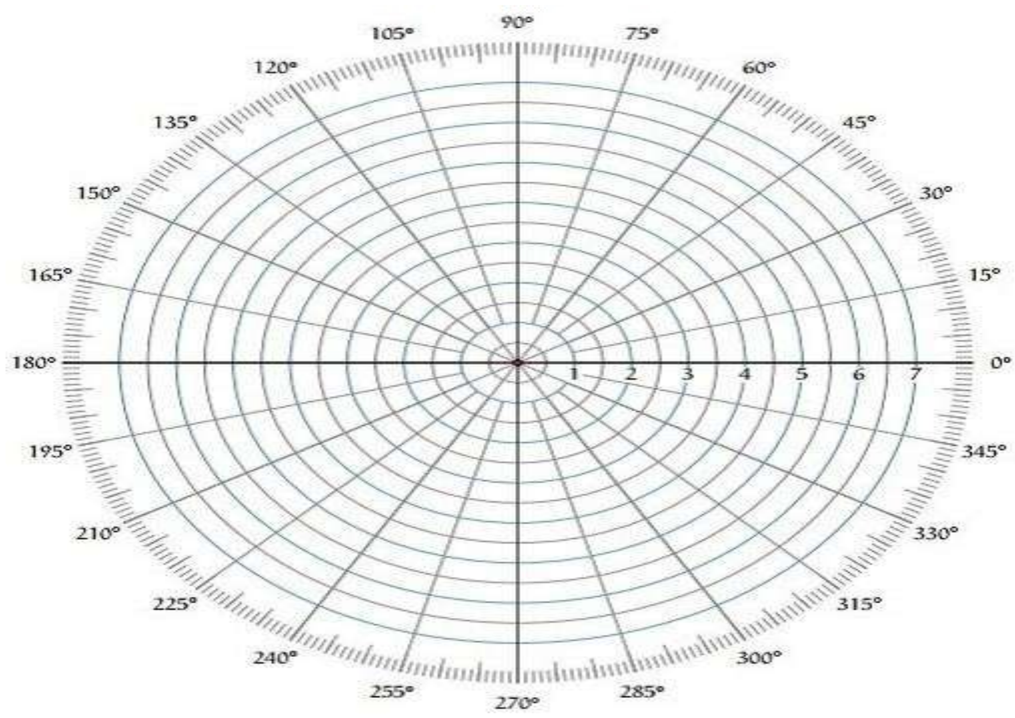
Polar plots

Polar plot is a plot which can be drawn between magnitude and phase. Here, the magnitudes are represented by normal values only.

The polar form of $G(j\omega)H(j\omega)$ is

$$G(j\omega)H(j\omega) = |G(j\omega)H(j\omega)|\angle G(j\omega)H(j\omega)$$

The **Polar plot** is a plot, which can be drawn between the magnitude and the phase angle of $G(j\omega)H(j\omega)$ by varying ω from zero to ∞. The polar graph sheet is shown in the following figure.



This graph sheet consists of concentric circles and radial lines. The **concentric circles** and the **radial lines** represent the magnitudes and phase angles respectively. These angles are represented by positive values in anti-clock wise direction. Similarly, we can represent angles with negative values in clockwise direction. For example, the angle 270^0 in anti-clock wise direction is equal to the angle -90^0 in clockwise direction.

Rules for Drawing Polar Plots

Follow these rules for plotting the polar plots.

- Substitute, s=jω in the open loop transfer function.
- Write the expressions for magnitude and the phase of G(jω)H(jω)
- Find the starting magnitude and the phase of G(jω)H(jω) by substituting ω=0. So, the polar plot starts with this magnitude and the phase angle.
- Find the ending magnitude and the phase of G(jω)H(jω) by substituting ω=∞ So, the polar plot ends with this magnitude and the phase angle.
- Check whether the polar plot intersects the real axis, by making the imaginary term of G(jω)H(jω) equal to zero and find the value(s) of ω.
- Check whether the polar plot intersects the imaginary axis, by making real term of G(jω)H(jω) equal to zero and find the value(s) of ω.
- For drawing polar plot more clearly, find the magnitude and phase of G(jω)H(jω) by considering the other value(s) of ω.

