INTRODUCTION TO PROCESS CONTROL

In process control, the basic objective is to regulate the value of some quantity. To regulate means to maintain the quantity at some desired value called reference value or set point regardless of external influences.

EXAMPLE FOR PROCESS CONTROL

PROCESS

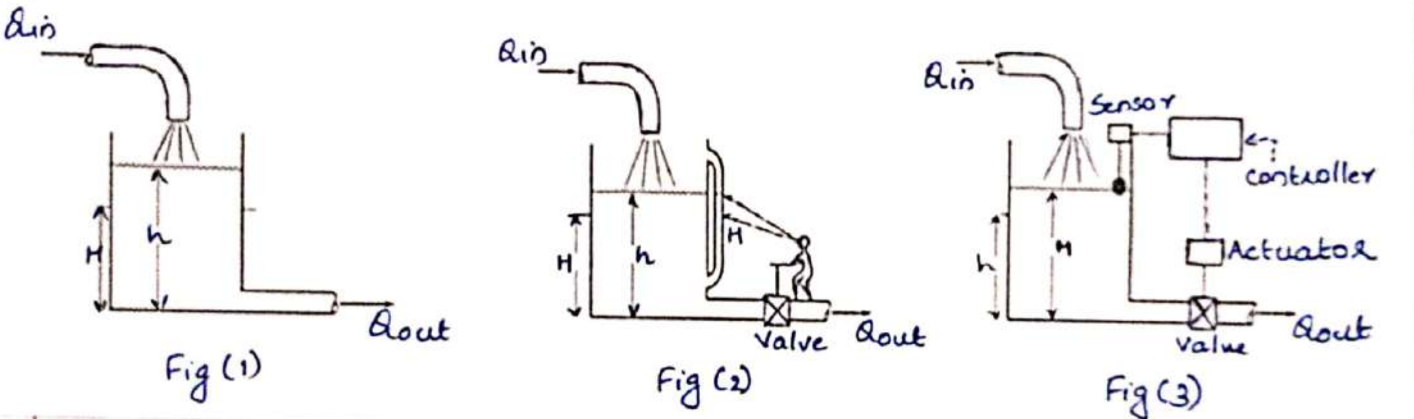


Fig (1) shows the process. Liquid is flowing into the tank at some rate Qin and out of the tank at a rate Qout. The liquid in the tank has some height or level (h). If Qout > Qin, then level 'h' will drop and if Qin > Qout, then level 'h' will rise. Suppose we want to maintain the level at some particular value 'H' regardless of input flow rate. This can be achieved either by human aided control or automatic control.

EEE, VJCET

HUMAN AIDED CONTROL

In this, a sight tube is used to measure the level as shown in fig (2). The actual liquid level (h) is called the controlled variable. A valve has been added so that the output flow rate can be changed by the human. The output flow rate is called the manipulated variable or controlling variable.

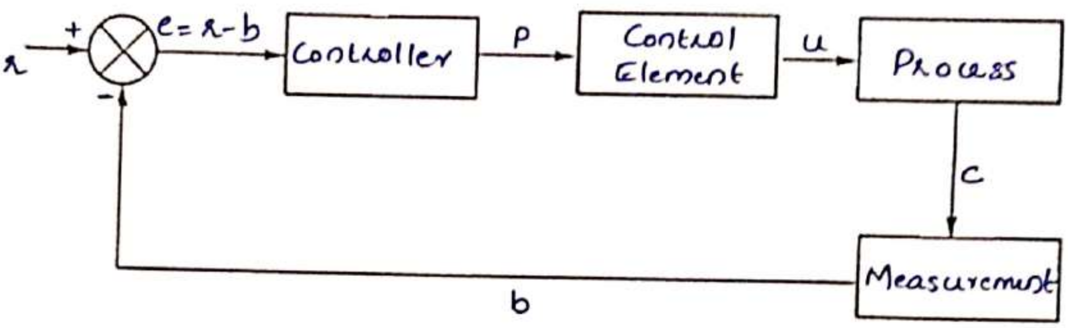
The human measures the height in the sight tube and compares the value with the setpoint. If the measured value is larger, the human opens the valve a little to let the flow out increase, and thus the level lowers towards the setpoint. If the measured value is less than the setpoint, human closes the valve to decrease the flow out and thus the level rises towards the setpoint. By a succession of incremental opening and closing of the value, the human can bring the level to the setpoint value H and maintain it thereby continuous monitoring of the sight tube and adjustment of the valve.

AUTOMATIC CONTROL

An instrument called sensor is added as in fig(3) to measure the value of the level and convert it to a proportional signal 's'. This signal is given as input to a machine, electronic circuit or computer called the controller. The controller performs the function of human is evaluating the measurement and providing the output signal to change the value setting via an actuator connected to the value by a mechanical linkage.

When automatic control is applied to systems as in fig.(3), which are designed to regulate the value of some variable to a setpoint, it is called process control.

PROCESS CONTROL BLOCK DIAGRAM



Fig(4)

i) PROCESS

In the previous example, flow of liquid in and out of the tank, tank itself and the liquid all constitute a process to be placed under control with respect to the fluid level. There are single variable processes, in which only one variable is to be controlled, as well as multivariable processes, in which many variables may require control. The process is also called the plant.

ii) MEASUREMENT

To control a variable in a process, we must have information about the variable itself. Such information is found by measuring the variable. In fig (2), sight tube is used to measure the level. In fig(3), sensor is the device that performs initial measurement.

iii) ERROR DETECTOR :

In fig(2), the human looked at the difference between the actual level 'h' and the setpoint 'H' and deduced an error. For the automatic control system in fig(3), this same kind of error determination must be done before any control action can be taken by the ~~transducer~~. controller. The error detector is often a physical part of the controller.

iv) CONTROLLER :

Controller is used to examine the error and determine what action should be taken. Controller is also called as compensator. The evaluation may be performed by an operator as in fig(2) or by a computer as in fig(3). The controller requires input of both a measured indication of the controlled variable and a representation of the reference value of the variable. Evaluation consists of determining the action required to drive the controlled variable to the setpoint value.

v) CONTROL ELEMENT :

The final element is a device that exerts direct influence on the process, that is it provides the required changes in the controlled variable to bring it to the setpoint. In the previous example the control element is the valve that controls the outflow of fluid in the tank. Often an intermediate operation referred to as an actuator ~~because~~ is required between the controller and control element. The actuator translates the small energy signal of the controller into a larger energy action on the processes.

EEE, VJCET

SENSOR TIME RESPONSE

The static transfer function of a process control loop element specifies how the output is related to the input if the input is constant. Dynamic transfer function specifies how the output changes in time when the input is changing in time. This dynamic transfer function is simply called as the time response. Time response is important for a sensor because they are the primary element providing knowledge of the controlled variable.

fig (1)

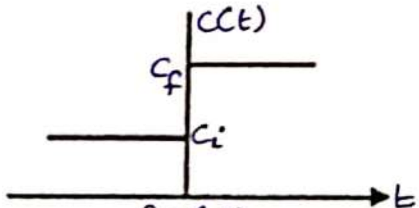
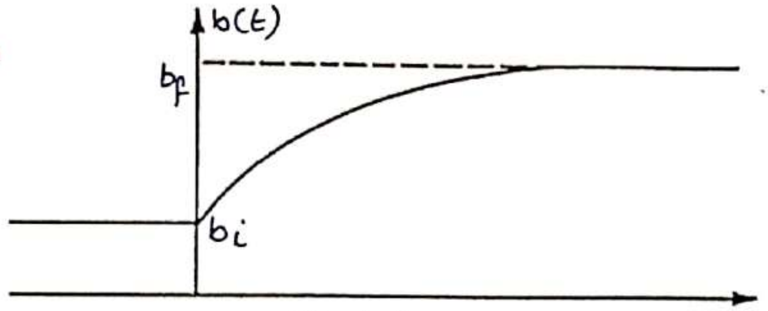


fig (2)

Fig (1) shows a sensor that produces an output b(t) as a function of the input c(t). To specify the time response, consider a ~~tonis~~ step input as shown in fig (2). At time t = 0, the input to the sensor is suddenly changed from an initial value c_i to a final value c_f. If the sensor is perfect. o/p can be determined by static transfer function. (i.e. b_i before t = 0 and b_f after t = 0). However all sensors exhibit some lag between the output and input and before settling on the final value. Sensor time response can be either first order response or second order response.

FIRST ORDER RESPONSE



fig(3)

The simplest time response for a step i/p is shown in fig(3). This is called first order response because, for all sensors of this type, the time response is determined by the solution of first order differential equation.

A general equation can be written for this response independent of the sensor, variable being measured or static transfer function. Equation for the sensor output shown in fig(3) is

$$b(t) = b(i) + (b_f - b_i)\left[1 - e^{-t/\tau}\right] \quad \text{——— (1)}$$

where,

b_i - initial sensor o/p from static transfer function and initial input

b_f - final sensor o/p from static transfer function and final i/p.

τ - sensor time constant.

The sensor o/p is in error during the transition time of o/p from b_i to b_f. Eqn.(1) represents the transducer o/p very well except during the initial time period i.e. at the start of response near $t=0$.

EEE, VJCET

Significance of time constant (τ)

Set $t = \tau$ in eqn (1)

$$b(\tau) - b_i = (b_f - b_i)\left[1 - e^{-1}\right]$$

$$\Rightarrow b(\tau) - b_i = 0.6321(b_f - b_i)$$

Thus one time constant represents the time at which the output value has changed by approximately 63% of the total change.

The time constant τ is referred to as the 63% response time For a step change, the output response has approximately reached its final value after five time constants as given below

$$b(5\tau) - b_i = 0.993(b_f - b_i)$$

SECOND ORDER RESPONSE

In some sensors, a step change in the input causes output to oscillate for a short period of time before settling down to a value that corresponds to the new input. This output transient generated by the transducer is an error. This is called a second order response because the time behaviour is described by a second order differential equation. Fig(4) shows the output curve from a transducer having a second order response. The general behaviour can be described as

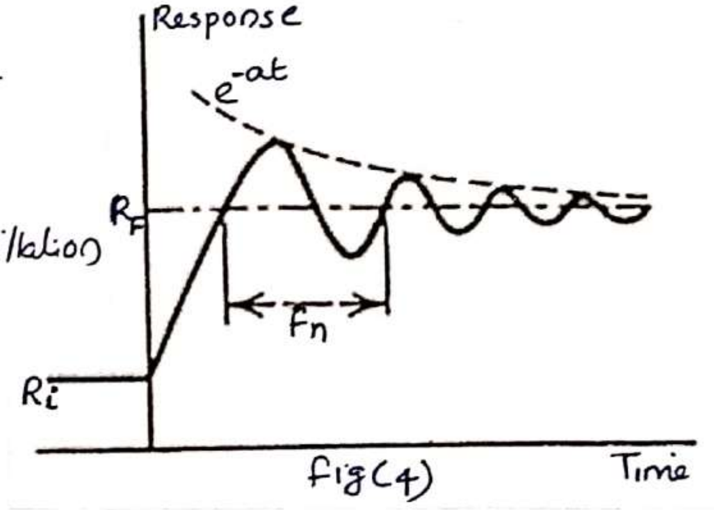
$$R(t) \propto R_0 e^{-at} \sin(2\pi f_n t)$$

where $R(t)$ - transducer output

a - o/p damping constant

f_n - natural frequency of oscillation

R_0 - amplitude.



Fig(4)

EEE, VICET

Significance of damping constant.

The damping constant defines the time one must wait after a disturbance at time $t = 0$ for the transducer o/p to be a true indication of the transducer i/p. Thus in a time of $(1/a)$, the amplitude of the oscillations would be down to e^{-1}, approximately 37%.

TRANSDUCERS

An electronic instrumentation system consists of a number of components to perform a measurement and record its results. A measurement system consists of 3 major components:

a) an input device

b) signal conditioning / processing device

c) an output device.

The input device receives the measurand or quantity under measurement and delivers a proportional or analougous electrical signal to the signal conditioning device. Here the signal is amplified, attenuated, filtered, modulated or modified in the format acceptable to the o/p device.

The input quantity for most instrumentation ~~sfm~~ system is a non electrical quantity. Inorder to use electrical methods, the non electrical quantity is converted into an electrical form by a device called transducer.

CLASSIFICATION OF TRANSDUCERS

The transducers can be classified

(i) on the basis of transduction form used

(ii) as primary and secondary transducers.

(iii) as passive and active transducers

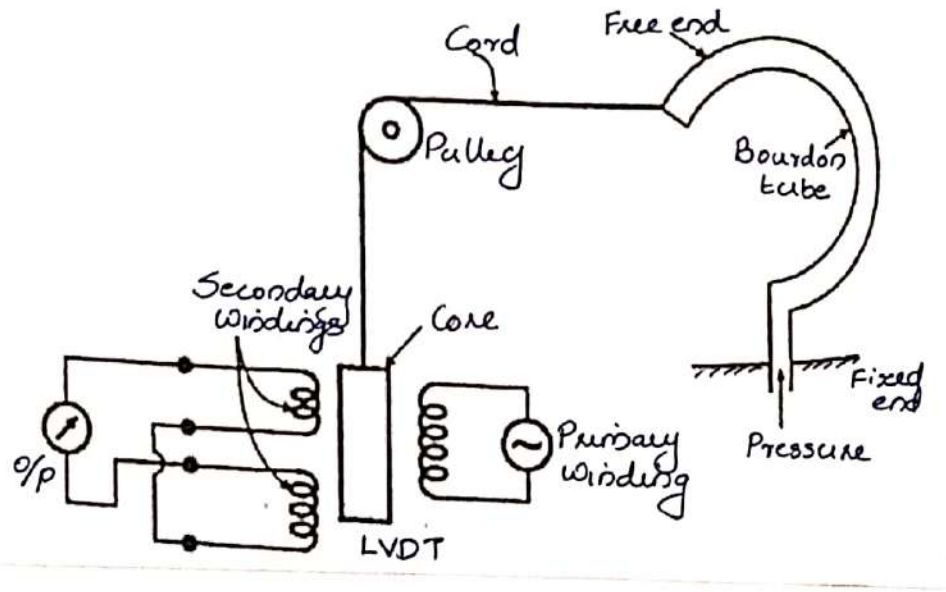
(iv) as analog and digital transducers

(v) as transducer and inverse transducers.

(i) Based upon principle of transduction

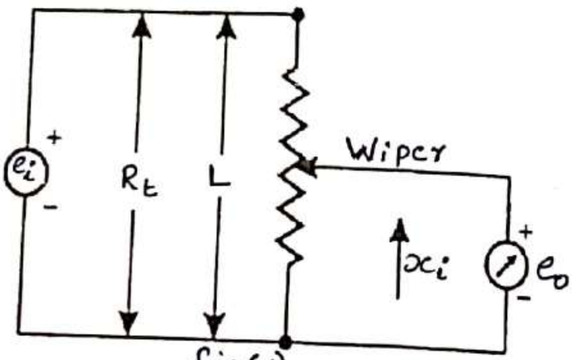
The transducers can be classified on the basis of principle of transduction as resistive, inductive, capacitive etc. depending upon how they convert the input quantity into resistance, inductance or capacitance respectively.

(ii) Primary and Secondary Transducers

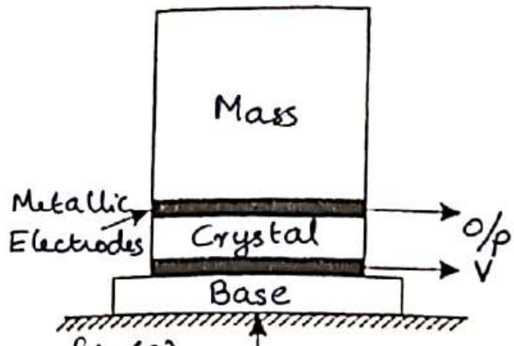


The Bourdon tube senses the pressure and converts the pressure into displacement of its free end. The displacement of the free end moves the core of the linear variable differential transformer (LVDT) which produces an output voltage which is proportional to the movement of core, which is proportional to the displacement of the free end which in turn is proportional to the pressure. Thus, there are two stages of transduction, firstly the pressure is converted into a displacement by Bourdon tube then the displacement is converted into an analogous voltage by LVDT. The Bourdon tube is called a primary transducer while the LVDT is called a secondary transducer.

(iii) Passive and active transducers.



fig(1)



fig(2)

Passive transducers: Passive transducers derive the power required for transduction from an auxillary power source. They are known as externally powered transducers. eg: Resistive, inductive & capacitive transducers. Typical example for a passive transducer is a POT which is used for measurement of displacement.

A 'POT' is a resistive transducer powered by a source voltage e_i as shown in fig (1)

o/p voltage $e_o = \frac{x_i}{L} \cdot e_i \Rightarrow x_i = \left(\frac{e_o}{e_i}\right) L$

where, L - length of POT, x_i - i/p displacement.

Active transducers: Active transducers are those which donot require an auxillary power source to produce their o/p. They are also known as self generating type since they develop their own voltage or current o/p. Consider a piezoelectric crystal shown in fig(2). The crystal is sandwiched b/w 2 electrodes A fixed mass is placed on the top of the sandwich.

The property of the piezoelectric crystal is that when a force is applied to them, they produce an o/p voltage.

(iv) Analog and Digital Transducers

Analog Transducers: These transducers convert input quantity into an analog output which is a continuous function of time. eg: Strain gauge, LVDT, thermocouple.

Digital Transducers: These transducers convert input quantity into an ~~analog~~ electrical o/p which is in the form of pulses.

(v) Transducers and Inverse Transducers

Transducers: Device which converts a non electrical quantity into electrical quantity. (fig 1)

Inverse Transducers: Device which converts electrical quantity into non electrical quantity. (fig 2) eg:- piezoelectric crystal

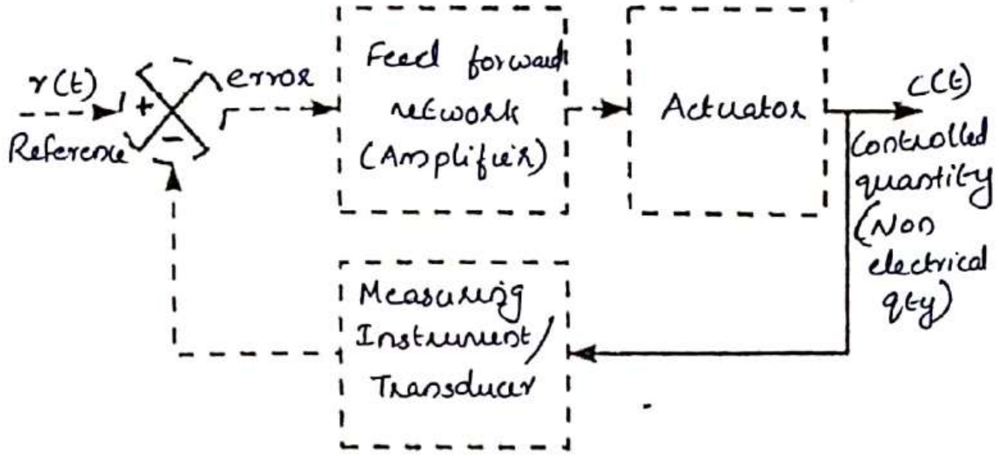


fig (1)

- - - - Electrical signal

——— Mechanical signal

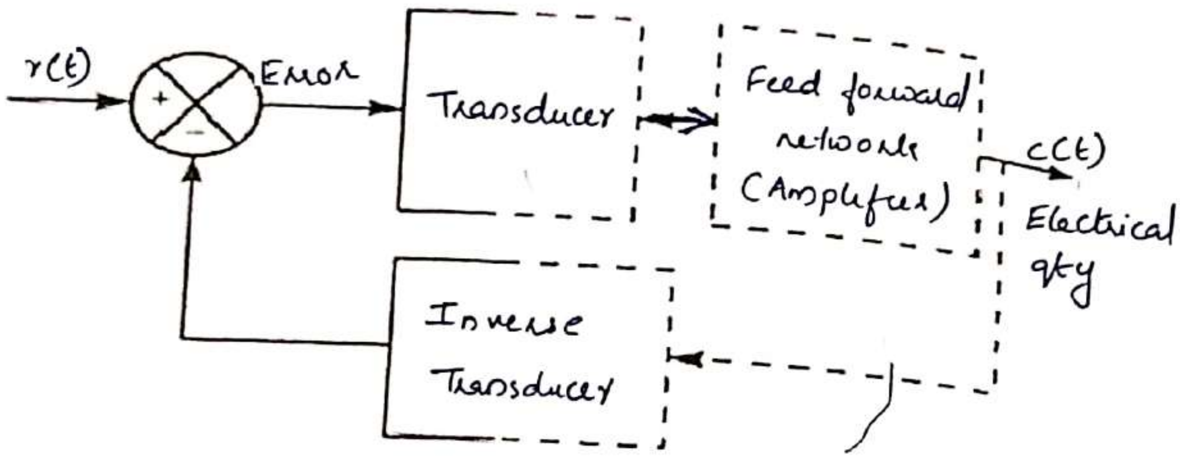


fig (2)

- - - -

———

CHARACTERISTICS & CHOICE OF TRANSDUCERS

When choosing a transducer for any application the input, transfer and output characteristics have been taken into account.

INPUT CHARACTERISTICS

i) Type of input and operating range

The foremost consideration for the choice of a transducer is the input quantity it is going to measure and its operating range. The upper limit of operating range is decided by the transducer capabilities and the

EEE, VJCET

lower limit is normally determined by the transducer error or by the unavoidable noise originating in the transducer.

ii) Loading effect :

Ideally a transducer should have no loading effect. But in practice, steps may be taken to reduce the loading effects The magnitude of the loading effects can be expressed in terms of force, power or energy extracted from the quantity under measurement for working of the transducers ∴ the transducer that is selected for a particular application should ideally extract no force, power or energy from the quantity under measurement.

TRANSFER CHARACTERISTICS

Transfer characteristics includes transfer function, error and response of a transducer to environmental influences.

i) Transfer Function:

Transfer function defines the relationship between the input quantity and output quantity.

The transfer function is $q_o = f(q_i)$ where, q_o and q_i are respectively output & input of the transducer.

Sensitivity $S = \frac{dq_o}{dq_i}$

In some cases, the relationship between output q_o and input q_i is linear. In that cases the sensitivity remains constant over the entire range. i.e. $S' = \frac{dq_o}{dq_i} = \frac{q_o}{q_i}$

ii) Error

The errors in transducers occur because in many situations the i/p - o/p relationship given by $q_o = f(q_i)$ If q_o' is the o/p instead of q_o then error of the instrument is $\epsilon = q_o' - q_o$ Errors can be either scale error, dynamic error and error on account of noise and drift. and error due to change of frequency.

a) Scale error

- Zero Error : In this case the o/p deviates from the correct value by a constant factor over the entire range of transducer. fig (1)
- Sensitivity Error : Sensitivity error occurs where the observed o/p deviates from the correct value by a constant value. If the correct o/p is q_o, observed o/p is kq_o over the entire range where k is a constant. fig (2)

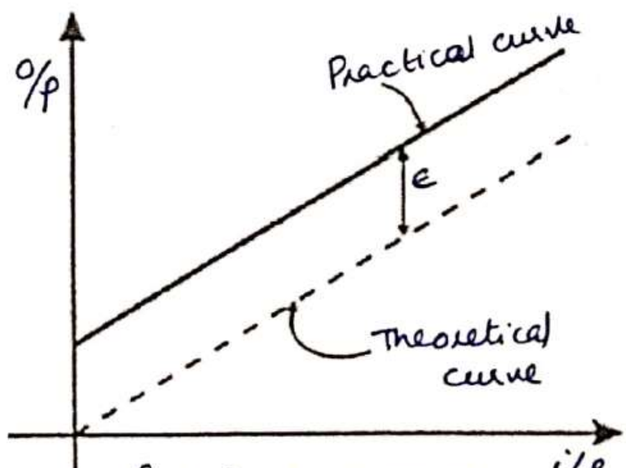


fig (1) : zero error

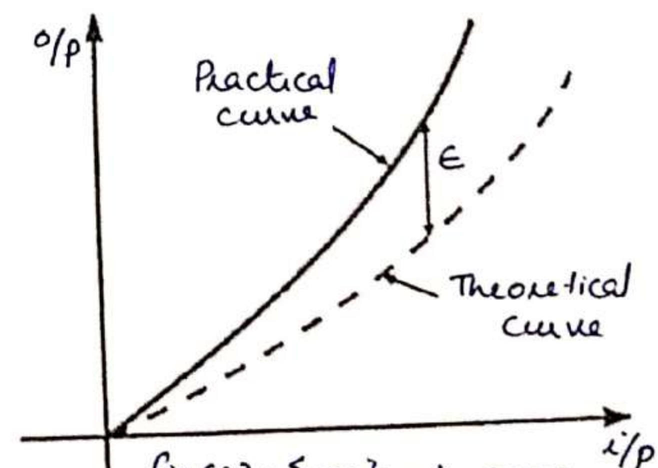
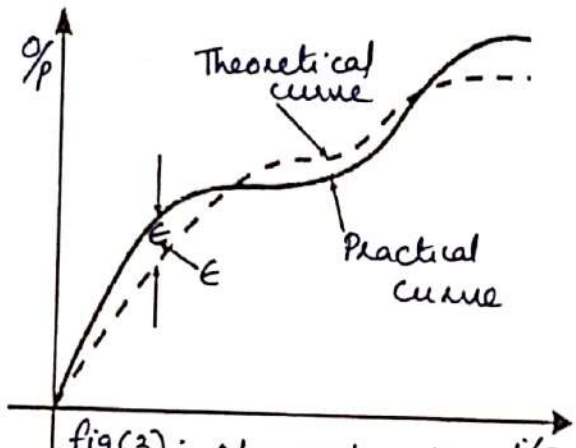


fig (2) : Sensitivity error

- Non conformity : This is a case in which the experimentally obtained transfer function deviates from theoretical transfer function for almost every input. This error is called non linearity or non linear distortion. fig(3)
- Hysterisis : The o/p of a transducer not only depends on upon the i/p quantity but also on the i/p previously applied. ∴ a different o/p is obtained when the same value of i/p quantity is applied depending upon whether it is increasing/decreasing fig(4)



fig(3): Non conformity

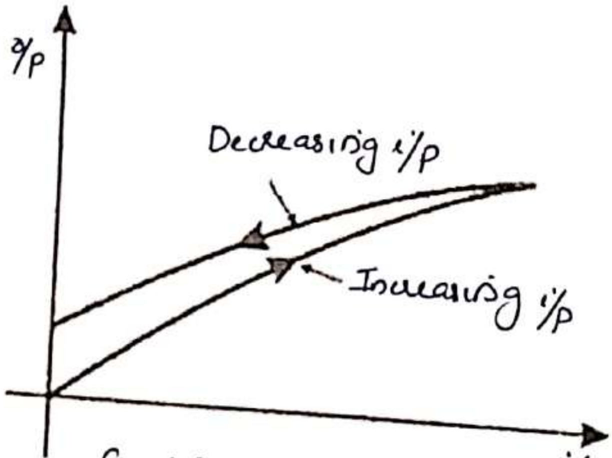


fig (4): Hysterisis

b) Dynamic error

Dynamic errors occurs only when i/p varies with time.
Consider an RC n/w to which a step of magnitude E is applied, the voltage across the capacitor after a time t $e_c = E[1 - e^{-t/\tau}]$

$\tau = RC$ - time constant

The response is shown in fig(5)

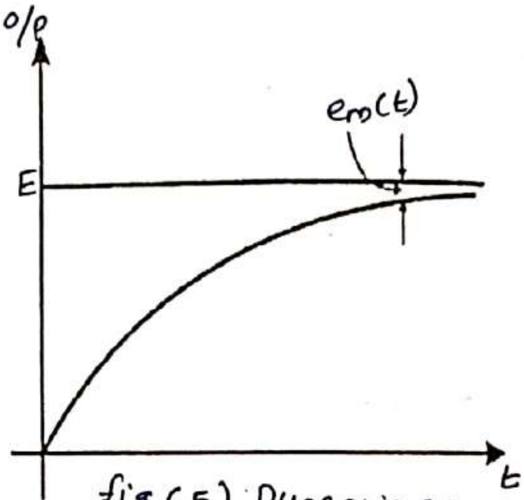


fig (5): Dynamic error

$$e_m(t) = E - e_c = e^{-t/\tau}$$

where, $e_m(t)$ - dynamic error / measurement error.

c) Errors due to noise and drift

Noise and drift signals originating from the transducers vary with time and are superimposed on the o/p signal. The difference b/w noise and drift is that noise consists of signal of random amplitude and random frequency while drift is a slow change with time.

d) Errors due to change in frequency

Frequency response and high frequency cutoff are the 2 specifications that describe the response of a transducer to a variable frequency sine wave i/p applied to it. For a linear transducer, the sinewave i/p yeilds sinewave o/p. As the frequency of sine wave is increased, the transducer is required to respond more and more quickly. Beyond a particular frequency, the transducer can no larger respond as rapidly as its sinusoidal i/p is changing. So the o/p of the transducer becomes smaller and also the phase shift between the i/p and o/p increases. Thus as the frequency increases, o/p falls. This rolloff of o/p with increase in i/p frequency is the frequency response. The frequency response is shown in fig.(6).

Gain in db = $20 \log \left(\frac{q_o}{q_i} \right)$

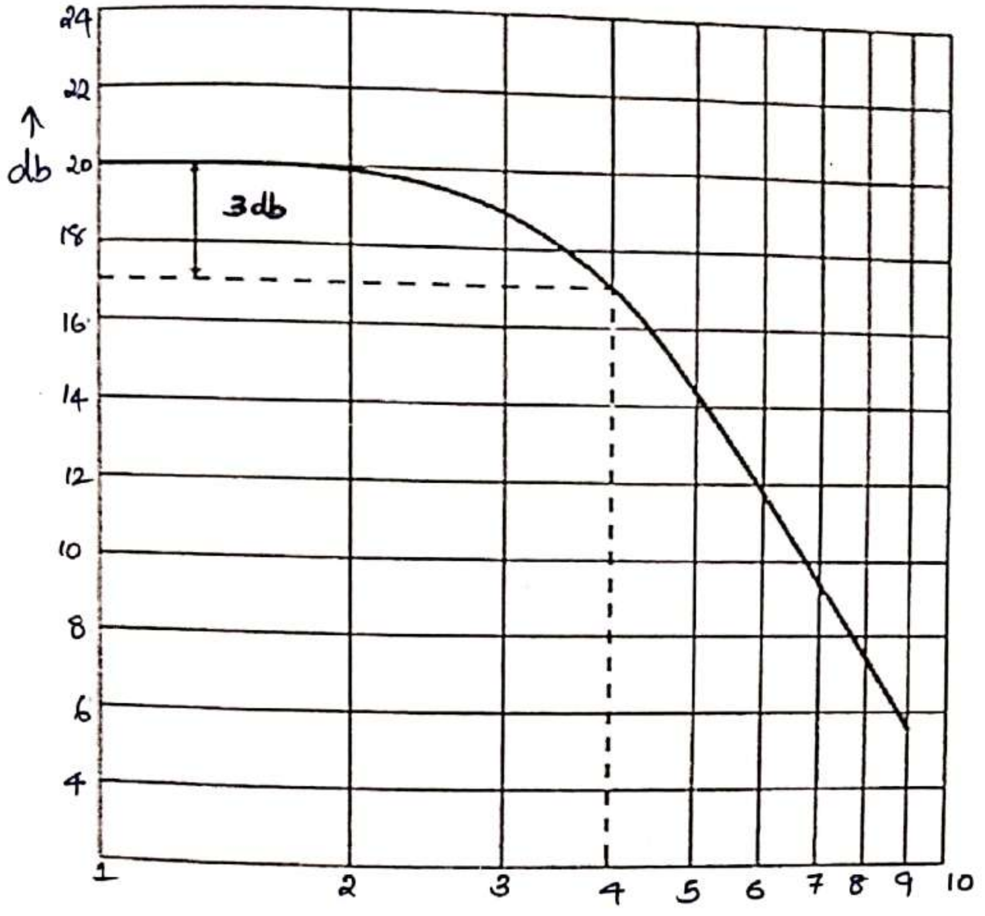


fig (6) : Frequency Response Plot

iii) Transducer Response

The response of a transducer to environmental influences is of great importance. This is often given insufficient attention when choosing the best transducer for a particular application. This give rise to unexpected results. The performance of a transducer is fully defined by its transfer function and errors, provided that the transducer is in constant environment.

OUTPUT CHARACTERISTICS

It includes type of electrical o/p, o/p impedance and useful range.

i) Type of electrical output

The type of o/p may be voltage, current, impedance or a time function of these amplitudes.

ii) Output Impedance:

The o/p impedance determines to the extent the subsequent ~~changes~~ stages of instrumentation is loaded. Its value should be kept as low as possible to minimize the loading effects.

iii) Useful output range:

Output range is limited at the lower end by noise signals. The upper limit is set by the maximum useful input level.

FACTORS INFLUENCING CHOICE OF TRANSDUCER

1) Operating Principle: The operating principles may be resistive, inductive, capacitive etc.

2) Sensitivity: The transducer must be sensitive enough to produce detectable o/p.

3) Operating Range: The rating of the transducer should be sufficient so that it does not breakdown ~~&~~ while operating in its specified range.

4) Accuracy: High degree of accuracy is assured if the transducer doesn't require frequent calibration and has a small value of repeatability.

5) Cross Sensitivity: This must be taken into account when measuring mechanical quantities. These are situations

where actual quantity is being measured is in one plane and the transducer is subjected to variations in another plane.

6) Errors: The transducers should maintain the expected input-output relationship.

7) Transcient and Frequency Response: It should ideally have a flat frequency response.

8) Loading Effects: The transducer should have a high input impedance and low output impedance.

9) Environment Compatability: It should be assured that the transducer selected to work under specified environmental conditions maintains its input-output relationship and doesn't breakdown.

10) Insensitivity to unwanted Signals: The transducer should be sensitive to unwanted signals and highly sensitive to desired signals.

11) Usage and Ruggedness: The ruggedness both of mechanical & electrical intensities of transducer versus its size and weight must be considered while selecting a suitable transducer.

12) Electrical aspects: The electrical aspects that need consideration while selecting a transducer include the length and type of cable required.

13) Stability and Reliability: The transducer should exhibit a high degree of stability to be operative during its operation and storage life.

Reliability should be assured in case of failure of transducer in order that the functioning of the instrumentation system continues uninterrupted.

14) Static Characteristics : Apart from low static error, the transducers should have a low non-linearity, low hysteresis, high resolution and a high degree of repeatability.

Signal Conditioning

26.1 SIGNAL CONDITIONING

The measurand, which is basically a physical quantity as is detected by the first stage of the instrumentation or measurement system. The first stage, with which we have become familiar, is the *"Detector Transducer stage"*. The quantity is detected and is transduced into an electrical form in most of the cases. The output of the first stage has to be modified before it becomes usable and satisfactory to drive the signal presentation stage which is the third and the last stage of a measurement system. The last stage of the measurement system may consist of indicating, recording, displaying, data processing elements or may consist of control elements.

In this chapter, methods used for modifying the transduced signal into a usable format for the final stage of the measurement system are described.

Measurement of dynamic physical quantities requires faithful representation of their analog or digital output obtained from the intermediate stage *i.e., signal conditioning stage* and this places a severe strain on the signal conditioning equipment. The signal conditioning equipment may be required to do linear processes like amplification, attenuation integration, differentiation, addition and subtraction. They are also required to do non-linear processes like modulation, demodulation, sampling, filtering, clipping and clamping, squaring, linearizing or multiplication by another function etc. These tasks are by no means simple. They require ingenuity in proper selection of components and the selection of most faithful methods of reproduction of output signals for the final data presentation stage.

The **signal conditioning or data acquisition equipment** in many a situation is an excitation and amplification system for **passive transducers**. It may be an **amplification system** for active transducers. In both the applications, the transducer output is brought up to a sufficient level to make it useful for conversion, processing indicating and recording. Excitation is needed for passive transducers because these transducers do not generate their own voltage or current. Therefore, passive transducers like strain gauges, potentiometers, resistance thermometers, inductive and capacitive transducers require excitation from external sources. The active transducers like technogenerators, thermocouples, inductive pick-ups and piezo-electric crystals, on the other hand, do not require an external source of excitation since they produce their own electrical output on account of application of physical quantities. But these signals usually have a low voltage level, and hence, need amplification.

The excitation sources may be an alternating or a d.c. voltage source. The d.c. system is comparatively simple as shown in Fig. 26.1.

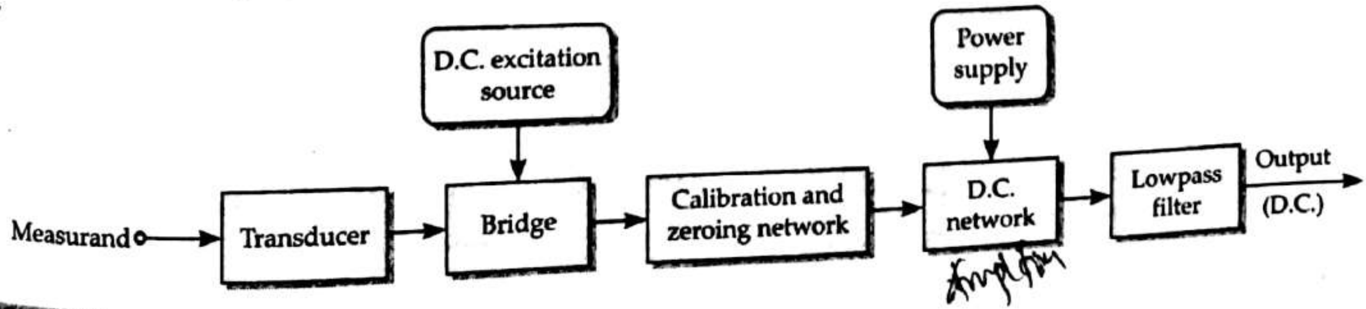


Fig. 26.1 D.C. Signal conditioning system.

The resistance transducers like strain gauges, constitute one arm or more than one arm of a wheatstone bridge which is excited by an isolated d.c. source. The bridge can be balanced by a potentiometer and can also be calibrated for unbalanced conditions.

The desirable characteristics of a d.c. amplifier are :

(*i*) It may need balanced differential inputs giving a high **Common mode rejection ratio** (CMRR). This is elaborated later on in this chapter.

(*ii*) It should have an extremely good thermal and long term stability.

The advantages of a d.c. amplifier are that :

(*i*) It is easy to calibrate at low frequencies.

(*ii*) It is able to recover from an overload condition unlike its a.c. counterpart.

But the greatest disadvantage of a d.c. amplifier is that it suffers from the problem of drift. Thus low frequency spurious signals come out as data information. For this reason special low drift d.c. amplifiers are used. The d.c. amplifier is followed by a lowpass filter which is used to eliminate high frequency components or noise from the data signal.

In order to overcome the problems that are encountered in d.c. systems, a.c. systems are used. In a.c. systems, the **Carrier-type a.c. Signal Conditioning systems** are used as shown in Fig. 26.2.

The transducers used are the variable resistance or variable inductance transducers. They are employed between carrier frequencies of 50 Hz to 200 kHz. The carrier frequencies are much higher, they are at least 5 to 10 times the signal frequencies.

Transducer parameter variations amplitude modulate the carrier frequencies at the bridge output and waveform is amplified and demodulated. The demodulation is **Phase Sensitive** so that the polarity of d.c. output indicates the direction of the parameter change in the bridge output.

In a carrier system amplifier drift and spurious signals are not of much importance unless they modulate the carrier. However, it is more difficult to achieve a stable carrier oscillator than a comparable d.c. stabilized source. In carrier systems, it is easy to obtain very high rejection of mains frequency pick up. Active filters can be used to reject this frequency and prevent overloading of a.c. amplifier. The phase-sensitive, demodulators filter out carrier frequency components of the data signal.

D.C. systems are generally used for common resistance transducers such as potentiometers and resistance strain gauges. A.C. systems have to be used for variable reactance transducers and for systems where signals have to be transmitted via long cables to connect the transducers to the signal conditioning equipment.

After the physical quantities like temperature, pressure, strain, acceleration etc. have been transduced into their analogous electrical form and amplified to sufficient current or voltage levels (say 1 V to 10 V), they are further processed by electronic circuits. In some applications, the signal does not need any further processing and the amplified signal may be directly applied to indicating or recording or control instruments. But many applications involve further processing of signals which involve linear and non-linear operations as mentioned earlier.

The signal may be applied to **sample and hold (S/H) circuit** as shown in Fig. 26.3. This may be fed to an **analog multiplexer** and **analog to digital (A/D) converter**. If the signal is in digital form it may be applied to a variety of digital systems like a digital computer, digital controller, digital data logger or a digital data transmitter.

The sample and hold units shown in Fig. 26.3 sample the different inputs at a specified time and then hold the voltage levels at their output while analog multiplexer performs the **time division multiplexing (TDM)** operation between different

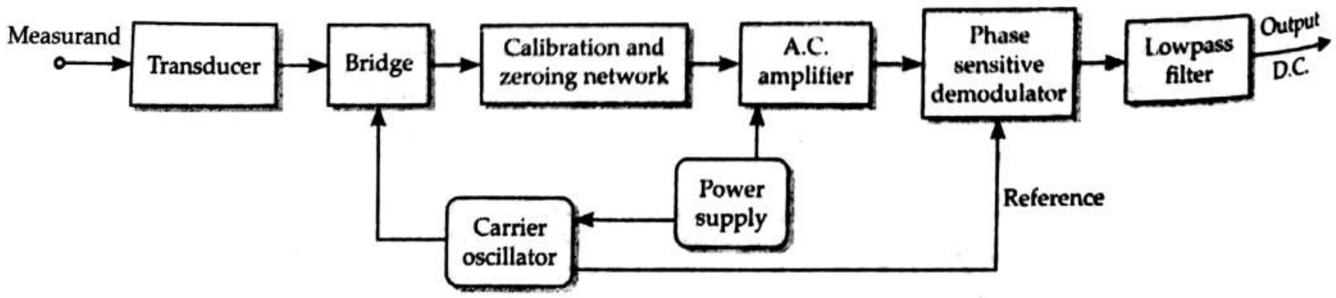


Fig. 26.2 A.C. Signal conditioning system

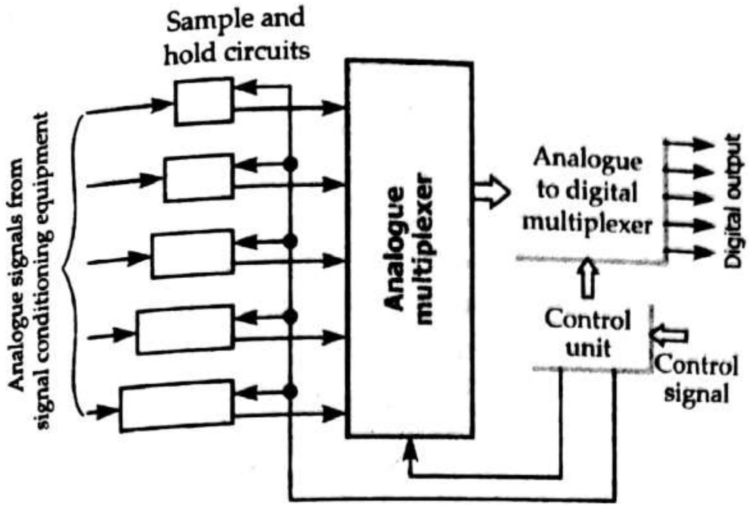


Fig. 26.3 Data acquisition and conversion system.

data inputs. Time Division Multiplexing means that each input channel is sequentially connected to the multiplexer for a certain specified time. (The input signals are not applied to the multiplexer continuously but are connected in turn to the multiplexer thereby sharing time). The timing of the various input channels is controlled by a **control unit**. This unit controls the sample and hold (S/H) circuits, the multiplexer and the analog to digital (A/D) converter. The control unit may be controller itself.

In case time division multiplexing is not used, the **frequency division multiplexer (FDM)** may be used. In this case, the multiple data analog inputs can remain in analog form and are transmitted all at the same time, using frequency division multiplexing (FDM). The voltage input from the signal conditioning equipment is converted into frequency. Thus any change in voltage input of the measurand produces a corresponding change in frequency.

Earlier circuits comprising of discrete electronic components were used where impedance transformations, amplification, and other signal conditioning were required. The requirement to produce designs from discrete components has given way to easier and more reliable methods of signal conditioning which use integrated circuits (ICs). Many special circuits and general purpose amplifiers are now contained in IC packages producing a quick solution to signal conditioning problems, together with small size, low power consumption and, low cost.

26.7 INSTRUMENTATION AMPLIFIERS

The instrumentation amplifier is a dedicated differential amplifier with extremely high input impedance. Its gain can be precisely set by a single internal or external resistor. The high common mode rejection makes this amplifier very useful in recovering small signals buried in large common-mode offsets and noise.

The instrumentation amplifier (*IA*) is not a general building block, like the OPAMP. The instrumentation amplifier (*IA*) is a *closed loop device* with carefully set gain. The OPAMP itself is an open loop device with some very large (but variable) gain. This allows the instrumentation amplifier (*IA*) to be optimized for its role as signal conditioner of low level (often dc.) signals in large amounts of noise. The OPAMP in contrast, can be used to build a wide variety of circuits but does not make as good as a difference amplifier as does an Instrumentation Amplifier (*IA*). Instrumentation amplifiers consist of two stages. The first stage offers very high input impedance to both input signals and allows to set the gain with a single resistor. The second stage is a differential amplifier with the output, negative feedback, and ground connections all brought out.

The input stage is shown in Fig. 26.47. It consists, of two carefully matched OPAMPS. Each input v_1 and v_2 is applied to the non-inverting input terminal of its OPAMP. This OPAMP configured as a voltage follower, produces the instrumentation amplifier's very high input impedance. The outputs of the OPAMPs are connected together through a string of resistors. The two resistors R are internal to integrated circuit while R_g is the gain setting resistor. It may be internal or connected externally.

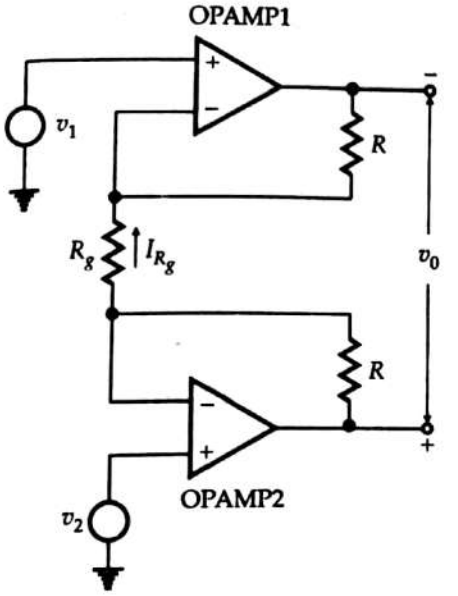


Fig. 26.47 Instrumentation amplifier input stage.

The output voltage is taken between outputs of two OPAMPs.

$$v_0 = (v_2 - v_1)\left(1 + \frac{2R}{R_g}\right) \quad ...(26.93)$$

Equation 26.93 can be derived first noting that there can be no difference in potential between the inverting and non-inverting inputs of OPAMPS. This means that v_1 is at the top of R_g and v_2 is at the bottom of R_g. That is, voltage across R_g is

$$v_{R_g} = v_2 - v_1 = I_g R_g$$

$$I_{R_g} = \frac{v_2 - v_1}{R_g} \quad ...(26.94)$$

This current must flow through all the three resistors because none of the current can flow into OPAMP's input. So the output voltage is,

$$v_0 = I_{R_g} \times (2R + R_g)$$

$$= \frac{v_2 - v_1}{R_g}(2R + R_g)$$

$$= (v_2 - v_1)\left(1 + \frac{2R}{R_g}\right) \quad ...(26.94)$$

Equation 26.95, it is clear that decrease in the value of R_g will increase the output voltage v_0. Therefore, to increase the value of gain, the value of R_g has to be decreased.

The second stage of the instrumentation amplifier is a unity gain differential amplifier. The complete diagram of an instrumentation amplifier (including the first and second stage) is shown in Fig. 26.48. Three terminals are brought out. The sense terminal gives an access to the feedback loop. The reference terminal allows to establish d.c. reference *i.e.,* ground potential of the output. For normal operation the sence terminal is directly connected to output (completing the feedback loop) and tie the reference to ground. This configures the output of a standard differential amplifier as shown in Fig. 26.48.

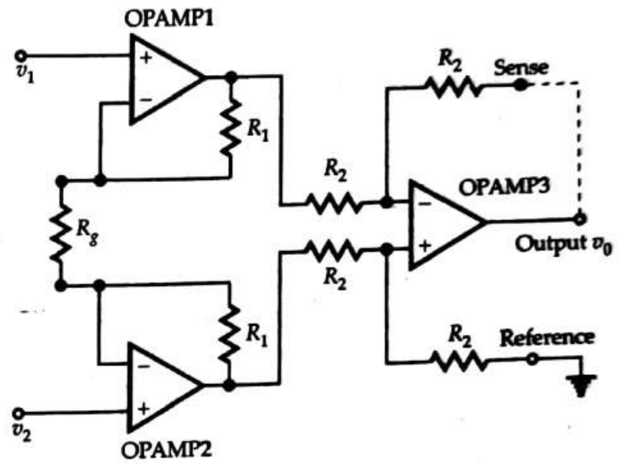


Fig. 26.48 Instrumentation amplifier full schematic.

