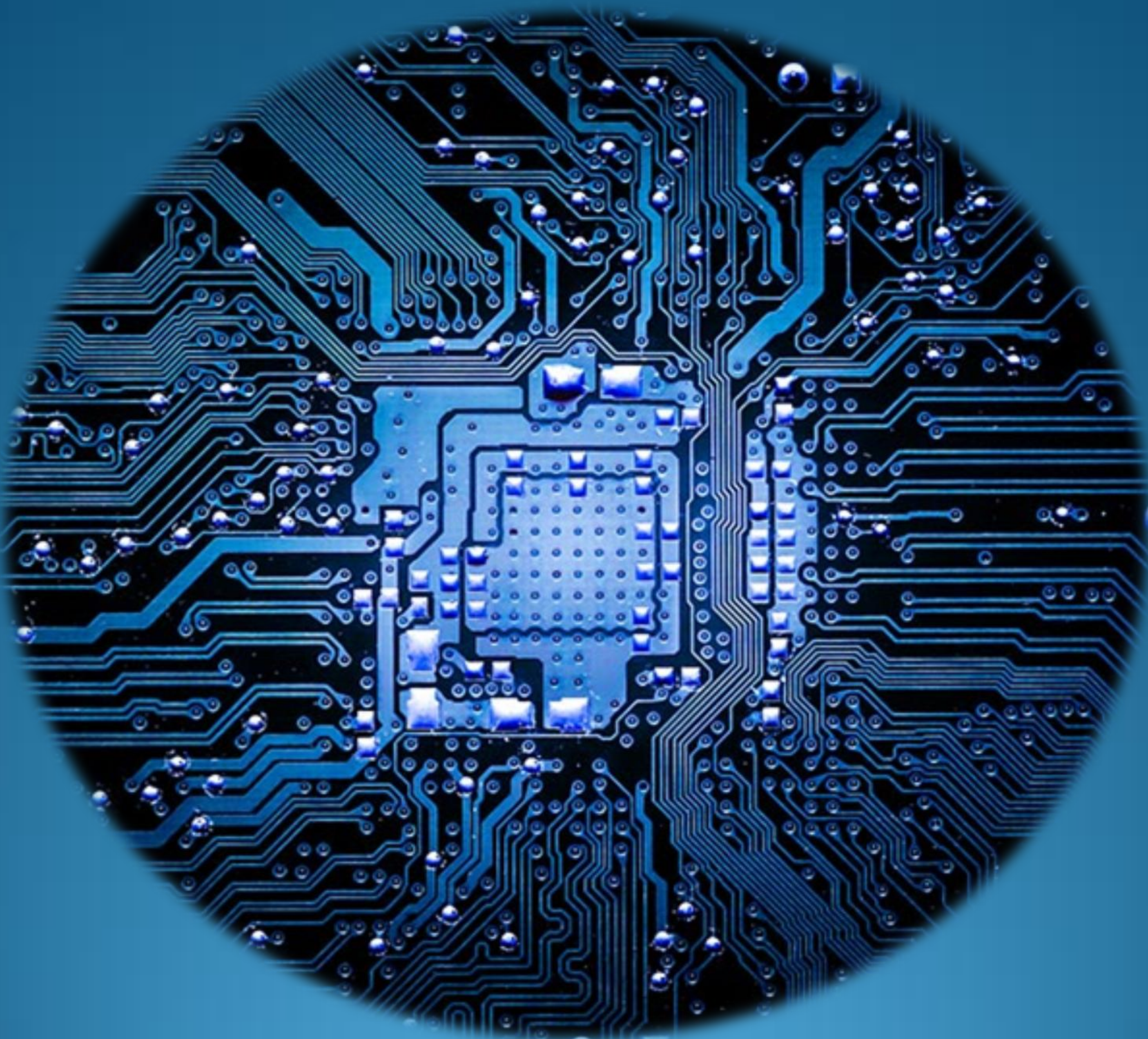
Applied Electronics and Instrumentation



Dr. B. Sridhar
Dr. K. V. Prakash

AgriMoon.Com

Applied Electronics and Instrumentation

-: Course Content Developed By :-

Dr. B. Sridhar

Professor and Head, Agricultural Machinery Research Centre

Agricultural Engineering College and Research Institute, TNAU, Coimbatore

-:Content Reviewed by :-

Dr. K. V. Prakash

Dept. of Farm Machinery, College of Agrl. Engg.

INDEX

MODULE 1.

LESSON 1. Electron- Energy bond in solids- Insulators-semiconductors and metals.

Atom

An atom is defined as the smallest particle into which an element can be divided and still retain the chemical properties of that element. Atom basically consists of electrons, proton and neutrons. The proton and neutron is altogether called as nucleus. The nucleus is the central portion of an atom. It is having positive charge. Around the nucleus small negatively charged particles called electrons revolve in different paths. The different paths are called either orbits or shells.

The electron

An electron is a negatively charged particle. It is moving around the nucleus. It is having negligible mass. Hence it is very mobile.

Properties

Charge of an electron = 1.6025×10^{-19} coulomb

Mass of an electron = 9.106×10^{-31} kg

The ratio of charge mass of an electron = 1.77×10^{11} coulombs / kg

Energy of electron

The electrons which are moving around the nucleus posses two types of energy.

1. Kinetic energy due to its motion
2. Potential energy due to the charge on the nucleus.

Hence, the total energy of the electrons are the sum of these two energies. The total energy increases as its distance from the nucleus increases. Thus the electrons in the last orbit posses very high energy as compared to the electrons in the inner orbits.

Nucleus

The central portion of an atom is called nucleus. It contains protons and neutrons. The proton is a particle carrying a positive charge. The positive charge is numerically equal to the negative charge of an electron. The neutron has no charge but it is having the same mass as the proton. So, the nucleus of an atom is positively charged, The mass of an electron is (1/1845 that of a porton) negligible. Hence the neutron and proton constitute the entire weight of an atom. This is called atomic weight. In an atom, the number of electrons are equal to the number of protons. This number is called as atomic number. The negative charges of the electrons are exactly balanced by the positive charges of the protons. Therefore net charge of an atom is zero.

The electrons are revolving around the nucleus in different orbits. Each orbit have different energy level. The electrons moving in the outermost orbit have highest potential energy. They can be easily disturbed by external influences. These electrons are known as free electrons or valance electrons. The last orbit is known as valance orbit. The electrons moving in the inner orbits which are closely situated to nucleus have very low potential energy. They are much influenced by the central portion nucleus. Hence, they cannot be much disturbed by external influences. They are known as bound electrons.

Electron arrangement in atoms

In general, electrons reside in groups of orbits called shells, The shells are elliptically shaped and are assumed to be located an fixed intervals. Thus the shells are arranged insteps that corresponding to fixed energy levels.

The shells and the number of electrons required to fill them may be predicted by the employment of Pauli's Exclusion Principle. Simply this principle specifies that each shall may contain no more than $2n^2$ electrons where 'n' corresponds to the small number starting with the one closest to the nucleus.

Starting with the shell closest to the nucleus and progressing outward, the shell are labeled K,L,M,N,O,P and Q respectively. The shells are considered to be full or complete when they contain the following quantities of electrons : 2 in K shell, 8 in the L shell, 18 in the M shell and 32 in the N shell. The formula $2n^2$ can be used to determine the number of electrons only in the four shells closest to the nucleus of an atom. Succeeding shells have as maximum number of electrons; O –shell-18 electrons. P shell-12 electrons, and Q shell –2 electrons.

Each of the shell is a major shell and can be divided into sub shells. They are labled S,P, d and f A sub-shell exists at a given energy (level that is, at a given distance from the nucleus.)

Like the major shells, the subshells are also limited as to the number of electrons which they can contain. Thus the S subshell is complete when it contains two electrons, the P subshell when it contain six, the d subshell when it contains ten and the last subshell f when it contains fourteen electrons.

The K shell can contain not more than two electrons, it must have one sub shell, the S subshell. The M shell is composed of 3 sub shells; S, P and d. The relationship exists between shells and sub shells upto and including the N shell. In such a way, the electrons are arranged in an atom.

Energy band in solids

1. Valence band
2. Conduction band
3. Forbidden energy gap

(i)Valence band

Electrons in the outer most orbit of an atom are called valence electrons. The range of energies possessed by valence electrons is known as valence band. This band may be completely or partially filled. It is the highest occupied band.

(ii) Conduction band

The electrons which left the valence band are called free electrons. The band occupied by these electrons is called the conduction band. This band is next in the valence band. It may either be empty or partially filled with electrons. In the conduction band, electrons move freely and conduct electric current through the solid.

(iii) Forbidden energy gap

The valence band and conduction band are separated by a gap on the energy band diagram is known as forbidden energy gap. There is no allowed energy state in the forbidden energy gap. If the width of the forbidden energy gap is greater means the valence electrons are tightly bounded to the nucleus and vice versa.

Insulators, Semiconductors and Metals

A semiconductor is a substance which conductivity lies in between these two extremes A material may be placed in one of these three class depending upon its energy band structure.

Insulator

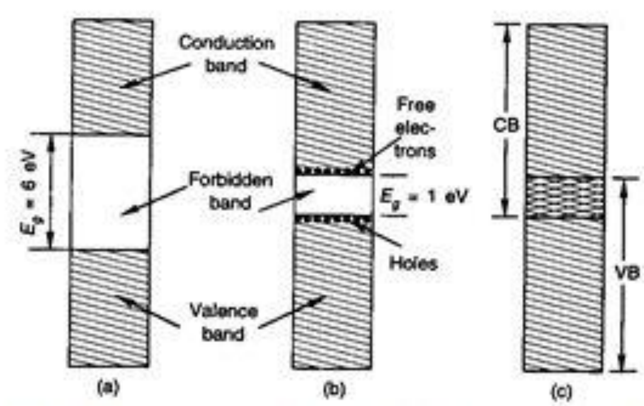


Fig 1.1. Energy band gap in (a) Insulation, (b) semiconductor (c) Metals

Insulator is a substance through which the passage of current is not allowed. The electrons in the valence band are bound very tightly to their parent atoms. Hence it requires very large electric field to remove them from their nuclei, in terms of energy band, the valence band is full, the conduction band is empty and the forbidden energy gap is very large between them. This is shown in Fig. 1.1(a). Therefore a very high electric field is required to carry the electron from the filled valence band into the empty conduction band. At higher temperature some electrons may go to the conduction band and in turn the insulator resistance decreases i.e., an insulator has negative temperature co-efficient of resistance.

Semiconductor

A semiconductor is a substance whose conductivity lies in between conductors and insulators.

At low temperature, the valence band is completely full and conduction band is completely empty. Hence a semiconductor behaves as an insulator at low temperatures. As the temperature is increased, more valence electrons cross over to the conduction band and the conductivity increases. Hence electrical conductivity of the semiconductor increases, with rising temperature i,e., a semi conductor has negative temperature co-efficient of resistance. In terms of energy band the valence band is filled and the conduction band is empty. Moreover, the energy gap between valence and conduction band is very small as shown in Fig. 1.1(b). Hence, it requires smaller electrical field to push the electron from the valence band to the conduction band.

Conductor

Conductor is a substance which easily allows the passage of electric current through it. It is because there are plenty of free electrons available in a conductor. In terms of energy band the valence and the conduction bands overlap each other as shown in Fig 1.1(c). In fact, there is no physical distinction between the two bands. A slight potential difference across a conductor is quite enough to cause the electrons to constitute electric current. There is no structure to establish holes since the forbidden energy gap is absent. Hence total current in such conductor is simply due to the flow of electrons.

Conduction in metals

In metals, atoms are kept very close to each other in a regular fashion, called crystal lattice. It is a repeating arrangement of atoms with in a crystal. The physical properties of a material are to a great extend depend upon the lattice structure of the material.

The atoms kept in the lattice have inter atomic reactions. Therefore the electrons in a particular orbit which have little bit difference in energy levels form large number of lines. That lines are regarded as continuous band of energy. A band is partially filled means that there is some free electrons. These free electrons move freely from one atom to another or a random manner inside the metal.

Depends upon the metal, one or two electrons are free to move in this manner. When an electric field is applied to a metal the free electrons are pulled towards the positive electrode. This drift towards the positive electrode by the external applied electrical field is superimposed on the random motion of the electrons due to thermal energy. Thus it constitutes an electric current that means, the material conduct electricity.

LESSON 2. Semiconductor-P& N type-drift and diffusion current- PN junction as diode.

SEMICONDUCTORS

A semiconductor is a substance which has receptivity in between conductors and insulators (e.g) germanium, silicon carbon etc.

Bonds in Semiconductor

The atoms of every element are held together by the bonding action of valence electrons. This makes the atom more active to enter into bargain with the other atom to acquire eight electrons in the last orbit. During this process the atom may lose, gain or share valence electrons with other atoms. In this case, there is only sharing of one or more valence electrons between the two atoms, each of which tries to fill up its outer most orbit. Such bonds are called covalent bonds. The germanium atom has four valence electrons in the outer most orbit. As seen each germanium atom shares one electron each with four surrounding atoms. In this way the central atom setup covalent bonds.

At 0^{O}K, all electrons in the covalent bonds are firmly held, there is no free carriers and hence becomes an insulator. However at room temperature some of the covalent bonds will be broken, because of the thermal energy applied to the crystal. In this way be the movement of free electron, holes are filled up and created.

The electron in the valence band moves away leaving a hole in its place. It is relatively easy for a valence electron in a neighboring atom to leave its covalent bond to fill this hole. An electron moving from a band to fill a hole leaves in its initial position. Hence the hole effectively moves in the direction opposite to electron and so electron-hole pairs are generated, when temperature of semiconductor is increased.

For each created electron-hole pair, two charge carrying particles are formed. One is free electron and the other is the hole. These particles move in opposite direction in an electric field, but since they are opposite sign, the current of each is in the same direction.

Conductivity of Pure Germanium

Fig. (2.1) shows the motion of free electrons and holes in a pure germanium when an external electrical field is applied.

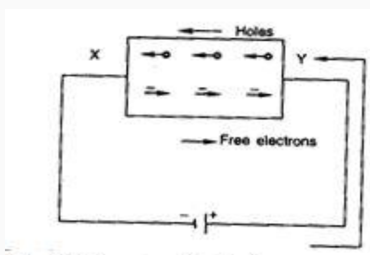


Fig. 2.1. Current conduction in

When the direction of the electrical field is from right to left as shown in Fig 2.1, the free electrons are pulled towards the right side whereas the holes are pulled towards the left side. The pulling forces are superimposed upon the thermal random motion of the free electrons and poles. The free electrons move from (point M) left to right and reaches point V, after some time. The net distance travelled. During the travel it collides with atoms at points N,O,P,Q,R,S,T and U. The net displacement of the electron is the cause for the flow of electric current. The conventional direction of electric current is just opposite to the flow of electrons.

In the same manner the holes are pulled towards the left side (negative electrodes). Assume that a hole is existing at point I Due to thermal energy the covalent band breaks. The electron escapes from the structure and thus a hole is formed. This is only for short time duration. Due to thermal agitation the electron kept at G may move to position I and thus once again the bond is completed. In turn a hole is created at junction G. In such a way hole is created at F,E,D, B and A. Thus the hole is displaced from I to A. This kind of displacement is called hole motion which is also the cause for the electric current.

Therefore, there exists two kind of charge namely electrons and holes in a pure germanium. The conductivity of pure germanium depends upon the mobile charge carriers and also their mobility.

Types of Semiconductor

Semiconductor may be classified as follows:

1. Intrinsic or pure semiconductor
2. Extrinsic or impure semiconductor

Intrinsic Semiconductor

Pure semiconductor is known as an intrinsic semi-conductor. Pure germanium and silicon are common examples. These semiconductors have four valence electrons which are easily affected by external influences. Covalent bond is formed only by means of these electrons with the neighboring atoms.

The last orbit electrons (valence electrons) energy are shared with neighboring atoms and hence each atom is surrounded by a completely filled orbits at 0°K. These materials behave as perfect insulator. But even at room temperature hole-electron pairs are produced. When external electric field is applied across the intrinsic semiconductor the current conduction take place. That current conduction is due to free electrons and holes. The number of electrons is equal to the number of holes in an intrinsic semiconductor at room temperature. The energy gap is moderate of an intrinsic semiconductor (pure. germanium) At 0°K the valence band is completely filled with electrons and the conduction, band is empty, Fermi level is nothing but the energy which corresponds to the centre of gravity of conduction electrons and holes.

Extrinsic Semiconductors

A pure semiconductor is called intrinsic semiconductor. The only current carriers in this pure semiconductor are electron hole pairs. In most of the applications, these produce feeble current. So, for increasing either the number of free electrons or the number of holes, some impurity atoms are added with the pure semiconductor. This adding process is called doping. After the impurity atoms are doped with pure semiconductor it is called an extrinsic Semiconductor.

Depending upon the types of impurity atoms that have been added with the pure semiconductor, the extrinsic semiconductor is classified as

a) P-type semiconductor

b) N-type semiconductor

P-type Semiconductor

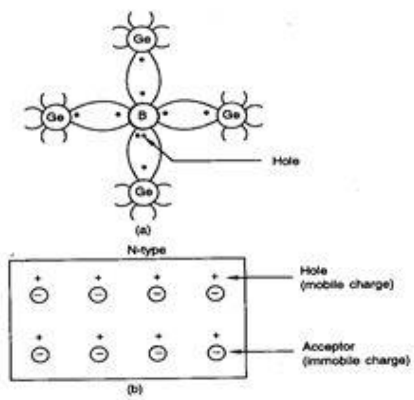


Fig.2.2. P-type semiconductor: (a) Formation of covalent bonds and (b) Charged carriers

Let us consider a germanium crystal. The germanium atom has four valence electrons. If we doped with an impurity atom of having only three electrons (Boron atom) in its valence band with this germanium atom, a new crystal is formed. The new crystal will have its structure and energy band diagram as shown in Fig.2.2. In this newly formed extrinsic semiconductor we find that a trivalent atom is in between four neighboring germanium atoms as shown in Fig. 2.2. It is further noted that only seven electrons travel in its valence orbits. In other words one hole appears in each trivalent atom. The number of holes in the doped crystal can be increased by adding more number of impurity atoms.

A pure semiconductor doped by trivalent (aluminium, boron and gallium) impurity is known as P-type semiconductor . In this P-type semiconductor, conduction is by means of holes in the valence band. So the holes are considered as majority carries in P-type semi-conductor whereas electrons are minority carriers. The trivalent provide room for the germanium atom. If the trivalent impurity is added with germanium the new extrinsic crystal is formed as shown in Fig. 1.3. The strong force holding the structure is called convalent bond. The minimum heat energy required to break the germanium extrinsic covalent bond is 0.1 eV and 0.05ev for silicon extrinsic semiconductor. In this type, concentration of electrons is in the conduction band. Conduction is by the hole movement in the valence band. The acceptor level accopts the electrons from the valence band.

N-type Semiconductor

If pentavalent impurities are added with extrinsic semiconductor a new crystal is formed.

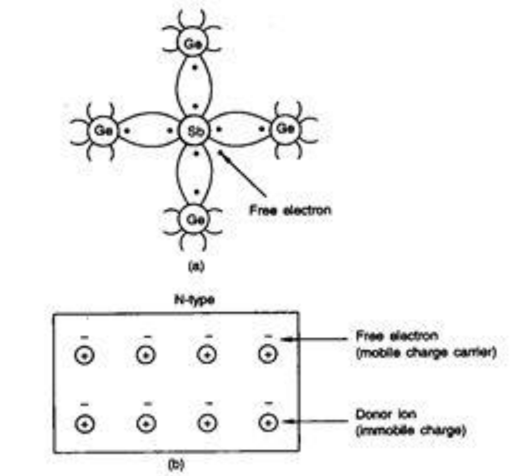


Fig.2.3. P-type semiconductor: (a) Formation of covalent bonds and (b) Charged carriers

The new crystal will have its structure and energy band diagram as shown in Fig 2.3. In this structure, we find that a pentavalent atom is in between four neighboring germanium atoms. After forming covalent bonds with four neighbors, this central atom has an extra electron left over. Since the valence orbit cannot hold more than eight electrons, the extra electron must travel in a conduction band orbit. Since each pentavalent atom denote one electron, the number of electrons in this structure can be increased by doping of more number of impurity atoms. Suitable pentavalent impurities are antimony, and arsenic.

In N-type extrinsic semiconductor adding of pentavalent impurity like antimony (Sb) increases the number of conduction electrons. Hence, concentration of electrons in the conduction band is increased and exceeds the concentration of holes in the valence band. Due to this 'Fermi level shifts upwards towards the bottom of the conduction band .The energy level for fifth electron (donor level) is 0.001eV below the conduction band for germanium and 0.05eV for silicon. In this type of extrinsic semiconductor, conduction is by means of electrons which are called the majority carriers whereas holes are called minority carriers.

Drift current

A pure semiconductor at 0^{o} Kevin behave as an insulator. When the temperature is raised above absolute zero more electrons are detached from the covalent bond. Thus the electron-hole pairs are increased, when the temperature is raised. These free electrons and holes move in random manner. When the electric field is applied across the semiconductor bar, it forces the randomly moving electron-hole pairs in its direction (field direction). This causes a current flow in the circuit. This current is called drift current.

Diffusion current

Let us assume that one type of charge carriers concentration is accumulated at one end of piece of semiconductor material. Since the charge carries are all the same polarity there is a force of repulsion between them. The result is that there is a tendency for the charge density to one of low density. This movement continues until all the carriers are evenly distributed throughout the material. Thus the movement of charge carriers constitutes an electric current. This type of current is known as diffusion current.

P-N Junction as diode

The simplest semiconductor device has to region, one regions doped with p-type impurity and the second region doped the n type impurity. The impurity concentrate changes from donar to acceptor at boundary.

a. Depletion region

Both the region have equal positive and negative charges and are electrically neutral. However the n-type has more free electron then holes and p-type has more holes and p-type has more holes than free electron. The free electron on n-type material tend to diffuse. Each electron diffuse across the juctia into the p-type one. Each electron diffusia from the n-region into the p-region leaves a positive charges behind in the n-region. Similarly the holes in the p-type material diffuse across the junction into the n-region leavings negatives charges behind. As this the n-region leavings negatives charges behind. As this diffusion occurs, then-region becomes positively charged and the p-region becomes negatively charged. Due to this displacement of charges, an electric field appears at the junction. Equilibrium is established when this field large enough to restrain further diffusion. The holes which neutralized the acceptor ion hear the junction in the p-type materials disappear due to recombinant with electron which have diffused from n-type materials across the junction. Later wise, the neutralizing electron in the n-type material combine with the holes which have crossed the junction from the p-type material. The electron charges are confined to the neighbourhood of the junction and consists of in mobile ions. Since the region around the function in depleted of mobile charges it is known a depletion region or transit region or the space charge region. The width of this depletion region is about 10^{-4} cm. There is an equilibrium potential difference vs across the depletion region.

b. Barrier potential

The equilibrium potential across the depletion region is known a barrier potential. At 25°C the barrier potential. At 25°C the barrier potential is 0.3 V for germanium and 0.74 for silica diode. The temperature at the junction of p and n-type materials is known as function temperature and more to different than the ambient temperature. The barrier potential depend as the junction temperature. At higher junction temperature more hole and free electron present. This mean a reduction in the width of the depletion layer, this decreasing the barrier potential. For each ie. Rise 1a junction temperature, the barrier potential of both silicon and germicum decreased by 2 mV.

When a diode is forward biased, the current remain zero till applied voltage overcomes the barrier potential. Therefore in a silicon diode biased in forward direction, the current flows in the circuit only when the applied voltage is more than 0.7 V.

c. Effect of forward bias

The forward bias is applied by connecting the terminal of a battery to p-type material and the negative battery terminal to n-type material. The potential and the P-type material is raised (art to n-type material) and an electric field is set up. The terminal of the battery removes electron from p material leaving holes there. The negative terminal of the battery inject electron into n material. As a result of the above, the width of the depletion bore is reduced. The free electron cross the junction from n-type to p-type and holes from p-type to n-type. These majority carriers travel around the closed loop and a steady current flows in the circuit through p-n junction. This connection is also known as the easy current direction.

Effect of reverse bias

The p-n junction is reverse biased when the terminal of the battery is connected to the n-type material and –ve terminal to P–type material. The polants the connector cause be holes in P–type and electron in P–type materials to move away from the junction. Here by increasing the depletion region. The current in normally zero.

Thus the P – type junction ha prospects of conducting current only when biased in the forward direct and water as a diode.

Effect of short circuit

When a P- type junction in short circuited no current can flow in the circuit, the electrostatic potential remain of the fans value as that rather open circuit condition.

LESSON 3. V I characteristics of diode-Zener diode- Zener and breakdown- V I characteristics of a Zener diode

Volt –ampere characteristics of semiconductor Diode

The figure shows the circuit for determining the volt-ampere characteristics of a semiconductor diode. By the use of a potential divider, the applied voltage can be varied both in the forward and revere direction.

Whom the applied voltage in few, a small equalization potential Vo exists across the depletion region. As the applied voltage in increased more and more, the correct goes an increasing. The current is bipolar the character since it consultant –ve and the carrier (electron and holes). The total current in constant throughout the circuit but the properties of current due to holes and electron varies with distance along the P-N bar.

Whom a reverse voltage is applied (b main the slide r s below 0), only a small current flows. This small current is the reverse ** current Io. IF the operating temp in increased, Io increase. When the reverse voltage reaches a critical value, the reverse current through the diode increase abruptly and a relatively large current can flow with little increase in voltage. This phenomenon, known as reverse break down, occurs because electron gain enough energy so that ion zeta by collision takes place and the covalent bonds are distributed. The release a large number of electron and a large reverse current occurs. This breakdown happens at the zener breakdown voltage and the diode in destroyed. However, same diodes are built to operate specifically in the zener breakdown region and are known as zener diodes.

EQUIVALENT CIRCUIT OF A PN JUNCTION DIODE

When forward biased the junction offers ac resistance r_{ac} and possesses diffusion capacitance C_d. C_d is taken into account at high frequencies only. Hence the equivalent circuit as shown in Fig. can represent it.

An opposing battery has been connected in series with r_{ac} to account for the junction barrier potential. Resistance R_R connected in parallel with capacitance C_T can be represent a reverse biased junction.

ZENER DIODE

It is a reversed biased, heavily doped silicon or germanium PN junction diode that is operated in the breakdown region where current is limited by both external resistance and power dissipation of the diode. Silicon is preferred to germanium because of its higher temperature and current capabilities.

If the reverse bias applied to a PN junction is increased, a point is reached when the junction breaks down and reverse current rises to a value limited only by the external resistance connected in series with the junction. This critical voltage is known as break down voltage

V_{BR}. Once break down has occurred, very little further increase in voltage is required to increase the current to relatively high values. The junction offers almost zero resistance at this point. The breakdown voltage depends on the width of the depletion region, which in turn depends on the doping level. Two mechanisms are responsible for break down under increasing reverse voltage.

Zener break down

This form of break down occurs in junctions which being heavily doped have narrow depletion layers. The break down voltage sets a very strong electric field about 10^8 V/m across this narrow layer. This field is strong enough to break or rupture the covalent bonds thereby generating electron –hole pairs. Even a small further increase in reverse voltage is capable of producing large number of current carriers. That is why the junction has very low resistance in the break down region.

Avalanche break down

This form of break down occurs in junctions which being lightly doped have wide depletion layers where the electric field is not strong enough to produce Zener break down. Instead the minority collide with the semiconductor atoms in the depletion region. Upon collision with valance electrons, covalent bonds are broken and electron-hole pairs are generated. These newly generated charge carriers are also accelerated by the electric field resulting in more collision and hence further production of charge carriers. This leads to an avalanche of charge carriers and subsequently, to a very low resistance.

V-I CHARACTERSTICS OF A ZENER DIODE

The forward characteristic is simply that of a ordinary forward biased junction diode.

The reverse characteristics indicates certain points.

1. Zener break down voltage
2. Minimum current to sustain break down.
3. Maximum Zener current limited by maximum power dissipation.

Since its reverse characteristics is not exactly vertical, the diode posse some resistance zener dynamic impedance.

$$Z_z = \frac{\Delta V_z}{\Delta I_z}$$

It is negligible compared to the large external resistance connected in the circuit.

MODULE 2.

LESSON 4. Diode circuits-rectifier circuits-Half and full wave rectifier-Bridge rectifier-comparison

Diode Circuits

Diode circuits to be considered perform functions such as rectification, clipping. and clamping. These functions are possible only because of the nonlinear properties of the pn junction diode. The conversion of an ac voltage to a dc voltage, such as for a dc power supply, is called rectification. Clipper diode circuits clip portions of a signal that are above or below some reference level. Clamper circuits shift the entire signal by some dc value, Zener diodes, which operate in the reverse-bias breakdown region, have the advantage that the voltage across the diode in this region is nearly constant over a wide range of currents. Such diodes are used in voltage reference or voltage regulator circuits.

Finally, we look at the circuits of two special diodes: the light-emitting diode (LED) and the photodiode. An LED circuit is used in visual displays, such as the seven-segment numerical display. The photodiode circuit is used to detect the presence or absence of light and convert this information into an electrical signal.

4.1 RECTIFIER CIRCUITS

One important application of diodes is in the design of rectifier circuits. A diode rectifier forms the first stage of a dc power supply as shown in Figure 4.1 below.

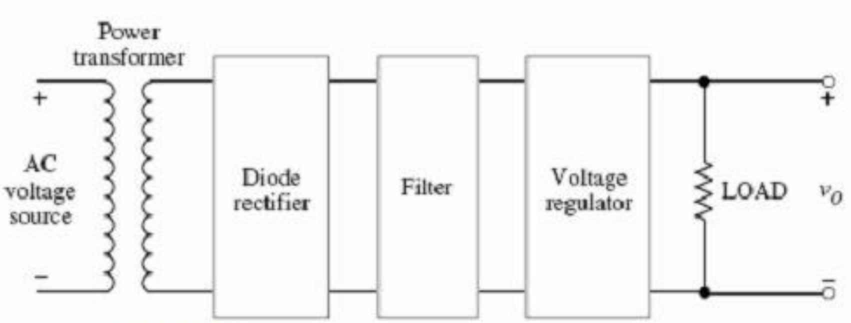


Figure 2.1 Block diagram of an electronic power supply

As we will see throughout the text, a dc power supply is required to bias all electronic circuits. The dc output voltage v_o will usually be in the range of 3 to 24V depending on the application.

Rectification is the process of converting an ac voltage to one polarity. The diode is useful for this function because of its nonlinear characteristics, that is, current exists for one voltage polarity, but is essentially zero for the opposite polarity. Rectification is classified as half-wave or full-wave, with half-wave being the simplest.

PN Junction diode as rectifier

The process in which alternating voltage or alternating current is converted in to direct voltage or direct current is known as rectification. The device used for this process is called as rectifier. The junction diode has the property of offering low resistance and allowing current to flow through it, in the forward biased condition. This property is used in the process of rectification.

Half wave rectifier

A circuit which rectifies half of the a.c. wave is called half wave rectifier.

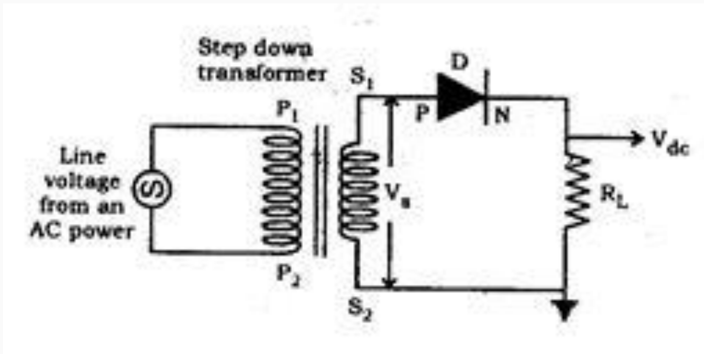


Fig.4.2. Half wave rectifier

Fig. 4.2 shows the circuit for half wave rectification. The a.c. voltage (V_S) to be rectified is obtained across the secondary ends S_1 S_2 of the transformer. The P-end of the diode D is connected to S_1 of the secondary coil of the transformer. The N-end of the diode is connected to the other end S_2 of the secondary coil of the transformer, through a load resistance R_L. The rectified output voltage V_{dc} appears across the load resistance R_L.

During the positive half cycle of the input a.c. voltage (V_S), S_1 will be positive and the diode is forward biased and hence it conducts. Therefore, current flows through the circuit and there is a voltage drop across R_L. This gives the output voltage as shown in Fig.4.3.

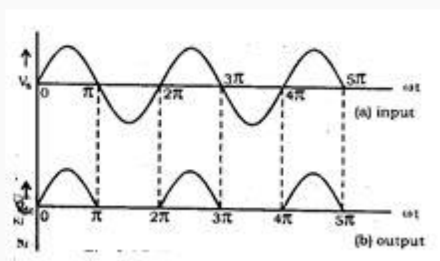


Fig.4.3. Half wave rectifier signals

During the negative half cycle of the input a.c. voltage (V_S), S_1 will be negative and the diode D is reverse biased. Hence the diode does not conduct. No current flows through the circuit and the voltage drop across R_L will be zero. Hence no output voltage is obtained. Thus corresponding to an alternating input signal, unidirectional pulsating output is obtained.

The ratio of d.c. power output to the a.c. power input is known as rectifier efficiency. The efficiency of half wave rectifier is approximately 40.6%.

Full wave rectifier

The circuit which rectifies both half cycles of the a.c. wave is called full wave rectifier.

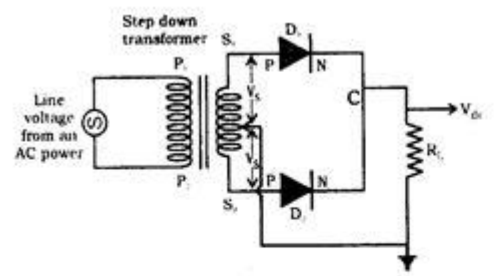


Fig.4.4. Full wave rectifier

Fig. 4.4 shows the full wave rectifier circuit using two diodes. The a.c. input voltage (V_S) to be rectified is obtained from the secondary ends S_1 and S_2 of the transformer. The P-ends of the diodes D_1 and D_2 are connected to the secondary ends S_1 and S_2 of the transformer. The centre tap C is connected to the load resistance, which in turn is connected to the junction of N-ends of the diodes. The rectified output voltage, V_{dc} appears across the load resistance R_L. Because of the centre tap, the circuit is equivalent to two half-wave rectifiers. The upper diode handles the positive half cycle of the secondary voltage, while lower diode handles the negative half cycle of the secondary voltage.

During positive half cycle of the a.c. input, diode D_1 is forward biased and it conducts and causes current flow in the load resistance R_L. Diode D_2 is reverse biased. Therefore, no current flows through it. During the negative half cycle of the a.c. input, diode D_2 is forward biased and it conducts. Hence, current flows through R_L. During this process, D_1 is reverse biased and no current flows through it.

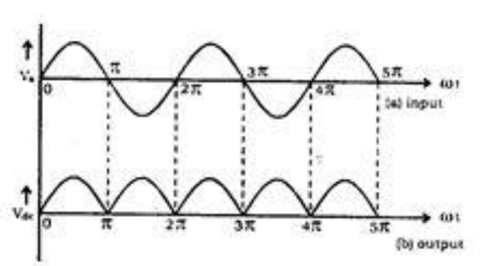


Fig.4.5. Full wave rectifier signal

Thus, whether the input signal is positive or negative, current always flows through the load resistance in the same direction. Hence full wave rectification is obtained. Fig. 4.5 shows the output voltage corresponding to the given input voltage. The efficiency of rectification is approximately 81.2%. The efficiency of full wave rectifier is twice that of half wave rectifier.

Bridge rectifier

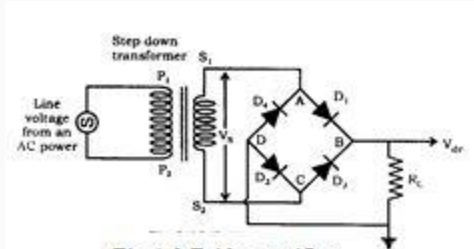


Fig.4.6. Bridge rectifier

A bridge rectifier is shown in Fig.4.6. There are four diodes D_1, D_2, D_3 and D_4 used in the circuit, which are connected to form a network. The input ends A and C of the network are connected to the secondary ends S_1 and S_2 of the transformer. The output ends B and D are connected to the load resistance R_L.

During positive input half cycle of the a.c. voltage, the point A is positive with respect to

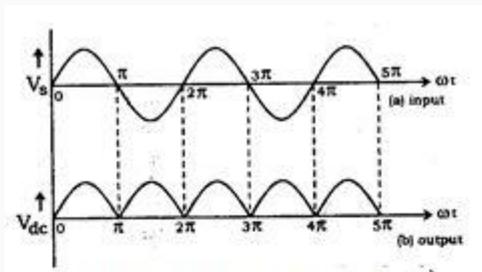


Fig.4.7. Bridge rectifier signal

C.

The diodes D_1 and D_3 are forward biased and conduct, whereas the diodes D_2 and D_4 are reverse biased and do not conduct. Hence current flows along S_1ABDCS_2 through R_L. During negative half cycle, the point C is positive with respect to A. The diodes D_2 and D_4 are forward biased and conduct, whereas the diodes D_1 and D_3 are reverse biased and they do not conduct. Hence current flows along S_2CBDAS_1 through R_L. The same process is repeated for subsequent half cycles. It can be seen that, current flows through R_L in the same direction, during both half cycles of the input a.c. signals. The output signal corresponding to the input signal is shown in Fig.4.7. The efficiency of the bridge rectifier is approximately 81.2% Because the full secondary voltage is applied to the conducting diodes in series with the load resistance, the load voltage is twice that of the full-wave rectifier.

Advantages of bridge rectifier

1. Centre tap on the secondary of the transformer is not necessary
2. Smaller transformer can be used
3. It is suited for high voltage applications

Some of the important terminologies related to Rectifier are:

Cut in or threshold voltage (Vr): It is the forward voltage in a diode under forward bias condition, below which the current is very small. It is 0.3V foir Germanium and 0.7V for Silicon. Vi = Vm Sinwt, where Vm >>Vr

Rectifier Efficiency (h): The ratio of dc output to ac input power is known as rectifier efficiency (h) h = (dc output power)/(ac input power)

Ripple Factor (t): The ratio of rms value of ac component to the dc component in the output is known as Ripple Factor (t)

Ripple Factor (t) = (rms value of ac component)/dc component)

Peak Inverse Value (PIV): It is defined as the maximum reverse voltage that a diode can withstand without destroying the junction. For half wave rectifier PIV = Vm.

Transformer Utilization factor (TUF):

TUF = (dc power delivered to the load)/(ac rating of the transformer secondary)

$= P_{dc}/P_{ac\ rated}$

Form factor:

$$\text{Form factor} = \frac{\textit{rms value}}{\textit{average value}}$$

Peak factor:

$$\text{Peak factor} = \frac{\textit{peak value}}{\textit{rms value}}$$

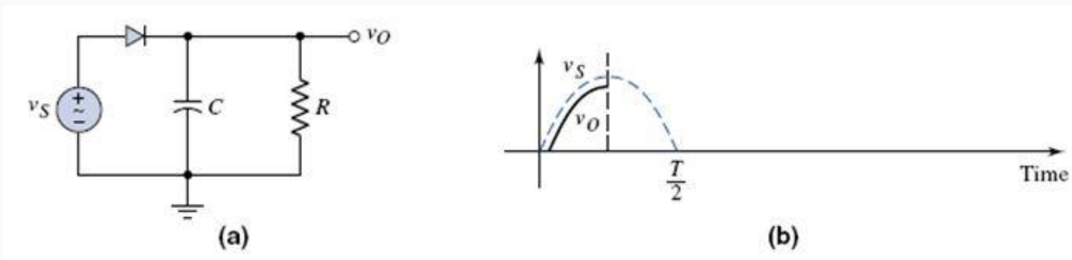
Comparison of rectifiers

SI.No	Parameter	Type of the rectifier		
		Halfwave	Fullwave	Bridge
1.	Number of diodes	1	2	4
2.	V_{dc}	V_m/π	$2V_m/\pi$	$2V_m/\pi$
3.	Peak inverse voltage	V_m	$2V_m$	V_m
4.	Ripple factor	1.21	0.48	0.48
5.	Rectifier efficiency	40.6%	81.2%	81.2%
6.	Transformer Utilization factor	0.287	0.693	0.812
7.	Form factor	1.57	1.11	1.11

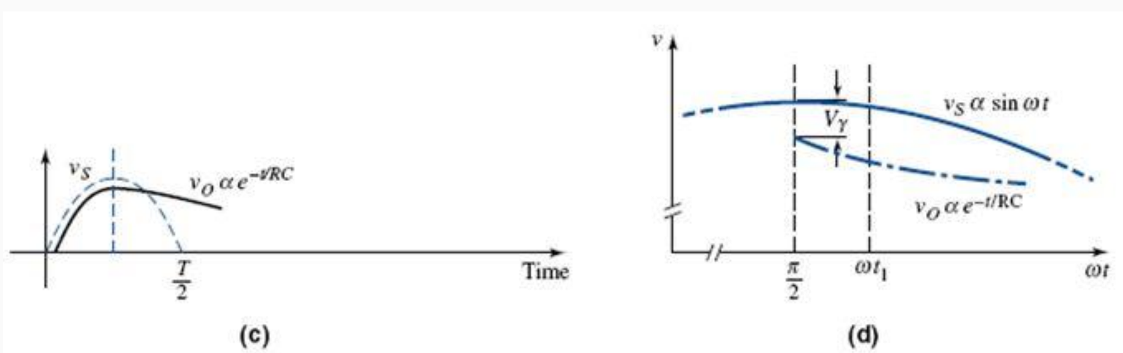
LESSON 5. Full wave rectifier with RC filter- Zener diode circuit- multiple diode circuits-photo diode and LED circuit.

Filters, Ripple Voltage, and Diode Current

If a capacitor is added in parallel with the load resistor of a half-wave rectifier to form a simple filter circuit (Figure 5.1(a)), we can begin to transform the half-wave sinusoidal output into a dc voltage. Figure 5.2(b) shows the positive half of the output sine wave, and the beginning portion of the voltage across the capacitor, assuming the capacitor is initially uncharged.



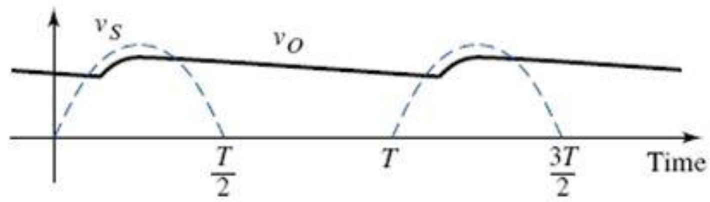
When the signal voltage reaches its peak and begins to decrease, the voltage across the capacitor also starts to decrease, which means the capacitor starts to discharge. The only discharge current path is through the resistor. If the RC time constant is large, the voltage across the capacitor discharges exponentially with time (Figure 2.8(c)). During this time period, the diode is cut-off.



A more detailed analysis of the circuit response when the input voltage is near its peak value indicates a subtle difference between actual circuit operation and the qualitative description. If we assume that the diode turns off immediately when the input voltage starts to decrease from its peak value, then the output voltage will decrease exponentially with time, as previously indicated. An exaggerated sketch of these two voltages is shown in Figure 5.1(d). The output voltage decreases at a faster rate than the input voltage,which means that at time t_1 the voltage across the diode, is greater than V_γ. However, this condition cannot exist and the diode does not turn off immediately. If the RC time constant is large, there is only a small difference between the time of the peak input voltage and the time the diode turns off.

During the next positive cycle of the input voltage, there is a point at which the input voltage is greater than the capacitor voltage, and the diode turns back on. The diode remains on until the input reaches its peak value and the capacitor voltage is completely recharged.

Since the capacitor filters out a large portion of the sinusoidal signal, it is called a filter capacitor. The steady-state output voltage of the RC filter is shown in Figure 5.1 e).



(e)

The ripple effect in the output from a full-wave filtered rectifier circuit can be seen in the output waveform in Figure 5.2.

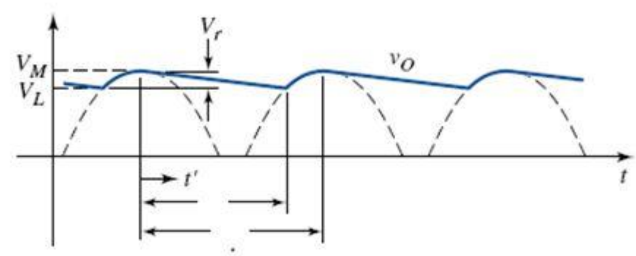


Figure 2.9 Output voltage of a full-wave rectifier with an *RC* filter

The capacitor charges to its peak voltage value when the input signal is at its peak value. As the input decreases, the diode becomes reverse biased and the capacitor discharges through the output resistance R. Determining the ripple voltage is necessary for the design of a circuit with an acceptable amount of ripple.

If we can assume that the ripple effect is small we get the following approximation

$$V_r \cong V_M\left(\frac{T_p}{RC}\right)$$

where T_p is the time between peak values of the output voltage.

For a full-wave rectifier T_p is one-half the signal period. Therefore, we can relate T_p to the signal frequency,

$$f = \frac{1}{2T_p}$$

and the ripple voltage becomes

$$V_r = \frac{V_M}{2fRC}$$

The diode in a filtered rectifier circuit conducts for a brief interval Δt near the peak of the sinusoidal input signal (Figure 5.3a).