Example

Consider the open loop transfer function of a closed loop control system.

$$G(s)H(s) = \frac{5}{s(s+1)(s+2)}$$

Let us draw the polar plot for this control system using the above rules.

Step 1 − Substitute, $s = j\omega$ in the open loop transfer function.

$$G(j\omega)H(j\omega) = \frac{5}{j\omega(j\omega+1)(j\omega+2)}$$

The magnitude of the open loop transfer function is

$$M = \frac{5}{\omega(\sqrt{\omega^2+1})(\sqrt{\omega^2+4})}$$

The phase angle of the open loop transfer function is

$$\phi = -90^0 - \tan^{-1}\omega - \tan^{-1}\frac{\omega}{2}$$

Frequency (rad/sec)	Magnitude	Phase angle(degrees)
0	∞	-90 or 270
∞	0	-270 or 90

So, the polar plot starts at $(\infty, -90^0)$ and ends at $(0, -270^0)$. The first and the second terms within the brackets indicate the magnitude and phase angle respectively.

Step 3 − Based on the starting and the ending polar co-ordinates, this polar plot will intersect the negative real axis. The phase angle corresponding to the negative real axis is -180^0 or 180^0. So, by equating the phase angle of the open loop transfer function to either -180^0 or 180^0, we will get the ω value as $\sqrt{2}$.
By substituting $\omega=\sqrt{2}$ in the magnitude of the open loop transfer function, we will get M=0.83. Therefore, the polar plot intersects the negative real axis when $\omega=\sqrt{2}$ and the polar coordinate is $(0.83, -180^0)$.
So, we can draw the polar plot with the above information on the polar graph sheet.

Nyquist Plots

Nyquist plots are the continuation of polar plots for finding the stability of the closed loop control systems by varying ω from −∞ to ∞. That means, Nyquist plots are used to draw the complete frequency response of the open loop transfer function.

Nyquist Stability Criterion

The Nyquist stability criterion works on the **principle of argument**. It states that if there are P poles and Z zeros are enclosed by the 's' plane closed path, then the corresponding G(s)H(s)G(s)H(s) plane must encircle the origin P−ZP−Z times. So, we can write the number of encirclements N as,

$$N=P-ZN=P-Z$$

- If the enclosed 's' plane closed path contains only poles, then the direction of the encirclement in the G(s)H(s)G(s)H(s) plane will be opposite to the direction of the enclosed closed path in the 's' plane.

- If the enclosed 's' plane closed path contains only zeros, then the direction of the encirclement in the G(s)H(s)G(s)H(s) plane will be in the same direction as that of the enclosed closed path in the 's' plane.

Let us now apply the principle of argument to the entire right half of the 's' plane by selecting it as a closed path. This selected path is called the **Nyquist** contour.

We know that the closed loop control system is stable if all the poles of the closed loop transfer function are in the left half of the 's' plane. So, the poles of the closed loop transfer function are nothing but the roots of the characteristic equation. As the order of the characteristic equation increases, it is difficult to find the roots. So, let us correlate these roots of the characteristic equation as follows.

- The Poles of the characteristic equation are same as that of the poles of the open loop transfer function.
- The zeros of the characteristic equation are same as that of the poles of the closed loop transfer function.

We know that the open loop control system is stable if there is no open loop pole in the the right half of the 's' plane.

i.e.,$P=0 \Rightarrow N=-Z P=0 \Rightarrow N=-Z$

We know that the closed loop control system is stable if there is no closed loop pole in the right half of the 's' plane.

i.e.,$Z=0 \Rightarrow N=P Z=0 \Rightarrow N=P$

Nyquist stability criterion states the number of encirclements about the critical point (1+j0) must be equal to the poles of characteristic equation, which is nothing but the poles of the open loop transfer function in the right half of the 's' plane. The shift in origin to (1+j0) gives the characteristic equation plane.

Rules for Drawing Nyquist Plots

Follow these rules for plotting the Nyquist plots.

- Locate the poles and zeros of open loop transfer function G(s)H(s) in 's' plane.