26.7.1 Three Amplifier Configuration

This circuit is being described to show the elimination of a common mode signal. The circuit of Fig. 26.49 is used for this purpose. This circuit uses three operational amplifiers and hence it is called "three amplifier configuration". v_1 and v_2 are the desired signals and v_{cm} is the common mode signal.

For simplicity it is assumed $R_2 = R_3 = R_4 = R_5$

Therefore, we can write,

$$v_3 = \left(1 + \frac{R_1}{R_g}\right)v_1 - \left(\frac{R_1}{R_g}\right)v_2 + v_{cm}$$

$$v_4 = \left(1 + \frac{R_1}{R_g}\right)v_2 - \left(\frac{R_1}{R_g}\right)v_1 + v_{cm}$$

and $\quad v_0 = v_4 - v_3$

∴ The output voltage is

$$v_0 = \left(1 + \frac{2R_1}{R_g}\right)(v_2 - v_1) \quad ...(29.96)$$

26

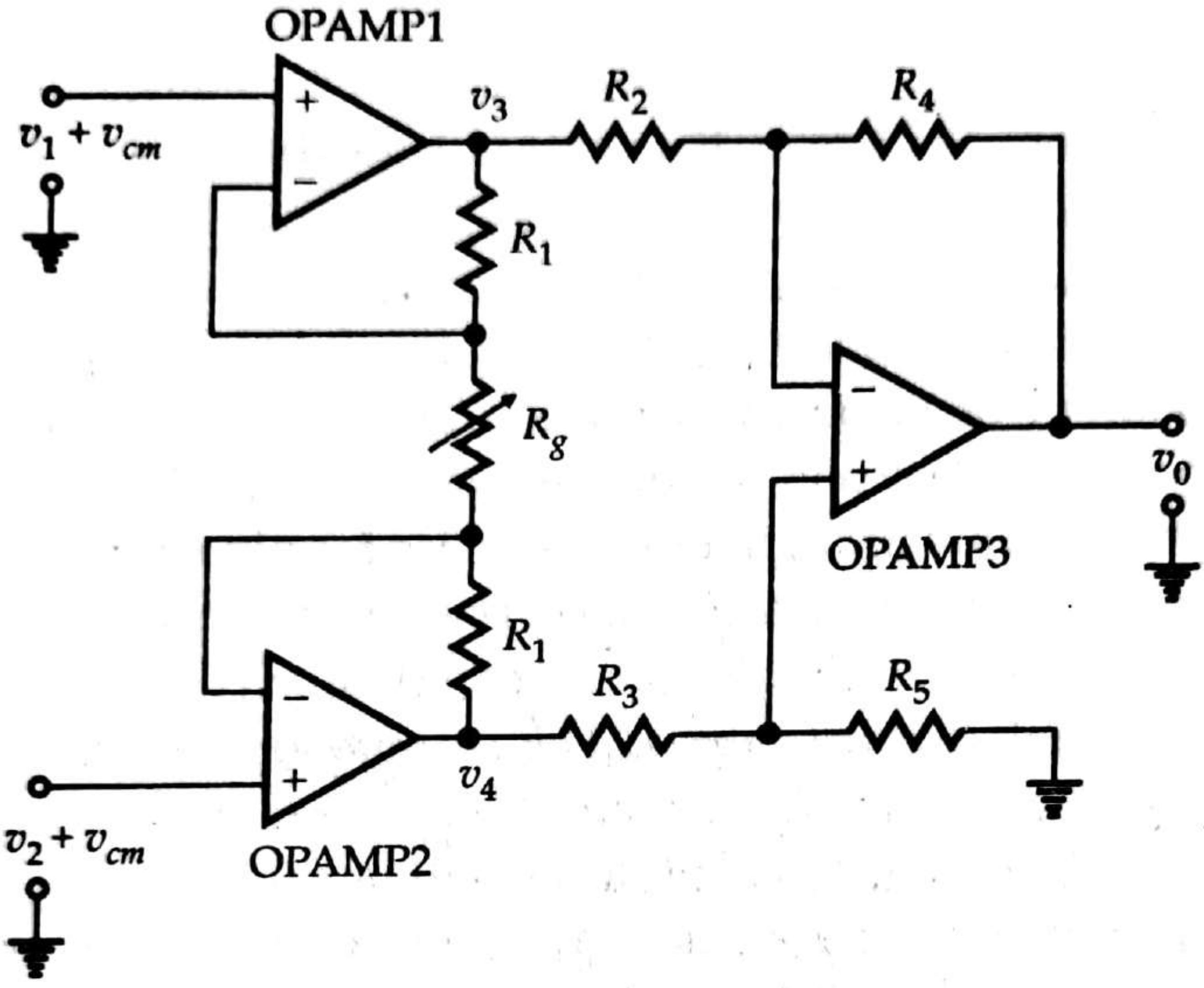


Fig. 26.49 Three amplifier configuration.

OPAMP 1 and OPAMP 2 act as buffers with unity gain for the common mode signal and with a gain of $(1+2\,R_g\,/\,R_1)$ for the differential inputs v_1 and v_2. The circuit has a high input impedance since the OPAMPs 1 and 2 operate in non-inverting mode for common mode signal.

It is clear from Eqn. 26.95 that there is no output corresponding to the common mode input signal. This is because OPAMP 3 acts as a difference amplifier with a high value of CMRR. The gain off course can be changed by changing resistance R_g.

BRIDGE CIRCUITS

10.1 EARLY DEVELOPMENTS

The concept of electric bridge circuits was quite a novelty at a certain point of time, specially because these circuits were found to be extremely useful for the measurement of components like resistance, inductance and capacitance. The earliest bridge circuit was the Wheatstone bridge which was and is still used, although sparingly, for the measurement of resistance only. The circuit has since then been modified for the measurement of very high and/or very low resistances. Bridge circuits are classified in a number of ways. Quite common classification is: (i) DC type and (ii) AC type. An alternative classification is (i) null balance type and (ii) deflecting type. Still another is (i) current sensitive type and (ii) voltage sensitive type.

Null balance type with automated adjustment facility is quite extensively used in many measurement systems. The basic resistance bridge, the Wheatstone bridge shown in Figure 10.1, is still used with necessary changes for accuracy and sensitivity. Wheatstone bridge can be used with both DC and AC supply. In the voltage sensitive type the detector D is virtually an open circuit with an output voltage V_D across it, while in the current sensitive type D is virtually short-circuited with a current I_D flowing through the detector. In the former case the detector is a voltmeter and in the latter it is a galvanometer. For null balance operation one derives the relation

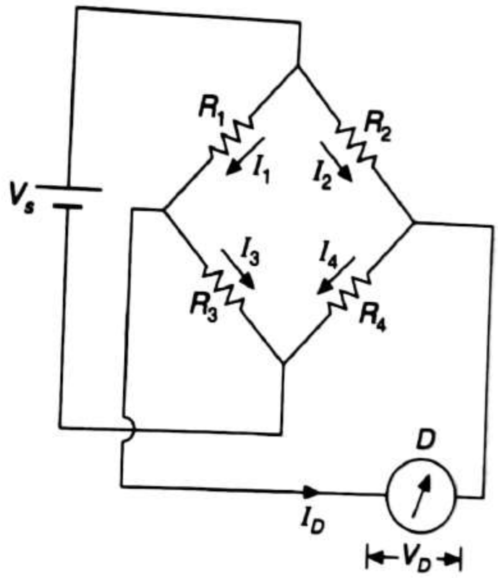


FIGURE 10.1 Basic Wheatstone bridge circuit.

$$I_1 R_1 = I_2 R_2 \tag{10.1}$$

Since no current flows through the voltmeter in case of voltage sensitive type and no current flows through the galvanometer in case of current sensitive type in null-balance operation:

$$I_1 = I_3 = \frac{V_s}{R_1 + R_3} \tag{10.2}$$

Similarly,

$$I_2 = I_4 = \frac{V_s}{R_2 + R_4} \tag{10.3}$$

from which the balance equation is derived as

$$R_1R_4 = R_2R_3 \tag{10.4}$$

For measuring resistance, the test resistance is put in one of the arms of the bridge, usually R_4 in this resistance, so that

$$R_4 = \frac{R_2R_3}{R_1} \tag{10.5}$$

For obtaining balance, R_3 is kept adjustable, and this arm is called the *standard arm*, whereas the arms containing R_2 and R_1 are called the *ratio arms* and lead to the ratio R_2/R_1.

10.2 UNBALANCED MODE

With the bridge used in the unbalanced mode or deflection mode $I_1 \neq I_3$, neither $I_1R_1 = I_2R_2$. However, in the voltage sensitive condition, current through the voltmeter is effectively negligible, so that one has

$$I_1 = \frac{V_s}{R_1 + R_3} \quad \text{and} \quad I_2 = \frac{V_s}{R_2 + R_4}$$

Hence

$$V_0 = I_1R_1 - I_2R_2 = V_s\left[\frac{R_1}{R_1 + R_3} - \frac{R_2}{R_2 + R_4}\right] \tag{10.6}$$

From Eq. (10.6), the null balance condition given by Eq. (10.4) is easily derived for $V_0 = 0$. With this condition, if now R_4 changes by a small amount ΔR_4, then one gets

$$V_0 = \frac{[R_1(R_4 + \Delta R_4) - R_2R_3]V_s}{[(R_1 + R_3)(R_2 + R_4 + \Delta R_4)]} = \frac{\left[1 + \frac{\Delta R_4}{R_4} - \frac{R_2R_3}{R_1R_4}\right]V_s}{\left[1 + \left(\frac{R_3}{R_1}\right)\left(1 + \frac{\Delta R_4}{R_4} + R_2R_4\right)\right]} \tag{10.7}$$

If V_0 changes to $V_0 + \Delta V_0$ and V_0 was initially adjusted to be zero, then

$$V_0 = \frac{\Delta R_4/R_4\ V_s}{\left[2 + \frac{R_3}{R_1} + \frac{R_2}{R_4} + \left(\frac{\Delta R_4}{R_4} + \frac{R_3\Delta R_4}{R_1R_4}\right)\right]} = \frac{\Delta R_4/R_4\ V_s}{\left[2 + \frac{R_3}{R_1} + \frac{R_2}{R_4} + \frac{2\Delta R_4}{R_4}\right]} \tag{10.8}$$

For best operating condition of the bridge, one makes all the bridge arm resistances equal to R initially, so that change ΔR in $R(= R_4)$ produces ΔV_0 in the output, which would then be given by

$$\Delta V_0 = \frac{V_s[\Delta R / R]}{\left[4 + \dfrac{2\Delta R}{R}\right]} \tag{10.9}$$

With $\Delta R/R << 2$, it is further simplified to

$$\Delta V_0 = \left(\frac{V_s}{4}\right)\left(\frac{\Delta R}{R}\right) \tag{10.10}$$

making the output linearly variable with ΔR. The normalized bridge sensitivity is defined as

$$\left(\frac{1}{V_s}\right)\frac{\Delta V_0}{(\Delta R / R)} = \frac{1}{4} \tag{10.11}$$

As such, the bridge sensitivity is directly proportional to $R(= R_4)$ and supply voltage V_s.

For the current sensitive condition the small meter resistance R_m is taken into consideration and the source resistance is assumed zero, so that the equivalent circuit is given by the circuit drawn in Figure 10.2. Obviously the meter sees and equivalent bridge resistance R_e

$$R_e = \frac{R_1 R_3}{R_1 + R_3} + \frac{R_2 R_4}{R_2 + R_4} \tag{10.12}$$

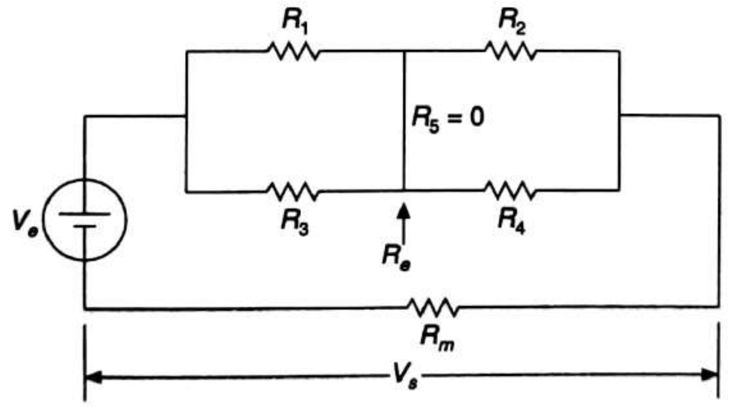


FIGURE 10.2 Equivalent circuit with a 'current' meter detector.

The equivalent source voltage V_e is also obtained by thevenizing, as

$$V_e = V_s \frac{R_1 R_4 - R_2 R_3}{(R_1 + R_3)(R_2 + R_4)} \tag{10.13}$$

From Eq. (10.12), one obtains for incremental change ΔR_4 in R_4,

$$R_e = \frac{R_1 R_2 R_3 + R_1 R_3 R_4 (1 + \Delta R_4/R_4) + (R_1 R_2 + R_2 R_3) R_4 (1 + \Delta R_4/R_4)}{(R_1 + R_3)[R_2 + R_4(1 + \Delta R_4/R_4)]} \tag{10.14}$$

However, the current through the meter I_m is given by

$$I_m = \frac{V_e}{R_e + R_m}$$

$$= \frac{V_s(R_1R_4 - R_2R_3)}{R_1R_2R_3 + R_m(R_1R_2 + R_2R_3) + R_4[R_1R_2 + R_2R_3 + R_3R_1 + R_m(R_1 + R_3)]} \quad (10.15)$$

Again, assuming that the initial setting ensures that all R's are identical and R_4 (=R) changes to $R_4 + \Delta R_4$(= $R + \Delta R$),

$$\Delta I_m = \frac{[V_s(\Delta R / R)]}{[4R + 4R_m + (3R + 2R_m)(\Delta R / R)]} \quad (10.16)$$

Making the valid assumption that $\Delta R/R \ll 1$, Eq. (10.16) reduces to

$$\Delta I_m = \frac{[V_s(\Delta R / R)]}{[4R(1 + R_m/R)]} \quad (10.17)$$

The voltage across the meter is thus

$$\Delta V_m = R_m \Delta I_m = \left[\left\{\frac{\Delta R}{4R}\right\}\left\{\frac{R_m R}{(1 + R_m/R)}\right\}\right] V_s \quad (10.18)$$

From Eqs. (10.10) and (10.18), one derives now

$$\frac{\Delta V_m}{\Delta V_0} = \left(\frac{R_m/R}{(1 + R_m/R}\right) = \frac{R_m}{(R + R_m)}$$

$$= \frac{1}{1 + R/R_m} \quad (10.19)$$

showing that $\Delta V_m = \Delta V_0$ when $R_m \to \infty$

EXAMPLE 10.1

A passive resistive sensor with a nominal value of 120 Ω is connected to the fourth arm of a Wheatstone bridge which has three 120 Ω resistances in the other three arms. The supply voltage is 12 V.

(a) For an ideal voltmeter as a detector, the output voltage varies between −0.3 V to +0.3 V over the operating range of the sensor. Obtain the percentage change of its resistance, and also the bridge sensitivity.

(b) If a galvanometer of resistance 100 Ω is used, what is the current variation for the above range?

Solution:

(a) We have $$\pm\Delta V_0 = \left(\frac{V_s}{4}\right)\left(\pm\frac{\Delta R}{R}\right)$$

10.7 ACTIVE BRIDGE AS A LINEARIZER

The bridge circuit with the insertion of an element can linearize the response characteristic of a passive element, specially a resistance like that of thermistor and RTD. The configuration proposed is a feedback bridge with an operational amplifier inserted. The generalized scheme is shown in Fig. 10.15(a) whose equivalent model is shown in Figure 10.15(b). The nonlinear resistance (with temperature as a functional parameter for example) may be either R_2 or R_3. If R_3 is chosen as the one, with the truncated characteristic, as that of an RTD

$$R_3 = R_{3i}(1 + \alpha_1 t + \alpha_2 t^2) \tag{10.67}$$

and if at $t = t_i$, the initial balance temperature, $R_1 R_7 = R_2 R_{3i}$ where $R_7 = R_4\beta$ and $\beta = R_5/(R_5 + R_6)$, then one obtains

$$\frac{V_0}{V_1} = \frac{(-R_2/R_1)(R_3/R_{3i} - 1)}{(1 + R_2/R_1)(1 - \beta)\left[1 + \dfrac{\{1 - \beta(1 + R_2 R_1)\}\{R_3/R_{3i} - 1\}}{(1 + R_2/R_1)(1 - \beta)}\right]} \tag{10.68a}$$

$$= \frac{-(R_2/R_1)(1 + \alpha_2 t/\alpha_1)\alpha_1 t}{(1 + R_2/R_1)(1 - \beta)\left[1 + \dfrac{\{1 - \beta(1 + R_2/R)\}\{1 + \alpha_2 t/\alpha_1\}\alpha_1 t}{(1 + R_2/R_1)(1 - \beta)}\right]} \tag{10.68b}$$

For RTD, coefficient α_1 is positive and coefficient α_2 is negative, and $|\alpha_2/\alpha_1|$ is quite small. Therefore, proper choice can be made to give an output voltage V_o to be nearly a linear function of temperature over a sufficiently wide range. This circuit also provides lead resistance compensation for both three-wire and four-wire RTD elements in cases of $R_2/R_1 = 1$, $R_2/R_1 = 2$ and $R_2/R_1 >> 2$. Figure 10.15(c) shows the scheme for the last case. The amplifier has to be a high performance one.

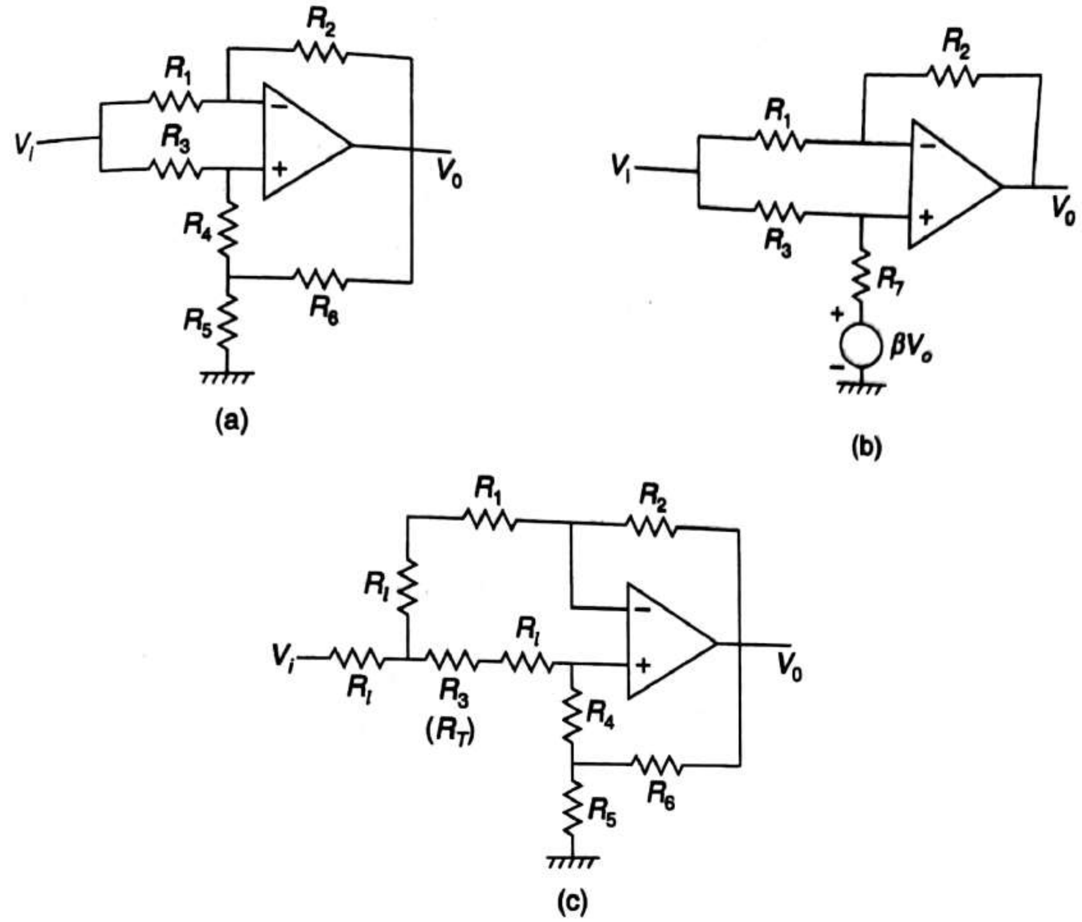


FIGURE 10.15 (a) Generalized configuration of an active bridge, (b) Equivalent model of Figure 10.15(a), (c) Active bridge circuit as used with 4-wire RTD element.

EXAMPLE 10.7

For a test resistance R_T i.e. R_3 with $R_{3i} = 1$ kΩ, where $\alpha_1 = 3.81 \times 10^{-3}/°C$ and $\alpha_2 = -6.17 \times 10^{-7}/(°C)^2$, a bridge circuit is designed with $R_1 = R_2/2 = 10$ kΩ, $R_4 = 1$ kΩ, $R_5 = 4$ kΩ and $R_6 = 20$ kΩ. Obtain the approximate range of linear relation between V_0 and t for a given V_i.

***Solution*:**

Using the relation given in Eq. (10.68b)

$$\frac{V_0}{V_1} = \frac{-2(1 + 1.62 \times 10 - 4t)\, 3.81 \times 10 - 3t}{3\left(1 - 0.1666\left[1 + \dfrac{(1 - 0.166 \times 3)(1 + 1.62 \times 10^{-4}t)3.81 \times 10^{-3}t}{3(1 - 0.166)}\right]\right)} \cong \frac{-3.05 \times 10^{-3}t}{(1 + 0.76 \times 10^{-3}t)}$$

Thus, for $t \leq 130°C$, V_0 is approximately linear. This can however be extended with proper choice i.e. R_5 and R_6 in relation to R_1, R_3 and R_4.

43.2. PRECISION RECTIFIERS

Precision rectifier is a small signal rectifier which is capable of rectifying signals of very small peaks (of the order of a few millivolts). Since such small ac signals are not able to drive a diode directly due to cut-in voltage of the diode, therefore, an op-amp is introduced prior to diode, as illustrated in Fig. 43.10. In the circuit shown diode is used in noninverting configuration. Diode is driven by the op-amp output. When diode is *off*, op-amp acts in open loop configuration, as illustrated in Fig. 43.11 (*a*). When diode is *on*, op-amp acts in voltage follower configuration as illustrated in Fig. 43.11 (*b*).

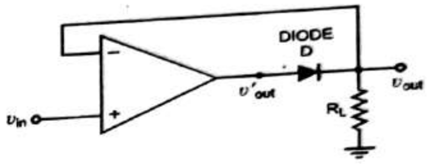
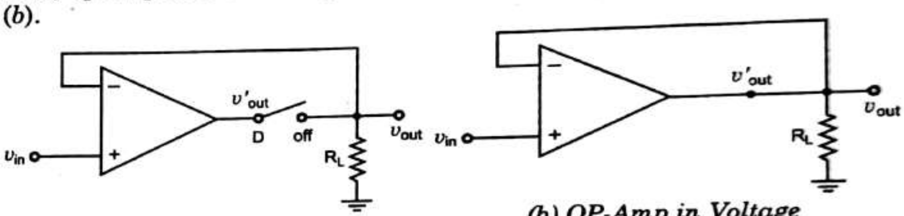


Fig. 43.10

When input signal v_{in} goes negative a high negative signal appears at the output of op-amp making the diode reverse-biased disconnecting v_{out} and v'_{out} (op-amp output). Thus during negative half cycle no output is obtained.

Now when the diode is *off*, the input is zero and op-amp acts in open-loop configuration, as shown in Fig. 43.11 (*a*). Thus just at the start of positive half cycle, op-amp is in open-loop. At the start of positive half cycle, when input goes positive, due to high open-loop gain op-amps output goes high driving the diode *on*. Op-amp comes in voltage follower configuration. Since in voltage follower configuration closed-loop gain of the device is unity, the output voltage equals input voltage. Thus output voltage equal to input voltage is available at the output terminal of the circuit.



(*a*) *OP-Amp in Open-Loop Configuration*

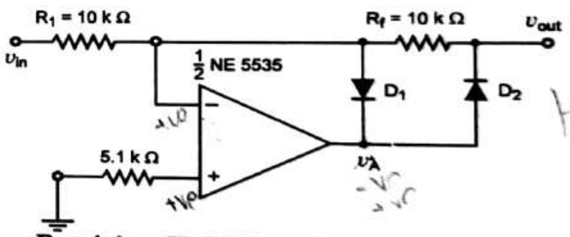
(*b*) *OP-Amp in Voltage Follower Configuration*

Fig. 43.11

Another half-wave rectifier using op-amp is depicted in Fig. 43.12. The op-amp used is high SRNE5535 dual op-amp.

When input voltage v_{in} positive, v_A is negative, diode D_1 being forward biased is *on* while diode D_2 being reverse-biased is *off*. No current flows through R_f, so that $v_{out} \cong v_{in} \cong 0$. The input current all flows through diode D_1.

When input voltage v_{in} is negative, v_A is positive, diode D_1, being reverse-biased, is off and diode D_2 being forward biased is *on*. The



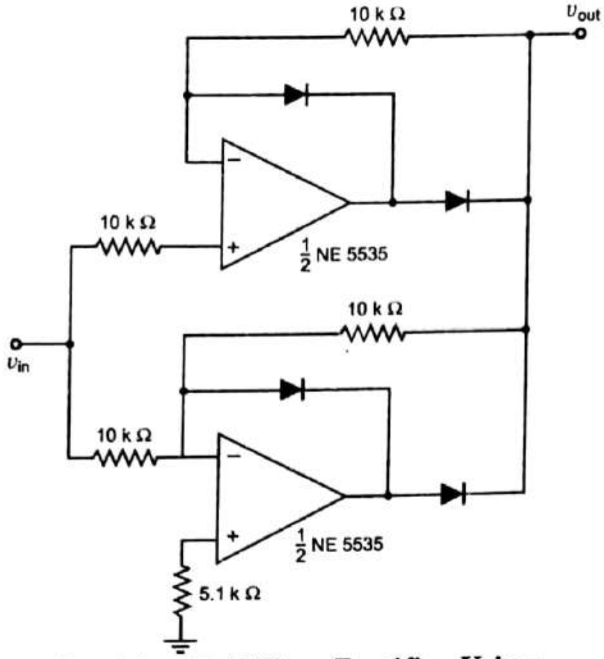
Precision Half-Wave Rectifier Using NE 5535 High SR Dual Op-Amp

Fig. 43.12

Scanned by CamScanner

currents through R_1 and R_f are approximately equal. Given $v_{in} \cong 0$, $v_{out} \cong -v_{in}$. The result is a half-wave rectified output, with no diode drops to reduce the signal.

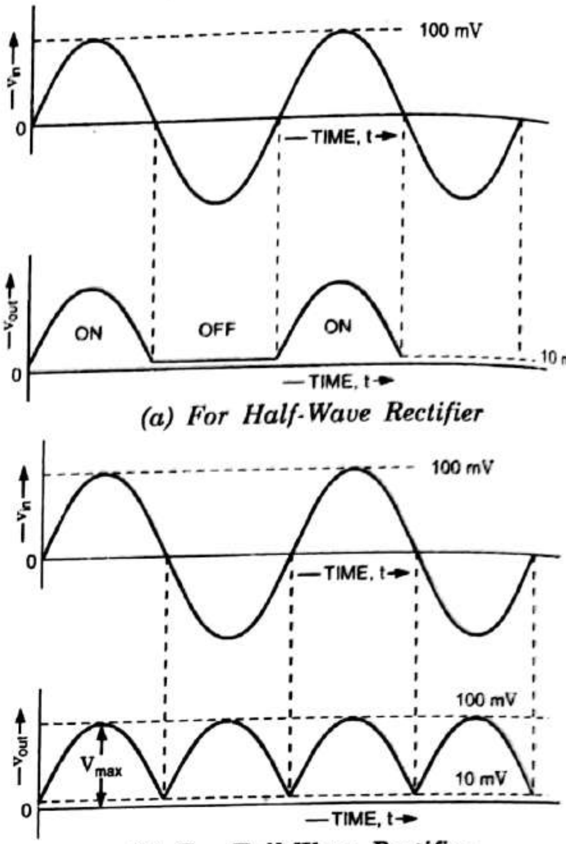
Two half-wave rectifier circuits can be combined, as shown in Fig. 43.13, to provide a precision full-wave rectifier. Note that input connections are reversed for the second op-amp.



Precision Full-Wave Rectifier Using NE 5535 High SR Dual Op-Amps

Fig. 43.13

Wave shapes of precision half-wave and full-wave rectifiers are shown in Figs. 43.14 (*a*) and 43.14 (*b*) respectively.



(a) For Half-Wave Rectifier

(b) For Full-Wave Rectifier

Wave Forms of Precision Rectifiers

Fig. 43.14

The advantages of precision rectifiers are that (*i*) they can rectify signals of very low amplitude (all the way down to the mV level) and (*ii*) they are non-saturating ones.

The inverting characteristics of the output V_{out} can be circumvented by the use of an additional inversion for achieving a positive output.

The precision rectifiers, however, have their speed limited by the slew rate of the op-amp. The operation of the circuit using 741 is usually restricted below 1 kHz. When JFET input op-amps are used, the operating range may be increased to the 10 kHz range.

Example 43.2. The input voltage v_{in} in the circuit shown in Fig. 43.15 is a 1 kHz sine-wave of 1 V amplitude. Assume ideal operational amplifiers with ± 15 V dc supply. Sketch on a single line diagram the waveforms of the voltage v_{in}, v_0 and v_1 shown, indicating the peak value of v_1 and the average value of v_0. [*GATE 1999*]

Solution : Input voltage, $v_{in} = V_{in} \sin 2\pi f t = 1 \sin 2\pi \times 1{,}000\, t = 1 \sin 6{,}283\, t$

First op-amp is operating as a half-wave rectifier with amplification factor

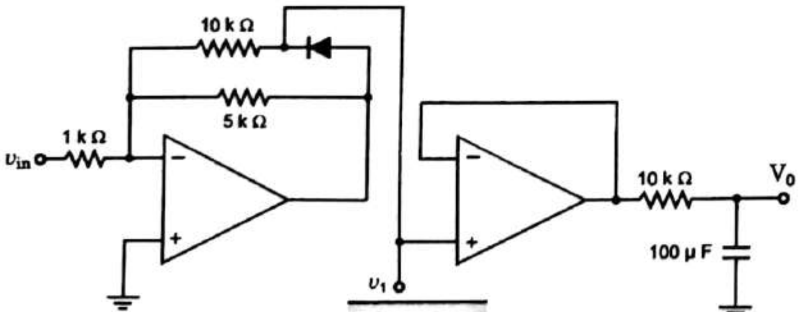


Fig. 43.15

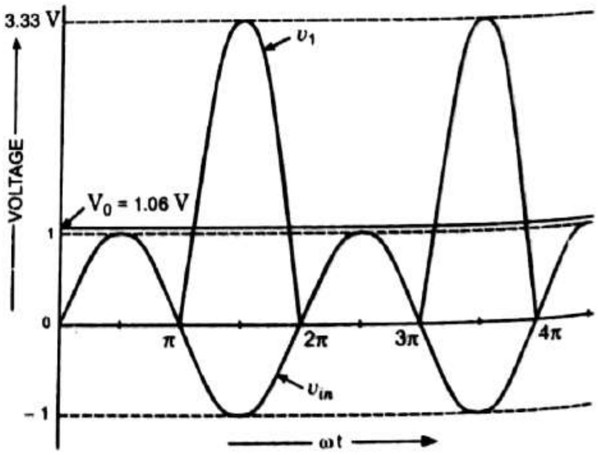


Fig. 43.16

$$= \frac{10\,k\Omega \,||\, 5\,k\Omega}{1\,k\Omega} = \frac{10 \times 5}{10 + 5} = 3.33$$

$$\text{Hence } V_1 = 3.33 \sin 6{,}283\, t \qquad \pi < \omega t < 2\pi$$
$$= 0 \qquad 0 < \omega t < \pi$$

Second op-amp network is operating as an average detector

$$\text{So } V_0 = \frac{\text{Peak value of } V_1}{\pi} = \frac{3.33}{\pi} = 1.06\text{ V}$$

Peak value of V_1 = 3.33 V **Ans.**

Average value of V_0 = 1.06 V **Ans.**

Waveforms of voltages v_{in}, v_1 and v_0 are shown in Fig. 43.16.

43.1. LOGARITHMIC AND ANTILOGARITHMIC AMPLIFIERS

Logarithmic amplifiers have several areas of applications. Analog computation may need $\log_e x$, $\log_{10} x$, or sin hx. This can all be performed continuously with analog / log amplifiers. It is very convenient to have direct dB displays on digital voltmeters and spectrum analyzers which employ log amplifiers. Signal dynamic range compressors enhances computer signal processing as well as audio recording. Logarithmic amplifiers can be employed to compress a signal's dynamic range.

43.1.1. Logarithmic Amplifiers

The fundamental log op-amp circuit is shown in Fig. 43.1(a) where a feedback resistor is replaced by a diode. Here the output voltage obtained is proportional to logarithm of the input voltage.

The V-I relationship in the feedback path is given by

$$I_f = I_0\, e^{(V_D/\eta V_T - 1)} \qquad ...(43.1)$$

$$\text{or } I_f \simeq I_0\, e^{V_D/\eta V_T} \qquad ...(43.2) \quad \left(\text{assuming } \frac{V_D}{\eta V_T} >> 1 \text{ or } I_f >> I_0\right)$$

$$\text{or } \frac{I_f}{I_0} = e^{V_D/\eta V_T}$$

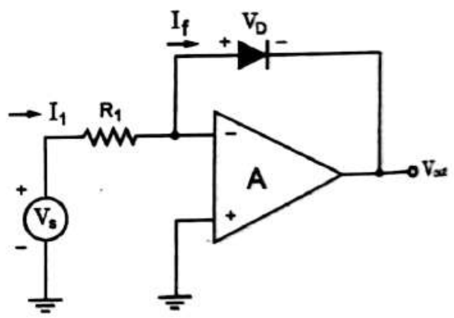
$$\text{or } \log\left(\frac{I_f}{I_0}\right) = \frac{V_D}{\eta V_T}$$

$$\text{or } V_D = \eta V_T (\log_e I_f - \log_e I_0) \qquad ...(43.3)$$

$$\text{Also } I_f = I_1 = \frac{V_s}{R_1} \qquad \text{(due to virtual ground)}$$

$$\text{and } V_{out} = -V_D$$

$$\text{or } V_{out} = -\eta V_T \left(\log \frac{V_s}{R_1} - \log I_0\right) \qquad ...(43.4)$$



(a) *Basic Log Op-amp*

Equation (43.4) gives us the result that output voltage V_{out} is proportional to logarithm of input voltage V_s.

We can also find $\log_{10} x$ by using relation

$$\log_{10} x = 0.4343 \log_e x.$$

The circuit, however, shows that V_{out} is dependent on ηV_T and on saturation current I_0 and both these have temperature dependency.

The temperature effects can be reduced by using circuit shown in Fig. 43.1 (*b*) where diodes D_1 and D_2 are matched, we have used a constant current source I which is independent of temperature, although R_T is a temperature dependent resistor.

Here in Fig. 43.1 (*b*) the input voltage at noninverting terminal of second op-amp is V_2 (say) and diode D_2 voltage say, V_{D2} so that

$$V_{out} = V_2 - V_{D2} \quad \text{...(43.5)}$$

or $$V_2 = V_{out} + V_{D2} \quad \text{...(43.6)}$$

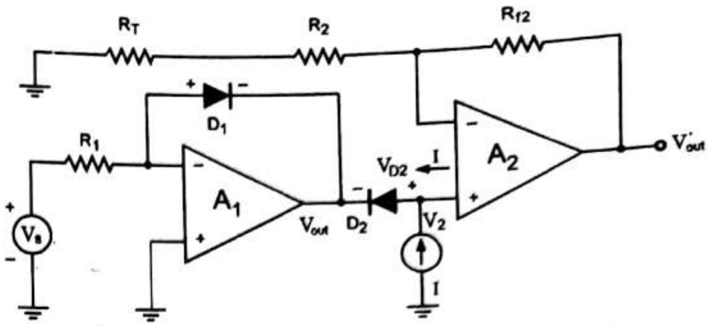
Now using Eq. (43.3) we can express V_{D2} as

$$V_{D2} = \eta V_T (\log_e I - \log_e I_0) \quad \text{...(43.7)}$$

Using Eqs. (43.4), (43.6) and (43.7) we have

$$V_2 = -\eta V_T \left(\log_e \frac{V_s}{R_1} - \log_e I_0 \right) + \eta V_T (\log I - \log I_0)$$

$$= \eta V_T \left(\log_e I - \log_e \frac{V_s}{R_1} \right) = -\eta V_T \log_e \left(\frac{V_s}{I R_1} \right) \quad \text{...(43.8)}$$



(*b*) *Log Op-amp With Temperature Compensation*

We can express V'_{out} as

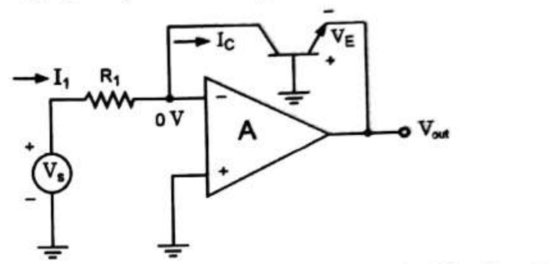
$$V'_{out} = \left(1 + \frac{R_{f2}}{R_T + R_2} \right) V_2 \quad \text{...(43.9)}$$

So that the output voltage

$$V'_{out} = -\left(1 + \frac{R_{f2}}{R_T + R_2} \right) \eta V_T \log_e \left(\frac{V_s}{IR_1} \right) \quad \text{...(43.10)}$$

Here, the resistance R_T is selected to compensate approximately for the factor ηV_T.

The diode of basic log op-amp circuit of Fig. 43.1(*a*) can also be replaced by a transistor, as shown in Fig. 43.1(*c*).



(*c*) *Log Op-amp Using Transistor in Feedback*

Fig. 43.1

It can be shown that circuit shown in Fig. 43.1(*c*) gives the same result as with diode

$$I_E = I_0 \, (e^{V_E/\eta V_T - 1})$$

As the base of transistor is grounded $I_C \simeq I_E$,

$$I_C = I_0 \, (e^{V_E/\eta V_T - 1})$$

Log Amplifier Circuit Using Two Difference Amplifiers. The temperature effects can also be reduced by using difference logarithmic amplifier circuit shown in Fig. 43.2.

The emitter voltage of transistor Q_1 can be had from Eq. (43.4). This is also V_A.

$$V_A = -\eta V_T \log \frac{V_s}{I_0 R_1} \quad \text{...(43.11)}$$

The base-emitter voltage for Q_2 is

$$V_{Q_2BE} = \eta V_T \log \frac{V_{REF}}{I_0 R_1} \quad \text{...(43.12)}$$

The base voltage of Q_2, then is

$$V_{Q_2 \, base} = V_{Q_2 \, emitter} + V_{Q_2BE}$$

But $V_{Q_2 \, base} = V_B$ and $V_{Q_2 \, emitter} = V_A$

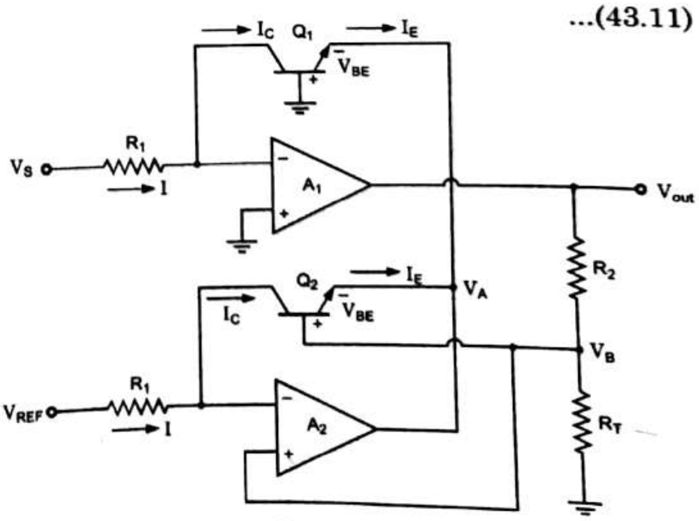
Thus, $$V_B = V_A + V_{Q_2BE} \quad \text{...(43.13)}$$

Now from Eqs. (43.11), (43.12) and (43.13) we have

$$V_B = -\eta V_T \log \frac{V_s}{I_0 R_1} + \eta V_T \log \frac{V_{REF}}{I_0 R_1}$$

$$= -\eta V_T \left(\log \frac{V_s}{I_0 R_1} - \log \frac{V_{REF}}{I_0 R} \right)$$

$$= -\eta V_T \log \frac{V_s}{V_{REF}} = -\eta V_T \log \frac{V_s}{IR_1} \quad \text{...(43.14)}$$



Difference Logarithmic Amplifier

Fig. 43.2

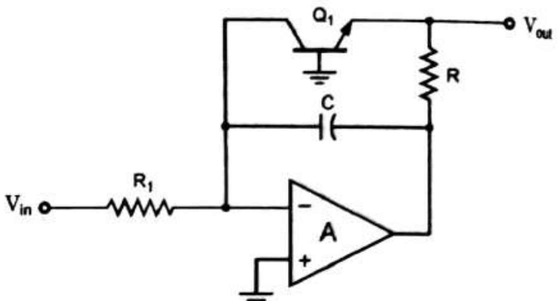
This is the voltage in the centre of the voltage divider

$$V_B = \frac{R_T}{R_2 + R_T} V_{out} \qquad ...(43.15)$$

From Eqs. (43.14) and (43.15) we have

$$-\eta V_T \log \frac{V_s}{IR_1} = \frac{R_T}{R_2 + R_T} V_{out}$$

or $$V_{out} = -\frac{R_2 + R_T}{R_T} \eta V_T \log \frac{V_s}{IR_1} \qquad ...(43.16)$$



Log Amplifier Frequency Compensation

Fig. 43.3

The worthnoting point is that the specifications of op-amp need careful consideration. At low levels of input current (and, therefore, input voltage), bias current and drift can cause linearity and log conformance errors. With bias currents as large as 10 percent of the input current, output errors of the order of 2.5 mV may be encountered.

Frequency stability considerations may need external compensation. This is normally had by placing a capacitor across the negative feedback element. This causes lowering in gain with rising frequency. However, the effective resistance in the negative feedback loop is

$$r_e \cong \frac{26 \text{ mV}}{I_C}$$

At high levels of input (collector) current, r_e becomes very small, shorting out the parallel compensation capacitor. The capacitor becomes ineffective. The solution is to include a resistor in series with the emitter, placing a lower limit on the feedback resistance through the transistor.

A typical frequency compensation network is given in Fig. 43.3. Values of R and C are best determined experimentally, as parasitic effects play a vital role.

2.7 Charge Amplifiers

Charge amplifiers make use of electrometer (FET input) op-amps, which have ultra-low dc bias currents at their inputs and ultra-high input resistances. They find primary application in isolating and conditioning the outputs of piezoelectric transducers used to measure pressure and force transients. They are also used to measure the coulombic charge stored in capacitors and indirectly, the electric field strength.

2.7.1 Charge Amplifiers Used with Piezoelectric Transducers

In Chapter 6, we show that the equivalent circuit of a piezoelectric pressure transducer operated well below its mechanical resonant frequency is given by the simple parallel current source, conductance and capacitance, shown in Figure 2.47. Note that d is a constant which depends on the piezoelectric material and how it is cut relative to the crystal axes to form the transducer and that in general, current is the rate of transfer of charge through the plane through which it is passing. By U.S. convention, current is conventionally taken as positive in the direction of motion of positive charges, such

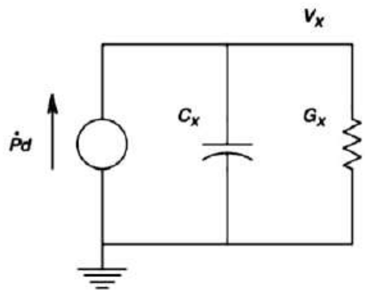


FIGURE 2.47
Equivalent circuit of a piezoelectric crystal responding to pressure on its active surface at frequencies well below its mechanical resonance frequency.

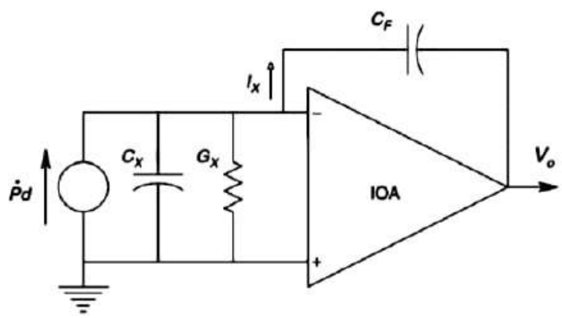


FIGURE 2.48
A charge amplifier circuit used to condition the output of a piezoelectric crystal sensor.

as holes in semiconductors, or positive ions in solutions or plasma discharges. Hence in a conductor, such as copper, where charge is carried by electrons, the current direction is opposite to the direction of electron motion.

Figure 2.48 illustrates an ideal op-amp used as a charge amplifier. The op-amp is assumed to have infinite gain, frequency response and input resistance, and zero bias current, offset voltage, noise and output resistance. Under these ideal op-amp assumptions, it is easy to see that because the summing junction is at zero volts, all of the current, $\dot{P}d$, flows into the feedback capacitor, C_F. Hence V_o is proportional to $\dot{P}(t)$, and is given by:

$$V_o(t) = -\dot{P}(t)\,d/C_F \tag{2.152}$$

In actual practice, this result of the ideal case is seen to become less simple and it should be noted that the electrometer op-amp has a non-infinite gain approximated by:

$$K_V = \frac{K_{Vo}}{(\tau s + 1)}(V_i - V_i') \tag{2.153}$$

Also, we assume that there is a small but finite conductance, G_F, in parallel with C_F. The G_F is of the order of 10^{-10}S, or smaller. Now, if we write the node equation at the summing junction and substitute the gain expression above to eliminate V_i', we can show that the charge amplifier transfer function is given by (approximately):

$$V_o/P = \frac{-s\,(d/G_F)}{s^2[\tau(C_T + C_F)/K_{Vo}G_F] + s[\tau(G_T + G_F) + K_{Vo}C_F]/K_{Vo}\,G_F + 1} \tag{2.154}$$

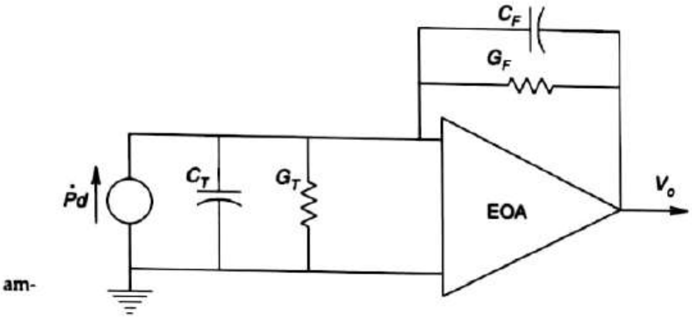


FIGURE 2.49
Circuit of a practical charge amplifier and crystal sensor.

This is the transfer function of a bandpass system. The low frequency pole is approximately at:

$$f_{\text{low}} \approx 1/(2\pi C_F R_F)\ \text{Hz} \tag{2.155}$$

The mid-band gain is:

$$A_{V\text{mid}} = -K_{Vo}d/[\tau(G_T + G_F) + K_{Vo}C_F] \cong -d/C_F \tag{2.156}$$

The high frequency pole is at:

$$f_{\text{hi}} = K_{Vo}C_F/[2\pi\tau(C_T + C_F)]\ \text{Hz} \tag{2.157}$$

Note that $K_{Vo}/(2\pi\tau)$ is simply the small signal GBWP of the electrometer op-amp. Figure 2.49 shows the circuit of the non-ideal charge amplifier. Here, C_T is the total capacitance to ground which includes the capacitances of the piezoelectric crystal, the cable connecting the transducer to the charge amp, and the charge amp input. Typically, C_T might be about 400 pF. The G_T is the total shunt conductance to ground at the op-amp summing junction. The G_T includes the leakage conductances of the crystal, the cable, and the op-amp input. Typically, G_T is about 10^{-13} S. If, in addition, $C_F = 2 \times 10^{-9}$ F, $G_F = 10^{-10}$ S, $K_{Vo} = 5 \times 10^4$, $d = 2 \times 10^{-12}$ amp.sec/psi for quartz, and the op-amp GBWP = 350 kHz, the system ends up having a mid-band gain of 1 mV/psi, a high frequency pole at 1.5×10^5 Hz, and a low frequency pole at 8×10^{-3} Hz. Note that the low frequency pole is set by C_F/G_F, and is independent of the input circuit. Thus, the piezoelectric transducer charge amplifier is not a system which can respond to static (dc) pressures; it must be used to measure transient pressure changes. In addition to the low frequency limitations imposed on the transfer function by non-ideal electrometer op-amp behavior, there are also dc errors at the output due to the bias current and dc offset voltage of the op-amp. Figure 2.50 shows the dc equivalent circuit of the charge amplifier. Note that the capacitors drop out from the circuit at dc because they carry no current. It is easy to see that the dc error in the output is:

$$V_{oDC} = I_B R_F + V_{OS}(1 + G_T/G_F) \tag{2.158}$$

Note that I_B and V_{OS} can have either a plus or a minus sign. Typical values of V_{OS} are about $\pm 200\ \mu$V, and I_B values are around 75 fA (75×10^{-15} A).