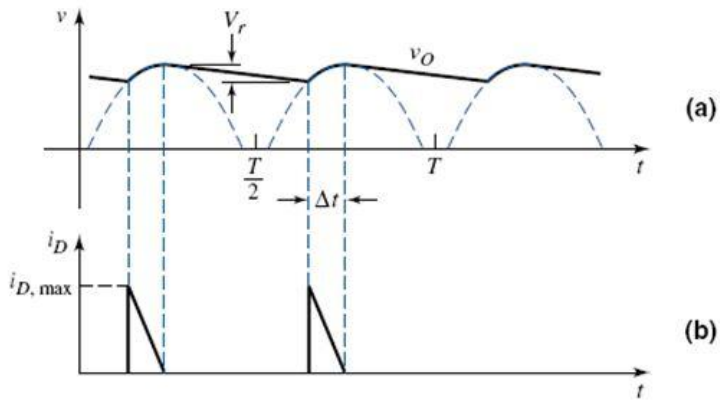


Figure 2.10 Output of a full-wave rectifier with an *RC* filter

ZENER DIODE CIRCUITS

The breakdown voltage of a Zener diode was nearly constant over a wide range of reverse-bias currents. This makes the Zener diode useful as a voltage regulator, or a constant-voltage reference circuit. In this chapter, we will look at an ideal voltage reference circuit, and the effects of including a non-ideal Zener resistance.

The results of this section will then complete the design of the electronic power supply .We should note that in actual power supply designs, the voltage regulator will be a more sophisticated integrated circuit rather than the simpler Zener diode design that will be developed here. One reason is that a standard Zener diode with a particular desired breakdown voltage may not be available.

Ideal Voltage Reference Circuit

Figure 5.4 shows a Zener voltage regulator circuit. For this circuit, the output voltage should remain constant, even when the output load resistance varies over a fairly wide range, and when the input voltage varies over a specific range.

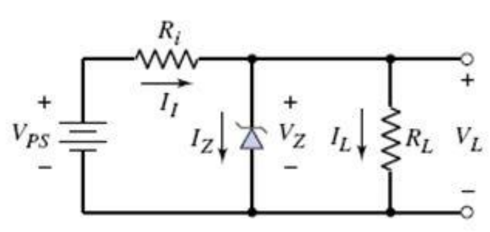


Figure 2.15 A Zener diode voltage regulator circuit

We first determine the proper input resistance R_i. This resistance limits the current through the Zener diode and drops the excess voltage between V_{PS} and V_Z.

$$R_i = \frac{V_{PS} - V_Z}{I_I} = \frac{V_{PS} - V_Z}{I_Z + I_L}$$

which assumes that the Zener resistance is zero for the ideal diode. Solving this equation for the diode current. I_z, we get

$$I_Z = \frac{V_{PS} - V_Z}{R_i} - I_L$$

where $1_L = V_z / R_L$, and the variables are the input voltage source V_{PS} and the load current 1_L.

For proper operation of this circuit, the diode must remain in the breakdown region and the power dissipation in the diode must not exceed its rated value. In other words:

1. The current in the diode is a minimum, I_z(rnin), when the load current is a maximum, I_L(max), and the source voltage is a minimum, V_{PS}(min).
2. The current in the diode is a maximum, I_z(max), when the load current is a minimum, I_L(min), and the source voltage is a maximum, V_{PS}(max).

Inserting these two specifications into the previous equation, we obtain

$$R_i = \frac{V_{PS}(\min) - V_Z}{I_Z(\min) + I_L(\max)}$$

and

$$R_i = \frac{V_{PS}(\max) - V_Z}{I_Z(\max) + I_L(\min)}$$

Equating these two expressions, we then obtain

$$[V_{PS}(\min) - V_Z] \cdot [I_Z(\max) + I_L(\min)] \\ = [V_{PS}(\max) - V_Z] \cdot [I_Z(\min) + I_L(\max)]$$

Reasonably, we can assume that we know the range of input voltage, the range of output load current, and the Zener voltage. The previous equation then contains two unknowns I_z(min) and I_z(max). Further, as a minimum requirement, we can set the minimum Zener current to be one-tenth the maximum Zener current, or $I_Z(\min) = 0.1\ I_Z(\max)$. We can then solve for I_Z(max), using the previous equation, as follows:

$$I_Z(\max) = \frac{I_L(\max) \cdot [V_{PS}(\max) - V_Z] - I_L(\min) \cdot [V_{PS}(\min) - V_Z]}{V_{PS}(\min) - 0.9V_Z - 0.1V_{PS}(\max)}$$

Use the maximum current obtained from the above equation, we can determine the maximum required power rating of the Zener diode. Then we can determine the required value of the input resistance using one of the previous equations.

MULTIPLE-DIODE CIRCUITS

Since a diode is a nonlinear device, part of the analysis of a diode circuit involves determining whether the diode is "on" or "off." If a circuit contains more than one diode, the analysis is complicated by the various possible combinations of "on" and "off."

In this section, we will look at several multiple-diode circuits. We will see, for example, how diode circuits can be used to perform logic functions.

Example Diode Circuits

To review briefly, consider two single-diode circuits. Figure 5.5(a) shows a diode in series with a resistor. A plot of voltage transfer characteristics, v_0 versus v_I, shows the piecewise linear nature of this circuit (Figure 5.5{b)).

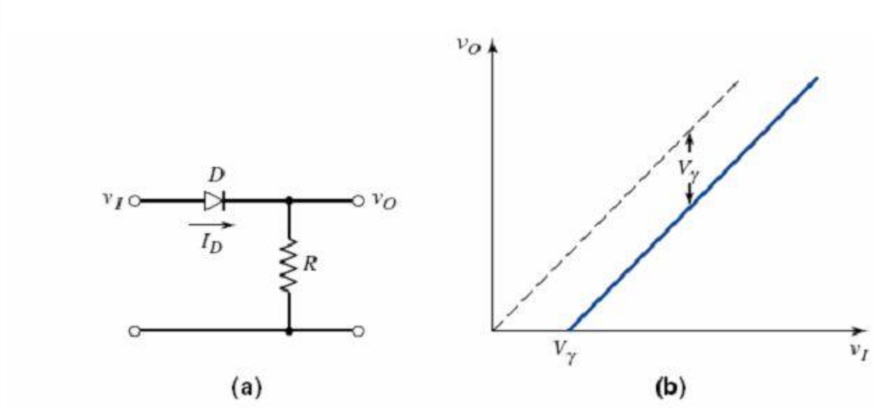


Figure 2.30 Diode and resistor in series: (a) circuit and (b) voltage transfer

The diode does not begin to conduct until $v_I = V_\gamma$. Consequently, for $v_I < V_\gamma$ the output voltage is zero; for $v_I > V_\gamma$ the output voltage is $v_o = v_I - V_\gamma$.

Figure 5.6(a) shows a similar diode circuit, but with the input voltage source explicitly included to show that there is a path for the diode current. The voltage transfer characteristic is shown in Figure 5.6(b).

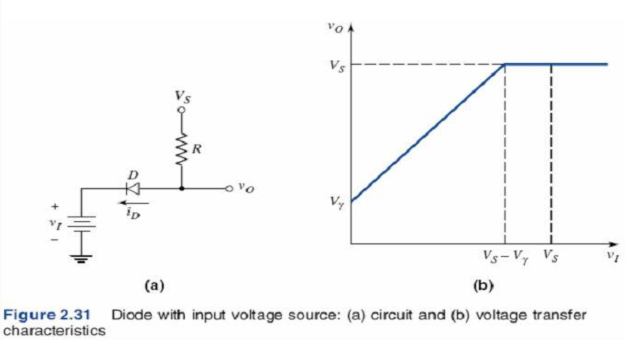


Figure 2.31 Diode with input voltage source: (a) circuit and (b) voltage transfer characteristics

In this circuit, the diode remains conducting for $v_I < V_S - V_\gamma$, and the output voltage is $v_0 = v_I + V_\gamma$. When $v_I > V_S - V_\gamma$, the diode turns off and the current through the resistor is zero; therefore, the output remains constant at V_S.

In multidiode circuits, each diode may be either "on" or "off." Consider the two-diode circuit in Figure 5.7. Since each diode may be either on or off, the circuit has four possible states. However, some of these states may not be feasible because of diode directions and voltage polarities.

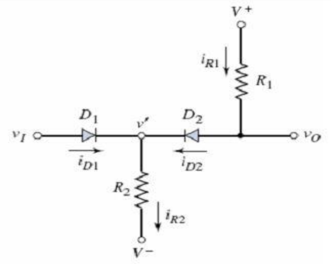


Figure 2.32 A two-diode circuit

If we assume that $V^+ > V^-$ and that $V^+ - V^- > V_\gamma$, there is a possibility that D_2 can be turned on. Figure 5.8 shows the resulting plot of v_o versus v_I .

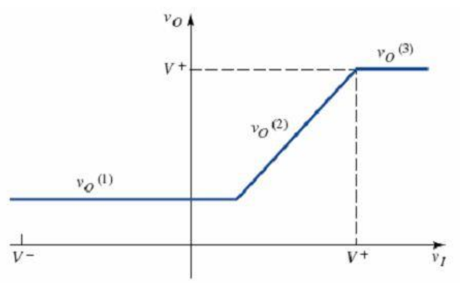


Figure 2.33 Voltage transfer characteristics for the two-diode circuit in Figure 2.32

Three distinct regions are shown, corresponding to the various conducting states of D_1 and D_2. The fourth possible state, corresponding to both diodes being off, is not feasible in this circuit

PHOTODIODE AND LED CIRCUITS

A photodiode converts an optical signal into an electrical current and a light-emitting diode (LED) transforms an electrical current into an optical signal.

Photodiode Circuit

Figure 5.9 shows a typical photodiode circuit in which a reverse-bias voltage is applied to the photodiode. If the photon intensity is zero, the only current through the diode is the reverse-saturation current, which is normally very small. Photons striking the diode create excess electrons and holes in the space-charge region, where the electric field quickly separates these excess carriers and sweeps them out of the space-charge region, creating a photocurrent in the reverse-bias direction.

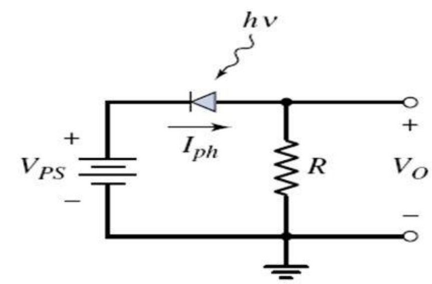


Figure 2.39 A photodiode circuit

LED Circuit

A light-emitting diode (LED) is the inverse of a photodiode; that is, a current is converted into an optical signal. If the diode is forward biased, electrons and holes are injected across the space-charge region, where they become excess minority carriers. These excess minority carriers diffuse into the neutral n and p regions, where they recombine with majority carriers, and the recombination can result in the emission of a photo (with a direct band gap material such as GaAs).

One application of LEDs and photodiodes is in opto-isolators, in which the input signal is electrically decoupled from the output (Figure 5.10). An input signal applied to the LED generates light, which is subsequently detected by the photodiode. The photodiode then converts the light back to an electrical signal. There is no electrical feedback or interaction between the output and input portions of the circuit.

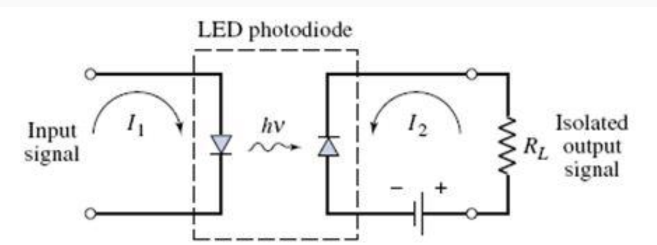


Figure 2.42 Optoisolator using an LED and a photodiode

MODULE 3.

Lecture 6 Active positive limiter, Active positive clamper,Active peak detector clipper and clamper circuit

Active positive limiter

Figure 6.1 shows an active positive limiter circuit, which clips off a part of the signal. The input and output signals are also shown. When the moving contact of the resistor V_{ref} is positive, the error voltage causes a negative output voltage and the diode is turned on. Since the feedback resistance is zero, a heavy negative feedback is produced. The output is zero for all positive values of v_i. When vi is negative, the output of op amp, is positive and the diode is truned off. Thus, the feedback loop is opened and the output v_a follows the negative half of the input v_i. Thus, the positive half is dipped off. To change the clipping level, we can adjust V_{ref}. Then the clipping occurs at V_{ref}.

Active positive clamper

Figure 6.2 shows a positive clamper circuit using op amp. This circuits adds a de component to the input signal. The input and output signals are also shown.

During the first negative half cycle of the input signal, we get a positive output and the diode is turned on. Due to the presence of the virtual ground, the capacitor C is charged to peak value V_m of the input signal with the polarity as shown. As the input signal turns positive, the diode is turned off (because the output is now negative). The virtual ground is lost. The output voltage is the sum of the signal and V_m.

Active peak detector

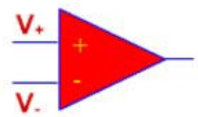
Figure 6.3 shows an op amp circuit for peak detection of low lwvel signals. The closed loop knee voltage is very small (in the mirco volt region). When the diode is on, the heavy voltage feedback producers an output impedance which is nearly zero. Thus, the charging time of the capacitor C is very small and the capacitor charges quickly to the positive peak value. When the diode is off, the capacitor tends to discharge through load resistor R_L. It time constant R_L C is much longer as compared to the period of the input signal, we can achieve peak detection of low level signals.

Comparator

A comparator is the simplest circuit that moves signals between the analog and digital worlds.

What does a comparator do?

Simply put, a comparator compares two analog signals and produces a one bit digital signal. The symbol for a comparator is shown below.



The comparator output satisfies the following rules:

- When V_+ is larger than V_- the output bit is 1.
- When V_+ is smaller than V_- the output bit is 0

When an op amp is used in open loop configuration, an input signal, as small as 1 m V (or even lower) van saturates the op amp. Therefore, the output will be between the maximum and minimum limits imposed by the dc supply. This action is unsuitable for linear circuits. However, in digital circuits only the highest and lowest voltages are the meaningful values as they can be used to represent the binary numbers 0 and 1.

Figure 6.4 (a) shows a simple comparator circuit. The inverting input is grounded and the input signal is fed to non-inverting input. If the dc supply is +/- 15 V, the output can swing only from –13.5 to +13.5V. The input voltage required to produce saturation (if open loop gain is 100000) is 13.5 /100000 or 0.135 mV. Let an input of 1V be applied to non-inverting terminal. The output will be limited to the maximum value of + 13.5 V. If the input signal is reduced, the output will remain at +13.5V. Only when the input is less than 0.135 mV will the output fall. For any input higher than 0.135 mV, the output will be 13.5 V. If the signal is negative, a similar action will occur. Any input signal more negative than –0.135 m V will produce an output o f-13.5V. Since 0.135mVis a very small value, the graph of output voltage will show a vertical transistor from –V to V as the input signal goes from negative to positive as shown in Fig. 6.4(b).

It is also possible to operate the circuit of Fig.6.4 (a) with non-inverting terminal grounded and the signal applied to inverting terminal. Now a positive input (more than 0.135 mV) will cause negative output saturation delivering an output of –13.5 V. Similarly, an input more negative than –0.135 mV will cause an output of +13.5V.

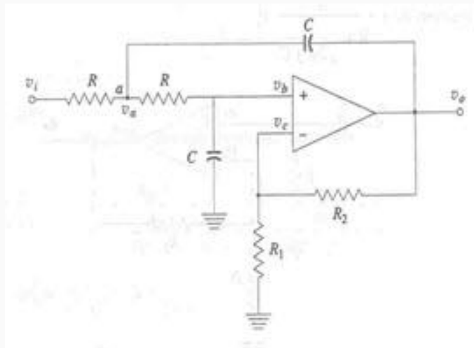
The Circuit of Fig 6.4 (a) does also know as zero crossing detectors.

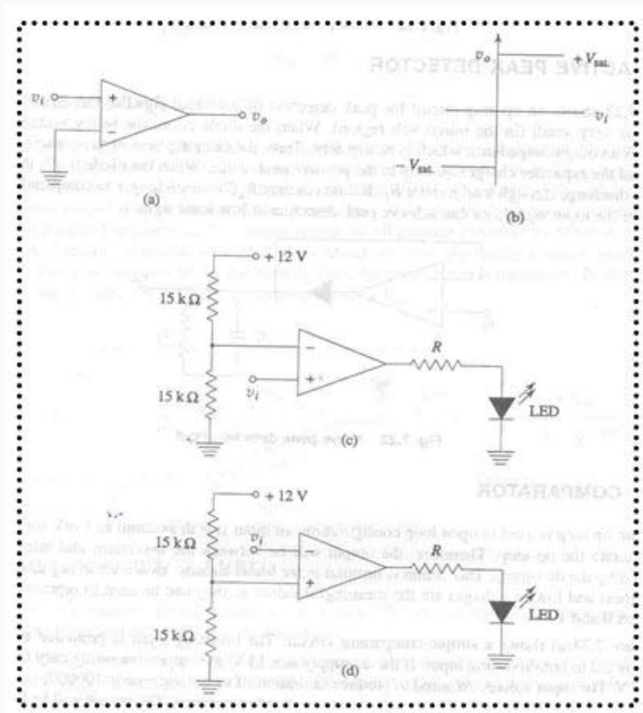
In Fig. 6.4(a), the reference level has been set at zero (by grounding the inverting terminal). In general the reference level need not be zero but can be any; desired positive or negative voltage. Moreover, the reference voltage can be connected to either the inverting terminal or non-inverting terminal and the input applied to the other terminal.

Figure 6.4(c) shows the reference voltage applied to minus input, (i.e., inverting input) and output connected to an LED. The reference voltage level is,

Since the reference level is connected to inverting, i.e., terminal, the output v_o will got to positive saturation level whenever the input v is more than +6V and LED will become on, thus, indicating that input is less positive than reference level.

Figure 6.4(d) shows the reference voltage connected to non-inverting (or +) terminal. Any input v less than reference level of +6V will cause the output to switch on the LED, indicating that input is less than reference level.The comparator has many applications in signal processing and wave shaping circuits. In each such application, the comparator detects whether the applied signal is more or less than the reference level.





Clampers:

A clamper is an electronic circuit that prevents a signal from exceeding a certain defined magnitude by shifting its DC value. The clamper does not restrict the peak-to-peak excursion of the signal, but moves it up or down by a fixed value. A diode clamp (a simple, common type) relies on a diode, which conducts electric current in only one direction; resistors and capacitors in the circuit are used to maintain an altered dc level at the clamper output.

General function:

A clamping circuit (also known as a clamper) will bind the upper or lower extreme of a waveform to a fixed DC voltage level. These circuits are also known as DC voltage restorers. Clampers can be constructed in both positive and negative polarities. When unbiased, clamping circuits will fix the voltage lower limit (or upper limit, in the case of negative clampers) to 0 Volts. These circuits clamp a peak of a waveform to a specific DC level compared with a capacitively coupled signal which swings about its average DC level

Types:

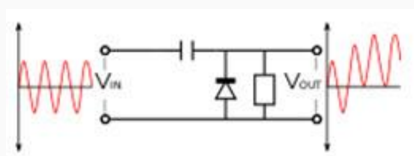
Clamp circuits are categorized by their operation; negative or positive and biased and unbiased. A positive clamp circuit outputs a purely positive waveform from an input signal; it offsets the input signal so that all of the waveform is greater than 0 V. A negative clamp is the opposite of this - this clamp outputs a purely negative waveform from an input signal.A bias voltage between the diode and ground offsets the output voltage by that amount.For example, an input signal of peak value 5 V (V_{IN} = 5 V) is applied to a positive clamp with a bias of 3 V (V_{BIAS} = 3 V), the peak output voltage will be

$V_{OUT} = 2V_{IN} + V_{BIAS}$

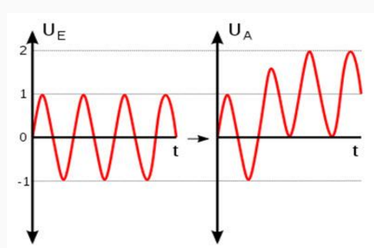
$V_{OUT} = 2 * 5\ V + 3\ V$

$V_{OUT} = 13\ V$

Positive unbiased:

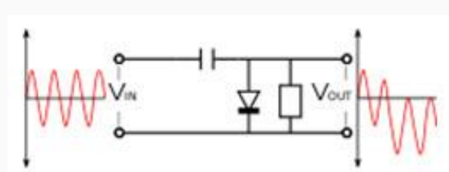


In the negative cycle of the input AC signal, the diode is forward biased and conducts, charging the capacitor to the peak positive value of V_{IN}. During the positive cycle, the diode is reverse biased and thus does not conduct. The output voltage is therefore equal to the voltage stored in the capacitor plus the input voltage gain, so $V_{OUT} = 2V_{IN}$



Positive unbiased voltage clamping shifts the amplitude of the input waveform so that all parts of it are greater than 0 V

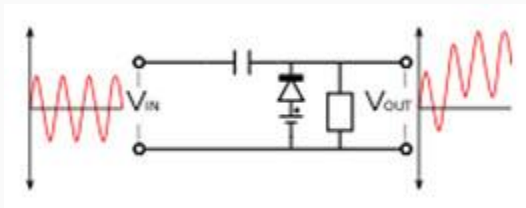
Negative unbiased:



A negative unbiased clamp is the opposite of the equivalent positive clamp. In the positive cycle of the input AC signal, the diode is forward biased and conducts, charging the capacitor to the peak value of V_{IN}. During the negative cycle, the diode is reverse biased and thus does not conduct. The output voltage is therefore equal to the voltage stored in the capacitor plus the input voltage again, so

$$V_{OUT} = -2V_{IN}$$

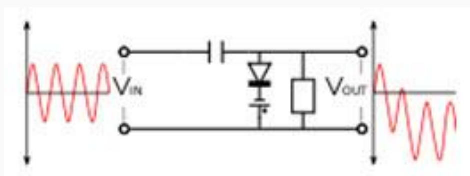
Positive biased:



A positive biased voltage clamp is identical to an equivalent unbiased clamp but with the output voltage offset by the bias amount V_{BIAS}. Thus,

$$V_{OUT} = 2V_{IN} + V_{BIAS}$$

Negative biased:



A negative biased voltage clamp is likewise identical to an equivalent unbiased clamp but with the output voltage offset in the negative direction by the bias amount V_{BIAS}. Thus,

$$V_{OUT} = -2V_{IN} - V_{BIAS}$$

MODULE 4.

LESSON 7. Filter- classification-components-Inductor, capacitor. Filter-LC, π filter-Active low pass and high pass filter.

Filters:

A filter is a circuit or device that allows some frequencies to pass through and attenuates other frequencies.

Based on the construction, filters can be classified as (i)Passive filters and (ii) Active filters.

1. Passive filters – Filters that use only passive elements such as resistors, capacitors and inductors are called Passive filters.
2. Active filters – Filters that use active devices such as transistors or op-amps are called Active filters.

The performance of active filters is better than passive filters. Further, low cost integrated op-amps have helped to make active filters popular. They have an advantage at lower frequencies.

Resistors:

Physical materials resist the flow of electrical current to some extent. Certain materials such as copper offers very low resistance to current flow, and hence they are called as conductors. Other materials such as ceramic which offer extremely high resistance to current flow are called as insulators. In electric and electronic circuits, there is a need for materials with specific values of resistance in the range between that of a conductor and insulator. These materials are called as resistors and their value of resistance are expressed in ohms.

Types of resistors:

1. Fixed resistors – Carbon composition, carbon film, metal film, wire wound etc
2. Variable resistors – Potentiometer, Rheostat, Trimmer

Capacitors:

Capacitors are the devices which can store electric charge. The energy stored in the capacitor is given by,

$$W = \frac{CV^2}{2}$$

Where, C – Capacitance, Farad

V – Voltage across the capacitor

A capacitor consists of two conductive plates separated by a dielectric material.

Inductors:

Inductors store energy in the form of magnetic field and deliver it as and when required. Whenever current passes through a conductor, lines of magnetic flux are generated around it. This magnetic flux opposes any change in current due to the induced emf. This opposition to the change in current is known as inductance and the component producing inductance is known as inductor. The unit of inductance is Henry (H). The inducexd emf is given by,

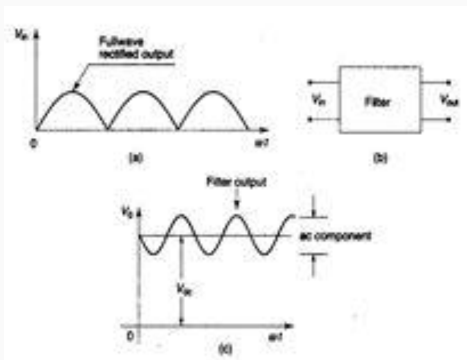
$$e = -L\frac{di}{dt}$$

where, e – induced emf in volts

L – Inductance in Henry

di/dt – rate of change of current

Filters:



Concept of a filter

The output of a rectifier contains dc component as well as ac component. Filters are used to minimize the undesirable ac ie., ripple leaving only the dc component to appear at the output ie., Filters are used to remove the ripple from the d.c output. The commonly used filters are listed below:

1. Inductor filter
2. Capacitor filter
3. LC filter
4. π - filter

Inductor filter

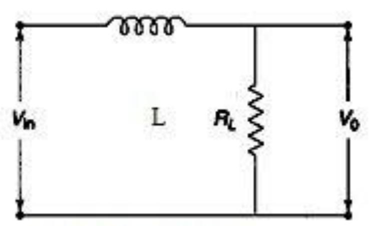


Fig.7.1. Inductor filter

In consists of a choke in series with the load as shown in the Figure 7.1. The Inductor has the inherent characteristic of opposing any change in the current. Hence the introduction of choke in the rectifier circuit will have a smoothing effect. The inductive filter is suitable for heavy loads.

Capacitor filter

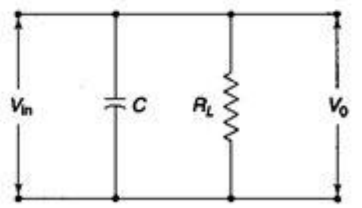


Fig.7.2. Capacitor filter

In consists of a choke in series with the load as shown in the Figure 7.1. The Inductor has the inherent characteristic of opposing any change in the current. Hence the introduction of choke in the rectifier circuit will have a smoothing effect. The inductive filter is suitable for heavy loads.

It consists of a capacitor directly across the load as shown in the Figure7.2. At light loads, the capacitor filter maintains the output voltage near to maximum voltage (V_m). The capacitor charges up to the maximum value of input voltage and maintains the value even as the fullwave voltage drops to zero.

The discharge of capacitor through load resistance takes place till the input voltage raises to a value more than the capacitor voltage. Thus the diode will again be forward biased causing recharging of capacitor due to diode current. However, as the load increases, the ripple also increases due to the greater discharge of the capacitor.

LC filter

We have seen that, in the case of L-filter the ripple increases with decrease in the load, where as the ripple increases with increases in load in the C-filter. Hence the combination of these two filters namely, LC filter has to make the ripple independent of the load. The LC filter circuit is shown in Figure 7.3.

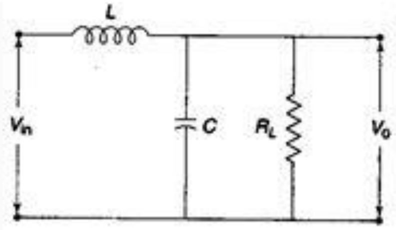


Fig.7.3. LC Filter

CLC or p-Filter

The use of CLC or p-filter improves the filtering process. The circuit arrangement is as shown in the Figure 7.4. In some cases, for light loads, the inductor may be replaced by a resistor.

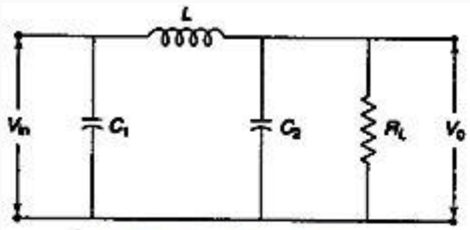


Fig.7.4. CLC or π filter

Active low pass filter

A filter circuit allows only a certain range of frequencies to pass freely (without attenuation) while suppressing other frequencies. Filter circuits are classified as low pass, (i.e., low frequencies from 0 to a certain cut off frequency fc allowed to pass), high pass, (ie., frequencies above a cut off frequency fc are allowed to pass), band pass, (i.e frequencies within a specified band are allowed to pass) and band stop, (i.e. frequencies within a specified band are suppressed). Filter circuits can be fabricated from passive components, i.e., resistors, capacitors and inductors. An active filter uses an operational amplifiers are very suitable due to good characteristics of operational amplifiers.

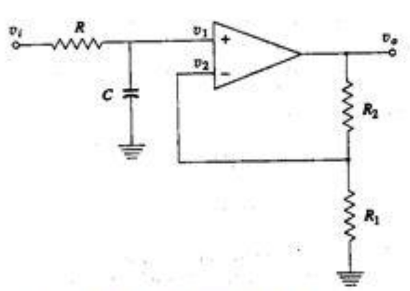


Fig.7.5(a).One pole active low pass

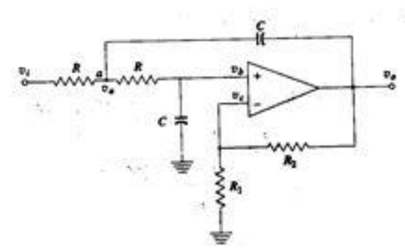


Fig.7.5(b).Two pole active low pass

Figure 7.5(a) shows a one pole active low pass filter and figure 7.5(b) shows a two pole active low pass filter. The closed loop gain A_{vf} from inverting input to the output is,

$$A_{vf} = \frac{v_0}{v_2} = 1 + \frac{R_2}{R_1}$$

The cut off frequency f_c of the by-pass circuit consisting of R and C is

$$f_c = \frac{1}{2\pi RC}$$

When A_{vf} = 1.586 the cut off frequency f_c is given by the equation. At cut off frequency the overall gain is down 3 dB. Above f_c the voltage gain decreases 40 dB per decade increase in frequency.

Active high pass filter

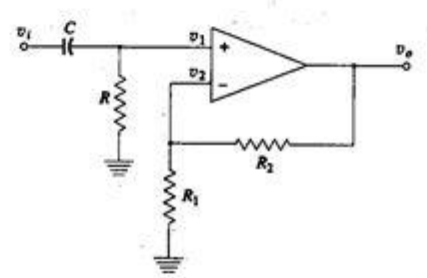


Fig.7.6(a) One pole active high pass filter

Figure 7.6(a) shows a one pole high pass filter and figure 6.6(b) shows a two pole high pass filter. It uses acoupling circuit consisting of R and C. It passes high frequencies but blocks the low frequencies. The closed loop gain A_{vf} from inverting input to output is

$$A_{vf} = \frac{v_0}{v_2} = 1 + \frac{R_2}{R_1}$$

The cut off frequency f_c of the by-pass circuit consisting of R and C is

$$f_c = \frac{1}{2\pi RC}$$

MODULE 5.

LESSON 8. Gates, logic circuits, Boolean expressions. Bipolar transistor basics-construction and configurations

NPN transistor configurates- α and β relationship- Input and output characteristics-PNP transistor-configurates and circuits

Gate

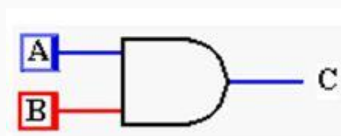
Gate is a device, which contains number of inputs and one or more outputs, the output occurs only for a well-defined condition of the input.

Truth table

It is the tabulation, which represents all possible input and output conditions in logic levels.

AND gate

The output of an AND gate will stand in its defined one level if and only if all the inputs stand at their defined one level.

Symbol 

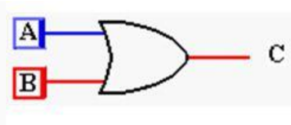
Boolean Expression: Y = A,B

Truth table

Input		output
A	B	Y= A,B
0	0	0
0	1	0
1	0	0
1	1	1

OR gate (inclusive OR gate)

The output of an OR gate will stand in its defined one level if any one or more or all the inputs stand at their defined one level.



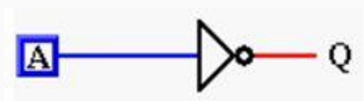
Boolean expression: Y=A+B

Truth table

Input		output
A	B	Y= A,B
0	0	0
0	1	1
1	0	1
1	1	1

NOT gate

It has a single input and a single output. It inverts the input and so it is also called an inverter.The output of the NOT gate is zero if the input is 1 and the output is 1 if the input is 0.



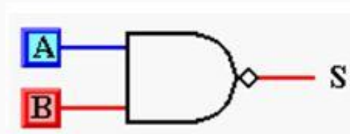
Truth Table

Input	Output
0	1
1	0

NAND gate (Universal gate)

The output of an NAND gate will stand in the defined zero level if and only if all inputs are at one level. (It is nothing but the combination of AND gate and NOT gate)

Symbol



Boolean expression: Y = A, B

Truth table

Input		Output
A	B	C=AB
0	0	1
0	1	1
1	0	1
1	1	0

It is the combination of OR gate and NOT gate.

The output of a NOR gate will stand in its defined one level if and only if all the inputs are at zero level.

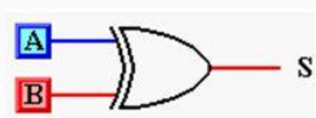
Boolean expression: Y = A+B

Truth table

Input		Output
A	B	C=A+B
0	0	1
1	0	0
0	1	0
1	1	0

Exclusive OR gate (X-OR gate)

The output of an X-OR gate will stand in its defined one level if and only if all the inputs are at different levels. **Exclusive OR gate (X-OR gate)**

Symbol 

Truth table

A	B	C=A+B
0	0	0
0	1	1
1	0	1
1	1	0

Boolean expression: Y=A+B (or) AB + AB

We know that the output of an X-OR gate will be at one level only if the inputs are at different levels. The Boolean expression for X-OR gate is Y=A, B+A. This statement can be proved using the truth table

A	B	A	B	A.B	A.B	AB + AB
0	0	1	1	0	0	0
0	1	1	0	0	1	1
1	0	0	1	1	0	1
1	1	0	0	0	0	0

De Morgan's theorems

First law

The complements of the sum of the variables are equal to the product of their complements. These statements can be proved using the truth table given below.

Figure shows the logic circuit to represent the 2nd law for two variables A and B. The truth table for the left and right side of the statement are represented in tables. Since the net result of tables are same, this statements is proved for different inputs

Truth table

Left hand side = Right hand side

Second law

The complement of the product of variables is equal to the sum of their complements

A, B = A + B

Figure shows the logic circuit to represent the II law for two variables A and B. The truth tables for each circuit proves that A.B = A+B.

Truth Table

A	B	A.B	A.B
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

A	B	A	B	A+B
0	0	1	1	1
0	1	1	0	1
1	0	0	-	1
1	1	0	0	0

Transistors are three terminal active devices made from different semiconductor materials that can act as either an insulator or a conductor by the application of a small signal voltage. The transistor's ability to change between these two states enables it to have two basic functions: "switching" (digital electronics) or "amplification" (analogue electronics). Then bipolar transistors have the ability to operate within three different regions:

Bipolar Transistor Basics

Diodes made up from two pieces of semiconductor material, either silicon or germanium to form a simple PN-junction and we also learnt about their properties and characteristics. If we now join together two individual signal diodes back-to-back, this will give us two PN-junctions connected together in series that share a common P or N terminal. The fusion of these two diodes produces a three layer, two junction, three terminal device forming the basis of a **Bipolar Junction Transistor**, or **BJT** for short.

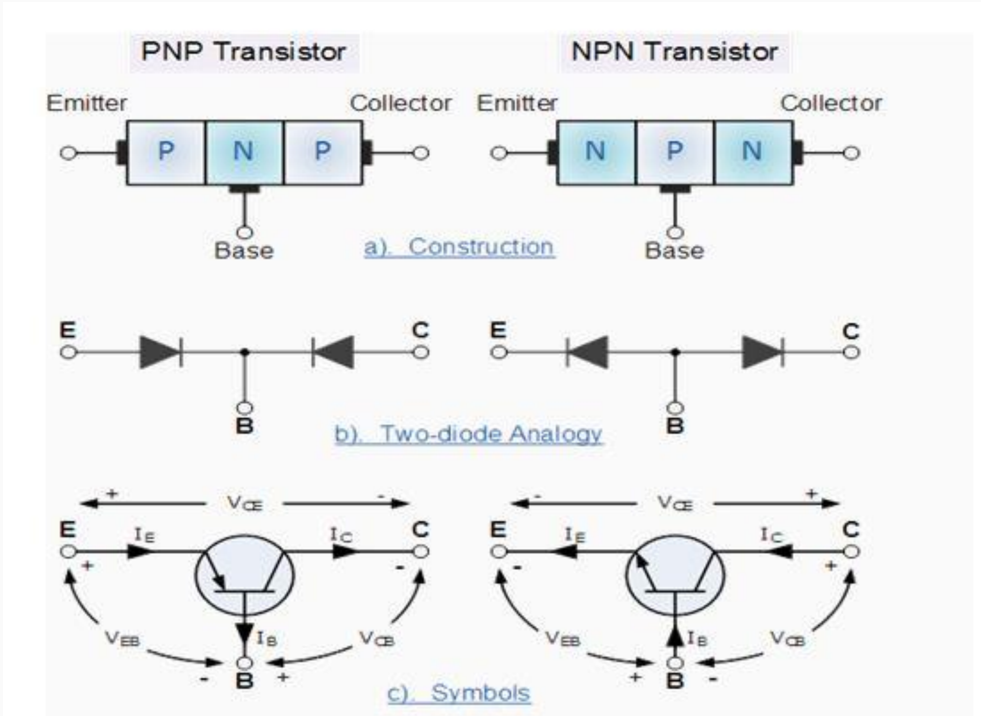
Transistors are three terminal active devices made from different semiconductor materials that can act as either an insulator or a conductor by the application of a small signal voltage. The transistor's ability to change between these two states enables it to have two basic functions: "switching" (digital electronics) or "amplification" (analogue electronics). Then bipolar transistors have the ability to operate within three different regions:

1. Active Region - the transistor operates as an amplifier and $I_c = \beta.I_b$
2. Saturation - the transistor is "fully-ON" operating as a switch and I_c = I(saturation)
3. Cut-off - the transistor is "fully-OFF" operating as a switch and $I_c = 0$

Typical Bipolar Transistor

The word Transistor is an acronym, and is a combination of the words Transfer Varistor used to describe their mode of operation way back in their early days of development. There are two basic types of bipolar transistor construction, PNP and NPN, which basically describes the physical arrangement of the P-type and N-type semiconductor materials from which they are made. The **Bipolar Transistor**basic construction consists of two PN-junctions producing three connecting terminals with each terminal being given a name to identify it from the other two. These three terminals are known and labelled as the Emitter (E), the Base (B) and the Collector(C) respectively. Bipolar Transistors are current regulating devices that control the amount of current flowing through them in proportion to the amount of biasing voltage applied to their base terminal acting like a current-controlled switch. The principle of operation of the two transistor types PNP and NPN, is exactly the same the only difference being in their biasing and the polarity of the power supply for each type.

Bipolar Transistor Construction



The construction and circuit symbols for both the PNP and NPN bipolar transistor are given above with the arrow in the circuit symbol always showing the direction of "conventional current flow" between the base terminal and its emitter terminal. The direction of the arrow always points from the positive P-type region to the negative N-type region for both transistor types, exactly the same as for the standard diode symbol.

Bipolar Transistor Configurations

As the **Bipolar Transistor** is a three terminal device, there are basically three possible ways to connect it within an electronic circuit with one terminal being common to both the input and output. Each method of connection responding differently to its input signal within a circuit as the static characteristics of the transistor vary with each circuit arrangement.

1. Common Base Configuration - has Voltage Gain but no Current Gain.

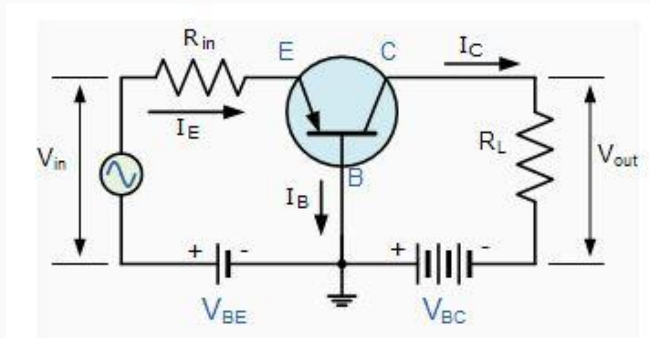
2. Common Emitter Configuration - has both Current and Voltage Gain.

3. Common Collector Configuration - has Current Gain but no Voltage Gain.

The Common Base (CB) Configuration

As its name suggests, in the **Common Base** or grounded base configuration, the BASE connection is common to both the input signal AND the output signal with the input signal being applied between the base and the emitter terminals. The corresponding output signal is taken from between the base and the collector terminals as shown with the base terminal grounded or connected to a fixed reference voltage point. The input current flowing into the emitter is quite large as its the sum of both the base current and collector current respectively therefore, the collector current output is less than the emitter current input resulting in a current gain for this type of circuit of "1" (unity) or less, in other words the common base configuration "attenuates" the input signal.

The Common Base Transistor Circuit



This type of amplifier configuration is a non-inverting voltage amplifier circuit, in that the signal voltages Vin and Vout are in-phase. This type of transistor arrangement is not very common due to its unusually high voltage gain characteristics. Its output characteristics represent that of a forward biased diode while the input characteristics represent that of an illuminated photo-diode. Also this type of bipolar transistor configuration has a high ratio of output to input resistance or more importantly "load" resistance (RL) to "input" resistance (Rin) giving it a value of "Resistance Gain". Then the voltage gain (Av) for a common base configuration is therefore given as:

Common Base Voltage Gain

$$A_V = \frac{Vout}{Vin} = \frac{I_C \times R_L}{I_E \times R_{IN}}$$

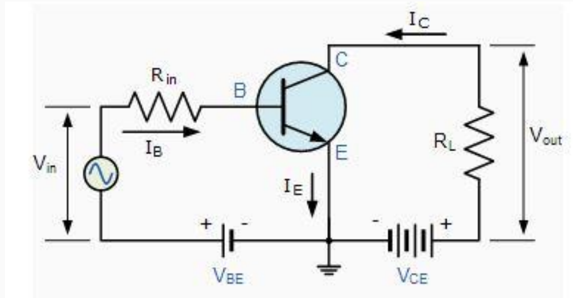
Where: Ic/Ie is the current gain, alpha (α) and RL/Rin is the resistance gain.

The common base circuit is generally only used in single stage amplifier circuits such as microphone pre-amplifier or radio frequency (Rf) amplifiers due to its very good high frequency response.

The Common Emitter (CE) Configuration

In the **Common Emitter** or grounded emitter configuration, the input signal is applied between the base, while the output is taken from between the collector and the emitter as shown. This type of configuration is the most commonly used circuit for transistor based amplifiers and which represents the "normal" method of bipolar transistor connection. The common emitter amplifier configuration produces the highest current and power gain of all the three bipolar transistor configurations. This is mainly because the input impedance is LOW as it is connected to a forward-biased PN-junction, while the output impedance is HIGH as it is taken from a reverse-biased PN-junction.

The Common Emitter Amplifier Circuit



In this type of configuration, the current flowing out of the transistor must be equal to the currents flowing into the transistor as the emitter current is given as Ie = Ic + Ib. Also, as the load resistance (RL) is connected in series with the collector, the current gain of the common emitter transistor configuration is quite large as it is the ratio of Ic/Ib and is given the Greek symbol of Beta, (β). As the emitter current for a common emitter configuration is defined as Ie = Ic + Ib, the ratio of Ic/Ie is called Alpha, given the Greek symbol of α. Note: that the value of Alpha will always be less than unity.

Since the electrical relationship between these three currents, Ib, Ic and Ie is determined by the physical construction of the transistor itself, any small change in the base current (Ib), will result in a much larger change in the collector current (Ic). Then, small changes in current flowing in the base will thus control the current in the emitter-collector circuit. Typically, Beta has a value between 20 and 200 for most general purpose transistors.

By combining the expressions for both Alpha, α and Beta, β the mathematical relationship between these parameters and therefore the current gain of the transistor can be given as:

$$\text{Alpha},(\alpha) = \frac{I_C}{I_E} \quad \text{and} \quad \text{Beta},(\beta) = \frac{I_C}{I_B}$$

$$\therefore I_C = \alpha . I_E = \beta . I_B$$

$$\text{as:} \quad \alpha = \frac{\beta}{\beta + 1} \qquad \beta = \frac{\alpha}{1 - \alpha}$$

$$I_E = I_C + I_B$$

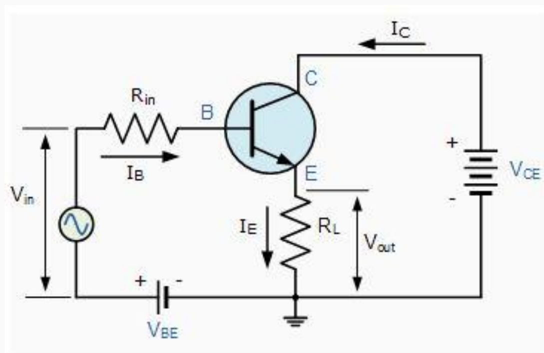
Where: "Ic" is the current flowing into the collector terminal, "Ib" is the current flowing into the base terminal and "Ie" is the current flowing out of the emitter terminal.

Then to summarise, this type of bipolar transistor configuration has a greater input impedance, current and power gain than that of the common base configuration but its voltage gain is much lower. The common emitter configuration is an inverting amplifier circuit resulting in the output signal being 180^{o} out-of-phase with the input voltage signal.

The Common Collector (CC) Configuration

In the **Common Collector** or grounded collector configuration, the collector is now common through the supply. The input signal is connected directly to the base, while the output is taken from the emitter load as shown. This type of configuration is commonly known as a **Voltage Follower** or **Emitter Follower** circuit. The emitter follower configuration is very useful for impedance matching applications because of the very high input impedance, in the region of hundreds of thousands of Ohms while having a relatively low output impedance.