- Draw the polar plot by varying ω from zero to infinity. If pole or zero present at s = 0, then varying ω from 0+ to infinity for drawing polar plot.
- Draw the mirror image of above polar plot for values of ω ranging from −∞ to zero (0^- if
- any pole or zero present at s=0).
- The number of infinite radius half circles will be equal to the number of poles or zeros at origin. The infinite radius half circle will start at the point where the mirror image of the polar plot ends. And this infinite radius half circle will end at the point where the polar plot starts.

After drawing the Nyquist plot, we can find the stability of the closed loop control system using the Nyquist stability criterion. If the critical point (-1+j0) lies outside the encirclement, then the closed loop control system is absolutely stable.

Stability Analysis using Nyquist Plots

From the Nyquist plots, we can identify whether the control system is stable, marginally stable or unstable based on the values of these parameters.

- Gain cross over frequency and phase cross over frequency
- Gain margin and phase margin

Phase Cross over Frequency

The frequency at which the Nyquist plot intersects the negative real axis (phase angle is 180^0) is known as the **phase cross over frequency**. It is denoted by ωpc.

Gain Cross over Frequency

The frequency at which the Nyquist plot is having the magnitude of one is known as the **gain cross over frequency**. It is denoted by ωgc.

The stability of the control system based on the relation between phase cross over frequency and gain cross over frequency is listed below.

- If the phase cross over frequency ωpc is greater than the gain cross over frequency ωgc, then the control system is **stable**.
- If the phase cross over frequency ωpc is equal to the gain cross over frequency ωgc, then the control system is **marginally stable**.

- If phase cross over frequency ωpc is less than gain cross over frequency ωgc, then the control system is **unstable**.

Gain Margin

The gain margin GM is equal to the reciprocal of the magnitude of the Nyquist plot at the phase

$$GM = \frac{1}{M_{pc}}$$

cross over frequency.

Where, Mpc is the magnitude in normal scale at the phase cross over frequency.

Phase Margin

The phase margin PM is equal to the sum of 180^0 and the phase angle at the gain cross over frequency.

$$PM=180^0+\phi gc$$

Where, ϕgc is the phase angle at the gain cross over frequency.

The stability of the control system based on the relation between the gain margin and the phase margin is listed below.

- If the gain margin GM is greater than one and the phase margin PM is positive, then the control system is **stable**.
- If the gain margin GMs equal to one and the phase margin PM is zero degrees, then the control system is **marginally stable**.
- If the gain margin GM is less than one and / or the phase margin PM is negative, then the control system is **unstable**.

UNIT-IV (PART-II)

SUB: STABILITY OF CONTROL SYSTEMS: (ME 210C); SEM-4th B.tech

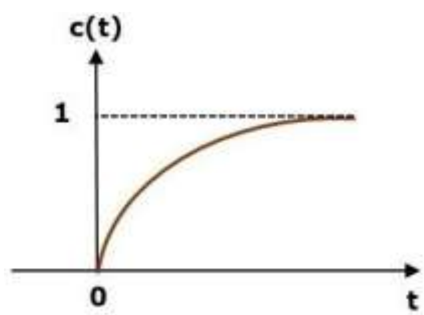
Stability is an important concept. In this chapter, let us discuss the stability of system and types of systems based on stability.

What is Stability?

A system is said to be stable, if its output is under control. Otherwise, it is said to be unstable. A **stable system** is a bounded output for a given bounded input.

The following figure shows the response of a stable system.

This is the response of first order control system for unit step input. This response has the values between 0 and 1. So, it is bounded output. We know that the unit step signal has the



value of one for all positive values of **t** including zero. So, it is bounded input. Therefore, the first order control system is stable since both the input and the output are bounded.

Types of Systems based on Stability

We can classify the systems based on stability as follows.

- Absolutely stable system
- Conditionally stable system
- Marginally stable system

Absolutely Stable System

If the system is stable for all the range of system component values, then it is known as the absolutely stable system. The open loop control system is absolutely stable if all the poles of the open loop transfer function present in left half of **'s' plane**. Similarly, the closed loop control system is absolutely stable if all the poles of the closed loop transfer function present in the left half of the 's' plane.