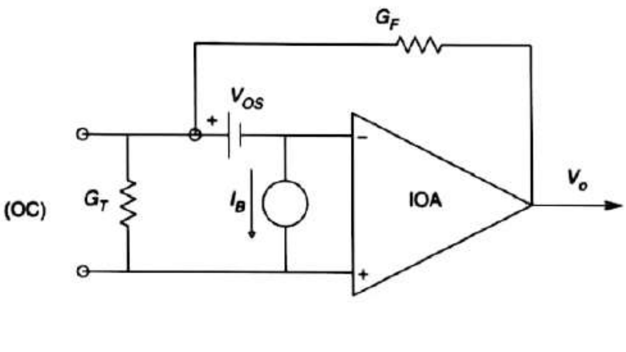


FIGURE 2.50
The dc equivalent circuit of a practical charge amplifier.

2.7.2 Charge Amplifier as Integrating Coulombmeter

A charge of Q_1 coulombs is trapped on a capacitor, C_1. The capacitor can be a real capacitor, or the capacitance of a charged object, such as the fuselage of a helicopter, or a human body. We wish to measure Q_1. The voltage across C_1 is given by the well-known relation, $V_1 = Q_1/C_1$. In Figure 2.51, C_1 is connected to the summing junction of an electrometer op-amp. At the same time, the short is removed from the feedback capacitor, C_F, allowing it to charge from the op-amp output to a value which maintains the summing junction at 0 V. Assuming the op-amp to be ideal, the output voltage of the op-amp, after three or four R_1C_1 time constants, is:

$$V_{o(SS)} = -Q_1/C_F \text{ dc V} \tag{2.159}$$

Hence, the system can measure the charge on a small capacitor, provided the peak current, $i_{1PK} = Q_1/R_1C_1$, does not exceed the current sourcing or sinking capability of the op-amp.

In practice, this circuit is subject to dc errors caused by the bias current and offset voltage of a practical op-amp. The output error caused by I_B and V_{OS} can be found by superposition and is given by:

$$\delta V_o = T(I_B/C_F) + V_{OS}(1 + C_1/C_F) \text{ V} \tag{2.160}$$

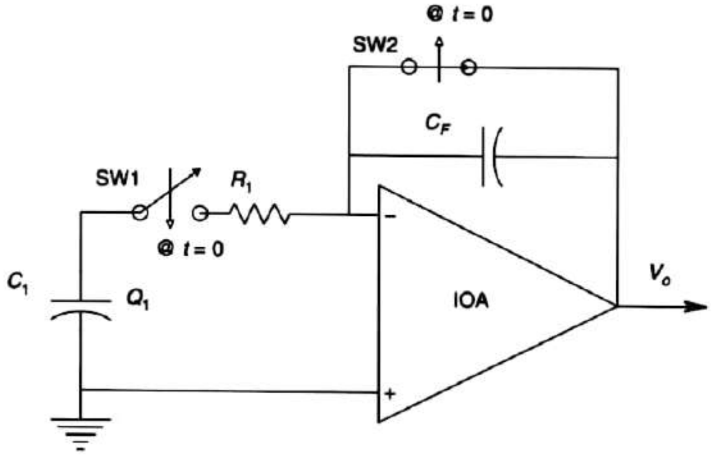


FIGURE 2.51
An integrating coulombmeter circuit. An electrometer op-amp is generally used.

where T is the time elapsed between opening SW1 and closing SW2, and reading the steady state, V_o. In practice, δV_o can be made negligible by the judicious choice of circuit parameters.

2.7.3 Summary

In this section, we have presented two examples of applications of the op-amp charge amplifier configuration. The basic charge amplifier consists of an ideal op-amp, with an ideal feedback capacitor to integrate charge transferred to the summing junction. In practice, we use electrometer amplifiers which have ultra-high input resistances and finite gains, bandwidths, offset voltages and bias currents. Since a pure C_F will integrate the bias current of the summing junction, a large resistance, R_F, must be put in parallel with C_F to give a fixed dc error, which is proportional to I_B. The parallel combination of R_F and C_F sets the low frequency pole of a bandpass frequency response for the piezoelectric transducer. In the mid-band range of frequencies, the output voltage is proportional to the pressure on the transducer, rather than its derivative.

2.5.2 Isolation Amplifiers

Isolation amplifiers are a special class of IAs which find extensive application in biomedical instrumentation, specifically, the recording of low level bioelectric signals from human and animal subjects. Their use is generally dictated by electrical safety considerations. They are also used when the input signal is found in the presence of a high common-mode voltage, as well as in a situation where ground loop currents can introduce errors in the signals under measurement. Isolation amplifiers have an extremely low conductance between the reference ground of the output signal and the

ground and input terminals of the differential headstage. The dc conductive isolation between the headstage ground and the output stage ground is generally accomplished by a magnetic transformer coupling or by opto-coupling techniques.

In the Analog Devices family of isolation amplifiers, the output of the input DA is coupled to the output stage of the isolation amplifier by an isolation transformer. The operating power for the input DA is also coupled with the transformer from a high frequency power converter oscillator, which derives its power from the non-isolated or normal supply for the isolation amplifier. The block diagram of an Analog Devices, AD295 precision hybrid IA, is shown in Figure 2.28. Note that in this design, both the input DA and the output circuit have isolated (internal) power supplies.

To illustrate the differences between ordinary instrumentation amplifiers and isolation amplifiers, some of the key specifications for the Analog Devices Model AD294A Medical Isolation Amplifier are reviewed below.

The overall gain is the product of G_{IN} and G_{OUT}.

$$G_{IN} = (1 + 10^5/R_G), \quad R_G \geq 1\,\text{k}\Omega$$

$$G_{INmax} = 100, \quad G_{OUT} = \left(1 + \frac{R_A}{R_B}\right), \quad 1 < G_{OUT} < 10$$

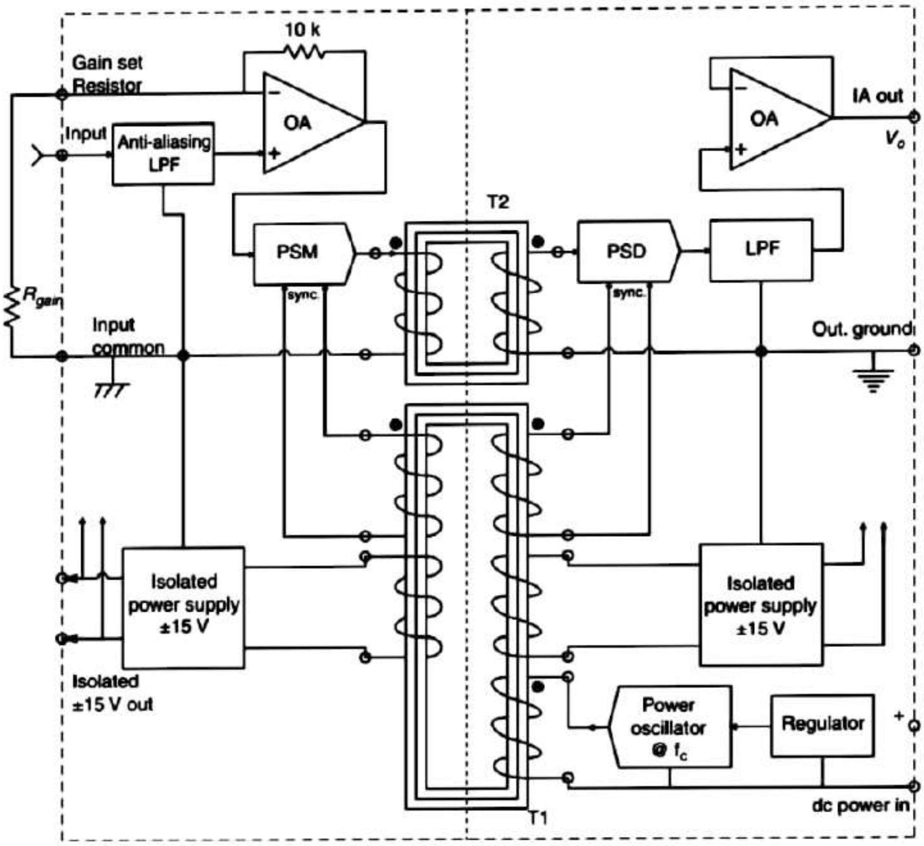


FIGURE 2.28
Simplified block diagram of an Analog Devices AD295 precision isolation amplifier. (Source: Figure used with permission of Analog Devices)

Hence the overall gain may be 1000. The CMRR of the AD294A is 100 dB, not exceptionally high for an IA. The maximum continuous common-mode voltage (CMV) is ±3500 Vpk. The amplifier is rated to stand a ±8000 Vpk, 10 ms pulse at the rate of 1 pulse every 10 seconds. The input impedance is $10^8\ \Omega \| 150$ pF. The input (dc) bias current I_B, is 2 nA. The small signal bandwidth is 2.5 kHz (gain 1–100 V/V). The slew rate is 9.1 V/μs. The maximum leakage current between headstage and output stage is 2 μA rms, given that 240 V rms 60 Hz is applied to the input. The AD 294A is typically used to record electrocardiographic or electromyographic signals using body surface electrodes.

Another medical isolation amplifier suitable not only for ECG recording, but also EEG signal conditioning, is the Intronics Model IA296. This isolation amplifier also uses the two-transformer architecture to couple power into the differential headstage and signals out. Its specifications are summarized below.

The gain of IA296 is fixed at 10. The CMRR is 160 dB with a 5 kΩ source resistance imbalance. The maximum safe dc common-mode input voltage is ±5000 V. The CM input impedance is $10^{11}\ \Omega \| 10$ pF and the DM input impedance is $3 \times 10^8\ \Omega \| 2.2$ pF. The input bias current is ±200 pA and the I_B tempco is ±5 pA/°C. The −3 dB bandwidth is 1 kHz. Input noise with a single-ended input with $R_s = 5\ \text{k}\Omega$ is 0.3 μV in a 10 to 1 kHz bandwidth (about $9.5\ \text{nV}/\sqrt{\text{Hz}}$); the current noise is about $0.13\ \text{pA}/\sqrt{\text{Hz}}$ in a 0.05 Hz to 1 kHz bandwidth. Maximum leakage current due to component failure is 10 μA. Due to its low noise, the Intronics IA296 is better suited for measurements of low level biological signals such as fetal ECGs, EEGs and evoked brain potentials.

Finally, as an example of a photo-optically coupled isolation amplifier, we examine the Burr-Brown Model 3652JG. Since this amplifier does not use a built-in transformer to couple power to the input DA, an external, isolating dc/dc converter must be used, such as the Burr-Brown Model 722. The BB722 supply uses a 900 kHz oscillator to drive a transformer which is coupled to two independent rectifiers and filters. It can supply ±5 to ±16 V dc, at about 16 mA to each of the four outputs. It has isolation test ratings of ±8000 Vpk for 5 seconds, and ±3500 Vpk continuous between inputs and outputs. The leakage current maximum for this power supply is 1 μA for 240 V rms at 60 Hz, insuring medical safety. The Model 3652 isolation amplifier has the following specifications when used with the BB722 dc/dc power supply.

Gain is settable in the range 1–1000. The CMRR is 80 dB. The BB3652 can block a CM voltage of ±2000 V dc and a pulse of ±5000 V for 10 ms. The input resistance is given as $10^{11}\ \Omega$, CM and DM. The isolation impedance between input and output is $10^{12}\ \Omega \| 1.8$ pF. The input bias current is 10 pA. The offset voltage tempco is 25 μ V/°C. The −3 dB bandwidth is 15 kHz at all gains. The short circuit voltage noise is 5 μV rms, measured over a 10 Hz to 10 kHz bandwidth. No figure is given for the input current noise. Maximum leakage current is 0.35 μA.

3.6.1 Phase-Sensitive Detectors

A phase-sensitive detector is used for detection and measurement of the amplitude and phase of a periodic signal with respect to another reference signal of the same frequency. It is also known as lock-in amplifier and synchronous demodulator. In instrumentation, a large number of situations exist where the desired information resides in both the amplitude and phase of a signal. For example, the unbalance voltage of an ac bridge network has its magnitude and phase dependent on the components of the unknown impedance. The excitation voltage of the bridge serves as the reference signal.

The basic scheme of operation of a phase-sensitive detector (p.s.d.) is shown in Figure 3.54(a). The reference signal is assumed to be a square wave driving the switch towards the contactor 1 during the positive half-cycle and to contactor 2 during the negative half-cycle. The sinusoidal test signal is amplified to assume reasonable amplitude and then applied to the switching system. The reference and test signals are of the same frequency and hence the output of the adder consists of full-wave rectified signal, if the signals are in phase. Then the average value is read on a pmmc system and is maximum. When the test and reference signals are in phase, the operation is similar to an ac-to-dc converter, but when there is a phase difference between them, the average output voltage and the meter reading become functions of the phase angle θ of the test voltage with respect to the reference voltage. When $\theta = 180°$, the reading reaches a negative maximum, and when $\theta = 90°$, the reading is zero as can be seen from Figure 3.54(b). The low-pass filter removes the ac ripple components and presents ripple-free dc output voltage for indication. The reading can be shown to be $(2/\pi)V_{i\ \max}\cos\theta$.

In case the frequency of the test signal voltage is not exactly the same as that of the reference signal, the output of the adder in each half-cycle becomes erratic, changing from positive to negative values, thus finally becoming zero when averaged over many cycles. The ability of the phase-sensitive detector to reject signals such as random noise which differs in frequency from that of the reference signal, depends on the averaging time of the smoothing filter. The longer the averaging time, the greater is its rejection capability. Thus the p.s.d. is used to process signals that are 'buried' in noise and present output voltages with high signal-to-noise ratio.

For use with test signals of low to moderate frequencies, the circuit of a p.s.d. with an FET acting as a single-pole, double-throw switch is shown in Figure 3.55. The output response is limited by the capacitor C in the low-pass filters. The op-amp 2 provides the output voltage with respect to ground.

Most of the phase-sensitive detectors may be designed with the additional provision for shifting the phase of the reference signal until a maximum or a zero reading is obtained. Thus the circuit also provides the means for searching components of the test signal, having the frequency of the reference signal, but with a different phase angle. It is sufficient that the noise components do not constitute so large an amplitude in comparison to the component being searched, as to cause saturation of the amplifier at the input end.

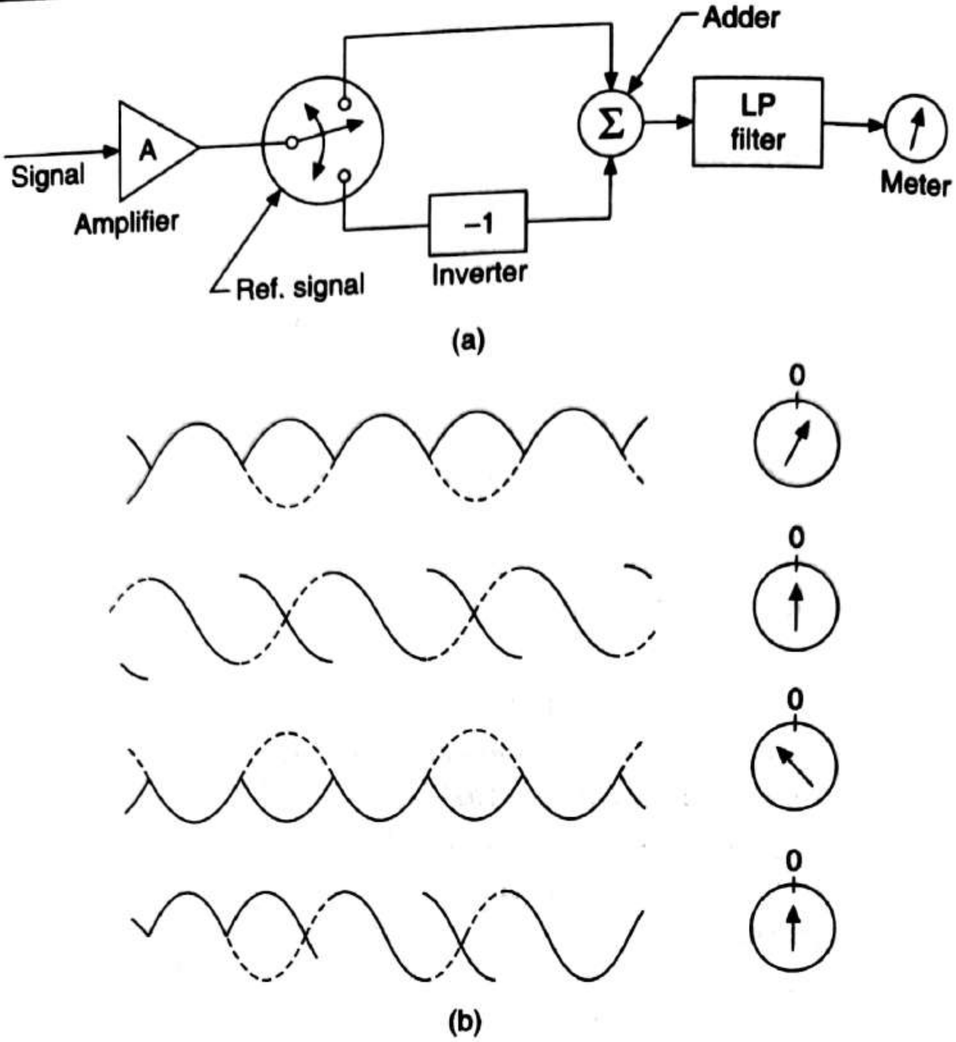


Figure 3.54 (a) Basic scheme of a phase-sensitive detector (p.s.d.); (b) principle of operation of p.s.d.

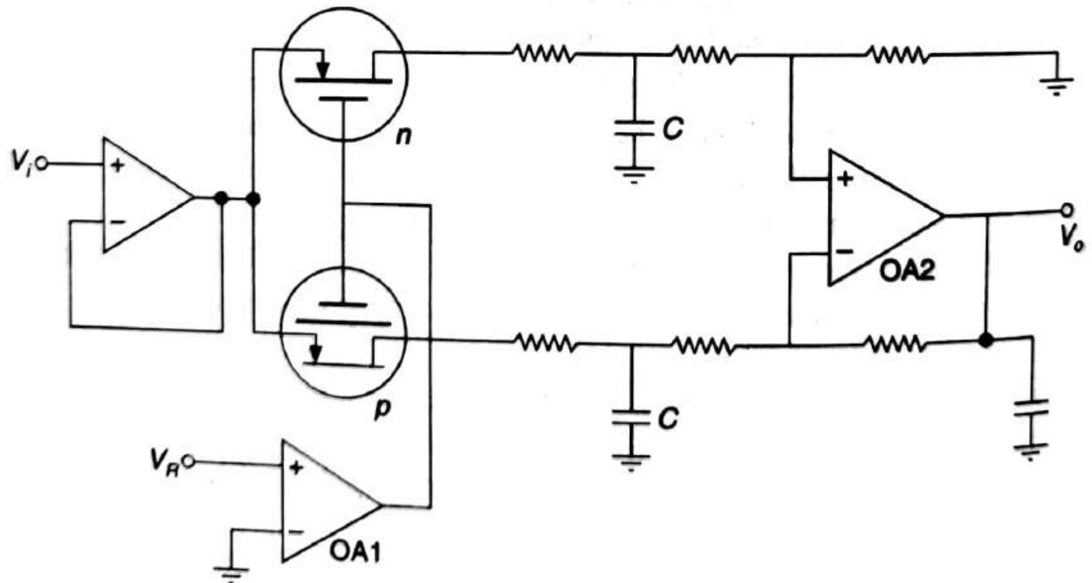


Figure 3.55 Circuit of a phase-sensitive detector.

The low-pass filter of the p.s.d. can be replaced by an integrator that integrates the output of the adder for a definite length of time. Such an arrangement has the advantage that the final output of the integrator is a definite value related to the amplitude and phase of the component of the test signal having the reference frequency, and a digital voltmeter can be used for indication.

9.2.4 Noise Problems, Shielding, and Grounding[15]

In the progression from single-amplifier op-amp circuits to instrumentation and finally isolation amplifiers, capability for successfully dealing with noisy environments significantly improves, at the expense of greater cost, complexity, and size. Clearly, you should use any available methods to minimize noise effects so that the simplest and cheapest type of amplifier, which provides satisfactory service can be utilized. Here, we cover only briefly some of the main considerations, leaving details to the listed references. We begin by stating that transducers of low impedance generally cause less severe noise problems and should be employed in preference to high-impedance devices, when possible.

Gradual changes in amplifier offset voltages and bias currents resulting from temperature, time, and line voltage sometimes are called drift, rather than noise, but we include them here for completeness. If the temperature and/or line voltage can be kept more nearly constant, errors caused by these disturbances are, of course, reduced. Drift performance of amplifiers varies greatly from model to model; thus a close check of specifications allows a proper choice. Remember that any junction of dissimilar metals in a circuit is an unintentional thermocouple capable of introducing temperature-related errors. [An *integrated-circuit* (IC) amplifier, its socket, and a pair of binding posts connected to one intentional thermocouple can have 18 *unintentional* thermojunctions in series.] If frequency response to dc is not required, use of ac coupling can greatly reduce all drift effects. In digital data systems, often *automatic zero* techniques are utilized to compensate drift. Here, since data are already intermittent because of the sampling required in digital systems, the input can be shorted periodically (made zero) and the resulting output (which must be the total drift error) temporarily put in memory. Then the next data reading can be corrected for drift by adding the memorized error.

Inductive pickup, electrostatic pickup, and ground loops can cause large error voltages, often concentrated at the power-line frequency. The changing magnetic field that surrounds any conductor carrying alternating current can link signal wires and produce large interfering voltages by inductive pickup. Because of the stray capacitance present between any adjacent conductors, a varying electric field can couple noise voltages to a signal circuit electrostatically. In solving (or preventing) such

---

[14]Model PA-138, Labworks Inc., Costa Mesa, CA, 714-549-1981 (www.labworks-inc.com).

[15]R. Morrison, "Grounding and Shielding Techniques in Instrumentation," Wiley, New York, 1977; D. H. Sheingold (ed.), "Transducer Interfacing Handbook," chap. 3, Analog Devices, 1980; "Elimination of Noise in Low-Level Circuits," Gould Inc., Cleveland, OH; "Noise Control in Strain Gage Measurements," *TN*- 501, Measurements Group, Raleigh, NC, 1980.

noise problems, start at the source by taking all measures possible to remove equipment such as power lines, motors, transformers, fluorescent lamps, relays, etc., from the near neighborhood of sensitive signal circuits, since the listed devices produce both inductive and electrostatic interference. If further measures are necessary, the use of an enclosing conductive shield (Fig. 9.17) can decrease electrostatic pickup. The shield functions by capturing charges that otherwise would reach the signal conductors. Once captured, these charges must be drained off to a satisfactory ground, or else they would be coupled to the signal conductors through the shield-to-cable capacitance.

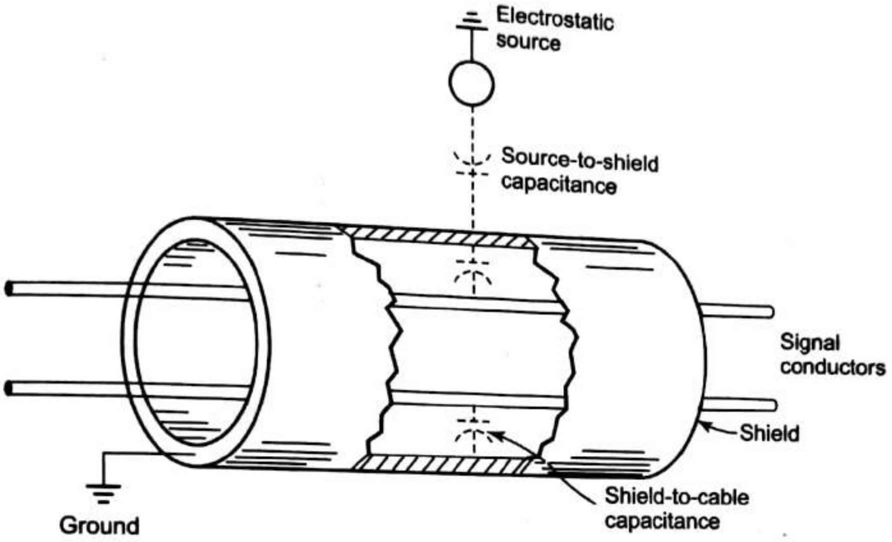


Fig. 9.17 *Electrostatic shielding.*

Shielding of the above type is *not* practical for magnetically induced noise, since magnetic shielding at low (60 Hz) frequencies requires thick (2.5 mm) shields of ferromagnetic metals. Rather, inductive noise is minimized by *twisting* the two signal conductors so that the "loop area" available for inducing error voltages is reduced and the mutual inductances between the noise source and each wire are balanced, to give a canceling effect. Commercially available cable provides twisted conductors, wrapped foil shields, and a grounding drain wire to meet the needs of both electrostatic and inductive noise reduction. Flexing of a cable can generate noise as a result of rubbing friction within the cable itself (triboelectric effect), and so cables should be properly secured. When cable flexing is unavoidable, a construction employing a conductive tape between the primary insulation and the metal-foil shield reduces triboelectric noise considerably.

A *ground loop* is created by connecting a signal circuit to more than one ground. If the multiple "ground" points were truly at identical potentials, no problem would arise. However, the conductor that serves as ground (piece of heavy wire, metal chassis of instrument, ground plane of printed-circuit board, etc.) generally carries intentional or unintentional currents and has some resistance; thus two points some distance apart will *not* have identical potentials. If this potential difference produces current flows through the shield and/or signal circuit, then large noise voltages may occur. In Fig. 9.18a, two ground loops, one through the shield and the other through a signal wire, are caused by improper grounding practices. Current in the signal wire directly causes error voltages (because of wire resistance), while shield current couples voltages into the signal circuit through shield-to-cable capacitance. In Fig. 9.18b, the shield ground loop is broken simply by grounding the shield at only one point (the signal source), while use of a floating-input (isolated) amplifier breaks the other loop, thus greatly reducing noise pickup. If the amplifier has a floating internal shield ("guard shield" or "guard") surrounding its input section, then the cable shield is still grounded only at the signal source; but the amplifier end of the cable is connected to this guard, effectively extending it to the signal source. A final technique for noise control is filtering (see Sec. 9.3), but usually this can be utilized only if signal and noise occupy different frequency ranges.

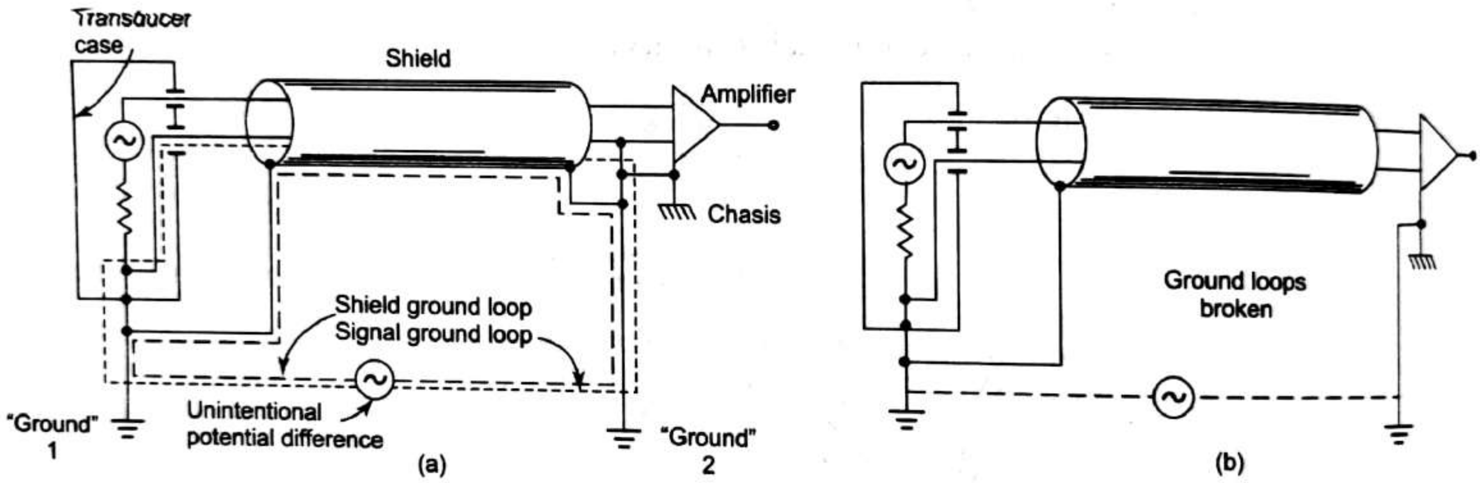


Fig. 9.18 *Ground-loop problems.*

Module IV
Micro Electromechanical System

MEMS = MicroElectroMechanical System

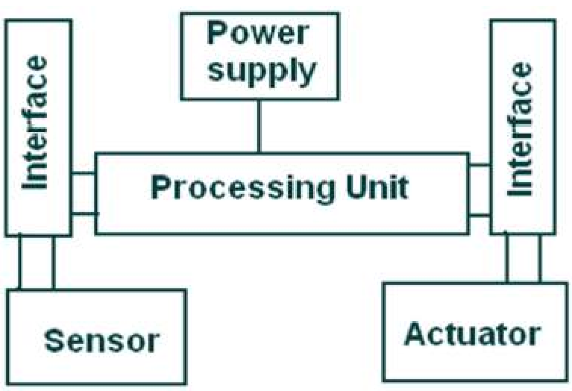
- Micro establishes a "dimensional scale"
- Electro suggests either "electricity or electronics (or both)
- Mechanical suggests "moving parts of some kind"

- Any engineering system that performs electrical and mechanical functions with components in micrometres is a MEMS. (1 μm = 1/10 of human hair)
- MEMS = a pioneer technology for Miniaturization
- Micro-Electro-Mechanical Systems (MEMS) is the integration of mechanical elements, sensors, actuators, and electronics on a common silicon substrate through micro fabrication technology.
- A technology to create any integrated devices or systems that combine mechanical and electrical components. These devices have the ability to sense, control and actuate on a micro scale and generate effects on macro scale.
- A system or a device that has static and movable components with some dimension on the scale of micron.
- By combining IC's with mechanical parts, MEMS are complete systems on a chip.
- MEMS is miniaturing of Electrical and Mechanical Systems.
- Electromechanical Devices on the Micron to Millimeter scale known as Micro Mechanical Systems (MEMS
- MEMS Passive devices: Transmission lines, filters, couplers
- Active devices: Switches, tuners and variable capacitors.

<u>Segments (Parts/Blocks) / Major constituents in a typical MEMS</u>

Micro Sensors + Micro Processors + Micro Actuators = MEMS

- Micro Sensors: Senses the change experienced by the system like change in pressure, change in velocity,
- acceleration etc.
- Micro Signal Conditioning and Processing Unit: Processes and conditions the input and gives a suitable output involves microelectronic circuits.
- Micro Actuator: Senses the signal output from microprocessor and gives a mechanical output
- Interface: Interfaces at the input and output
- Power supply: To energize the system.



Why MEMS is useful?

- Micro Electro Mechanical Systems are physically small; this is the reason why MEMS is useful.
- MEMS sensors interfere less with the environment and they sense the signals better than larger devices.
- The motion of the MEMS actuators is very precise.
- MEMS devices can also be placed in small spaces such as inside automobile engines, small appliances, and living organisms to measure the various parameters.
- MEMS devices offer tight integration of electronics with mechanical components, which results in very in very low power consumption and fast response times.
- Cost Savings: MEMS devices are fabricated using micro-machining process. Therefore, the efficiencies and economies of scale that the semiconductor industry has enjoyed is leveraged to fabricate MEMS devices as well. In addition, MEMS devices do not require any assembly; all components are manufactured in their assembled positions.
- Reliability and Redundancy: MEMS devices are less susceptible to the wear and tear that is experienced by macro devices. MEMS devices are manufactured in large numbers on a single silicon wafer. Several MEMS components can be used to provide redundancy for critical applications. Therefore, when a single MEMS device fails, the overall performance of the system may only degrade slightly, instead of experiencing a complete failure.
- Performance: MEMS devices offer much better performance: energy consumption ratios. Precision control/actuation is possible with MEMS devices.

Comparison of Microelectronics and Microsystems

- Microsystems are “very small systems” or “systems made of very small components”.
- MEMS involve both electronic and non-electronic components and perform functions that can include signal acquisition (sensing), signal processing, actuation, display and control. MEMS involve large arrays of micro fabricated elements
- MEMS also have important system issues such as packaging, system partitioning into components, calibration, signal to noise ratio, stability and reliability.

Microelectronics	Microsystems (silicon based)
Primarily 2-dimensional structures	Complex 3-dimensional structure
Stationary structures	May involve moving components
Transmit electricity for specific electrical functions	Perform a great variety of specific biological, chemical, electromechanical and optical functions
Use single crystal silicon dies, silicon compounds, ceramics and plastic materials	Use single crystal silicon dies and few other materials, e.g. GaAs, quartz, polymers, ceramics and metals
Fewer components to be assembled	Many more components to be assembled
Mature IC design methodologies	Lack of engineering design methodology and standards
Complex patterns with high density of electrical circuitry over substrates	Simpler patterns over substrates with simpler electrical circuitry
Large number of electrical feed-through and leads	Fewer electrical feed-through and leads
Industrial standards available	No industrial standard to follow in design, material selections, fabrication processes and packaging
Mass production	Batch production, or on customer-need basis
Fabrication techniques are proven and well documented	Many microfabrication techniques are used for production, but with no standard procedures
Manufacturing techniques are proven and well documented	Distinct manufacturing techniques
Packaging technology is relatively well established	Packaging technology is at the infant stage
Primarily involves electrical and chemical engineering	Involves all disciplines of science and engineering

Available MEMS products include:

- Micro sensors (acoustic wave, biomedical, chemical, inertia, optical, pressure, radiation, thermal, etc.)
- Micro actuators (valves, pumps and microfluidics; electrical and optical relays and switches; grippers, tweezers and tongs; linear and rotary motors, etc.)
- Read/write heads in computer storage systems.
- Inkjet printer heads.
- Micro device components (e.g., palm-top reconnaissance aircrafts, mini robots and toys, micro surgical and mobile telecom equipment, etc.)

Applications of MEMS

a. Safety:

- Air bag deployment system
- Antilock braking systems (position sensor)
- Suspension systems (displacement, position and pressure sensors, microvalves.)
- Object avoidance (pressure and displacement sensor)
- Navigation (micro-gyroscope)

b. Sensors used in engine and power train

- MAP- Manifold control with pressure sensor: This measures the engine speed in rpm to determine the ignition advance. This ignition timing, with optimum air/fuel ratio can optimize the power performance of the engine with low emission.

c. Sensors used in Engines

- Air flow control
- Exhaust gas analysis and control
- Crankshaft positioning
- Fuel pump pressure and fuel injection control
- Pressure control

d. Comfort and convenience

- Seat control
- Security - (remote status monitoring and access control sensors)
- Sensor for Defogging of wind shield
- Satellite navigation sensors
- Rider’s comfort- sensors for air quality, air flow, temperature and humidity control

e. Vehicle diagnostics

- Engine coolant temperature and quality
- Engine oil pressure, level, quality.
- Tire pressure
- Brake oil pressure
- Fuel pressure
- Transmission fluid

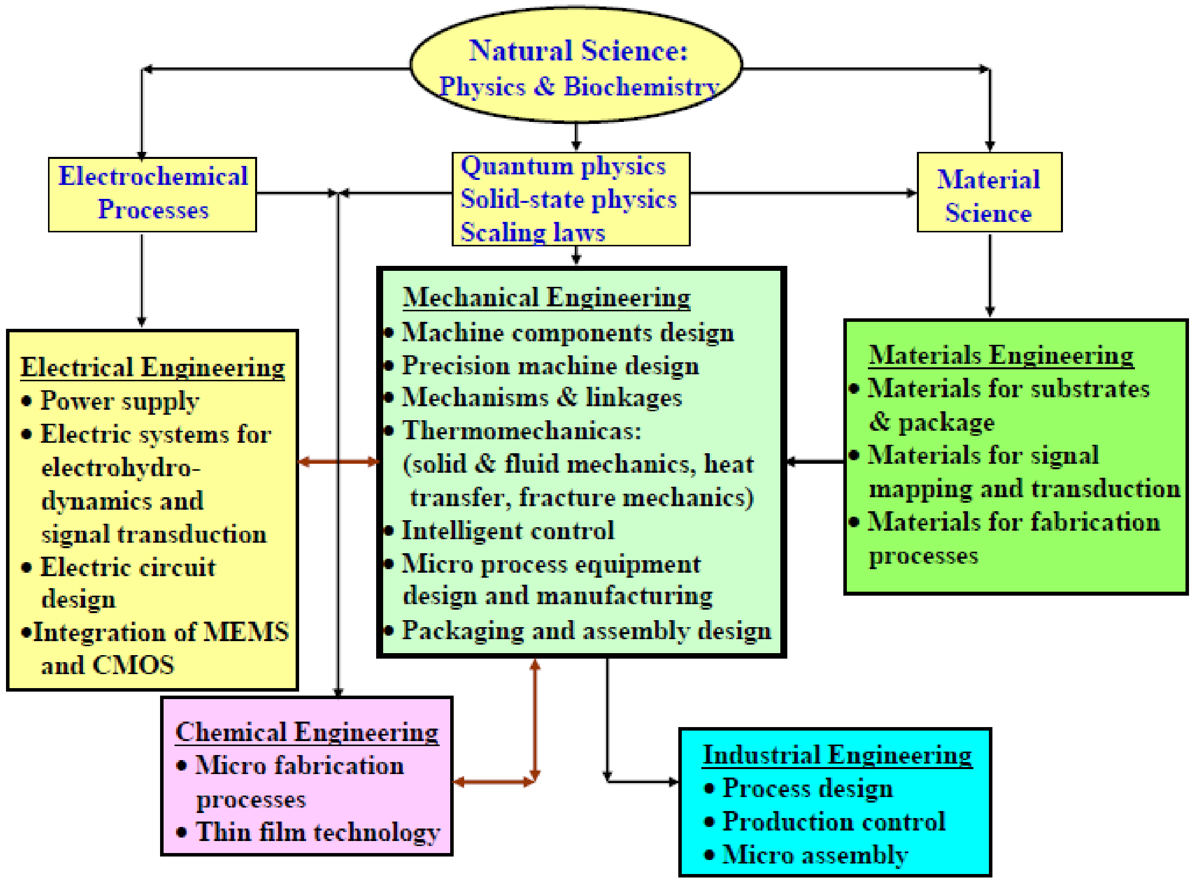
f. Application of MEMS in bio medical industry

- Disposable blood pressure transducer(DPT)
- Angioplasty pressure sensor: To monitor the pressure inside the balloon once it is inside the blood vessel.
- Catheter tip pressure sensor
- Sphygmomanometers
- Respirators and Lung capacity meters
- Barometric correction instrumentation
- Medical process monitoring (drug production by growth of bacteria)
- Kidney dialysis equipment

g. Applications of MEMS in automobiles

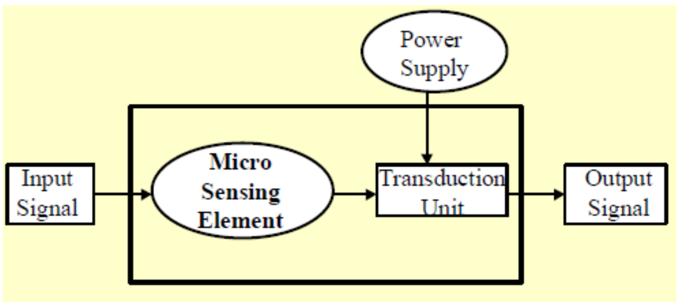
- To make automobiles safer and more comfortable for the riders and to meet the high fuel efficiency and low emission standards required by the government.
- Make the automobiles smarter for the consumers' needs.
- Smart means with various sensors
- Detect the environment / road conditions
- Micro-sensors and actuators allow the makers to use smaller devices in effective ways
    - Temperature manifold absolute pressure sensor
    - Exhaust gas differential pressure sensor
    - Fuel rail pressure sensor
    - Barometric pressure sensor
    - Combustion sensor
    - Gasoline direct injection pressure sensor
    - Fuel tank evaporative fuel pressure sensor
    - Engine oil sensor
    - Transmission sensor
    - Tire pressure sensor.

Automotive	Electronics	Medical	Communications	Defence
Internal navigation sensors	Disk drive heads	Blood pressure sensor	Fibre-optic network components	Munitions guidance
Air conditioning compressor sensor	Inkjet printer heads	Muscle stimulators & drug delivery systems	RF Relays, switches and filters	Surveillance
Brake force sensors & suspension control accelerometers	Projection screen televisions	Implanted pressure sensors	Projection displays in portable communications devices and instrumentation	Arming systems
Fuel level and vapour pressure sensors	Earthquake sensors	Prosthetics	Voltage controlled oscillators (VCOs)	Embedded sensors
Airbag sensors	Avionics pressure sensors	Miniature analytical instruments	Splitters and couplers	Data storage
"Intelligent" tyres	Mass data storage systems	Pacemakers	Tuneable lasers	Aircraft control



MEMS as Microsensors

Sensor is a device converts one form of energy into another form of energy and provides the user with a usable energy output in response to a specific measurable input.



Basic components of microsensors are microsensing element, transduction element and power supply (if needed). Microsensing element senses the physical quantity of interest. It can be pressure, force, temperature or any other quantity. Transduction element converts the sensed physical quantity into equivalent electrical quantity, usually current or voltage. In some sensors, external power supply may be needed for the sensing element as well as the transduction element.

Example: Pressure Sensor- The input pressure is sensed by silicon diaphragm- Micro-sensing device. The deflection of diaphragm is converted into resistance by piezo-resistor. Change in resistance is converted into change in voltage- Transduction element.

Microsensors are expected to detect a variety of signals associated with:

- Accelerations (velocity and forces),
- Biological and biomedical
- Chemical composition,
- Forces (e.g., micro-accelerometers and gyroscopes)
- Optical,
- Pressure,
- Thermal (temperatures)
- Magnetic flux

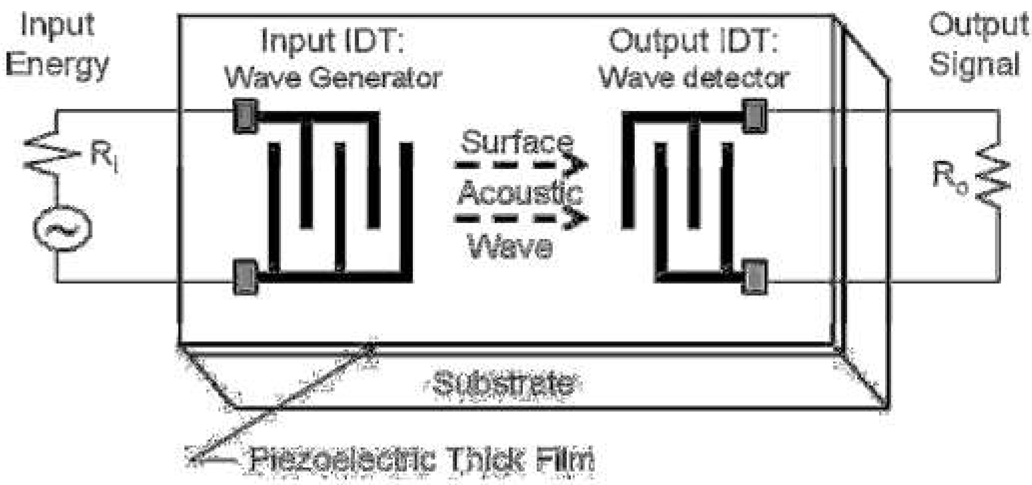
Input samples may be: motion of a solid, pressurized liquids or gases, biological and chemical substances.

Working Principles for Microsensors

Micro Sensors: Senses the change experienced by the system like change in pressure, change in velocity, acceleration etc. It also senses the existence of chemical, physical or biological quantities such as pressure, force, sound light, nuclear radiation, magnetic flux and chemical composition.

A smart sensor unit includes

- Automatic calibration,
- Interference signal reduction,
- Compensation for parasitic effects
- Offset correction and self-test
- Acoustic Wave Sensors



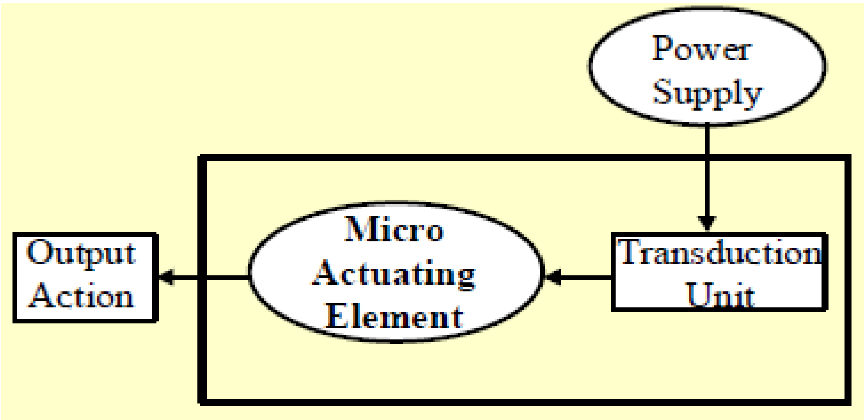
Example of MEMS acoustic wave sensor

Other types of microsensors are:

- Accelerometer
  - Measures acceleration
- Gyroscope
  - Measures angular rate
- Pressure sensor
  - Measures pressure of a fluid
- Viscosity meter
  - Measures viscosity of a fluid
- Anemometer
  - Measures wind speed
- Bolometer
  - Measures radiation
- Blood analyser
  - Measures the presence or quantity of a chemical species
- Virus detector
  - Detects the presence of a virus
- Etc.

Microactuators

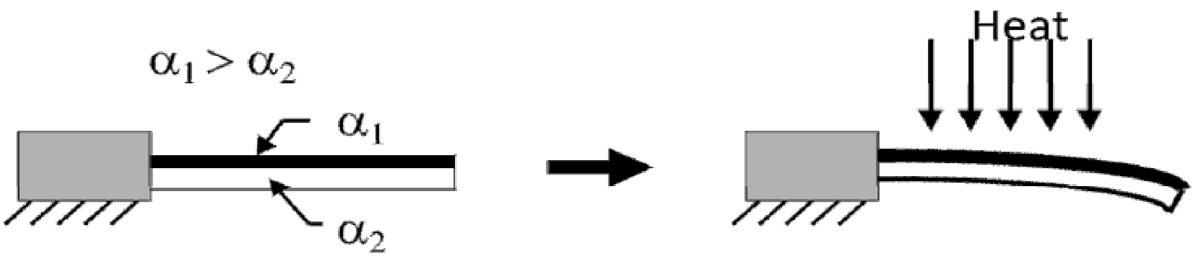
Microactuators are the micro devices which converts the electrical signal to a physical quantity to produce an actuation.



- Power supply: Electrical current or voltage
- Transduction unit: To covert the appropriate form of power supply into the desired form of actions of the actuating element
- Actuating element: A material or component that moves with power supply
- Output action: Usually in a prescribed motion
- Due to the minute sizes, microactuators work on radically different principles than the conventional electromagnetic means, such as solenoids and ac/dc motors.
- Instead, electrostatic, thermal, piezoelectric and shape-memory alloys are extensively used in microactuations.