The Common Collector Transistor Circuit



The common emitter configuration has a current gain approximately equal to the β value of the transistor itself. In the common collector configuration the load resistance is situated in series with the emitter so its current is equal to that of the emitter current. As the emitter current is the combination of the collector AND the base current combined, the load resistance in this type of transistor configuration also has both the collector current and the input current of the base flowing through it. Then the current gain of the circuit is given as:

The Common Collector Current Gain

$$I_E = I_C + I_B$$

$$A_i = \frac{I_E}{I_B} = \frac{I_C + I_B}{I_B}$$

$$A_i = \frac{I_C}{I_B} + 1$$

$$A_i = \beta + 1$$

$$A_i = \beta + 1$$

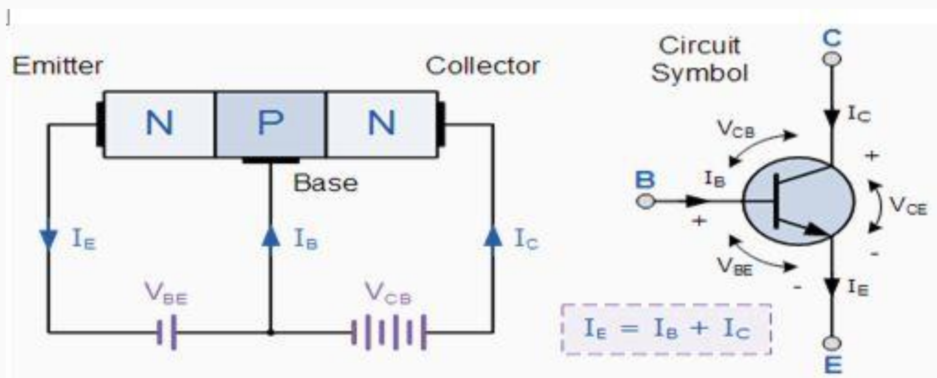
This type of bipolar transistor configuration is a non-inverting circuit in that the signal voltages of Vin and Vout are in-phase. It has a voltage gain that is always less than "1" (unity). The load resistance of the common collector transistor receives both the base and collector currents giving a large current gain (as with the common emitter configuration) therefore, providing good current amplification with very little voltage gain.

LESSON 9. NPN transistor configurations- α and β relationship- Input and output characteristics-PNP transistor-configurations and circuits

The NPN Transistor

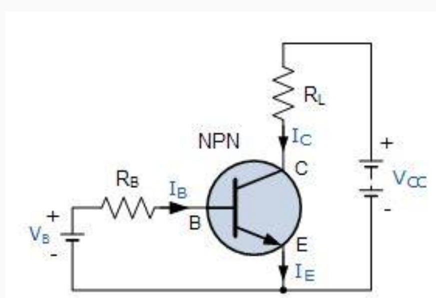
Bipolar Transistor or BJT, comes in two basic forms. An NPN (Negative-Positive-Negative) type and a PNP (Positive-Negative-Positive) type, with the most commonly used transistor type being the **NPN Transistor**. We also learnt that the transistor junctions can be biased in one of three different ways - **Common Base**, **Common Emitter** and **Common Collector**. In this tutorial we will look more closely at the "Common Emitter" configuration using **NPN Transistors** with an example of the construction of a NPN transistor along with the transistors current flow characteristics is given below.

An NPN Transistor Configuration



(Note: Arrow defines the emitter and conventional current flow, "out" for an NPN transistor.)

The construction and terminal voltages for an NPN transistor are shown above. The voltage between the Base and Emitter (V_{BE}), is positive at the Base and negative at the Emitter because for an NPN transistor, the Base terminal is always positive with respect to the Emitter. Also the Collector supply voltage is positive with respect to the Emitter (V_{CE}



NPN Transistor Connections

). So for an NPN transistor to conduct the Collector is always more positive with respect to both the Base and the Emitter. Then the voltage sources are connected to an NPN transistor as shown. The Collector is connected to the supply voltage V_{CC} via the load resistor, RL which also acts to limit the maximum current flowing through the device. The Base supply voltage V_B is connected to the Base resistor R_B, which again is used to limit the maximum Base current.

We know that the transistor is a **"current"** operated device (Beta model) and that a large current (Ic) flows freely through the device between the collector and the emitter terminals when the transistor is switched "fully-ON". However, this only happens when a small biasing current (Ib) is flowing into the base terminal of the transistor at the same time thus allowing the Base to act as a sort of current control input.

The transistor current in an NPN transistor is the ratio of these two currents (Ic/Ib), called the *DC* Current Gain of the device and is given the symbol of hfe or nowadays Beta, (β). The value of β can be large up to 200 for standard transistors, and it is this large ratio between Ic and Ib that makes the NPN transistor a useful amplifying device when used in its active region as Ib provides the input and Ic provides the output. Note that Beta has no units as it is a ratio.

Also, the current gain of the transistor from the Collector terminal to the Emitter terminal, Ic/Ie, is called Alpha, (α), and is a function of the transistor itself (electrons diffusing across the junction). As the emitter current Ie is the product of a very small base current plus a very large collector current, the value of alpha α, is very close to unity, and for a typical low-power signal transistor this value ranges from about 0.950 to 0.999

α and β Relationship in a NPN Transistor

$$\text{DC Current Gain} = \frac{\text{Output Current}}{\text{Input Current}} = \frac{I_C}{I_B}$$

$$I_E = I_B + I_C \; ..... \text{(KCL)} \quad \text{and} \quad \frac{I_C}{I_E} = \alpha$$

$$\text{Thus:} \quad I_E = I_B + I_C = \frac{I_C}{\alpha}$$

$$\text{and} \quad I_B = I_C\left(1 - \frac{1}{\alpha}\right)$$

$$\therefore \quad \beta = \frac{I_C}{I_B} = \frac{1}{\left(1 - \frac{1}{\alpha}\right)} = \frac{\alpha}{1-\alpha}$$

By combining the two parameters α and β we can produce two mathematical expressions that gives the relationship between the different currents flowing in the transistor.

$$\beta = \frac{\alpha}{1-\alpha} \quad \text{and} \quad \alpha = \frac{\beta}{\beta+1}$$

$$\text{If} \quad \alpha = 0.99 \quad \beta = \frac{0.99}{0.01} = 99$$

The values of Beta vary from about 20 for high current power transistors to well over 1000 for high frequency low power type bipolar transistors. The value of Beta for most standard NPN transistors can be found in the manufactures datasheets but generally range between 50 - 200.

The equation above for Beta can also be re-arranged to make Ic as the subject, and with a zero base current (Ib = 0) the resultant collector current Ic will also be zero, (β x 0). Also when the base current is high the corresponding collector current will also be high resulting in the base current controlling the collector current. One of the most important properties of the **Bipolar Junction Transistor** is that a small base current can control a much larger collector current. Consider the following example.

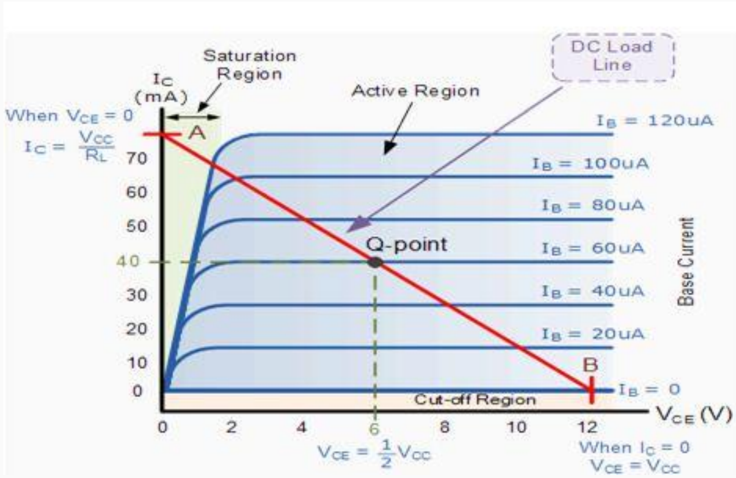
The Common Emitter Configuration.

As well as being used as a semiconductor switch to turn load currents "ON" or "OFF" by controlling the Base signal to the transistor in ether its saturation or cut-off regions, **NPN Transistors** can also be used in its active region to produce a circuit which will amplify any small AC signal applied to its Base terminal with the Emitter grounded. If a suitable DC "biasing" voltage is firstly applied to the transistors Base terminal thus allowing it to always operate within its linear active region, an inverting amplifier circuit called a single stage common emitter amplifier is produced.

One such Common Emitter Amplifier configuration of an NPN transistor is called a **Class A Amplifier**. A "Class A Amplifier" operation is one where the transistors Base terminal is biased in such a way as to forward bias the Base-emitter junction. The result is that the transistor is always operating halfway between its cut-off and saturation regions, thereby allowing the transistor amplifier to accurately reproduce the positive and negative halves of any AC input signal superimposed upon this DC biasing voltage. Without this "Bias Voltage" only one half of the input waveform would be amplified. This common emitter amplifier configuration using an NPN transistor has many applications but is commonly used in audio circuits such as pre-amplifier and power amplifier stages.

With reference to the common emitter configuration shown below, a family of curves known as the **Output Characteristics Curves**, relates the output collector current, (Ic) to the collector voltage, (Vce) when different values of Base current, (Ib) are applied to the transistor for transistors with the same β value. A DC "Load Line" can also be drawn onto the output characteristics curves to show all the possible operating points when different values of base current are applied. It is necessary to set the initial value of Vce correctly to allow the output voltage to vary both up and down when amplifying AC input signals and this is called setting the operating point or Quiescent Point, **Q-point** for short and this is shown below.

Output Characteristics Curves of a Typical Bipolar Transistor



The most important factor to notice is the effect of Vce upon the collector current Ic when Vce is greater than about 1.0 volts. We can see that Ic is largely unaffected by changes in Vce above this value and instead it is almost entirely controlled by the base current, Ib. When this happens we can say then that the output circuit represents that of a "Constant Current Source". It can also be seen from the common emitter circuit above that the emitter current Ie is the sum of the collector current, Ic and the base current, Ib, added together so we can also say that Ie = Ic + Ib for the common emitter (CE) configuration.

By using the output characteristics curves in our example above and also Ohm´s Law, the current flowing through the load resistor, (RL), is equal to the collector current, Ic entering the transistor which inturn corresponds to the supply voltage, (Vcc) minus the voltage drop between the collector and the emitter terminals, (Vce) and is given as:

$$\text{Collector Current,}\quad I_C = \frac{V_{CC} - V_{CE}}{R_L}$$

Also, a straight line representing the **Dynamic Load Line** of the transistor can be drawn directly onto the graph of curves above from the point of "Saturation" (A) when Vce = 0 to the point of "Cut-off" (B) when Ic = 0 thus giving us the "Operating" or **Q-point** of the transistor. These two points are joined together by a straight line and any position along this straight line represents the "Active Region" of the transistor. The actual position of the load line on the characteristics curves can be calculated as follows:

$$\text{When:}\ (V_{CE} = 0)\quad I_C = \frac{V_{CC} - 0}{R_L},\quad I_C = \frac{V_{CC}}{R_L}$$

$$\text{When:}\ (I_C = 0)\quad 0 = \frac{V_{CC} - V_{CE}}{R_L},\quad V_{CC} = V_{CE}$$

Then, the collector or output characteristics curves for **Common Emitter NPN Transistors** can be used to predict the Collector current, Ic, when given Vce and the Base current, Ib. A Load Line can also be constructed onto the curves to determine a suitable Operating or **Q-point** which can be set by adjustment of the base current. The slope of this load line is equal to the reciprocal of the load resistance which is given as: $-1/R_L$

Then we can define a **NPN Transistor** as being normally "OFF" but a small input current and a small positive voltage at its Base (B) relative to its Emitter (E) will turn it "ON" allowing a much large Collector-Emitter current to flow. NPN transistors conduct when Vc is much greater than Ve.

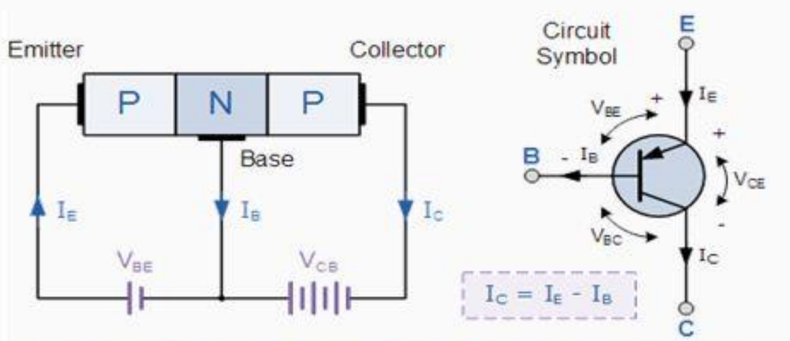
The PNP Transistor

The **PNP Transistor** is the exact opposite to the **NPN Transistor** device we looked at in the previous tutorial. Basically, in this type of transistor construction the two diodes are reversed with respect to the NPN type giving a Positive-Negative-Positive configuration, with the arrow which also defines the Emitter terminal this time pointing inwards in the transistor symbol.

Also, all the polarities for a PNP transistor are reversed which means that it "sinks" current into its Base as opposed to the NPN transistor which "sources" current through its Base. The main difference between the two types of transistors is that holes are the more important carriers for PNP transistors, whereas electrons are the important carriers for NPN transistors. Then, PNP transistors use a small base current and a negative base voltage to control a much larger emitter-collector current. In other words for a PNP transistor, the Emitter is more positive with respect to the Base and also with respect to the Collector.

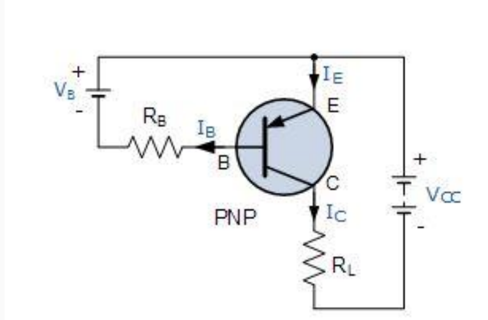
The construction of a "PNP transistor" consists of two P-type semiconductor materials either side of an N-type material as shown below.

A PNP Transistor Configuration



(Note: Arrow defines the emitter and conventional current flow, "in" for a PNP transistor.)

The construction and terminal voltages for an NPN transistor are shown above. The **PNP Transistor** has very similar characteristics to their NPN bipolar cousins, except that the polarities (or biasing) of the current and voltage directions are reversed for any one of the possible three configurations looked at in the first tutorial, Common Base, Common Emitter and Common Collector.



PNP Transistor Connections

The voltage between the Base and Emitter (V_{BE}), is now negative at the Base and positive at the Emitter because for a PNP transistor, the Base terminal is always biased negative with respect to the Emitter. Also the Emitter supply voltage is positive with respect to the Collector (V_{CE}). So for a PNP transistor to conduct the Emitter is always more positive with respect to both the Base and the Collector.

The voltage sources are connected to a PNP transistor are as shown. This time the Emitter is connected to the supply voltage V_{CC} with the load resistor, RL which limits the maximum current flowing through the device connected to the Collector terminal. The Base voltage V_B which is biased negative with respect to the Emitter and is connected to the Base resistor R_B, which again is used to limit the maximum Base current.

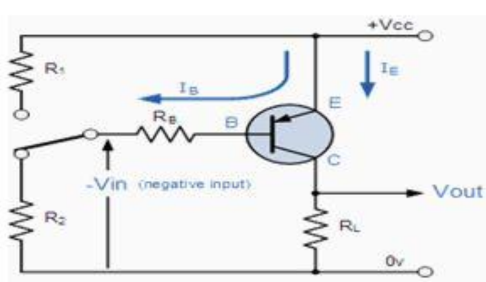
To cause the Base current to flow in a PNP transistor the Base needs to be more negative than the Emitter (current must leave the base) by approx 0.7 volts for a silicon device or 0.3 volts for a germanium device with the formulas used to calculate the Base resistor, Base current or Collector current are the same as those used for an equivalent NPN transistor and is given as.

$$I_C = I_E - I_B$$

$$I_C = \beta . I_B \qquad I_B = \frac{I_C}{\beta}$$

Generally, the PNP transistor can replace NPN transistors in most electronic circuits, the only difference is the polarities of the voltages, and the directions of the current flow. PNP transistors can also be used as switching devices and an example of a PNP transistor switch is shown below.

A PNP Transistor Circuit



The **Output Characteristics Curves** for a PNP transistor look very similar to those for an equivalent NPN transistor except that they are rotated by 180^{o} to take account of the reverse polarity voltages and currents, (the currents flowing out of the Base and Collector in a PNP transistor are negative). The same dynamic load line can be drawn onto the I-V curves to find the PNP transistors operating points..

MODULE 6.

LESSON 10. Amplifiers-general characteristics-classification-Common Emitter amplifier-characteristics-Common Collector amplifier-characteristics.

AMPLIFIER:

- The important function of an amplifier is the amplification.
- Circuit that increases the amplitude of the weak signals.
- The important parameters of an amplifier are input impedance, output impedance, current gain and voltage gain.

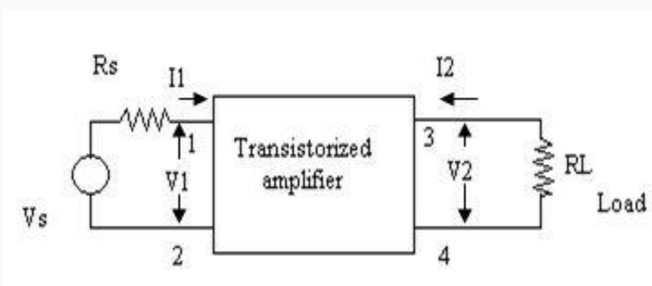
A good design of an amplifier circuit must possess high input impedance, low output impedance and high current gain.

- Constitute an essential part of radio, television and communication devices.
- In discrete circuits, bipolar JFET, FET is used as amplifying element.

General Characteristics of Amplifier

The amplifier is a two port network having two input terminals and two output terminals, as shown in Fig. 10.1.

Figure 10.1.Two port network of transistorized amplifier



The signal to be amplified is applied to the input terminals and the load is connected to the output terminals. The amplified signal output is available across the load RL. One of the input and output terminals are made generally common, i.e. a straight through connection. The signal to be amplified may be an ac or dc voltage. However we will assume the signal to be an a.c. signal so that the input and output voltages are sinusoidal, at some fixed or variable frequency

The signal is a low level voltage such as obtained from a microphone, tape head, or a transducer. The output load may be a loudspeaker in an audio amplifier, a motor in a servo amplifier, a relay in control application, etc. In any case, the output of the amplifier is an enlarged version of the input. To amplify means to increase the amplitude of, raise the level of, or magnify input. Note that in the amplification process, the frequencies of the input and output signals are exactly identical. If not, there is " distortion" present in the amplifier. Then the signal is not faithfully reproduced at the output.

The amplifier is constructed using transistors. For their operation as an amplifier, transistors requires proper d.c. biasing. The necessary d.c. voltage is provided by a battery (usually a dry battery) or a d.c. source resulting from a rectifier and filter combination. In this case, the amplifier operates from 230 V,50 Hz a.c. mains supply. Very often, many amplifiers have the facility to operate from a.c. mains or battery.

Referring to Fig. 10.1,

V_S : is the signal voltage,

R_S : is the internal resistance of the source,

V_1 : is the actual input voltage to the amplifier,

I_1 : is the input current to the amplifier,

V_2 : is the output voltage across the load R_L,

I_2 : is the output current flowing through the load R_L.

The ratio of V_1 to I_1 is called input resistance, R_i , of the amplifier.

$R_i = \frac{V_1}{I_1}$

V_2 is the output voltage across the load R_L, and I_2 is the output current flowing through the load R_L. R_o is the output (or internal) resistance of the amplifier.

The ratio of output current to input current is called current gain, A_I, of the amplifier.

Current gain, $A_I = \frac{I_2}{I_1}$

The ratio of output voltage to input voltage is called voltage gain, $A_V = \frac{V_2}{V_1}$

The ratio of signal power delivered to the load to the signal power at the input of the amplifier is the power gain.

$$A_P = \frac{P_2}{P_1} = \frac{V_2 I_2}{V_1 I_1}$$

Power gain, $$A_P = \frac{V_2}{V_1} * \frac{I_2}{I_1} = A_V A_I$$

An amplifier may or may not exhibit both a voltage and a current gain, but in general will show a power gain. However, whether voltage or power gain is more important depends upon the application. The amplifier, in which voltage gain is more important than the power gain, is called a voltage amplifier, while in which power gain is more important than voltage gain is known as power amplifier.

CLASSIFICATION:

Based on transistor configuration:

a. Common emitter amplifier.

b. Common collector amplifier.

c. Common base amplifier.

SINGLE STAGE AMPLIFIERS:

Single stage amplifiers have only one amplifying device, say BJT in CE, CC, or CB configuration or FET in CS, CD, or CS configuration.

Based on the active device amplifiers are classified as:

- BJT Amplifier.
- FET Amplifier.

Common Emitter Amplifier:

- NPN Transistor, emitter base junction forward biased by power supply Vbb, collector base junction is reverse biased by power supply Vcc.
- Transistor is active zone throughout its operation.
- Input signal is applied to base emitter circuit and output signal is taken from the collector emitter circuit,

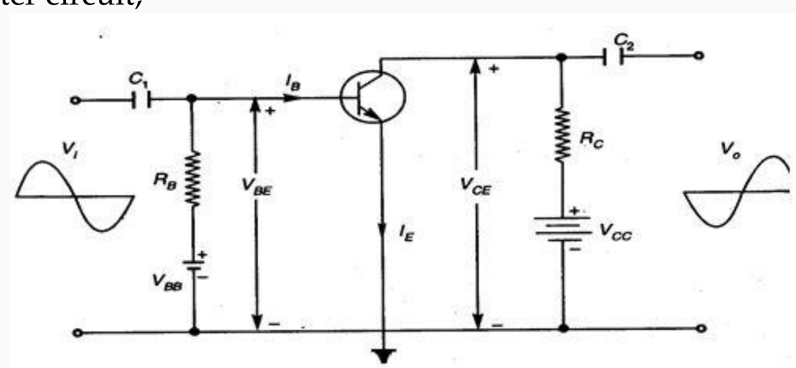


Fig. 10.1 Common emitter amplifier

C1, C2 are the coupling capacitors to provide DC isolation at the input and output of the amplifier.

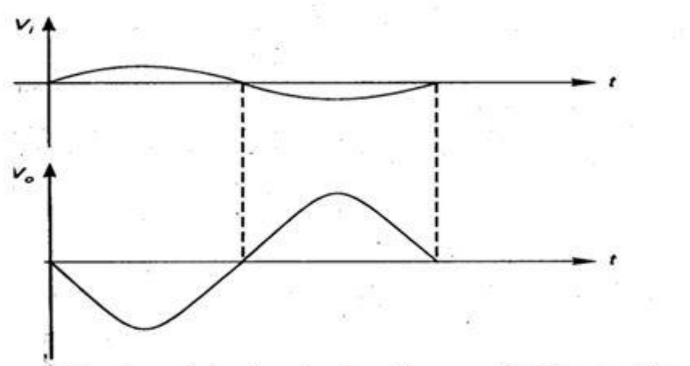
+ Ve input signal is converted into – Ve going output signal,

Ie. 180` phase shift.

Characteristics of CE Amplifier:

- Large current gain Ai.
- Large voltage gain Av.
- Large power gain Ap = Ai. Av.
- Voltage phase shift of 180°.
- Moderate input impedance.
- Moderate output impedance.

Input & Output voltages of CE Amplifiers:



10.3 Input and output voltages of CE amplifier

Common Collector (CC) Amplifier:

- NPN Transistor, emitter base junction forward biased by power supply Vee, collector base junction is reverse biased by power supply Vcc,
- Transistor is active zone throughout its operation.
- Input signal is applied to base collector circuit and output signal is taken from the emitter collector circuit,

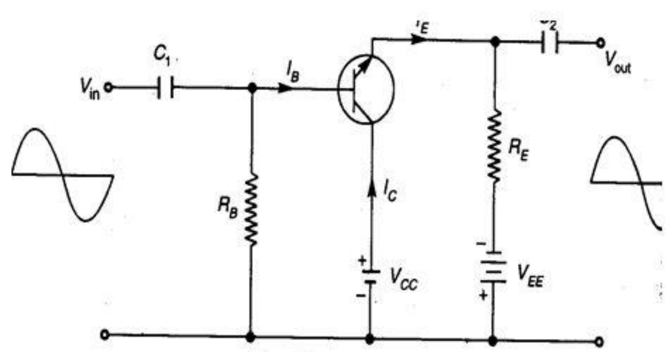


Fig. 10.4 CC amplifier

- C1, C2 are the coupling capacitors to provide DC isolation at the input and output of the amplifier.

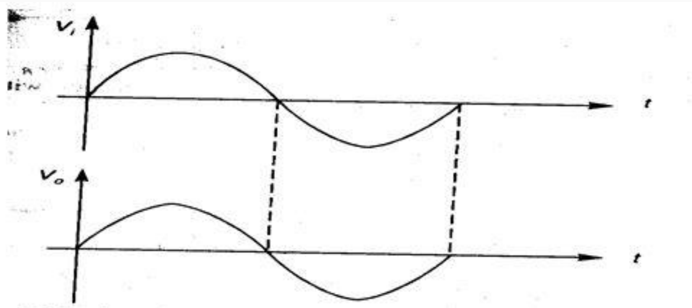
+ Ve input signal is converted into + ve going output signal,

Hence CC amplifier is also called as emitter follower.

Characteristics of CC Amplifier:

- High current gain
- Voltage gain approximately unity
- Power gain approximately equal to current gain
- No current or voltage phase shift
- Large input impedance
- Small output impedance

Input & Output voltages of CC Amplifier:



10.6 Input and output voltages of CC amplifier

LESSON 11. Common Base amplifier-characteristics-Feedback amplifier- types of feedback

Common Base (CB) Amplifier:

- NPN Transistor, emitter base junction forward biased by power supply Vee, collector base junction is reverse biased by power supply Vcc,
- transistor is active zone throughout its operation,
- Input signal is applied to emitter base circuit and output signal is taken from the collector base circuit.

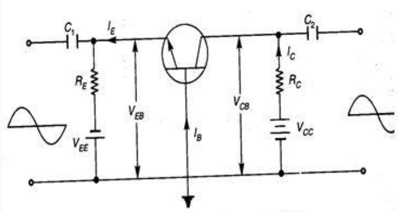


Fig. 10.7 Common base amplifier

- C1,C2 are the coupling capacitors to provide DC isolation at the input and output of the amplifier,
- \+ Ve input signal is converted into + ve going output signal without any phase reversal.

Characteristics of CB Amplifier:

- Current gain of less than unity
- High voltage gain
- Power gain approximately equal to voltage gain
- No phase shift for current or voltage
- Small input impedance
- Large output impedance

Input & Output voltages of CB Amplifier:

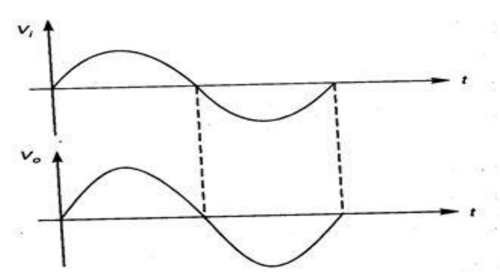
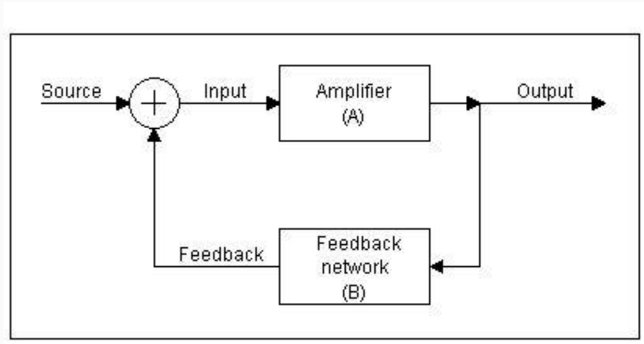


Fig. 10.8 Input and output voltages of CB amplifier

FEEDBACK AMPLIFIERS

Feedback is a mechanism, process or signal that is looped back to control a system within itself. Such a loop is called a feedback loop. In systems containing an input and output, feeding back part of the output so as to increase the input is positive feedback (regeneration); feeding back part of the output in such a way as to partially oppose the input is negative feedback (degeneration).



Generally, a control system has input from an external signal source and output to an external load; this defines a natural sense (or direction) or path of propagation of signal; the feedforward sense or path describes the signal propagation from input to output; feedback describes signal propagation in the reverse sense. When a sample of the output of the system is fed back, in the reverse sense, by a distinct feedback path into the interior of the system, to contribute to the input of one of its internal feed forward components, especially an active device or a substance that is consumed in an irreversible reaction, it is called the "feedback". The propagation of the signal around the feedback loop takes a finite time because it is causal.

The natural sense of feed forward is defined chemically by some irreversible reaction, or electronically by an active circuit element that has access to an auxiliary power supply, so as to be able to provide power gain to amplify the signal as it propagates from input to output. For example, an amplifier can use power from its controlled power reservoir, such as its battery, to provide power gain to amplify the signal; but the reverse is not possible: the signal cannot provide power to re-charge the battery of the amplifier. Feed forward, feedback and regulation are self related. The feed forward carries the signal from source to load.

Negative feedback helps to maintain stability in a system in spite of external changes. It is related to homeostasis. For example, in a population of foxes (predators) and rabbits (prey), an increase in the number of foxes will cause a reduction in the number of rabbits; the smaller rabbit population will sustain fewer foxes, and the fox population will fall back. In an electronic amplifier feeding back a negative copy of the output to the input will tend to cancel distortion, making the output a more accurate replica of the input signal

Positive feedback amplifies possibilities of divergences (evolution, change of goals); it is the condition to change, evolution, growth; it gives the system the ability to access new points of equilibrium.

When a public-address system is used with a microphone to amplify speech, the output from a random sound at the microphone may produce sound at a loudspeaker that reaches the microphone such as to reinforce and amplify the original signal (positive feedback), building up to a howl (of frequency dependent upon the acoustics of the hall). A similar process is used deliberately to produce oscillating electrical signals.

Feedback is distinctly different from reinforcement that occurs in learning, or in conditioned reflexes. Feedback combines immediately with the immediate input signal to drive the responsive power gain element, without changing the basic responsiveness of the system to future signals. Reinforcement changes the basic responsiveness of the system to future signals, without combining with the immediate input signal. Reinforcement is a permanent change in the responsiveness of the system to all future signals. Feedback is only transient, being limited by the duration of the immediate signal

Types of feedback

When feedback acts in response to an event/phenomenon, it can influence the input signal in one of two ways:

1. An in-phase feedback signal, where a positive-going wave on the input leads to a positive-going change on the output, will amplify the input signal, leading to more modification. This is known as positive feedback.

2. A feedback signal which is inverted, where a positive-going change on the input leads to a negative-going change on the output, will dampen the effect of the input signal, leading to less modification. This is known as negative feedback.

Positive feedback tends to increase the event that caused it, such as in a nuclear chain-reaction. It is also known as a self-reinforcing loop. An event influenced by positive feedback can increase or decrease its output/activation until it hits a limiting constraint. Such a constraint may be destructive, as in thermal runaway or a nuclear chain reaction. Self-reinforcing loops can be a smaller part of a larger balancing loop, especially in biological systems such as regulatory circuits.

Negative feedback, which tends to reduce the input signal that caused it, is also known as a self-correcting or balancing loop. Such loops tend to be goal-seeking, as in a thermostat, which compares actual temperature with desired temperature and seeks to reduce the difference. Balancing loops are sometimes prone to hunting: an oscillation caused by an excessive or delayed negative feedback signal, resulting in over-correction, wherein the signal becomes a positive feedback.

ADVANTAGES

a) Increased stability in the amplification. The gain is less dependent on the parameters of the amplifier elements.

b) Feedback reduces distortion in the amplifier.

c) The bandwidth of the amplifier is increased.

d) It is easier to achieve desired input and output impedances

Types of Amplifiers

- There are four common classes of amplifier in the high-fidelity reproduction of audio:
- Class A
- Class B
- Class AB
- Class D

Class A

- This is the most linear of the classes, meaning the output signal is a truer representation of what was imputed.
- characteristics of the class:
- they reproduce the entire waveform in its entirety.
- Class A is the most inefficient of all power amplifier designs, averaging only around 20.

They are the most accurate of all amps available, but at significant cost to manufacture, because of tight tolerances, and the additional components for cooling and heat regulation.

Class B

- In this amp, the positive and negative halves of the signal are dealt with by different parts of the circuit.
- Class B operation has the following characteristics:
- The input signal has to be a lot larger in order to drive the transistor appropriately.
- This is almost the opposite of Class A operation.

Class AB

- This is the compromise of the bunch. Class AB operation has some of the best advantages of both Class A and Class B built-in.
- Its main characteristics are:
- The output bias is set so that current flows in a specific output device for more than a half the signal cycle but less than the entire cycle.

Class D

- These amplifiers are erroneously called "digital" amplifiers by the press and many audio "experts." Here's the skinny on Class D:
- While some Class D amps do run in true digital mode, using coherent binary data, most do not.
- This efficiency gain is at the cost of high-fidelity.

MODULE 7.

LESSON 12. Transistor amplifier- h parameter-benefits-h parameter equivalent circuit for CE configurates- comparison.

Small Signal – Low Frequency h – Parameter Model

Let us consider transistor amplifier as a block box as shown in the Fig. 12.1.

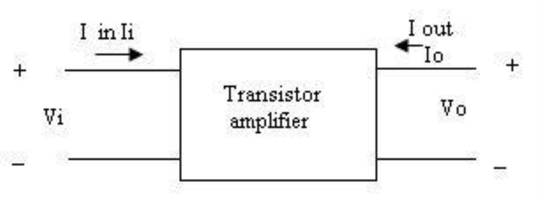


Figure 12.1 Transistor amplifier

Here, I_i : is the input current to the amplifier

V_i : is the input voltage to the amplifier

I_o : is the output current to the amplifier and

V_o : is the output voltage to the amplifier

As we know transistor is a current operated device, input current is an independent variable. The input current, Ii and output voltage V_o devices the input voltage V_i as well as the output current I_o. Hence input voltage V_i and outpur current I_o are the dependent variables, whereas input current I_i and output voltage V_o are independent variables. Thus we can write

$$V_i = f_1 (I_i, V_o) \qquad ... (1)$$

$$I_o = f_2 (I_i, V_o) \qquad ... (2)$$

This can be written in the equation form as follows

$$V_i = h_{11} I_i + h_{12} V_o \qquad ... (3)$$

$$I_o = h_{21} I_i + h_{22} V_o \qquad ... (4)$$

The above equations can also be written using alphabetic notations,

$$V_i = h_i . I_i + h_r . V_o \qquad ... (5)$$

$$I_o = h_f . I_i + h_o . V_o \qquad ... (6)$$

Definitions of h – parameter

The parameters in the above equation are defined as follows:

$h_{11} = \frac{V_1}{I_1} \,|V_1o = 0$ = input resistance with output short – circuited, in ohms..

$h_{12} = \frac{V_1}{V_o} \,|I_1i = 0$ = Fraction of output voltage at input with input open circuited.

This parameter is ratio of similar quantities, hence unitless

$h_{21} = \frac{I_o}{I_1} \,|V_1o = 0$ = Forward current transfer ratio or current gain with output short circuited.

This parameter is a ratio of similar quantities, hence unitless.

$h_{22} = \frac{I_o}{V_o} \,|I_1i = 0$ = Output admittance with input open-circuited, in mhos.

From the above discussion we can say that, these four parameters are not same. They have different units. In other words, they are mixture of different units and hence referred to as hybrid parameters. As we use small letter for ac analysis, these are commonly known as h-parameters. The standard notations can be given as

i = 11= input   o = 22 = output

f = 21 = forward transfer   r = 12 = reverse transfer

Thus we can write h-parameters as follows.

a) With output short circuited :

$h_{11} = h_i$: Input resistance

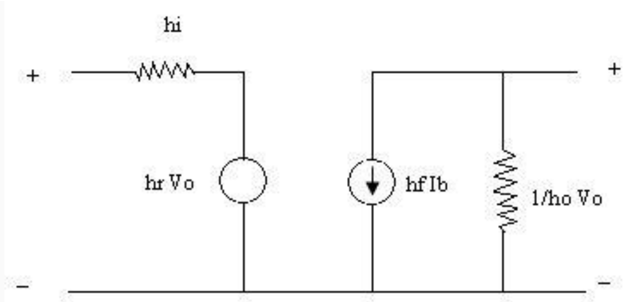
$h_{21} = h_f$: Short circuit current gain

b) With input open circuited :

$h_{12} = h_r$: Reverse voltage transfer ratio

$h_{22} = h_o$: output admittance.

H- parameter equivalent circuit for transistor is shown in the following figure



In order to analyze transistorized amplifier circuit and calculate its input impedance, output impedance, current gain and voltage gain, it is necessary to replace transistor circuit with its equivalent. The equivalent circuit can be drawn with the help of two equations, as shown in Fig. 1.10.

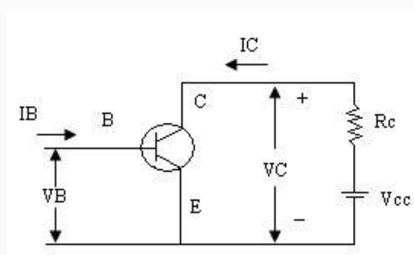
$$V_i = h_i I_i + h_r V_o \qquad I_o = h_f I_i + h_o V_o$$

Many transistor models have been proposed, each one having its advantages and disadvantages. The transistor model used in this text is in terms of h-parameters.

Benefits of h-parameters

1. Real numbers at audio frequencies.
2. Easy to measure.
3. Can be obtained from the transistor static characteristics curves.
4. Convenient to use in circuit analysis and design.
5. Most of the transistor manufacturers specify the h-parameters.

H – parameters equivalent circuit for CE configuration in the following figure



Simple common emitter configuration

To see how we can derive a hybrid model for a transistor, let us consider the common emitter configuration as shown in the above figure.. The variables I_b, I_c, V_b and V_c represent total instantaneous currents and voltages.

I_b = input current

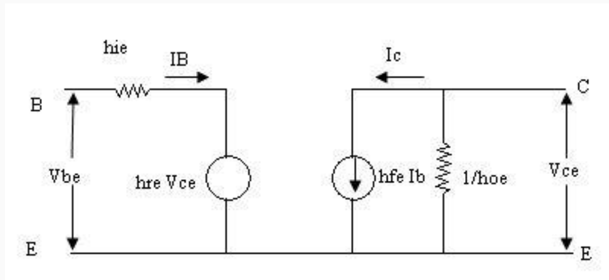
I_c = output current

V_{be} = input voltage

V_{ce} = output voltage.

The following figure shows the h-parameter equivalent circuit for the common emitter configuration.

h-parameter equivalent circuit for the common emitter configuration



From the h-parameter equivalent circuit of the common emitter configuration we can write,

$$V_{be} = h_{ie} I_b + h_{re} V_{ce} \qquad \ldots(7)$$

$$I_c = h_{fe} I_b + h_{oe} V_{ce} \qquad \ldots(8)$$

Where $h_{ie} = \dfrac{\Delta V_{BE}}{\Delta I_B} \,|V_1 (CE\ constant)$ ...(9)

$h_{re} = \dfrac{\Delta V_{BE}}{\Delta V_{CE}} \,|I_1 (B\ constant)$ ... (10)

$h_{fe} = \dfrac{\Delta I_C}{\Delta I_B} \,|V_1 (CE\ constant)$ ...(11)

$h_{oe} = \dfrac{\Delta I_C}{\Delta V_C} \,|I_1 (B\ constant)$ ... (12)

The quantities ΔV_{BE} (V_{be}), ΔV_{CE} (V_{ce}), ΔI_B (I_b) and ΔI_C (I_c) represent the small change in base and collector voltages and currents.

H-parameters for all three configurations

As mentioned earlier, transistor can be represented as a two port network by making any one terminal common between input and output. Since there are three possible configurations in which a transistor can be used, there is a change in terminal voltage and current for different transistor configurations. For different configurations the relation between input parameters and output parameters also differs. Therefore, one needs to define different set of h-parameters for different configurations. To designate the type of configuration another subscript is added to the h-parameters.

For example :

$h_{ie} = h_{11e}$ = input resistance in common emitter configuration.

$h_{fb} = h_{21b}$ = short-circuit current gain in common base configuration.

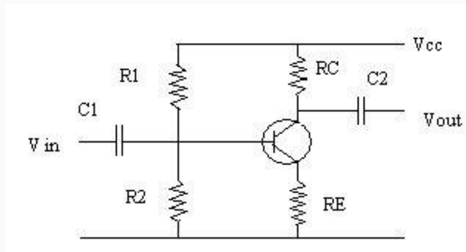
The following table summarizes the h-parameters for all the three configurations.

parameter	CB	CE	CC
Input resistance	h_{ib}	h_{ie}	h_{ic}
Reverse voltage gain	h_{rb}	h_{re}	h_{rc}
Forward transfer current gain	h_{fb}	h_{fe}	h_{fc}
Output admittance	h_{ob}	h_{oe}	h_{oc}

MODULE 8.

LESSON 13. Common emitter amplifier-practical circuits-phase reversal

Common emitter amplifier



An amplifier is used to increase the small signal level; i.e. the amplifier is used to get a larger signal output from a small signal input. We will assume a sinusoidal signal at the input of the amplifier. At the output, signal must remain sinusoidal in waveform, with frequency same as that of the input.

To make the transistor work as an amplifier, it is to be biased to operate in the active region, i.e. base-emitter junction is to be forward biased, while base-collector junction to be reversed biased.

Let us consider the common emitter amplifier circuit using self bias or voltage divider bias as shown in the above Figure.

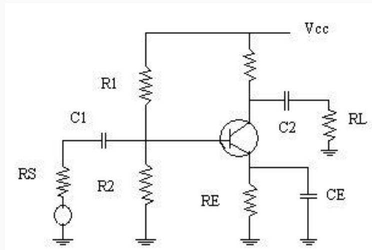
I_{SQ} is quiescent DC base current

In the absence of input signal, only dc voltage are present in the circuit. This is known as **zero-signal** or **no-signal** condition or quiescent condition for the amplifier. The dc collector-emitter voltage, V_{CE}, the dc collector current I_C and dc base current I_B is the **quiescent** operating point for the amplifier. On this dc quiescent operating point, we superimpose ac signal by application of ac sinusoidal voltage at the input. Due to this base current varies sinusoidally, as shown in Figure.

Since the transistor is biased to operate in the active region, the output is linearly proportional to the input. The output current i.e. the collector current is β times larger than the input base current in common emitter configuration. Hence the collector current will also vary sinusoidally about its quiescent value, I_{CQ}. The input voltage will also vary sinusoidally as shown in the Figures..

The variations in the collector current and the voltage between collector and emitter due to change in the base current are shown graphically with the help of load line in .The collector current varies above and below its Q point value in-phase with the base current, and the collector-to-emitter voltage varies above and below its Q point value 180° out-of-phase with the base voltage. When one cycle of input is completed, one cycle of output will also be completed. This means the frequency of output sinusoidal is the same as the frequency of input sinusoid. Thus in the amplification process, frequency of the output signal does not change, only the magnitude of the output is larger than that of the input.

Common Emitter Amplifier Circuit



Practical common emitter amplifier circuit

The above figure shows the practical circuit of common emitter transistor amplifier. It consists of different circuit component. The functions of these components are as follows:

1. **Biasing Circuit**

The resistances R_1, R_2 and R_E forms the voltage divider biasing circuit for the CE amplifier. It sets the proper operating point for the CE amplifier.

2. **Input Capacitor C_1**

This capacitor couples the signal to the base of the transistor. It blocks any dc component present in the signal and passes only ac signal for amplification. Because of this biasing conditions are maintained constant.

3. Emitter Bypass Capacitor C_E

An emitter bypass capacitor C_E is connected in parallel with the emitter resistance, RE to provide a low resistance path to the amplified ac signal. If it is not inserted, the amplified ac signal passing through RE will cause a voltage drop across it. This will reduce the output voltage, reducing the gain of the amplifier.

4. **Output Coupling Capacitor C_2**

The coupling capacitor C_2 couples the output of the amplifier to the load or to the next stage of the amplifier. It blocks dc and passes only ac part of the amplified signal.

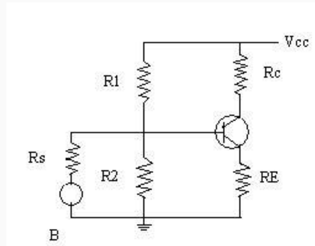
Need for C_1, C_2 and C_E

We know that, the impedance of capacitor is given as

$$X_C = \frac{1}{2\pi f C}$$

Thus, at signal frequencies all the capacitors have extremely small impedance and it can be treated as an ac short circuit. For bias / dc conditions of the transistor all the capacitors act as a dc open circuit. With this knowledge we will see the importance of C_1, C_2 and C_E.

Consider that the signal source is connected directly to the base of the transistor as shown in the following figure. Looking at the Figure, we can immediately notice that source resistance R_S is in parallel with R_2. This will reduce the bias voltage at the transistor base and, consequently alter the collector current, which is not desired. Similarly, by connecting R_L directly, the



dc levels of V_C and V_{EC} will change. To avoid this and maintain this stability of bias condition coupling capacitors are connected. As mentioned earlier, coupling capacitors act as open circuits to dc, maintain stable biasing conditions even after connection of R_S and R_L. Another advantages of connecting C_1 is that any dc component in the signal is opposed and only ac signal is routed to the transistor amplifier.

The emitter resistance R_E is one of the component which provides bias stabilization. But it also reduces the voltage swing at the output. The emitter bypass capacitor C_E provides a low reactance path to the amplified a.c. signal increasing the output voltage swing.