Conditionally Stable System

If the system is stable for a certain range of system component values, then it is known as conditionally stable system.

Marginally Stable System

If the system is stable by producing an output signal with constant amplitude and constant frequency of oscillations for bounded input, then it is known as **marginally stable system**. The open loop control system is marginally stable if any two poles of the open loop transfer function is present on the imaginary axis. Similarly, the closed loop control system is marginally stable if any two poles of the closed loop transfer function is present on the imaginary axis.

n this chapter, let us discuss the stability analysis in the **'s'** domain using the RouthHurwitz stability criterion. In this criterion, we require the characteristic equation to find the stability of the closed loop control systems.

Routh-Hurwitz Stability Criterion

Routh-Hurwitz stability criterion is having one necessary condition and one sufficient condition for stability. If any control system doesn't satisfy the necessary condition, then we can say that the control system is unstable. But, if the control system satisfies the necessary condition, then it may or may not be stable. So, the sufficient condition is helpful for knowing whether the control system is stable or not.

Necessary Condition for Routh-Hurwitz Stability

The necessary condition is that the coefficients of the characteristic polynomial should be positive. This implies that all the roots of the characteristic equation should have negative real parts.

Consider the characteristic equation of the order 'n' is -

$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \ldots + a_{n-1} s^1 + a_n s^0 = 0$$

Note that, there should not be any term missing in the **nth** order characteristic equation. This means that the **nth** order characteristic equation should not have any coefficient that is of zero value.

Sufficient Condition for Routh-Hurwitz Stability

The sufficient condition is that all the elements of the first column of the Routh array should have the same sign. This means that all the elements of the first column of the Routh array should be either positive or negative.

Routh Array Method

If all the roots of the characteristic equation exist to the left half of the 's' plane, then the control system is stable. If at least one root of the characteristic equation exists to the right half of the 's' plane, then the control system is unstable. So, we have to find the roots of the characteristic equation to know whether the control system is stable or unstable. But, it is difficult to find the roots of the characteristic equation as order increases.

So, to overcome this problem there we have the **Routh array method**. In this method, there is no need to calculate the roots of the characteristic equation. First formulate the Routh table and find the number of the sign changes in the first column of the Routh table. The number of sign changes in the first column of the Routh table gives the number of roots of characteristic equation that exist in the right half of the 's' plane and the control system is unstable.

Follow this procedure for forming the Routh table.

- Fill the first two rows of the Routh array with the coefficients of the characteristic polynomial as mentioned in the table below. Start with the coefficient of sn and continue up to the coefficient of s0.
- Fill the remaining rows of the Routh array with the elements as mentioned in the table below. Continue this process till you get the first column element of **row** s0s0 is an. Here, an is the coefficient of s0 in the characteristic polynomial.

Note − If any row elements of the Routh table have some common factor, then you can divide the row elements with that factor for the simplification will be easy.

The following table shows the Routh array of the n^{th} order characteristic polynomial.

$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \ldots + a_{n-1} s^1 + a_n s^0$$

s^n	a_0	a_2	a_4	a_6	$\ldots$	$\ldots$
s^{n-1}	a_1	a_3	a_5	a_7	$\ldots$	$\ldots$
s^{n-2}	$b_1 = \frac{a_1 a_2 - a_3 a_0}{a_1}$	$b_2 = \frac{a_1 a_4 - a_5 a_0}{a_1}$	$b_3 = \frac{a_1 a_6 - a_7 a_0}{a_1}$	$\ldots$	$\ldots$	$\ldots$
s^{n-3}	$c_1 = \frac{b_1 a_3 - b_2 a_1}{b_1}$	$c_2 = \frac{b_1 a_5 5 - b_3 a_1}{b_1}$	$\vdots$			
$\vdots$	$\vdots$	$\vdots$	$\vdots$			
s^1	$\vdots$	$\vdots$				
s^0	a_n					

Example

Let us find the stability of the control system having characteristic equation,

$$s^4 + 3s^3 + 3s^2 + 2s + 1 = 0$$

Step 1 − Verify the necessary condition for the Routh-Hurwitz stability.

$$s^4 + 3s^3 + 3s^2 + 2s + 1$$

All the coefficients of the characteristic polynomial,

are positive. So, the control system satisfies the necessary condition.