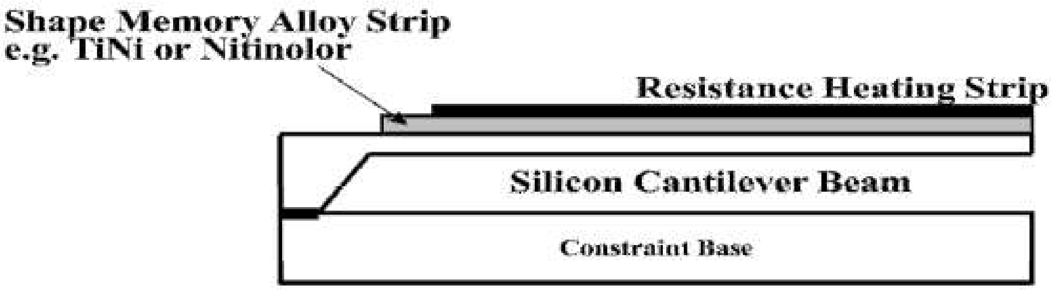
Actuation Using Thermal Forces

- Solids deform when they are subjected to a temperature change (ΔT)
- A solid rod with a length L will extend its length by $\Delta L = \alpha \Delta T$, in which α = coefficient of thermal expansion (CTE) – a material property.
- These compound beams are commonly used as micro switches and relays in MEMS products



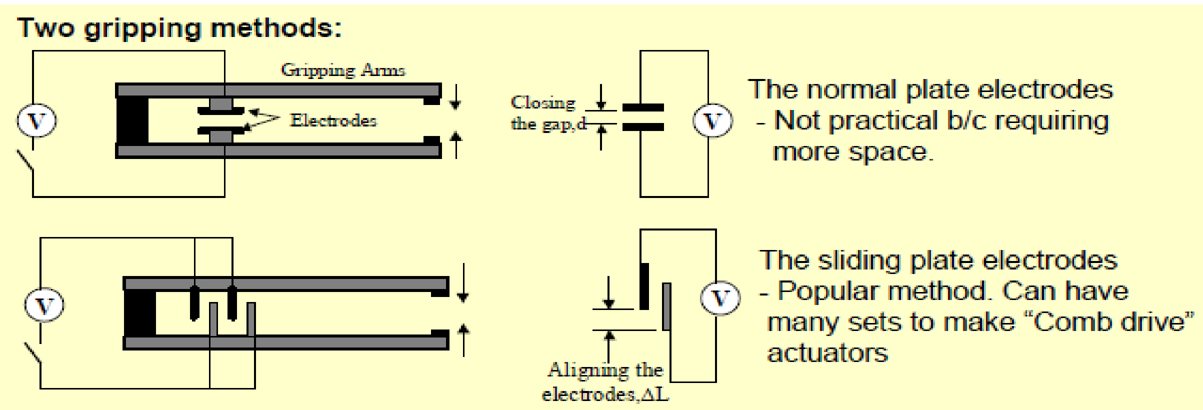
Actuation Using Shape Memory Alloys (SMA)

- SMA are the materials that have a "memory" of their original geometry (shape) at a typically elevated temperature of production.
- These alloys are deformed into different geometry at typically room temperature.
- The deformed SMA structures will return to their original shapes when they are heated to the elevated temperature at their productions.
- Ti-Ni is a common SMA

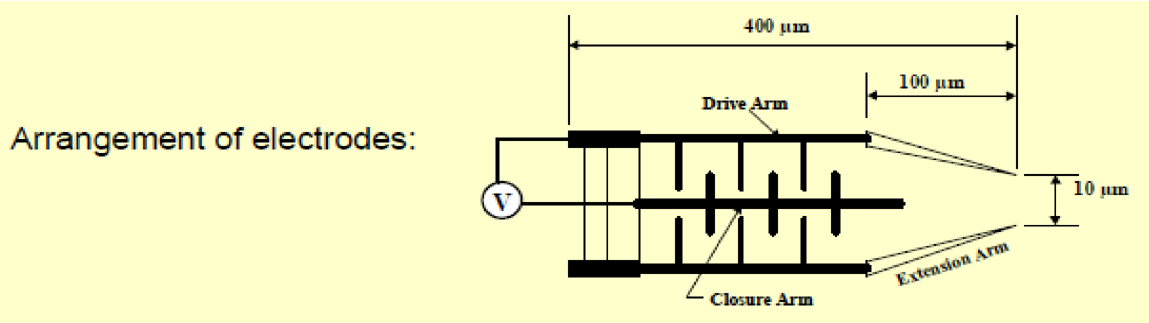


Actuation Using Electrostatic Forces:

Consists of electrodes separated by a dielectric material. The input voltage to the plates can result in electrostatic forces that prompt relative motion of the plates. These motions are set to required actions- E.g.: Microgrippers.



- Arrangement of the fingers are like comb. Two sets of comb architectures are arranged in such a way that the fingers are inter-digital fashion.
- While we apply the external force, the combs come closer and thereby the arms connected to the combs come closer and work like a gripper.
- Drastic reduction in required actuation voltage with increase of number of pairs of electrodes:



In this given figure above, a comb architecture with flexible beam is placed with another fixed comb mechanism. While apply the external force the flexible comb mechanism moves in a horizontal direction and thereby reducing the gap between inter digital electrodes This structure is called transverse comb drive architecture.

Other types of microactuators are Micro motors, micro-valves, micro-pumps, Micro Heat Pipes, Actuation Using Piezoelectric Crystals etc.

Microfabrication Technologies

- There is no machine tool with today's technology can produce any device or MEMS component of the size in the micrometer scale (or in mm sizes).
- The complex geometry of these minute MEMS components can only be produced by various physical-chemical processes – the microfabrication techniques originally developed for producing integrated circuit (IC) components.
- For MEMS and microsystems components, the sizes are so small that no machine tools, e.g. lathe, milling machine, drilling press, etc. can do the job.
- There is simply no way one can even grip the work piece.
- Consequently, radically different techniques, non-machine-tool techniques need to be used for such purpose.
- Most physical-chemical processes developed for "shaping" and fabricating ICs are adopted for microsystems fabrications. This is the principal reason for using silicon and silicon compounds for most MEMS and microsystems – because these are the materials used to produce ICs.
- Two techniques: Bulk micromanufacturing and Surface micromachining

Overview of Micromanufacturing.

MEMS is the acronym for micro electromechanical systems. MEMS uses a lot of microstructures like cantilevers, membranes, nozzles, microsensors, actuators etc. In mechanical engg these structures are usually made using certain machines like lathe machine, grinding, polishing etc. But in MEMS we need structures of micrometer to range. Hence we use materials such as silicon, quartz and techniques which are compatible with VLSI processes for making microstructures in MEMS. That is why it is called micromachining / micromanufacturing.

There are 3 distinct micromachining techniques used by current industry

(1) Bulk micromanufacturing
(2) Surface micromachining
(3) LIGA processes.

Bulk Micromanufacturing

→ widely used in the production of microsensors and accelerometers.

→ Bulk micromanufacturing / micromachining involves the removal of materials from the bulk substrates, usually silicon wafers to form the desired three dimensional geometry of the microstructures (Eg beams, membranes)

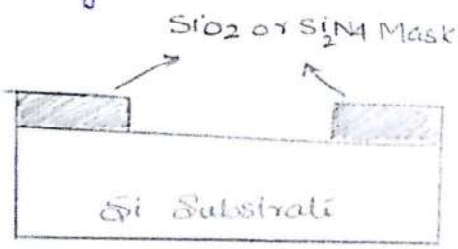
→ Shaping of microsystem components of size 0.1um to 1mm are done by etching (dry / wet etching)

→ wet etching uses liquid chemical solution while dry etching process use plasma (high energy gas containing ionized radicals) or vapour phase etchants to remove materials.

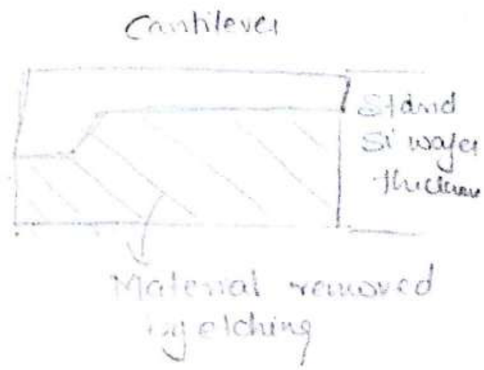
Isotropic and Anisotropic Etching:

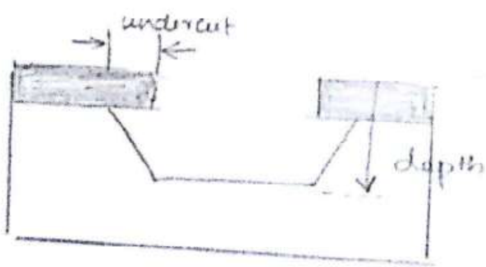
Isotropic etching : The etch rate in all directions are identical (lateral and Vertical etch rate are equal). i.e etch rate is independent of direction.

Anisotropic Etching: here the etch rate is orientation dependent (NO etching in the lateral direction)

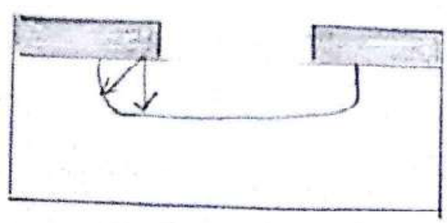


Ideal Etching





Anisotropic etching (Under cut)



Isotropic etching

Wet Etchants :

→ There are a number of different types of etchants that can be used to etch different substrate materials.

→ The common isotropic etchant for Si is called HNA which stands for $HF/HNO_3/CH_3COOH$ (Hydrofluoric / Nitrite / acetic acid. (acidic agents)

→ Alkaline chemicals with PH > 12 are used for anisotropic etching

→ Eg of anisotropic etchants for Si are potassium hydroxide (KOH), ethylene-diamine and pyrocatecol (EDP), tetramethyl ammonium hydroxide (TMAH) and hydrazine.

→ The selectivity ratio of a material is defined as the ratio of the etching rate of Silicon to the etching rate of another material using the same etchant.

Eg: SiO_2 has ~~etch~~ selectivity ratio of 10^3 in KOH, which means that the etching rate of SiO_2 in KOH is 10^3 times slower than that of Si.

Thus higher the selectivity ratio of the material, the better it is for masking material.

→ The timing of etching and the agitated flow patterns of the etchants over the substrate surface need to be carefully controlled in order to avoid serious under etching and undercutting.

→ One must also take special caution in selecting the masking materials.

A common practice is to use SiO_2 layer as mask for Silicon substrate in KOH etchants for trenches of modest depth.

SiO_2 masking is relatively inexpensive in an etching process. but SiO_2 mask itself can be attacked by the etchants if the system is left in the etchant for a long period of time. as in case of deep etching. Hence Silicon nitride should be used for deep etching.

An efficient way to control the shape of the etched Silicon substrate is to apply etch stop.
[In many cases it is necessary to stop etching when a certain cavity depth or certain membrane thickness is reached.

Etch Stop

There are 2 popular technique used in etch stop.

(1) Dopant - Controlled etch stop

(2) Electrochemical etch stop.

Dopant - Controlled Etch stop / Boron Etch stop

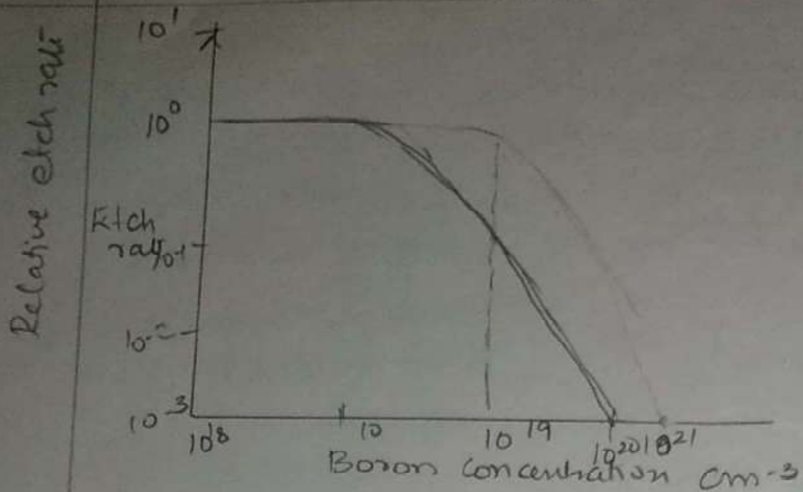
→ It is the most widely used etch stop technique

→ Based on the fact that anisotropic etchants such as EDP do not attack heavily doped boron doped (p++) Si layers.

→ Selective p++ doping is implemented using gaseous or solid boron diffusion source with a mask (such as SiO_2).

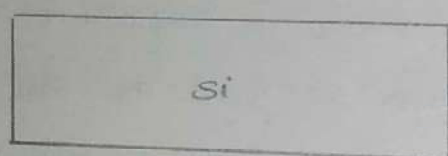
→ Silicon membranes are usually fabricated using boron etch-stop technique.

Scanned by CamScanner



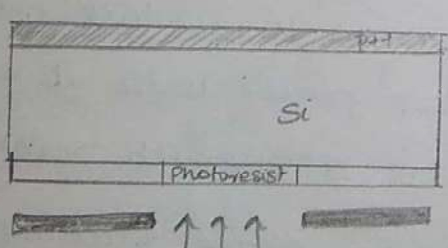
for EDP etchants

→ A simple boron diffusion or implantation is introduced from the front of the wafer and etching is done from back to create beams and diaphragms.

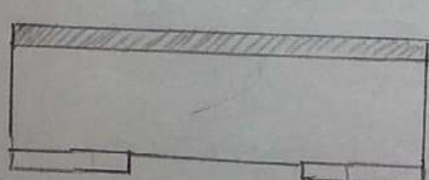


(1) Si wafer

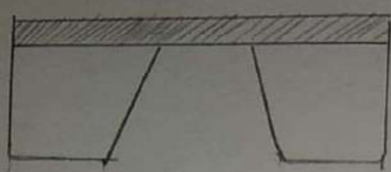
(2) boron doping creates p+ type film on surface



(3) Back patterning - A window is opened at back with lithography

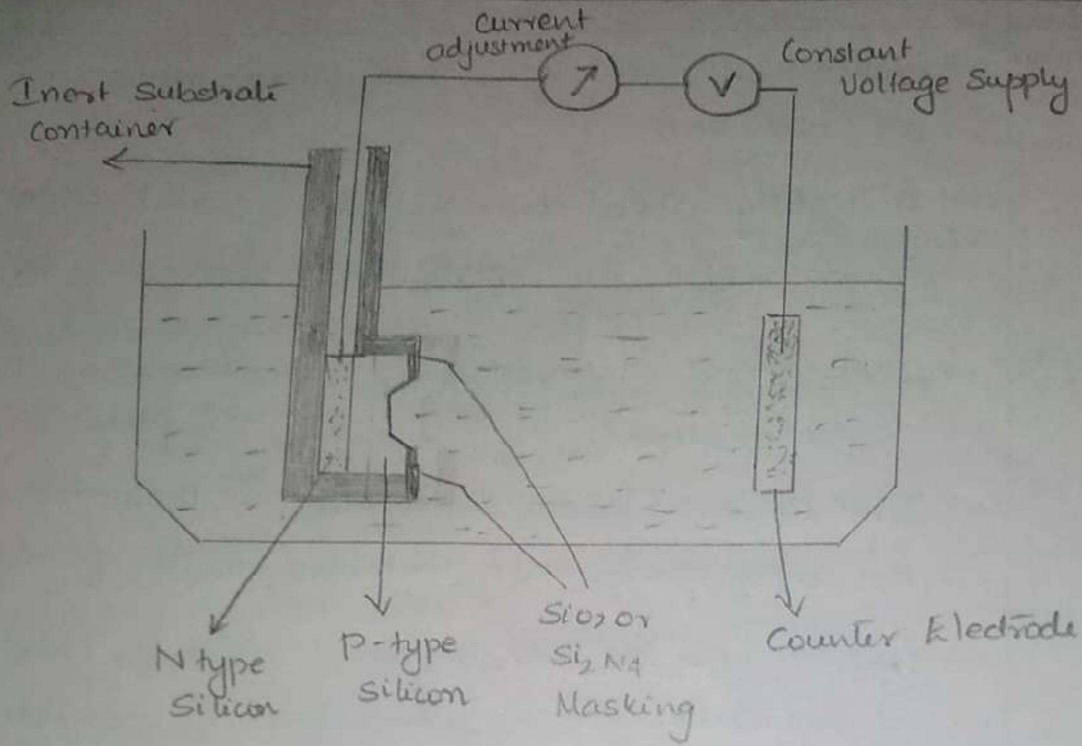


(4) Etch the wafer



(5) Etching stoped at the
p-type layer
Creating a membrane
(beam membrane).
that is used in MEMS

Electro Chemical Etch Stop

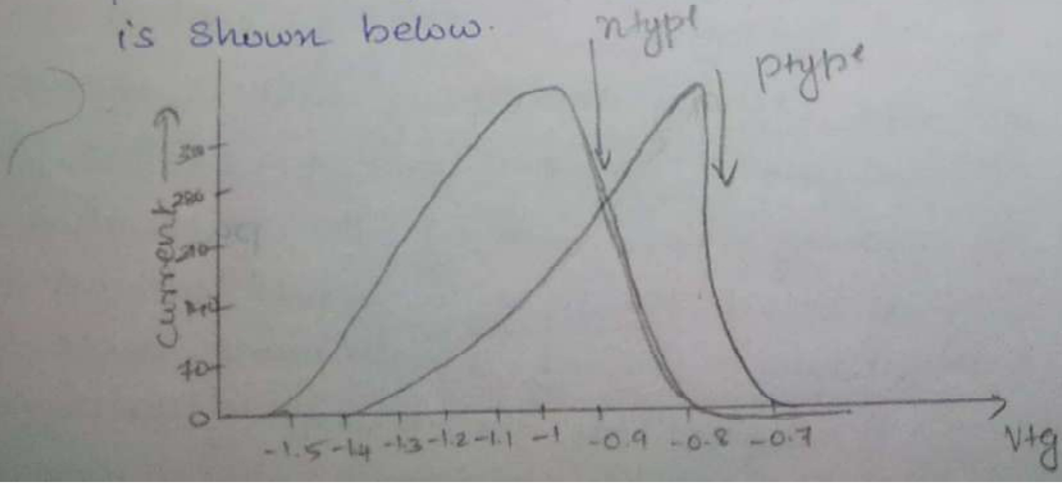


Electrochemical etch stop

→ The etch rate of Si in Electrochemical etching depends on the applied potential.

→ Etching stops if the applied potential is greater than the passivation potential. (is the potential after which a thin layer of SiO_2 is formed which passivates the anode and thus prevents etching)

→ passivation vtg for n type & p type si in KOH is shown below.



→ With KOH etching, this phenomena is used as etch stop.

→ A lightly doped p-n junction is first produced in Silicon wafer by diffusion process.

→ The ntype is phosphorous doped while the p-type is boron doped.

→ The doped silicon substrate is then mounted on an inert substrate container made of sapphire.

→ The n-type silicon layer is used as one electrode of the electrolyte system and is connected to one terminal of a vtg source, while the other terminal of the vtg source is connected via a current meter to a counter electrode in the etching solution.

→ The n layer is biased at its passivation potential.

→ Since the p-type substrate is not biased, the potential on it will float (open ckt potential) (unmasked p-region) which means that it will etch as normal in the etching solution.

→ The etching takes place until it reaches the interface of the p-n junction, at this point etching stops, leaving (the passivation potential at which the n-layer is held prevents further etching) leaving a membrane whose thickness is defined by the thickness of the n-region.

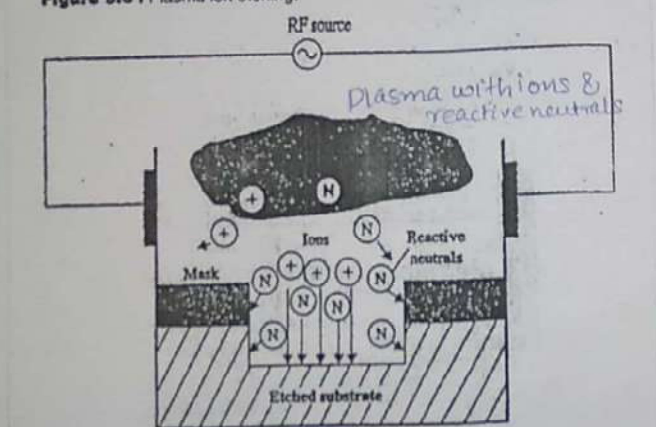
F + Si ⟶ SiF_4 (gas) vapourised)

neutral species will react with the material to be etched Eg Si producing byproducts which are volatile. (Easily vapourized or evaporated into gas phase)

→ The high energy reaction causes local evaporation and thus removal of the substrate material.

→ Etching moves more rapidly in depth direction than in the direction of the sidewalls. This is due to larger number of high energy particles involving both neutrals and charged ions bombarding the normal surface, while the sidewalls are bombarded only by neutrals.

Figure 9.6 | Plasma ion etching.



Advantage of plasma Etching

1. Conventional dry etching is very slow process at a rate of 0.1 µm/min or 100 Å/min while plasma etching etch rate is 2000 Å/min.

2. plasma etching is faster and cleaner than wet etching since it is done in high vaccum.

3.

Dry Etching

Involves the removal of substrate materials by gaseous etchants without the use of wet chemicals.

Different types
(1) plasma etching
(2) deep reactive ion etching (DRIE)

Plasma Etching

→ plasma is a neutral ionized gas carrying a large number of free electrons and positively charged ions.

→ A common source of generating plasma is a radio frequency (RF) source.

→ The process involves adding a chemically reactive gas such as CCl_2F_2 to the plasma (one which contains ions and has its own carrier gas (inert gas such as argon gas).

→ The reactive gas produces reactive neutrals when it is ionized in the plasma.

→ The reactive neutrals bombard the target on both the sidewalls as well as the normal surface, where as the charged ions bombard only the normal surface of the substrate

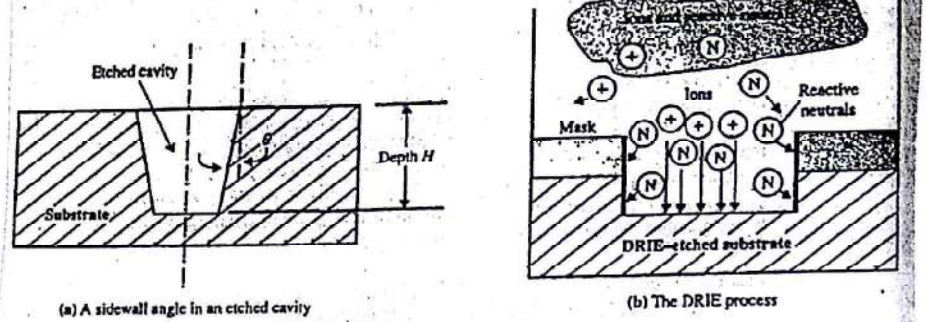
→ Etching of the substrate materials is accomplished by the high energy ions in the plasma bombarding the substrate surface as well as the chemical reaction b/w the substrate & free radical (neutrals)

Eg: neutral fluoride CF_3 is created in plasma.

$e^- + CF_4 \longrightarrow CF_3 + F + e^-$

Deep Reactive Ion Etching (DRIE)

Figure 9.7 | The deep reactive ion etching process.



(a) A sidewall angle in an etched cavity

(b) The DRIE process

→ Although plasma etching produces a significant increase in etching rate and increase in depth of the etched cavity (trench), the etched walls in the trenches remain at a wide angle (θ) to its depth.

→ The cavity angle θ is critical in many MEMS structures such as comb electrodes in microgrippers. These structures require the faces of electrodes or fingers to be parallel to each other. It is highly desirable that the angle θ is kept at a minimum in deep treches etched treches that separate the parallel electrodes.

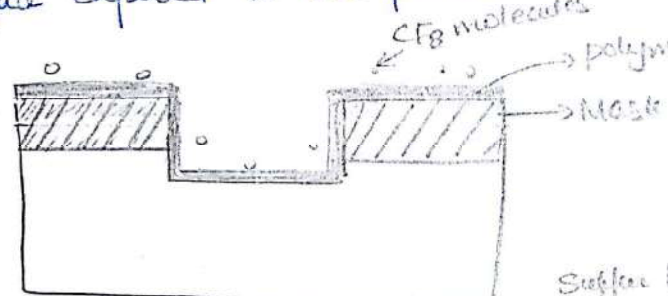
→ Deep reactive ion etching (DRIE) is a process that can overcome these problems. It is used to produce MEMS components with high aspect ratio (defined as the ratio of dimension in depth of MEMS components to those on surface) with virtually vertical side walls i.e $\theta = 0$.

→ DRIE process differs from dry plasma etching in that it produces thin protective films of a few micrometers on the side walls during the etching processes.

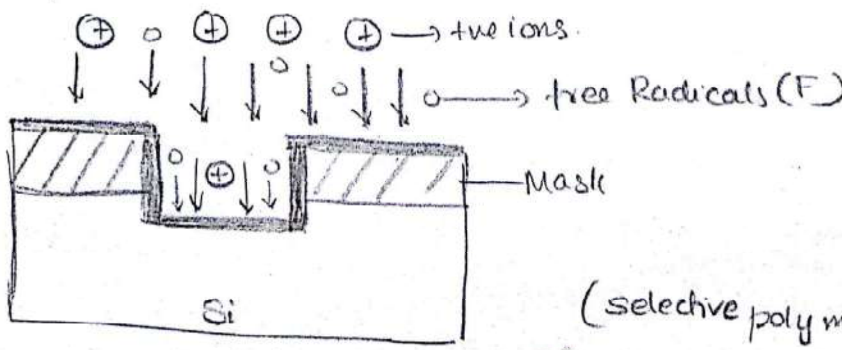
→ It involves the use of high density plasma source which allows alternating processes of plasma ion etching of the substrate material and deposition of etching protective material on the sidewall.

→ DRIE – Bosh process.

1. C_4F_8 based plasma is used to deposit a few um thickness layer of florocarbon polymer across all the surface exposed to the plasma.

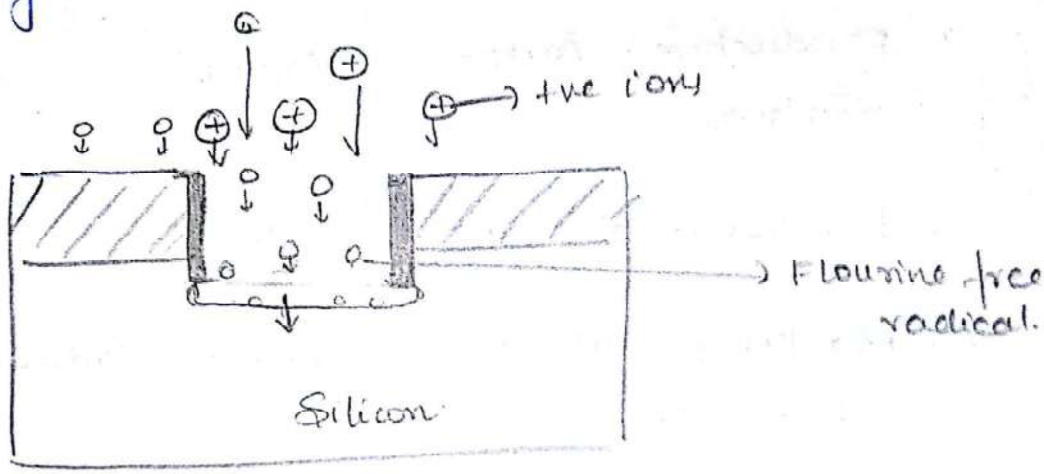


→ The plasma gas is switched to Sulfur Fluoride SF_6 to create ions that isotropically etches si. Through the application of dc bias to the plate, ions from plasma bombard the surface of the wafer removing the polymer. ↑sed ion energy in the vertical direction results in a much higher rate of removal of flurocarbon from surface parallel to the wafer surface



(selective polymer Removal)

3. Following the selective polymer removal, the Silicon surface at the base of the trench is exposed to reactive fluorine- based species that etch the unprotective Silicon. The remaining flaoro carbon polymer protects the vertical walls of the trench from etching.

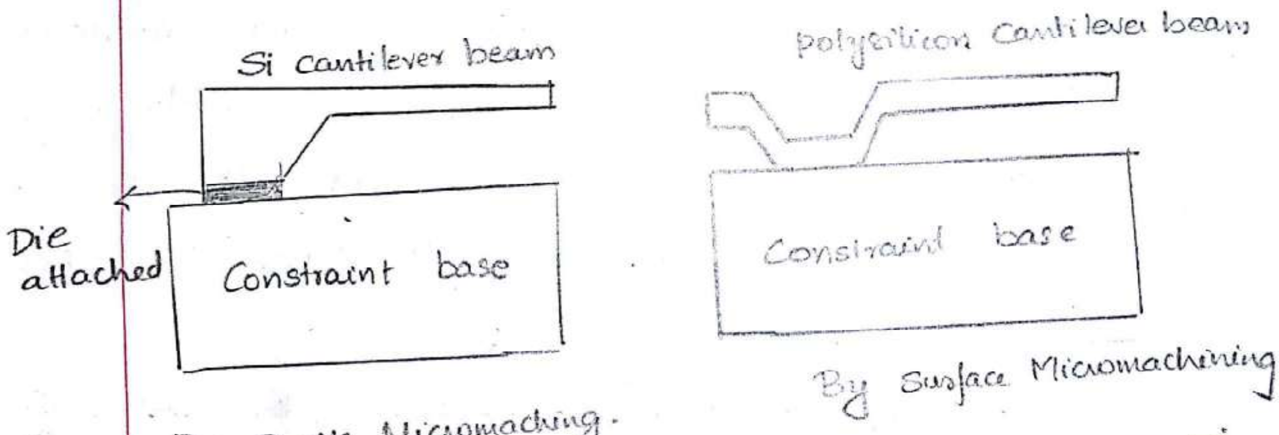


Surface Micromachining

→ In bulk micromanufacturing, the substrate material is removed by physical or chemical means. where as in Surface micromachining technique builds micro structures by adding materials layer by layer on top of the substrate

→ Deposition tech used is Low pressure chemical Vapour deposition (LPCVD) technique.

→ polycrystalline Silicon (polysilicon) is common material for the layer material
Sacrificial layer is usually made of SiO_2
(Sacrificial layer is used for creating void space)
wet etching is common method used for this purpose.



By Bulk Micromaching.

By Surface Micromachining

→ The substrate material is single crystal silicon in most cases while the added layers need not be single crystal.

→ Fig shows a cantilever beam made by bulk micro machu

& surface micromachining, It is seen that in bulk micromachining significant amount of material is etched away. while the same cantilever beam produced by polysilicon with surface micromachining tech not only saves material, but also eliminates the need for a die attach since polysilicon beam can be built on the top of the constraint base directly.

Process

Surface micromachined devices are made of 3 types of components

(1) A Sacrificial component (also called spacer layer)

(2) A Microstructural component

(3) An Insulator component.

The process step in manufacturing a microcantilever beam using surface machining is as follows.

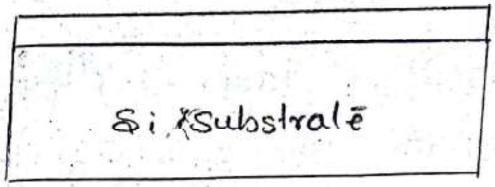
Step 1 : Deposition of Sacrificial layer
We begin with a Silicon substrate base with PSG (phosphosilicate glass) deposited on its surface.

Step 2 : ~~A mask 1 is made to cover the surface of PSG) for subsequent etching to in order to make an attachment for microcantilever.~~

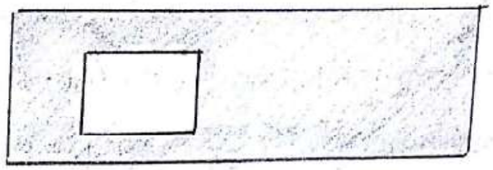
Step 3 : Deposit photoresist, place a mask 1, Expose to UV light, develop to create a window for etching. After etching an attachment is made for microcantilever

Sacrificial layer

Step 1



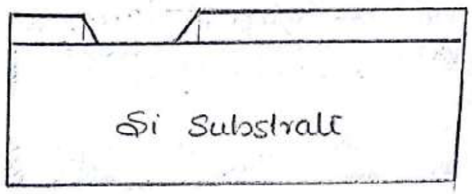
→ PSG Sacrificial layer

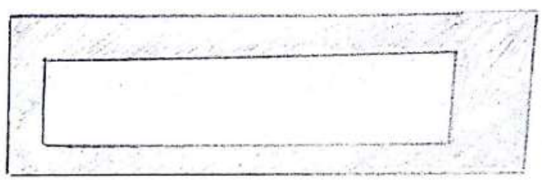


Mask 1

for Etching.

Step 2
After patterning of Sacrificial layer

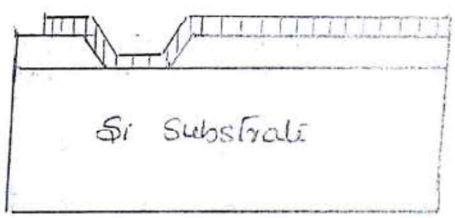




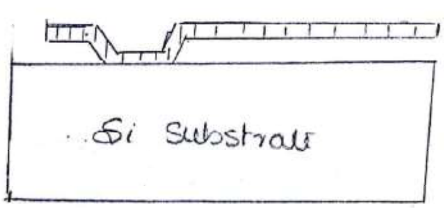
Mask-2

for deposition

Step3
After dep of Structural layer



Step4



After etching sacrificial layer

Scanned by CamScanner

process:

Step1 : A sacrificial layer is deposited on a Silicon Substrate. The sacrificial layer is phosphosilicate glass (PSG)

Step2 : with mask1, a window is patterned (photoresist + optical lithography) and etching of the sacrificial layer is done.

Step3 : Using mask 2, a layer of polysilicon Structural layer is deposited.

Step4 : The PSG (Sacrificial layer) is etched away Using etchants 1:1 HF which is made of 1:1 HF : H_2O + 1:1 HCl : H_2O. to produce the desired cantilever beam.

Step 5 : After etching, the structure is rinsed in deionized water thoroughly followed by drying under infrared lamps.

Mechanical problems associated with Surface Micromachining

There are 3 major problems of mechanical nature that result from surface micromachining

(1) Adhersion of Layers

(2) Interfacial Stresses

(3) Stiction.

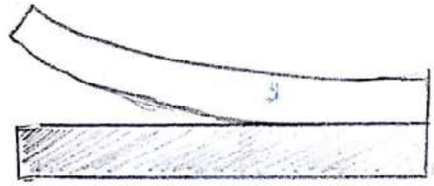
Adhersion of Layers

whenever 2 layers of materials either similar or dissimilar are bonded together, a possibility of delamination exists.

A bilayer Structure can delaminate at the interface either by.

(1) peeling of one layer from other

(2) Severing along the interface by Shear.
(To cut something into 2 pieces)



Peeling off

Shear

Severing along the interface by Shear

→ The main cause is excessive thermal and mechanical Stress

Interfacial Stresses

There are 3 types of stress that exist in bilayer structures.

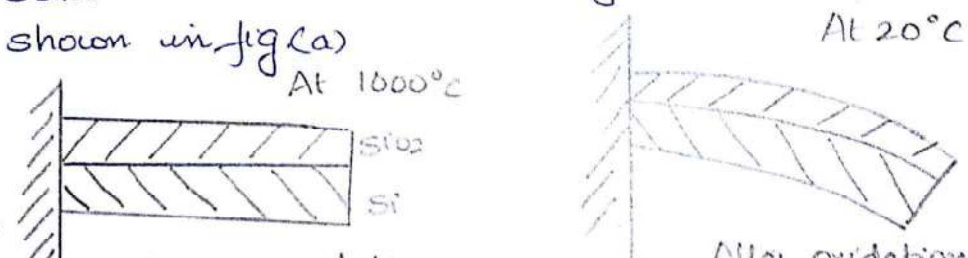
(1) Thermal stresses : resulting from the mismatch of the coefficients of thermal expansion (CTE) of the component materials

Eg: CTE for Silicon is about 5 times that of SiO_2 hence when the bilayer structure of Si and SiO_2 is subjected to high operating temperature, severe thermal stresses causes delamination of SiO_2 layer from silicon substrate.

(2) The second type of interfacial stress is the residual stresses that are inherent in microfabrication processes.

Eg: SiO_2 layer grown on the top surface of silicon substrate beam at 1000°C by thermal oxidation is shown in fig (a)

Residual Stress & Strain in bilayer beam



During oxidation   After oxidation

The resultant bilayer beam at room temp is as in fig (b). This is becoz of the significant difference in the CTE for both materials.

Also the residual stress and associated residual strain in the SiO_2 layer after it is cooled to room temp (20°C) can cause multiple cracks in the layer.

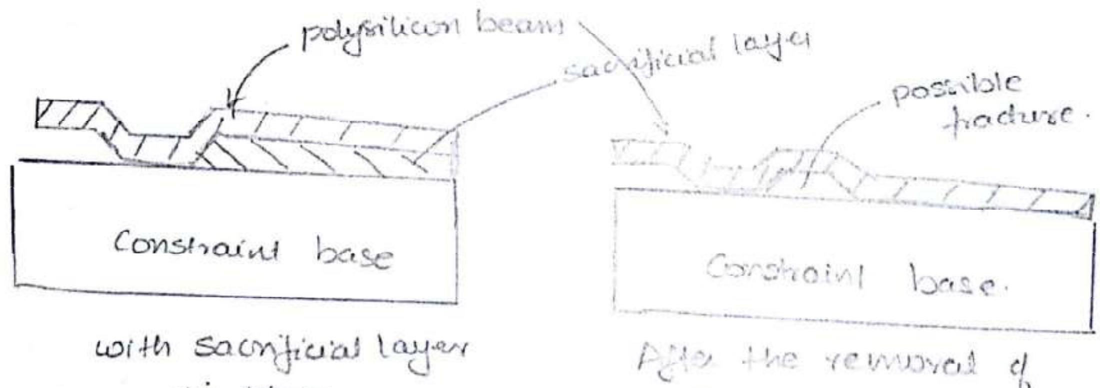
3. The third type of stress in thin film structures is the intrinsic stress due to local change of atomic structure during microfabrication processes.

Eg: Excessive doping can introduce substantial residual stresses in the structure after surface micromachining.

Stiction:

Many have experienced the difficulty in separating 2 transparencies after the thin dividing paper is pulled out. A similar phenomenon occurs in surface micromachining.

The phenomenon of 2 separate pieces sticking together after the sacrificial layer is removed is called stiction



with sacrificial layer in place

After the removal of sacrificial layer

For Eg: consider the production of a thin beam as in figure. Stiction could happen with the thin polysilicon beam dropping on to the top surface of the silicon substrate after the removal of the sacrificial layer PSG.

The two materials would then stick together after the joint. Considerable machanical forces are required to separate the 2 stuck layers, these excessive forces can break the delicate microstructure.

Stiction occurs as a result of hydrogen bonding of surfaces during rinsing of the interface after etching of PSG sacrificial layer or by forces such as the van der Waals forces.

Virtual Instrumentation

- Instrument is a device which collect the data, analyse data and display Information. E.g.: Transducer, Oscilloscopes, Digital Multi-meter
- Virtual instrumentation is the use of customizable software and modular measurement hardware to create user-defined measurement systems, called virtual instruments.
- A virtual instrument (VI) is defined as an industry-standard computer equipped with user friendly application software, cost-effective hardware and driver software that together perform the functions of traditional instruments.
- Simulated physical instruments are called virtual instruments (VIs). Virtual instrumentation software based on user requirements defines general-purpose measurement and control hardware functionality.
- To test, control and design applications, making accurate analog and digital measurements.
- Traditional hardware instrumentation systems are made up of pre-defined hardware components, such as digital multimeters and oscilloscopes that are completely specific to their stimulus, analysis, or measurement function.
- Because of their hard-coded function, these systems are more limited in their versatility than virtual instrumentation systems.
- The primary difference between hardware instrumentation and virtual instrumentation is that software is used to replace a large amount of hardware.
- The software enables complex and expensive hardware to be replaced by already purchased computer hardware.
- Using VI, can also control external hardware devices from desktop computer and for displaying unit.
- It extends to computerized systems for controlling processes based on data collected and processed by a computerized instrumentation system.
- Virtual instrumentation combines mainstream commercial technologies, such as the PC, with flexible software and a wide variety of measurement and control hardware.
- Engineers use virtual instrumentation to bring the power of flexible software and PC technology to test, control and design applications making accurate analog and digital measurements.
- Engineers and scientists can create user-defined systems that meet their exact application needs.
- Industries with automated processes, such as chemical or manufacturing plants use virtual instrumentation with the goal of improving system productivity, reliability, safety, optimization and stability.
- The front panel control function of the existing instrument is duplicated through the computer interface.
- The flexibility is possible as the capabilities of a virtual instrument depend very little on dedicated hardware. A customer can use his own computer.
- The application ranges from simple laboratory experiments to large automation application.

- The concept was born in late 1970s.
- With virtual instrumentation, engineers and scientists reduce development time, design higher quality products, and lower their design costs.
- Virtual instrumentation is necessary because it is flexible.
- It delivers instrumentation with the rapid adaptability required for today's concept, product and process design, development and delivery.
- Virtual instruments are defined by the user while traditional instruments have fixed vendor-defined functionality
- Every virtual instrument consists of two parts—**software and hardware**.
- A virtual instrument provides all the software and hardware needed to accomplish the measurement or control task.
- In addition, with a virtual instrument, engineers and scientists can customize the acquisition, analysis, storage, sharing and presentation functionality using productive, powerful software.

Historical Development

- First Phase: Early pure Analog Measurement Devices such as oscilloscopes. They were completely closed dedicated systems
- Second Phase: Instruments started to digitalize measured signals allowing Digital Data Acquisition and Processing Devices
- Third Phase: Measurement Instruments became computer based. Digital processing based on general purpose computing platform. Began to include interfaces that enabled communication between the instrument and the computer.
- Fourth Phase: Distributed virtual instrumentation

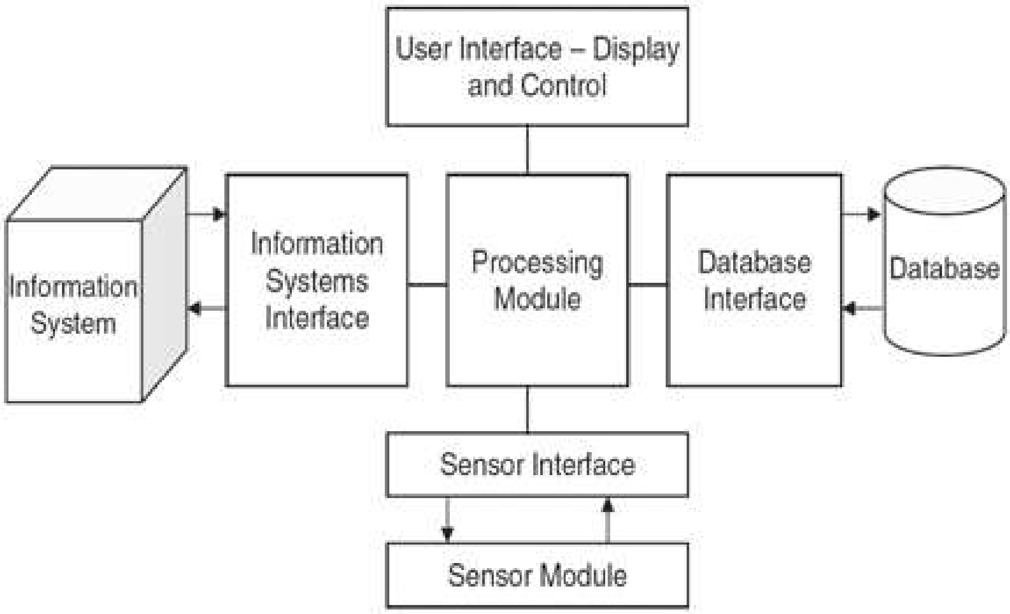
Components of Virtual Instrumentation

Virtual instrumentation combines productive software, modular I/O and scalable platforms.

- Virtual instrumentation uses highly productive software.
- Uses symbolic or graphical representations.

- With the right software tool, engineers and scientists can efficiently create their own applications by designing and integrating the routines that a particular process requires.
- Can also create an appropriate user interface that best suits the purpose of the application and those who will interact with it.
- You can define how and when the application acquires data from the device, how it processes, manipulates and stores the data, and how the results are presented to the user.
- With powerful software, we can build intelligence and decision-making capabilities into the instrument so that it adapts when measured signals change inadvertently or when more or less processing power is required.
- Another virtual instrumentation component is modular I/O, designed to be rapidly combined in any order or quantity to ensure that virtual instrumentation can both monitor and control any development aspect.
- Using well-designed software drivers for modular I/O, engineers and scientists quickly can access functions during concurrent operation.
- The third virtual instrumentation element using commercial platforms, often enhanced with accurate synchronization, ensures that virtual instrumentation takes advantage of the very latest computer capabilities and data transfer technologies.
- This element delivers virtual instrumentation on a long-term technology base that scales with the high investments made in processors, buses and more.

<u>Architecture of Virtual Instrumentation</u>



A virtual instrument is composed of

- Sensor Module,
- Sensor Interface,
- Information Systems Interface,
- Processing Module,
- Database Interface and
- User Interface.

Sensor Module:

- The sensor module detects physical signal and transforms into electrical form, conditions the signal, and then into a digital form for further manipulation.
- A sensor module principally consists of three main parts: the sensor, the signal conditioning part, and the A/D converter
- Once the data are in a digital form on a computer, they can be displayed, processed, mixed, compared, stored in a database, or converted back to analog form for further process control.

Sensor Interface:

- Interfaces used for communication between sensor modules and the computer
- Can be classified into wired and wireless.
- Wired Interfaces are usually either standard parallel interfaces, like GPIB, SCSI, System buses PCI eX(PXI), VME eX (VXI) or serial buses like RS232, USB interfaces.
- Wireless interfaces are increasingly used because of convenience, typical interfaces include 802.11 family of standards, Bluetooth or GPRS/GSM interface.

Processing Module:

- Integration of Microprocessors allowed flexible implementation of sophisticated processing functions.
- Since the little dependency of VI's on dedicated hardware, functionality & appearance of VI may be completely changed utilizing various processing functions.
- Processing functions used in Virtual Instrumentation may be classified as "Analytical processing" and "Artificial Intelligence Techniques".

Data Base & Information System Interface Modules:

- Computerized instrumentation allows measured data to be stored for offline processing, or to keep records.
- Several currently available database technologies are Xtensible Markup Language (XML) and SQL Server and Oracle.
- Virtual instruments are increasingly integrated with other information systems, such as real time alerts, and predictive warnings etc.

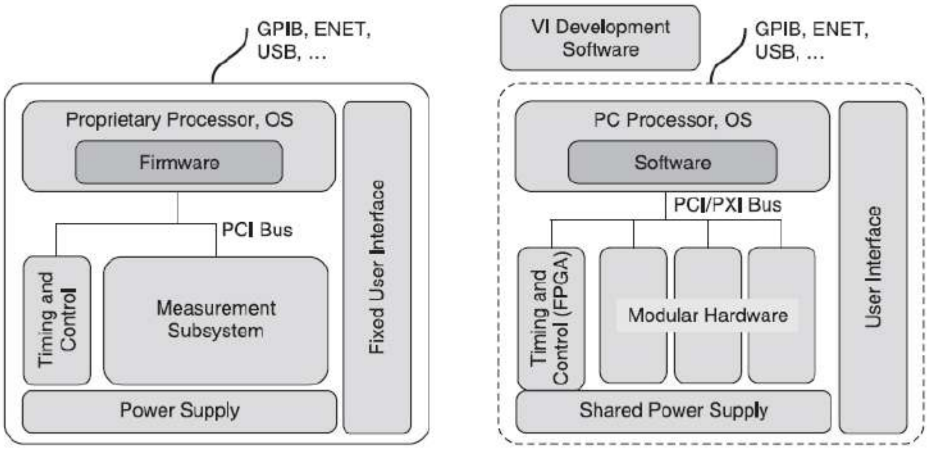
User Interface- Display & Control:

- Terminal user interfaces, graphical user interfaces, multi-model user interfaces and virtual & augmented reality interfaces.
- An effective user interface for presentation and control of a virtual instrument affects efficiency and precision of an operator do the measurements and facilitates to the result interpretation.
- The main application domain for this is Graphical user interfaces.

Traditional vs Virtual Instruments

- A traditional instrument is designed to collect data from an environment, or from a unit under test, and to display information to a user based on the collected data.
- Such an instrument may employ a transducer to sense changes in a physical parameter such as temperature or pressure, and to convert the sensed information into electrical signals such as voltage or frequency variations.
- The term "instrument" may also cover a physical or software device that performs an analysis on data acquired from another instrument and then outputs the processed data to display or recording means. This category of instruments includes oscilloscopes, spectrum analyzers and digital millimetres.
- A virtual instrument (VI) is defined as an industry-standard computer equipped with user-friendly application software, cost-effective hardware and driver software that together perform the functions of traditional instruments. Simulated physical instruments are called virtual instruments (VIs).
- Virtual instrumentation software based on user requirements defines general-purpose measurement and control hardware functionality. With virtual instrumentation, engineers and scientists reduce development time, design higher quality products, and lower their design costs.

Traditional Instruments	Virtual Instruments
Vendor-defined	User-defined
Function-specific, stand-alone with limited connectivity	Application-oriented system with connectivity to networks, peripherals, and applications
Hardware is the key	Software is the key
Expensive	Low-cost, reusable
Closed, fixed functionality	Open, flexible functionality leveraging off familiar computer technology
High development and maintenance costs	Software minimizes development and maintenance costs
Minimal economics of scale	Maximum economics of scale



Advantages of Virtual Instruments

- Improved Performance
- Lower cost of instrumentation
- Easy-to-use graphical interface
- Portability between various computer platforms
- Increase the utility of computer
- Flexibility
- Plug-In and Networked Hardware
- The Costs of a Measurement Application
- Minimizing Set-Up and Configuration Time Costs
- Platform-Independent Nature

Disadvantages

- Security: Sensitive information may be accessible to public users
- Power Consumption: VI demands that many devices run simultaneously and can consume a lot of power. Each computer will consume a large amount of power in addition to any external hardware.