For the proper operation of the circuit, polarities of the capacitors must be connected correctly. The curve bar which indicates negative terminal must always be connected at a dc voltage level lower than (or equal to) the dc level of the positive terminal (straight bar). For example, C_1 in Fig. 1.4 has its negative terminal at dc ground level, because it is grounded through the signal source resistance R_S. The positive terminal of C_1 is at $+V_B$ with respect to ground.

Phase Reversal

The phase relationship between the input and output voltages can be determined by considering the effect of a positive half cycle and negative half cycle separately. Consider the positive half cycle of input signal in which terminal A is positive w.r.t B. Due to this, two voltages, ac and dc will be adding each other, increasing forward bias on base emitter junction. This increases base current. The collector current is β times the base current, hence the collector current will also increase. This increases the voltage drop across R_C. Since $V_C = V_{CC} - I_C R_C$, the increases in I_C results in a drop in collector voltage V_C, as V_{CC} is constant. Thus, as V_i increases in a positive direction, V_o goes in a negative direction and we get negative half cycle of output voltage for positive half cycle at the input.

In the negative half cycle of input, in which terminal A becomes negative w.r.t. terminal B, the ac and dc voltages will oppose each other, reducing forward bias on base-emitter p-n junction. This reduces base current. Accordingly collector current and drop across R_C both reduce, increasing the output voltage. Thus, we get positive half cycle at the output for negative half cycle at the input. Therefore, we can say that there is a phase shift of 180° between input and output voltages for a common emitter amplifier.

LESSON 14. Oscillator- LC oscillator-Armstrong oscillator-Hartley oscillator-Colpitt oscillator-Crystal oscillator.

Oscillators

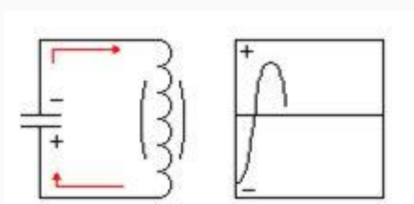
An Electronic device, that generate oscillations (Signals), is called an oscillator. Simply says an oscillator receives DC energy and converts it into AC energy of desired frequency. The frequency of oscillations depends up on the constants of the device.

Oscillators are extensively used in electronic equipments.Oscillators can produce sinusoidal or non sinusoidal signals. Electronic oscillators are commonly found in everyday circuits, ranging from antique radios to the transmitter in your TV remote. The basic job of an electronic oscillator is to generate an oscillating output often at a constant frequency. Various outputs can be sine, square, sawtooth, triangle, or complex waveforms.

The LC Oscillator

This oscillator consists of a capacitor and a coil connected in parallel. To understand how the LC oscillator basically works, let's start off with the basics. Suppose a capacitor is charged by a battery. Once the capacitor is charged, one plate of the capacitor has more electrons than the other plate, thus it is charged. Now, when it is discharged through a wire, the electrons return to the postive plate, thus making the capacitor's plates neutral, or discharged. However, this action works differently when you discharge a capacitor through a coil. When current is applied through a coil, a magnetic field is generated around the coil. This magnetic field generates a voltage across the coil that opposes the direction of electron flow. Because of this, the capacitor does not discharge right away. The smaller the coil, the faster the capacitor discharges. Now the interesting part happens. Once the capacitor is fully discharged through the coil, the magnetic field starts to collapse around the coil. The voltage induced from the collapsing magnetic field recharges the capacitor oppositely. Then the capacitor begins to discharge through the coil again, generating a magnetic field.

This process continues until the capacitor is completely discharged due to resistance.



Technically this basic LC circuit generates a sine wave that loses voltage in every cycle. To overcome this, additional voltage is applied to keep the oscillator from losing voltage.

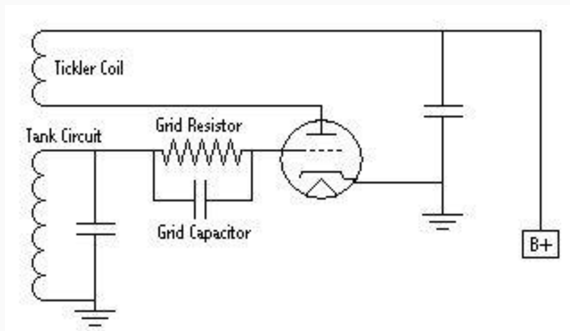
However, to keep this oscillator going well, a switching method is used. A vacuum tube (or a solid-state equivalent such as a FET) is used to keep this LC circuit oscillating.

The advantage of using a vacuum tube is that they can oscillate at specified frequencies such as a thousand cycles per second.

There are several different types of LC oscillators. A little off subject; one well known and entertaining oscillator is known as the tesla coil, which uses an unique LC circuit using the spark gap to oscillate.

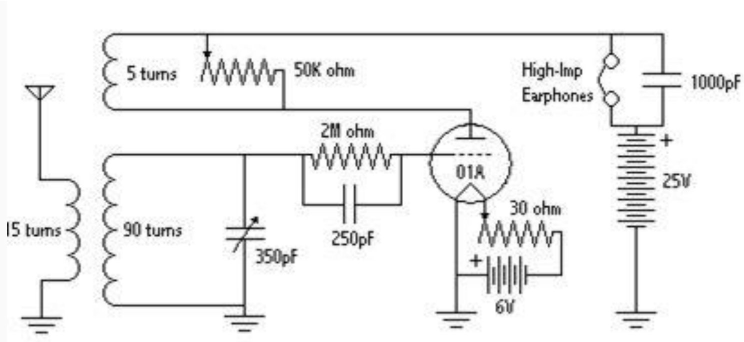
Armstrong Oscillator

This oscillator is very much like a RF amplifier, but a new coil, called the tickler coil, is connected between the plate and the B+, or high voltage supply. This coil is generally wound next to the main LC coil, or tank coil.



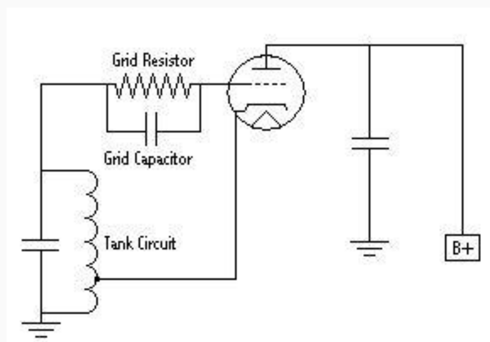
When current flows in the plate, a electromagnetic field gives feedback to the tank coil, which keeps the oscillations going. The grid resistor drops the voltage, thus the grid is very negative with respect to the cathode. The grid capacitor keeps enough charge to keep the grid negative for at least one cycle of oscillation, it helps keep the grid negative when either side of the LC circuit is postive. When the LC circuit's postive charge is at it's maximum, the charge will balance with the grid capacitor, causing plate current to flow because there is no negative on the grid. This is the point when the tickler coil provides feedback to the LC circuit. The grid controls the plate current in all vacuum tubes, thus if the grid oscillates this number of times, the plate will oscillate the the same number of times, the tickler coil will give feedbacks at the same number of times. This is because of the tank circuit, which specifies the frequency. The frequency can be adjusted when the coil and/or capacitor are adjusted. The bigger the coil and capacitor are, the lower the frequency. The smaller the coil and capacitor are, the higher the frequency. However, in an antique radio, it is handier to adjust the capacitor than the coil to specify a frequency so a variable capacitor is used. This is a very typical feedback oscillator, however the greatest disadvantage of this oscillator are unstable frequencies.

Some of you may be familar with this circuit. This circuit is also known as the regenerative receiver Edwin H. Armstrong developed. Below is a basic regenerative radio circuit you can construct that is similar to the Armstrong oscillator.



Hartley Oscillator

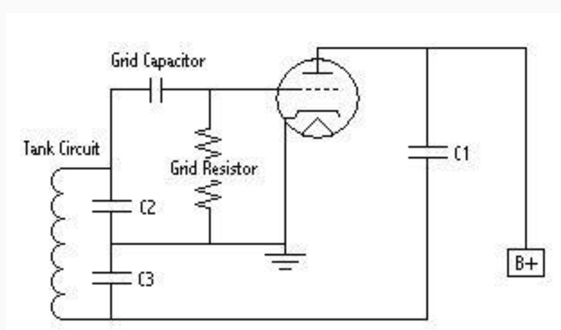
This oscillator is very similar to the Armstrong oscillator and is commonly used. The difference between the Armstrong oscillator and the Hartley oscillator is that the tickler coil is part of the LC circuit. This oscillator is easier to tune compared to the Armstrong oscillator.



The cathode is tapped to the coil so when current flows through the coil, there is a voltage kick in the grid coil. The amount of feedback is controlled by changing the cathode tap. Most of the LC circuit works in a manner like the Armstrong oscillator. The Hartley oscillator is an improvement on the Armstrong oscillator, however it has some frequency instabilities.

Colpitts Oscillator

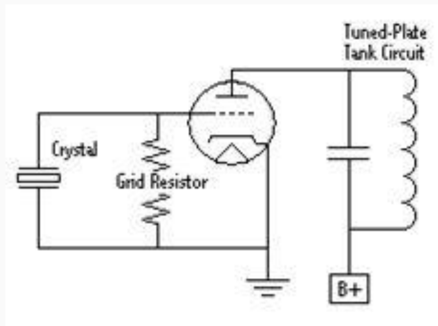
The Colpitts oscillator is very similar to the Hartley oscillator, but instead of a tapped grid coil, it has tapped capacitance.



The tap between the two capacitors is grounded and the feedback is obtained from the coupling capacitor, C1. The amount of feedback depends on the ratio of C2 to C3. The capacitor part of the LC circuit consists of both C2 and C3, which determines the oscillating frequency. This oscillator has more frequency stablities than the Hartley oscillator.

Crystal Oscillator

This is a type of oscillator that is controlled by a crystal. The big advantage of a crystal oscillator is high frequency stablility. Common crystals used are tourmaline, Rochelle salts, and quartz. The crystal makes a voltage difference when voltage is applied to the two plates on the crystal. When AC is applied, the crystal compresses and stretches, in other words it vibrates. The natural frequency of a crystal's vibrations is found to be more constant than the oscillations in a LC circuit. The thinner the crystal is, the faster it vibrates. The LC circuit is the electricial equilavent of a crystal.

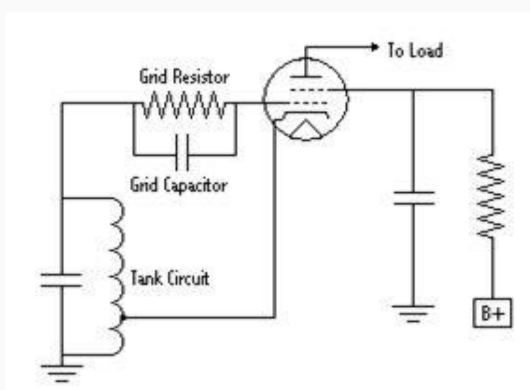


Notice there is a LC circuit on the plate circuit now. As said earlier, the crystal vibrates at its own frequency, so the LC circuit is on the plate to adjust the amplitude of oscillations. However, more components are suggested in this circuit to maintain the voltages and RF in the circuit. The disadvantage of this oscillator is its limited power output.

Caring and feeding of a crystal is important. Crystals will overheat or crack when fed with too much voltage. The current flowing through a crystal generally should not be more than 100mA (.1A).

Electron-Coupled Oscillator (ECO)

This is a unique oscillator that generally uses a pentode, sometimes a tetrode. This oscillator has advantages over the previous oscillators, such as good frequency stabilities. Below is a ECO equilavent of the Hartley oscillator.



The circuit functions very much like a Hartley oscillator except feedback is obtained from the screen grid instead of plate. A dropping resistor is used in series with the screen grid and B+ because the screen grid uses less power compared to the plate.

MODULE 9.

LESSON 15. Integrated circuit-advantages-classification-types of amplifiers-operational amplifiers-characteristics of ideal operational amplifier

Integrated Circuit is an Improvement in the characterisation and minimisation of solid state devices and components (transistors, diodes etc.,)

It is a single – crystal chip of silicon, containing both active (transistors and diodes) and passive (resistors and capacitors) elements and their interconnections

Advantages of Integrated Circuit

1. Extremely small in size
2. Low power consumption
3. Reliability
4. Reduced cost
5. Very small weight
6. Easy replacement

Classification of ICs Based on type of signal processed

1. Linear or Analog ICs – Op-amps, voltage regulators, voltage comparators, timers etc., - signals represented by continuous or analog variables
2. Digital ICs – logic gates, counters, digital clock chips, calculator chips, memory chips, memory processors etc., - represented by binary digits and involve logic and memory

Classification of ICs Based on fabrication process

1. Monolithic ICs
2. Hybrid ICs

Types of amplifiers

1. Operational Amplifier
2. Buffer amplifier
3. Differential amplifier
4. Instrumentation amplifier

Operational Amplifiers

The operational amplifier is a direct-coupled high gain (negative feedback) amplifier. It can amplify signals having frequency ranging from 0H_Z to MH_Z. It is used to perform a wide variety of linear functions and also some non-linear functions. Therefore it is also referred to as the basic linear integrated circuit. An operational amplifier is so named because it was originally designed to perform mathematical operations like subtraction, summation, differentiation, integration and multiplication etc., in analog computer.

Basic operational Amplifiers

Fig 15.1 shows the schematic diagram operational amplifier. It contains tow inputs called as inverting and non-inverting input.

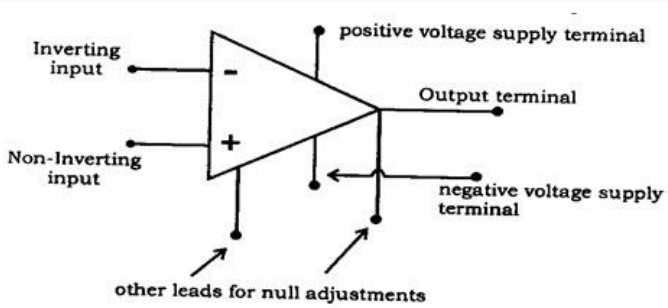


Fig Symbol for Op-Amp

1. Inverting input

The signal given to this input is always inverted at the output. i.e. When positive supply is given, the output will be negative. When negative supply is given, the output will be positive.

2. Non-inverting input

When the signal is given to this non-inverting input terminal, output is available without any change in sign.

Characteristics Ideal operational amplifier:

1. Input resistance R_i = α

2. Output resistance R_0 = O

3 Voltage gain V_A = α

4, Bandwidth = α

5. Perfect balance V_0 = O when $V_1 = V_2$

6. Characteristics do not drift with temperature.

Block diagram of an Op-amp.

An integrated amplifier is a four stage cascaded amplifier as shown in Fig. 15.2

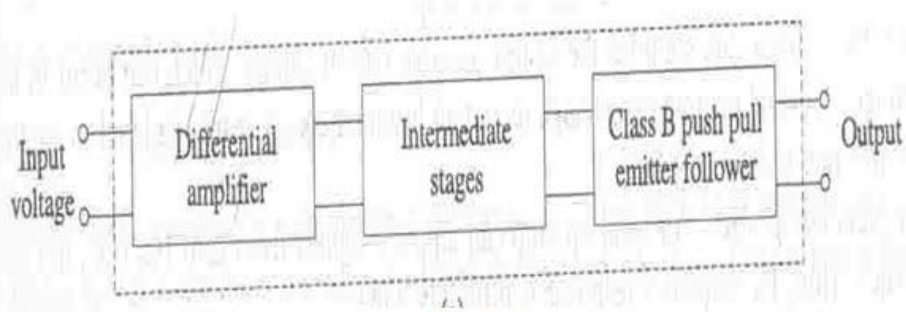


Fig Op-Amp block diagram

The first stage is a differential amplifier with a double ended output. This stage provides the maximum voltage gain. The second stage is also a differential amplifier. But is has single ended output driver. Because of this stage, the output voltage is zero when the input is zero. And also output resistance is very low and hence it supplies large output voltage or current.

Output resistance (Ro)

This is the effective resistance between the two inputs when it is operated open loop system.

Differential input resistance (Ri)

This is the effective resistance between the two inputs when it is operated in open loop system.

Differential voltage gain (Ad)

This is the ratio of change in output voltage to the corresponding change in differential input voltage.

Slew rate (Sr)

Slew rate is defined as the maximum rate of change of output voltage, when the input is over driven (at large input signal condition). Its unit is volt/micro second.

Input offset voltage (Vor)

This is the differential DC input voltage required to provide zero output voltage with zero input signal. It is usually represented at room temperature and at rated power supply voltage.

Common mode voltage gain (Ac)

This is the ratio of the output signal voltage to the signal voltage applied to the two input terminals in parallel.

Common Mode Rejection Ratio (CMRR)

In an ideal op-amp the common mode input signal is rejected and the output is zero. But real op-amps have a common mode gain Ao.

The common mode rejection ratio is the ratio of the differential mode voltage gain A_d to the common mode voltage gain A_c

CMRR = Ad / Ac

Large value of a CMRR is always preferred. It is expressed in decibels.

Open loop voltage gain: (A)

This is the effective output signal voltage to differential input signal voltage without feed back.

Pin details of Op-Amp

The pin details of the p-Amp is shown in fig.15.3.

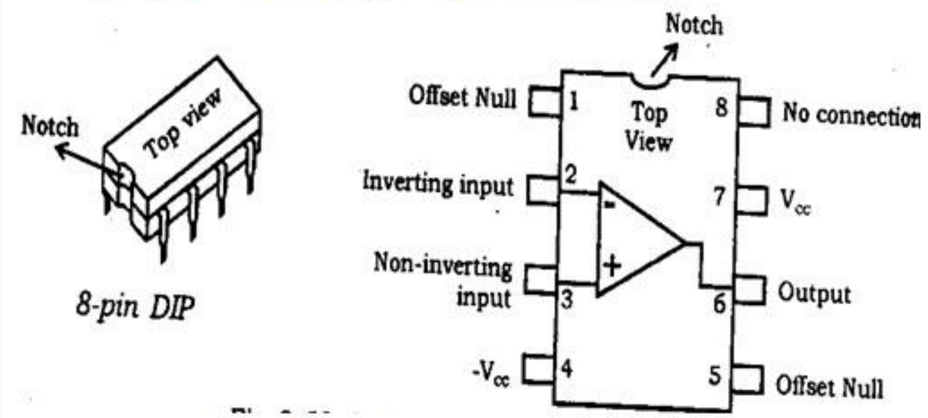


Fig. Op-Amp Pin-out configuration

The top pin on the left side of the notch indicates Pin 1. The pin number 2 is inverting input terminal and 3 is non-inverting input terminal. Pin 6 is output terminal. A d.c. voltage or a.c. signal placed on the inverting input will be 180° out of phase at the output. A d.c. voltage or a.c. signal placed on the non-inverting input will be inphase at the output. Pins 7 and 4 are the power supply terminals. Terminals 1 and 5 are used for null adjustment. Null adjustment pins are used to null the output voltage when equal voltages are applied to the input terminals for perfect balance. Pin 8 indicates no connection.

MODULE 10.

LESSON 16. Application of operated amplifier-inverting-non inverting- difference amplifier.

1) Inverting amplifier

The basic OP-AMP inverting amplifier is shown in Fig.16.1. The input voltage V_{in} is applied to the inverting input through the input resistor R_{in}. the non inverting input is grounded. The feedback resistor R_f is connected between the output and the inverting input.

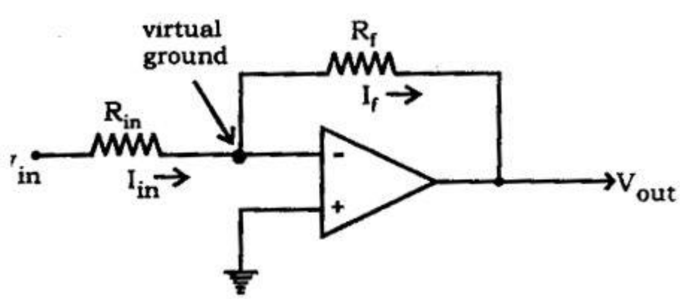


Fig.16.1. Inverting amplifier

Since the input impedance of an op-amp is considered very high, no current can flow into or out of the input terminals. Therefore I_{in} must flow through R_f and is indicated by I_f (the feedback current). Since R_{in} and R_f are in series, then $I_{in} = I_f$. The voltage between inverting and non-inverting inputs is essentially equal to zero volt. Therefore, the inverting input terminal is also at 0 volt. For this reason the inverting input is said to be at virtual ground. The output voltage (V_{out}) is taken across R_f.

It can be proved that

$$I_f = - V_{out} / R_f$$

Since $I_{in} = I_f$ then

$$V_{in} / R_{in} = -V_{out} / R_f$$

Rearranging the equation, we obtain

$$-V_{out} / V_{in} = R_f / R_{in}$$

Therefore the voltage gain of an inverting amplifier can be expressed as

$$A_v = -R_f / R_{in}.$$

The amplifier gain is the ratio of R_f to R_{in}

Finally, the output voltage can be found by

$V_{out} = -(R_f / R_{in}) \times V_{in}$

The output voltage is out of phase with the input voltage.

2) Non-inverting amplifier

The basic OP-AMP non-inverting amplifier is shown in Fig.16.2. The input signal V_{in} is applied to the non-inverting input terminal. The resistor R_{in} is connected from the inverting input to ground. The feedback resistor R_f is connected between the output and the inverting input.

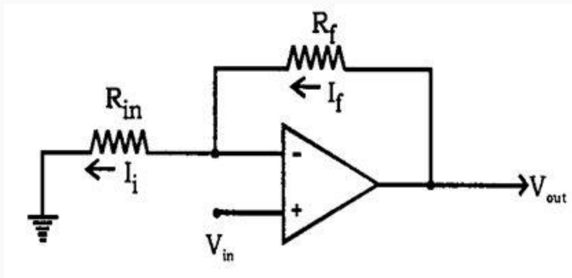


Fig.16..2. Non-inverting amplifier

Resistors R_f and R_{in} form a resistive ratio network to produce the feedback voltage (V_A) needed at the inverting input. Feedback voltage (V_A) is developed across R_{in}. Since the potential at the inverting input tends to be the same as the non-inverting input (as pointed out with the description of virtual ground), $V_{in} = V_A$.

Since $V_A = V_{in}$, the gain of the amplifier can be expressed as

$A_v = V_{out} / V_A$

However, V_A is determined by the resistance ratio of R_{in} and R_f ;

Thus,

$$V_A = \left\{ \frac{R_{in}}{R_f + R_{in}} \right\} V_{out}$$

(Or) $$V_{out} / V_A = \frac{R_f + R_{in}}{R_{in}}$$

$$V_{out} / V_A = 1 + \left\{ \frac{R_f}{R_{in}} \right\}$$

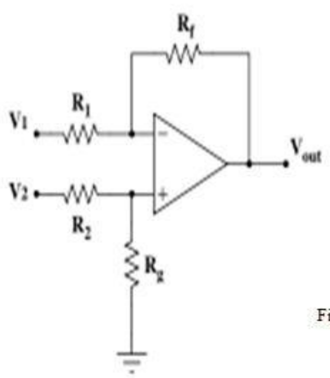
$$A_v = 1 + \left\{ \frac{R_f}{R_{in}} \right\}$$

Finally, the output voltage can be found by, $V_{out} = [1 + (R_f / R_{in}) V_{in}$

It is seen that the input and output voltages are in phase.

3) Difference amplifier:-

A differential amplifier is a type of an electronic amplifier that multiplies the difference between two inputs by some constant factor (the differentialgain). The basic OP-AMP Differential amplifier is shown in Fig.16.3.



$$V_{out} = V_2\left(\frac{(R_f + R_1)R_g}{(R_g + R_2)R_1}\right) - V_1\left(\frac{R_f}{R_1}\right)$$

Fig.16.3. Difference amplifer

By convention, the net difference of two voltages measured with respect to a common reference is called the differential-mode voltage, while the sum of the voltages, usually divided by two to give an average value, is called the common-mode voltage.

An ideal differential amplifier produces an output that is directly proportional to its differential-mode voltage.

The amplifier delivers zero output in response to common-mode

voltages.

The common-mode gain, the ratio of the output response of a real differential amplifier to the input signal applied equally to each input terminal, is a measure of this gain mismatch.

Differential amplification is very useful when the signal to be amplified exists in an electrically noisy environment. The most important requirement for a differential amplifier is that it be constructed with transistors with closely matched electrical characteristics .

Integrated circuits with amplifier transistors physically close to each other meet the required close matching requirement and are ideally suited for the production of differential amplifiers.

A differential amplifier is the input stage of operational amplifier or op-amps, and emitter coupled logic gates. Given two inputs Vin+ and Vin- , a practical differential amplifier gives an output *V*out:

$$V_{\rm{out}} = A_{\rm{d}}(V_{\rm{in}}^+ - V_{\rm{in}}^-) + A_{\rm{c}}\left({{V_{\rm{in}}^+ + V_{\rm{in}}^-} \over 2} \right)$$

The common mode rejection ratiois usually defined as the ratio between differential-mode gain and common-mode gain:

$$\mathrm{CMRR}=\frac{A_{\mathrm{d}}}{A_{\mathrm{c}}}$$

For a perfectly symmetrical differential amplifier with $A_c = 0$, the output voltage is given by,

$$V_{\mathrm{out}}=A_{\mathrm{d}}(V_{\mathrm{in}}^{+}-V_{\mathrm{in}}^{-})$$

Differential amplifiers are found in many systems that utilise negative feed back, where one input is used for the input signal, the other for the feedback signal.

A common arrangement for implementing a differential amplifier is the long-tailed pair, which is also usually found as the differential element in most op-amp integrated circuits

Applications of Difference amplifiers:-

When we have an input which has come from some distance and may have had some added interference. Using a pair of wires to send the signal we can then take the difference in potential between them as the signal and reject any ‘common mode’ voltages on both wires as being induced by interference. In feedback arrangements we can use the second input to control the behaviour of the amplifier. When we wish to combine two signals we can feed one into one transistor, and the second signal into the other.

LESSON 17. Application of operated amplifier-Summing-Integrating-Differentiating amplifier

Summing amplifier

The summing amplifier provides an output voltage equal to the algebraic sum of the input voltages.

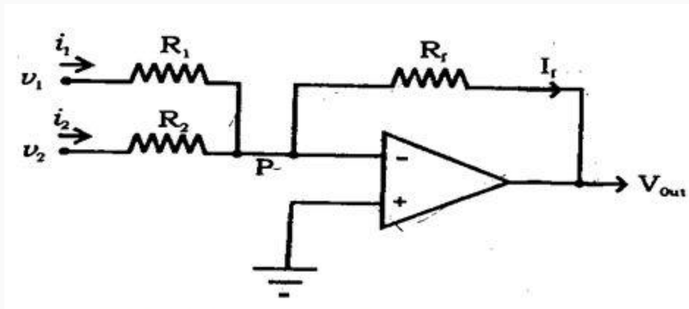


Fig.17.1. Summing amplifier

Fig.17.1 shows an inverting amplifier, used to sum two input voltages. The input voltages v_1 and v_2 are applied through the resistors R_1 and R_2 to the summing junction (P) and R_f is the feedback resistor. At the point P,

$$i_1 + i_2 = i_f$$

Since the voltage at the point P is ideally 0,

$$(V_1 / R_1) + (V_2/R_2) = V_{out} / R_f$$

Hence the output voltage,

$$V_{out} = (R_f / R_1) V_1 + (R_f / R_2)V_2$$

If $R_1 = R_2 = R_f = R_v$ then $V_{out} = - (V_1 + V_2)$

Hence the output voltage is equal to the sum of the input voltages and the circuit acts as a summing amplifier. The negative sign indicates than OP-AMP is used in the inverting mode.

Difference amplifier

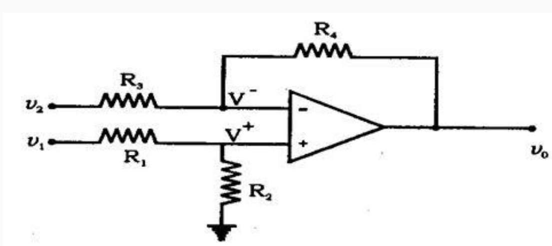


Fig.17.2. Difference amplifier

The difference amplifier is shown in Fig.17.2. The output voltage can be obtained by using superposition principle. To find the output voltage V_{01} due to V_1 alone, assume that V_2 is shorted to ground. Then

$$V+ = [R_2 / (R_1 + R_2)]\, V_1$$

And $V_{01} = [(R_3 + R_4) / R_3]\, V+ = [(R_3 + R_4) / R_3]\, [R_2 / (R_1 + R_2)]\, V_1$

Now assuming that V_1 is shorted to ground, the output voltage V_{02} due to V_2 alone is given by

$$V_{02} = (R_4/R_3)\, V_2$$

Therefore, with both inputs present, the output is

$$V_0 = V_{01} + V_{02}$$

$$= [(R_3 + R_4) / R_3]\, [R_2 / (R_1 + R_2)]\, V_1 - (R_4/R_3)V_2$$

If $R_1 = R_2 = R_3 = R_4 = R$

Then $V_o = V_1 - V_2$

If all the external resistors are equal, the voltage difference amplifier functions as a voltage subtractor.

Integrating amplifier

If a capacitor is inserted in the feedback path, we got an output, which is the time, integral of the input signal. Figure 17.3 (a) shows the circuit and Fig.17.3 (b) shows the equivalent circuit (with virtual ground).

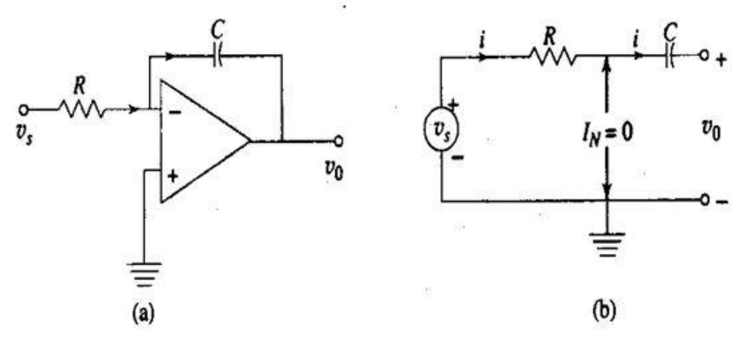


Fig.17.3. Integrating amplifier (a) Circuit (b) Equivalent circuit

Thus, the output voltage is proportional to the integral of input voltage.

$$V_{\text{out}} = \int_0^t -\frac{V_{\text{in}}}{RC}\, dt + V_{\text{initial}}$$

If the input voltage is constant, i.e., $V_{in} = V_s = V$, the output will be $-Vt/RC$ which is a ramp. Thus, the circuit of Fig 17.3 (a) can be used as a ramp generator.

Differentiating amplifier

The insertion of a capacitor in the input path and resistor in the feed back path (Fig.17.4 (a)) gives a differentiating amplifier. Figure17.4 (b) shows the equivalent current.

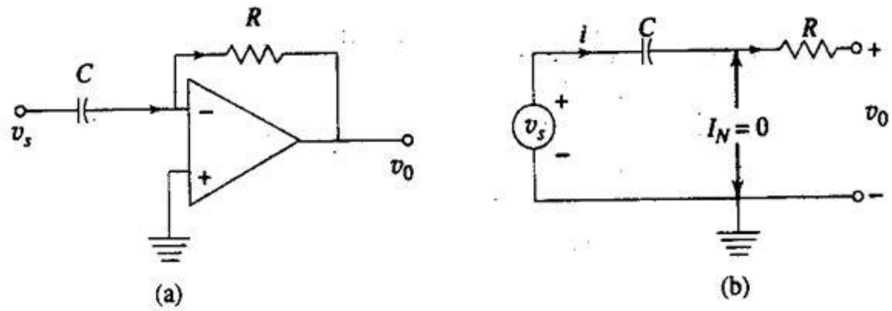


Fig17.4. Differentiating amplifier (a) Circuit (b) Equivalent circuit

Where, $V_{in} = V_s$ is the input voltage.

MODULE 11.

LESSON 18. Boolean Algebra-Fundamental postulate-Demorgan's theorem.

BOOLEAN ALGEBRA AND THEOREMS

Introduction

In 1854, George Boole introduced a systematic treatment of logic and developed for this purpose an algebraic system now called Boolean Algebra. In 1938 C.E. Shannon introduced a two-valued Boolean algebra called switching algebra. Boolean algebra is a system of mathematical logic. It differs from both ordinary algebra and the binary number system. As an illustration, in Boolean, 1 + 1 = 1, in binary arithmetic the result is 10. Thus although there are similarities, Boolean algebra is a unique system.

- The symbol which represent an arbitrary elements of an Boolean algebra is known as variable. Any single variable or a function of several variables can have either a 1 or 0 value. For example, in expression Y = A + BC, variables A, B and C can have either a 1 or 0 value, and function Y also can have either a 1 or 0 value; however its value depends on the value of Boolean expression.
- A complement of a variable is represented by a "bar" over the letter. For example, the complement of a variable A will be denoted by

$\overline{A}$. So if A = 1,

$\overline{A} = 0$ and if A =0, $\overline{A} = 1$. Sometimes a prime symbol (') is used to denote the complement. For example, the complement of A can be written as A'.

- The logical AND operator of two variables is represented either by writing a dot (·) between two variables, such as A · B or by simply writing two variables, such as AB. Similarly, AND operation between three variables can be represented as A·B·C or ABC.
- The logical OR operator of two variables is represented by writing a '+' sign between the two variables such as A+B. Similarly, OR operation between three variables can be represented as A+B+C.

Fundamental Postulates of Boolean Algebra

The postulates of a mathematical system from the basic assumption from which it is possible to deduce the theorems, laws and properties of the system. Boolean algebra is formulated by a defined set of elements, together with two binary operators, + and ·, provided that the following postulates are satisfied.

- Closure (a) : Closure with respect to the operator +

When two binary elements are operated by operator + the result is a unique binary element.

- Closure (b) : Closure with respect to the operator · (dot)

When two binary elements are operated by operator · (dot) the result is a unique binary element.

- An identity element with respect to +, designated by 0:

A+ 0 = 0 + A = A

- An identify element with respect to (·) , designated by 1: A · 1 = 1 · A = A
- Commutative with respect to + : A + B = B + A
- Commutative with respect to · : A · B = B · A
- Distributive property of · over + :

A · (B + C) = (A ·B) + (A ·C)

- Distributive property of + over · :

A + (B · C) = (A + B) · (A + C)

- For every binary element, there exists complement element. For example, if A is an element, we have A is a complement of A. i.e., if A = 0, A = 1 and if

$A = 1, \overline{A} = 0.$

- There exists at least two elements, say A and B in the set of binary elements such that A ≠ B.

From the above discussion we can summarize the postulates of Boolean algebra as shown in the following table 18.1

No.	Postulates	Comment
1	Result of each operation is either 0 or 1	1, 0 ε B
2	a) 0 + 0 = + 0 +1 = 1 + 0 =1 b) 1 · 1 = 1 1 · 0 = 0 · 1 = 0	Identify elements 0 for + and 1 for ·
3	a) (A+B) = (B+A) b) (A · B) = (B · A)	Commutative law
4	a) A · (B + C) = (A · B) + (A · C) b) A + (B · C) = (A + B) · (A + C)	Distributed law
5	a) $A + \overline{A} = 1$ since $0 + \overline{0} = 0 + 1 = 1$ and $1 + \overline{1} = 1 + 0 = 1$ b) $A \cdot \overline{A} = 0$ since $0 \cdot \overline{0} = 0 \cdot 1 = 1$ and $1 \cdot \overline{1} = 1 \cdot 0 = 1$	Complement

Table 18.1 Fundamental postulates of Boolean algebra

Laws of Boolean Algebra

Three of basic laws of Boolean algebra: the commulative laws, associative laws, and the distributive law.

Commulative laws

LAW 1: A + B = B + A : This states that the order in which the variables are ORed makes no difference in the output. The truth tables are identical. Therefore, A OR B is same as B OR A.

A	B	A + B
0	0	0
0	1	1
1	0	1
1	1	1

≡

B	A	B + A
0	0	0
0	1	1
1	0	1
1	1	1

Table 18.2 Truth table for commutative law for OR gates

LAW 2: AB = BA :

The commutative law of multiplication states that the order in which the variables are AND ed makes no difference in the output. The truth tables are identical. Therefore, A AND B is same as B AND A.

A	B	A B
0	0	0
0	1	0
1	0	0
1	1	1

≡

B	A	B A
0	0	0
0	1	0
1	0	0
1	1	1

Table 18.2 Truth table for commutative law for AND gates

It is important to note that the commutative laws can be extended to any number of variables. For example, since A + B = B + A, it follows that A + B + C = B + A + C, and since A + C = C + A, it is true that B + A + C = B + C + A. Similarly,

ABCD = BACD = BADC = ABDC, and so on.

Associate Laws

Law 1: A + (B +C) = (A + B) + C:

This law states that in the ORing of several variables, the result is the same regardless of the grouping of the variables. For three variables, A OR B ORed with C is the same as A ORed with B OR C.

A	B	C	A + B	(A + B) + C
0	0	0	0	0
0	0	1	0	1
0	1	0	1	1
0	1	1	1	1
1	0	0	1	1
1	0	1	1	1
1	1	0	1	1
1	1	1	1	1

≡

A	B	C	B + C	A + (B + C)
0	0	0	0	0
0	0	1	1	1
0	1	0	1	1
0	1	1	1	1
1	0	0	0	1
1	0	1	1	1
1	1	0	1	1
1	1	1	1	1

Truth Table for associative law for OR gates

Law 2 : (AB) C = A (BC):

The associative law of multiplication states that it makes no difference in what order the variables are grouped when ANDing several variables. For three variables, A AND B ANDed with C is the same as A ANDed with B and C.

A	B	C	AB	(AB) C
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	0	0
1	0	0	0	0
1	0	1	0	0
1	1	0	1	0
1	1	1	1	1

≡

A	B	C	BC	A (BC)
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	1	0
1	0	0	0	0
1	0	1	0	0
1	1	0	0	0
1	1	1	1	1

Truth Table for associative law for AND gates

Distributive Law:

Law: A (B+C) = AB + AC :

The distributive law states that ORing several variables and ANDing the result with a single variable is equivalent to ANDing the result with a single variable with each of the several variables and then ORing the products.

A	B	C	(B+C)	A (B+C)
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	1	0
1	0	0	0	0
1	0	1	1	1
1	1	0	1	0
1	1	1	1	1

≡

A	B	C	AB	AC	AB+AC
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	0
0	1	1	0	0	0
1	0	0	0	0	0
1	0	1	0	1	1
1	1	0	1	0	1
1	1	1	1	1	1

It is important to note that the distributive property is often used in reverse; i.e., given AB + AC, we replace it by its equivalent, A (B+C). As in ordinary algebra, this process is called factoring. We factored A out of the expression AB + AC.

DeMorgan's Theorems

DeMorgan suggested two theorems that form an important part of Boolean algebra. In the equation form, they are:

1) $\overline{AB} = \overline{A} + \overline{B}$

The complement of a product is equal to the sum of complements. This is illustrated by following truth table.

TRUTH TABLE

A	B	$\overline{AB}$	$\overline{A} + \overline{B}$
0	0	1	1
0	1	1	1
1	0	1	1
1	1	0	0

2) $\overline{A + B} = \overline{A}\ \overline{B}$

The component of a sum is equal to the product of the complements. The truth Table illustrates this law.

TRUTH TABLE

A	B	$\overline{A + B}$	$\overline{A}\ \overline{B}$
0	0	1	1
0	1	1	1
1	0	0	0
1	1	0	0

MODULE 12.

LESSON 19. Sequential circuit-Flip flops-types- SR flip flop-JK flip flop-T flip flop.

FLIP FLOPS

Flip-flops are the basic elements for storing information. One latch or flip-flop can store one bit of information. The main difference between latches and flip-flops is that for latches, their outputs are constantly affected by their inputs as long as the enable signal is asserted. In other words, when they are enabled, their content changes immediately when their inputs change. Flip-flops, on the other hand, have their content change only either at the rising or falling edge of the enable signal. This enable signal is usually the controlling clock signal. After the rising or falling edge of the clock, the flip-flop content remains constant even if the input changes.

Flip-flops are the fundamental element of sequential circuits and they are bistable whereas gates are the fundamental element for combinational circuits. Flip-flops are essentially 1-bit storage devices and outputs can be set to store either 0 or 1 depending on the inputs even when the inputs are de-asserted, the outputs retain their prescribed value. Flip-flops have (normally) 2 complimentary outputs and three main types of flip-flop are viz., R-S J-K D-type

SR Latch

The bistable element is able to remember or store one bit of information. However, because it does not have any inputs, we cannot change the information bit that is stored in it. In order to change the information bit, we need to add inputs to the circuit. The simplest way to add inputs is to replace the two inverters with two NAND gates as shown in Figure 4(a). This circuit is called a SR latch. In addition to the two outputs Q and Q', there are two inputs S' and R' for set and reset respectively. Following the convention, the prime in S and R denotes that these inputs are active low. The SR latch can be in one of two states: a set state when Q = 1, or a reset state when Q = 0.

If both S' and R' are asserted, then both Q and Q' are equal to 1 as shown at time t4. If one of the input signals is de-asserted earlier than the other, the latch will end up in the state forced by the signal that was de-asserted later as shown at time t5. At t5, R' is de-asserted first, so the latch goes into the normal set state with Q = 1 and Q' = 0. A problem exists if both S' and R' are de-asserted at exactly the same time as shown at time t6. If both gates have exactly the same delay then they will both output a 0 at exactly the same time. Feeding the zeros back to the gate input will produce a 1, again at exactly the same time, which again will produce a 0, and so on and on. This oscillating behavior, called the critical race, will continue forever.

If the two gates do not have exactly the same delay then the situation is similar to de-asserting one input before the other, and so the latch will go into one state or the other. However, since we do not know which is the faster gate, therefore, we do not know which state the latch will go into. Thus, the latch’s next state is undefined.

In order to avoid this indeterministic behavior, we must make sure that the two inputs are never de-asserted at the same time. Note that both of them can be de-asserted, but just not at the same time. Inpractice, this is guaranteed by not having both of them asserted. Another reason why we do not want both inputs to be asserted is that when they are both asserted, Q is equal to Q', but we usually want Q to be the inverse of Q'.

Bistable Element

Simplest sequential circuit or storage element is a bistable element, which is constructed with two inverters connected sequentially in a loop as shown in Figure 1. It has no inputs and two outputs labeled Q and Q′. Since the circuit has no inputs, we cannot change the values of Q and Q′. However, Q will take on whatever value it happens to be when the circuit is first powered up. Assume that Q = 0 when we switch on the power. Since Q is also the input to the bottom inverter, Q′, therefore, is a 1. A 1 going to the input of the top inverter will produce a 0 at the output Q, which is what we started off with. Similarly, if we start the circuit with Q = 1, we will get Q′ = 0, and again we get a stable situation. A bistable element has memory in the sense that it can remember the content (or state) of the circuit indefinitely. Using the signal Q as the state variable to describe the state of the circuit, we can say that the circuit has two stable states:

Q = 0, and Q = 1; hence the name "bistable."

An analog analysis of a bistable element, however, reveals that it has three equilibrium points and not two as found from the digital analysis. Assuming again that Q = 1, and we plot the output voltage (Vout1) versus the input voltage (Vin1) of the top inverter.The dotted line shows the operation of the bottom inverter where Vout2 and Vin2 are the output and input voltages respectively for that inverter. Figure 2 shows that there are three intersection points, two of which corresponds to the two stable states of the circuit where Q is either 0 or 1. The third intersection point labeled metastable, is at a voltage that is neither a logical 1 nor a logical 0 voltage. Nevertheless, if we can get the circuit to operate at this voltage, then it can stay at that point indefinitely. Practically, however, we can never operate a circuit at precisely a certain voltage. A slight deviation from the metastable point as cause by noise in the circuit or other stimulants will cause the circuit to go to one of the two stable points. Once at the stable point, a slight deviation, however, will not cause the circuit to go away from the stable point but rather back towards the stable point because of the feedback effect of the circuit.

Flip-Flop Types

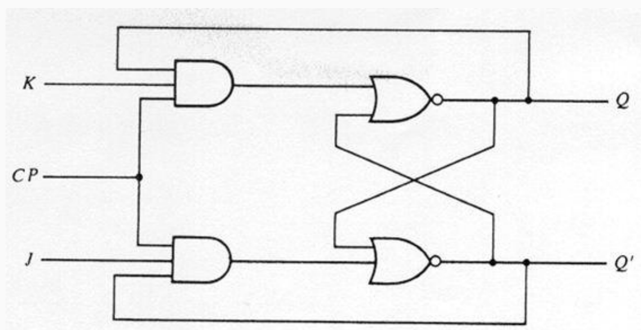
There are basically four main types of flip-flops: SR, D, JK, and T. The major differences in these flip-flop types are in the number of inputs they have and how they change state. Each type can have different variations such as active high or low inputs, whether they change state at the rising or falling edge of the clock signal, and whether they have asynchronous inputs or not. The flip-flops can be described fully and uniquely by its logic symbol, characteristic table, characteristic equation, state diagram, or excitation table, and are summarized.

SR Flip-Flop

We can replace the D latches in the D flip-flop of Figure 10(a) with SR latches to get a master-slave SR flipflop shown in Figure 16. Like SR latches, SR flip-flops are useful in control applications where we want to be able to set or reset the data bit. However, unlike SR latches, SR flip-flops change their content only at the active edge of the clock signal. Similar to SR latches, SR flip-flops can enter an undefined state when both inputs are asserted simultaneously.