Step 2 − Form the Routh array for the given characteristic polynomial.

s^4	1	3	1
s^3	3	2	
s^2	$\frac{(3\times 3)-(2\times 1)}{3} = \frac{7}{3}$	$\frac{(3\times 1)-(0\times 1)}{3} = \frac{3}{3}$ $= 1$	
s^1	$\frac{\left(\frac{7}{3}\times 2\right)-(1\times 3)}{\frac{7}{3}}$ $= \frac{5}{7}$		
s^0	1		

Step 3 − Verify the sufficient condition for the Routh-Hurwitz stability.

All the elements of the first column of the Routh array are positive. There is no sign change in the first column of the Routh array. So, the control system is stable.

Special Cases of Routh Array

We may come across two types of situations, while forming the Routh table. It is difficult to complete the Routh table from these two situations.

The two special cases are −

- The first element of any row of the Routh's array is zero.
- All the elements of any row of the Routh's array are zero.

Let us now discuss how to overcome the difficulty in these two cases, one by one.

First Element of any row of the Routh's array is zero

If any row of the Routh's array contains only the first element as zero and at least one of the remaining elements have non-zero value, then replace the first element with a small positive integer, ϵ. And then continue the process of completing the Routh's table. Now, find the number of sign changes in the first column of the Routh's table by substituting $\epsilon\epsilon$ tends to zero.

Example

Let us find the stability of the control system having characteristic equation,

$$s^4 + 2s^3 + s^2 + 2s + 1 = 0$$

Step 1 − Verify the necessary condition for the Routh-Hurwitz stability.
All the coefficients of the characteristic polynomial,

$$s^4 + 2s^3 + s^2 + 2s + 1 = 0$$

are positive. So, the control system satisfied the necessary condition.

Step 2 − Form the Routh array for the given characteristic polynomial.

s^4	1	1	1
s^3	$\cancel{2}\ 1$	$\cancel{2}\ 1$	
s^2	$\frac{(1\times1)-(1\times1)}{1} = 0$	$\frac{(1\times1)-(0\times1)}{1} = 1$	
s^1			
s^0			

The row s^3 elements have 2 as the common factor. So, all these elements are divided by 2. **Special case (i)** − Only the first element of row s^2 is zero. So, replace it by ϵ and continue the process of completing the Routh table.

s^4	1	1	1
s^3	1	1	
s^2	ϵ	1	
s^1	$\frac{(\epsilon \times 1)-(1\times 1)}{\epsilon} = \frac{\epsilon - 1}{\epsilon}$		
s^0	1		

Step 3 − Verify the sufficient condition for the Routh-Hurwitz stability.

As ϵ tends to zero, the Routh table becomes like this.

s^4	1	1	1
s^3	1	1	
s^2	0	1	
s^1	-∞		
s^0	1		

There are two sign changes in the first column of Routh table. Hence, the control system is unstable.

All the Elements of any row of the Routh's array are zero

In this case, follow these two steps −

- Write the auxilary equation, A(s) of the row, which is just above the row of zeros.
- Differentiate the auxiliary equation, A(s) with respect to s. Fill the row of zeros with these coefficients.

Example

Let us find the stability of the control system having characteristic equation,

$$s^5 + 3s^4 + s^3 + 3s^2 + s + 3 = 0$$

Step 1 − Verify the necessary condition for the Routh-Hurwitz stability.

All the coefficients of the given characteristic polynomial are positive. So, the control system satisfied the necessary condition.

Step 2 − Form the Routh array for the given characteristic polynomial.

s^5	1	1	1
s^4	~~3~~ 1	~~3~~ 1	~~3~~ 1
s^3	$\frac{(1 \times 1)-(1 \times 1)}{1}=0$	$\frac{(1 \times 1)-(1 \times 1)}{1}=0$	
s^2			
s^1			
s^0			

The row s^4 elements have the common factor of 3. So, all these elements are divided by 3.

Special case (ii) − All the elements of row s^3 are zero. So, write the auxiliary equation, A(s) of the row s^4.

$$A(s) = s^4 + s^2 + 1$$