Graphical Programming and Textual Programming

- Graphical programming is a visually-oriented approach to programming.
- Graphical programming is easier and more intuitive to use than traditional textual programming.
- Textual programming requires the programmers to be reasonably proficient in the programming language.
- Non-programmers can easily learn the graphical approach faster at less amount of time.
- The main advantage of textual languages like C is that they tend to have faster graphical approach execution time and better performance than graphical programs.
- Textual programming environments are typically used in determining high throughput virtual instrumentation systems, such as manufacturing test systems.
- Textual programming environments are popular and many engineers are trained to use these standardized tools.
- Graphical environments are better for nonprogrammers and useful for developing virtual instruments quickly and need to be reconfigured rapidly.
- Virtual instrumentation is not limited to graphical programming but can be implemented using a conventional programming language.
- The most important task is to understand how to use standard analysis packages that can directly input data from the instruments and can be used to analyse, store and present the information in a useful format. Irrespective of whether it is classical or graphical environment any system with a graphical system design can be looked at as being composed of two parts—the user interface and the underlying code. The code in a conventional language like C comprises a number of routines while in the graphical language G it is a collection of icons interconnected by multi-coloured lines.

Text-based programming	Graphical programming
• *Syntax* must be known to do programming.	• *Syntax* is knowledge but is not required for programming.
• The execution of the program is from *top to bottom*.	• The execution of program is from *left to right*.
• To check for the *error* the program has to be compiled or executed.	• *Errors* are indicated as we wire the blocks.
• *Front panel* design needs extra coding or needs extra work.	• *Front panel* design is a part of programming.
• Text-based programming is *non interactive*.	• Graphical programming is highly *interactive*.
• This is text-based programming where the programming is a *conventional method*.	• The programming is *Data Flow Programming*.
• *Logical Error* finding is easy in large programs.	• *Logical Error* finding in large programs is quiet complicated.
• Program flow is *not visible*.	• Data flow is *visible*.
• It is *test-based* programming.	• It is *icon-based* programming and wiring.
• Passing parameters to *sub routine* is difficult.	• Passing parameters to *sub VI* is easy.

Introduction to LabVIEW

- Laboratory Virtual Instrument Engineering Workbench (LabVIEW).
- LabVIEW 1.0 was launched in 1986.
- LabVIEW is a graphical programming language (G) that uses icons instead of lines of text to create applications.
- LabVIEW (Laboratory Virtual Instrument Engineering Workbench) is a graphical programming environment which has become prevalent throughout research labs, academia and industry.
- It is a powerful and versatile analysis and instrumentation software system for measurement and automation. Its graphical programming language called G programming is performed using a graphical block diagram that compiles into machine code and eliminates a lot of the syntactical details.
- LabVIEW Programs are called Virtual Instruments (VIs) because their appearance and operation imitate physical instruments like Oscilloscopes, Digital Multi-meter.
- LabVIEW contains a set of VIs and functions for acquiring, analysing, displaying and storing data.
- Uses symbolic or graphical representations to speed up development.
- Includes analysis functions for differential equations, optimization, curve fitting, calculus, linear algebra, statistics and so on.
- Includes the tools to present the data on the computers: Charts, Graphs, Tables, Gauges, Meters, Tanks, 3D controls, 3D Graphs, Picture Control etc.
- LabVIEW offers more flexibility than standard laboratory instruments because it is software based.
- Using LabVIEW, the user can originate exactly the type of virtual instrument needed and programmers can easily view and modify data or control inputs.
- LabVIEW programs are called virtual instruments (VIs), because their appearance and operation imitate physical instruments like oscilloscopes.
- LabVIEW is designed to facilitate data collection and analysis, as well as offers numerous display options.
- With data collection, analysis and display combined in a flexible programming environment, the desktop computer functions as a dedicated measurement device.
- LabVIEW contains a comprehensive set of VIs and functions for acquiring, analyzing, displaying, and storing data, as well as tools to help you troubleshoot your code.
- LabVIEW can communicate with hardware such as data acquisition, vision, and motion control devices, and GPIB, PXI, VXI, RS-232, and RS-485 devices.
- LabVIEW also has built-in features for connecting your application to the Web using the LabVIEW Web Server and software standards such as TCP/IP networking and ActiveX.
- Using LabVIEW, you can create test and measurement, data acquisitions, instrument control, data logging, measurement analysis, and report generation applications.
- LabVIEW can command plug-in data acquisition, or DAQ, devices to acquire or generate analog and digital signals

- Using DAQ devices and LabVIEW, to monitor a temperature, send signals to an external system, or determine the frequency of an unknown signal.
- LabVIEW also facilitates data transfer over the General Purpose Interface Bus (GPIB), or through your computer's built-in USB, Ethernet, Firm wire (also known as IEEE 1394), or serial port.
- GPIB is frequently used to communicate with oscilloscopes, scanners, and multi meters, and to drive instruments from remote locations.

LabVIEW Software

LabVIEW software consists of two parts: Front Panel and Block Diagram Panel

Front Panel

- Controls = Inputs
- Indicators = Outputs

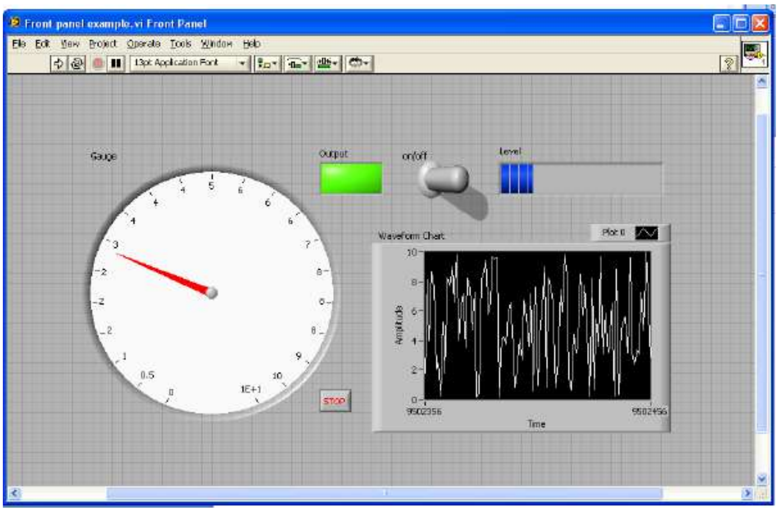
Block Diagram

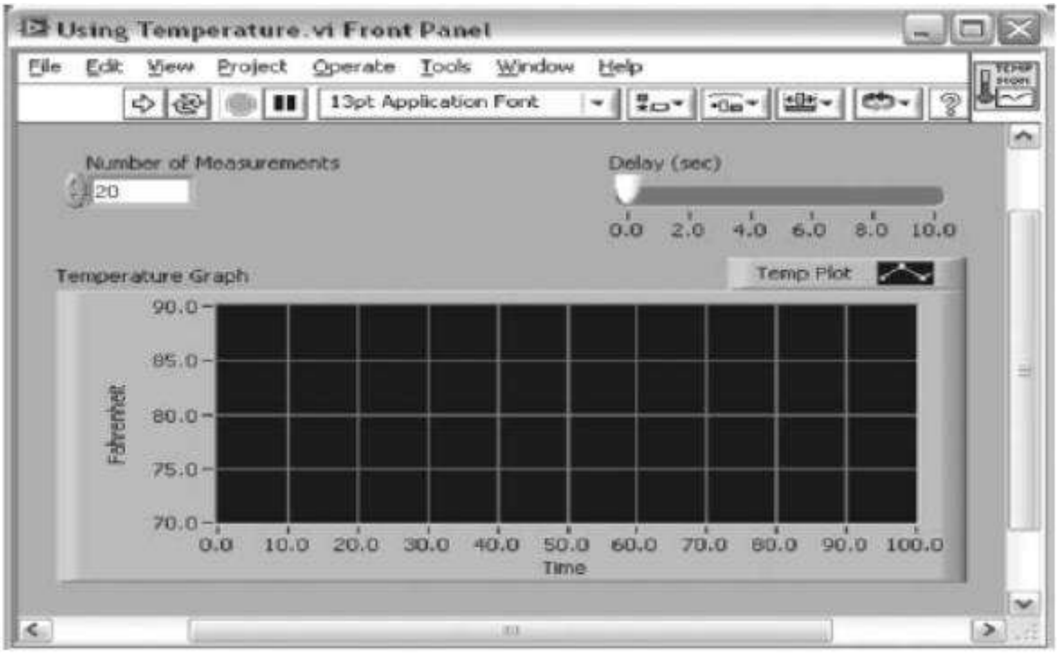
Accompanying "program" for front panel

Components "wired" together

LabVIEW Front Panel:

- The front panel is the user interface of the VI.
- Build with controls and indicators, which are the interactive input and output terminals of the VI, respectively.
- Controls are knobs, push buttons, dials, and other input devices.
- Indicators are graphs, LEDs, and other displays.
- Controls simulate instrument input devices and supply data to the block diagram of the VI.
- Indicators simulate instrument output devices and display data the block diagram acquires or generates.

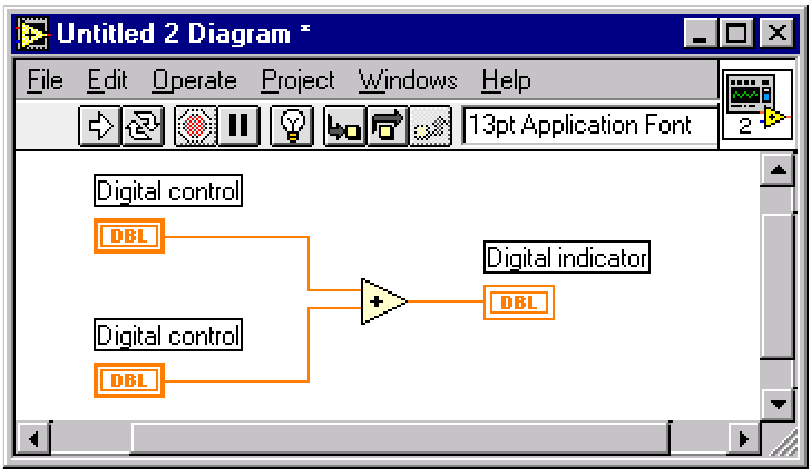




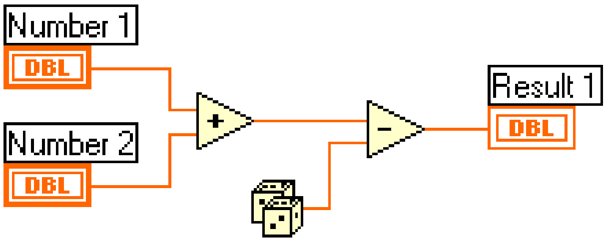
The front panel can include knobs, push buttons, graphs and various other controls (which are user inputs) and indicators (which are program outputs).

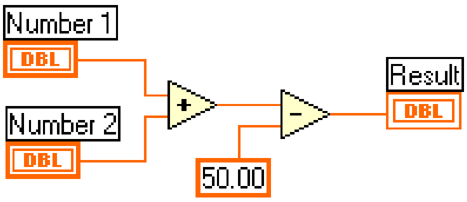
<u>Block Diagram</u>

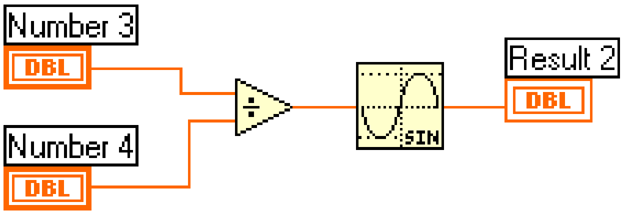
- Every control or indicator on the front panel has a corresponding terminal on the block diagram.
- It contains terminals, nodes, wires etc.
- Accompanying "program" for front panel. I.e., the programme for front panel is done in the block diagram panel. Components are selected from the library and wired together as per the requirement of the user.
- Wires connect each of the nodes on the block diagram, including control and indicator terminals, functions and structures.



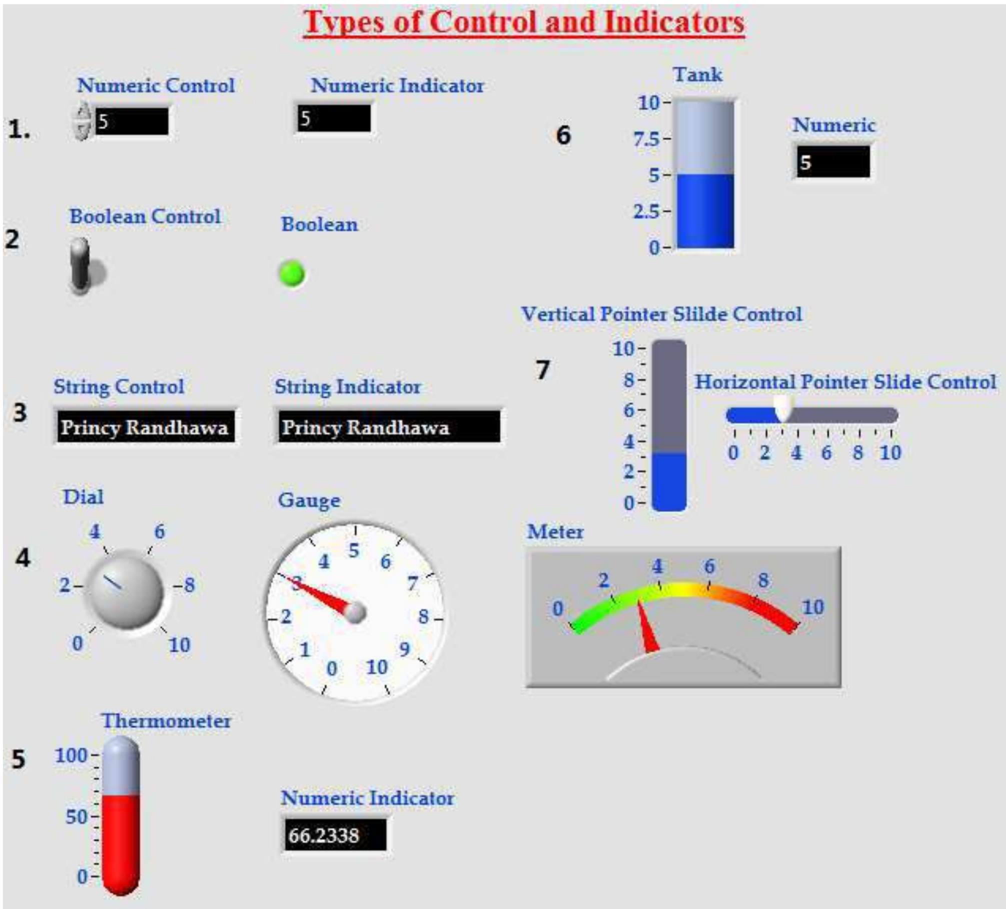
- Block diagram does NOT execute left to right
- Node executes when data is available to ALL input terminals
- Nodes supply data to all output terminals when done



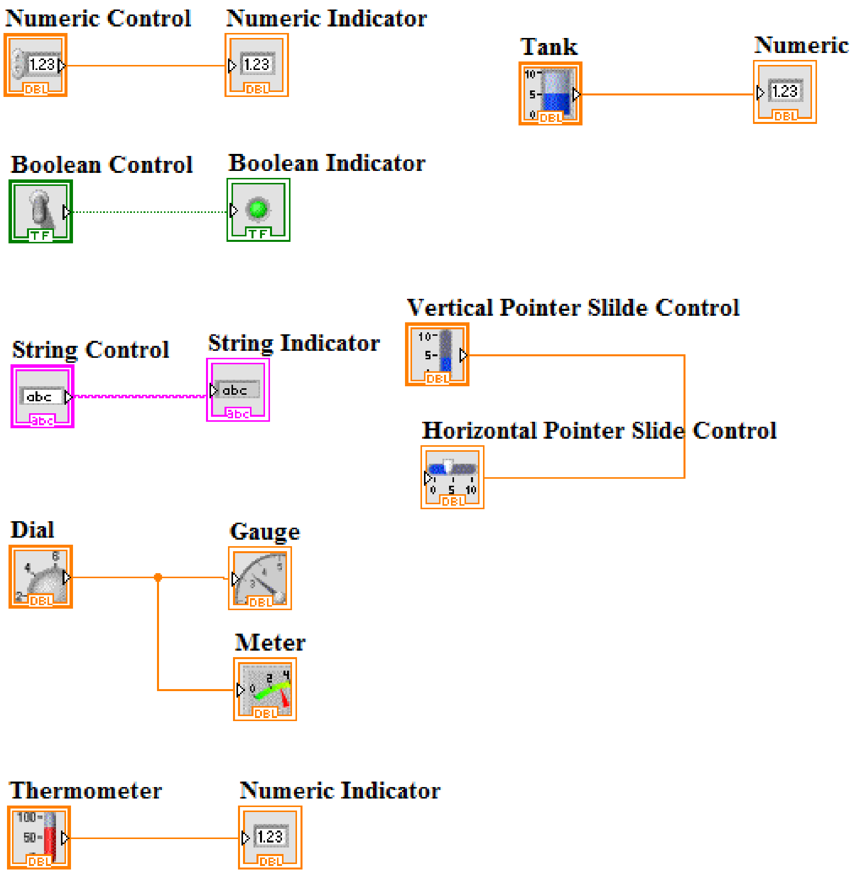




Front Panel and Corresponding Block Diagram of LabVIEW



Front Panel



Block Diagram

Advantages of LabVIEW

- Graphical user interface: Design professionals use the drag-and-drop user interface library by interactively customizing the hundreds of built-in user objects on the controls palette.
- Drag-and-drop built-in functions: Thousands of built-in functions and IP including analysis and I/O, from the functions palette to create applications easily.
- Modular design and hierarchical design: Run modular LabVIEW VIs by themselves or as subVIs and easily scale and modularize programs depending on the application.
- Multiple high level development tools: Develop faster with application specific development tools, including the LabVIEW Statechart Module, LabVIEW Control Design and Simulation Module and LabVIEW FPGA Module.
- Professional Development Tools: Manage large, professional applications and tightly integrated project management tools; integrated graphical debugging tools; and standardized source code control integration.
- Multi platforms: The majority of computer systems use the Microsoft Windows operating system. LabVIEW works on other platforms like Mac OS, Sun Solaris and Linux. LabVIEW applications are portable across platforms.

- Reduces cost and preserves investment: A single computer equipped with LabVIEW is used for countless applications and purposes—it is a versatile product. Complete instrumentation libraries can be created for less than the cost of a single traditional, commercial instrument.
- Flexibility and scalability: Engineers and scientists have needs and requirements that can change rapidly. They also need to have maintainable, extensible solutions that can be used for a long time
- Connectivity and instrument control: LabVIEW has ready-to-use libraries for integrating stand-alone instruments, data acquisition devices, motion control and vision products,
- Open environment: LabVIEW provides the tools required for most applications and is also an open development environment.
- Visualization capabilities: LabVIEW includes a wide array of built-in visualization tools to present data on the user interface of the virtual instrument as chart, graphs, 2D and 3D visualization.
- Compiled language for fast execution: LabVIEW is a compiled language that generates optimized code with execution speeds comparable to compiled C and develops high-performance code.
- Object–oriented design: Use object-oriented programming structures to take advantage of encapsulation and inheritance to create modular and extensible code.
- Algorithm design: Develop algorithms using math-oriented textual programming and interactively debug .m file script syntax with LabVIEW MathScript.

Automation System

In general sense the term “Industry” is defined as systematic Economic Activity that could be related to Manufacture/Service/ Trade.

The word Automation is derived from Greek words “Auto” (self) and “Matos” (moving)- Automation therefore is the mechanism for systems that “move by itself”. Apart from this original sense of the word, automated systems also achieve significantly superior performance than what is possible with manual systems, in terms of power, precision and speed of operation, because automated system cannot get bored, no mistakes due to fatigue, always same quality, handle larger things that cannot be done by human. Using computing techniques, much more sophisticated and efficient operational solutions can be derived and applied in real-time in automation system.

Automation systems are the set of technologies that results in operation of machines and systems without significant human intervention and achieves performance superior to manual operation.

It includes the use of various control systems for operating equipment such as machinery, processes in factories, boilers and heat treating ovens, switching on telephone networks, steering and stabilization of ships, aircraft and other applications and vehicles with minimal or reduced human intervention and is implemented using a program of instructions combined with a control system that executes the instructions.

To automate a process power is required, both to drive the process itself and to operate the program and control system.

Although automation can be applied in a wide variety of areas, it is most closely associated with the manufacturing industries.

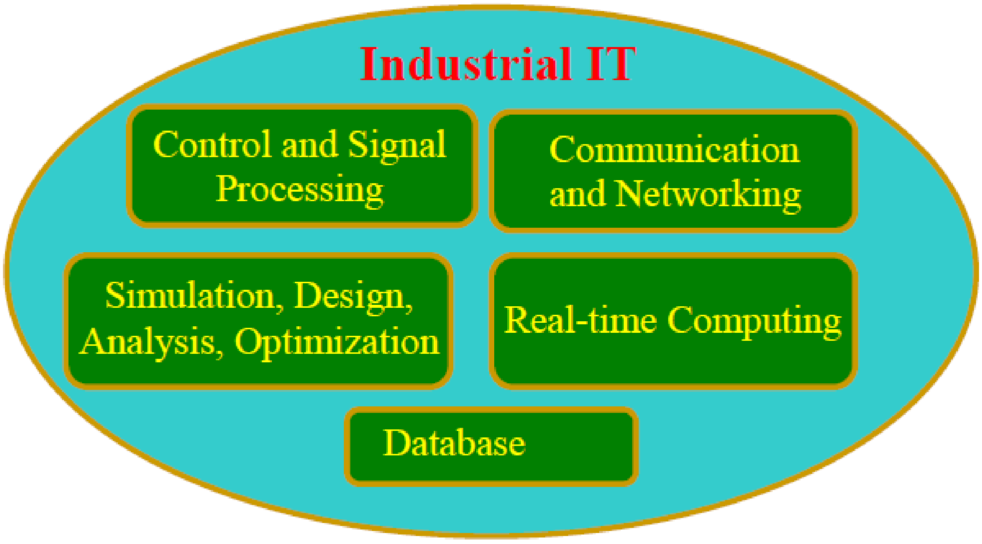
Control: Control is a set of technologies that achieves desired patterns of variations of operational parameters and sequences for machines and systems by providing the input signals necessary. Automation Systems may include Control Systems but the reverse is not true. Control Systems may be parts of Automation Systems.

The main function of control systems is to ensure that outputs follow the set points. However, Automation Systems may have much more functionality, such as computing set points for control systems, monitoring system performance, plant start up or shut down, job and equipment scheduling etc. Automation Systems are essential for most modern industries.

An Industrial Automation System consists of numerous elements that perform a variety of functions related to Instrumentation, Control, Supervision and Operations Management related to the industrial process. Such physical control inputs are provided by the actuation elements.

Industrial Automation vs. Industrial Information Technology

Industrial Automation makes extensive use of Information Technology. Fig. below shows some of the major IT areas that are used in the context of Industrial Automation.



Industrial Automation also involves significant amount of hardware technologies, related to Instrumentation and Sensing, Actuation and Drives, Electronics for Signal Conditioning, Communication and Display, Embedded as well as Stand-alone Computing Systems etc.

As Industrial Automation systems grow more sophisticated in terms of the knowledge and algorithms they use, as they encompass larger areas of operation comprising several units or the whole of a factory, or even several of them, and as they integrate manufacturing with other areas of business, such as, sales and customer care, finance and the entire supply chain of the business, the usage of IT increases dramatically.

Both control and IT are used in automation systems to realize one or more of its functionalities.

While Control Technology is used for operation of the individual machines and equipment, IT is used for coordination, management and optimized operation of overall plants.

Role of Automation in Industry

Manufacturing processes, basically, produce finished product from raw/unfinished material using energy, manpower and equipment and infrastructure. Since an industry is essentially a "systematic economic activity", the fundamental objective of any industry is to make profit.

Profit = (Price/unit – Cost/unit) x Production Volume

Profit can be maximized by producing good quality products, which may sell at higher price, in larger volumes with less production cost and time. Automation affects every aspect of profit making- Crucial for the success of an industry

Reduction of Cost/Unit

- Cost/unit of a mass-manufactured industrial product is affected by Material, Energy, Manpower and Infrastructure.
- Automation reduces the scraps and non-quality products there by reducing the material cost.
- Automated machines are programmed to work with optimum energy.
- Purpose of automation is to reduce man-power.
- Using automation, much larger number of units can be produced in a given time- life of the infrastructure: Reduces the infrastructure cost.
- Infrastructure cost/unit= Total IC/Number of units
- Total infrastructure cost will go up, but IC per unit will come down.

Increasing Production Volume

- Production volume is affected by Production Time, Material Handling Time, Idle Time and Quality Assurance Time.
- Firstly, automated machines have significantly lower production times- less production time/ unit
- For example, in machine tools, manufacturing a variety of parts, significant setup times are needed for setting the operational configuration and parameters whenever a new part is loaded into the machine.
- This can lead to significant unproductive for expensive machines when a variety of products is manufactured.
- In Computer Numerically Controlled (CNC) Machining Centers set up time is reduced significantly with the help of Automated Tool Changers, Automatic Control of Machines from a Part Program loaded in the machine computer.
- Similarly, systems such as Automated Guided Vehicles, Industrial Robots, Automated Crane and Conveyor Systems reduce material handling time.
- Idle time can be reduced by automation and high coordination leading to maximum capacity utilization of the equipment.
- Automated quality assurance can cut down the Quality Assurance time.

Increasing Price/Unit

- As quality increases, demand increases and thus price also can be increased.
- Quality of a product can be improved by the material used, process of manufacturing and control used out of which the latter two are enhanced by automation.
- The product quality that can be achieved with automated precision machines and processes cannot be achieved with manual operations.
- Moreover, since operation is automated, the same quality would be achieved for thousands of parts with little variation.

Industrial Products go through their life cycles, which consists of various stages.

- At first, a product is conceived based on Market feedbacks, as well as Research and Development Activities.
- Once conceived the product is designed. Prototype Manufacturing is generally needed to prove the design.
- Once the design is proved, Production Planning and Installation must be carried out to ensure that the necessary resources and strategies for mass manufacturing are in place.
- This is followed by the actual manufacture and quality control activities through which the product is mass-produced.
- This is followed by a number of commercial activities through which the product is actually sold in the market.
- Automation also reduces the overall product life cycle i.e., the time required to complete
    - Product conception and design
    - Process planning and installation
    - Production, Quality Control etc.
- Automation can increase profitability in multiple ways by reducing labour, material and energy requirements, by improving quality as well as productivity.

Advantages of Automation

- Increased productivity.
- Improved quality or increased predictability of quality.
- Increased consistency of output.
- Reduced direct human labour costs and expenses.
- Installation in operations reduces cycle time.
- Can complete tasks where a high degree of accuracy is required.
- Replaces human operators in tasks that involve hard physical or monotonous work (e.g., using one forklift with a single driver instead of a team of multiple workers to lift a heavy object)
- Reduces some occupational injuries (e.g., fewer strained backs from lifting heavy objects)
- Replaces humans in tasks done in dangerous environments (i.e. fire, space, volcanoes, nuclear facilities, underwater, etc.)
- Performs tasks that are beyond human capabilities of size, weight, speed, endurance, etc.
- Reduces operation time and work handling time significantly.
- Frees up workers to take on other roles.
- Provides higher level jobs in the development, deployment, maintenance and running of the automated processes.

Types of Automation Systems

Automation systems can be categorized based on the flexibility and level of integration in manufacturing process operations.

Fixed Automation:

- It is used in high volume production with dedicated equipment, which has a fixed set of operation and designed to be efficient for this set.
- Continuous flow and Discrete Mass Production systems use this automation. e.g. Distillation Process, Conveyors, Paint Shops, Transfer lines etc.
- A process using mechanized machinery to perform fixed and repetitive operations in order to produce a high volume of similar parts.
- Fixed automation is appropriate in the following circumstances.
    - Low variability in product type as also in size, shape, part count and material
    - Predictable and stable demand for 2- to 5-year time period, so that manufacturing capacity requirement is also stable
    - High production volume desired per unit time
- Significant cost pressures due to competitive market conditions. So automation systems should be tuned to perform optimally for the particular product.

Programmable Automation:

- It is used for a changeable sequence of operation and configuration of the machines using electronic controls.
- However, significant programming effort may be needed to reprogram the machine or sequence of operations.
- Investment on programmable equipment is less, as production process is not changed frequently.
- It is typically used in Batch process where job variety is low and product volume is medium to high, and sometimes in mass production also. e.g. in Steel Rolling Mills, Paper Mills etc.

Flexible Automation:

- It is used in Flexible Manufacturing Systems (FMS) which is invariably computer controlled.
- Human operators give high-level commands in the form of codes entered into computer identifying product and its location in the sequence and the lower level changes are done automatically.
- Each production machine receives settings/instructions from computer.
- These automatically loads/unloads required tools and carries out their processing instructions.
- After processing, products are automatically transferred to next machine.

- It is typically used in job shops and batch processes where product varieties are high and job volumes are medium to low. Such systems typically use Multi-purpose CNC machines, Automated Guided Vehicles (AGV) etc.

Flexible automation is used in the following situations.

- Significant variability in product type. Product mix requires a combination of different parts and products to be manufactured from the same production system
- Product life cycles are short. Frequent upgradation and design modifications alter production requirements
- Production volumes are moderate, and demand is not as predictable

Integrated Automation:

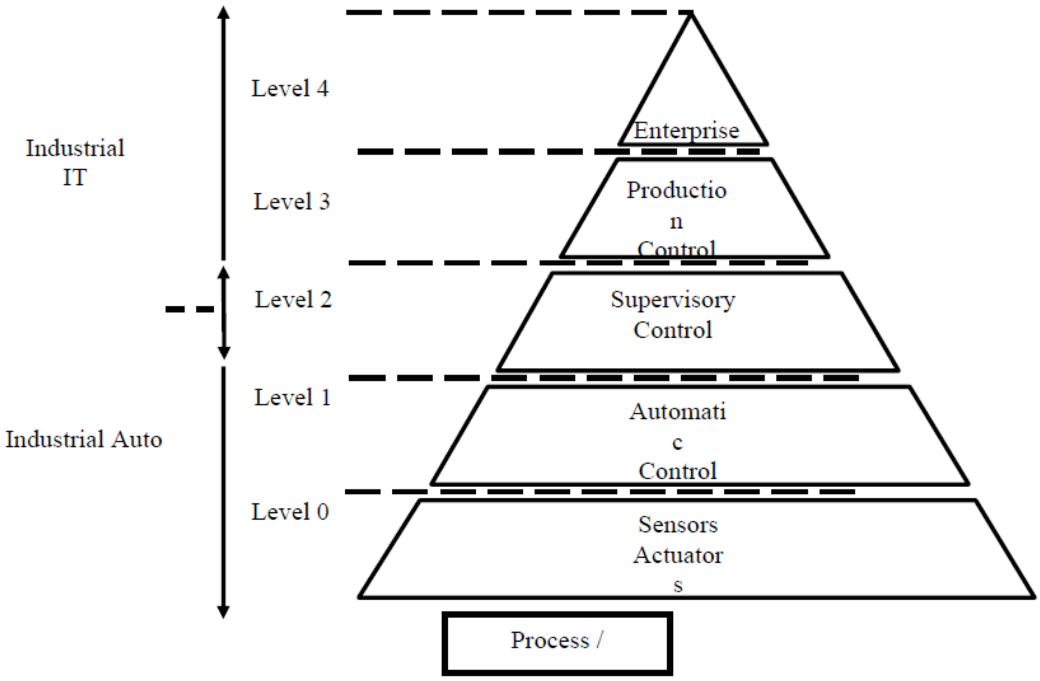
- It denotes complete automation of a manufacturing plant, with all processes functioning under computer control and under coordination through digital information processing.
- It includes technologies such as computer-aided design and manufacturing, computer-aided process planning, computer numerical control machine tools, flexible machining systems, automated storage and retrieval systems, automated material handling systems such as robots and automated cranes and conveyors, computerized scheduling and production control.
- It may also integrate a business system through a common database. In other words, it symbolizes full integration of process and management operations using information and communication technologies.
- Typical examples of such technologies are seen in Advanced Process Automation Systems and Computer Integrated Manufacturing (CIM)

Architecture of Automation System: The Automation Pyramid

Industrial automation systems are very complex having large number of devices with a number of technologies working in synchronization. In order to know the performance of the system we need to understand the various parts of the system.

Industrial automation systems are organized hierarchically. Various components in an industrial automation system can be explained using the automation pyramid.

Here, various layers represent the wideness (in the sense of no. of devices), and fastness of components on the time-scale.



Sensors and Actuators Layer: This layer is closest to the processes and machines, used to translate signals so that signals can be derived from processes for analysis and decisions and hence control signals can be applied to the processes. This forms the base layer of the pyramid also called ‘level 0’ layer.

Automatic Control Layer: This layer consists of automatic control and monitoring systems, which drive the actuators using the process information given by sensors. This is called as ‘level 1’ layer.

Supervisory Control Layer: This layer drives the automatic control system by setting target/goal to the controller. Supervisory Control looks after the equipment, which may consist of several control loops. This is called as ‘level 2’ layer.

Production Control Layer: This solves the decision problems like production targets, resource allocation, task allocation to machines, maintenance management etc. This is called ‘level 3’ layer. Here most of the operations are performed by humans with the aid of tools that help people to perform the production control functions.

Enterprise Control Layer: This deals less technical and more commercial activities like management of supply, demand, cash flow, product marketing etc. This is called as the 'level 4' layer.

The spatial scale increases as the level is increased e.g. at lowest level a sensor works in a single loop, but there exist many sensors in an automation system which will be visible as the level is increased. As the level increases, the number of devices controlled by the layer increases. Each device in sensor and actuator layer is dedicated for a small area, but as the level increases the area controlled by each device increases.

The lowest level is faster in the time scale and the higher levels are slower. The aggregation of information over some time interval is taken at higher levels.

From level 2 upward, it is not true that computer based automation is needed/present for functioning the performing on that level, some functions may be done by human being

All the layers are connected by various types of communication systems. For example, the sensors and actuators may be connected to the automatic controllers using a point-to-point digital communication, while the automatic controllers themselves may be connected with the supervisory and production control systems using computer networks. Over the last decade, with emergence of embedded electronics and computing, standards for low level network standards (CANBus, Fieldbus etc.) for communication with low level devices, such as sensors and actuators are also emerging.

Functional Elements of Industrial Automation

An Industrial Automation System consists of numerous elements that perform a variety of functions related to Instrumentation, Control, Supervision and Operations Management related to the industrial process. These elements may also communicate with one another to exchange information necessary for overall coordination and optimized operation of the plant/factory/process.

Layer 0: Sensing and Actuation Elements

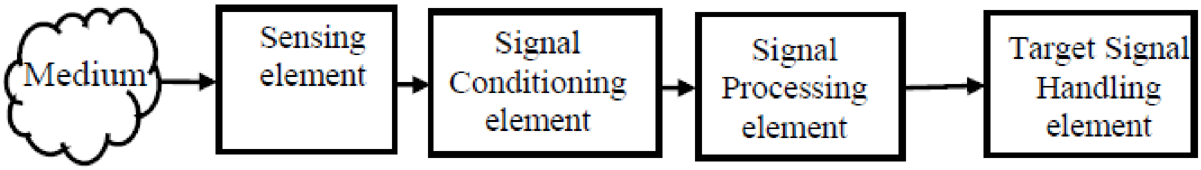
These elements interface directly and physically to the process equipment and machines.

The sensing elements translate the physical process signals such as temperature, pressure or displacement to convenient electrical or pneumatic forms of information, which can be used for analysis, decisions and finally, computation of control inputs.

These computed control inputs, which again are in convenient electrical or pneumatic forms of information, need to be converted to physical process inputs such as, heat, force or flow-rate, before they can be applied to effect the desired changes in the process outputs. Such physical control inputs are provided by the actuation elements.

Industrial Sensors and Instrument Systems

Scientific and engineering sensors and instrument systems of a spectacular variety of size, weight, cost, complexity and technology are used in the modern industry. However, a close look would reveal that all of them are composed of a set of typical functional elements connected in a specified way to provide signal in a form necessary.



A sensor system is shown, decomposed into its major functional components, along with the medium in which the measurement takes place.

The **Physical Medium** refers to the object where a physical phenomenon is taking place and we are interested in the measurement of some physical variable associated with the phenomenon. E.g.: the physical medium may stand for the hot gas in a furnace in the case of temperature measurement or the fluid in a pipe section in the case of measurement of liquid flow rate.

The **Sensing Element** is affected by the phenomenon in the physical medium either through direct or physical contact or through indirect interaction of the phenomenon in the medium with some component of the sensing element. E.g.: A thermocouple probe as the sensing element that often comes in physical contact with the hot object such as the flue gas out of a boiler-furnace or an optical pyrometer which compares the brightness of a hot body in the furnace with that of a lamp from a distance through some window and does not come in direct contact with the furnace. In the more common case where the sensing element comes in contact with the medium, often some physical or chemical property of the sensor changes in response to the measurement variable. This change then becomes a measure of the physical variable of interest. E.g.: Change in resistivity due to heat in a resistance thermometer wire or generation voltage in a thermocouple in response to a difference in temperature between its two ends.

The **Signal-conditioning Element** serves the function of altering the nature of the signal generated by the sensing element. Since the method of converting the nature of the signal generated in the sensor to another suitable signal form (usually electrical) depends essentially on the sensor, individual signal conditioning modules are characteristic of a group of sensing elements.

E.g.: An RTD whose output response is a change in its resistance due to change in temperature of its environment. This change in resistance can easily be converted to a voltage signal by incorporating the RTD in one arm of a Wheatstone's bridge. The bridge therefore serves as a signal-conditioning module. These typically involve analog electronic circuits that finally produce electrical signals in the form of voltage or current in specific ranges.

The **Signal Processing Element** is used to process the signal generated by the first stage for a variety of purposes such as, filtering (to remove noise), diagnostics (to assess the health of the sensor), linearization (to obtain an output which is linearly related with the physical measurand etc. Signal processing systems are therefore usually more general purpose in nature.

The **Target Signal-handling Element** may perform a variety of functions depending on the target application. It may therefore contain data/signal display modules, recording or/storage modules, or simply a feedback to a process control system.

E.g.: A temperature chart recorder, an instrumentation tape recorder, a digital display or an Analog to Digital Converter (ADC) followed by an interface to a process control computer.

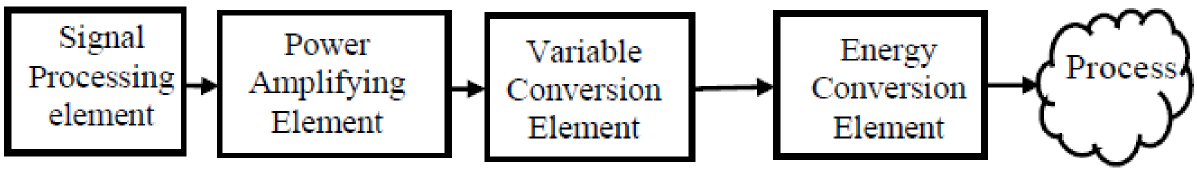
Modern sensors often have the additional capability of digital communication using serial, parallel or network communication protocols. Such sensors are called “smart” and contain embedded digital electronic processing systems.

Industrial Actuator Systems

Actuation systems convert the input signals computed by the control systems into forms that can be applied to the actual process and would produce the desired variations in the process physical variables. In the same way as in sensors but in a reverse sense, these systems convert the controller output, which is essentially information without the power, and in the form of electrical voltages (or at times pneumatic pressure) in two ways.

- Firstly, it converts the form of the variable into the appropriate physical variable, such as torque, heat or flow.
- Secondly, it amplifies the energy level of the signal manifold to be able to causes changes in the process variables.

Thus, while both sensors and actuators cause variable conversions, actuators are high power devices while sensors are not. In most cases, actuators are devices that first produce motion from electrical signal, which is then further converted to other forms.



The **Electronic Signal-processing Element** accepts the command from the control system in electrical form. The command is processed in various ways. For example, it may be filtered to avoid applying input signals of certain frequencies that may cause resonance.

Many actuators are themselves closed feedback controlled units for precision of the actuation operation. Therefore, the electronic signal-processing unit often contains the control system for the actuator itself.

The **Electronic Power Amplification Element** sometimes contains linear power amplification stages called servo-amplifiers. In other cases, it may comprise power electronic drive circuits such as for motor driven actuators.

The **Variable Conversion Element** serves the function of altering the nature of the signal generated by the electronic power amplification element from electrical to non-electrical form, generally in the form of motion.

E.g.: Electrohydraulic servo valve, stepper/servo motors, Current to Pneumatic Pressure converters etc.

The **Non-Electrical Power Conversion Elements** are used to amplify power further, if necessary, typically using hydraulic or pneumatic mechanisms.

The **Non-Electrical Variable Conversion Elements** may be used further to transform the actuated variable in desired forms, often in several stages.

E.g.: Motion-to-flow rate conversion in flow-valves, rotary to linear motion converters using mechanisms, flow-rate to heat conversion using steam or other hot fluids etc.

Other Miscellaneous Elements such as Auxiliaries for Lubrication/Cooling/Filtering, Reservoirs, Prime Movers etc., sensors for feedback, components for display, remote operations, as well as safety mechanisms since the power handling level is significantly high.

Layer 1: Industrial Control Systems

Controllers are essentially (predominantly electronic, at times pneumatic/hydraulic) elements that accept command signals from human operators or Supervisory Systems, as well as feedback from the process sensors and produce or compute signals that are fed to the actuators.

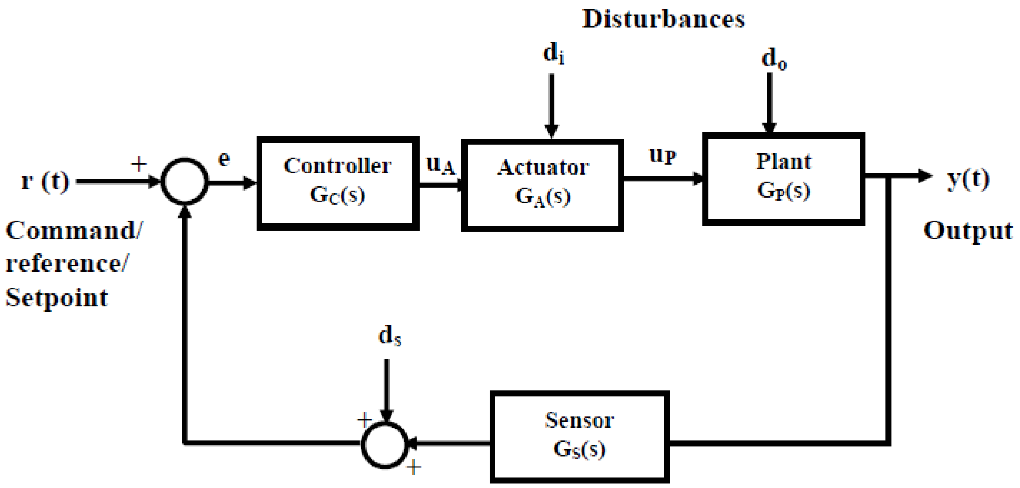
Control Systems can be classified into two kinds.

- Continuous Control
- Sequence / Logic Control

Continuous Controller

This is also often termed as Automatic Control, Process Control, Feedback Control etc. Here the controller objective is to provide such inputs to the plant such that the output y(t) follows the input r(t) as closely as possible, in value and over time.

Figure shows the Common control loop with its constituent elements, namely the Controller, the Actuator, the Sensor and the Process. In addition, the signals that exist at various points of the system are also marked. These include the command (alternatively termed the set point or the reference signal), the exogenous inputs (disturbances, noise).



The difficulties in achieving the performance objective is mainly due to the unavoidable disturbances due to load variation and other external factors, as well as sensor noise, the complexity, possible instability, uncertainty and variability in the plant dynamics, as well as limitations in actuator capabilities.

Most industrial control loop command signals are piecewise constant signals that indicate desirable levels of process variables, such as temperature, pressure, flow, level etc., which ensure the quality of the product in Continuous Processes. In some cases, such as in case of motion control for machining, the command signal may be continuously varying according to the dimensions of the product. Therefore, here deviation of the output from the command signal results in degradation of product quality. It is for this reason that the choice of the feedback signals, that of the controller algorithm (such as, P, PI or PID), the choice of the control loop structure (normal feedback loop, cascade loop or feedforward) as well as choice of the controller gains is extremely important for industrial machines and processes. Typically, the control configurations are well known for a given class of process, however, the choice of controller gains have to be made from time to time, since the plant operating characteristics changes with time. This is generally called controller tuning.

Features of Continuous Control are

- Controls analog continuous process variables- Temperature, Pressure
- Closed loop control
- Track/hold set point: Track if set point changes and hold if constant
- Reject disturbances
- Generic P-I-D/Special purpose
- Tuneable
- Implemented using Dedicated Digital Real Time System
- Comparatively inexpensive: Less than 5% cost of plant

Sequence / Logic Controller

Many control applications do not involve analog process variables, that is, the ones which can assume a continuous range of values, but instead variables that are set valued, that is they only assume values belonging to a finite set.

The simplest examples of such variables are binary variables- 1 or 0, on or off, open or closed. These control systems operate by turning on and off switches, motors, valves, and other devices in response to operating conditions and as a function of time.

In the control of a conveyor system, analog motor control is not applied. Simple on-off control is adequate. Therefore, for this application, the motor-starter actuation system may be considered as discrete having three modes, namely, start, stop and run.

A modern controller device used extensively for sequence control today in transfer lines, robotics, process control, and many other automated systems is the Programmable Logic Controller (PLC).

Features of Sequence/Logic Controller are:

- Control of discrete valued process variables
- On/Off Control
- High/Medium/Low Control
- Discrete Sensing: Limit switch, Pressure Switch, Photo Switch
- Alarm
- PLC/Industrial PC/Dedicated Processor Based devices are used
- No tuning needed unless control logic changes
- Status/sequence/timing control is used

Level 2: Supervisory Controller

Supervisory control systems perform, typically the following functions:

- **Set point computation:** Set points for important process variables are computed depending on factors such as nature of the product, production volume, mode of processing. This function has a lot of impact on production volume, energy and quality and efficiency.
- **Performance Monitoring / Diagnostics:** Process variables are monitored to check for possible system component failure, control loop detuning, actuator saturation, process parameter change etc. The results are displayed and possibly archived for subsequent analysis.
- **Start-up / Shut down / Emergency Operations:** Special discrete and continuous control modes are initiated to carry out the intended operation, either in response to operator commands or in response to diagnostic events such as detected failure modes.
- **Control Reconfiguration / Tuning:** Structural or Parametric redesign of control loops are carried out, either in response to operator commands or in response to diagnostic events such as detected failure modes. Control reconfigurations may also be necessary to accommodate variation of feedback or energy input e.g. gas fired to oil fired.
- **Operator Interface:** Graphical interfaces for supervisory operators are provided, for manual supervision and intervention.

Level 3: Production Controller

Various functions of production controller are

- Process Scheduling: Depending on the sequence of operations to be carried on the existing batches of products, processing resource availability for optimal resource utilization.
- Material Handling
- Maintenance Management: Decision processes related to detection and deployment of maintenance operations.
- Inventory Management: Decision processes related to monitoring of inventory status of raw material, finished goods etc. and deployment of operations related to their management.
- Quality Management: Assessment, Documentation & Management of Quality

- Typically, the algorithms make use of Resource Optimization Technology and are non-real-time although they may be using production data on-line.

Level 4: Enterprise controller

Enterprise controller deals less technical and more commercial activities like

- Management of supply
- Management of demand
- Management of cash flow
- Product marketing

Actuators

Actuation systems convert the input signals computed by the control systems into forms that can be applied to the actual process and would produce the desired variations in the process physical variables.

Convert the controller output, which is essentially information without the power, and in the form of electrical voltages (or at times pneumatic pressure) in two ways.

Firstly, it converts the form of the variable into the appropriate physical variable, such as torque, heat or flow. Secondly, it amplifies the energy level of the signal manifold to be able to causes changes in the process variables.

In industrial control systems, an actuator is a hardware device that converts a controller command signal into a change in a physical parameter. The change in the physical parameter is usually mechanical such as position or velocity change.

The controller command signal is usually low level, and so an actuator may also include an amplifier to strengthen the signal sufficiently to drive the actuator.

With the increasing utilization of automated machinery/systems, there has also been a rise in the demands for actuators, which play a vital role in the automation process.

Actuators, which are responsible for moving, controlling or positioning a mechanism or system, make the working of automated equipment seamless and easy

There are several areas wherein actuators play a crucial role in automated systems.