JK Flip-Flop

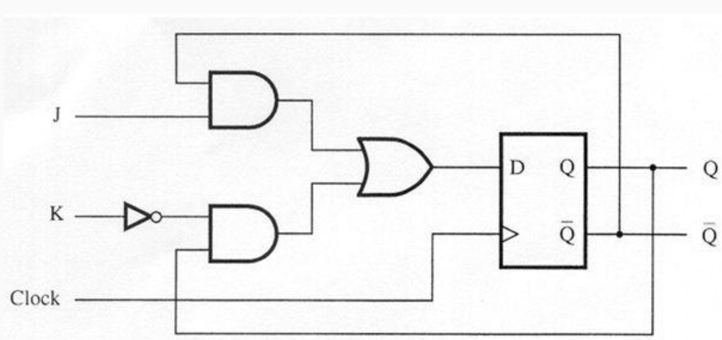
JK flip-flops are very similar to SR flip-flops. The J input is just like the S input in that when asserted, it sets the flip-flop. Similarly, the K input is like the R input where it clears the flip-flop when asserted.The only difference is when both inputs are asserted. For the SR flip-flop, the next state is undefined, whereas, for the JK flip-flop, the next state is the inverse of the current state.



In other words, the JK flip-flop toggles its state when both inputs are asserted. The circuit, truth table and the logic symbol for the JK flip-flop.

T Flip-Flop

The T flip-flop has one input in addition to the clock. T stands for toggle for the obvious reason. When T is asserted (T = 1), the flip-flop state toggles back and forth, and when T is de-asserted,the flip-flop keeps its current state. The T flip-flop can be constructed using a D flip-flop with the two outputs Q and Q' feedback to the D input through a multiplexer that is controlled by the T input.



Logic Symbol

The logic or graphical symbol describes the flip-flop's inputs and outputs, the names given to these signals, and whether they are active high or low. All the flip-flops have Q and Q' as their outputs. All of them also have a CLK input. The small triangle at the clock input indicates that the circuit is a flip-flop and so it is triggered by the edge of the clock signal; if there is a circle in front, then it is the falling edge, otherwise, it is the rising edge of the clock.

Characteristic Table

The characteristic table is just the truth table but usually written in a shorter format. For example, compare the characteristic table for the JK flip-flop in Figure 20 with the truth table in Figure 17(b). The truth table, as we have seen, simply lists all possible combinations of the input signals, the current state (or content) of the flip-flop, and the next state that the flip-flop will go to at the next active edge of the clock signal. The characteristic table answers the question of what is the next state when given the inputs and the current state, and is used in the analysis of sequential circuits.

Characteristic Equation

The characteristic equation is the functional Boolean equation that is derived from the characteristic table. This equation formally describes the functional behavior of the flip-flop. Like the characteristic table, it specifies the flipflop's next state as a function of its current state and inputs. For example, the characteristic equation for the JK flipflop can be derived from the truth table as follows:

$$\begin{aligned} Qnext &= J'K'Q + JK'Q + JK'Q' + JKQ' \\ &= K'Q(J'+J) + JQ'(K'+K) \\ &= K'Q + JQ' \end{aligned}$$

The characteristic equation can also be obtained from the truth table using the K-map method as follows for the SR flip-flop.

State Diagram

A state diagram is a graph that shows the flip-flop's operations in terms of how it transitions from one state to another. The nodes are labeled with the states and the directed arcs are labeled with the input signals that cause the transition to go from one state to the next. Figure 21 shows the state diagram for the SR flip-flop. For example, to go from state Q = 0 to the state Q = 1, the two inputs S and R have to be 1 and 0 respectively. Similarly, if the current state is Q = 0 and we want to remain in that state, then SR need to be 00 or 01.

LESSON 20. Counters-Asynchronous and synchronous counter-decade counter-up down counter- ring and Johnson counter.

COUNTER

Electronic counters:

In electronics, counters can be implemented quite easily using register-type circuits such as the flip-flop, and a wide variety of classifications exist:

- Asynchronous (ripple) counter – changing state bits are used as clocks to subsequent state flip-flops
- Synchronous counter – all state bits change under control of a single clock
- Decade counter – counts through ten states per stage
- Up/down counter – counts both up and down, under command of a control input
- Ring counter – formed by a shift register with feedback connection in a ring
- Johnson counter – a twisted ring counter
- Cascaded counter

Each is useful for different applications. Usually, counter circuits are digital in nature, and count in natural binary. Many types of counter circuits are available as digital building blocks, for example a number of chips in the 4000 series implement different counters.

Occasionally there are advantages to using a counting sequence other than the natural binary sequence such as the binary coded decimal counter, a linear feedback shift register counter, or a Gray-code counter.

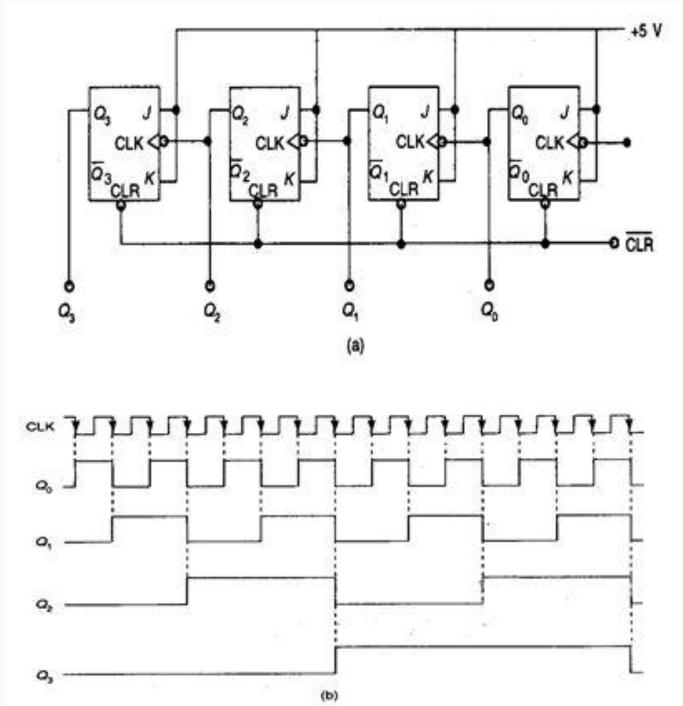
Counters are useful for digital clocks and timers, and in oven timers, VCR clocks, etc.

Counters

The function of a counter is to count the number of clock pulses, which have arrived at the clock input. A counter consists of a number of flip-flops. Counters are classified as ripple counter (or asynchronous counter) and synchronous counter.

Ripple Counter

Figure shows the circuit of a 4-bit ripple counter consisting of 4 –edge triggered JK flip-flop. As indicated by small circles at the CLK input of flip-flop, the triggering occurs when CLK input gets a negative edge. Q is the least significant bit (LSB) and Q is the most significant bit (MSB). The flip-flops are connected in series. The Q output is connected to CLK terminal of second flip-flop.

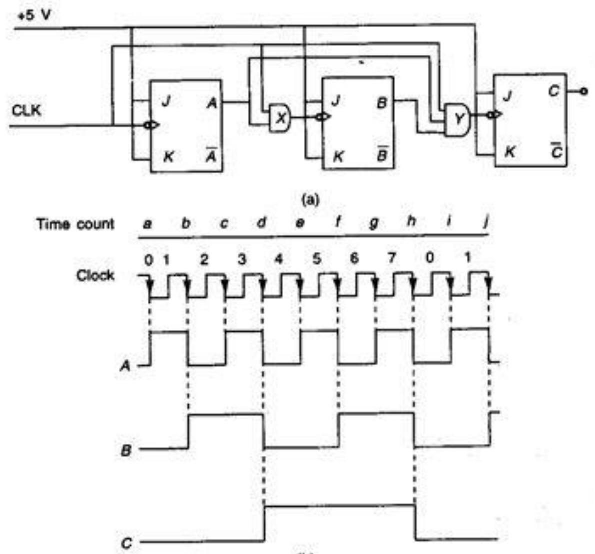


The Q output is connected to CLK terminal of third flip-flop and so on. By adding more flip-flops a counter of any length can be built. It is know as a ripple counter because the carry moves through the flip-flops like a ripple on water. Initially, CLR is made low and all flip-flops reset giving an output Q =0000. When CLR becomes high, the counter is ready to start. As LSB receives its clock pulse, its output changes from 0 to1 and the total ouptu Q = 0001. When second clock pulse arrives, Q resets and carries (i.e., Q goes from 1 to 0 and the second flip flop will receive CLKK input). Now the output is Q = 0010. The third CLK pulse changes Q to 1 giving a total output Q = 0100 and the process goes on. The number of output states of a counter is known as modulus (or mod). A ripple counter with 4 flip-flops can count from 0-15 and is , therefore, known as mod-16 counter while one with 6 flip-flops can count from 0 to 63 and is a mod-64 counter and so on.

Ripple counters are simple to fabricate but have the problem that the carry has to propagate through a number of flip-flops. The delay; time of all the flip-flops are added. Therefore, they are very slow for some applications. Another problem is that unwanted pulses occur at the output of gates.

Synchronous Counter

In a synchronous counter, all the flip-flops are clocked together. Figure 8.36 shows a synchronous counter having positive edge triggered JK flip-flops. Since all the flip-flops are clocked together, the delay time is less. The flip-flop corresponding to least significant bit (LSB) has its inputs J K fed from voltage + V. Therefore, it responds to each positive clock edge. However, the other three flip-flops can respond to the positive clock pulse.



(a)

(b)

C	B	A	Count
0	0	0	0
0	0	1	1
0	1	0	2
0	1	1	3
1	0	0	4
1	0	1	5
1	1	0	6
1	1	1	7
0	0	0	0

under some certain conditions. The Q flip-flop toggles on positive clock edge only if Q is 1. The Q flip-flop toggles on positive clock edge only when Q and Q are 1 (due to presence of AND circuit) and so on.

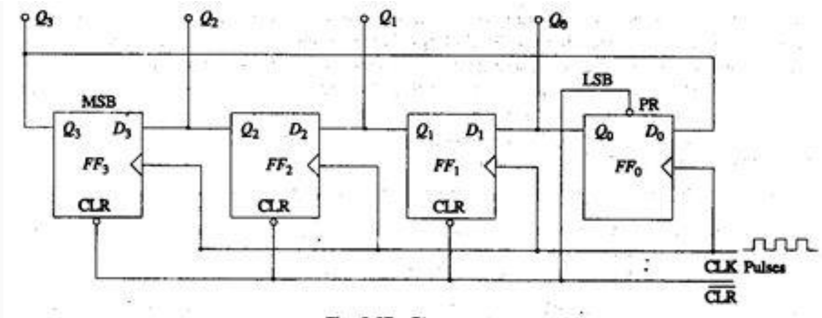
Thus, each flip-flop toggles on the next positive clock edge if all lower bits are 1.

A low (CLR) signal resets the counter so that Q=0000. When CLR goes high, the counter is ready to start. The first positive clock edge sets Q to 1 so that Q=0001. At second positive clock edge Q and Q toggle and Q=0010. The third positive clock edge increases the count by 1 so that Q =0011.

The successive Q outputs are 0100, 0101 and so on upto 1111 (i.e., decimal 15). The next positive clock edge resets the counter to 000 and the cycle is repeated. More flip-flops can be added to increase the count.

RING COUNTER

The following figure is the circuit of a ring counter. It uses D flip-flops. The output Q sets D input, Q sets D, Q sets D and Q is fed back to D. Because of these connections, bits are shifted left one position per positive clock edge and fed back to the input. As seen, all the flip-flops are clocked together.



When CLR goes low then back to high, the output is Q=0000. The first positive clock edge shifts MSB to LSB position and other bits to one position left so that the output becomes Q = 0010. This process continues on second and third positive clock edge so that successive outputs are 0100 and 1000. The fourth positive clock edge starts the cycle all over again and output Q=0001. Thus, the stored 1 bits follow a circular path (i.e., the stored 1 bits move left through all flip-flops and the final flip flop sends it back to the first flip-flop. This action has given it the name of ring counter. It is used to control a sequence of operations in a sequence, control stepper motors

USES:

The use of flip-flop outputs as clocks leads to timing skew between the count data bits, making this ripple technique incompatible with normal synchronous circuit design styles.

Synchronous counter:

A 4-bit synchronous counter using JK flip-flops. A simple way of implementing the logic for each bit of an ascending counter (which is what is depicted in the image to the right) is for each bit to toggle when all of the less significant bits are at a logic high state. For example, bit 1 toggles when bit 0 is logic high; bit 2 toggles when both bit 1 and bit 0 are logic high; bit 3 toggles when bit 2, bit 1 and bit 0 are all high; and so on.Synchronous counters can also be implemented with hardware finite state machines, which are more complex but allow for smoother, more stable transitions. Hardware-based counters are of this type.

DECADE COUNTER:

A decade counter is one that counts in decimal digits, rather than binary. A decade counter may have each digit binary encoded (that is, it may count in binary-coded decimal, as the 7490 integrated circuit did) or other binary encodings (such as the bi-quinary encoding of the 7490 integrated circuit).

Alternatively, it may have a "fully decoded" or one-hot output code in which each output goes high in turn (the 4017 is such a circuit). The latter type of circuit finds applications in multiplexers and demultiplexers, or wherever a scanning type of behavior is useful. Similar counters with different numbers of outputs are also common.

The decade counter is also known as a mod-counter when it counts to ten (0, 1, 2, 3, 4, 5, 6, 7, 8, 9). A Mod Counter that counts to 64 stops at 63 because 0 counts as a valid digit.

UP/DOWN COUNTER:

A counter that can change state in either direction, under the control of an up/down selector input, is known as an up/down counter. When the selector is in the up state, the counter increments its value. When the selector is in the down state, the counter decrements the count.

RING COUNTER:

A ring counter is a circular shift register which is initiated such that only one of its flip-flops is the state one while others are in their zero states.

A ring counter is a Shift Register (a cascade connection of flip-flops) with the output of the last one connected to the input of the first, that is, in a ring. Typically, a pattern consisting of a single bit is circulated so the state repeats every n clock cycles if n flip-flops are used. It can be used as a cycle counter of n states.

Johnson counter:

A Johnson counter (or switchtail ring counter, twisted-ring counter, walking-ring counter, or Moebius counter) is a modified ring counter, where the output from the last stage is inverted and fed back as input to the first stage.[2][3][4] The register cycles through a sequence of bit-patterns, whose length is equal to twice the length of the shift register, continuing indefinitely.

These counters find specialist applications, including those similar to the decade counter, digital-to-analog conversion, etc. They can be implemented easily using D- or JK-type flip-flops.

Computer science counters

In computability theory, a counter is considered a type of memory. A counter stores a single natural number (initially zero) and can be arbitrarily many digits long.

A counter is usually considered in conjunction with a finite-state machine (FSM), which can perform the following operations on the counter:

- Check whether the counter is zero
- Increment the counter by one.
- Decrement the counter by one (if it's already zero, this leaves it unchanged).

The following machines are listed in order of power, with each one being strictly more powerful than the one below it:

1. Deterministic or non-deterministic FSM plus two counters
2. Non-deterministic FSM plus one stack

3. Non-deterministic FSM plus one counter
4. Deterministic FSM plus one counter
5. Deterministic or non-deterministic FSM

For the first and last, it doesn't matter whether the FSM is a deterministic finite automaton or a nondeterministic finite automaton. They have equivalent power. The first two and the last one are levels of the Chomsky hierarchy.

MODULE 13.

LESSON 21. Digital to Analog converter- transfer characteristics- conversion techniques. Binary weighted resister DAC-drawback.

Digital to Analog Converter (DAC) and Analog to Digital Converter (ADC)

Introduction

The digital to analog conversion and analog to digital conversion techniques are important tools in the digital data processing systems. The digital to analog conversion technique involves conversion of digital information into equivalent analog information. Digital to analog converts acts as a decoding device since it operates on the output of a digital system. Using analog to digital conversion techniques, the analog information is converted into its equivalent binary number which is in the digital form. In general, the analog to digital converter acts as an encoder. In this chapter, we will discuss various DAC and ADC techniques in details. The discussion of the performance parameters of ADC and DAC is included in the chapter. As the digital to analog conversion is the integral part of an analog to digital conversion, let us discuss first the DAC.

Digital to Analog Converter (DAC)

It is clear from name itself that digital to analog converters are the circuits which convert digital signals into analog converters are very important components inside analog to digital converters. The digital to analog converters (DACs) are also used in many other applications such as cathod ray tube display systems, voice synthesizers, automatic test systems and process control actuators. It is the most important component in computers.

The DAC can perform its conversion in either serial or parallel form depending upon the application in which it is used. For example, process control applications such as bottle filling systems require slow or serial operation, while modern instrumentation such as digital storage oscilloscopes and military weapon systems required very fast or parallel operation.

In general, the input quantity to the DAC is a digital number, the conversion techniques convert the number into corresponding number of units of current, voltage or charge and then these units are added together with an analog summing circuits.

A basic block diagram of commercial DAC is shown in the fig.

A basic block diagram of commercial DAC is as shown in the Fig. 8.1.

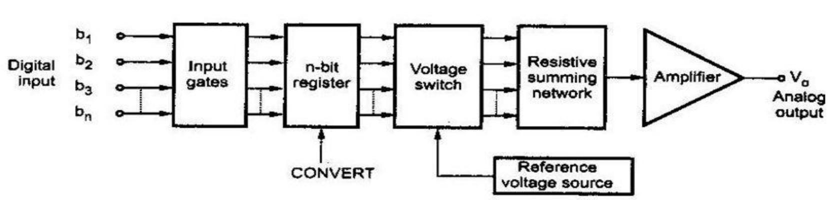


Fig. 8.1 Basic block diagram of commercial DAC

The n-bit digital input number which is to be converted is fed to the n-bit register through the input gates of DAC, upon the execution of CONVERT command. Here b_1 to b_n are binary digits; either 0 or 1. Note that the register accepts only a digital input during the CONVERT command. This register holds this inputed digital number constant till next CONVERT command is received. The outputs of the register are fed to the voltage switches that provide two possible outputs either 0 V or voltage equal to the value of precision reference voltage source. Thus these switches are exactly similar to the single pole double through (SPDT) switches. The voltage switches are realized in the form of transistor switches. These switches provide access to the resistive summing network which converts each bit into its weighted current value and then adds to get a total current. The total value is then fed to an amplifier which performs current to voltage conversion and output scaling.

In general, a DAC accepts an n – bit digital input word say b_1 , b_2 , b_3 ,…… b_n in binary format and produces an analog output signal V_{out} proportional to input digital signal B_{in}. B_{in} is defined as n-bit digital word given by

$$B_{in} = b_1 2^{-1} + b_2 2^{-2} + b_3 2^{-3} + \ldots\ldots\ldots + b_n 2^{-n} \qquad \ldots.(1)$$

Note that each digital input can be represented by an electrical signal representing either logic 1 and logic 0. Also b_1 is defined as most significant bit (MSB), while b_n is defined as least significant bit (LSB). A basic circuit symbol for n-bit D/A converter is as shown in fig.

The relationship between digital input word B_{in}, analog output V_o and reference voltage V_R is given by,

$$V_o = V_R \,.\, B_{in} = V_R \; (b_1 2^{-1} + b_2 2^{-2} + b_3 2^{-3} + \ldots\ldots\ldots + b_n 2^{-n}) \qquad \ldots..(2)$$

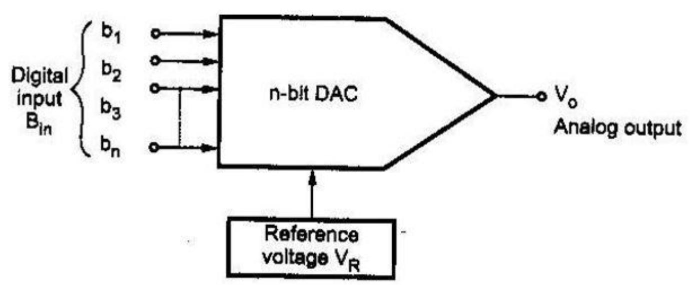


Fig. 8.2 Basic circuit symbol for n-bit DAC

Transfer Characteristic of DAC

The transfer characteristics of 4-bit DAC is shown in fig. In the transfer characteristic, analog output voltage V_o is plotted against all 16 possible inputs.

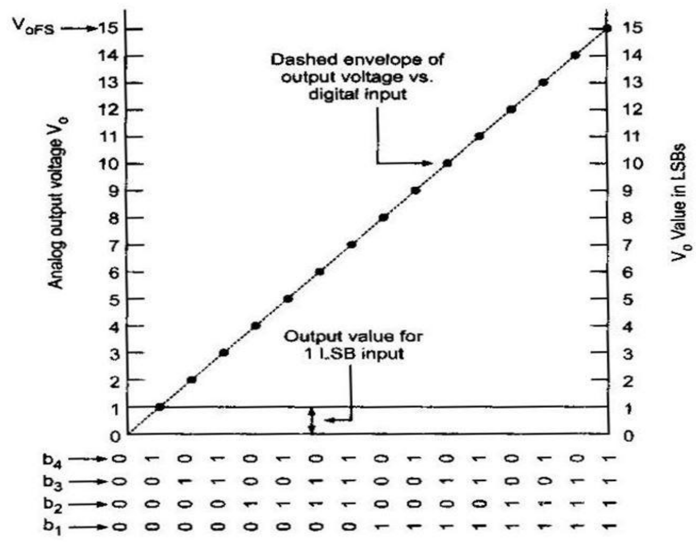


Fig. 8.3 Transfer characteristics of 4-bit DAC

Digital to Analog Conversion Techniques

There are mainly two techniques used for digital to analog conversion.

- Binary weighted resistor DAC
- R/2R ladder DAC

In these techniques, the shunt resistors are used to generate n binary weighted currents. These currents are added according to switch positions controlled by the digital input and then converted into voltage to give analog voltage equivalent to the digital input. Therefore, such digital to analog converters are called current driven DACs.

Binary Weighted Resistor DAC

The binary weighted resistor DAC uses an op-amp to sum n binary weighted currents derived from a reference voltage V_R via current scaling resistors 2R, 4R, 8R, ….2^n R, as shown in fig.

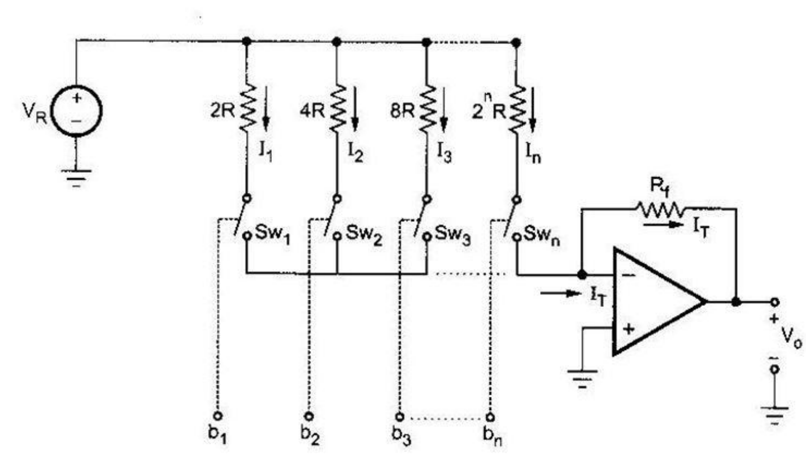


Fig. 8.4 Binary weighted resistor DAC

As shown in fig. switch positions are controlled by the digital inputs. When digital input is logic 1, it connects the corresponding resistance to the reference voltage V_R; otherwise it leaves resistor open. Therefore,

For ON-Switch, $I = \dfrac{V_R}{R}$ and

For OFF – Switch $I = 0$

Here operational amplifier is used as a summing amplifier. Due to high input impedance of op-amp, summing current will flow through R_f. Hence the total current through R_f can be given as

$$I_T = I_1 + I_2 + I_3 + \ldots\ldots I_n$$

The output voltage is the voltage across R_f and it is given as

$$\begin{aligned} V_o &= -I_T R_f = -(I_1 + I_2 + I_3 + \cdots + I_n) R_f \\ &= -\left(b_1 \frac{V_R}{2R} + b_2 \frac{V_R}{4R} + b_3 \frac{V_R}{8R} + \cdots + b_n \frac{V_R}{2^n R}\right) R_f \\ &= -\frac{V_R}{R} R_f \left(b_1 2^{-1} + b_2 2^{-2} + b_3 2^{-3} + \cdots + b_n 2^{-n}\right) \qquad \ldots (1) \end{aligned}$$

When $R_f = R$, V_o is given as

$$V_o = -V_R \left(b_1 2^{-1} + b_2 2^{-2} + b_3 2^{-3} + \cdots + b_n 2^{-n}\right) \qquad \ldots (2)$$

The equation (1) indicates that the analog output voltage is proportional to the input digital word.

The simplicity of the binary weighted DAC is offset by drawbacks associated with it.

Drawbacks

1. wide range of resistor values are required. For 8-bit DAC, the resistors required are 2^1 R, 2^2 R, 2^3 R , …… and 2^8 R. Therefore the largest resistor is 128 times the smallest one.

2. This wide range of resistor values has restrictions on both, higher and lower ends. It is impracticable to fabricate large values of resistor in IC and voltage drop across such a large resistor due to the bias current also affects the accuracy. For smaller values of resistors, the loading effect may occur.

3. The finite resistance of the switches disturbs the binary-weighted relationship among the various currents, particularly in the most significant bit positions, where the current setting resistance are smaller.

All these drawbacks, especially the requirement of wide range of resistors restricts the use of binary weighted resistor DACs below 8 bits.

LESSON 22. R& 2R ladder DAC-Advantages-performance

R/2R Ladder DAC

In this type, reference voltage is applied to one of the switch positions and other switch position is connected to ground, as shown in the fig.

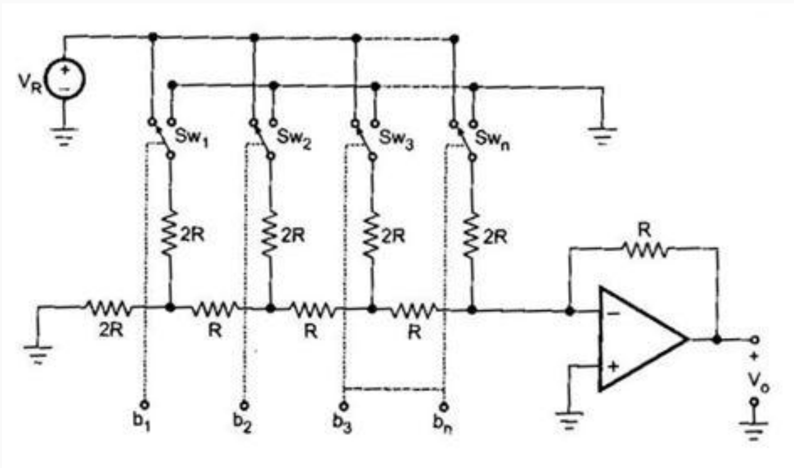
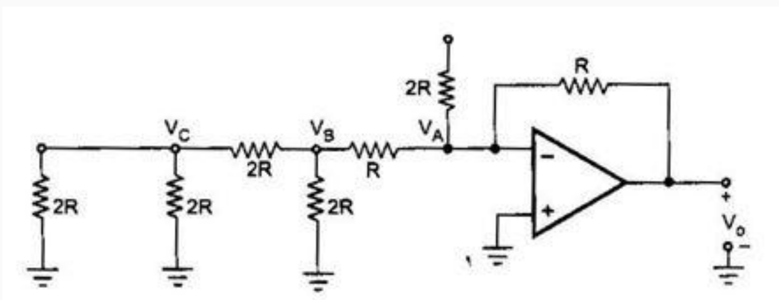


Fig. 8.7 R/2R Ladder DAC



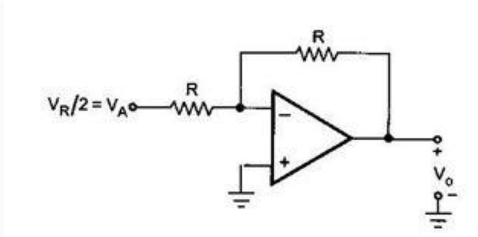
In general, the expression for V_o can be obtained as,

Let

I_{out} = Output current

R_f = feedback resistance of op-amp

@ $V_o = - I_{out} R_f$



Now I_{out} = current resolution x D

@ V_o = - (current resolution x D) R_f

@ V_o = - (current resolution x R_f) D

The coefficient of D is the voltage resolution and can be called as simple resolution.

@ V_o = - resolution x D

In terms of actual circuit elements, output can be written as,

V_o =

The resolution of R/2R ladder type DAC with current output is,

$$\text{resolution} = \frac{1}{2^n} \times \frac{V_R}{R}$$

While the resolution for R/2R ladder type DAC with voltage output is

resolution =

Example:

Suggest the values of resistors and reference voltage if resolution required is 0.5 V for 4 bit r/2R ladder type DAC.

Solution:

resolution

let V_R = 10 V, n = 4 and resolution = 0.5

@ 0.5 =

Advantages of R/2R ladder DACs:

1) Easier to build accurately as only two precision metal film resistors are required.
2) Number of bits can be expanded by adding more sections of same R/2R values.
3) In inverted R/2R ladder DAC, node voltages remain constant with changing input binary words. This avoids any slowdown effects by stray capacitances.

Performance Parameters of DAC

The various performance parameters of DAC are,

Resolution

Resolution is defined in two ways.

- Resolution is the number of different analog output values that can be provided by DAC. For an n-bit DAC.

resolution = 2^n

- Resolution is also defined as the ratio of a change in output voltage resulting from a change of 1 LSB at the digital inputs. For an n-bit DAC it can be given as

$$\text{resolution} = \frac{V_{oFS}}{2^n - 1}$$

Where, V_{oFS} = Full scale output voltage

From equation (2) we can say that, the resolution can be determined by the number of bits in the input binary word. For an 8 – bit DAC resolution can be given as

resolution = $2^n = 2^8 = 256$

If the full scale output voltage is 10.2 V then by second definition the resolution for an 8 –bit DAC can be given as

$$\text{resolution} = \frac{V_{oFS}}{2^n - 1} = \frac{10.2}{2^8 - 1} = \frac{10.2}{255} = 40 \text{ mV/LSB}$$

Therefore, we can say that an output change of 1 LSB causes the output equation for a DAC.

Thus V_o = resolution x D

where D = decimal value of the digital input

and V_o = output voltage

The resolution takes care of changes in the input.

Accuracy

It is a comparison of actual output voltage with expected output. It is expressed in percentage. Ideally the accuracy of DAC should be, at worst, ± ½ of its LSB. If the full scale output voltage is 10.2 V then for an 8-bit DAC accuracy can be given as

$$\text{Accuracy} = \frac{V_{oFS}}{(2^n - 1)\,2}$$

$$\text{Accuracy} = \frac{10.2}{255 \times 2} = 20 \text{ mV}$$

Monotonicity

A converter is said to have good monotonicity if it does not miss any step backward when stepped through its entire range by a counter.

Conversion Time

It is a time required for conversion of analog signal into its digital equivalent. It is also called as setting time. It depends on the response time of the switches and the output of the amplifier.

Setting Time

This is the time required for the output of the DAC to settle to within ± ½ LSB of the final value for a given digital input i.e., zero to full scale.

Stability

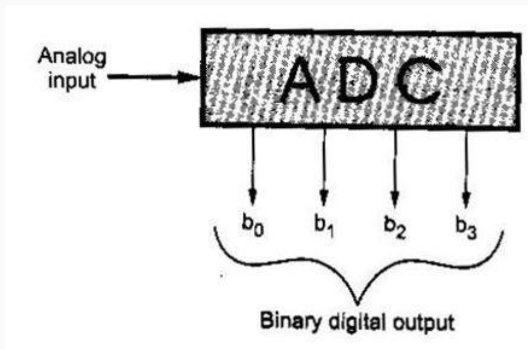
The performance of converter changes with temperature, age and power supply variations. So all the relevant parameters such as offset, gain, linearity error and monotonicity must be specified over the full temperature and power supply ranges. These parameters represent the stability of the converter.

MODULE 14.

LESSON 23. Analog to digital converter-transfer characteristics-conversion techniques-successive approximate ADC.

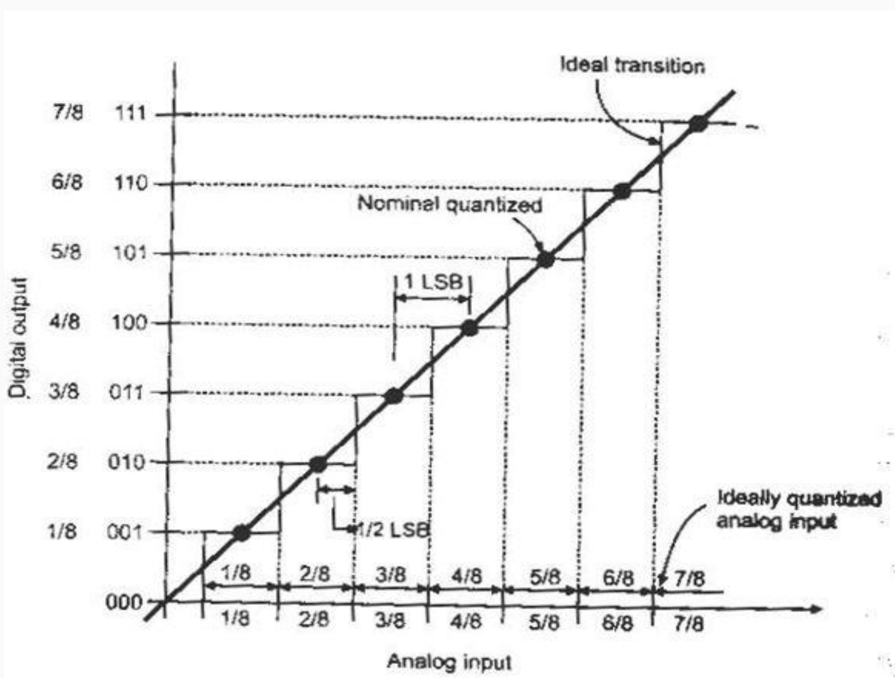
Analog to Digital Converters (ADC)

The analog to digital conversion is a quantizing process where by an analog signal is converted into an equivalent binary word. The fig. shows the symbol for A/D converter.



Transfer Characteristics of ADC

The graph of output plotted against the input is called transfer characteristics. Thus transfer characteristics of ADC is the graph of digital output against the analog input. The fig. shows the transfer characteristics of a 3-bit ADC.



The transfer characteristics shows 2^3 (as 3 bit) discrete output states i.e., from binary 000_2 to 111_2. The suffix 2 indicates binary nature of the output. The each step is being 1/8 V apart.From this transfer characteristics, we can define following performance parameters of ADC.

Resolution

The resolution of ADC is defined as the number of different digital outputs states which can be provided by ADC. Thus for a n bit ADC.

resolution $= 2^{n}$ ….. (1)

Resolution is also called as the ratio of a change in value of input voltage, V_i , needed to change the digital output by 1 LSB. If the full scale input voltage required to cause a digital output of all 1's is V_{iFS} , then resolution can be given as

$$\text{resolution} = \frac{V_{iFS}}{2^n - 1} \qquad \text{.....(2)}$$

Quantization Error

The fig. shows that the binary output is 011 for all values of V_i between ¼ and ½ V. There is an unavoidable uncertainly about the exact value of V_i when the output is 011. This uncertainly is specified as quantitization error. Its value is ± ½ LSB.

It is given as,

$$Q_E = \frac{V_{iFS}}{(2^n - 1)\,2} \qquad \text{...(3)}$$

Increasing the number of bits results in a finer resolution and a smaller quantization error.

Conversion Time

It is an important parameter for ADC. It is defined as the total time required to convert an analog signal into its digital output. It depends on the conversion technique used and the propagation delay of circuit components.

Example:

An 8-bit ADC outputs all 1's when $V_i = 5.1$ V. Find it's a) Resolution and

b) Digital output when $V_i = 1.28$ V.

Solution:

a) From equation (1) we have,

resolution $= 2^8 = 256$

and from equation (2) we have,

$$\text{resolution} = \frac{5.1 \text{ V}}{2^8 - 1}$$

$= 20$ mV/LSB

Therefore, we can say that to change output by 1 LSB we have to change input by 20 mV.

b) For 1.28 V analog input, digital output can be calculated as,

$$D = \frac{1.28 \text{ V}}{20 \text{ m V / LSB}} = 64 \text{ LSBs}$$

The binary equivalent of 64 is $0100\ 0000_2$

Example:

Calculate the quantizing error for 12-bit ADC with full scale input voltage 4.095 V.

Solution

From equation (3) get

$$Q_E = \frac{4.095}{(2^{12} - 1) \times 2}$$

$$= \frac{4.095}{(4096 - 1) \times 2} = 0.5 \text{ mV}$$

Analog to Digital conversion Techniques

Analog to digital converters are classified into two general groups based on the conversion techniques. One technique involves comparing a given analog signal with the internally generated reference voltages. This group includes successive approximation, flash, delta modulated (DM), adaptive delta modulated and flash type converters. The another technique involves changing an analog signal into time or frequency and comparing these new parameters against known values. This group includes integrator converters and voltage-to-frequency converters.

Types of ADCs using various conversion techniques:

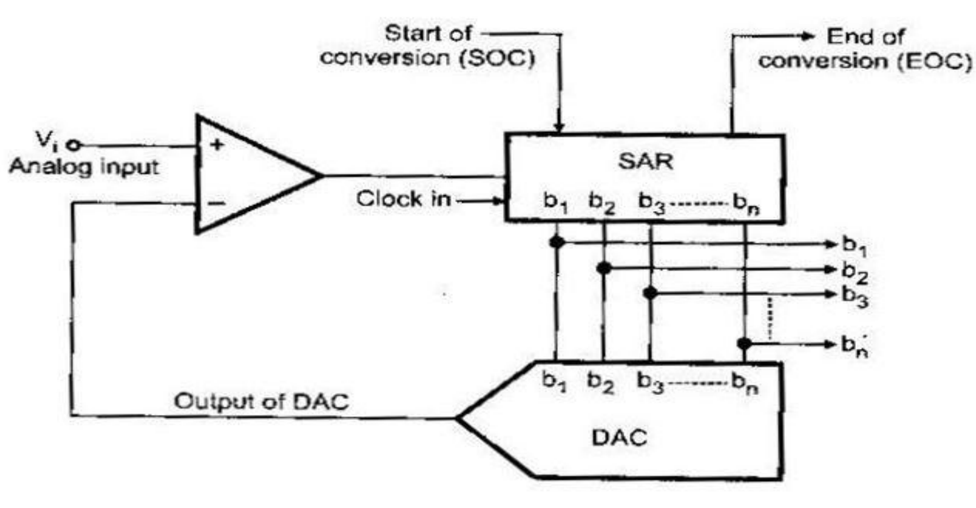
1. Single ramp or single slope
2. Dual slope
3. Successive approximation
4. Flash
5. Delta Modulation
6. Adaptive delta modulation

Successive Approximation ADC

In this technique, the basic idea is to adjust the DAC's input code such that its output us within ± ½ LSB of the analog input V_i to be A/D converted. The code that achieves this represents the desired ADC output.

The successive approximation method uses very efficient code searching strategy called binary search. It completes searching process for n-bit conversion in just n clock periods.

Fig. shows the block diagram of successive approximation ADC. It consists of a DAC, a comparator and a successive approximation register (SAR).

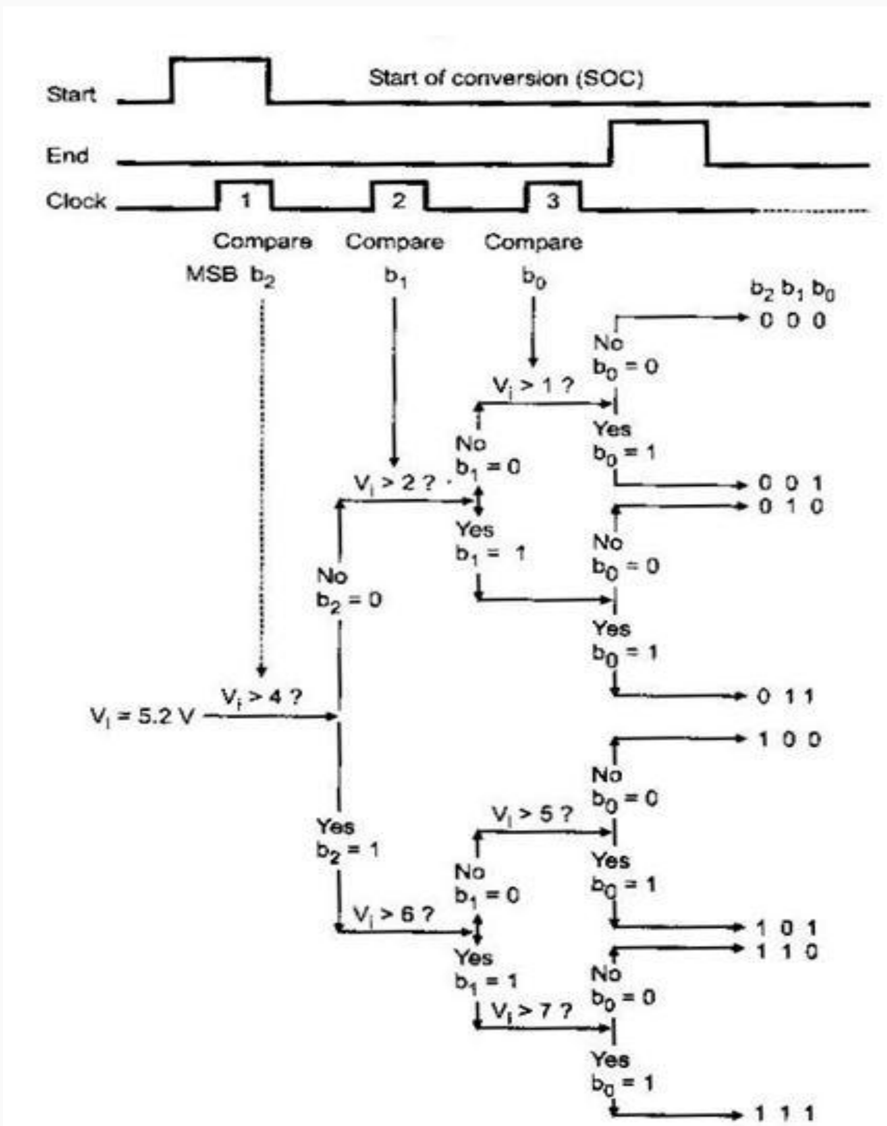


The external clock input sets the internal timing parameters. The control signal start of conversion (SOC) initiates an A/D conversion process and end of conversion signal is activated when the conversion is completed.

Operation:

The searching code process in successive approximation method is similar to weighing an unknown material with a balance scale and a set of standard weights. Let us assume that we have 1 kg, 2 kg and 4 kg weights (SAR) plus a balance scale (comparator and DAC). Now we will see the successive approximation analogy for 3-bit ADC.

The analog voltage V_{in} is applied at one input of comparator. On receiving start of conversion signal (SOC) successive approximation register sets 3-bit binary code 100_2 ($b_2 = 1$) as an input of DAC. This is similar process of placing the unknown weight on one platform of the balance and 4kg weight on the other. The DAC converts the digital word 100 and applies it equivalent analog output at the second input voltage is greater than the analog output of DAC, successive approximation register keeps $b_2 = 1$ and makes $b_1 = 1$ (addition of 2 kg weight to have total 6 kg weight) otherwise it resets $b_2 = 0$ and makes $b_1 = 1$ (replacing 2 kg weight). The same process is replaced for b_1 and b_0. The status of b_0, b_1 and b_2 bits gives the digital equivalent of the analog input.



The dark lines in the fig. shows setting and resetting actions of bits for input voltage 5.2 V, on the basis of comparison. It can be seen from the fig. that one clock pulse is required for the successive approximation register to compare each bit. However an additional clock pulse is usually required to reset the register prior to performing a conversion. The time for one analog to digital conversion must depend on both the clock's period T and number of bits n. It is given as,

$$T_C = T\,(n+1) \qquad \ldots(1)$$

Where T_C = conversion time

T = clock period

n = number of bits

Example:

An 8-bit successive approximation ADC is driven by a 1 MHz clock. Find its conversion time.

Solution:

$$f = 1\text{ MHz}$$

$$@ \quad T = \frac{1}{f} = \frac{1}{1 \times 10^{6}} = 1\ \mu\text{sec}$$

$$n = 8$$

$$@T_C = T\,(n+1) = 1\,(8+1) = 9\ \mu\text{sec}$$

LESSON 24. Measurement-methods-Instrument and classification of instruments-static and dynamic characteristics of instrument.General measurement system-Functional elements of a measurement system-

Measurements

Measurement of a given quantity is essentially an act or result of comparison between the quantity (whose magnitude is unknown) and a predefined standard.

Two basic requirements

- Standard used for comparison should be accurately defined.
- The apparatus used and the method adopted must be provable.

Methods of measurement

• Direct methods – unknown quantity directly compared against a standard.

• Indirect methods –direct measurement is not always possible, they are inaccurate and less sensitive. Hence this method is followed.

Instrument

An instrument may be defined as a device for determining the value or magnitude of a quantity or variable.

Development phases of an instrument

1) Mechanical instrument

2) Electrical instrument

3) Electronic instrument.

Classification of instruments

- Absolute instruments – these give magnitude of the quantity under measurements in terms of physical constants of the instrument.
- Secondary instruments –quantity being measured can only be measured by observing the output indicated by the instrument calibrated by comparison with an absolute instrument /secondary instrument.

Generalized input-output configuration

In the design and or use of measuring instruments a number of methods for nullifying or reducing these interfering and modifying inputs are available (For eg. Use of (i) feed back / closed lap system reduces these errors (ii) filtering also reduces these errors.

Static characteristics

Applications involve the measurement of quantities that are constant or vary only quite slowly. Under these conditions, it is possible to define a set of performance criteria that give a meaningful description of the quality of measurement without becoming concerned with dynamic description.

Dynamic characteristics

Dynamic relations between the instrument input and output must be examined, generally by the use of differential equation. Performance criteria based on these dynamic relations constitute the dynamic characteristics.