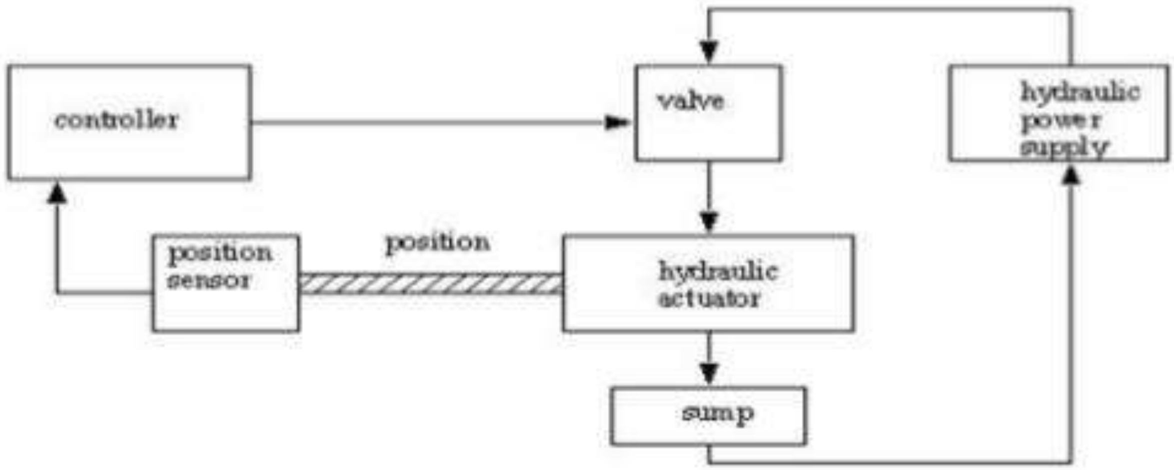
- **Process Industries:** Control the operation of various equipment in an industry.
- **Solar:** The industry utilizes several types of actuators including ball screw and ACME actuators to automate the functioning of various solar tracking devices, solar power concentration equipment, photovoltaic device, and photovoltaic concentration equipment, among several others.
- **Military:** These military actuators are basically equipped in tanks, cannon carriers, airplanes, fighter planes, helicopters, ships, and robots, because they offer complete movement. Actuators for the military industries are ruggedly built to withstand harsh environmental conditions, extreme weathers, and severe attacks.
- **Satellite:** The use of several linear and rotational actuators help automate the working of several equipment. These include dish antennas, sensors, earth-imaging equipment, and so on. Specially designed and high quality satellite actuators are sought after for precise positioning and motion control for the industry.
- **Construction:** The industry depends on high speed automated machines/systems to carry out several dangerous and tedious tasks. Actuators of varied types are used in these systems, because they are known to delivering powerful movements to heavy duty machineries like loading trucks, towing trucks, and cranes. Actuators for the construction industry are built to offer durable, reliable and long-lasting performance.

Hydraulic Actuating System

It uses fluid power to produce mechanical force. Fluid power is technology that deals with the generation, control and transmission of forces responsible for movement of mechanical element.

The fluid used in hydraulic actuator is highly incompressible so that pressure applied can be transmitted instantaneously to the member attached to it.

Hydraulics is based on Pascal's law which states that when there is an increase in pressure at any point in a confined fluid, there is an equal increase at every other point in the container. Pressure applied to a confined fluid at any point is transmitted undiminished and equally throughout the fluid in all directions and acts upon every part of the confining vessel at right angles to its interior surfaces. In simple words intensity of pressure in fluid is equal in all direction. Hydraulic actuators are used where high speed and large forces are required.



Various components of Hydraulic Actuators are given in brief below:

Pump: Hydraulic power generation system. Pump pressurizes fluid to the level required by an actuating system.

Valve: Controls the direction, pressure, flow rate of hydraulic fluid. It basically acts as a regulating system.

Actuator: Converts the fluid pressure into mechanical movement either linear or rotary.

Pipes: Used to connect various elements of actuating system with each other.

Accumulator, Reservoir and Filter: Filter is used to filter the hydraulic fluid to prevent damage to system components by contaminants. Reservoir & accumulator is use for storage purposes.

Sensors: Crucial part of any automation process. Provide some kind of feedback signal to evaluate the end result by controller.

Control Device: Control whole actuating system by providing signals. (Microcontroller or PLC are popularly used)

Hydraulic Actuators, as used in industrial process control, employ hydraulic pressure to drive an output member. The fluid used in hydraulic actuator is highly incompressible so that pressure applied can be transmitted instantaneously to the member attached to it.

A hydraulic actuator consists of a cylinder or fluid motor that uses hydraulic power to facilitate mechanical operation. The mechanical motion gives an output in terms of linear, rotary or oscillatory motion. Because liquids are nearly impossible to compress, a hydraulic actuator can exert considerable force.

Principle Used in Hydraulic Actuator System

Pascal's Law:

Pressure applied to a confined fluid at any point is transmitted undiminished and equally throughout the fluid in all directions and acts upon every part of the confining vessel at right angles to its interior surfaces.

Amplification of Force:

Since pressure P applied on an area A gives rise to a force F, given as,

$$F = P \times A$$

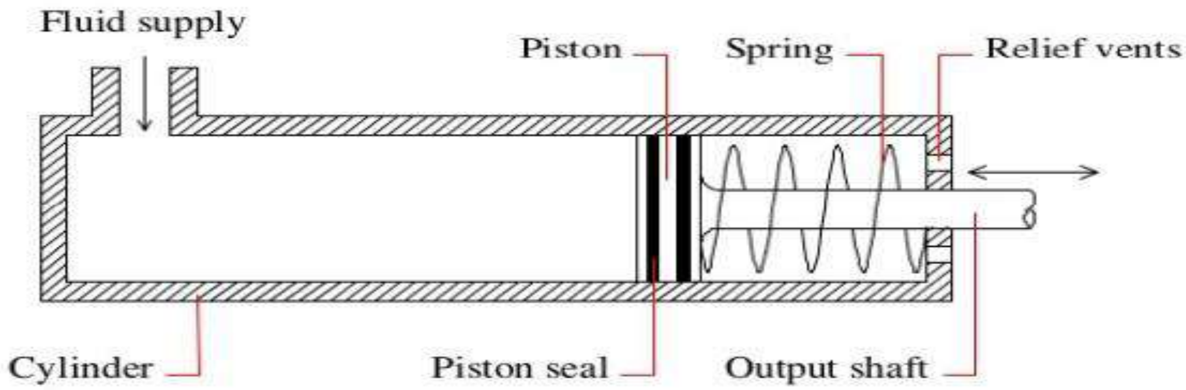
Thus, if a force is applied over a small area to cause a pressure P in a confined fluid, the force generated on a larger area can be made many times larger than the applied force that crated the pressure. This principle is used in various hydraulic devices to such hydraulic press to generate very high forces.

Hydraulic Actuators are of three types

- Linear Actuator
  - Single Acting cylinder
  - Double Acting cylinder
- Rotary Actuator
- Oscillatory Actuator

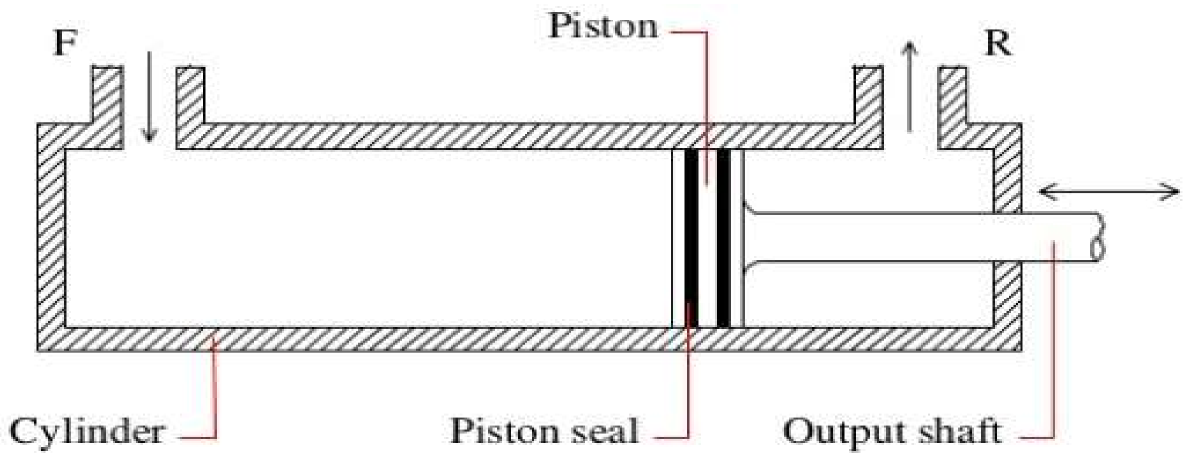
Linear Actuator: Single Acting Cylinder

In single acting cylinder actuator, there is only one opening for the fluid to enter and leave. The forward motion is due to the hydraulic pressure of the fluid entering through the opening, while the reverse motion is due to the spring.



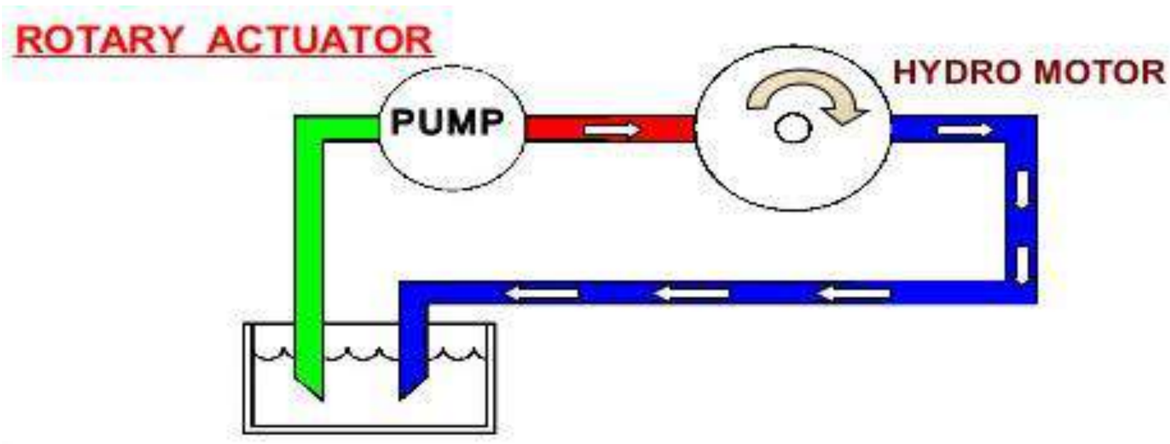
Double Acting Cylinder

In double acting cylinder actuator, there are two openings for the fluid to enter and leave. The forward motion is due to the hydraulic pressure of the fluid entering through one opening, while the reverse motion is due to the hydraulic pressure of the fluid entering through one opening.



Rotary Actuator

In rotary actuator, the linear motion of fluid inside the pipe is converted into rotary motion by a hydro motor.

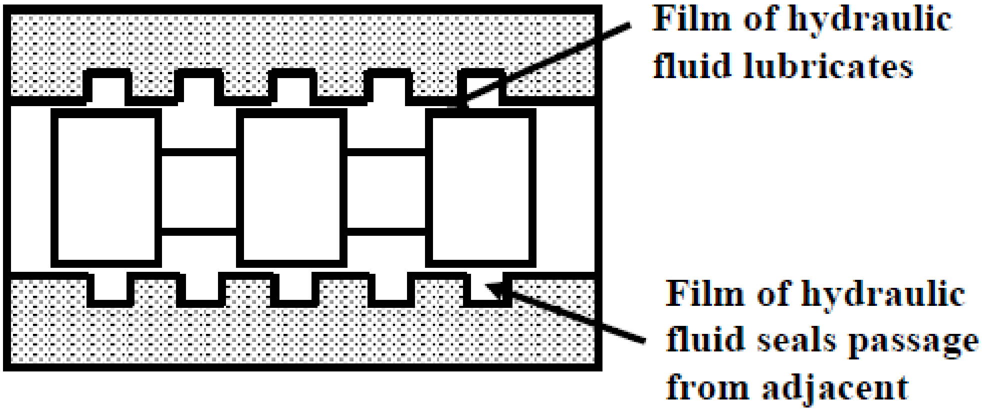


Components of Hydraulic Actuation Systems

Hydraulic Fluid

- Hydraulic Fluid transmit the fluid power from the reservoir to the actuator cylinder.
- Hydraulic fluid must be essentially non-compressible to be able to transmit power instantaneously from one part of the system to another.
- It should lubricate the moving parts to reduce friction loss and cool the components so that the heat generated does not lead to fire hazards.
- It also helps in removing the contaminants to filter.
- The other desirable property of oil is its lubricating ability.
- The fluid also acts as a seal against leakage inside a hydraulic component.

- The most common liquid used in hydraulic systems is petroleum oil because it is only very slightly compressible.

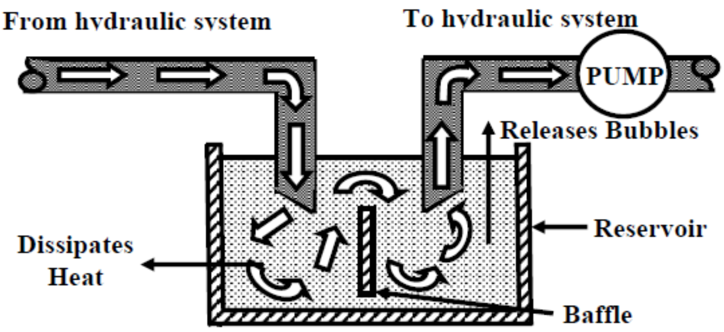


The Fluid Delivery Subsystem

It consists of the components that hold and carry the fluid from the pump to the actuator. It consists of reservoir, filter, line, fittings and seals, pump, motor, accumulator etc.

Reservoir

It holds the hydraulic fluid to be circulated and allows air entrapped in the fluid to escape. This is an important feature as the bulk modulus of the oil, which determines the stiffness of hydraulic system, deteriorates considerably in the presence of entrapped air bubbles. It also helps in dissipating heat.



Filter

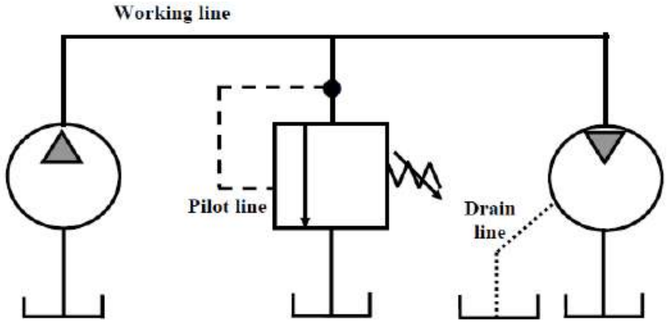
The hydraulic fluid is kept clean in the system with the help of filters and strainers. It removes minute particles from the fluid, which can cause blocking of the orifices of servo-valves or cause jamming of spools.

Line

- Pipe, tubes and hoses, along with the fittings or connectors, constitute the conducting lines that carry hydraulic fluid between components.
- Lines convey the fluid and also dissipate heat.
- Lines are one of the disadvantages of hydraulic system that we need to pay in return of higher power to weight ratio. They need return path
- In contrast, for Pneumatic Systems, no return path for the fluid, which is air, is needed, since it can be directly released into the atmosphere.

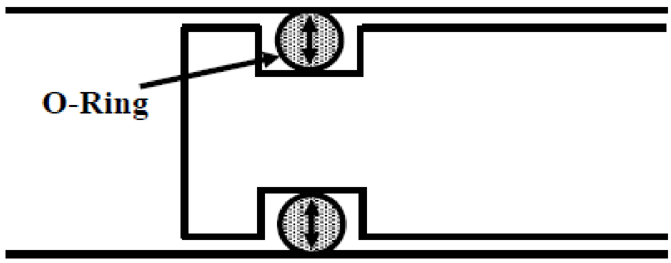
There are various kinds of lines in a hydraulic system.

- The working lines carry the fluid that delivers the main pump power to the load.
- The pilot lines carry fluid that transmit controlling pressures to various directional and relief valves for remote operation or safety.
- Drain lines carry the fluid that inevitably leaks out, to the tank.



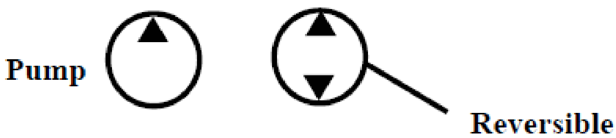
Fittings and Seals

- Various additional components are needed to join pipe or tube sections, create bends and also to prevent internal and external leakage in hydraulic systems.
- Although some amount of internal leakage is built-in, to provide lubrication, excessive internal leakage causes loss of pump power since high pressure fluid returns to the tank, without doing useful work.
- External leakage, on the other hand, causes loss of fluid and can create fire hazards, as well as fluid contamination.
- Various kinds of sealing components are employed in hydraulic systems to prevent leakage.



Hydraulic Pumps

The pump converts the mechanical energy of its prime-mover to hydraulic energy by delivering a given quantity of hydraulic fluid at high pressure into the system. Generically, all pumps are divided into two categories, namely, hydrodynamic or non-positive displacement and hydrostatic or positive displacement. Hydraulic systems generally employ positive displacement pumps only.



Hydrostatic or Positive Displacement Pumps: These pumps deliver a given amount of fluid for each cycle of motion, that is, stroke or revolution. Their output in terms of the volume flow rate is solely dependent on the speed of the prime-mover and is independent of outlet pressure notwithstanding leakage. There are various types of pumps used in hydraulic systems such as Gear Pumps, Vane Pumps, Piston Pumps, Radial Piston Pumps etc.

Motors

Motors work exactly on the reverse principle of pumps. In motors fluid is forced into the motor from pump outlets at high pressure. This fluid pressure creates the motion of the motor shaft and finally go out through the motor outlet port and return to tank. All three variants of motors, already described for pumps, namely Gear Motors, Vane Motors and Piston motors are in use.

Accumulators

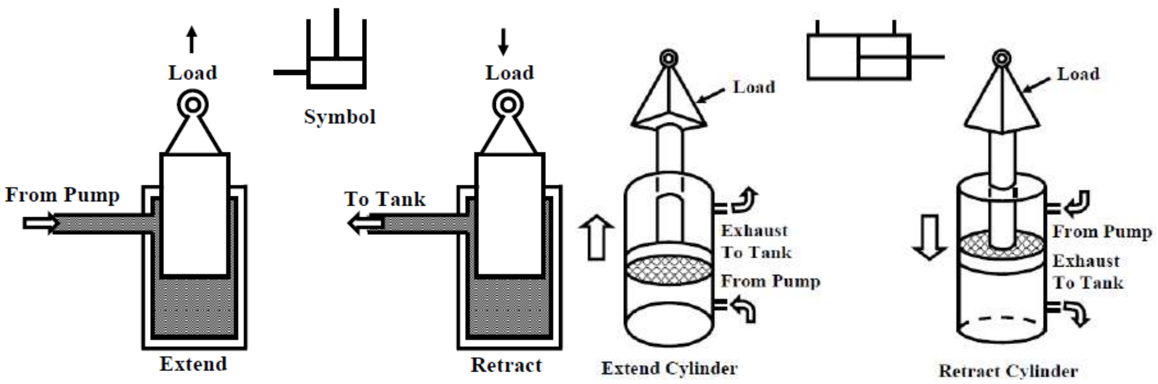
Unlike gases the fluids used in hydraulic systems cannot be compressed and stored to cater to sudden demands of high flow rates that cannot be supplied by the pump. An accumulator in a hydraulic system provides a means of storing these incompressible fluids under pressure created either by a spring, or compressed gas. Any tendency for pressure to drop at the inlet causes the spring or the gas to force the fluid back out, supplying the demand for flow rate.

There are two types of accumulators: Spring-Loaded Accumulators, Gas Charged Accumulator

Cylinders

Cylinders are linear actuators, that is, they produce straight-line motion and/or force. Cylinders are classified as single-or double-acting with the graphical symbol for each type.

There are two types of cylinders such as Single Acting Cylinder and Double-Acting Cylinder



Advantages of Hydraulic Actuator Systems

- Variable Speed and Direction: The actuator (linear or rotary) of a hydraulic system, can be driven at speeds that vary by large amounts and fast, by varying the pump delivery or using a flow control valve. A hydraulic actuator can be reversed instantly while in full motion without damage. This is not possible for most other prime movers.
- Power-to-weight ratio: Hydraulic components, because of their high speed and pressure capabilities, can provide high power output with vary small weight and size.
- Stall Condition and Overload Protection: A hydraulic actuator can be stalled without damage when overloaded, and will start up immediately when the load is reduced. The pressure relief valve in a hydraulic system protects it from overload damage. During stall, or when the load pressure exceeds the valve setting, pump delivery is directed to tank with definite limits to torque or force output.
- Hydraulic actuators are rugged and suited for high-force applications. They can produce forces 25 times greater than pneumatic cylinders of equal size. They also operate in pressures of up to 4,000 psi.
- Hydraulic motors have high horsepower-to-weight ratio by 1 to 2 hp/lb greater than a pneumatic motor.
- A hydraulic actuator can hold force and torque constant without the pump supplying more fluid or pressure due to the incompressibility of fluids
- Hydraulic actuators can have their pumps and motors located a considerable distance away with minimal loss of power.

Drawbacks

- Hydraulics will leak fluid. Like pneumatic actuators, loss of fluid leads to less efficiency. However, hydraulic fluid leaks lead to cleanliness problems and potential damage to surrounding components and areas.
- Hydraulic actuators require many companion parts, including a fluid reservoir, motors, pumps, release valves, and heat exchangers, along with noise-reduction equipment. This makes for linear motions systems that are large and difficult to accommodate.

Pneumatic Actuating System

- Devices used for converting pressure energy of compressed air into the mechanical energy to perform useful work.
- Actuators are used to perform the task of exerting the required force at the end of the stroke or used to create displacement by the movement of the piston.
- The pressurized air from storage is supplied to pneumatic actuator to do work.
- The air cylinder is a simple and efficient device for providing linear thrust or straight line motions.
- Pneumatic actuator consists of a piston and rod moving inside a closed cylinder.
- This actuator can be sub-divided into two types based on the operating principle: single acting and double acting.

The principle is based on the concept of pressure as force per unit area.

If we imagine that a net pressure difference is applied across a diaphragm of surface area A, then a net force acts on the diaphragm given by $F= A(P_1-P_2)$.

P_1-P_2 = pressure difference (Pa)

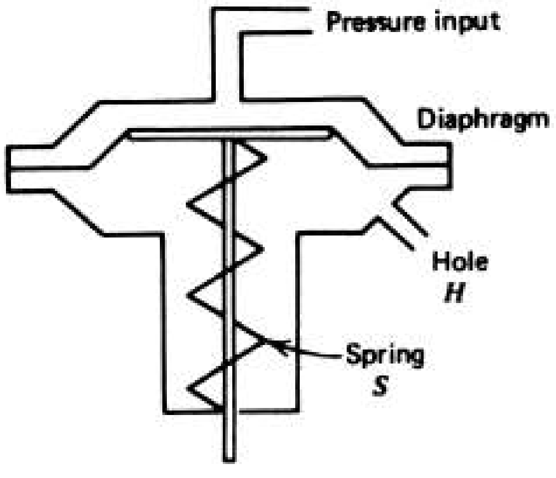
A = diaphragm area (m^2)

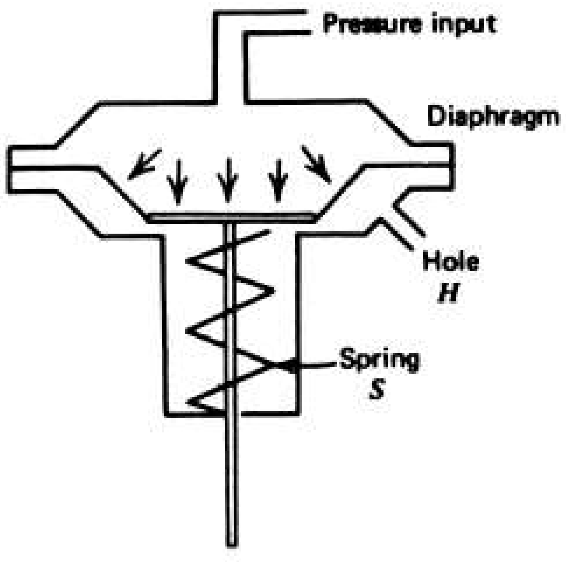
F = force (N)

If we need to double the available force for a given pressure, it is merely necessary to double the diaphragm area.

Very large forces can be developed by standard signal-pressure ranges of 3 to 15 psi (20 to 100 kPa).

Many types of pneumatic actuators are available, but perhaps the most common are those associated with control valves.

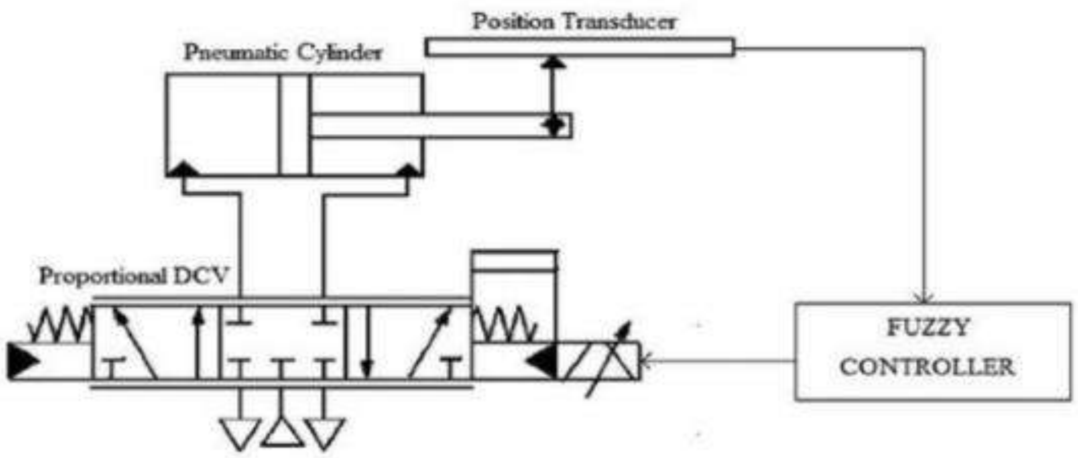




Widely used in automation filed because of its low weight and compact size.

Control of pneumatics system is slightly difficult as compare to hydraulic system because of compressed air governing equations of pneumatics are nonlinear in nature.

More used in automation industry as discrete device such as gripper

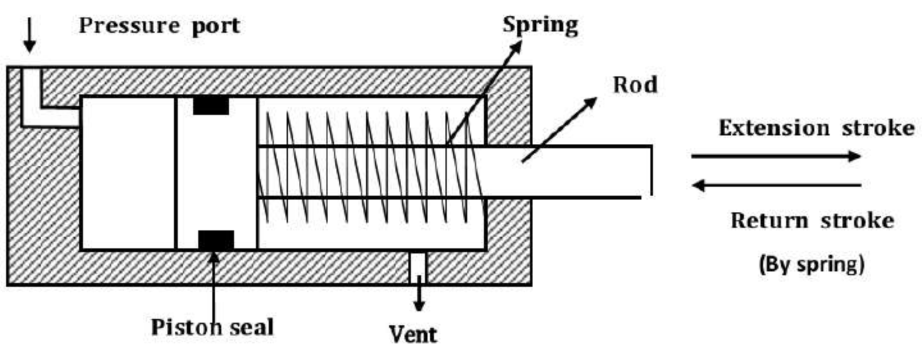


Types of Pneumatics Actuators are:

- Linear Actuator or Pneumatic cylinders
    - Single acting cylinder
    - Double acting cylinder
- Rotary Actuator or Air motors
- Limited angle Actuators

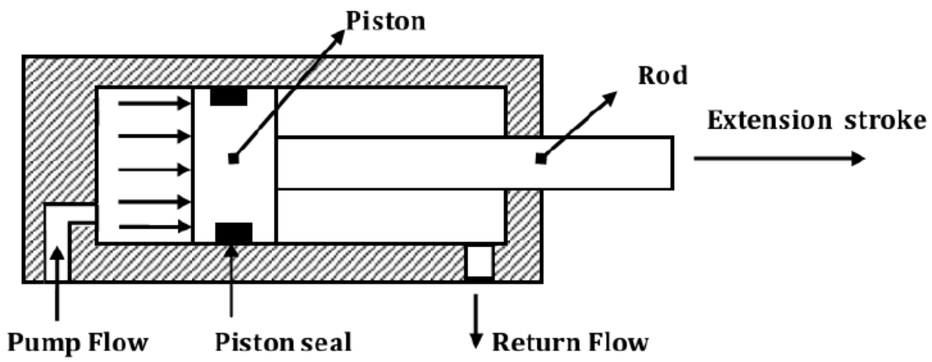
Single Acting Cylinder

- Cylinders have single air inlet line- One working port.
- Forward motion of the piston is obtained by supplying compressed air to working port while return motion of piston is obtained by spring placed on the rod side of the cylinder.
- Used where force is required to be exerted only in one direction such as clamping, feeding, sorting, locking, ejecting, braking etc.
- Available in short stroke lengths [maximum length up to 80 mm] due to the natural length of the spring.
- Single acting cylinders require only about half the air volume consumed by a double acting cylinder for one operating cycle.

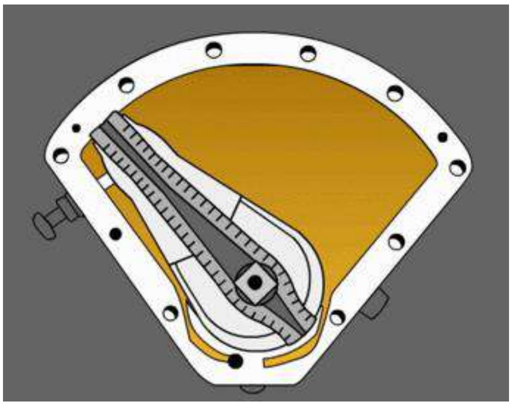


Double Acting Cylinder

- Cylinders have two air inlet lines- Two working ports- one on the piston side and the other on the rod-side.
- To achieve forward motion of the cylinder, compressed air is admitted on the piston side and the rod side is connected to exhaust.
- During return motion supply air admitted at the rod side while the piston side volume is connected to the exhaust.
- Force is exerted by the piston both during forward and return motion of cylinder.
- Double acting cylinders are available in diameters from few mm to around 300 mm and stroke lengths of few mm up to 2 meters.



Limited Angle Actuators



Components of Pneumatic Actuating System

Compressor: Acts as power source for pneumatic system.

Air Treatment Unit: Consist of Filter Regulator Lubricator(FRL) unit which comprises of filter to filter the compressed air, a pressure regulator to regulate a flow of compressed air and a lubricator to lubricate air.

Valve: Controls the direction and regulates the air flowing to actuating system

Actuator: Convert the pneumatic force or compressed air pressure into mechanical fore or movement.

Pipes: Used to connect various elements of actuating system with each other.

Sensors: It act as feedback element for a controller in pneumatic system.

Controller: Monitors whole system to provide required output.

Applications

- Tied rod cylinders,
- Rotary actuators, grippers,
- Rod-less actuators with magnetic linkage or rotary cylinders,
- Rod-less actuators with mechanical linkage,
- Pneumatic artificial muscles,
- Speciality actuators that combine rotary and linear motion (frequently used for clamping operations)
- Vacuum generators.

Advantages

- High force rating
- Pneumatic actuators generate precise linear motion by providing accuracy, for example, within 0.1 inches and repeatability within .001 inches.
- Pneumatic actuators typical applications involve areas of extreme temperatures. A typical temperature range is -40°F to 250°F.
- In terms of safety and inspection, by using air, pneumatic actuators avoid using hazardous materials. They meet explosion protection and machine safety requirements because they create no magnetic interference due to their lack of motors.
- In recent years, pneumatics has seen many advances in miniaturization, materials, and integration with electronics and condition monitoring.
- The cost of pneumatic actuators is low compared to other actuators.
- Pneumatic actuators are also lightweight, require minimal maintenance, and have durable components that make pneumatics a cost-effective method of linear motion.

Drawbacks

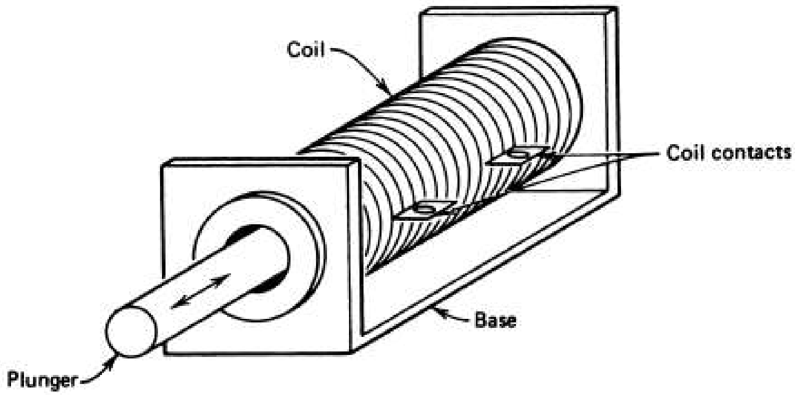
- Pressure losses and air's compressibility make pneumatics less efficient than other linear-motion methods. Compressor and air delivery limitations mean that operations at lower pressures will have lower forces and slower speeds. A compressor must run continually operating pressure even if nothing is moving.
- To be truly efficient, pneumatic actuators must be sized for a specific job. Hence, they cannot be used for other applications. Accurate control and efficiency requires proportional regulators and valves, but this raises the costs and complexity.
- Even though the air is easily available, it can be contaminated by oil or lubrication, leading to downtime and maintenance. Companies still have to pay for compressed air, making it a consumable, and the compressor and lines are another maintenance issue.

Electromechanical/Electrical Actuating System

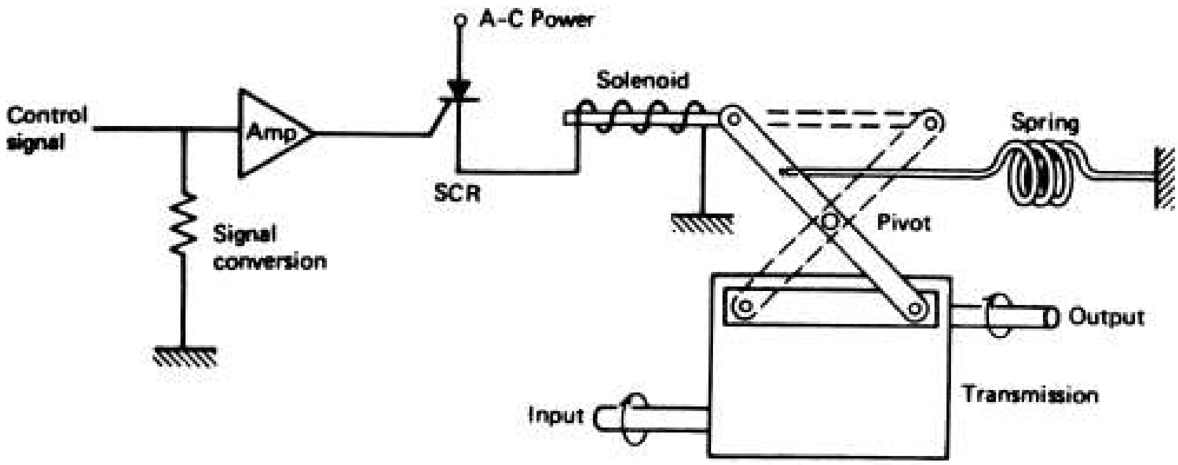
- This type of actuators converts electrical energy into mechanical energy.
- There are different ways to achieve this conversion. Includes all motors, responsible for moving or controlling a mechanism.
- One of the popular method which is used by many electromechanical actuators is by generating magnetic field in which "when current carrying conductor is placed in magnetic field it experiences mechanical force" - Lorentz's Law of Electromagnetic Force. This generated force will further be converted into mechanical motion.
- DC and AC motors, solenoids, voice coils, Active materials (piezoelectric, electrostrictive etc.) and MEMS.

Solenoid

- A solenoid is an elementary device that converts an electrical signal into mechanical motion, usually rectilinear.
- Consists of a coil and plunger. The plunger may be freestanding or spring loaded.
- The coil will have some voltage or current rating and may be dc or ac.
- Solenoid specifications include the electrical rating and the plunger pull or push force when excited by the specified voltage.
- Are used when a large, sudden force must be applied to perform some job.



Solenoid is used to change the gears of a two-position transmission as shown below.



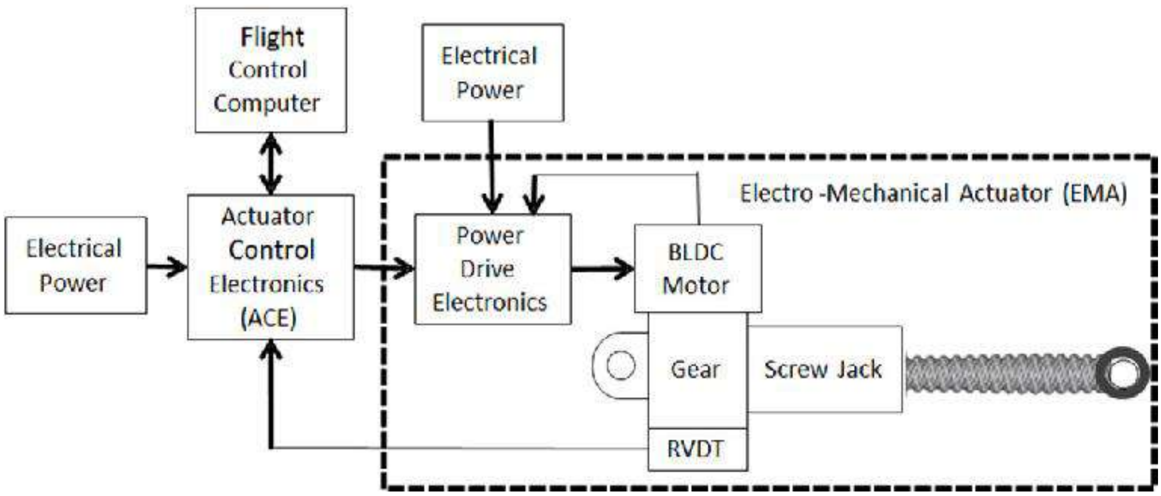
Electrical Motors

- Electrical motors are devices that accept electrical input and produce a continuous rotation as a result.
- Motor styles and sizes vary as demands for rotational speed (revolutions per minute, or rpm), starting torque, rotational torque, and other specifications vary.
- Most common control situation is where motor speed drives some part of a process, and must be controlled to control some variable in the process- e.g.: the drive of a conveyor system
- There are many types of electrical motors, each with its special set of characteristics- dc motor, ac motor, stepping motor etc.

Components of Electric Motor

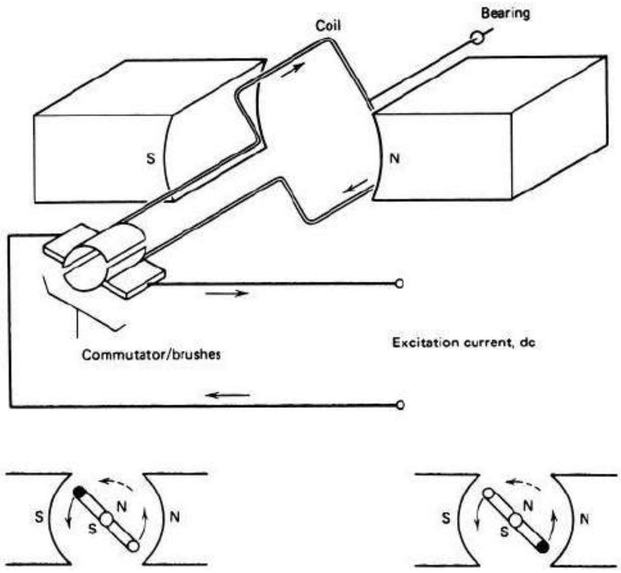
- Electric Motor- Electric motors has wide applications in the industrial as well as in automation sector and popularly used in robotics application because of its high speed of response, finer speed control capability. The magnetic field of stator interacts with magnetic field of rotor to produce rotating torque. There are two basic types of electric motor - AC & DC motor which consist of basic element such as
- Stator– It is stationery outer part of motor made up of permanent magnet or coli winding.
- Rotor – It is rotating part of motor made up of permanent magnet or ferromagnetic coil.
- Armature–Rotor winding that generates rotor magnetic field.
- Field Coil – Part of stator which generates stator magnetic field.
- Air Gap – Small gap between stator and rotor where two magnetic fields interact with each other to generate mechanical torque.

Block Diagram of Motor Based Electrical Actuator System



DC Motor

The rotation of a dc motor is produced by the interaction of two constant magnetic fields.



Many dc motors use an electromagnet instead of a PM to provide the static field. The coil used to produce this field is called the field coil. This kind of dc motor is called a wound field motor. There are different types of wound field DC motor based on the way in which the field winding is connected.

- Series Field DC motors: This motor has large starting torque but is difficult to speed control. Good in applications for starting heavy, non-mobile loads and where speed control is not important, such as for quick-opening valves.
- Shunt Field DC motors: This motor has a smaller starting torque, but good speed-control characteristics produced by varying armature excitation current. Good in applications where speed is to be controlled, such as in conveyor systems.
- Compound Field DC motors: This motor attempts to obtain the best features of both of the two previous types. Generally, starting torque and speed-control capability fall predictably between the two pure cases.

Applications of DC Motor

- The use of dc motors in control systems ranges from very low energy, delicate control applications, to heavy-duty control operations in elevators and vehicles.
- In general, PM types are used for motors of less than 10 hp (7.5 kW) and wound field types for units up to about 100 to 200 hp (75 to 150 kW).
- Control of the speed and torque of these large machines requires very high power dc electricity. Such power is derived from the power electronics devices.
- In general, three-phase ac power is rectified using switching technology to produce the required high-voltage, high-current dc electricity.
- Control is often made possible by variation of the voltage amplitude.

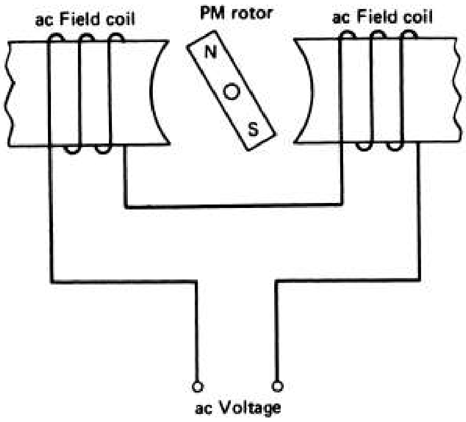
AC Motors

The basic operating principle of ac motors still involves the interaction between two magnetic fields. In this case, however, both fields are varying in time in consonance with the ac excitation voltage. Therefore, the force between the fields is a function of the angle of the rotor but also the phase of the current passing through the coils.

There are two basic types of ac motors, synchronous and induction. The primary motor for application to the control industry is the induction motor.

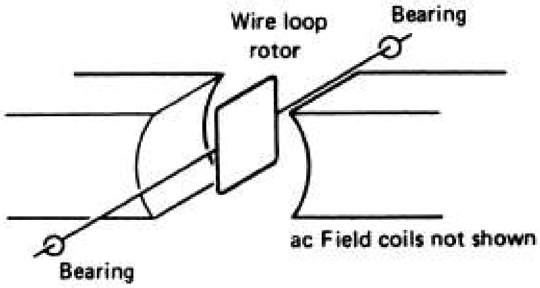
Synchronous Motor

- In a synchronous motor the ac voltage is applied to the field coils, called the stator in an ac motor.
- This means the magnetic field is changing in time in phase with the impressed ac voltage.
- The armature, called the rotor for ac motors, is either a permanent magnet or a dc electromagnet, and possesses a fixed magnetic field.
- Synchronous motors can be operated using single-phase ac but such units are used for only very low power (0.1 hp) and suffer from very low starting torque.
- When operated from three-phase, ac synchronous motors can be operated at very high power, up to 50,000 hp.



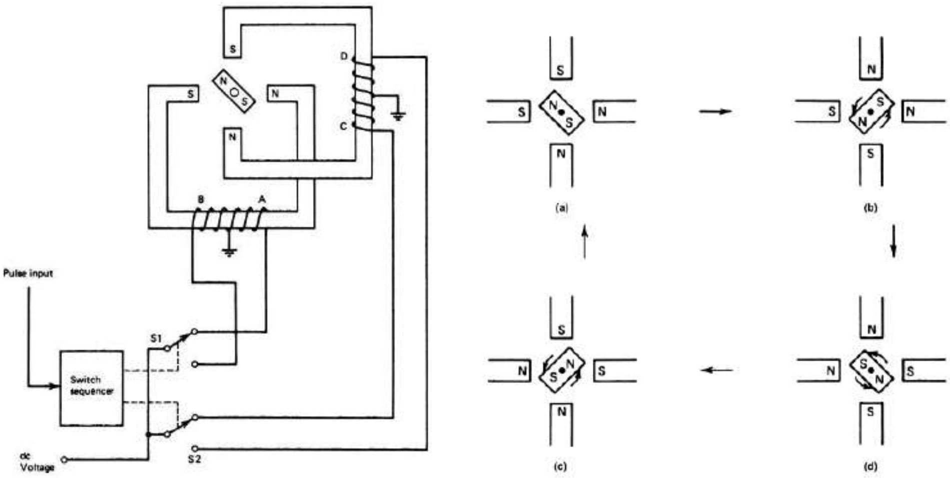
Induction Motors

- Induction ac motors are characterized by a rotor which is neither a PM nor a dc excited electromagnet.
- Instead current induced in a coil wound on the rotor generates the interacting magnetic field of the rotor.
- This current is induced from the stator coils.
- Single-phase induction motors are used for applications of relatively low power, say less than 5 hp (< 3.7 kW). Such motors are typical of those found in household appliances, for example.
- For higher power we use three-phase ac excitation. Such motors are available up to 10,000 hp.



Stepping Motor

- The stepping motor has increased in importance in recent years because of the ease with which it can be interfaced with digital circuits.
- A stepping motor is a rotating machine that actually completes a full rotation by sequencing through a series of discrete rotational steps.
- Each step position is an equilibrium position in that, without further excitation, the rotor position will stay at the latest step.
- Thus, continuous rotation is achieved by the input of a train of pulses, each of which causes an advance of one step.
- It is not really continuous rotation, but discrete, stepwise rotation.
- The rotational rate is determined by the number of steps per revolution and the rate at which the pulses are applied.
- A driver circuit is necessary to convert the pulse train into proper driving signals for the motor.



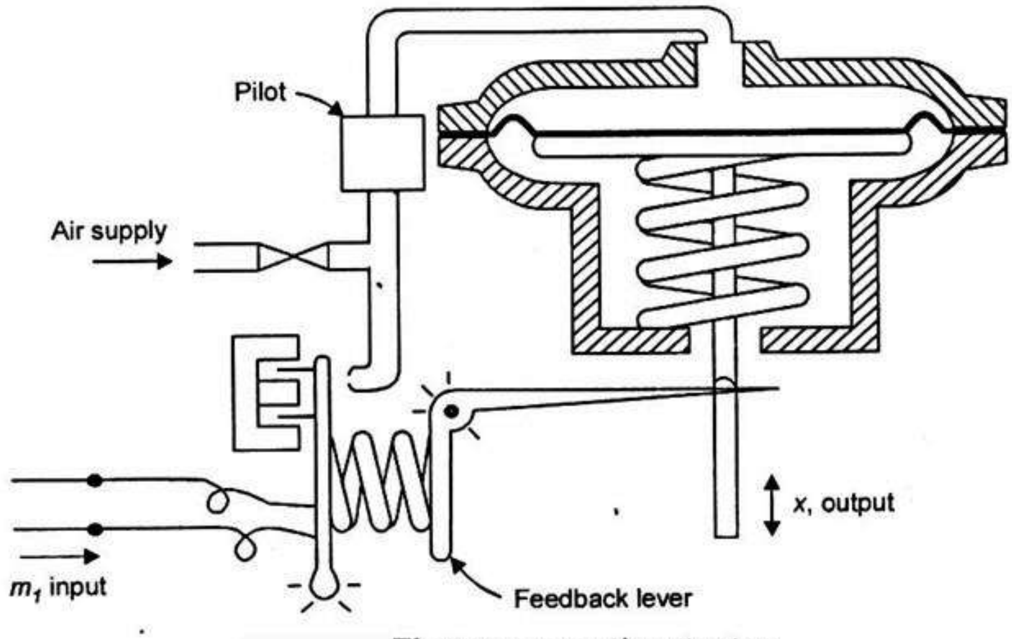
- Most common stepper motor does not use a PM, but rather a rotor of magnetic material (not a magnet) with a certain number of teeth.
- This rotor is driven by a phased arrangement of coils with a different number of poles so that the rotor can never be in perfect alignment with the stator.
- The direction of rotation of stepper motors can be changed just by changing the order in which different poles are activated and deactivated.

Electro-Pneumatic Actuator

Electro pneumatics is now commonly used in many areas of Industrial low cost automation. They are also used extensively in production, assembly, pharmaceutical, chemical and packaging systems.

There is a significant change in controls systems. Relays have increasingly been replaced by the programmable logic controllers in order to meet the growing demand for more flexible automation.

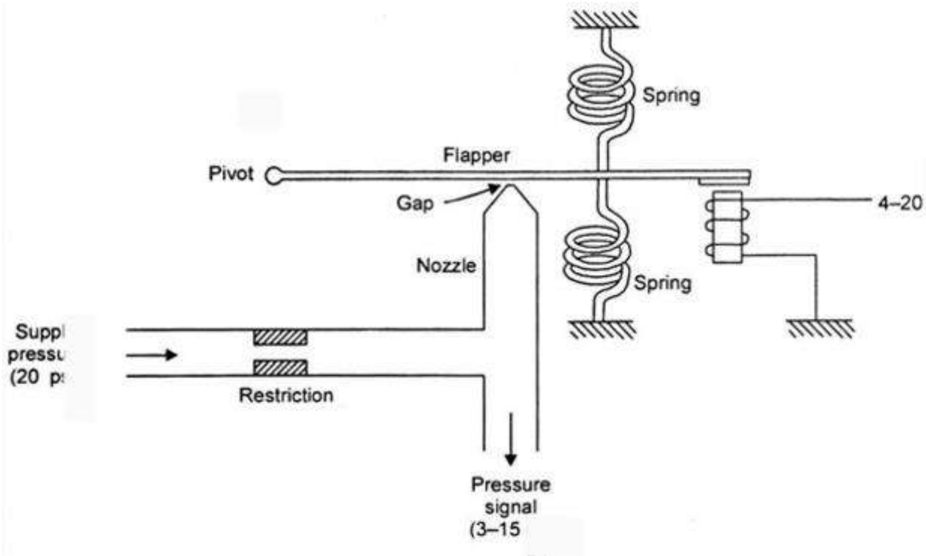
Electro-pneumatic control consists of electrical control systems operating pneumatic power systems. In this solenoid valves are used as interface between the electrical and pneumatic systems. Devices like limit switches and proximity sensors are used as feedback elements.



Electro-pneumatic actuator

Electro Pneumatic control integrates pneumatic and electrical technologies, is more widely used for large applications. In Electro Pneumatics, the signal medium is the electrical signal either AC or DC source is used. Working medium is compressed air. Operating voltages from around 12 V to 220 Volts are often used. The final control valve is activated by solenoid actuation The resetting of the valve is either by spring [single Solenoid] or using another solenoid [Double solenoid Valve]. More often the valve actuation/reset is achieved by pilot assisted solenoid actuation to reduce the size and cost of the valve

Control of Electro Pneumatic system is carried out either using combination of Relays and Contactors or with the help of Programmable Logic Controllers [PLC]. A Relay is often used to convert signal input from sensors and switches to number of output signals [either normally closed or normally open]. Signal processing can be easily achieved using relay and contactor combinations A Programmable Logic Controller can be conveniently used to obtain the out puts as per the required logic, time delay and sequential operation. Finally, the output signals are supplied to the solenoids activating the final control valves which controls the movement of various cylinders.



In electro-pneumatic actuators, the current in the solenoid is converted into pressure ***(I to P conversion)*** using the arrangement shown above. When an electrical actuating current of 4-20 mA is passed through the solenoid, it attracts the flapper fixed on a pivot. So, when the solenoid is activated, the flapper moves as per the magnitude of current through the solenoid, which will control the opening of the nozzle. Depending on the opening and closing of nozzle, the pressure signal is varied between 3-15 psi. When the electrical signal in the solenoid is 4 mA, the nozzle output is minimum, thus the pressure signal is maximum (15psi). When the current signal in the solenoid is maximum, i.e., 20 mA, the nozzle is opened to maximum leading to maximum fluid flow through the nozzle and thus minimum pressure (3 psi) of the pressure signal.

The greatest advantage of electro pneumatics is the integration of various types of proximity sensors [electrical] and PLC for very effective control. As the signal speed with electrical signal, can be much higher, cycle time can be reduced and signal can be conveyed over long distances.

In Electro pneumatic controls, mainly three important steps are involved:

- Signal input devices -Signal generation such as switches and contactor, Various types of contact and proximity sensors
- Signal Processing – Use of combination of Contactors of Relay or using Programmable Logic Controllers
- Signal Out puts – Out puts obtained after processing are used for activation of solenoids indicators or audible alarms

Seven basic electrical devices used in electro pneumatic actuators are

- Manually actuated push button switches
- Limit switches
- Pressure switches
- Solenoids
- Relays
- Timers
- Temperature switches

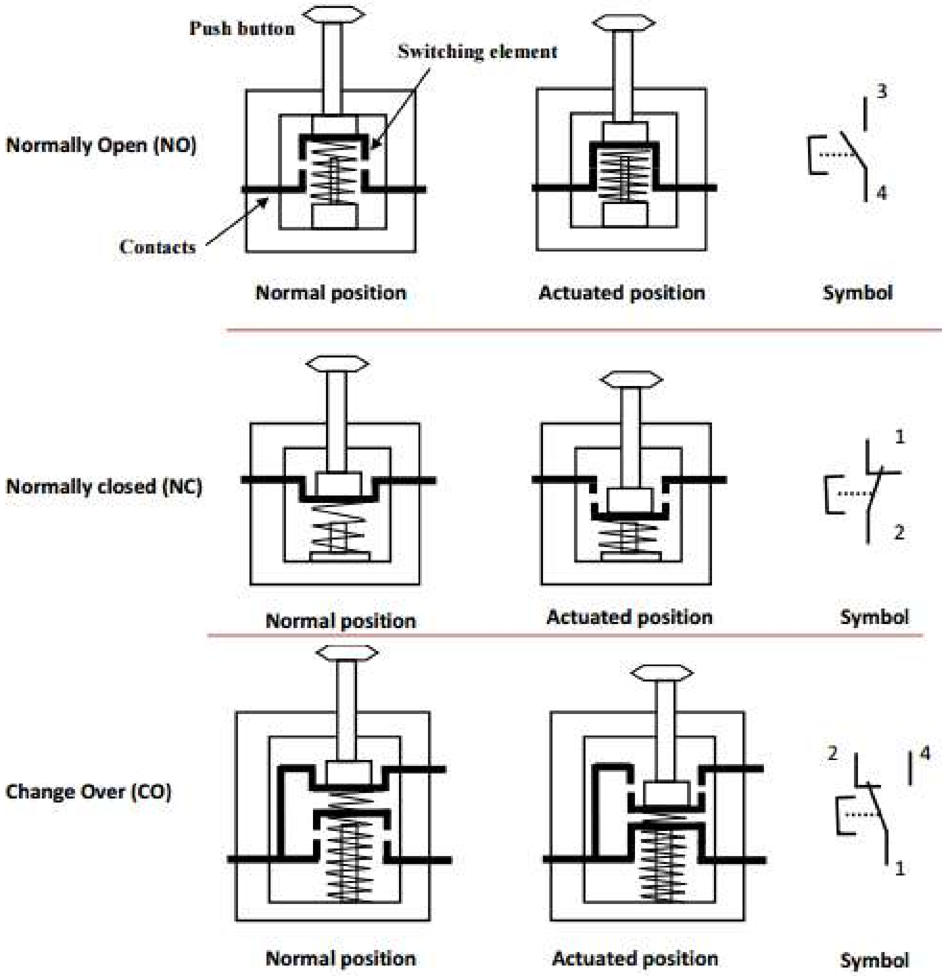
Push button switches: A push button is a switch used to close or open an electric control circuit. They are primarily used for starting and stopping of operation of machinery. They also provide manual override when the emergency arises. Push button switches are actuated by pushing the actuator into the housing. This causes set of contacts to open or close.

Push buttons are of two types

- Momentary push button: return to their unactuated position when they are released
- Maintained contact or detent push button: Has a latching mechanism to hold it in the selected position

The contact of the push buttons, distinguished according to their functions,

i. Normally open (NO) type
ii. Normally closed (NC) type
iii. Change over (CO) type.



Limit Switches: Any switch that is actuated due to the position of a fluid power component (usually a piston rod or hydraulic motor shaft or the position of load is termed as limit switch. The actuation of a limit switch provides an electrical signal that causes an appropriate system response. Limit switches perform the same function as push button switches. Push buttons are manually actuated whereas limit switches are mechanically actuated.

There are two types classification of Limit switches depending upon method of actuations of contacts

i. Lever actuated contacts
ii. Spring loaded contacts

Pressure switches: A pressure switch is a pneumatic-electric signal converter. Pressure switches are used to sense a change in pressure, and opens or closes an electrical switch when a predetermined pressure is reached. Bellow or diaphragm is used to sense the change of pressure. Bellows or Diaphragm is used to expand or contract in response to increase or decrease of pressure. Figure 1.3 shows a diaphragm type of pressure switch. When the pressure is applied at the inlet and when the pre-set pressure is reached, the diaphragm expands and pushes the spring loaded plunger to make/break contact.

Solenoids: Electrically actuated directional control valves form the interface between the two parts of an electro-pneumatic control.

The most important tasks of electrically actuated DCVs include.

- Switching supply air on or off
- Extension and retraction of cylinder drives

Electrically actuated directional control valves are switched with the aid of solenoids. They can be divided into two groups:

- Spring return valves only remain in the actuated position as long as current flows through the solenoid
- Double solenoid valves retain the last switched position even when no current flows through the solenoid.

Relays: A relay is an electro magnetically actuated switch. It is a simple electrical device used for signal processing. Relays are designed to withstand heavy power surges and harsh environment conditions. When a voltage is applied to the solenoid coil, an electromagnet field results. This causes the armature to be attracted to the coil core. The armature actuates the relay contacts, either closing or opening them, depending on the design. A return spring returns the armature to its initial position when the current to the coil is interrupted.

Timer or Time delay relays: Timers are required in control systems to effect time delay between work operations. This is possible by delaying the operation of the associated control element through a timer. Most of the timers we use is Electronic timers.

There are two types of time relay

i. Pull in delay (on –delay timer)
ii. Drop –out delay (off delay timer)

Temperature Switch: Temperature switches automatically senses a change in temperature and opens or closes an electrical switch when a predetermined temperature is reached. This switch can be wired either normally open or normally closed.

Temperature switches can be used to protect a fluid power system from serious damage when a component such as a pump or strainer or cooler begins to malfunction.

Electro-pneumatic actuator can be briefed as below:

- Integrates pneumatic and electrical technologies, consists of electrical control systems operating pneumatic power systems.
- The signal medium is the electrical signal either AC or DC source is used. Working medium is compressed air.
- In this solenoid valves are used as interface between the electrical and pneumatic systems. i.e.: The final control valve is activated by solenoid actuation.
- Operating voltages from around 12 V to 220 Volts are often used.
- Devices like limit switches and proximity sensors are used as feedback elements.
- More widely used for large applications.

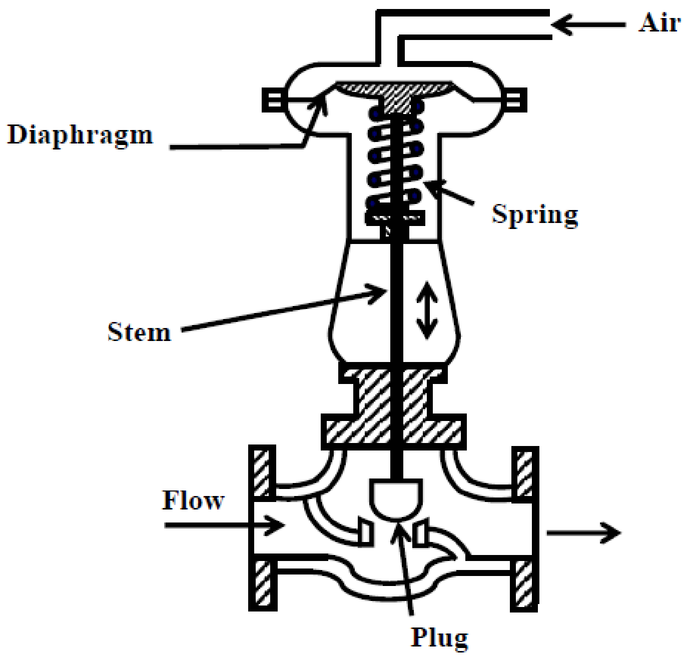
Types of Actuators and Their Characteristics

Criteria	**Mechanical actuators**	**Electromechanical actuators**	**Pneumatics actuators**	**Hydraulics actuators**
I/O energy source	Electric Motor, IC Engines.	Water/gas turbine	Presser tank/Compressor.	IC engines, electric motor, Hydraulic Pump
Energy transmission Element	Gears, shafts, levers.	Electric cables	Pipes & hoses	Pipe & hoses
Speed of response	Fair	Best	Good	Good
Control	Fair	Best	Fair	Good
Cost	High cost	Very cheap	Cheap	Very costly
Output motion	Liner/rotary	Mostly Rotary	Liner/rotary	Liner/rotary

Actuator types	Advantages	Disadvantages
Electrical(servomotor or stepping motor)	Direct interface with computer system. Simple design.	Low thrust. Slow speed. No mechanical fail safe hazardous.
Electromechanical(Motors combined with gear boxes)	High thrust High stiffness coefficient Flexible adaptation	Complex design No mechanical fail safe Large, heavy structure Hazardous
Hydraulic and Electro hydraulic	High thrust Fast speed High stiffness coefficient Self lubrication	Complex design Large, heavy structure Hazardous Fluid Viscosity Sensitive
Pneumatic and Electro-pneumatic	Low cost Mechanical Fail safe Simple design Small package Suitable for highly hazardous areas also Good control with control device	Slow speed Lack of stiffness Instability Moderate trust Quality air requirement

Control Valves

- The control action in any control loop system, is executed by the final control element.
- The most common type of final control element used in chemical and other process control is the control valve.
- A control valve is normally driven by a diaphragm type pneumatic actuator that throttles the flow of the manipulating variable for obtaining the desired control action.
- A control valve essentially consists of a plug and a stem.
- The stem can be raised or lowered by air pressure and the plug changes the effective area of an orifice in the flow path.
- When the air pressure increases, the downward force of the diaphragm moves the stem downward against the spring.



Classification of Control Valves

Control valves are available in different types and shapes.

They can be classified in different ways based on

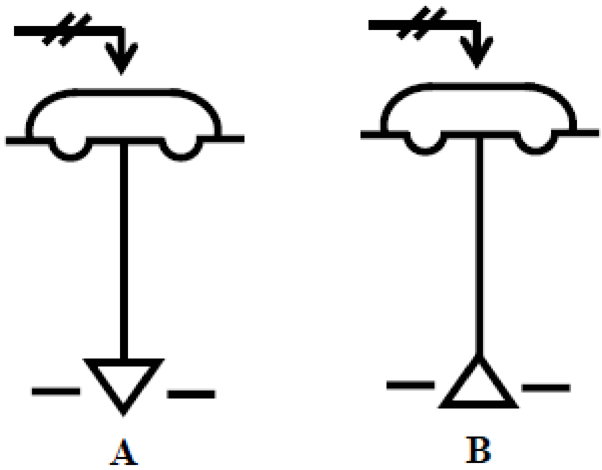
- Action
- Number of plugs
- Flow characteristics.

<u>Classification of Control Valves based on Action</u>

Air to Open: If the air supply fails, the control valve will be fully closed

Air to Close: If the air supply fails, the control valve will be fully open.

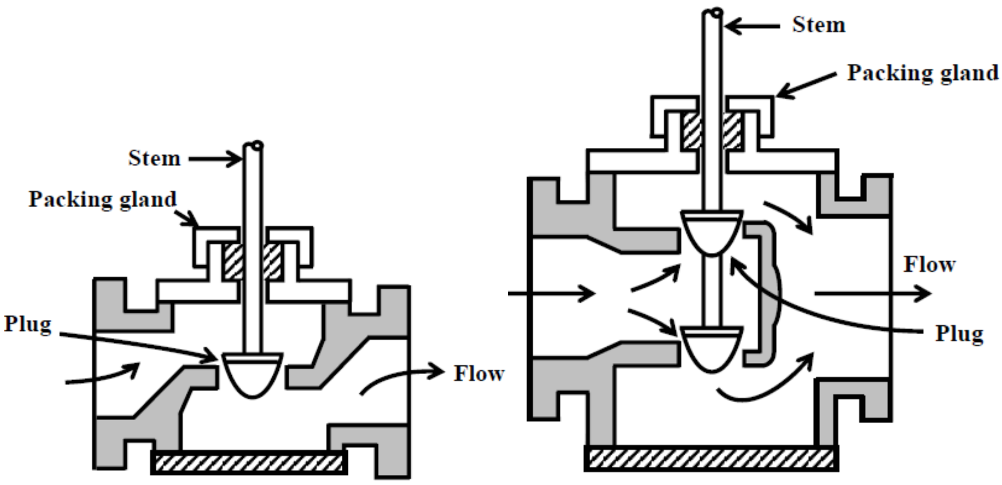
Control valves operated through pneumatic actuators can be either (i) air to open, or (ii) air to close. They are designed such that if the air supply fails, the control valve will be either fully open, or fully closed, depending upon the safety requirement of the process. For example, if the valve is used to control steam or fuel flow, the valve should be shut off completely in case of air failure. On the other hand, if the valve is handling cooling water to a reactor, the flow should be maximum in case of emergency. The schematic arrangements of these two actions are shown in Figure below. Valve A are air to close type, indicating, if the air fails, the valve will be fully open. Opposite is the case for valve B.



- **Fail open or Air to close : A**
- **Fail closed or Air to open : B**

Classification of Control Valves based on Number of Plugs

Control valves can also be characterized in terms of the number of plugs present, as single-seated valve and double-seated valve. The difference in construction between a single seated and double-seated valve are illustrated in Fig.



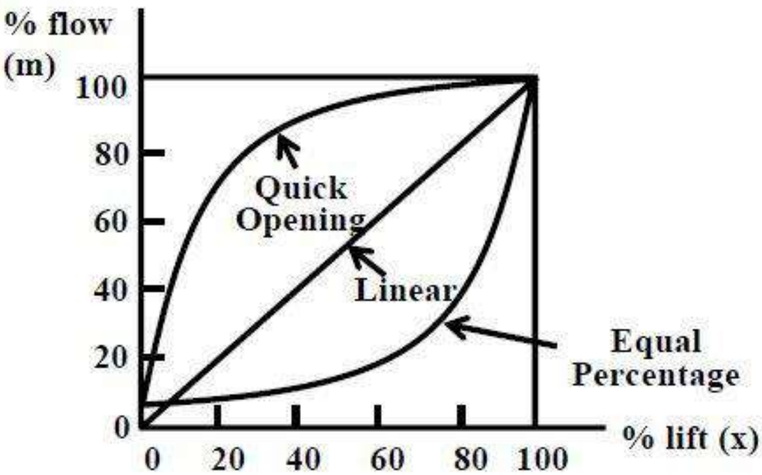
Here, only one plug is present in the control valve, so it is single seated valve. The advantage of this type of valve is that, it can be fully closed and flow variation from 0 to 100% can be achieved. But looking at its construction, due to the pressure drop across the orifice a large upward force is present in the orifice area, and as a result, the force required to move the valve against this upward thrust is also large. Thus this type of valves is more suitable for small flow rates. On the other hand, there are two plugs in a double-seated valve; flow moves upward in one orifice area, and downward in the other orifice. The resultant upward or downward thrust is almost zero. As a result, the force required to move a double-seated valve is comparatively much less.

But the double-seated valve suffers from one disadvantage. The flow cannot be shut off completely, because of the differential temperature expansion of the stem and the valve seat. If one plug is tightly closed, there is usually a small gap between the other plug and its seat. Thus, single-seated valves are recommended for when the valves are required to be shut off completely. But there are many processes, where the valve used is not expected to operate near shut off position. For this condition, double-seated valves are recommended.

Classification of Control Valves based on Flow Characteristics

Flow Characteristics describes how the flow rate changes with the movement or lift of the stem. The shape of the plug primarily decides the flow characteristics. The flow characteristic of a valve is normally defined in terms of Inherent characteristics and Effective characteristics. An inherent characteristic is the ideal flow characteristics of a control valve and is decided by the shape and size of the plug. When the valve is connected to a pipeline, its overall performance is decided by its effective characteristic.

Based on inherent/ideal characteristics, control valves are classified into three such as Quick opening, Linear and Equal Percentage.

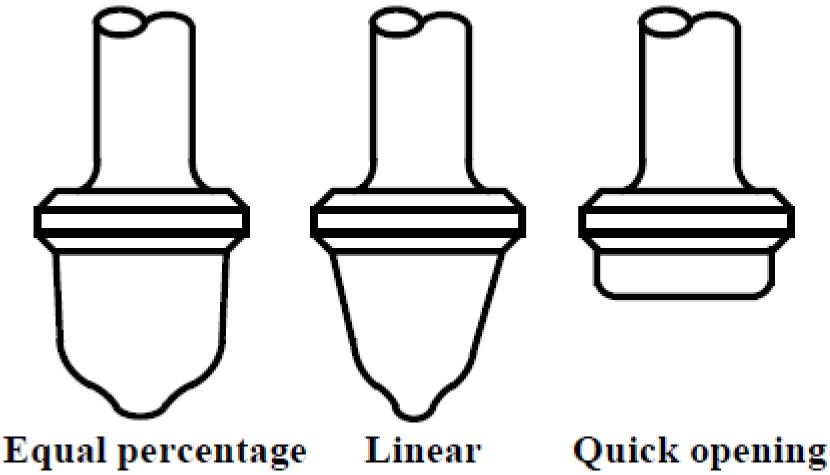


The shape of the plug decides, how the flow rate changes with the stem movement, or lift when the pressure drop across the valve is assumed to be held constant. This classification is basically done on relationship between the valve stem position and the flow rate through the valve. As the stem and plug move with respect to the seat/orifice, the shape of the plug determines the amount of actual opening of the valve.

Quick Opening Type: This type of valve is used predominantly for full ON/full OFF control applications. The valve characteristics shows that a relatively small motion of the valve stem results in maximum possible flow rate through the valve. Such a valve, for example, may allow 90% of maximum flow rate with only a 30% travel of the stem.

Linear Type: This type of valve has a flow rate that varies linearly with the stem position. It represents the ideal situation where the valve alone determines the pressure drop.

Equal Percentage Type: A very important type of valve employed in flow control has a characteristic such that a given percentage change in stem position produces an equivalent change in flow—that is, an equal percentage. Generally, this type of valve does not shut off the flow completely in its limit of stem travel. Thus, Q_{min} represents the minimum flow when the stem is at one limit of its travel. At the other extreme, the valve allows a flow Q_{max} as its maximum, open-valve flow rate.



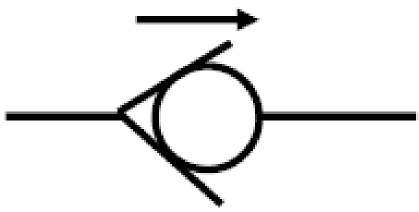
Directional Control Valves

In directional control valves, it is basically the directions of the flows that are controlled and not the magnitudes.

Directional valves can be characterized depending on the number of ports, the number of directions of flow that can be established, number of positions of the valve etc. They are mainly classified in terms of the number of flow directions, such as one-way, two way or four-way valves. Directional valves are often operated in selected modes using hydraulic pressure from remote locations. Such mechanisms are known as pilots. Thus a valve that may be blocking the flow in a certain direction in absence of pilot pressure, may be allowing flow, when pilot pressure is applied.

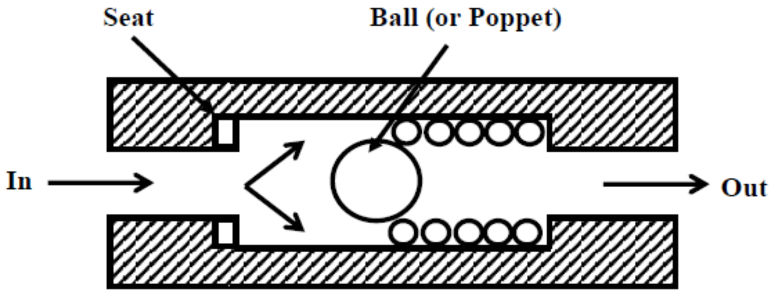
Check Valve

In its simplest form, a check valve is a one-way directional valve. It permits free flow in one direction and blocks flow in the other. It is analogous to the electronic diode. Symbol of check valve is shown below. The direction of the arrow shows the direction for free flow.

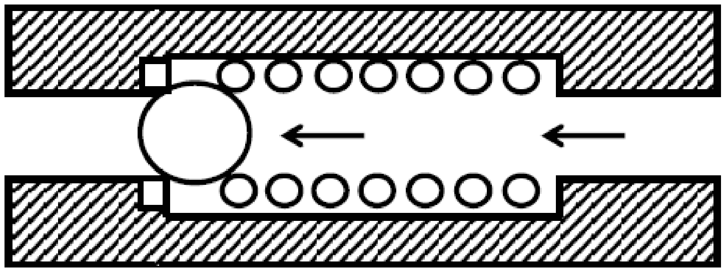


In its simplest form of construction, a check valve is realized as an in-line ball and spring.

Pressure from the left moves the ball from its seat so to permit unobstructed flow. Pressure from right pushes the ball tight on to the seat, and flow is blocked. In some valves a poppet is used in place of the ball. In some other construction, the valve inlet and outlet ports are made at right angles.



When flow is from left to right, valve is open

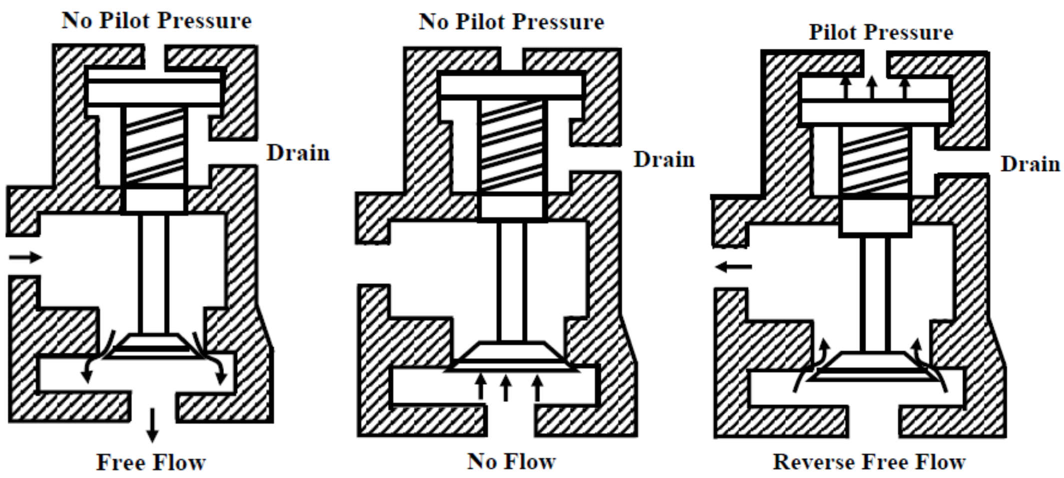


When flow is from right to left, the valve is closed

<u>Pilot-operated Check Valves</u>

Pilot-operated check valves are designed to permit free flow in one direction and to block return flow, unless pilot pressure is applied. However, under pilot pressure, flow is permitted in both directions. They are used in hydraulic presses as prefill valves – to permit the main ram to fill by gravity during the "fast approach" part of the stroke. They also are used to support vertical pistons which otherwise might drift downward due to leakage past the directional valve spool.

With no pilot pressure, the valve functions as a normal check valve. Flow to bottom is permitted but the reverse is blocked. If pilot pressure is applied, the valve is open at all times, and flow is allowed freely in both directions.

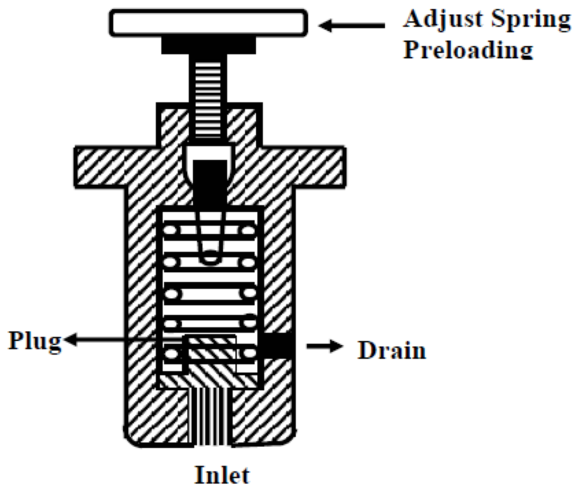


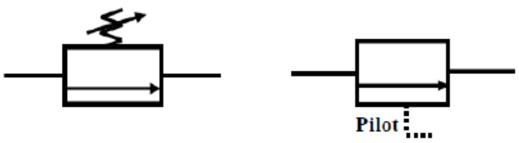
The check valve poppet has the pilot piston attached to the poppet stem. A light spring holds the poppet seated in a no-flow condition by pushing against the pilot piston. A separate drain port is provided to prevent oil from creating a pressure build-up on the underside of the piston. Reverse flow can occur only when a pressure that can overcome the pressure in the outlet chamber is applied.

Possible application of the valve can be to permit free flow to the accumulator, while blocking flow out of it. If the pilot is actuated the accumulator can discharge if the pressure at the inlet port is lower than the accumulator pressure.

Relief Valves

Relief valves are used for regulation of pressure in hydraulic systems for protection of equipment and personnel. The spring keeps the valve shut until a pressure set by an adjustable spring tension is reached which pushes the spring up to relief the pressure by connecting the inlet to the drain.



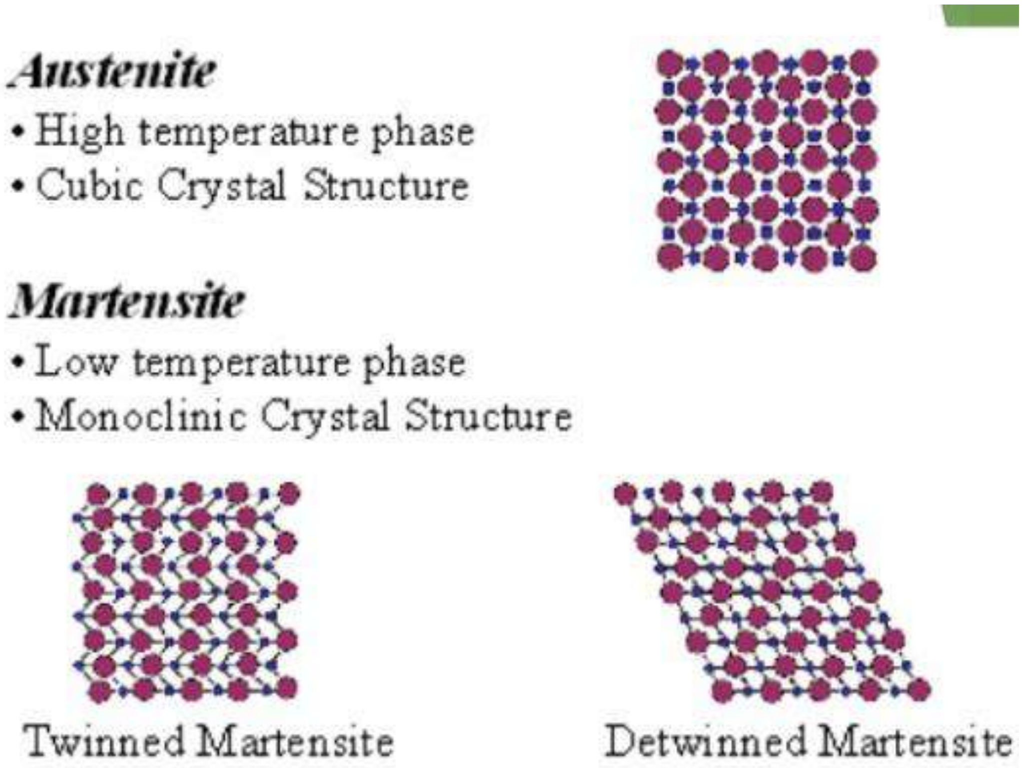


Symbol of relief valve

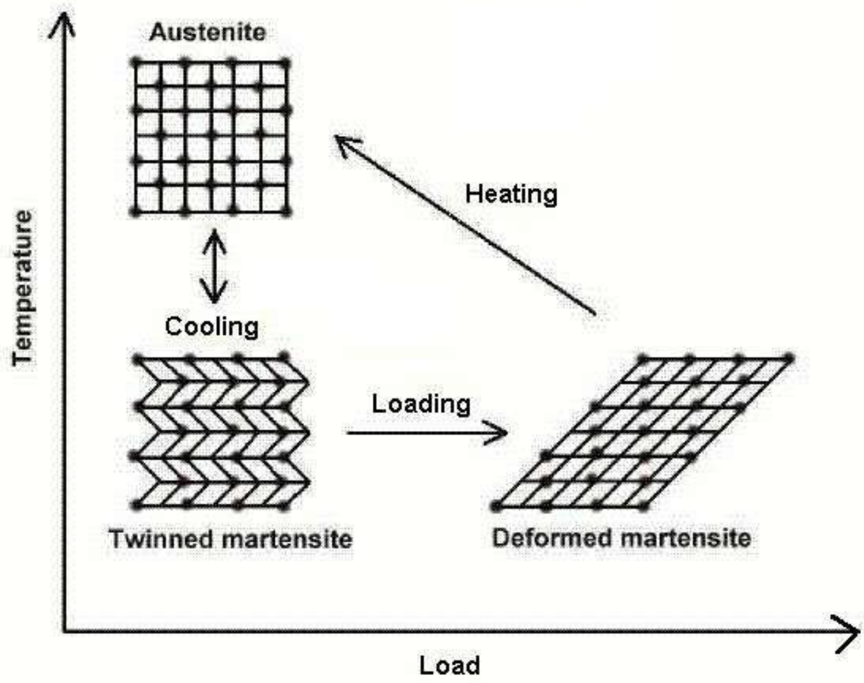
Shape Memory Alloys

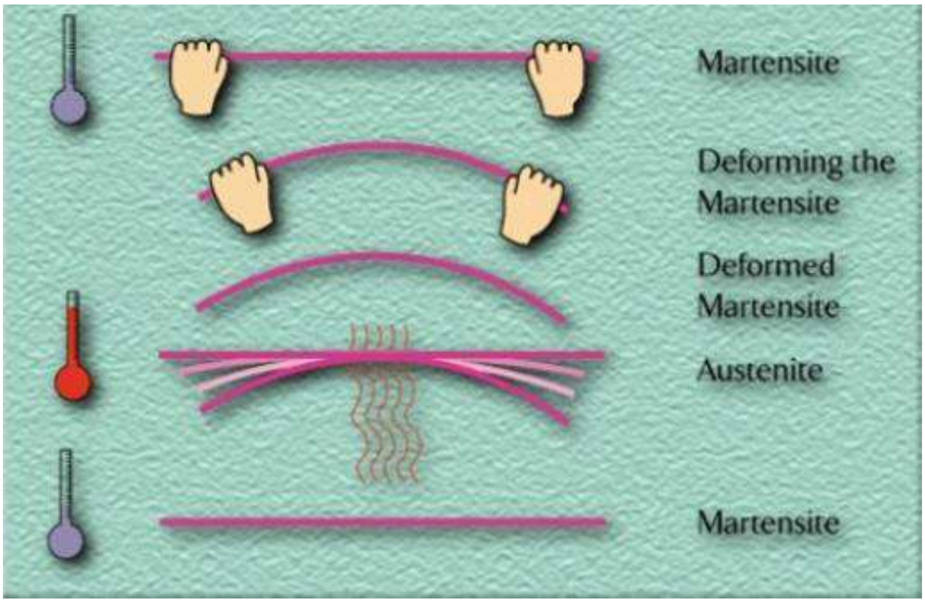
- Certain classes of metallic alloys have a special ability to memorize their shape at a high temperature, and recover large deformations imparted at a low temperature on thermal activation.
- The recovery of strains imparted to the material at a lower temperature, as a result of heating, is called the Shape Memory Effect(SME).
- They can remember the parent state with respect to the high temperature.
- Whatever is done with the metallic alloy at low temperature, the moment it is taken back to high temperature, it will get back to the parent shape
- Happens because of very regular crystal structure- most of the time a body centred cubic.
- SME 1st observed in 1932 in Gold Cadmium Alloy. They were expensive and not much large shape memory effect was seen in them
- Three types of SMA are currently popular
    - Cu Zn Al
    - Cu Al Ni
    - Ni Ti- Commercially available as NiTiNOL. (NOL: Naval Ordinance Laboratory)
- The shape memory alloys have two stable phases- the high temperature phase called austenite (After English metallurgist William Chandler Austen) and the low temperature phase called martensite (After German metallographer Adolf Martens)
- The martensite phase can be of two forms
    - Twinned martensite
    - Detwinned martensite
- A phase transformation which occurs between these two phases upon heating/cooling is the basis for the unique properties of the SMAs.
- Deformed shape + Heat = Original Shape
- The shape change involves a solid state phase change involving a molecular rearrangement between martensite and austenite.
- Upon cooling in the absence of applied load, the material transforms from austenite to twinned martensite (No observable macroscopic change occurs)
- If a mechanical load is applied to the material in twinned martensite phase (at low temperature), it is possible to detwin the martensite.
- Upon releasing of the load, the material remains deformed.
- A subsequent heating of the material to a temperature above the austenite finish temperature (Af) will result in a reverse phase transformation (martensite to austenite) and will lead to complete shape recovery.
- SMA remembers the shape when it has austenitic structure.
- So, if we need SMA to remember and regain/recover certain shape, the shape should be formed when the structure is austenite.
- The un-deformed Martensite phase is the same size and shape as the cubic Austenite phase on a macroscopic scale.
- No change in size or shape is visible in shape memory alloys until the Martensite is deformed.

- The high temperature causes the atoms to arrange themselves into the most compact and regular pattern possible resulting in a rigid cubic arrangement (austenite phase).
- They can be formed into various shapes like bars, wires, plates and rings thus serving various functions

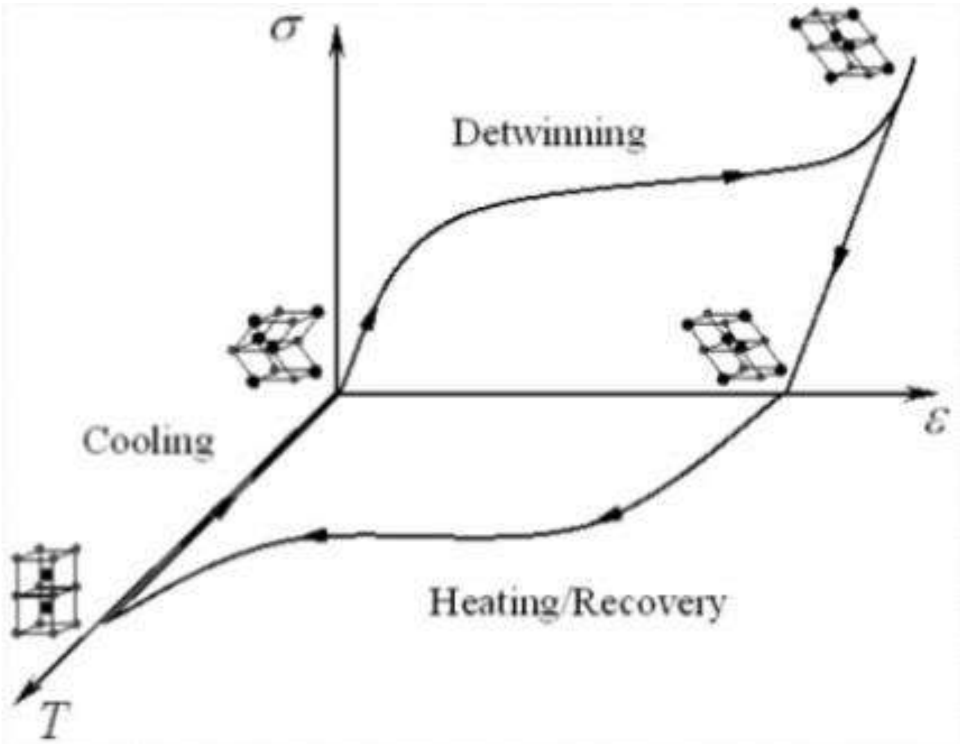


Phase Transformation in SMA





Hysteresis Curve of SMA



SME Temperatures

M_s: Temperature at which austenite begins to transform to martensite upon cooling

M_f: Temperature at which transformation of austenite to martensite is complete upon cooling

A_s: Temperature at which martensite begins to transform to austenite upon heating

A_f: Temperature at which transformation of martensite to austenite is complete upon heating

$$M_f < M_s < A_s < A_f$$

Types of SME

One-way shape memory: The material always remembers the shape at Parent always remembers the shape at Parent State (Austenite Phase). When the metal cools again it will remain in the hot shape, until deformed again.

Two-way shape memory: The deformed shape is remembered during cooling, in addition to the original shape being remembered during heating, i.e., memory is with both austenite and martensite phases. The material remembers two different shapes: one at low temperatures, and one at the high-temperature shape.

Advantages

Bio-compatibility

Diverse field of application

Good mechanical properties(strong, corrosion resistant)

Disadvantages

Expensive

Poor fatigue properties: Repeated use of shape memory effect leads to functional fatigue.

Overstress

Applications of Shape Memory Alloys

- The one well developed application of SMA is for simple and leak proof coupling of pneumatic and hydraulic lines.
- The alloys have also been exploited in mechanical and electromechanical control systems to provide a precise mechanical response to a small and repeated temperature changes.
- Also used in a wide range of medical and dental applications like healing broken bones, misaligned teeth etc.
- Various thermal actuators
- Biomedical
- Civil engineering of mega structures
- Self-expandable cardiovascular stent
- Blood clot filters
- Actuators for smart systems
- Flaps that change direction of airflow depending upon temperature (for air conditioners)
- Couplings
- Control of aerodynamic surfaces

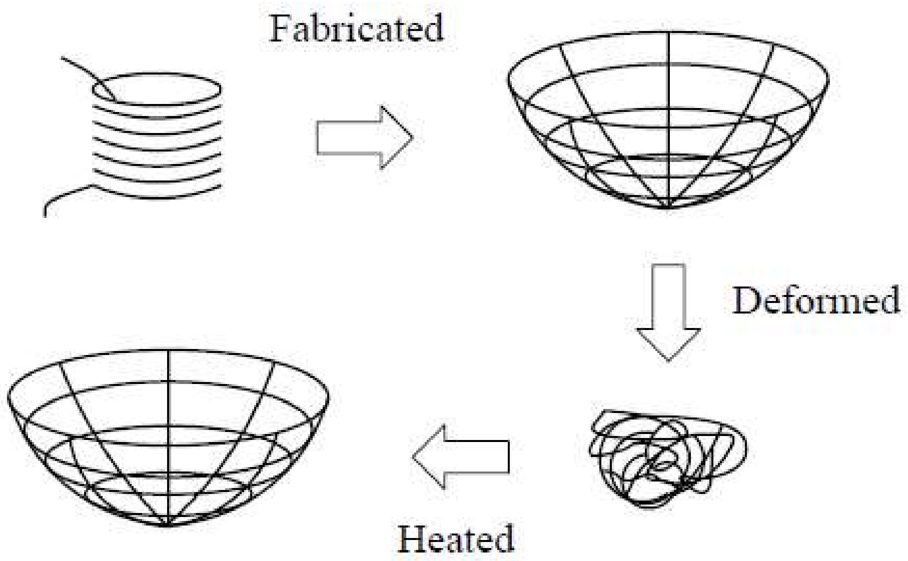


Fig: Satellite Antenna using SMA: Antenna is made into regular form in Austenite phase, it is cooled and brought to martensite phase and deformed so that it can be packed easily. When the satellite is exposed to high temperature at the orbit, the antenna regains its initial shape.

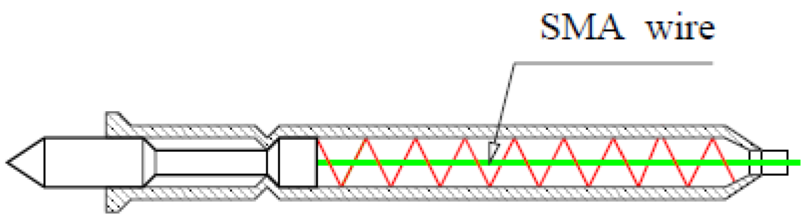


Fig: Actuator using SMA: The SMA wire is stretched or compressed during low temperature martensite phase. When it is exposed to high temperature, it regains its initial length which will lead to a movement of the actuator element.

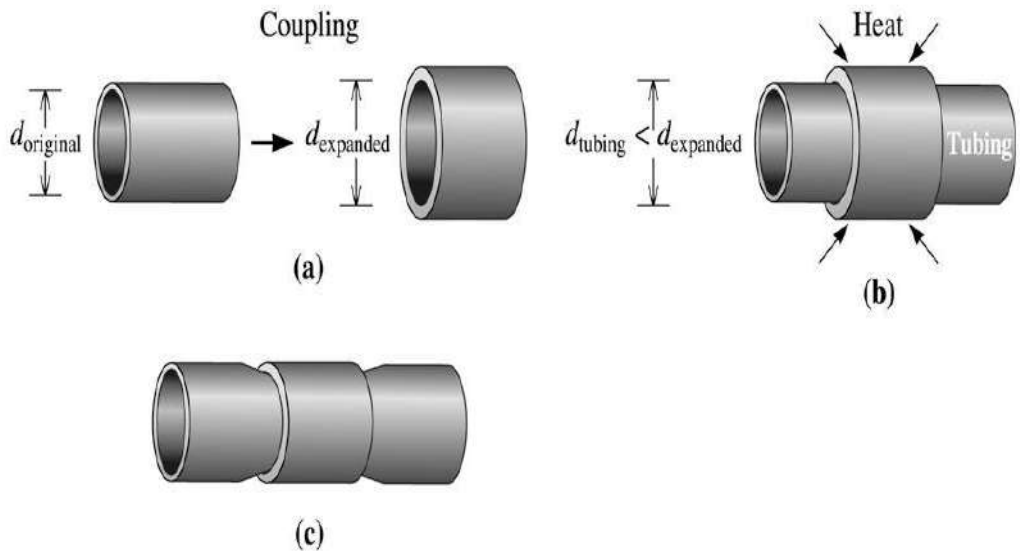


Fig: Coupling using SMA: A memory alloy coupling is expanded (a) so it fits over the tubing (b). When the coupling is reheated, it shrinks back to its original diameter (c), squeezing the tubing for a tight fit

Biological Applications

- Bone Plates
  - Memory effect pulls bones together to promote healing.
- Surgical Anchor
- Clot Filter
  - Does not interfere with MRI from non-ferromagnetic properties.

- Catheters

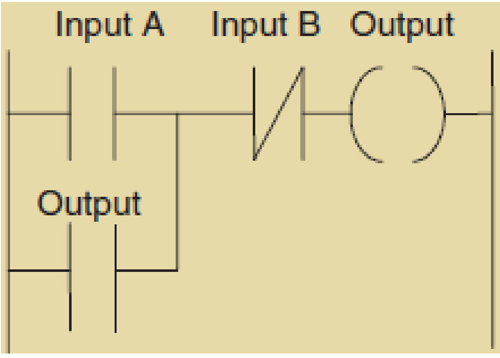
- Stent in artries
- Eyeglasses

Module VI

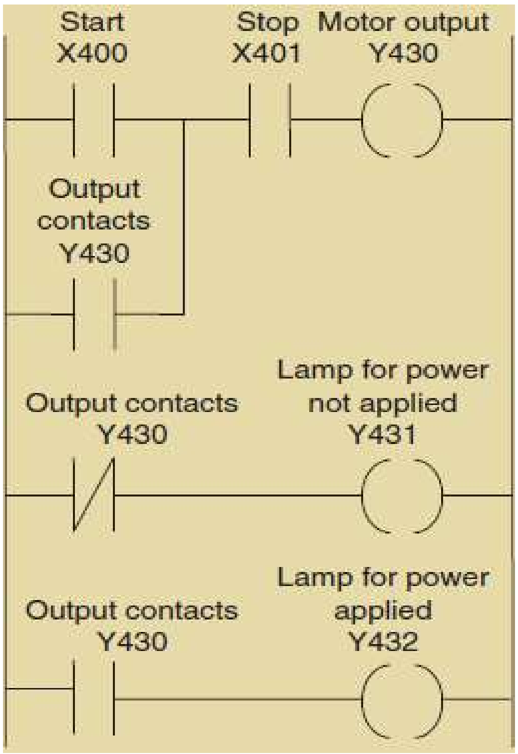
Latching in PLC

There are often situations in which it is necessary to hold an output energized, even when the input ceases. A simple example of such a situation is a motor that is started by pressing a push-button switch. Though the switch contacts do not remain closed, the motor is required to continue running until a stop push-button switch is pressed. The term latch circuit is used for the circuit that carries out such an operation. It is a self-maintaining circuit in that, after being energized, it maintains that state until another input is received.

An example of a latch circuit is shown in Figure. When the input A contacts close, there is an output. However, when there is an output, another set of contacts associated with the output closes. These contacts form an OR logic gate system with the input contacts. Thus, even if input A opens, the circuit will still maintain the output energized. The only way to release the output is by operating the normally closed contact B.



As an illustration of the application of a latching circuit, consider a motor controlled by stop and start push-button switches and for which one signal light must be illuminated when the power is applied to the motor and another when it is not applied. Figure shows a ladder diagram with Mitsubishi notation for the addresses. X401 is closed when the program is started. When X400 is momentarily closed, Y430 is energized and its contacts close. This results in latching as well as the switching off of Y431 and the switching on of Y432. To switch the motor off, X401 is pressed and opens. Y430 contacts open in the top rung and third rung but close in the second rung. Thus Y431 comes on and Y432 goes off. Latching is widely used with startups so that the initial switching on of an application becomes latched on.



*****Refer the given textbook for the remaining portions of PLC*****

"Programmable Logic Controllers and Industrial Automation: An Introduction"

Madhuchhanda Mitra & Samarjit Sengupta- Chapters 1, 2 & 3

Supervisory Control and Data Acquisition (SCADA)

SCADA is an acronym that stands for Supervisory Control and Data Acquisition. SCADA refers to a system that collects data from various sensors at a factory, plant or in other remote locations and then sends this data to a central computer which then manages and controls the data. SCADA systems are used not only in industrial processes: e.g. steel making, power generation (conventional and nuclear) and distribution, chemistry, but also in some experimental facilities such as nuclear fusion. The size of such plants range from a few 1000 to several 10 thousand input/output (I/O) channels.