Analog and digital modes of operation

Signals that vary in a continuous fashion and take on infinity of values in any given range are called analog signals.

Signals which vary in discrete steps and thus take up only finite different values in a given range are called digital signals.

Readability: indicates the closeness with which the scale of the instrument may read.

Least count: the smallest difference between two indications that can be defected on the instrument scale.

Sensitivity: ratio of the linear movement of the painter on the instrument to the change in the measured variable causing this motion.

Hysteresis effect

An instrument said to exhibit hysteresis when there is a difference in readings as whether the value of the measured quantity is approached from above / below. Hysteresis may be the result of mechanical friction, magnetic effects elastic deformation or thermal effects.

Accuracy of an instrument indicates the deviators of the reading from a known input. Precision of an instrument indicates its ability to reproduce certain reading with a given accuracy.

System response

In a system having both inputs and outputs suppose a steper instantaneous input signal is applied to a system. In general there will be a slight delay in the output response and this delay is called the ‘rise time’ or delay of the system.

Distortion

Distortion is a very general term that may be used to describe the variation of a signal from its true form. Depending on the system, the distortion may result from either poor frequency response or poor phase shift response.

Impedance matching

In al electrical circuits proper care must be taken to avoid impedance mistaking.

Deflection methods

Null type device attempts to maintain deflection at zero by suitable application of an effect opposing that generated by the measured quantity. Deflection instrument –

Accuracy depends on the calibrator of the spring null instrument it depends on the accuracy of the standard weights

Measurement system

- Measurement system consists of a transducing element which converts the quantity to be measured in an analogous form.
- The analogous signal is then processed by some intermediate means and is then fed to the end devices which present the results of the measurement.

General measurement system

Instruments classification according to applications.

1. Monitoring of processes and operation
2. Control of processes and operations
3. Experimental engineering analysis

A.Monitoring of processes and operation

- monitoring the function indicates the conditions.

Eg. : Thermometer, barometers and anemometer, electric meter, water meter, garmete.

B.Control of processes and operations

- component of an automatic control system.

Eg. : Thermostatic control, Air conditioner, thermocouples, accelerometers, altimeters.

C.Experimental Engineering analysis

(i) Theoretical methods

(ii) Experimental methods

Functional elements of an instrument or a measurement system

- to describe both the operations and the performance of measuring instruments

- in terms of functional elements of instrument system

- performance is defined in terms of static and dynamic characteristics.

Primary Sensing element

Receives energy from the measured medium and produces on output depending in some way on the measured quantity.

Variable conversion element

To convert the physical or any variable to another more suitable variable while preserving the information content of the original signal.

Variable manipulation system / element

To perform the intended task, an instrument may require a signal represented by some physical variable be manipulated in some way. Ie. Change in numerical value according to some definite rule with physical preservation of the nature of the variable. The manipulation element may appear / may not appear / elsewhere in the chain.

Data transmission element

For transmitting the data from one to another, when the functional elements of an instrument are actually physically separated.

Data Presentation Element

- Should be recognizable by one of the human senses. An element that performs the 'translation' function is called as data presentation element.

Data storage / playback function

For storing the data and get back the data

LESSON 25. Basic input circuits-Active and passive transducer-Error-types-calibration.

Basic input circuits

The output available from the sensor or transducer is often not compatible with the desired transmission system or display unit. This can be made to be so by appropriate signal conditioning and processing of data thereafter. Depending on the type and mode of the output from the sensor, the intermediate stage of conditioning of signal is decided. There may be a number of units in this conditioning stage. A very simple instrumentation system is represented in the figure.

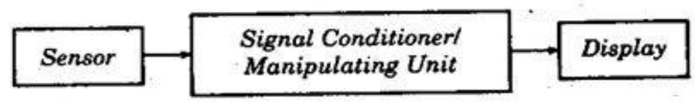


Fig A simple instrumentation system.

The manipulating unit consists mainly of a conversion circuit (a bridge circuit), an amplifier, if necessary an impedance matching circuit, for transmission a modulation circuit and a detector fro display. All units are not always necessary and sometimes some more and different units are needed.

The present day trend is to convert the sensor output into an electrical signal first whatever be the form of its original output as conditioning and data processing are much more convenient by electrical means.

Basic input circuits (Interfacing circuits)

Interfacing circuit or input circuit is an electrical circuit that interfaces the sensor/transducer (where electrical output is available) with the more demanding signal conditioning equipment like amplifier, modulator, processor, etc. There are a number of such circuits performing appropriate functions in the interfacing stage. One major reason of such circuits is that the transducer may be a passive type and requires an auxiliary source of energy for giving a signal in voltage or in current. As mentioned earlier, bridge circuit is a very common interfacing circuit. There are others like ballast circuits current sensitive circuits, voltage divider circuits etc.

Current sensitive circuit

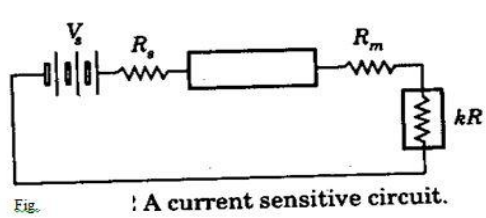


Fig. : A current sensitive circuit.

It is basically a series circuit of the voltage supply, a current indicating/recording device and the transducer of the resistance type. The supply source has a internal resistance R_s. The meter has a resistance R_m and the transducer resistance may vary between 0 and R (say) – let it be represented by kR when 0 £ k £1. The circuit schematic is shown in Fig. 13.2. The meter current at any stage is i with a maximum of i_x.

Let $R_m + R_s = R_T$ then the circuit current is given by

$$i = \frac{V_S}{R_T + kR}$$

The maximum current occurs when k = 0, so that

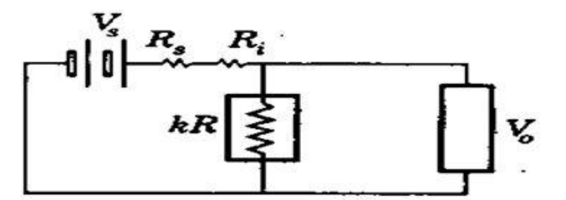
$$i_x = \frac{V_S}{R_T}$$

giving the ratio

$$\frac{i}{i_x} = \frac{1}{1 + k\frac{R}{R_T}}$$

The output, I is a function of i_x i.e. of V_S and it should be maintained constant. The sensitivity of the system increases with increasing R/R_T i.e. system is nonlinear.

Ballast circuit



A ballast circuit.

It is a voltage sensing circuit instead of the current sensing type. The scheme is shown in Fig. 13.3. V_s is the supply voltage, R_v is the supply or voltage source resistance, V_0 is the supply voltage, R_v is the supply or voltage source resistance, V_o is the voltage indicated or recorded for a variation in R again given as kR, R_i is a resistance introduced in the position as shown and $R_i >> R_s$. The resistance R_i is known as the ballast resistance.

For the voltmeter having a very high resistance, current through $R_i + R_s$ is

$$\frac{V_s}{R_i + R_s + kR}$$

and the voltage across kR is also the output voltage V_o

$$V_o = ikR = \frac{V_s kR}{R_i + R_s + kR}$$

or

$$\frac{V_o}{V_s} = \frac{kR/(R_i + R_s)}{1 + kR/(R_i + R_s)} = \frac{kR/R_b}{1 + kR/R_b}$$

With change in measurand k changes so that change in V_o with k is given by

$$\frac{dV_o}{dk} = \frac{V_s(R_i + R_s)R}{(R_i + R_s + kR)^2}$$

There is, however, an optimum choice of R_i (or, $R_i + R_s = R_b$ say) for which sensitivity can be made maximum. By differentiating V_o with respect to R_b, this is obtained as

$$R_b = kR$$

Obviously, this also varies with sensor output resistance

Loading Error

If the voltmeter draws a current, there is an error in measurement called the loading error. If the meter has a resistance R_m then the total resistance seen by the source V_s is

$$R_t = R_b + \frac{kR \times R_m}{kR + R_m}$$

and the current is now

$$i = \frac{V_s}{R_t} = V_s \Big/ \left[\frac{(kR + R_m)R_b + R_m kR}{kR + R_m}\right]$$

so that the output voltage is now

$$V_o = V_s - iR_b$$

$$= V_s - V_s R_b \Big/ \left[\frac{R_b(kR + R_m) + kRR_m}{kR + R_m}\right]$$

$$= V_s \frac{kR/R_b}{1 + kR/R_m + kR/R_b}$$

By taking the difference one can show that

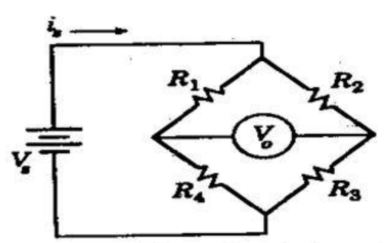
$$\Delta V_o = V_s \left[\frac{1}{1 + \frac{1 + kR/R_b}{kR/R_b}} \right]$$

so that the % error is

$$E\% = \frac{100}{1 + \frac{1 + kR/R_b}{kR/R_m}} \times \frac{1 + kR/R_b}{kR/R_b}$$

It is worthwhile to note that (percentage) error also is a function of k.

Bridge Circuits



The wheatstone bridge circuit.

The most common interfacing circuit between a passive sensor and the measurement system is the bridge circuit. Wheatstone bridge is a circuit that is extensively used till today although it was devised as early as 1833. A lot of variation has since been in operation of this bridge without altering the basic structure which is shown in Figure.

The resistance bridge supply is V_s and output V_o. Solving for $V_o = 0$, the condition in terms of R_1, R_2, R_3 and R_4 is obtained as

$$\frac{R_1}{R_4} = \frac{R_2}{R_3} \ or \ \frac{R_1}{R_2} = \frac{R_4}{R_3}$$

This is called the null-balance bridge. It is possible to take output V_o for different values of resistance. It is initially balanced satisfying the above condition and subsequently the unbalanced voltage is measured b a voltmeter of very high impedance. However, alternative to this, a current meter may also be connected and response observed.

For voltage detector scheme with very high impedance of the detector and R_4 as the sensor resistance that changes to R_4 + DR_4 and initially with $R_1 = R_2 = R_3 = R_4 = R$ one derives.

$$\frac{\Delta V_o}{V_s} = \frac{\Delta R_s / R}{4 + 2\frac{\Delta R_s}{R}}$$

which changes to

$$\frac{\Delta V_o}{V_s} = \frac{\Delta R_s}{4R}$$

for DR_s<<R.

If, instead, a galvanometer is used as a detector with its resistance R_g often chosen to be equal to R_v and, again with all R/s equal initially to R, one derives the incremental detecting current through the galvanometer for an incremental change in R_1 equal to DR_1 say, and is given as

$$\frac{\Delta i_g}{i_g} = \frac{-\Delta R_1 / R_1}{4\left(1 + \frac{R_g}{R}\right) + \left(2 + \frac{R_g}{R}\right)}$$

where i_g is the current supplied by the source.

Bridge circuits have been used in ac operation for resistance, capacitance as well as inductance measurement. Impedance bridges, as they are called, are listed in figure

Active and passive transducers

To perform all the general functions, a physical component may act as an active transducer or a passive transducer.

A component whose output energy is supplied entirely or almost entirely by its input signal is commonly called a passive transducer. The output and input signals may involve energy of the same from or there may be an energy conversion from one form to another farm.

An active transducer has an auxillary source of power which supplies a major part of the output while the input signal supplies only an insignificant portion.

Error Analysis

Types of Errors

a) Gross errors- largely human errors, among them misreading of instruments, incorrect adjustment and improper application of instruments and computational mistakes.

b) Random errors –those errors that cannot be directly established because of random variations in the parameter or the system of measurement.

c) Systematic errors- Shortcomings of the instruments such as defective or worn parts and effects of the environment on the equipment or the user.

Types of systemic errors

- Instrumental errors
- Environmental errors
- Static errors
- Dynamic errors

The uncertainty in the final result is due to the uncertainties in the primary measurements. This may be done by a commonsense analysis of the data which may take many forms. One rule of thumb that could be used that the error in the result is equal to the maximum error in any parameter used to calculate the result. Another commonsense analysis would combine all the errors in the most detrimental way in order to determine the maximum error in the final result. Consider the calculation of electric power from

$P=EI$

Where E and I are measured as

$E=100V\pm2V$

$I=10A\pm0.2A$

The nominal value of the power is 100*10=1000W.by taking the worst possible variations in the voltage and current, we could calculate

$P_{max}= (100+2)\ (10+0.2) =1040.4W$

$P_{min}= (100-2)\ (10-0.2) =960.4W$

Thus, using this method of calculation, the uncertainty in the power is +4.04 percent,-3.96%.it is quite unlikely that the power would be in the error by these amounts because the voltmeter variations would probably not correspond with the ammeter variations. when the voltmeters reads an extreme “ high” there is no reason that the ammeter must also read an extreme “high” at that particular instant, indeed, this combination is most unlikely.

Calibration

Calibration: To check the instrument against a known standard and subsequently to reduce errors in accuracy

Calibration procedures involve a comparison of the particular instrument with either a primary standard or a known input source.

NIST-This institute defines the standard for length, mass, time, temperature and force.

Static calibration -All the static performance characteristics are obtained in one form or another by a process called static calibration. All working instruments must be calibrated against some reference instruments which have a higher accuracy.

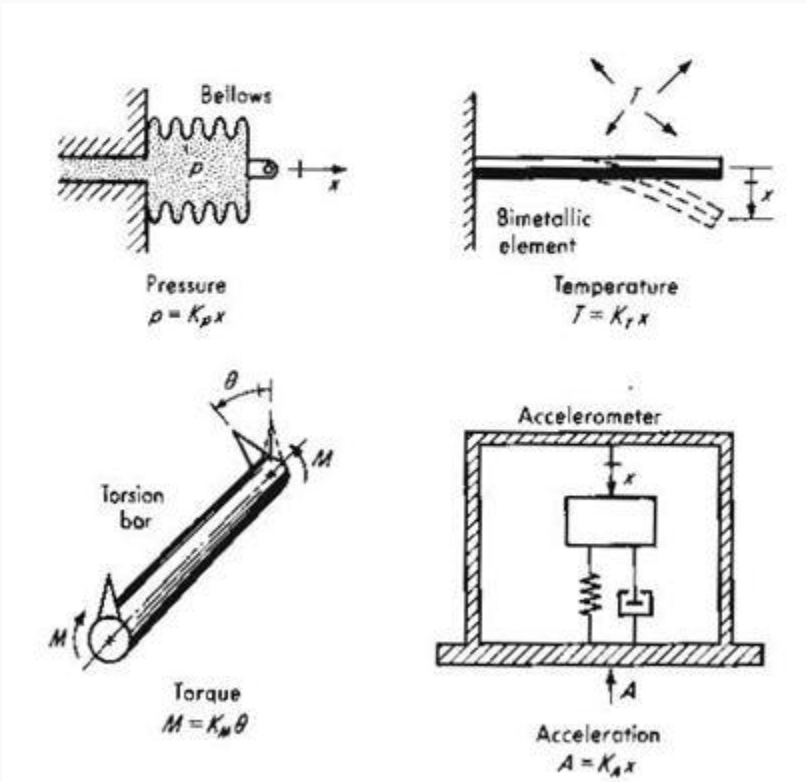
Error calibration-Error calibration means that an instrument has been calibrated against a suitable standard.

MODULE 16.

LESSON 26. Relative displacement, translational and rotational transducer-calibration

INTRODUCTION

We consider here devices for measuring the translation along a line of one point relative to another and the plane rotation about a single axis of one line relative to another. Such displacement measurements are of great interest as such and because they form the basis of many transducers for measuring pressure, force, acceleration, temperature, etc., as shown in the following figure



CALIBRATION

Static calibration of translational devices often can be satisfactorily accomplished by using ordinary dial indicators or micrometers as the standard. When used directly to measure the displacement of the transducer, these devices usually are suitable to read to the nearest 0.0001 in or 0.01 mm. If smaller increments are necessary, lever arrangements (about a 10:1 ratio is fairly easy to achieve) or wedge-type mechanisms (about 100:1) can be employed for motion reduction. The Mikrokator, a unique mechanical gage of high sensitivity, may also be useful in measuring small motions down to a few millionths of an inch.

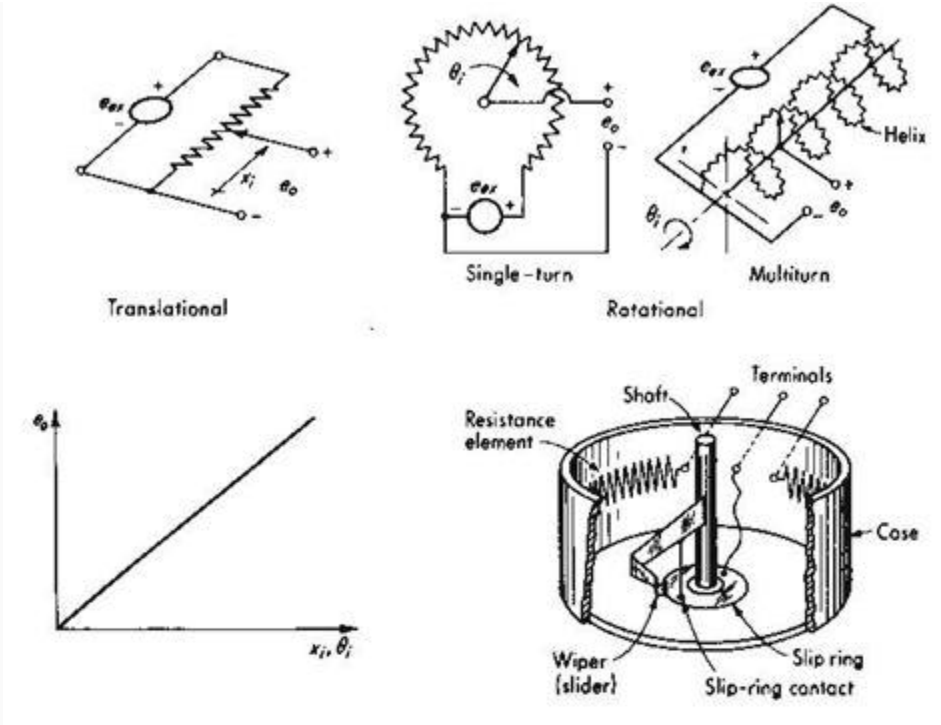
If accuracy to 0.0001 in or better is required, such equipment should itself be calibrated against gage blocks, or (for maximum accuracy) gage blocks of hard, used directly to calibrate the transducer. Gage blocks are small blocks of hard, dimensionally stable steel or other material, made up in sets which can be stacked up to provide accurate dimensions over a wide range and in small steps. They are the basic working length standards of industry. As purchased from the manufacturer, their dimensions are accurate to ± 8 μin for working grade blocks, ± 4 μin for reference grade, and 2 μin for all blocks up to 1 in (± 2 μin / in for blocks longer than 1 in) for master blocks. If these tolerances are too large, the blocks can be sent to NIST and calibrated against light wavelengths to the nearest 10^{-7} in. Some precision – manufacturing operations currently require and use the latter calibration service. When transducers are calibrated to very high accuracies, it is extremely important to control all interfering and/or modifying inputs such as ambient temperature, electrical excitation to the transducer, etc.

Rotational or angular displacement is not itself a fundamental quantity since it is based on length, and so a fundamental standard is not necessary. However, reference and working standards for angles (and thus angular displacement) are desirable and available. The basic standards (against which other standards or instruments) may be calibrated are called angle blocks. These are carefully made steel blocks about 5/8 in wide and 3 in long, with a specified angle between the two contact surfaces. Just as for length gage blocks, these angle blocks can be stacked to "build up" any desired angle accurately and in small increments. The blocks can be calibrated to an accuracy of 0.1 second of arc by NIST.

Rotational transducers rarely require such accuracy for calibration, nor can the laborious and expensive techniques necessary to realize these limits be economically justified. Thus most static calibration of angular – displacement transducers can be adequately carried out by using more convenient and readily available equipment. Examples of such equipment which should be available in a precision machine shop are the circular division taster (range 360^o, microscope reads to 0.1 minute of arc, precision of scale disk ± 20 seconds of arc), the optional dividing head (range 360^o, scale reads to 1.0 minute of arc, working accuracy ± 20 seconds of arc), the optical dividing head (range 360^o, scale reads to 1.0 minute of arc, working accuracy ± 20 seconds of arc), and the division tester with telescope and collimator (accuracy ± 2 seconds of arc). In some applications, even cruder devices such as ordinary machine-tool index heads, calibrated dials, etc., may be perfectly adequate.

RESISTIVE POTENTIOMETERS

Basically, a resistive potentiometer consists of a resistance element provided with a movable contact. The contact motion can be translation, rotation, or a combination of the two (helical motion in a multiturn rotational device), thus allowing measurement of rotary and translatory displacements. Translatory displacements. Translatory devices have strokes from about 0.1 to 20 in and rotational ones range from about 10^o to as much as 60 full turns. The resistance element is excited with either dc or ac voltage, and the output voltage is (ideally) a linear function of the input displacement. Resistance elements in common use may be classified as wire-wound, conductive plastic, hybrid, or cermet.

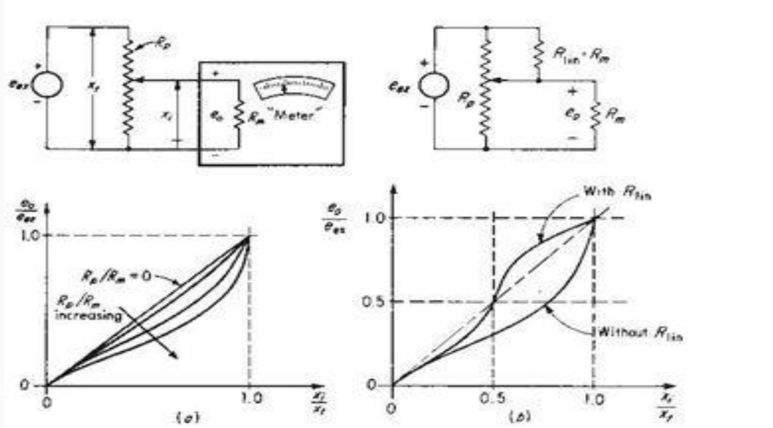


If the distribution of resistance with respect to translational or angular travel of the wiper (moving contact) is linear, the output voltage e_o will faithfully duplicate the input motion x_i or θ_i if the terminals at e_o are open-circuit (no current drawn at the output). (For ac excitation, x_i or θ_i amplitude-modulate e_{ex} , and e_o does not look like the input motion.) The usual situation, however, is one in which the potentiometer output voltage is the input to a meter or recorder that draws some current from the potentiometer. Thus a more realistic circuit is as shown in fig. Analysis of this circuit gives

$$\frac{e_o}{e_{ex}} = \frac{1}{1/(x_i/x_t) + (R_p/R_m)(1 - x_i/x_t)}$$

which becomes for ideal ($R_p/R_m = 0$ for an open circuit) conditions

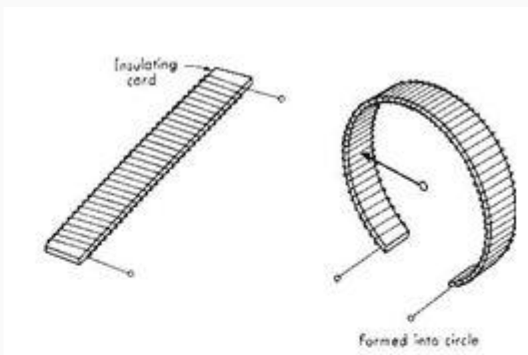
$$\frac{e_o}{e_{ex}} = \frac{x_i}{x_t}$$



thus for no "loading" the input –output curve is a straight line. In actual practice, $R_m \neq \alpha$ and Eq.(4.1) shows a nonlinear relation between e_o and x_i. This deviation from linearity is shown in Fig.4.5a. The maximum error is about 12 percent of full scale if $R_p / R_m = 1.0$ and drops to about 1.5 percent when $R_p/R_m = 0.1$. For values of $R_p / R_m < 0.1$, the position of maximum error occurs in the neighbourhood of $x_i/x_t = 0.67$, and the maximum error is approximately 15 R_p / R_m percent of full scale.

We see that to achieve good linearity, for a "meter" of a given resistance R_m, we should choose a potentiometer of sufficiently low resistance relative to R_m. This requirement conflicts with the desire for high sensitivity. Since e_o is directly proportional to e_{ex}, it would seem possible to get any sensitivity desired simply by increasing e_{ex}. This is not actually the case, however, since potentiometers have define power ratings related to their heat-dissipating capacity. Thus a manufacturer may design a series of potentiometers, say single-turn 2-in-diameter, with a wide range (perhaps 100 to 100,000 Ω) of total resistance R_p, but all these will be essentially the same size and mechanical configuration, giving the same heat-transfer capability and thus the same power rating, say about 5 W at 20°C ambient. If the heat dissipation is limited to P watts, the maximum allowable excitation voltage is given by

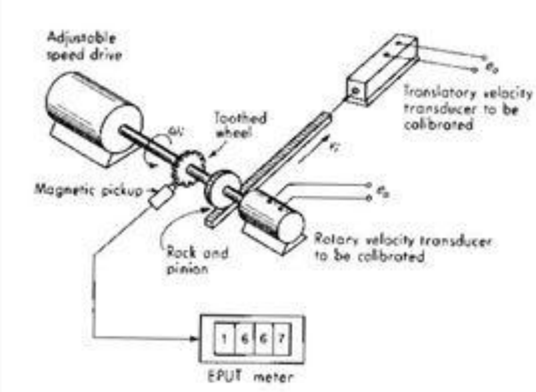
$$\max e_{ex} = \sqrt{PR_p}$$



Thus a low value of R_p allows only a small e_{ex} and therefore a small sensitivity. Choice of R_p Thus must be influenced by a tradeoff between loading and sensitivity considerations. Thus maximum available sensitivity of potentiometers varies considerably from type to type and also with size in a given type. It can be calculated from the manufacturer's data on maximum allowable voltage, current, or power and the maximum stroke. The shorter-stroke devices generally have higher sensitivity. Extreme values are of the order of 15 V/deg for short-stroke rotational types ("sector" potentiometers) and 300 V/in for short-stroke (about ¼ in) translational pots. It must be emphasized that these are maximum values and that the usual application involves a much smaller (10 to 100 times smaller) sensitivity. Fig. 4.5b shows a method for improving linearity without increasing R_m.

CALIBRATION

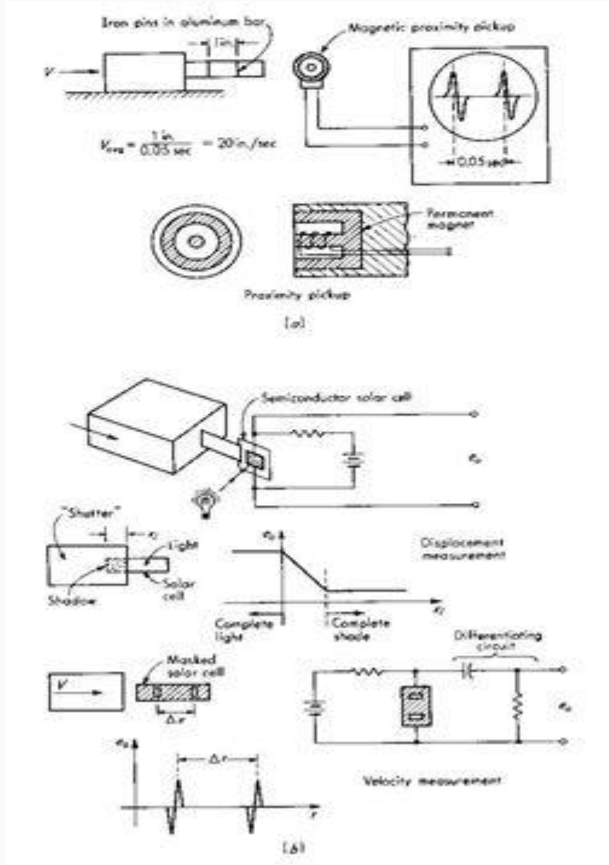
The measurement of rotational (angular) velocity is probably more common than that of translational velocity. Since translation generally can be obtained from rotation devices. For angular and linear velocities, perhaps the most convenient calibration scheme uses a combination of a toothed wheel, a single magnetic proximity pickup and an electronic EPUT (Events per unit time) meter .



The angular rotation is provided by some adjustable-speed drive of adequate stability. The toothed iron wheel passing under the proximity pickup produces an electric pulse each time one tooth passes. These pulses are fed to the EPUT meter, which counts them over an accurate period (say 1.00000 s), displays the result visually for a few seconds to enable reading, and then repeats the process. The stability of the rotational drive is easily checked by observing the variation of the EPUT meter readings from one sample to another. The inaccuracy of pulse counting is ± 1 pulse plus the error in the counter time base, which is of the order of 1 ppm. The overall accuracy achieved depends on the stability of the motion source, the angular velocity being measured, and the number of teeth on the wheel. If the motion source were absolutely stable (no change in velocity what-ever), very accurate measurement could be achieved simply by counting pulses over a long time, since then the average velocity and the instantaneous velocity would be identical. If the motion source has some drift, however, the time sample must be fairly short. For example, a shaft rotating at 1,000 r/m with a 100-tooth wheel produces 1,667 pulses in a 1-s sample period. The inaccuracy here would be 1 part in 1,667 (the 1-ppm time-base error is totally negligible), or 0.06 percent. If the shaft rotated at 10 r/min, the error would be 6 percent. Slow rotations can be measured accurately by such means if the toothed wheel is placed on a shaft which is sufficiently geared up from the shaft driving the transducer being calibrated.

The above procedure uses relatively simple equipment and generally provides entirely adequate accuracy. Other simpler and less accurate procedures can be employed if they are adequate for their intended purpose. These usually consists of simply comparing the reading of a velocity transducer known to be accurate with the reading of the transducer to be calibrated when both are experiencing the same velocity input.

Velocity by Electrical Differentiation of Displacement Voltage Signals



The output of any displacement transducer may be applied to the input of a suitable differentiating circuit to obtain a voltage proportional to velocity. The main problem is that differentiation accentuates any low-amplitude, high-frequency noise present in the displacement signal. Thus a carbon-film potentiometer would be preferable to the wire wound type, and demodulated and filtered signals from ac transducers may cause trouble because of the remaining ripple at carrier frequency.Workable systems using electrical differentiation are possible, however, with adequate attention to details.Displacement is the vector representing a change in position of a body or a point with respect to a reference. It may be linear or rotational motion, expressed in absolute or relative terms. Many of the modern scientific and industrial observations need a very accurate measurement of this parameter. Being a fundamental quantity, the basic sensing device is widely adapted with suitable linkages for the measurement of many derived quantities, such as force, stress, pressure, velocity, and acceleration. The magnitude of measurement ranges from a few microns to a few centimetres in the case of linear displacement and a few seconds to 3600 in the case of angular displacement. A majority of displacement transducers sense the static or dynamic displacement by means of a sensing shaft or similar links mechanically coupled to the point or body whose motion is measured. Such attachments of both linear and angular transducers are usually of simple mechanical configurations, but the coupling must be primarily designed to avoid any slippage after it is fastened and thereby keep the back-lash minimum. For linear-displacement measurements, the common types employed are the thread- ed end, lug, clevis, and bearing couplings. Spring-Ioaded shafts may also be used for certain applica- tions. A number of specialized types of displacement transducers operate without use of a mechanical linkage between the transducer and the object whose displacement is to be measured, as in the case of some of the electromagnetic, capacitive, and optical transducers.

PRINCIPLES OF TRANSDUCTION

Displacement transducers can be classified primarily on the basis of the transduction principle employed for the measurement. In this chapter only the electromechanical transducers, which convert displacement quantities into electrical voltages/currents, are dealt with.

The major electrical transduction principles used are: (i) Variable resistance-potentiometric/strain gauge

(i) Variable inductance /linear variable differential transformer/variable reluctance (iii) Variable capacitance

(ii) Synchros and resolvers.

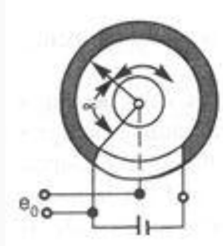
A number of additional types are also designed, depending upon the convenience and measurement accuracy required, such as digital output transducers, electro-optical devices, and the radioactive devices. In practice, potentiometric- and inductive-type devices are most widely used in scientific and engineering applications. The performance characteristics of a few selected variety displacement transducers are shown in Table

Transduction Principle	Range mm	Linearity % F.S.	Repeatability (microns)	Temperature Range (°C)	Resolution (microns)	Frequency Response (Hz)	Remarks
1. Resistive wire-wound potentiometer	100	0.25	50	−10 to 75	50	5	Long ranges, economical, high output, and minimum electronics
Conductive strip potentiometer	100	0.5	5	−10 to 75	10	10	Poor resolution and high noise
Cantilever with strain gauges	10	0.5	50	−10 to 70	10	100	Versatile, small size, low range, and large reaction forces
2. Inductive/variable reluctance	5	0.5	0.5	−20 to 75	2	100	Small size, high resolution, non-contact type, and low range
Linear variable differential transformer	50	0.1	0.5	−10 to 75	1	1000	Good linearity, high resolution, but interference due to magnetic field
Eddy current (proximity)	10	0.75	5	−20 to 80	2	5000	Non-contact type, and high output
3. Capacitive							
(a) Variable area	50	0.1	0.5	− 40 to 200	0.1	50	Easy mechanical design, high temperature operation, error due to stray capacitance
(b) Variable gap	5	0.5	2	−10 to 500	0.1	2000	
4. Digital transducer	10 to 500	0.1	0.5	−0 to 55	0.5	100	Long range, digital output, high accuracy and resolution, expensive and bulky

Variable Resistance Device

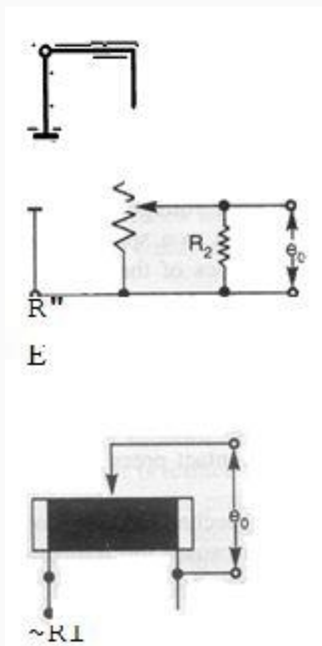
Displacement transducers using potentiometric variable resistance transduction elements are invariably shaft-coupled devices. The sensing element is basically a resistance-potentiometer with a movable wiper contact attached to an insulated plunger type shaft, mechanically

linking the point under measurement.



The contact motion can be translation, rotation, or a combination of the two, thus allowing measurements of rotary or translatory displacements. They are relatively simple in construction, in the sense that a sliding contact (wiper) is made to move linearly over a resistance el~ment which may be in the form of a wire or a conductive plastic film. The resistivity and temperature coefficient of the resistance element should be of such a value that the device operates with appreciable constant sensitivity over a wide temperature range. The three major elements critical in a potentiometric device are the winding wire, winding former, and wiper

Potentiometer displacement transducer. {a) linear motion; {b) angular motion; {c) circuit arrangement



The winding wire is a precision drawn resistance wire with a diameter of about 25 to 50 microns and is wound over a cylindrical or a flat mandrel of ceramic, glass, or anodized aluminium. The wire is annealed in a reducing atmosphere to avoid any surface oxidation. Resistivity may vary normally from 0.4 ~-m to 1.3 !lQ-m, and temperature coefficient may vary from *0.002%IOC* to 0.01 *%loC.* The wire should be strong, ductile, and protected from surface corrosion by enamelling or oxidation. The dimen- sional tolerance should be less than 1 %, and the resistance stability with time should be of a very high order. The materials commonly employed are the alloys of copper-nickel, nickel-chromium, and silver- palladium. The winding can be linear, toroidal, or helical, and should possess uniform spacing and constant tension. The outer surface, except for the linear track of the wiper, is covered with a suitable insulating material to protect against dust and abrasion.

The wipers are spring elements made from tempered phosphor-bronze, beryllium-copper, or other precious metal alloys and are suitably shaped to move over the resistance element with minimum friction. The wiper contact force and contact resistance are important factors in the overall accuracy of the device. In some cases conductive lubricants are also used to reduce the friction. Leaf spring and dual wipers are designed for better contact and ability to withstand high shocks and vibration. The main requirements for winding formers are good dimensional stability and surface insulation. Some of the recommended materials are ceramic, steatite, anodized aluminium, and moulded epoxies.

The whole transduction element described above can be employed for linear as well as angular displacement. The linear range depends very much on the mechanical design. The resistance value and the current-carrying capacity are chosen to suit the desired application; normal values are range 2 to 10 cm F.S., resistance 100 to 50,000 ohms, and current capacity 0.5 to 5 mA. The resolution of the device depends upon the wiper width, diameter of the resistance wire, and spacing between the windings. An optimum choice is sought for the highest precision and resolution. In the case of the wire wound element, the wiper wire diameter to spacing ratio is normally 10, and the resolution achievable is 0.05% to 0.1 %. A linearity of the order of 0.1% can be achieved easily, as wires of uniform diameter and specific resistance are available. The plastic.film type is ideal for infinite resolution purposes, even though it is difficult to get a resolution better than 5 microns in practice.

Electrical noise is another factor normally exhibited by these devices and they are random in nature. Further, they depend on the current and speed of motion of the wiper. Wire.wound devices are relatively free from Johnson noise. But the contact noise caused by variation in contact resistance when the wiper moves along the potentiometer track is not negligible in many cases. The noise level increases with wear and tear and also with the contamination or oxidation of the track and the wiper surfaces. Sometimes, thermoelectri~ effects due to dissimilar materials used for the wiper and wire can also generate a voltage acting as a noise source. This is particularly true when the device is operating at higher temperatures. Yet another type of noise exhibited is the vibrational noise or high.velocity noise caused by jumping or bouncing movements of the wiper. This is reduced by adjusting the contact pressure and oscillatol"y characteristics of the wiper structure.

The sensitivity of the device is normally given as volts per full-scale mechanical travel of the wiper. The input excitation voltage is limited by the dissipating wattage which causes the temperature of the winding wire to rise to a specified level. This voltage level depends upon the cooling conditions, the thermal characteristics of the potentiometer wire, and the transducer housing design.

Inherent linearity depends on the minimum resolution achievable in the device. If the apparent resolution is $n\%$ of F.S., the linearity error cannot be smaller than :t *Ifm*% of F.S. Further, the linearity is a function of the winding pitch, variations in wire diameter, and any irregularity in former dimensions and wiper movements. The normal value achievable for a standard unit is 0.1 %

The resistance measurement can be carried out with a simple circuitThe circuit linearity is determined by the ratio of the total potentiometer resistance *R1* to the load resistance *R2* as indicated in the figure.

The major disadvantages of the potentiometer-type displacement transducer are poor dynamic re- sponse, susceptibility to vibration and shock, poor resolution, and presence of noise in signal. Displacement transducers for very short stroke lengths can be designed with high precision using a bonded/unbonded strain.gauge type sensor. The motion to be measured is transferred to an elastic element, such as a cantilever beam, and the stresses developed on application of displacement is related to the motion. This principle is extended very much in the design of force, pressure, and acceleration transducers.

Variable Inductance Transducer

A simple and more popular type of displacement sensor is the variable-inductance type wherein the variation of inductance as a function of displacement is achieved either by variation in mutual inductances or self inductances. Devices operating on these principles are more widely known as linear variable differential transformers and variable reluctance sensors respectively.

(a) Linear Variable Differential Transformer *(*LVDT*)* Linear variable differential transformer type of transducers find a number of applications in both measurement and control systems. The extremely fine resolution, high accuracy, and good stability make the device particularly suitable as a short-stroke, po~ition-measuring device. Since a number of physical quantities, such as pressure, load, and acceleration can be measured in terms of mechanical deflection. L VDT forms the basic sensing element in all such measurements. The L VDT device is widely used as the basic element in extenso- meters, electronic comparators, thickness-measuring units, and level indicators. Some of the other important applications aI.e in numerically controlled machines and creep-testing machines. The linear variable differential transformer consists of a primary coil and two identical secondary coils, axially-spaced and wound on a cylindrical-coil former, with a rod-shaped magnetic core positioned centrally inside the coil assembly providing a preferred path for the magnetic flux linking the coils. The displacement to be measured is transferred to the magnetic core through suitable linkages. When the primary coil is energized with an ac carrier wave signal, voltages are induced in each secondary section, the exact value depending upon the position of the magnetic core with respect to the centre of the coil assembly. If the core is symmetrically placed (electrically) with respect to the two secondary coils, equal voltages are induced in the two coils. When these two outputs are connected in phase opposition as shown in Fig. 4.2(b), the magnitude of the resultant voltage tends to a zero value. Such a balancepoirit is termed 'the null position'. In practice, a small residual voltage is always present at a null position due to the presence of harmonics in the excitation signal and stray capacitance coupling between the primary and secondary windings. When the core is now displaced from the null position the induced voltage in the secondary towards which the core has moved increases while that in the other secondary decreases. This results in a differential voltage output from the transformer. where *f* = excitation signal frequency, *I,* = primary current, *np* = number of turns in primary, *n.f* = number of turns in secondary. *b* = width of[.the primary coil, *w* = width of the secondary coil, *x* = core displacement, *r* = outer radius of the coil, and *r.* = inner radius of the coil.

With proper design of coils, the magnitude of the output signal is made to vary linearly with the mechanical displacement of the core on both sides with respect to the null position. as shown in Fig.. While the magnitude of the output voltages are ideally the same for equal core displacements on either side of the null, the phase difference between the output and input voltages changes by 1800 when the core moves through the null position. In actual measurement. this phase change-over is measured with a phase-sensitive detector. The sensitivity is proportional to the frequency *f* and the primary current *I.* and for best linearity *x* « *b.* However. larger *I* produces core saturation and an increase in the temGerature of the coil, and hence results in larger harinonics at null position, making adjustment difficult. An increase in frequency produces a greater effect of stray capacitance. and in turn a large null voltage. In practice, the design is optimized for the lowest null voltage, highest linearity, and appropriate size.

The coils are wound on phenolic or ceramic formers to improve the dimensional stability. The coil former material should be strong and mechanically stable to guard against temperature effects and should be able to withstand elevated temperature and thermal shock. The coils are wound with an enamelled copper wire possessing an insulation suitable for the ambient temperature specified. The transformer is then enclosed in ferromagnetic cases, providing full electro- static and electromagnetic shielding. The moving core is made of ferromagnetic material of high per- meability. selected for optimum performance in general use and heat treated to provide the best magnetic properties. The normal excitation voltage is 1.0 Vat a carrier frequency of 2 kHz to 10 kHz. The carrier frequency is suitably chosen for optimum sensitivity and proper demodulation. The dynamic response of the L VDT is limited mainly by the excitation frequency; faithful linear characteristics are obtained for frequencies up to 0.1 times the carrier frequency. The normal ranges are :t 10 microns to :t10 mm, operating over a temperature range of -40° to + 100° C. In general, the linear range is primarily dependent on the length of the primary and secondary coils. The instrumentation can be carried out with a suitable carrier wave amplifer, followed by a phase sensitive detector and a filter. as described in Chapter 12. Phase detector is invariably used in all the measurement systems to avoid the ambiguity in the direction of motion. With the availability of miniature integrated chips. it is feasible to incorporate the oscillator, the demodulator and the associated electronic circuitry within the transducer housing itself, thereby enabling the device to operate as a dc-dc system. An output voltage of 0 to 5 V can be obtained with an input supply of :t 15 V dc.

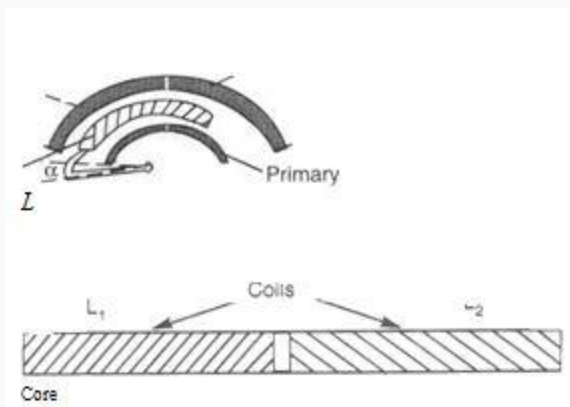
The main advantages of the LVDT type of displacement sensors are:

(i) Mechanical*:* Simplicity of design and ease of fabrication and installation, wide range of displace- ment; frictionless movement of core and hence infinite resolution; rugged construction; negligible operating force (core weight being low), and ability to operate even at higher temperatures.

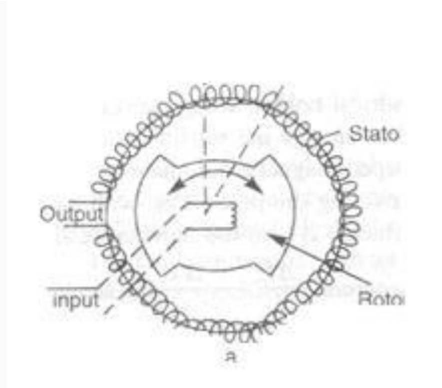
(ii) Electrical*:* Output voltage is a linear and continuous function of mechanical displacement (linear- ity better than 0.25%), high sensitivity (2 mV/volt/10 microns at 4 kHz excitation); low output impedance (100 ohms); ability to operate over a wide range of carrier frequencies (50 Hz to 20 kHz); infinite resolution in output (theoretically limiting factors being signal to noise ratio and input stability conditions); and very low cross sensitivity.

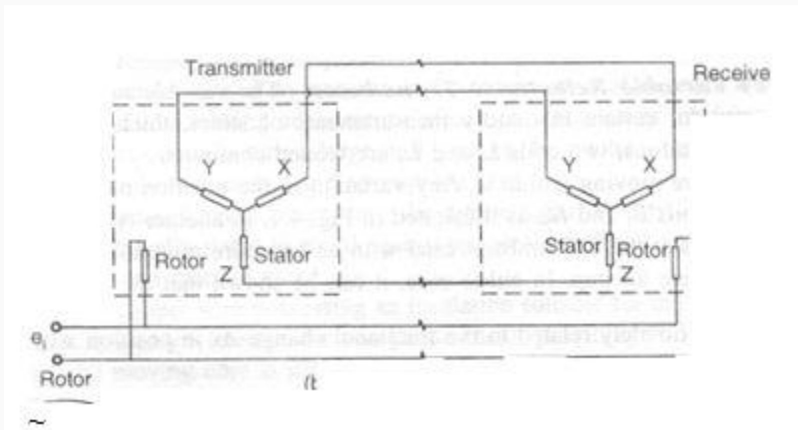
(b) Angular Displacement Measurement: A rotary variable differential transformer is a very convenient device for the measurement of angular displacement. The device operates on the same principle as L VDT explained earlier, where the output yoltage varies linearly with the angular position of the shaft. A cardioid-shaped cam (rotor) of a magnetic material is used as the core. The input shaft fastened to the core is mounted at the Sec. 1 centre of the coil former on which the primary and secondary are wound symmetrically. The cardioid shape of the rotor is so chosen Core as to produce a highly linear output over a specified angle of rotation. Main advantages of the unit are infinite resolution and linear operation (better than:!: 0.5% of the full range). The signal conditioner employed here is the same as that for LVDT type transducers.

(c) Variable Reluctance Transducer: The variable reluctance type displacement sensor is very useful for certain laboratory measurements of stress, thickness, vibration, and shock. In one of the configurations, two coils *Lt* and *L2* are wound continuously over a cylindrical bobbin with a ferromag- netic core moving within it. Any variation in the position of the core will change the self-inductances of the coils *Ll* and *L2,* as illustrated in Fig. 4.4. In another type, an E-shaped magnetic core having two windings at the end limbs is used with an armature mounted suitably covering the pole faces, with an appropriate air gap. In either case, it can be shown that the fractional change .1 *L* in the inductance *L*



Measurements are carried out with a Wheatstone's bridge network wherein the coils *L1* and *L2* form half of the bridge and the other two arms are completed with two fixed resistors *R1* and *R2* with capacitors C1 and C2 in parallel to achieve both amplitude and phase balance. The main advantages of the device are its high sensitivity and good linearity of 0.5 to 1% even for long stroke lengths of 50 cm (overall length should be double that of a stroke length). The device can be operated under severe environmental conditions, with encapsulation of the coils in an epoxy resin and hermet- ically-sealed bobbin. With appropriate modification, it can be employed for velocity and acceleration measurements also.





Displacement transducer

In the displacement transducer, mechanical elements like diaphragm, bellows, bourdon tube and single or double suspension cantilever are used to convert the applied mechanical force into displacement. These mechanical elements shown in Figure 15.1. are called force – summing devices. The displacement activated by the force – summing devices is converted into electrical signal using a transducer the resistance strain gauge, capacitive transducer, LVDT, piezolectric and photoelectric etc. The strain gauge and LVDT are dealt in detail below.

Linear variable differential transformer (LVDT)

The linear variable differential transformer (LVDT) is most widely used inductive transducer.. The LVDT consists of a primary coil (P), two identical secondary coils (S_1 and S_2) and a rod shaped magnetic core (A) at the centre. The magnetic core is made of nickel iron alloy and is slotted longitudinally to reduce eddy current loss. The primary coil is connected to an alternating current source. The displacement to be measured is applied to the arm attached to the core. When the core is placed symmetrically with respect to the two recording coils, equal voltages are induced in the two coils. When these two voltages are in phase opposition, the resultant becomes zero. This is called mull position of the core. When the core moves from its null position due to the displacement of the object linked mechanically to it, the voltage induced in the secondary coil (towards which the core has moved) increase, simiultaneously reducing the voltage in the other secondary winding. Thus the amount of voltage change in wither secondary winding is proportional to the amount of movement by the core. The difference of the two voltages induced in the secondary, appears across the output terminals of the transducer giving a measure of the displacement.

Advantages

1. LVDT has rugged construction and can withstand high degree of shock and vibration.
2. It consumes very less power
3. It does not possess sliding contacts and hence no friction.
4. It possesses a high sensitivity
5. It also possesses a very high resolution
6. Its output voltage is practically linear
7. It has excellent repeatability.

Disadvantages

1. LVDT is sensitive to stray magnetic field
2. The receiving instrument must be in a position to operate on a.c. signal.
3. For appreciable differential output, relatively large displacement is required.
4. Its performance is affected by temperature
5. The mass of the core and the frequency or the applied voltage limits the dynamic response of the transducer.

Uses of LVDT

Some of the major muses of LVDT are listed below

1. LVDT can be used for measuring displacements ranging from fraction of a millimetre to a few centimetres.
2. As a primary transducer, it converts the displacement directly into a proportional electric signal,
3. As a secondary transducer, it can be used to measure force with load cell as the primary transducer.
4. It can also be used to measure pressure with bourdon to be as primary transducer.

MODULE 17.

LESSON 27. Basic methods of force measurement.

Basic methods of force measurement.

Force F = MA mass x acceleration

Torque T = Force x length

Methods of force measurement

1. Balancing it against the known gravitational force as a standard mass, either directly or through a system of levers.
2. Measuring the acceleration of a body of known mass to which the unknown force is applied.
3. Balancing it against a magnetic force developed by interaction of a current carrying coil and a magnet.
4. Transducing the force to a fluid pressure and then measuring the pressure.
5. Applying the force to some elastic member and measuring the resulting deflection.
6. Measuring the change in processing of a gyroscope caused by an applied torque related to the measured force.
7. Measuring the change in natural frequency of a wire tensioned by the force.

Method 1 :

Analytical balance, the pendulum scale and the platform scale.

Analytical balance : Beam is designed so that the center of mass is only slightly below the knife edge pivot and thus barely in stable equilibrium. Pendulum scale in a deflector type instrument in which the unknown force is converted to a torque that is then balanced by the torque of a fixed standard mass arranged as a pendulum.

The pendulum scale is a deflection-type instrument in which the unknown force is converted to a torque that is then balanced by the torque of a fixed standard mass arranged as a pendulum. The practical version of this principle utilizes specially shaped sectors and steel tapes to linearize the inherently nonlinear torque-angle relation of a pendulum. The unknown force F_i may be applied directly through a system of levers, such as that shown for the platform scale, to extend the range. An electrical signal proportional to force is easily obtained from any angular-displacement transducer attached to measure the angle θ_o.

The platform scale utilizes a system of levers to allow measurement of large forces in terms of much smaller standard weights. The beam is brought to null by a proper combination of pan weights and adjustment of the poise-weight lever arm along its calibrated scale. The scale can be made self-balancing by adding an electrical displacement pickup for null detection and an amplifier-motor system to position the poise weight to achieve null. Another interesting feature is that if a/b = c/d, the reading of the scale is independent of the location of F_i on the platform. Since this is quite convenient, most commercial scales provide this feature by use of the suspension system shown or others that allow similar results. While analytical balances are used almost exclusively for "weighing" (really determining the mass of) objects or chemical samples, platform and pendulum scales are employed also for force measurements, such as those involved in shaft power determinations with dynamometers. All three instruments are intended mainly for static force measurements.

Commercially available analytical balances may be classified as follows:

Description	Range, g	Resolution, g
Macro analytical	200 – 1,000	10^{-4}
Semimicro analytical	50 – 100	10^{-5}
Micro analytical	10 – 20	10^{-6}
Micro balance	less than 1	10^{-6}
Ultramicro balance	less than 0.01	10^{-7}

Method 2

The use of an accelerometer for force measurement, is of somewhat limited application since the force determined is the resultant force on the mass. Often several unknown forces are acting, and they cannot be separately measured by this method.

Method 3

The The electromagnetic balance (method 3) utilizes a photoelectric (or other displacement sensor) null detector, an amplifier, and a torquing coil in a servo-system to balance the difference the difference between the unknown force F_i and the gravity force on a standard mass. Its advantages relative to mechanical balances are ease of use, less sensitivity to environment, faster response, smaller size, and ease of remote operation. Also, the electric output signal is convenient for continuous recording and/or automatic-control applications. Balances with built-in microprocessors allow even greater convenience, versatility, and speed of use by automating many routine procedures and providing features not formerly feasible. Automatic tare-weight systems subtract container weight from total weight to give net weight when material is placed in the container. Statistical routines allow immediate calculation of mean and standard deviation for a series of weighings. "Counting" of small parts by weighing is speeded by programming the microprocessor to read out the parts by weighing is speeded by programming the microprocessor to read out the parts count directly, rather than the weight. Accurate weighting of live laboratory animals (difficult on an ordinary balance because of animal motion) is facilitated by averaging scale readings over a preselected time. Interfacing the balance to (external to built-in) printers for permanent recording also is eased by the microprocessor.

Method 4

Hydraulic cells are completely filled with oil and usually have a preload pressure of the order of 30 lb/in^2. Application of load increase the oil pressure, which is read on an accurate gage. Electrical pressure transducers can be used to obtain an electrical signal. The cells are very stiff, detecting only a few thousandths of an inch under full load. Capacities to 100,000 lbf are available as standard while special units up to 10 million lbf are obtainable. Accuracy is of the order of 0.1 percent of full scale; resolution is about 0.02 percent. A hydraulic totalizer is available to produce a single pressure equal to the sum of up to 10 individual pressures in multiple-cell systems used for tank weighing, etc.,

The pneumatic load cell shown uses a nozzle-flapper transducer as a igh-gain amplifier in a servoloop. Application of force F_i causes a diaphragm deflection x, which in turn causes an increase in pressure P_o since the nozzle is more nearly shut off. This increase in pressure acting on the diaphragm to its former position. For any constant F_i, the system will come to equilibrium at a specific nozzle opening and corresponding pressure P_o. The static behaviour is given by,

$$(F_i - p_o A)K_d K_n = p_o$$

where $K_d \triangleq$ diaphragm compliance, in/lbf

$K_n \triangleq$ nozzle-flapper gain, (lb/in^2)/in

Solving for p_o, we get

$$p_o = \frac{F_i}{1/(K_d K_n) + A}$$

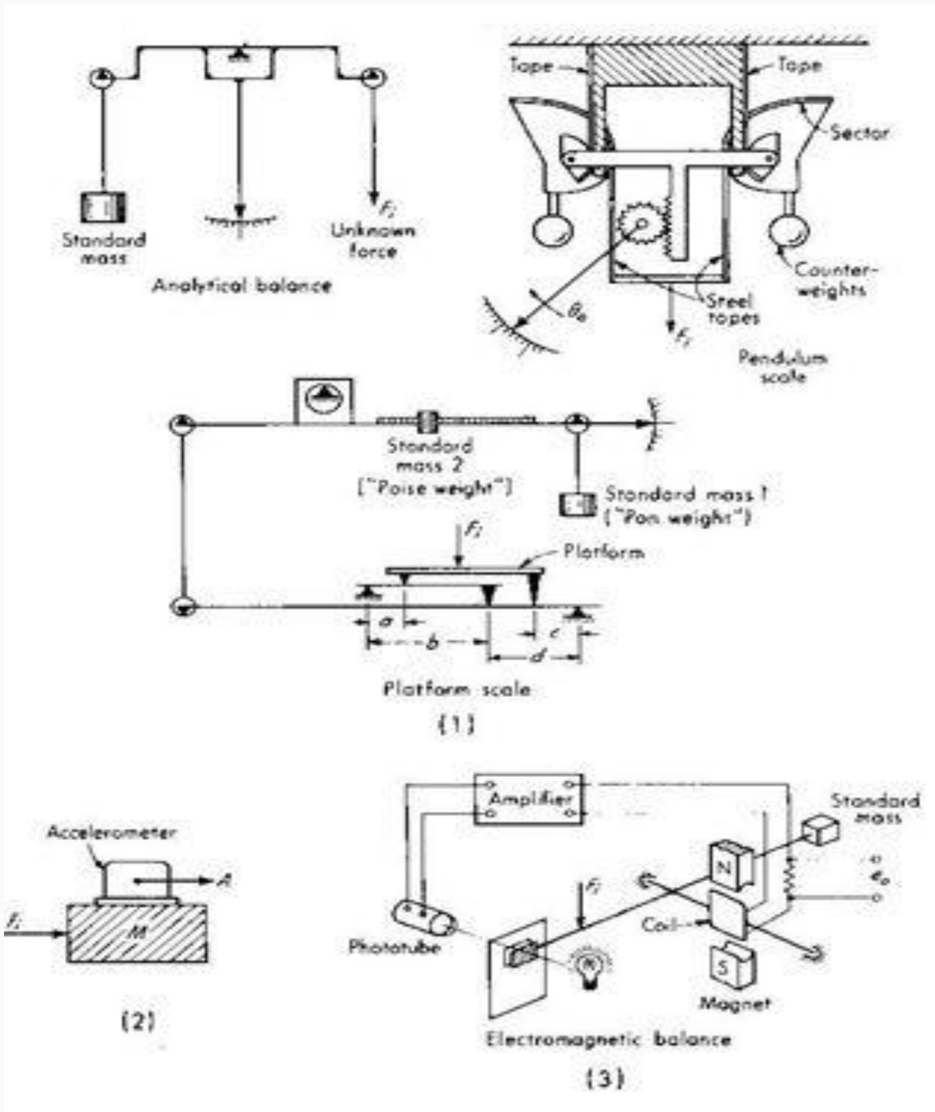
Now K_n is not strictly constant, but varies somewhat with x, leading to a nonlinearity between x and P_o. However, in practice, the product $K_d K_n$ is very large, so that $1/(K_d K_n)$ is made negligible compared with A, which gives

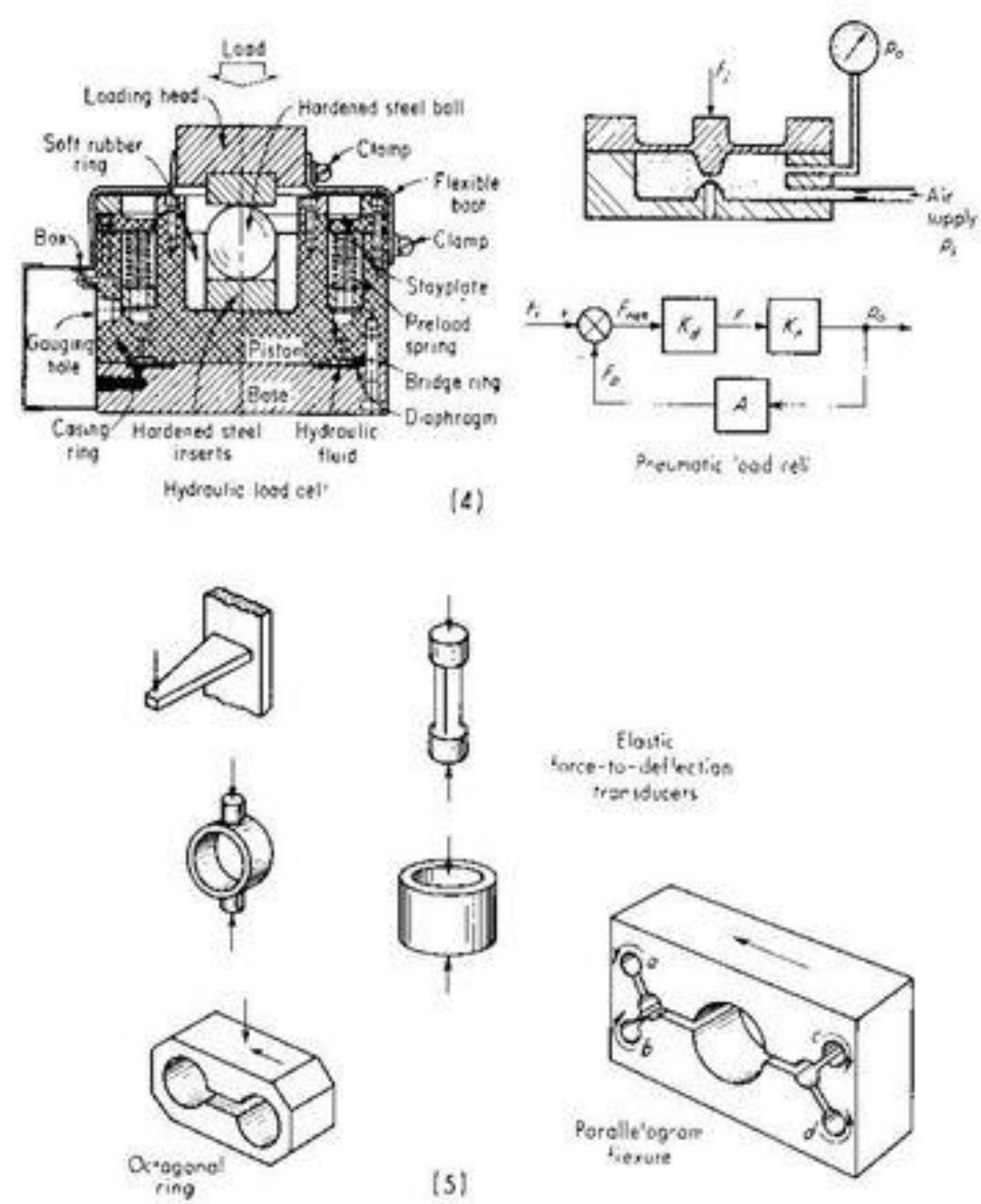
$P_o = F_i / A$

Which is linear since A is constant. As in any feedback system, dynamic instability limits the amount of gain that actually can be used. A typical supply pressure P_s is 60 lb/in^2, and since the maximum value of P_o cannot exceeded P_s , this limits F_i to somewhat less than 60A. A line of commercial pneumatic weighing systems using similar principles (combined with lever/knife-edge methods) is available in standard ranges to 110,000 lbf.

Method 5

While all the previously described force-measuring devices are intended mainly for static or slowly varying loads, the elastic deflection transducers of method 5 are widely used for both static and dynamic loads of frequency content up to many thousand hertz. While all are essentially spring-mass systems with (intentional or unintentional) damping, they differ mainly in the geometric form of “spring” employed and in the displacement transducer used to obtain an electrical signal. The displacement sensed force in terms of strain. Bonded strain gages have been found particularly useful in force measurements with elastic elements. In addition to serving as force-to-deflection transducers, some elastic elements perform the function of resolving vector forces or moments into rectangular components. As an example, the parallelogram flexure of Fig. is extremely rigid (insensitive) to all applied forces and moments except in the direction shown by the arrow. A displacement transducer arranged to measure motion in the sensitive direction thus will measure only that component of an applied vector force which lies along the sensitive axis. Perhaps the action of this flexure may be most easily visualized by considering it as a four-bar linkage with flexure hinges at the thin sections a, b, c, and d. Robotic manufacturing and assembly operations may be improved by adding various sensing functions, such as force sensing at the end of the robot’s arm. A six-axis (x, y, z force; x, y, z torque) sensor, complete with signal-processing hardware and software to interface with the robot controller, is available for connection between the robot arm and end-of-arm tooling . Simpler force sensors can also perform useful functions in robotic systems, sometimes obviating the need for expensive vision systems





Elastic Elements For Force Measurements

Elastic elements are frequently employed to furnish an indication of the magnitude of an applied force through a displacement measurement. The simple spring is an example of this type of force-displacement transducer. In this case the force is given by

$$F = ky$$

Where k is the spring constant and y is the displacement from the equilibrium position. For the simple bar shown in Fig. the force is given by

$$F = \frac{AE}{L} y$$

Where A = cross-sectional area

L = length

E = Young's modulus for the bar material

The deflection of the cantilever beam shown in Fig. is related to the loading force by

$$F = \frac{3\,EI}{L^3} y$$

Where I is the moment of inertia of the beam about the centroidal axis in the direction of the deflection. Any one of the three devices mentioned above is suitable for use as a force transducer provided that accurate means are available for indicating the displacements. The differential transformer , for example, may be useful for measurement of these displacements, as well as capacitance and piezoelectric transducers .

Another elastic device frequently employed for force measurements is the thin ring shown in Fig.. The force-deflection relation for this type of elastic element is

$$F = \frac{16 \qquad EI}{p/2 - 4/pd^3} y$$

where d is the outside ring diameter and I is the moment of inertia about the centroidal axis of the ring section. The proving ring is a ring transducer that employs a sensitive micrometer for the deflection measurement, as shown in Fig.. To obtain a precise measurement, one edge of the micrometer is mounted on a vibrating reed device R, which is plucked to obtain a vibratory motion. The micrometer contact is then moved forward until a noticeable damping of the vibration is observed. Deflection measurements may be made within ±0.00002 in (0.5 μm) with this method. The proving ring is widely used as a calibration standard for large tensile-testing machines.

LESSON 28. Torque measurement on rotating shaft.

Torque, or moment, may be measured by observing the angular deformation of a bar or hollow cylinder, as shown in Fig.. The moment is given by

$$M = \frac{\pi G(r_o^4 - r_i^4)}{2L} \varnothing$$

where G = shear modulus of elasticity

r_i = inside radius

r_o = outside radius

L = length of the cylinder

Ø = angular deflection

strain gages attached at 45° angles as shown will indicate strains of

$$\theta_{45^\circ} = \pm \frac{Mr_o}{\pi G(r_o^4 - r_i^4)}$$

Either the deflection or the strain measurement may be taken as an indication of the applied moment. Multiple strain gages may be installed and connected so that any deformation due to axial or transverse load is canceled out in the final readout circuit. A rather old device for the measurement of torque and dissipation of power from machines is the Prony brake. A schematic diagram is shown in Fig. Wooden blocks are mounted on a flexible band or rope, which is connected to the arm. Some arrangement is provided to tighten the rope to increase the frictional resistance between the blocks and the rotating flywheel of the machine. The torque exerted on the Prony brake is given by

$$T = FL$$

The force F may be measured by conventional platform scales or other methods discussed in the previous paragraphs.

The power dissipated in the brake is calculated from

$$2\pi TN$$

P = -------- hp

33,000

where the torque is in foot-pounds-force and N is the rotational speed in revolutions per minute.

TORQUE MEASUREMENT ON ROTATING SHAFTS

Measurement of the torque carried by a rotating shaft is of considerable interest for its own sake and as a necessary part of shaft power measurements. Torque transmission through a rotating shaft generally involves both a source of power and a sink (power absorber or dissipater), as in figure. Torque measurement may be accomplished by mounting either the source or the sink in bearings (“cradling”) and measuring the reaction force F and arm length L; or else the torque in the shaft itself is measured in terms of the angular twist or strain of the shaft (or a torque sensor coupled into the shaft).

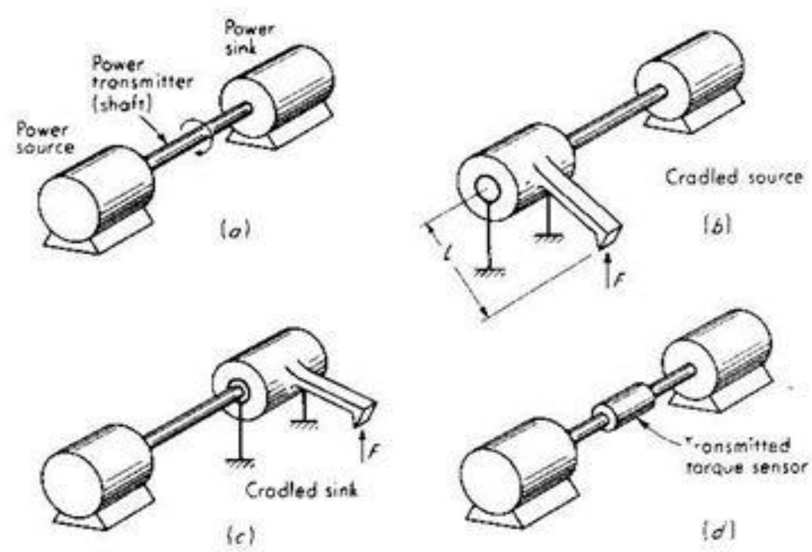


Figure 5.16 Torque measurement of rotating machines.

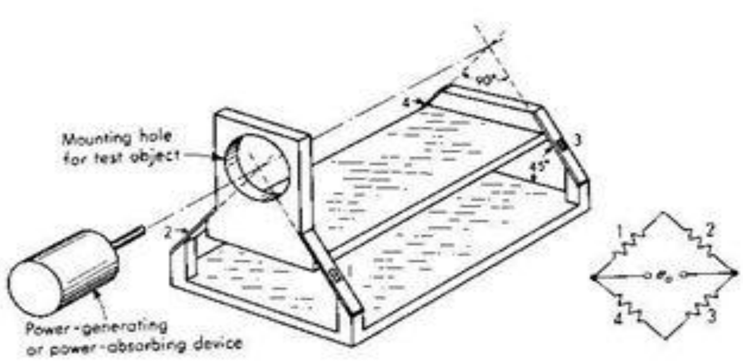


Figure 5.17 Strain-gage torque table.

The cradling concept is the basis of most shaft power dynamometers. These are utilized mainly for measurements of steady power and torque, by using pendulum or platform scales to measure F. A free-body analysis of the cradled member reveals error sources resulting from friction in the cradle bearings, static unbalance of the cradled member, windage torque (if the shaft is rotating) , and forces due to bending and / or stretching of power lines (electric, hydraulic, etc.) attached to the cradled member. To reduce frictional effects and to make possible dynamic torque measurements, the cradle-bearing arrangement may be replaced by a flexure pivot with strain gages to sense torque, as in the above figure. The crossing point of the flexure plates defines the effective axis of rotation of the flexure pivot. Angular deflection under full load is typically less than 0.5^{o}. This type of cross-spring flexure pivot is relatively very stiff in all directions other than the rotational one desired, just as in an ordinary bearing. The strain-gage bridge arrangement also is such as to reduce the effect of all forces other than those related to the torque being measured. Speed-torque curves for motors may be obtained quickly and automatically with such a torque sensor by letting the motor under test accelerate inertia from zero speed up to maximum while measuring speed with a dc tachometer. The torque and speed signals are applied to an XY recorder to give automatically the desired curves.

Even though they require additional equipment to transmit power and signal between rotating shaft and stationary readout, strain-gage torque sensors are very widely used. The following figure shows the basic principle. This arrangement (given accurate gage placement and matched gage characteristics) is temperature-compensated and insensitive to bending or axial stresses. The gages must be precisely at 45^{o} with the shaft axis, and gages 1 and 3 must be diametrically opposite, as must gages 2 and 4. Accurate gage placement is facilitated by the availability of special rosettes in which two gages are precisely oriented on one sheet of backing material. In some cases the shaft already present in the machine to be tested may be fitted with strain gages. In other cases a different shaft or a commercial torquemeter must be used to get the desired sensitivity or other properties. Of the various configuration shown in the figure., one manufacturer uses the hollow cruciform for low-range units and the solid, square shaft for high-range ones. Placement of the gages on a square, rather than round, cross section of the shaft has some advantages. The gages are more easily and accurately located and more firmly bonded on a flat surface. Also, the corners of a square section in torsion are stress-free and thus provide a good location for solder joints between lead wires and gages. These joints are often a source of fatigue failure if located in a high-stress region. Also, for equivalent strain/torque sensitivity, a square shaft is much stiffer in bending than a round one, thus reducing effects of bending forces and raising shaft natural frequencies.

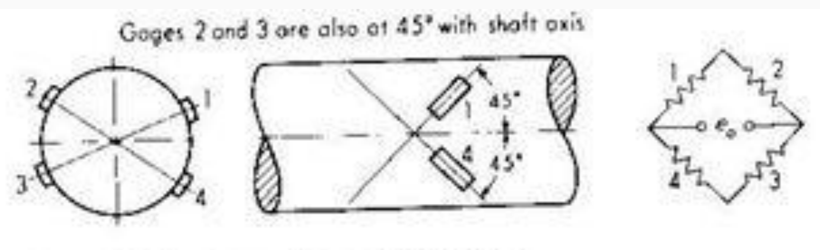
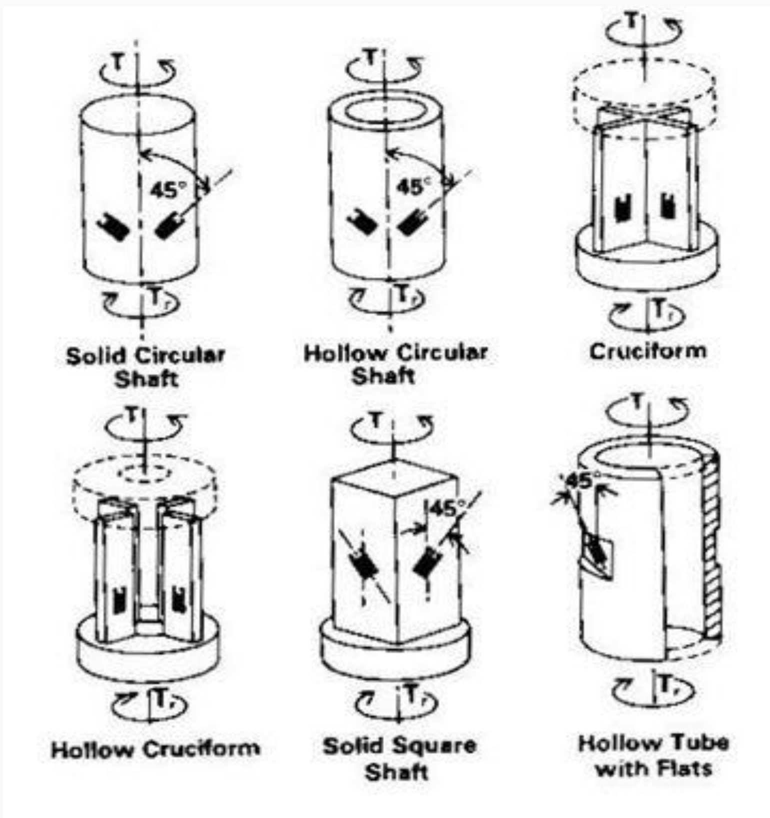


Figure 5.20 Strain-gage torque measurement.

The torque of many machines, such as reciprocating engines, is not smooth even when the machine is running under "steady-state" conditions. If we wish to measure the average torque so as to calculate power, the higher frequency response of strain-gage torque pickups may be somewhat of a liability since the output voltage will follow the cyclic pulsations and some sort of averaging process must be performed to obtain average torque. If exceptional accuracy is not needed, the low-pass filtering effect of a dc meter used to read e_o may be sufficient for this purpose. In the cradled arrangements of the following figure (employed in many commercial dynamometers for engine testing, etc.,) the inertia of the cradled member and the low-frequency response of the platform or pendulum scales used to measure F perform the same averaging function.



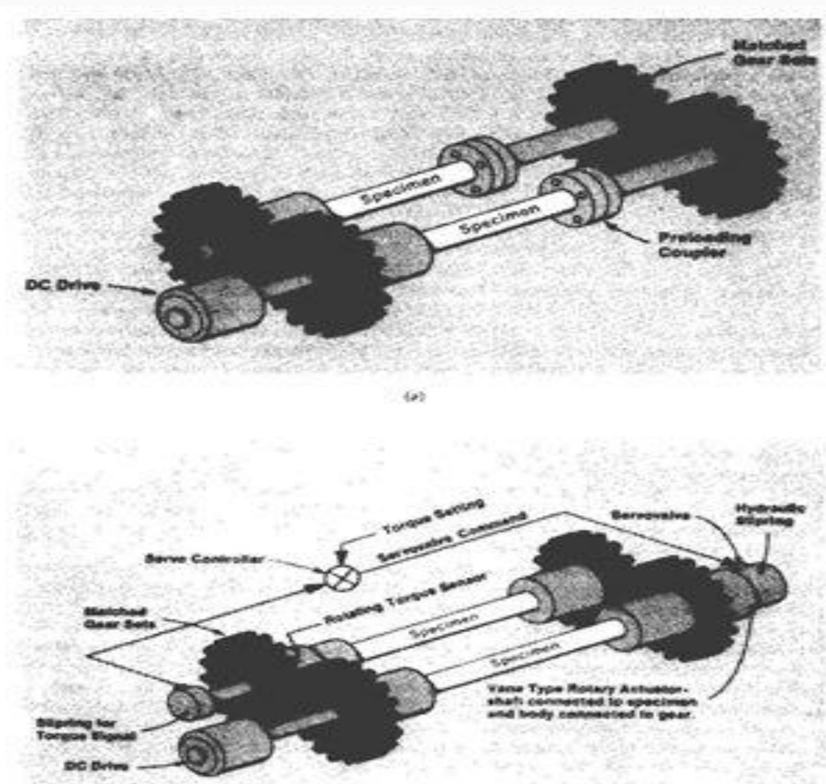
Commercial strain-gage torque sensors are available with built-in slip rings and speed sensors. A family of such devices covers the range 10 oz. in to 3×10^6 in · lbf with full-scale output of about 40 mV. The smaller units may be used at speeds up to 24,000 r/min. Torsional stiffness of the 10 oz·in unit is 112 in·lbf / rad while a 600,000 in · lbf unit has 4.0×10^6 in · lbf. Nonlinearity is 0.1 percent of full scale while temperature effect on zero is 0.002 percent of full scale / F° and temperature effect on sensitivity is 0.002 percent / F° over the range 70 to 170 °F .

The dynamic response of elastic deflection torque transducers is essentially slightly damped second-order, with the natural frequency usually determined by the stiffness of the transducer and the inertia of the parts connected at either end. Damping of the transducers themselves usually is not attempted, and any damping present is due to bearing friction, windage etc., of the complete test setup.

A good example of the application of torque measurement (and several other measurement system concepts of general interest) is found in a dynamic test system used in the design and development of front-wheel-drive components for automobiles (see Fig. 5.22 through 5.24). An alternative to conventional dynamometer loading systems, called the four-square principle, is utilized here to conserve energy and allow fast-response testing. Rather than connecting the shaft test specimens between a power source and a power sink, as in fig. 5.16 (an arrangement which wastes all the shaft-transmitted power sink and couples large, slow responding inertias to the specimen), the four-square principle (Fig. 5.22 a) includes a "looked-in" torque into the system by counter-twisting the flanges of a preloading coupler at assembly. Now the geared assembly can be driven at any desired speed by, say, a dc motor drive, and the shaft specimens will be subjected to the desired torque/speed conditions, but the drive supplies only friction losses (plus accelerating power if speed is changed), not the shaft-transmitted power(product of locked-in torque and shaft speed).

The basic four-square principle just described has been widely utilized for some time; however, several new features were needed in the present application. It was desired to be able to maintain a certain level of locked-in torque in the face of inevitable component wear, creep, slippage, or fatigue, which all cause a progressive "relaxation" of torque with time in the basic four-square system. Also, it was necessary to dynamically vary the locked-in torque in response to electrical commands. By replacing the mechanical preloading coupler with a rotary hydraulic actuator , measuring shaft torque with a rotating torque sensor, and driving the actuator from an electro hydraulic servo valve responding to the error between commanded and measured torque, the desired features are achieved.

To simulate actual driving conditions, the testing machine includes means for subjecting the front-wheel-drive shafts to "jounce" (up and down) motions.



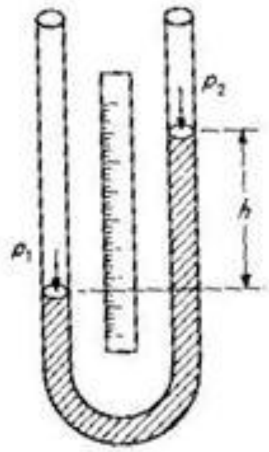
LESSON 29. BASIC METHODS OF PRESSURE MEASUREMENT

Since pressure usually can be easily transduced to force by allowing it to act on a known area, the basic methods of measuring force and pressure are essentially the same, except for the high-vacuum region where a variety of special methods not directly related to force measurement are necessary. These special methods are described in the section on vacuum measurement. Other than the special vacuum techniques, most pressure measurement is based on comparison with known deadweights acting on known areas or on the deflection of elastic elements subjected to the unknown pressure. The deadweight methods are exemplified by manometers and piston gages while the elastic deflection devices take many different forms.

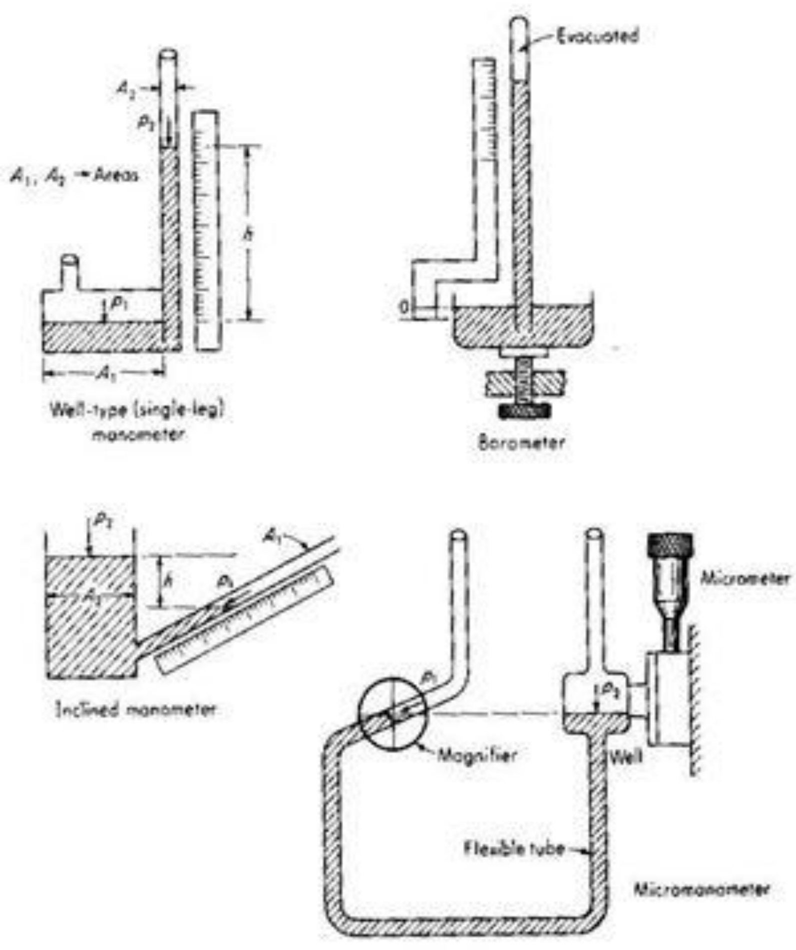
The manometer in its various forms is closely related to the piston gage, since both are based on the comparison of the unknown pressure force with the gravity force on a known mass. The manometer differs, however, in that it is self – balancing, is a deflection rather than a null instrument, and has continuous rather than stepwise output. The accuracies of deadweight gages and manometers of similar ranges are quite comparable; however, manometers become unwieldy at high pressures because of the long liquid columns involved. The U-tube manometer usually is considered the basic form and has the following relation between input and output for static conditions:

$$h = \frac{p_1 - p_2}{pg}$$

where g Δ local gravity and p Δ mass density of manometer fluid. If p_2 is atmospheric pressure, than h is a direct measure of p_1 as a gage pressure. Note that the cross-sectional area of the tubing (even if not uniform) has no effect. At a given location (given value of g) the sensitivity depends on only the density of the manometer fluid. Water and mercury are the most commonly used fluids. To realize the high accuracy possible with manometers, often a number of corrections must be applied. When visual reading of height h is employed, the engraved scale’s temperature expansion must be considered. The variation of p with temperature for the manometer fluid used must be corrected and the local value of g determined. Additional sources of error are found in the nonverticality of the tubes and the difficulty in reading h because of the meniscus formed by capillarity. Considerable care must be exercised in order to keep inaccuracies as small as 0.01 mm Hg for the overall measurement.



A number of practically useful variations on the basic manometer principle are shown in the following. . The cistern or well-type manometer is widely utilized because of its convenience in requiring reading of only a single leg. The well area is made very large compared with the tube; thus the zero level moves very little when pressure is applied. Even this small error is compensated by suitably distorting the length scale. However, such an arrangement, unlike a U tube, is sensitive to non uniformity of the tube cross-sectional area and thus is considered some-what less accurate.

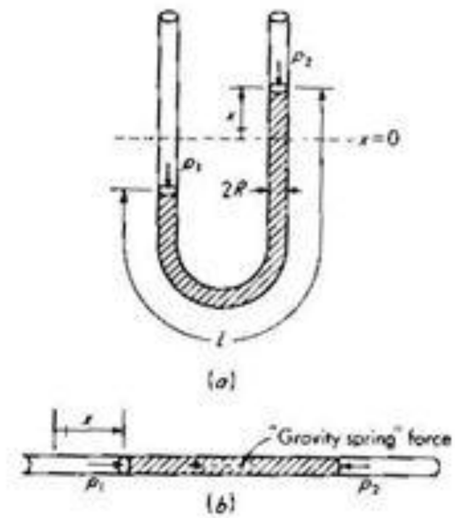


Given that manometers inherently measure the pressure difference between the two ends of the liquid column, if one end is at zero absolute pressure, then h is an indication of absolute pressure. This is the principle of the barometer . Although it is a “single-leg” instrument, high accuracy is achieved by setting the zero level of the well at the zero level of the scale before each reading is taken. The pressure in the evacuated portion of the barometer is not really absolute zero, but rather the vapor pressure of the filling fluid, mercury, at ambient temperature. This is about 10^{-4} lb/in^2 absolute at 70^{o}F and usually is negligible as a correction.

To increase sensitivity, the manometer may be titled with respect to gravity, thus giving a greater motion of liquid along the tube for a given vertical-height change. The inclined manometer (draft gage) exemplifies this principle. Since this is a single-leg device, the calibrated scale is corrected for the slight changes in well level so that rezeroing of the scale for each reading is not required.

The accurate measurement of extremely small pressure differences is accomplished with the micromanometer, a variation on the inclined-manometer principle. In the above figure. the instrument is initially adjusted so that when $p_1 = p_2$, the meniscus in the inclined tube is located at a reference point given by a fixed hairline viewed through a magnifier. The reading of the micrometer used to adjust well height is now noted. Application of the unknown pressure difference causes the meniscus to move off the hairline, but it can be restored to its initial position by raising or lowering the well with the micrometer. The difference in initial and final micrometer readings gives the height change h and thus the pressure. Instruments using water as the working fluid and having a range of either 10 or 20 in of water can be read to about 0.001 in of water. In another instrument in which the inclined tube (rather than the well) is moved and which uses butyl alcohol as the working fluid, the range is 2 in of alcohol, and readability is 0.0002 in. This corresponds to a resolution of 6 x 10^{-6} lb/in^2.

While manometers usually are read visually by a human operator, various schemes for rapid and accurate automatic readout are available, mainly for calibration and standards work using gaseous media. The sonar manometer employs a piezoelectric transducer at the bottom of each 1.5-in-diameter glass tube to launch ultrasonic pulses, which travel up through the mercury columns, are reflected at the meniscus, and return to the bottom to be received by the transducers. The pulse from the shorter column turns on a digital counter, while that from the longer one turns it off. Thus a digital reading is obtained that is proportional to the difference in column height and thus to pressure. Resolution is 0.0003 in Hg, and accuracy is 0.001 in Hg or 0.003 percent of reading, whichever is greater. Since temperature effects on sonic velocity and column length cause additive errors, a feedback control system keeps instrument temperature at 95 ± 0.05^{o}F.



Another instrument employs two large mercury cisterns (one fixed, one vertically movable by an electromechanical servosystem) connected by flexible tubing to create a U-tube manometer. Each cistern has a capacitor formed by a metal plate, the mercury surface, and a small air gap between them. The two capacitors are connected in an electric circuit which exhibits a null reading when they are not. This error voltage causes the servosystem to drive the movable cistern to an elevation where balance is again achieved. A digital counter on the servosystem motor shaft reads out position to the nearest 0.0001 in Hg. System accuracy is ± 0.0003 in Hg ±0.003 percent of reading. For manometers such as the two above, accessory automatic systems for generating and regulating the pressures of the gaseous (usually air or nitrogen) calibration media usually are available.

MODULE 18.

LESSON 30. Temperature measurement by mechanical effect and electrical effect

Temperature measurement by mechanical effects

Several temperature measurement devices may be classified as mechanically operative. In this sense we shall be concerned with those devices operating on the basis of a change in mechanical dimension with a change in temperature.

The liquid in glass thermometer is one of the most common types of temperature measurement devices. The construction details of such an instrument are shown in the following figure-. A relatively large bulb at the lower portion of the thermometer holds the major portion of the liquid, which expands when heated and rises in the capillary tube, upon which are etched appropriate scale markings. At the top of the capillary tube another bulb is placed to provide a safety feature in case the temperature range of the thermometer is inadvertently exceeded. Alcohol and mercury are the most commonly used liquids. Alcohol has the advantage that it has a higher coefficient of expansion than mercury, but it is limited to low temperature measurements because it tends to boil away at high temperatures. Mercury cannot be used below its freezing point of –38.78°F (-37.8°C). the size of the capillary depends on the size of the sensing bulb, the liquid, and the desired temperature range for the thermometer.

In operation, the bulb of the liquid-in-glass thermometer is exposed to the environment whose temperature is to be measured. A rise in temperature causes the liquid to expand in the bulb and rise in the capillary, thereby indicating the temperature. It is important to note that the expansion registered by the thermometer is the difference between the expansion of the liquid and the expansion of the glass. The difference is a function not only of the heat transfer to the bulb from the environment, but also of the heat conducted into the bulb from the stem; the more the stem conduction relative to the heat transfer from the environment, the larger the error. To account for such conduction effects the thermometer is usually calibrated for a certain specified depth of immersion. High grade mercury in glass thermometers have the temperature scale markings engraved on the glass along with a mark which designates the proper depth of immersion. Very precise mercury-in-glass thermometers may be obtained from the National Bureau of Standards with calibration information for each thermometer.