A SCADA system refer to a system consisting of a number of remote terminal units (or RTUs) collecting field data connected back to a master station via a communications system. The master station displays the acquired data and also allows the operator to perform remote control tasks. The accurate and timely data (normally real-time) allows for optimization of the operation of the plant and process. A further benefit is more efficient, reliable and most importantly, safer operations. This all results in a lower cost of operation compared to earlier non-automated systems.

SCADA systems are used to monitor or to control chemical or transport processes in municipal water supply systems, to control electric power generation, transmission and distribution, gas and oil pipelines, and other distributed processes. Supervisory control and data Acquisition (SCADA)achieves this requirement collecting reliable field data through remote terminal units (RTUs) Intelligent Electric Devices (IEDs) and presenting them to user requirement. The user interface or the man machine interface (MMI) provides various options of data presentation according to specific application and user needs. There are many parts of a working SCADA system. A SCADA system usually includes signal hardware (input and output), controllers, networks, user interface (HMI), communications equipment and software. All together, the term SCADA refers to the entire central system. The central system usually monitors data from various sensors that are either in close proximity or off site.

The RTU provides an interface to the field analog and digital signals situated at each remote site. The communications system provides the pathway for communications between the master station and the remote sites. This communication system can be radio, telephone line, microwave and possibly even satellite. Specific protocols and error detection philosophies are used for efficient and optimum transfer of data. The master station (and sub-masters) gather data from the various RTUs and generally provide an operator interface for display of information and control of the remote sites. In large telemetry systems, sub-master sites gather information from remote sites and act as a relay back to the control master station.

Contemporary SCADA systems exhibit predominantly open-loop control characteristics and utilize predominantly long distance communications, although some elements of closed-loop control.

Objectives of SCADA:

- Monitoring: Continuous monitoring of the parameters of voltage, current, etc.
- Measurement: Measurement of variables for processing.
- Data Acquisition: Frequent acquisition of data from RTUs and Data Loggers / Phasor data Concentrators (PDC).
- Data Communication: Transmission and receiving of large amounts of data from field to control centres.
- Control: Online real time control for closed loop and open loop processes.
- Automation: Automatic tasks of switching of transmission lines, CBs, etc.

Functions of SCADA:

- Data Acquisition
- Information Display
- Supervisory Control
- Alarm Processing
- Information Storage and Reports
- Sequence of Event Acquisition
- Data Calculation
- Special RTU Processing/Control

Advantages of SCADA System

The advantages of Supervisory Control and Data Acquisition system include:

- Improvement in Service Quality
- Improvement in Reliability
- Continuous monitoring of process.
- Real time control.
- Automation and Protection.
- Remote control and operation.
- Reduction in operation and maintenance costs
- Easy to monitor large system parameters
- Real time information on demand
- Reduction in Manpower
- Value added services
- Ease in Fault Detection and Fault Localization (FDFL)
- Reduction in Repair Time (System Down Time)

Applications/Usage of SCADA:

SCADA systems are widely used for control in the following domains

1. *Electric power generation, transmission and distribution:* Electric utilities use SCADA systems to detect current flow and line voltage, to monitor the operation of circuit breakers, and to take sections of the power grid online or offline.
2. *Water and sewage:* State and municipal water utilities use SCADA to monitor and regulate water flow, reservoir levels, pipe pressure and other factors.
3. *Industries, buildings, facilities and environments:* Facility managers use SCADA to control HVAC, refrigeration units, lighting and entry systems.
4. *Manufacturing:* SCADA systems manage parts inventories for just-in-time manufacturing, regulate industrial automation and robots, and monitor process and quality control.
5. *Mass transit:* Transit authorities use SCADA to regulate electricity to subways, trams and trolley buses; to automate traffic signals for rail systems; to track and locate trains and buses; and to control railroad crossing gates.
6. *Traffic signals:* SCADA regulates traffic lights, controls traffic flow and detects out-of-order signals.
7. *Communication Networks:* Used for monitoring and controlling servers, networks and nodes.

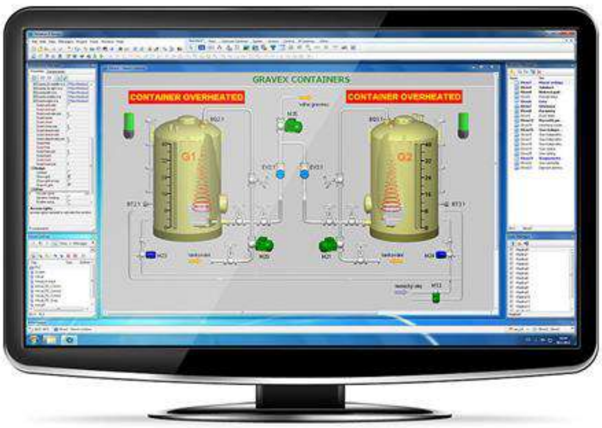
A SCADA system performs four basic functions:

- Data acquisition
- Networked data communication
- Data presentation
- Control

Data Acquisition: First, the systems need to be monitored are much more complex than just one machine with one output. So a real-life SCADA system needs to monitor hundreds or thousands of sensors. Some sensors measure inputs into the system (for example, water flowing into a reservoir), and some sensors measure outputs (like valve pressure as water is released from the reservoir). Some of those sensors measure simple events that can be detected by a straightforward on/off switch, called a discrete input (or digital input). In real life, discrete inputs are used to measure simple states, like whether equipment is on or off, or tripwire alarms, like a power failure at a critical facility. Some sensors measure more complex situations where exact measurement is important. These are analog sensors, which can detect continuous changes in a voltage or current input. Analog sensors are used to track fluid levels in tanks, voltage levels in batteries, temperature and other factors that can be measured in a continuous range of input. For most analog factors, there is a normal range defined by a bottom and top level. For example, you may want the temperature in a server room to stay between 60 and 85 degrees Fahrenheit. If the temperature goes above or below this range, it will trigger a threshold alarm.

Data Communication: If multiple systems have to be monitored from a central location, a communications network to transport all the data collected from the sensors is needed. Early SCADA networks communicated over radio, modem or dedicated serial lines. Today the trend is to put SCADA data on Ethernet and IP over SONET. For security reasons, SCADA data should be kept on closed LAN/WANs without exposing sensitive data to the open Internet. Real SCADA systems don't communicate with just simple electrical signals, either. SCADA data is encoded in protocol format. The remote telemetry unit (RTU) is needed to provide an interface between the sensors and the SCADA network. The RTU encodes sensor inputs into protocol format and forwards them to the SCADA master; in turn, the RTU receives control commands in protocol format from the master and transmits electrical signals to the appropriate control relays. Communication gateways are also used for conversion between electrical signals and protocols.

Data Presentation: A SCADA system reports to human operators over a specialized computer that is variously called a master station, an HMI (Human-Machine Interface) or an HCI (Human-Computer Interface). The SCADA master station has several different functions. The master continuously monitors all sensors and alerts the operator when there is an "alarm" — that is, when a control factor is operating outside what is defined as its normal operation. The master presents a comprehensive view of the entire managed system, and presents more detail in response to user requests. The master also performs data processing on information gathered from sensors — it maintains report logs and summarize historical trends. An advanced SCADA master can add a great deal of intelligence and automation to your systems management, making your job much easier.



The SCADA system uses human machine interface. The information is displayed and monitored to be processed by the human. HMI provides the access of multiple control units which can be PLCs and RTUs. The HMI provides the graphical presentation of the system. For example, it provides the graphical picture of the pump connected to the tank. The user can see the flow of the water and pressure of the water. The important part of the HMI is an alarm system which is activated according to the predefined values.

Control: If there is a sufficiently sophisticated master unit, the controls can run completely automatically, without the need for human intervention. The automatic controls can still be manually override from the master station. In real life, SCADA systems automatically regulate all kinds of industrial processes. For example, if too much pressure is building up in a gas pipeline, the SCADA system can automatically open a release valve. Electricity production can be adjusted to meet demands on the power grid. Even these real-world examples are simplified; a full-scale SCADA system can adjust the managed system in response to multiple inputs.

Architecture of SCADA

Generally, SCADA system is a centralized system which monitors and controls entire area. It is purely software package that is positioned on top of hardware. A supervisory system gathers data on the process and sends the commands control to the process. The SCADA has a remote terminal unit which is also known as RTU. Most control actions are automatically performed by RTUs or PLCs. The RTUs consist of programmable logic converter which can be set to specific requirement. For example, in the thermal power plant the water flow can be set to specific value or it can be changed according to the requirement.

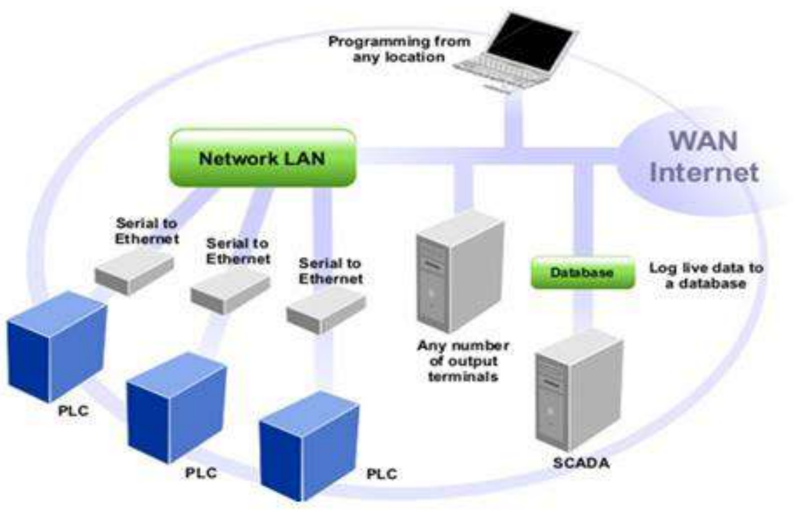
The SCADA system allows operators to change the set point for the flow, and enable alarm conditions in case of loss of flow and high temperature and the condition is displayed and recorded.

The SCADA system monitors the overall performance of the loop. The SCADA system is a centralized system to communicate with both wire and wireless technology to Clint devices. The SCADA system controls can run completely all kinds of industrial process. E.g.: If too much pressure in building up in a gas pipe line the SCADA system can automatically open a release valve.

Generally, SCADA system can be classified into two parts:

- The Clint layer which caters for the man machine interaction.
- The data server layer which handles most of the process data activities.

The SCADA station refers to the servers and it is composed of a single PC. The data servers communicate with devices in the field through process controllers like PLCs or RTUs. The PLCs are connected to the data servers either directly or via networks or buses. The SCADA system utilizes a WAN and LAN networks, the WAN and LAN consists of internet protocols used for communication between the master station and devices. The physical equipment's like sensors connected to the PLCs or RTUs. The RTUs convert the sensor signals to digital data and sends digital data to master. According to the master feedback received by the RTU, it applies the electrical signal to relays. Most of the monitoring and control operations are performed by RTUs or PLCs as we can see in the figure.



Types of SCADA System

There are four different types of SCADA systems from four generations. They are:

- Early or Monolithic SCADA Systems (First Generation)
- Distributed SCADA Systems (Second Generation)
- Networked SCADA Systems (Third Generation)
- IoT SCADA Systems (Fourth Generation)

Elements of SCADA/SCADA Block Diagram:

- Sensors and actuators
- RTUs/PLCs
- Communication
- Master Terminal Unit
  - Front End Processor
  - SCADA server
  - Historical/Redundant/Safety Server
  - HMI computer
  - HMI software

Sensors and Actuators: (Explain briefly about sensors and actuators)

Remote Terminal Unit(RTU)

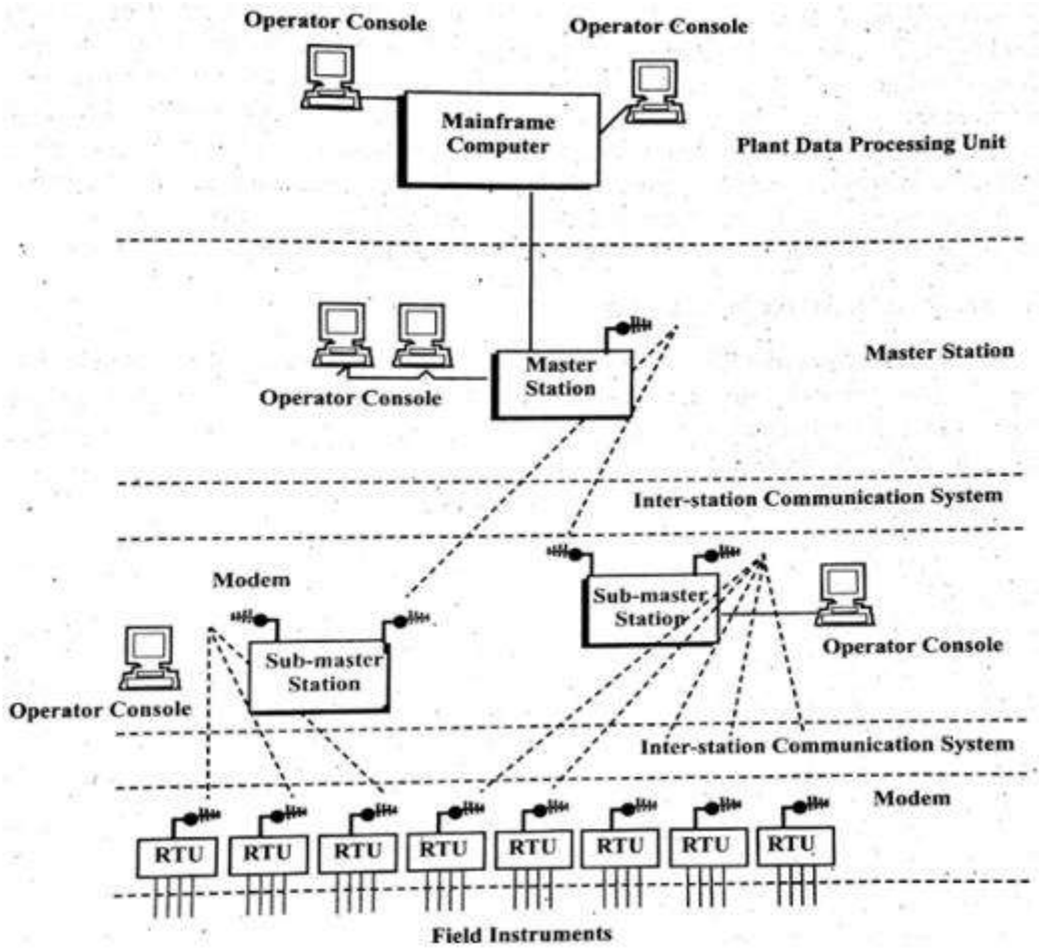
An RTU is a stand-alone data acquisition and control unit, which monitors and control equipment at some remote location from the central station. Its primary task is to control and acquire data from process equipment at the remote location and to transfer these data back to the central station. A SCADA system consists of a number of RTUs collecting field data and sending these data back to master station via a communication system. RTU provides an interface to the field analog and digital sensors.

PLCs are sometimes used as RTUs because a PLC has a lot of advantages over RTUs such as

i. Versatility and flexibility: PLCs can easily have their logic or hardware modified to cope with modified requirements of control.
ii. Enhanced reliability: When correctly installed, PLCs are far more reliable solution than a traditional hardwired relay solution.
iii. Sophisticate control: PLCs allows for far more sophisticated control mainly due to the software capability which is greater than that of RTUs.
iv. Easier troubleshooting and diagnostics: Software and clear cut reporting of problems allow easy and swift diagnosis of hardware/firmware/software problems on the system as well as identifying problems with the process and automation system.

Communication System:

The communication system provides pathway for communication between the master station and the remote sites. This communication system can be wire, fibre optic, radio, telephone line, microwave and possibly even a satellite. The communication interfaces that are generally used in RTUs are RS-232/RS-442/RS-485, dialup telephone lines/dedicated landlines, microwave, satellite etc.



<u>Master Station</u>

RTUs collect field data send these data to a master station via a communication system. The master station displays the acquired data and allows the operator to perform remote control tasks. In large telemetry systems, sub-master sites gather information from remote sites and acts as a relay back to the control master station.

The central site/master station can be pictured as having one or more operator stations tied together with a LAN connected to a communication system consisting of a modern and radio receiver/transmitter. Normally there are no input/output modules connected directly to the master stations although there may be RTUs located in close proximity to the master room. These master stations are provided with many features that should be available for an industrial system. If the geographical distance is large or data traffic is more, it is preferred that sub-master stations are set-up between the RTUs and the master station. The sub-master station may have functions like acquiring data from the RTUs within the region, logging and displaying the data on a local operator station, passing data back to the master station and passing on control requests from the master station to the RTUs in its region.

If at any level, any station malfunctions, the transmission process is interrupted. To avoid such situation, a number of stations at each level (master or sub-master station level) are provided to operate in parallel so that in case of failure of any one station, the other station takes charge so that the transmission process remains uninterrupted. This is called redundancy stations.

Human-machine interface:

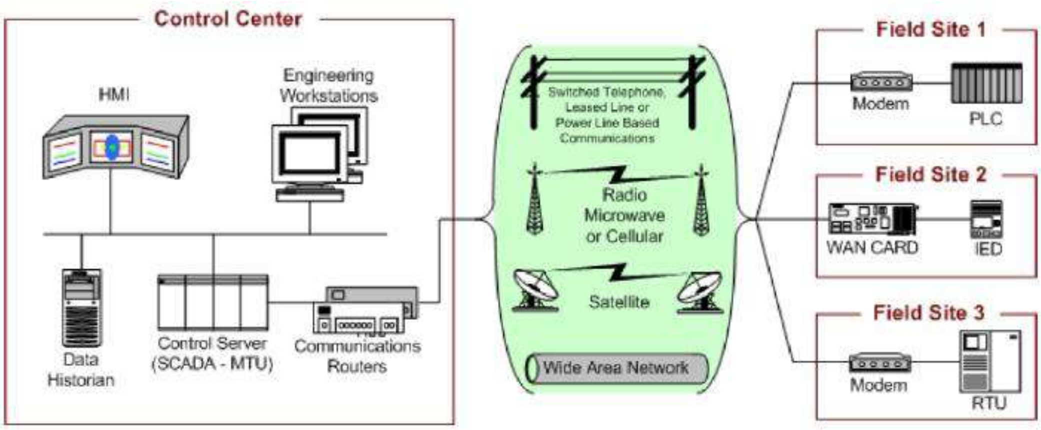
The human-machine interface (HMI) is the operator window of the supervisory system. It presents plant information to the operating personnel graphically in the form of mimic diagrams, which are a schematic representation of the plant being controlled, and alarm and event logging pages. The HMI is linked to the SCADA supervisory computer to provide live data to drive the mimic diagrams, alarm displays and trending graphs. In many installations the HMI is the graphical user interface for the operator, collects all data from external devices, creates reports, performs alarming, sends notifications, etc.

Supervisory operation of the plant is by means of the HMI, with operators issuing commands using mouse pointers, keyboards and touch screens. For example, a symbol of a pump can show the operator that the pump is running, and a flow meter symbol can show how much fluid it is pumping through the pipe. The operator can switch the pump off from the mimic by a mouse click or screen touch. The HMI will show the flow rate of the fluid in the pipe decrease in real time.

The HMI package for a SCADA system typically includes a drawing program that the operators or system maintenance personnel use to change the way these points are represented in the interface.

- SCADA systems are used to control dispersed assets where centralized data acquisition is as important as control (i.e., SCADA is a method of monitoring and controlling large processes, often scattered over thousands of square kilometres).
- These systems are used in distribution systems such as water distribution and wastewater collection systems, oil and natural gas pipelines, electrical utility transmission and distribution systems, and rail and other public transportation systems.
- SCADA systems integrate data acquisition systems with data transmission systems and HMI software to provide a centralized monitoring and control system for numerous process inputs and outputs.
- SCADA systems are designed to collect field information, transfer it to a central computer facility, and display the information to the operator graphically or textually, thereby allowing the operator to monitor or control an entire system from a central location in real time.
- Based on the sophistication and setup of the individual system, control of any individual system, operation, or task can be automatic, or it can be performed by operator commands.
- SCADA is not a full control system, but rather focuses on the supervisory level.
- SCADA is used for gathering, analysing and to storage real time data.
- SCADA systems consist of both hardware and software.
- Typical hardware includes an MTU placed at a control centre, communications equipment (e.g., radio, telephone line, cable, or satellite), and one or more geographically distributed field sites consisting of either an RTU or a PLC, which controls actuators and/or monitors sensors.
- The MTU stores and processes the information from RTU inputs and outputs, while the RTU or PLC controls the local process.
- The communications hardware allows the transfer of information and data back and forth between the MTU and the RTUs or PLCs.
- The software is programmed to tell the system what and when to monitor, what parameter ranges are acceptable, and what response to initiate when parameters change outside acceptable values.
- An IED, such as a protective relay, may communicate directly to the SCADA Server, or a local RTU may poll the IEDs to collect the data and pass it to the SCADA Server.
- IEDs provide a direct interface to control and monitor equipment and sensors.

- IEDs may be directly polled and controlled by the SCADA Server and in most cases have local programming that allows for the IED to act without direct instructions from the SCADA control centre.
- SCADA systems are usually designed to be fault-tolerant systems with significant redundancy built into the system architecture.



SCADA System General Layout

- The control centre houses a SCADA Server (MTU) and the communications routers.
- Other control centre components include the HMI, engineering workstations, and the data historian, which are all connected by a LAN.
- The control centre collects and logs information gathered by the field sites, displays information to the HMI, and may generate actions based upon detected events.
- The control centre is also responsible for centralized alarming, trend analyses, and reporting.
- The field site performs local control of actuators and monitors sensors.
- Field sites are often equipped with a remote access capability to allow field operators to perform remote diagnostics and repairs usually over a separate dial up modem or WAN connection.
- Standard and proprietary communication protocols running over serial communications are used to transport information between the control centre and field sites using telemetry techniques such as telephone line, cable, fibre, and radio frequency such as broadcast, microwave and satellite.
- MTU-RTU communication architectures vary among implementations. The various architectures used, including point-to-point, series, series-star, and multi-drop.

- An RTU (sometimes referred to as a remote telemetry unit) as the title implies, is a standalone data acquisition and control unit, generally microprocessor based, which monitors and controls equipment at some remote location from the central station.
- Its primary task is to control and acquire data from process equipment at the remote location and to transfer this data back to a central station.
- It generally also has the facility for having its configuration and control programs dynamically downloaded from some central station.
- There is also a facility to be configured locally by some RTU programming unit.
- Although traditionally the RTU communicates back to some central station, it is also possible to communicate on a peer-to-peer basis with other RTUs.
- The RTU can also act as a relay station (sometimes referred to as a store and forward station) to another RTU, which may not be accessible from the central station.
- Small sized RTUs generally have less than 10 to 20 analog and digital signals, medium sized RTU shave 100 digital and 30 to 40 analog inputs.
- RTUs, having a capacity greater than this can be classified as large.

Distributed Control Systems (DCS)

Systems similar to SCADA systems are routinely seen in factories, treatment plants etc. These are often referred to as Distributed Control Systems (DCS). They have similar functions to SCADA systems, but the field data gathering or control units are usually located within a more confined area. Communications may be via a local area network (LAN), and will normally be reliable and high speed. A DCS system usually employs significant amounts of closed loop control. SCADA systems on the other hand generally cover larger geographic areas, and rely on a variety of communications systems that are normally less reliable than a LAN. Closed loop control in this situation is less desirable in SCADA.

DCS are typically used for single-point processing and are employed in a limited geographic area. On the other hand, SCADA systems are used for large-scale, distributed management of critical infrastructure systems and are often geographically dispersed. For example, in a power utility, DCS may be used for generation of power, while SCADA is used for the distribution and transmission of power.

A distributed control system (DCS) refers to a control system usually of a manufacturing system, process or any kind of dynamic system, in which the controller elements are not central in location (like the brain) but are distributed throughout the system with each component sub-system controlled by one or more controllers. The entire system of controllers is connected by networks for communication and monitoring. DCS is a very broad term used in a variety of industries, to monitor and control distributed equipment.

A DCS typically uses custom designed processors as controllers and uses both proprietary interconnections and communications protocol for communication. Input and output modules form component parts of the DCS. The processor receives information from input modules and sends information to output modules. The input modules receive information from input instruments in the process (or field) and transmit instructions to the output instruments in the field. Computer buses or electrical buses connect the processor and modules through multiplexer or demultiplexers. Buses also connect the distributed controllers with the central controller and finally to the Human–machine interface (HMI) or control consoles.

DCS is a computerised control system for a process or plant that consist of a large number of control loops, in which autonomous controllers are distributed throughout the system, but there is a central operator supervisory control

These systems are used on large continuous process plants where high reliability and security is required.

As DCS contains distribution of the control processing around nodes in the system the complete system is reliable and mitigates a single processor failure. It will affect only one section of the plant process only, if a processor fails while the whole plant will be affected when the central computer fails.

This distribution of computing power to the field input/output field connection rack also ensures fast processing by removing possible network and central processing delays.

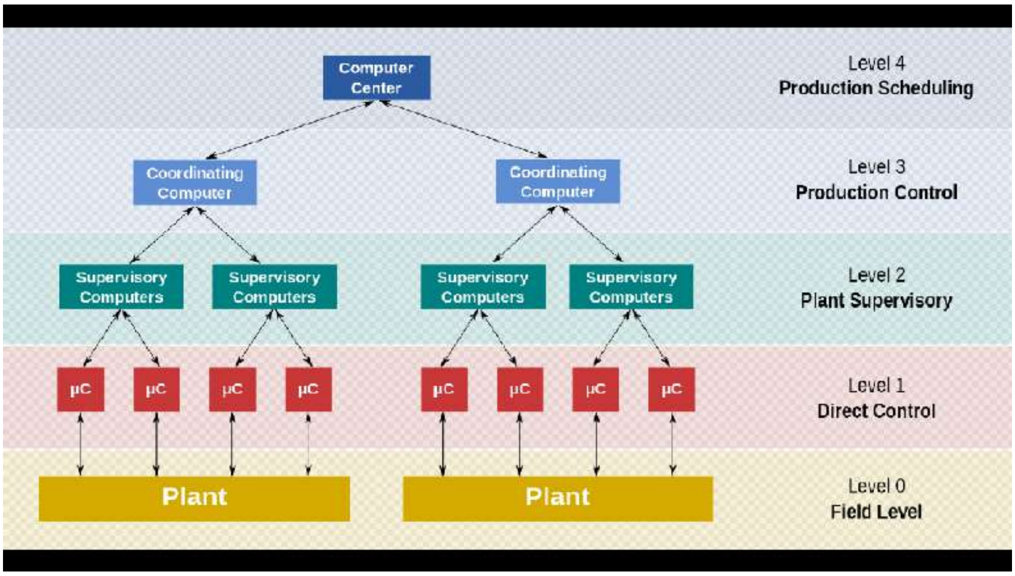


Figure above shows the architecture of computerized control in a process or industry.

- Level 0: field devices such as temp sensors, flow and final control elements such as control valves
- Level 1: Industrial Input/Output modules and their associated distributed electronic processors
- Level 2: Included with Supervisory computers that help to gather information from processor nodes on the system, and provide the operator control screens
- Level 3: production control level, which does not directly control the process but is concerned with monitoring production and monitoring targets
- Level 4: production scheduling level

Level 1,2 are the functional levels of a traditional DCS, in which all equipment's are part of an integrated system.

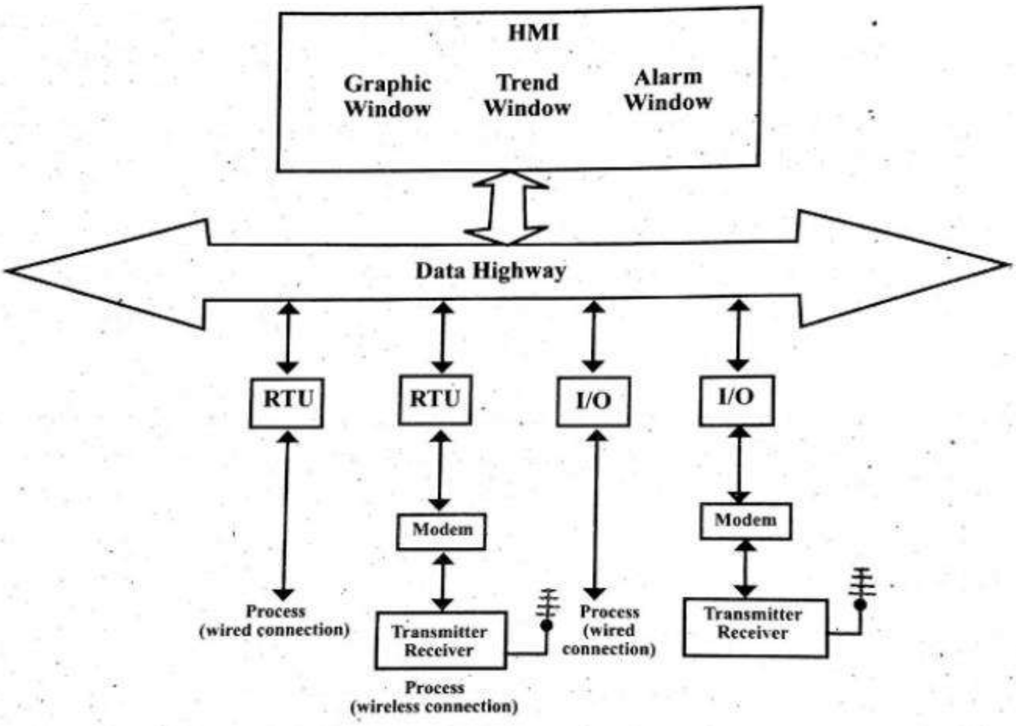


Fig. 7.5 Block Diagram of a DCS

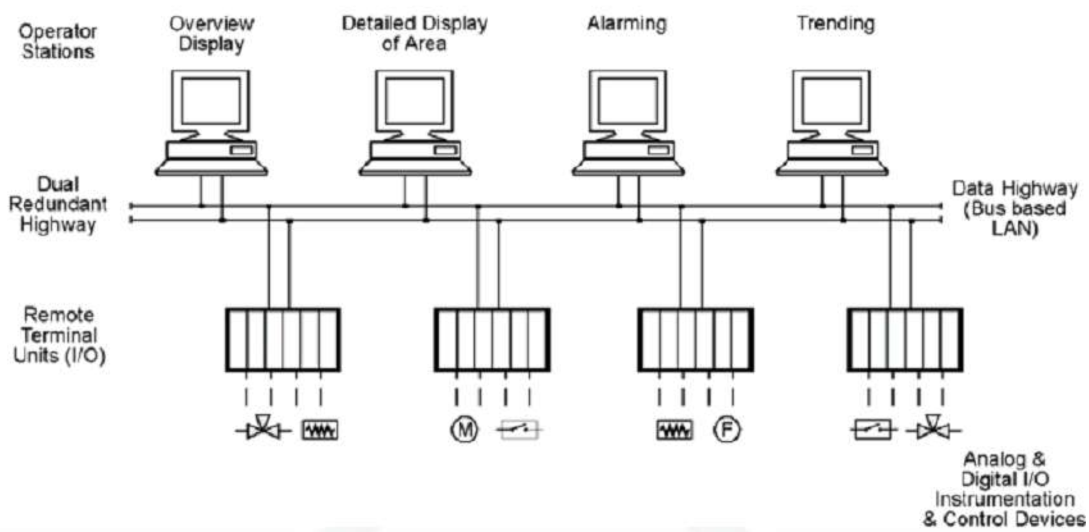


Figure 9.8
Distributed control system (DCS)

Features of DCS

- System redundancy
- More sophisticated HMI
- Scalable Platform
- System Security

Benefits of DCS

- Various control functions can be distributed into small sets of subsystems that are semi-autonomous.
- These are but interconnected by a high speed communication bus. (functions include data acquisition, process control, process supervision, reporting information, storing and retrieval of information)
- Automation of manufacturing process by integrating advanced control strategies.
- Arranging the things in a system; DCS unites sub-systems through a proper command structure and information flow.
- Requires minimum troubleshooting.
- Contains HMI graphics and faceplates
- Batch management is possible
- Reduce engineering time
- Redundant operator system servers

Applications of DCS

- Chemical plants
- Petrochemical (oil) and refineries
- Pulp and Paper Mills
- Boiler controls and power plant systems
- Nuclear power plant
- Environmental control systems
- Water management systems
- Water treatment plants
- Sewage treatment plants
- Food and food processing
- Agrochemical and fertilizer
- Metal and mines
- Automobile manufacturing
- Metallurgical process plants
- Pharmaceutical manufacturing
- Sugar refining plants
- Agriculture Applications

<u>DCS vs PLC</u>

To differentiate between PLC & DCS, there are five key factors to be considered: They are:

1) Response Time:

PLCs are fast, which make the PLC an ideal controller for real-time actions such as a safety shutdown or firing control.

A DCS takes much longer time to process data, which made it not the right solution when response times are critical.

2) Scalability:

A few thousand I/O points can be handled by a PLC whereas DCS can handle many thousands of I/O points and accommodate new equipment, process enhancements, and data integration.

DCS can be preferred when advanced process control is required and contain a large facility that's spread out over a wide geographic area with thousands of I/O points.

3) Complexity

The advanced process control capabilities of the DCS are required to carry out the complex nature of many continuous production processes such as oil and gas, water treatment and chemical processing.

4) Frequent process changes

PLCs are used for processes that will not change often. And, a DCS is the better solutions when the process is complex and requires frequent adjustments or must analyse a large amount of data.

5) Vendor support

To provide integration services and implement process changes, DCS vendors require users to employ them. System integrators perform similar functions for PLC-based systems. It has also become common for PLC vendors to offer support services through their network of system integrator partners.

Overall, DCS tends to be used in large continuous process plants where high reliability and security is important, and the control room is not geographically remote.

PLC	DCS
1. High level programming languages are available for creating custom logic	1. Custom logic created from existing function blocks
2. Customized routines usually required-	2. Many algorithms (i.e. PID) are complex and do not vary among application
3. Standard libraries considered nice features	3. Standard application libraries are expected (function blocks and faceplates)
4. Provisions must be available to integrate functions/ products into an integrated architecture	4. Entire system is expected to function as a Complete solution
5. Redundancy may not be cost justified	5. System redundancy is often required
6. Analog Control: Simple PID only	6. Analog Control: Simple to advanced PID control up to Advanced Process Control
7. Diagnostics to tell you when something is broken	7. Asset Management alerts you to what might break before it does
8. System designed to be flexible	8. System designed to make it "easy" to engineer process applications

DCS vs SCADA

There is considerable confusion today about the difference between DCS ("Distributed Control Systems") and SCADA ("Supervisory Control and Data Acquisition") systems. As you can tell from expanded acronyms above, SCADA includes "Data Acquisition" in addition to "Control". DCS, on the other hand, contains only "Control".

Historically, when computer networks either did not yet exist or had very low bandwidth, a SCADA system was the top-level controller for many lower-level intelligent agents. It was simply impractical to have a single system controlling every minute aspect of a system. In this technical environment, DCS devices did most of the detail work and simply reported to (and took high-level orders from) the SCADA system.

Today, computer networks have become so fast that there's no practical reason for SCADA and DCS to be separate. That's why they have blurred together into a single monitoring and control system. The choice of name - SCADA vs. DCS - largely depends on the region where you work. Some areas favour SCADA, others favour DCS. Occasionally, some people who worked with the systems before they effectively merged or who have moved from another region will use a term different than their co-workers. This again leads to confusion when new employees must learn to manage SCADA/DCS.

DCS is process oriented, while SCADA is data acquisition oriented.

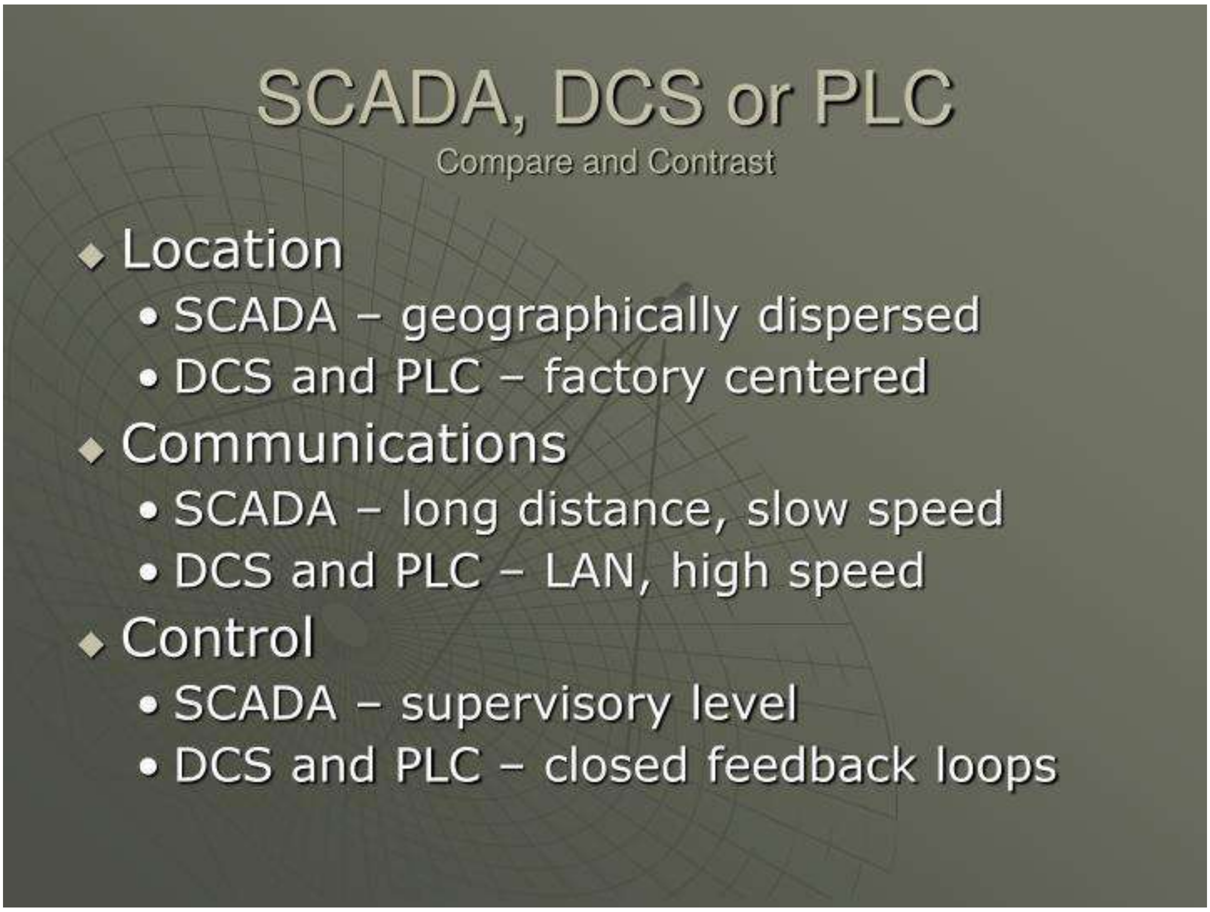
DCS is process state driven, while SCADA is event driven.

DCS is commonly used to handle operations on a single locale, while SCADA is preferred for applications that are spread over a wide geographic location.

DCS operator stations are always connected to its I/O, while SCADA is expected to operate despite failure of field communications.

DCS	**SCADA**
DCS is process oriented	Data-gathering oriented
Emphasizes more on control of the process, and it also consists of supervisory control level	Concentrates more on acquisition process data and presenting it to the operators and control center
Controllers are usually located within a more confined area, and the communication is through LAN	Generally covers large geographical areas, that uses different communication systems that are generally less reliable
Employs close loop control at Process control station and at RTU’s	There is no such close loop control
DCS is process state driven, where it scans the process in in regular basis and displays the result to the operator	SCADA is event driven, where it does not scan the process sequentially, but it waits for an event that causes process parameter to trigger certain actions.
DCS does not keep a database of process parameter values, as it is always in connection with its data source	SCADA maintains a database to log the parameter values, which can be further retrieved for operator display
DCS is used for installations within a confined area, like a single plant or factory E.g.: chemical plants, power generation stations, pharmaceutical manufacturing, oil and gas industries.	SCADA is used for much larger geographical locations E.g.: Water management systems, power transmission and distribution control, transport

	applications and small manufacturing and process industries.

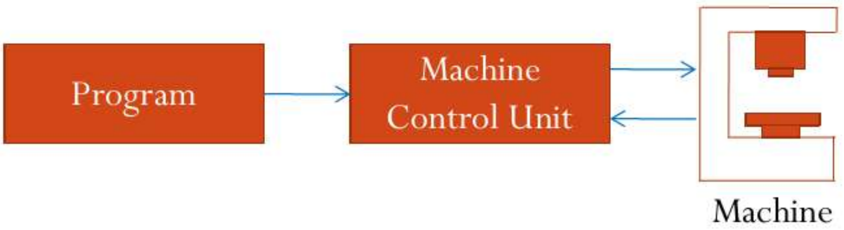


Computer Numerical Control (CNC)

Computer Numerical Control (CNC) is a specialized and versatile form of Soft Automation and its applications cover many kinds, although it was initially developed to control the motion and operation of machine tools.

A system in which actions are controlled by the direct insertion of numerical data at some point. The system must automatically interpret at least some portion of this data.

Computer Numerical Control may be considered to be a means of operating a machine through the use of discrete numerical values fed into the machine, where the required 'input' technical information is stored on a kind of input media such as floppy disk, hard disk, CD ROM, DVD, USB flash drive, or RAM card etc. The machine follows a predetermined sequence of machining operations at the predetermined speeds necessary to produce a work piece of the right shape and size and thus according to completely predictable results. A different product can be produced through reprogramming and a low-quantity production run of different products is justified.

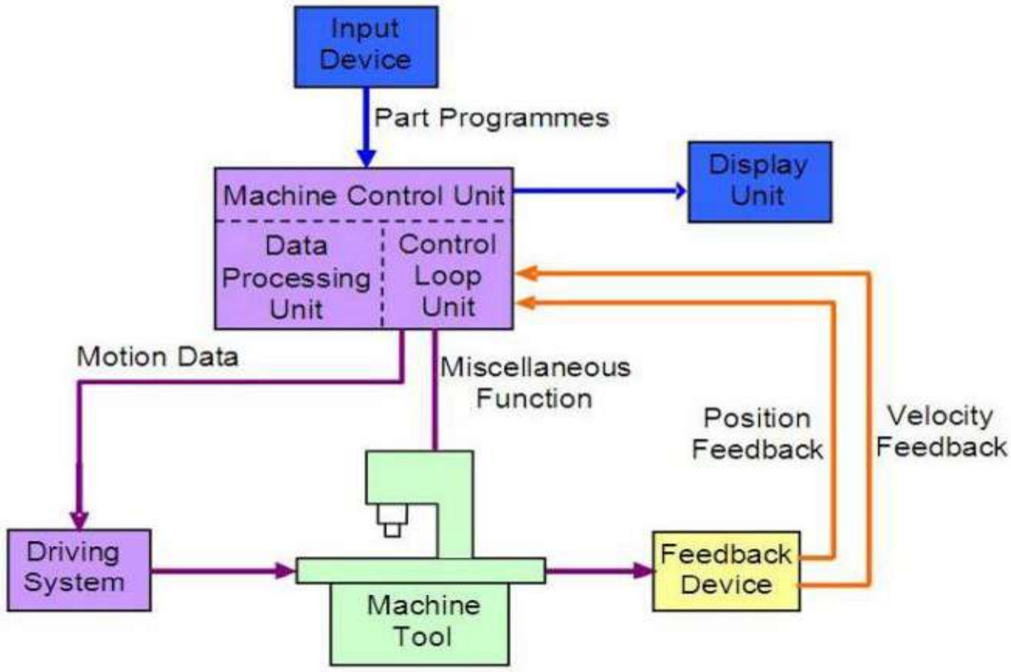


<u>Basic Principle of CNC</u>

Elements of a CNC System

A CNC system consists of the following 6 major elements:

a. Input Device
b. Machine Control Unit
c. Machine Tool
d. Driving System
e. Feedback Devices
f. Display Unit



a. Input Device

Floppy Disc Drive: Floppy disk is a small magnetic storage device for CNC data input. It had been most common storage media up to 1970s, in terms of data transfer speed, reliability, storage size, data handling and the ability to read and write. Furthermore, the data within a floppy disc could be easily edited.

USB Flash Drive: Removable and rewritable portable hard drive with compact size and bigger storage size than floppy disc.

Serial Communication: The data transfer between a computer and a CNC machine tool is often accomplished through a serial communication port. International standards for serial communications are established so that the information can be exchanged in an orderly way. The most common interface referred to the EIA Standard RS-232.

Ethernet Communication: Due to advancement of computer technology and the drastic reduction of the cost of the computer, it is becoming more practical and economical to transfer part programmes via an Ethernet communication cable. Provides more efficient and reliable means in part programming transmission and storage.

Conversational Programming: Part programmes can be input to the controller via the key board. Built-in intelligent software inside the controller enables the operator to enter the required data step by step.

b. Machine Control Unit (MCU)

The machine control unit is the heart of the CNC system. It accepts the information stored in the memory as part program.

- This data is decoded and transformed into specific position control and velocity signals.
- It also oversees the movement of the control axis or spindle and whenever this does not match with the programmed values, a corrective action is taken. There are two sub-units in the machine control unit: The Data Processing Unit (DPU) and the Control Loop Unit (CLU).

i. Data Processing Unit

On receiving a part programme, the DPU firstly interprets and encodes the part programme into internal machine codes. The interpolator of the DPU then calculate the intermediate positions of the motion in terms of BLU (basic length unit) which is the smallest unit length that can be handled by the controller. The calculated data are passed to CLU for further action.

ii. Control Loop Unit

The data from the DPU are converted into electrical signals in the CLU to control the driving system to perform the required motions. Other functions such as machine spindle ON/OFF, coolant ON/OFF, tool clamp ON/OFF are also controlled by this unit according to the internal machine codes.

c. Machine Tool

This can be any type of machine tool or equipment. In order to obtain high accuracy and repeatability, the design and make of the machine slide and the driving leadscrew of a CNC machine is of vital importance. The slides are usually machined to high accuracy and coated with anti-friction material such as PTFE and Turcite in order to reduce the stick and slip phenomenon. Large diameter recirculating ball screws are employed to eliminate the backlash and lost motion. Other design features such as rigid and heavy machine structure; short machine table overhang, quick change tooling system, etc also contribute to the high accuracy and high repeatability of CNC machines.

d. Driving System

The driving system is an important component of a CNC machine as the accuracy and repeatability depend very much on the characteristics and performance of the driving system. The requirement is that the driving system has to response accurately according to the programmed instructions. This system usually uses electric motors although hydraulic motors are sometimes used for large machine tools. The motor is coupled either directly or through a gear box to the machine leadscrew to moves the machine slide or the spindle. Three types of electrical motors are commonly used- DC Servo Motor, AC Servo Motor and Stepping Motor

Ball lead screw is the heart of the drive system. Advantages of ball lead screw are:

- Precise position and repeatability
- High Speed capability
- Less Wear
- Longer life

e. Feedback Device

In order to have a CNC machine operating accurately, the positional values and speed of the axes need to be constantly updated. Two types of feedback devices are normally used, positional feedback device and velocity feedback device.

i. Positional Feed Back Devices:

Linear Transducers - a device mounted on the machine table to measure the actual displacement of the slide in such a way that backlash of screws; motors etc. would not cause any error in the feedback data.

Rotary Encoders: A device to measure the angular displacement. It cannot measure linear displacement directly so that error may occur due to the backlash of screw and motor etc.

ii. Velocity Feedback Device:

The actual speed of the motor can be measured in terms of voltage generated from a tachometer mounted at the end of the motor shaft.

The voltage generated by the DC tachometer is compared with the command voltage corresponding to the desired speed. The difference of the voltages is used to actuate the motor to eliminate the error.

f. Display Unit

It is the interface between the machine and the operator. The Display Unit displays:

- Position of the machine slide
- Spindle rpm
- Feed rate
- Part programs
- Graphics simulation of the tool path.

Applications of CNC Machines

CNC machines are widely used in the metal cutting industry and are best used to produce the following types of product:

- ✓ Parts with complicated contours
- ✓ Parts requiring close tolerance and/or good repeatability
- ✓ Parts requiring expensive jigs and fixtures if produced on conventional machines
- ✓ Parts that may have several engineering changes, such as during the development stage of a prototype
- ✓ In cases where human errors could be extremely costly
- ✓ Parts that are needed in a hurry
- ✓ Small batch lots or short production runs

Some common types of CNC machines and instruments used in industry are as following:

- ➢ Drilling Machine
- ➢ Lathe / Turning Centre
- ➢ Milling / Machining Centre
- ➢ Turret Press and Punching Machine
- ➢ Grinding Machine
- ➢ Laser Cutting Machine
- ➢ Water Jet Cutting Machine
- ➢ Electro Discharge Machine
- ➢ Coordinate Measuring Machine
- ➢ Industrial Robot

Advantages:

- ✓ High Repeatability and Precision. e.g.: Aircraft parts.
- ✓ Volume of production is very high.
- ✓ Complex contours/surfaces can be easily machined.
- ✓ Flexibility in job change, automatic tool settings, less scrap.
- ✓ More safe, higher productivity, better quality.
- ✓ Less paper work, faster prototype production, reduction in lead times.

Disadvantages:

- ✗ Costly setup, skilled operators.
- ✗ Computer programming knowledge required.
- ✗ Maintenance is difficult.