Mercury-in-glass thermometers are generally applicable upto about 600°F (315°C), but their range may be extended to 1000°F (538°C) by filling the space above the mercury with a gas like nitrogen. This increases the pressure on the mercury, raises its boiling point, and thereby permits the use of the thermometer at higher temperatures.

A very widely used method of temperature measurement is the bimetallic strip. Two pieces of metal with different coefficients of thermal expansion are bonded together to form the device shown in Fig. When the strip is subjected to a temperature higher than the bonding temperature, it will bend in one direction; when it is subjected to a temperature lower than the bonding temperature, it will bend in the other direction. Eskin and Fritze have given calculation methods for bimetallic strips. The radius of curvature r may be calculated as

$$r = \frac{T[3(1+m)^2 + (1+mn)[m^2 + (1/mn)]]}{6(x_2-x_1)(T-T_0)(1+m)^2}$$

where t = combined thickness of the bonded strip

m = ratio of thickness of low- to high-expansion materials

n = ratio of moduli of elasticity of low- to high- expansion materials

x_1 = lower coefficient of expansion

a_2 = higher coefficient of expansion

T = temperature

T_0 = initial bonding temperature

Fluid expansion thermometers represent one of the most economical, versatile, and widely used devices for industrial temperature measurement applications. The principle of operation is indicated in Fig. A bulb containing a liquid, gas, or vapor is immersed in the environment. The bulb is connected by means of a capillary tube to some type of pressure measuring device, such as the Bourdon gage shown. An increase in temperature causes the liquid or gas to expand, thereby increasing the pressure on the gage; the pressure is thus taken as an indication of the temperature. The entire system consisting of the bulb, capillary, and gage may be calibrated directly. It is clear that the temperature of the capillary tube may influence the reading of the device because some of the volume of fluid is contained therein. If an equilibrium mixture of liquid and vapor is used in the bulb, however, this problem may be alleviated, provided that the bulb temperature is always higher than the capillary tube temperature. In this circumstance the fluid in the capillary will always be in a subcooled liquid state, while the pressure will be uniquely specified for each temperature in the equilibrium mixture contained in the bulb.

Capillary tubes as long as 200 ft (60 m) may be used with fluid-expansion thermometers. The transient response is primarily dependent on the bulb size and the thermal properties of the enclosed fluid. Highest response may be achieved by using a small bulb connected to some type of electric pressure transducer through a short capillary.

Temperature measurement by electrical effects

Electrical methods of temperature measurement are very convenient because they furnish a signal that is easily detected, amplified, or used for control purposes. In addition, they are usually quite accurate when properly calibrated and compensated.

Electrical resistance thermometer

One quite accurate method of temperature measurement is the electrical resistance thermometer. It consists of some type of resistive element, which is exposed to the temperature to be measured. The temperature is indicated through a measurement of the change in resistance of the element. Several types of materials may be used as resistive elements, and their characteristics are given in Table . The linear temperature coefficient of resistance x is defined by

$$x = \frac{R_2 - R_1}{R_1T_2 - R_2T_1}$$

Where R_2 and R_1 are the resistances of the material at temperatures T_2 and T_1, respectively. The relationship in Equation is usually applied over a narrow temperature range such that the variation of resistance with temperature approximates a linear relation. For wider temperature ranges the resistance of the material is usually expressed by a quadratic relation

$$R = R_0\ (1+aT + bT^2)$$

Where R = resistance at temperature T

R_0 = resistance at 0°F

a,b = experimentally determined constants

It may be noted that the platinum resistance thermometer is used for the International Temperature Scale between the oxygen point and the antimony point, as described in Chap.

Various methods are employed for construction of resistance thermometers, depending on the application. In all cases care must be taken to ensure that the resistance wire is free of mechanical stresses and so mounted that moisture cannot come in contact with the wire and influence the measurement.

The resistance measurement may be performed with some type of bridge circuit, as described in Chap.4. For steady-state measurements a null condition will suffice, while transient measurements will usually require the use of a deflection bridge. One of the primary sources of error in the electrical resistance thermometer is the effect of the resistance of the leads which connect the element to the bridge circuit. Several arrangements may be used to correct for this effect, as shown in Figure. The Siemen's three lead arrangement is the simplest type of corrective circuit. At balance conditions the center lead carries no current and the effect of the resistance of the other two leads is canceled out. The Callender four-lead arrangement solves the problem by inserting two additional lead wires in the adjustable leg of the bridge so that the effect of the lead wires on the resistance thermometer is canceled out. The floating-potential arrangement in Fig.--- is the same as the Siemen's connection, but an extra lead is inserted. This extra lead may be used to check the equality of lead resistance. The thermometer reading may be taken in the position shown, followed by additional readings with the two right and left leads interchanged, respectively. Through this interchange procedure the best average reading may be obtained and the lead error minimized.

Thermoelectric effects

The most common electrical method of temperature measurement uses the thermocouple. When two dissimilar metals are joined together as in fig, an emf will exist between the two points A and B, which is primarily a function of the junction temperature. This phenomenon is called the seebeck effect. If the two materials are connected to an external circuit in such a way that a current is drawn, the emf may be altered slightly owing to a phenomenon called the Peltier effect. Further, if a temperature gradient exists along either or both of the materials, the junction emf may undergo an additional slight alteration. This is called the Thomson effect. There are, then, three emfs present in a thermoelectric circuit; the seebeck emf, caused by the junction of dissimilar metals; the Peltier emf, caused by a current flow in the circuit; and the Thomson emf, which results from a temperature gradient in the materials. The seebeck emf is of prime concern since it is dependent on junction temperature. If the emf generated at the junction of two dissimilar metals is carefully measured as a function of temperature, then such as junction may be utilized for the measurement of temperature. The main problem arises when one attempts to measure the potential. When the two dissimilar materials are connected to a measuring device, there will be another thermal emf generated at the junction of the materials and the connecting wires to the voltage measuring instrument. This emf will be dependent on the temperature of the connection, and provision must be made to take account of this additional potential.

Two rules are available for analysis of thermoelectric circuits :

1. If a third metal is connected in the circuit as shown in fig., the net emf of the circuit is not affected as long as the new connections are at the time temperature. This statement may be proved with the aid of the second law of thermodynamics and is known as the law of intermediate metals.

2. Consider the arrangements shown in fig. The simple thermocouple circuits are constructed of the same materials but operate between different temperature limits. The circuit in Fig. develops an emf of E_1 between temperatures T_1 and T_2; the circuit in Fig. develops an emf of E_2 between temperatures T_2 and T_3. The law of intermediate temperatures states that this same circuit will develop an emf of $E_3 = E_1 + E_2$ when operating between temperatures T_1 and T_3, as shown in fig.

It may be observed that all thermocouple circuit must involve at least two junctions. If the temperature of one junction is known, then the temperature of the other junction may be easily calculated using the thermoelectric properties of the materials. The known temperature is called the reference temperature. A common arrangement for establishing the reference temperature is the ice bath shown in fig. An equilibrium mixture of ice and air saturated distilled water at standard atmospheric pressure produces a known temperature of 32°F. when the mixture is contained in a Dewar flask, it may be maintained for extended periods of time. Note that the arrangement in Fig maintains both thermocouple wires at a reference temperature of 32°F, whereas the arrangement in Fig. maintains only one at the reference temperature. The system in Fig. would be necessary if the binding posts at the voltage measuring instrument were at different temperatures, while the connection in Fig would be satisfactory if the binding posts were at the same temperature. To be effective the system in Fig. must have copper binding posts; the binding posts and leads must be of the same material.

It is common to express the thermoelectric emf in terms of potential generated with a reference junction at 32°F. Standard thermocouple tables have been prepared on this basis, and a summary of the output characteristics of the most common thermocouple combinations is given in Table . These data are shown graphically in, alongwith the behaviour of some of the more exotic thermocouple materials. The output voltage E of a simple thermocouple circuit is usually written in the form

$$E = AT + \tfrac{1}{2} BT^2 + \tfrac{1}{2} CT^3$$

Where T is the temperature in degrees Celsius and E is based on a reference junction temperature of 0°C. the constants, A, B and C are dependent on the thermocouple material. Powell gives an extensive discussion of the manufacture of materials for thermocouple use, inhomogeneity ranges and power series relationships for thermoelectric voltages of various standard thermocouples.

The sensitivity, or thermoelectric power, of a thermocouple is given by

$$S = \frac{dE}{dT} = A + BT + CT^2$$

Reference Junction Considerations

For the most precise work, reference junctions should be kept in a triple-point of water apparatus whose temperature is 0.01±0.0005°C. Such accuracy is rarely needed, and an ice bath is used much more commonly. A carefully made ice bath is reproducible to about 0.001C°, but a poorly made one may have an error of 1C°. Figure shows one method of constructing an ice-bath reference junction. The main sources of error are insufficient immersion length and an excessive amount of water in the bottom of the flask. Automatic ice baths that use the Peltier cooling effect as the refrigerator, rather than relying on externally supplied ice (which must be continually replenished), are available with an accuracy of 0.05C°. These systems use the expansion of freezing water in a sealed bellows as the temperature sensing element that signals the Peltier refrigerator when to turn on or off by displacing a microswitch.

Since low-power heating is obtained more easily than low-power cooling, some reference junctions are designed to operate at a fixed temperature higher than any expected ambient. A feedback system operates an electric heating element to maintain a constant and known temperature in an enclosure containing the reference junctions. Since the reference junction is not at 32°F, the thermocouple-circuit net voltage must be corrected by adding the reference junction voltage before the measuring junction temperature can be found. This correction is, however, a constant.

Electrical resistance sensors

The electrical resistance of various materials in a reproducible manner with temp, form the basis of the temp. sensing method. Resistance temperature detector (RTD) has come into use. Semiconductor types appeared later and have been given the generic name thermistor. Any of the various established techniques of resistance measurement may be employed to measure the resistance of these devices, with both bridge and 'ohmmeter' methods being common.

MODULE 19.

LESSON 31. Strain gauge-metallic sensing elements-Unbonded strain gauges-problem associated with strain gauge installation. Type of strain gauge.

STRAIN GAGE

The strain gage is an example of a passive transducer that converts a mechanical displacement into a change of resistance. A strain gage is a thin, wafer-like device that can be attached (bonded) to a variety of materials to measure applied strain. Metallic strain gages are manufactured from small diameter resistance wire, such as Constantan, or etched from thin foil sheets. The resistance of the wire or metal foil changes with length as the material to which gage is attached undergoes tension or compression. This change in resistance is proportional to the applied strain and is measured with a specially adapted Wheat-stone bridge.

The sensitivity of a strain gage is described in terms of a characteristic called the gage factor, K, defined as the unit change in resistance per unit change in length, or

$$\text{gage factor } K = \frac{\Delta R/R}{\Delta l/l}$$

where K = gage factor

R = nominal gage resistance

DR = change in gage resistance

l = normal specimen length (unstressed condition)

Dl = change in specimen length

The term Dl/l in the denominator of Eq.(11-1) is the strain s, so that Eq.(11-1) can be written as

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

where s is the strain in the lateral direction.

The resistance change DR of a conductor with length l can be calculated by using the expression for the resistance of a conductor of uniform cross section :

Length p x l

$$R = p \frac{}{\text{Area}} = \frac{}{(\pi/4)d^2}$$

Where p = specific resistance of the conductor material

l = length of the conductor

d = diameter of the conductor

Tension on the conductor causes an increase Dl in its length and a simultaneous decrease Dd in its diameter. The resistance of the conductor then changes to

$$R_S = p \frac{(l + \Delta l)}{(\pi/4)(d-\Delta d)^2} = p \frac{l(l + \Delta l/l)}{(\pi/4)d^2\,(1-2\Delta d/d)}$$

Equation ---- may be simplified by using Poisson's ratio, μ, defined as the ratio of strain in the lateral direction to strain in the axial direction. Therefore,

$$\mu = \frac{\Delta d/d}{\Delta l/l}$$

substitution yields

$$R_S = p \frac{l}{(\pi/4)d^2} \left(\frac{(l + \Delta l/l)}{(1-2\mu\,\Delta l/l)} \right.$$

which can be simplified to

$$R_S = R + \Delta R = R\,[1+(1+2\mu)\Delta l/l]$$

The increment of resistance DR as compared to the increment of length Dl can then be expressed in terms of the gage factor K where

$$K = \frac{\Delta R/R}{\Delta\, l/l} = 1 + 2\mu$$

Poissons' ratio for most metals lies in the range of 0.25 to 0.35, and the gage factor would then be on the order of 1.5 to 1.7.

For strain gage applications, a high sensitivity is very desirable. A large gage factor means a relatively large resistance change, which can be more easily measured than a small resistance change. For Constantan wire, K is about 2, whereas Alloy 479 gives a K value of about 4.It is interesting carry out a simple calculation to find out what effect an applied stress has on the resistance change of a strain gage. Hooke's law gives the relationship between stress and strain for a linear stress strain curve, in terms of the modulus of elasticity of the material under tension. Defining stress as the applied force per unit area and strain as the elongation of the stressed member per unit length, Hooke's law is written as

$$\sigma = \frac{S}{E}$$

Where s =strain, Δl/l (no units)

S = stress (kg/cm^2)

E = Young's modulus (kg/cm^2)

Metallic sensing elements

Metallic strain gages are formed from thin resistance wire or etched from thin sheets of metal foil. 1. Wire gages are generally small in size, are subject to minimal leakage, and can be used in high temperature applications. 2. Foil elements are somewhat larger in size and are more stable than wire gages. They can be used under extreme temperature conditions and under prolonged loading, and they dissipate self-induced heat easily.

Various resistance materials have been developed for use in wire and foil gages. Some of these are described in the following paragraphs.

Constantan is a copper-nickel alloy with a low temperature coefficient. Constantan is commonly found in gages that are used in dynamic strain measurements, where alternating strain levels do not exceed ± 1,500 μcm/cm. Operating temperature limits are from 10°C to 200°C.

Nichrome V is a nickel-chrome alloy used for static strain measurements to 375°C. with temperature compensation, the alloy may be used for static measurements to 650°C and dynamic measurements to 1000°C.

Dynaloy is a nickel-iron alloy with a high gage factor and a high resistance to fatigue. This material is used in dynamic strain applications when high temperature sensitivity can be tolerated. The temperature range of dynaloy gages is generally limited by the carrier materials and the bonding cement.

Stabiloy is a modified nickel-chrome alloy with a wide temperature compensation range. These gages have excellent stability from cryogenic temperatures to approimately 350°C and good fatigue life.

Platinum tungsten alloys offer excellent stability and high resistance to fatigue at elevated temperatures. These gages are recommended for static tests to 700°C and dynamic measurements to 850°C. because the material has a relatively large temperature coefficient, some form of temperature compensation must be used to correct this error.

Semiconductor strain gages are often used in high-output transducers such as load cells. These gages have very high sensitivities, with gage factors from 50 to 200. They are, however, sensitive to temperature fluctuations and often behave in a nonlinear manner.

The size of the finished gage, and the manner in which the wire or foil pattern is arranged, varies with the application. Some bonded gages can be as small as 1/8 in by 1/8 in., although they are generally somewhat larger, and are manufactured to a maximum size of 1 in. long by ½ in. wide. In the usual application, the strain gage is cemented to the structure whose strain is to be measured. The problem of providing a good bonding between the gage and the structure is very difficult. The adhesive material must hold the gage firmly to the structure, yet it must have sufficient elasticity to give under strain without losing its adhesive properties. The adhesive should also be resistant to temperature, humidity and other environmental conditions.

Gage configuration

The shape of the sensing element is selected according to the strain to be measured : uniaxial, biaxial, or multidirectional. Uniaxial applications most often use long, narrow sensing elements, as in fig. To maximize the strain sensing material in the direction of interest. End loops are few and short, so that sensitivity to transverse strains is low. Gage length is selected according to the strain field to be investigated. For most strain measurements, the 6mm gage length offers good performance and easy installation.

Simultaneous measurement of strains in more than one direction can be accomplished by placing single-element gages at the proper locations. However, to simplify this task and provide greater accuracy, multielement, or rosette, gages are available.

Two-element rosettes, shown in fig, are often used in force transducers. The gages are wired in a Wheatstone bridge circuit to provide maximum output. For stress analysis, the axial and transverse elements may have different resistances that can be so selected that the combined output is proportional to stress while the output of the axial element alone is proportion to strain. Three element rosettes are often used to determine the direction and magnitude of principal strains resulting from complex structural loading. The most popular types have 45° - or 60°-angular displacements between the sensing elements, as shown in fig. The 60° rosettes are used when the direction of the principal strains is unknown. The 45° rosettes provide greater angular resolutions and are normally used when the directions of the principal strains are known.

Unbonded strain gage

The unbonded straing gage consists of a stationary frame and an armature that is supported in the centre of the frame. The armature can move only in one direction. Its travel in that direction is limited by four filaments of strain sensitive wire, wound between rigid insulators that are mounted on the frame and on the armature. The filaments are of equal length and arranged as shown in fig.

When an external force is applied to the strain gage, the armature moves in the direction indicated. Elements A and D increase in length, whereas elements B and C decrease in length. The resistance change of the four filaments is proportional to their change in length, and this change can be measured with a Wheat-stone bridge, as shown in fig. The unbalance current, indicated by the current meter, is calibrated to read the magnitude of the displacement of the armature.

The unbonded strain gage transducer can be constructed in a variety of configurations, depending on the required use. Its principal use is a displacement transducer. A linkage pin can be attached to the armature in order to measure displacement directly. The unit of Fig. allows an armature displacement of 0.004 cm on each side of its center position. Using the same construction, the unit will functions as a dynamometer, capable of measuring force. Depending on the number of turns and the diameter of the strain wires, the transducer will measure forces from ±40 g to ±2 kg, full scale.

The transducer becomes a pressure pickup when its armature is connected to metallic bellows or diaphragm. When a bellows is used, force on the end of the bellows is transmitted by a pin to the armature, and the unit functions as a dynamometer. By applying pressure to one side of the bellows and venting the other side to the atmosphere, gage pressures may be read. If the bellows is evacuated and sealed, absolute pressure is measured.

Another modification is provided by two pressure connections, one to each side of the bellows or diaphragm, for the measurement of differential pressure.

When strain gages are mounted as a specimen, two notes of cauha should be followed : (1) the surface should be absolutely cleaning with emery cloth followed by acetone is usually satisfactory, (2) Sufficient time must be allowed for the cement to dry and harden completely. Even though the cement is dry around the edge of the gage, it may still be wet under the gage. If possible, 24h should be allowed for drying at room temperature. Drying time may be reduced for higher temperatures.

Several different cements are available for mounting strain gages. These cements are discussed in Ref.(2) alongwith rather detailed instructions for mounting the various types of gages. The interested reader should consult this reference and the literature of various manufacturers of strain gages for more information.

Problems associated with strain-gage installations generally fall into three categories : (1) temperature effects (2) moisture effects and (3) wiring problems. It is assumed that the gage is properly mounted. Temperature problems arise because of differential thermal expansion between the resistance element and the material to which it is bonded. Semiconductor gages offer the advantage that they have a lower expansion coefficient than either wire or foil gages. In addition to the expansion problem, there is a change in resistance of the gage with temperature, which must be adequately compensated for. We shall see how this compensation is performed in a subsequent paragraph. Moisture absorption by the paper and cement can change the electrical resistance between the gage and the ground potential and thus affect the output-resistance readings. Methods of moisture proofing are discussed in Refs. Wiring problems are those situations that arise because of faulty connections between the gage-resistance element and the external readout circuit. These problems may develop from poorly soldered connections or from inflexible wiring, which may pull the gage loose from the test specimen or break the gage altogether. Proper wiring practices are discussed in Res.

Electrical resistance strain gages cannot be easily calibrated because once they are attached to a calibration work piece, removal cannot be made without destroying the gage. In practice, then, the gage factor is taken as the value specified by the manufacturer and a semi calibration effected by checking the bridge measurement and readout system.

In the design and construction of machines and structures, the strength of the material plays a *very* important role. A theoretical knowledge of this property is essential to estimate whether the mechanical components can carry the loads demanded of them, without excessive deformation or failure. These load-carrying abilities are normally characterized in terms of stress, which is defined as the force experienced per unit area, and is expressed in pressure units. Stress itself cannot be measured directly and is normally deduced from the changes in mechanical dimensions and the applied load. The mechan- ical deformation formed due to stress is measured with strain-gauge elements. The relationship betweer load and elongation is characterized in terms of strain which is defined as the change *L1t* in length *t* pel unit length and is expressed as At */t* in microstrains.

The precise measurement of the parameter 'strain' is an important aspect in measurement engineering as it is very often encountered in many fields of engineering and technology, especially in experimenta stress analysis. Further, a large number of mechanical and physical parameters can be related witt devices operating on the principle of strain measurement.

The stress to strain relationship in a simple tension or compression test is expressed as linear so long as the stress is kept below the elastic limit.

The deformation of a strut under simple ten- 9 si on loading is given in Fig.. The length is <*n increased by L1x* and the cross-section is de- creased by *L1y* x Liz. If the strain is measured either of the planes perpendicular to the applied load, a strain with a lesser magnitude and with opposite sign is developed in this plane and this effect is known as Poisson's effect. The magni- tude is expressed as the Poisson's ratio *v* which

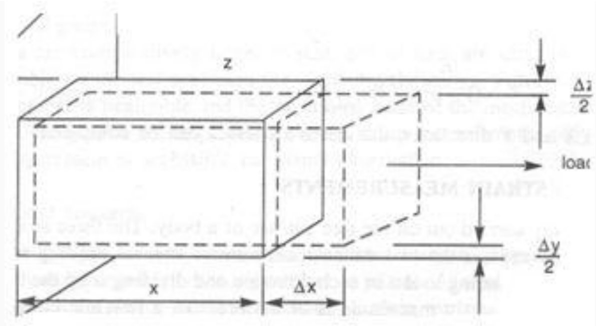


Fig. Deformation of a **strut**

FACTORS AFFECTING STRAIN MEASUREMENTS

Strain measurements are normally carried out on the free surface of a body. The three strains *Ex*' *E* " and *El* defined in Sec. 5.1 adequately express the two-dimensional state of stresses existing on the surface. By measuring displacements corresponding to 11 *x* in each direction and dividing it by the original length *x,* the strain can be determined. The strain magnitude is of the order of a few micrometers per metre expressed as microstrains. But it is difficult to measure such displacements directly, except in certain isolated cases, since the magnitude involved is very small. Therefore, a device or gauge, which can yield surface strains directly, is preferred. Such a device is popularly known as the "strain gauge". An accurate definition of strain on the surface requires the determination of the slopes of the displacement of the surfaces. The strains as such are likely to vary from point to point. Direct measurements of such small displacements over the entire surface of the body are also difficult. This is overcome by measuring one displacement component over a small portion of the body, along a short line segment. The strain measured in this manner may not be the true value, since the measurement is made over a finite length and not at a point. The error produced by this approach depends upon the strain gradient and the length of the line segment.

It is imperative to reduce the size of a strain gauge to improve the accuracy of measurements, but as the size is very much reduced, dimensional tolerances become very critical. If a strain value of the order of 1000 microstrains is to be measured to an accuracy of 5 microstrains over a gauge length of 5 mm, a displacement of 25 x *10-5* mm has to be accurately determined.

The basic characteristics of a strain gauge that one looks for are the gauge length, gauge width, gauge sensitivity, range of measurement, accuracy, frequency response and the ambient environmental conditions it can withstand. Since strain cannot be measured at a point, non-linear stress fields can give rise to errors, depending upon the gauge dimensions. Sensitivity, which is defined as the smallest value of strain that can be read, is of the order of microstrain and can be attained with the aid of modem instrumentation techniques. The maximum strain measurable and the accuracy achievable depends very much upon the type of gauges used and the method of gauging employed.

LESSON 32. Bonded strain gauge transducer-Tacho generator Strain Gages

TYPES OF STRAIN GAUGES

Strain gauges can be classified as mechanical, optical, or electrical depending upon the principle of operation and their constructional features. Of these, the electrical strain gauges and that too the electrical resistance type gauges, are the most popular because of the many advantages they offer in the process of measurement.

Mechanical Gauges

In mechanical gauges, the change in length L\ 1 is magnified mechanically using levl'.s or gears. Among them the Huggenburger type of extensometer is the most popular, wherein a lever system is employed to obtain the magnification of the movable knife-edge of the extensometer with respect to a fixed knife- edge. In a demountable type of strain gauge, the actual movement of the pivot is' transferred to the spindle of a dial gauge, where the movement is magnified by a rack-and-pinion arrangement. Mechanical strain gauges are comparatively larger in size, and as such are suitable only in cases where sufficient area is available on the test specimen for mounting the gauge. Further, they are useful in cases where the strain gradient is negligible and the additional mass of the mechanical gauge does not contribute to any error. These gauges are employed for static strain measurements only and also in cases where the point of measurement is accessible for visual observation.

Optical Gauges

Optical strain gauges are very similar to mechanical strain gauges except that the magnification is achieved with multiple reflectors using mirrors or prisms. As such the inertia of the system is very much reduced. In Martin's mirror-type extensometer, a plain mirror is rigidly attached to a movable knife- edge. When subjected to stress the mirror rotates through an angle, and the reflected light beam from the mirror subtends an angle twice that of the incident light. The measurement accuracy is high and independent of temperature variations.

Electrical Strain Gauges

The principle of an electrical strain gauge is based upon the measurement of the changes in resistance, capacitance, or inductance that are proportional to the strain transferred from the specimen to the basic gauge element. The most versatile device for experimental determination of strain for the purpose of stress analysis is the bonded resistance type of strain gauge. Capacitance and inductance type are only employed for special applications. Therefore, the rest of the treatment in this book is mainly concerned with resistance gauges only.

The basic concept of an electrical resistance strain gauge is attributed to Lord Kelvin who in 1856 expounded the theory that the resistance of a copper or iron wire changes when subjected to tension. The resistance of the wire changes as a function of strain. increasing with tension and reducing with compression. Sensitivity differs from material to material. Such a change in resistance can be measured accurately using a Wheatstone bridge. Developed on this principle, the electrical resistance strain gauge is basically a metal wire or foil subjected to the same strain as that of the specimen under test, achieved through suitable bonding of the gauge to the specimen.

Another class of strain gauge which is of a recent origin is the semiconductor type, piezoresistive strain gauge. This gauge has the advantages of high sensitivity, small size, and adaptability for both static and dynamic measurements.

THEORY OF OPERATION OF RESISTANCE STRAIN GAUGES

Assume a conductor of length L and cross-sectional area A. If this conductor is strained axially in tension. causing an increase in length, the lateral dimensions will reduce as a function of the Poisson's ratio of the wire material. Since the resistance of the wire is dependent on its length, area of cross- section, and also its specific resistivity, the resultant change in resistance due to strain can be interpreted as due to a dimensional change of the wire or due to a change in the specific resistivity.

The gauge factor indicates the strain sensitivity of the gauge in terms of the change in resistance per unit resistance per unit strain. It can be seen that the resistance change in a metal wire due to strain is produced by two factors, namely, the change in specific resistance 11 and the change in dimensions of the wire expressed by the factor (1 + 2v). In the elastic range, the Poisson's ratio v is nearly constant and is equal to 0.3 for most metals. The gauge factor G for various materials ranges from -12 for pure nickel to + 3.6 for isoelastic material, which indicates that the contribution due to the changes in the resistivity of the wire material can be considerable. The apparent reason for the change in resistivity with applied strain is due to changes in ~obility and the number of free electrons in the material. However, the gauge factor determined experimentally is reasonably constant for a given material. In the purely elastic region of deformation of any material, a change in volume is not possible as the wire cannot store energy in these conditions, which means that there cannot be any change in resistivity. Therefore, under constant volume conditions, the gauge factor takes a different value, which is nearly 2.0. The factor is constant over its elastic region, thus providing a wide linear stress-strain relationship.

The expression for gauge factor given by Eq. (5.19) is for a single uniform length of a conductor. Usually the strain gauge used for actual measurement is in the form of a grid. This causes certain sections of the strain gauge to be located in a direction transverse to the direction of the actual strain. The transverse strain also produces a change in resistance of the wire in addition to the axial strain. The calibrated gauge factor given by the manufacturer is normally valid only when the transverse strain is related to the axial strain as given by the equation J " "'

Such inaccuracies may be insignificant in many cases, since the Poisson's ratio is nearly the same for most metals.

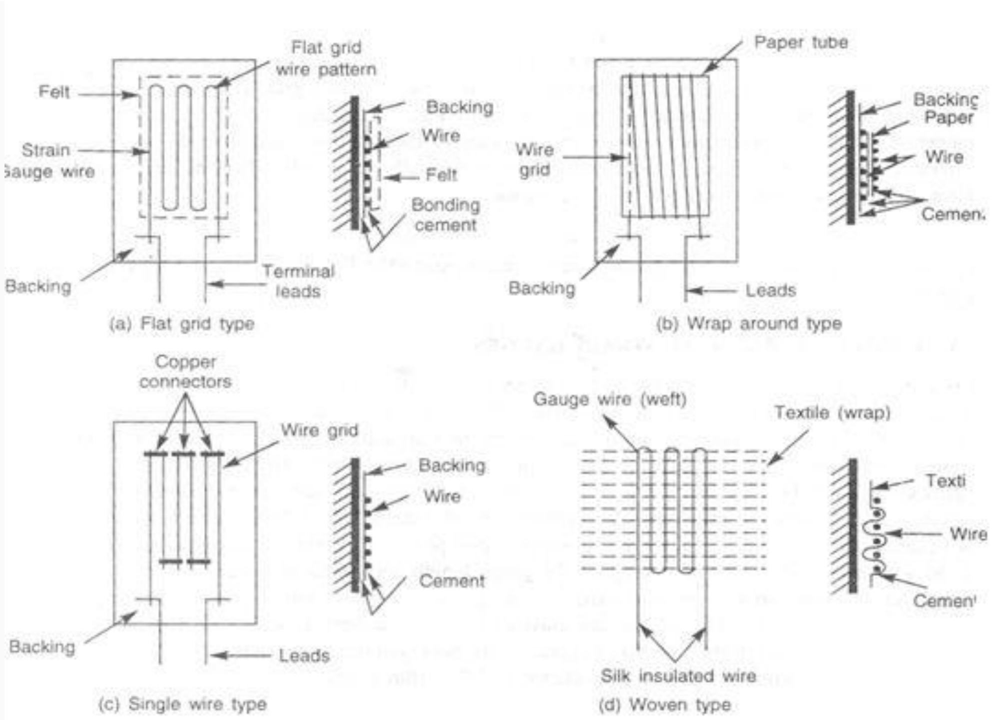
TYPES OF ELECTRICAL STRAIN GAUGES

It is apparent from the analysis presented in the previous section that a single length of wire can as well be used as the sensing element in a strain gauge. However, the circuits, which are used for measuring the resistance changes, impose certain restrictions on the minimum resistance that a strain gauge should possess. This value depends upon the gauge current and gauge length. Higher resistance gauges offer higher changes in the resistance for a given gauge factor, and at the same time draw lower current and have less heat dissipation problems. A resistance of the order of 60 to 1000 ohms is normally chosen for optimum performance. To achieve this value, a grid pattern is formed, thereby increasing the length of the wire and at the same time keeping the gauge length and width minimum. The sensing wires used in these electrical strain gauges are drawn out of special metal alloys which are discussed in Sec. 5.6 The gauges are classified into a number of categories depending upon the method of fabrication, but the two major types are the wire and the foil gauges. Two other types, which are of very recent origin are the semiconductor and thin, film gauges.

Wire Gauges

Wire strain gauges are normally of two types,. namely, bonded and unbonded gauges depending on the method of fabrication. **In** the first type, the strain gauge is bonded directly to the surface of the specimen being tested with a thin layer of adhesive cement \vhich serves to transmit the strain from the specimen to the gauge wires and at the same time serves as an electrical insulator. Keeping the surface area of varieties, viz. flat grid, wrap around, single wire, and woven.

Flat Grid Type In this type the wire is wound back and forth as a grid, as illustrated in Fig. 5.4(a). This grid structure is bonded to a backing material, such as paper or epoxy, with an adhesive that can hold the wire element to the base firmly, permitting a good transference of strain from the base to the wires. Since the ends of each section of the wire are looped around, transverse strains also cause changes in resistance in such sections of the wire. In order to reduce the cross sensitivity, such loop lengths should be minimized or joined through a different material having a lower sensitivity to strain than that of the actual material used for the strain gauge. In a standard gauge the cross sensitivity should not be greater than 2% of the sensitivity of the major axis. The wire grid plane should be as close to the specimen surface as possible to achieve maximum transfer of strain from the specimen and to keep the creep and hysteresis minimum.



Wrap-Around Type This type of gauge is wound on a flattened tube of paper, or alternately, on a thin strip or card as shown in Fig. 5.4(b). Gauge lengths smaller than that of the flat-grid type can be achieved for the same resistance value, but the gauge exhibits greater surface thickness since the grid wire is in two planes, introducing different transfer characteristics from that of the flat type and resulting in larger hysteresis and creep.

Single-Wire Gauges Single-wire types were developed to eliminate the cross-sensitivity factor. In this device single wires are stretched across and laid as shown in Fig. 5.4(c). Instead of loops formed by the same wires, thick copper wires are welded at the ends, reducing the cross sensitivity considerably. These gauges are not very popular and are intended for large gauge lengths only.

Woven Type This method of fabrication, as- shown in Fig. 5.4(d), is employed in gauges intended for the measurement of large strains. A silk-insulated Eureka wire is wound as the weft on a rayon wrap to form a woven-type gauge, which is useful for tests on fabrics and leather. High-temperature gauges of this type are developed with a glass fibre weave but they are not popular for common engineering applications.

For good sensitivity and faithful transmission of strain to the gauge, it is essential that the gauge wire should have high resistivity and large surface area. Such a high resistance can be achieved only with a thin wire of long length. The gauge is usually fabricated with a constantan wire of 20 microns diameter, wound in a grid format with as many loops as possible, laid side by side. In spite of its small diameter, the wire can withstand tension and compression easily, mainly because of the fact that the surface area is very large compared to its cross-section. This large bonded area controls the movements of the wire almost perfectly with no buckling. The sensitivity of the bonded wire gauge under compression is lower than that at tension by I to 2% only. A large length to width ratio in the grid structure is also desirable, to keep the transverse sensitivity minimum.

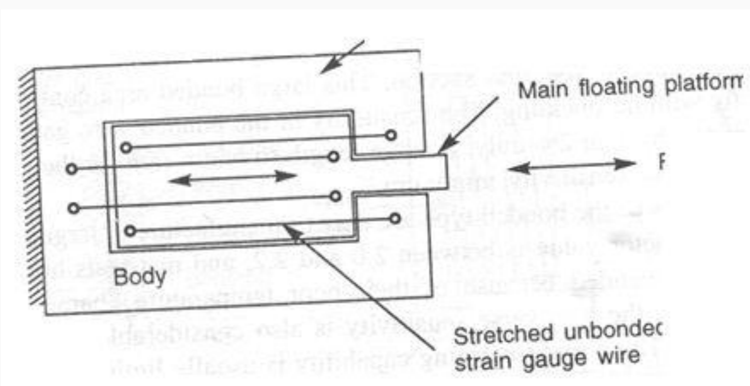
Wire strain gauges of the bonded type are easy to manufacture in large numbers at relatively low cost. The normal gauge factor; value is between 2.0 and 2.2, and materials having a higher gauge factor are not normally recommended because of their poor temperature characteristics. For materials having higher gauge factors, the transverse sensitivity is also considerable, unless they are specially reduced during manufacture. The current-carrying capability is usually limited because of the low area of cross- section of the wire, the normal value being 10 to 20 mA. It may be noted that stress concentrations can occur at the terminal and wire joints, causing fatigue failures.

Unbonded Strain Gauges

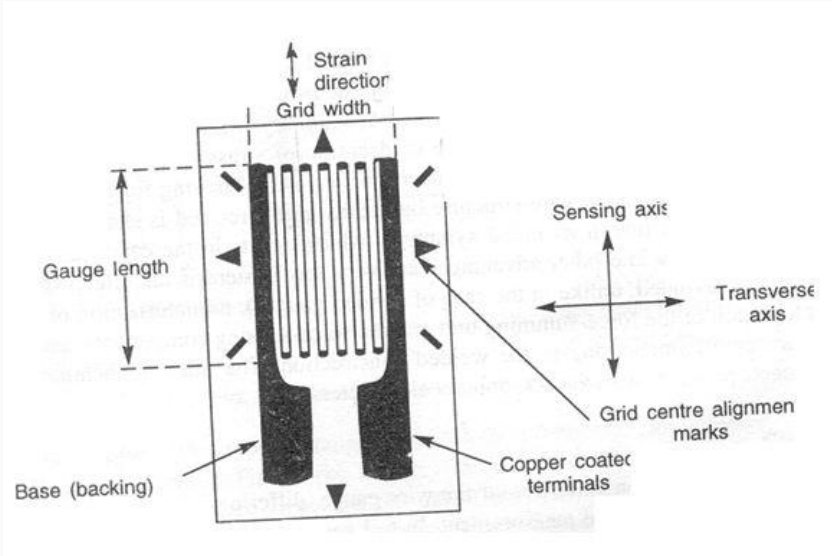
An unbonded strain gauge device is basically a free filament-sensing element where strain is transferred to the resistance wire directly without any backing. In a typical device, a few number of loops of high tensile strength resistance wire of 25 micron diameter of such materials as platinum tungsten alloy is wound bet\,,:een insulated pins, one of them attached to a stationary frame and the other to amovable frame, so that the winding experiences an increase or decrease of stress for a given force input. The schematic diagram of a typical displacement transducer wherein the measuring forces are transmitted to the platform containing the unbonded wire ~tructure by means of a force rod is shown in Fig. 5.5. The main advantage of this device lies in its radial symmetry which assists in the cancellation of spurious signals from transverse forces. The other advantages are very low hysteresis and creep (since the wire backing and bonding are avoided, unlike in the case of bonded gauges), miniaturization of the transduc- ing element by integration of the force-summing unit and the force-sensing components, and adaptability for high-temperature environments due to the welded construction. The main applications are in dis- placement transducers, pressure transducers, and accelerometers.

Foil Gauges

The foil strain gauge is basically an extension of the wire gauge, differing in its constructional features and having certain advantages in actual measurement. In foil gauges, the required grid pattern is formed



Unbonded strain gauge



hence better thermal stability. a larger ratio of the bonding area to the cross-sectional area is achieved compared to the wire gauges. This in turn enables a higher heat dissipation and better bonding properties. The strain repro- ducibility is also excellent. By making the perpendicular sections of the foil wide, the response of the gauge to the traverse strain can be considerably reduced. In the foil type of gauges, there is no stress concentration at the terminals due to the absence of joints, thereby extending the life of the gauge. By virtue of the versatile photo-chemical etching process, any type of complex pattern of any small size can be fabricated easily, such as circular gauges, diaphragm gauges, orthogonal arrays for shear measurement and transducer applications as well as many types of Rosettes for two-dimensional stress analysis.

Semiconductor Strain Gauges

Semiconductor strain gauges employ the piezoresistive property of doped silicon and germanium. In the metal alloy strain gauges described earlier, the strain sensitivity is mainly due to the dimensional change with a lesser contribution due to the resistivity changes. Semiconductor strain gauges, developed as an offshoot of the integrated circuit technology, are fabricated from single crystals of silicon and germa- nium, where the strain sensitivity is mainly due to resistivity changes in the semiconductor material itself. A unique feature of the device is that the change in resistance due to strain is 40 to 100 times more than that of the conventional metal alloy types. In addition to the high gauge factor, other advantages are che- mical inertness, freedom from hysteresis and creep effects, good fatigue life, and low cross sensitivity.

Since the electrical resistivity varies with the degree of doping, the type of semiconductor material is usually conductivity and conduction mechanism. used for general purpose gauges is the p (iii)-type silicon doped with 2 x 10-4 ohm-m, at room temperature. The material is first groove-cut into slices, This is followed by cutting the pieces to thin filaments of about 150 microns thickness. The electrodes are funned by vapour deposition and ohmic electrical contacts are made with gold wires, attached by means of a thermocompression method. The strain gauges are then brought to their nominal resistance by electrolytic etching. For use as a strain gauge element, they are embedded on a film backing of phenolic. bakelite or epoxy.

The main characteristics specified fur a semiconductor gauge are: (a) Filament material-p- or n-type silicon (b) Gauge factor-positive or negative (c) Gauge length (d) Gauge resistance (e) Temperature coefficient (f) Backing or encapsulation (g) Bonding (cementing or welding) (h) Lead geometry .

Special solder material (cadmium-tin) is used for soldering the leads to the circuit. The semiconductor strain gauges are available with both positive and negative gauge factors, for p- and n-type silicon respectively. This enables to form bridge circuits with two active arms at one location itself, even when both the gauges are subjected to the same strain value. In the application of these gauges in transducers, the strain-sensitive diaphragm and other elements are fabricated as bonded gauges, or by diffusion techniques, as p-n junctions on silicon substrates. Some of the properties of these gauges are discussed below.

(a) Gauge Sensitivity The variation of gauge factors of p- and n-type silicon as a function of resistivity and crystal orientation, is illustrated in Fig. 5.8. The sensitivity of semiconductor gauges at low strain limits at room temperature is computed from the equation, Since the product 1t' is well in excess of 100, the contribution towards the sensitivity from the gauge geometry is insignificant when compared with the piezoresistive effect. Unlike wire gauges, a small The bonding techniques employed are similar to that for wire or foil gauges, except that greater care must be taken in

Bonded Strain Gage Transducers

A typical construction for a strain-gage load cell for measuring compressive forces is shown in Fig. (Cells to measure both tension and compression require merely the addition of suitable mechanical fittings at the ends). The load-sensing member is short enough to prevent column buckling under the rated load and is proportioned to develop about 1,500 µÎ at full-scale load (typical design value for al forms of foil gage transducers). Materials used include SAE 4340 steel, 17-4 pH stainless steel, and 2002-T4 aluminium alloy, with the last being quite popular for 'homemade' transducers. Foil-type metal gages are bonded on all four sides; gages 1 and 3 sense the direct stress due to F_i and gages 2 and 4 the transverse stress due to Poisson's ratio µ. This arrangement gives a sensitivity 2(1+µ) times that achieved with a single active gage in the bridge. It also provides primary temperature compensation since all four gages (at least for steady temperatures) at the same temperature.

Strain Measurements

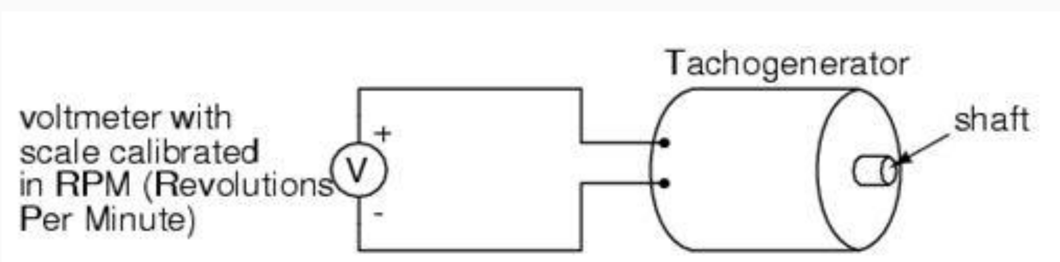
Let us first consider some basic definitions. Any strain measurement must be made over a finite length of the work piece. The smaller this length, the more nearly the measurement will approximate the unit strain at a point. The length over which the average strain measurement is taken is called the base length. The deformation sensitivity is defined as the minimum deformation that can be indicated by the appropriate gage. Strain sensitivity is the minimum deformation that can be indicated by the gage per unit base length.

A simple method of strain measurement is to place some type of grid marking on the surface of the work piece under zero-load conditions and then measure the deformation of this grid when the specimen is subjected to a load. The grid may be scribed on the surface, drawn with a fine ink pen, or photoetched. Rubber threads have also been used to mark the grid. The sensitivity of the grid method depends on the accuracy with which the displacement of the grid lines may be measured. A micrometer microscope is frequently employed for such measurements. An alternative method is to photograph the grid before and after the deformation and make the measurements on the developed photograph. Photographic paper can have appreciable shrinkage, so that glass photographic plates are preferred for such measurements. The grid may also be drawn on a rubber model of the specimen and the local strain for the model related to that which would be present in the actual work piece. Grid methods are usually applicable to materials and processes having appreciable deformation under load. These methods might be applicable to a study of the strains encountered in sheet-metal forming processes. The grid could be installed on a flat sheet of metal before it is formed. The deformation of the grid after forming gives the designer an indication of the local stresses induced in the material during the forming process.

Brittle coatings offer a convenient means for measuring the local stress in a material. The specimen or work piece is coated with a special substance having very brittle properties. When the specimen is subjected to a load, small cracks appear in the coating. These cracks appear when the state of tensile stress in the work piece reaches a certain value and thus may be taken as a direct indication of this local stress. The brittle coatings are valuable for obtaining an overall picture of the stress distribution over the surface of the specimen. They are particularly useful for determination of stresses at stress concentration points that are too small or inconveniently located for installation of electrical resistance or other types of strain gages. In some instances, stress data obtained from brittle – coating tests may be used to plan more precise strain measurements with resistance strain gages.

Tachogenerator

An electromechanical generator is a device capable of producing electrical power from mechanical energy, usually the turning of a shaft. When not connected to a load resistance, generators will generate voltage roughly proportional to shaft speed. With precise construction and design, generators can be built to produce very precise voltages for certain ranges of shaft speeds, thus making them well-suited as measurement devices for shaft speed in mechanical equipment. A generator specially designed and constructed for this use is called a tachometer or tachogenerator. Often, the word "tach" (pronounced "tack") is used rather than the whole word.



By measuring the voltage produced by a tachogenerator, you can easily determine the rotational speed of whatever its mechanically attached to. One of the more common voltage signal ranges used with tachogenerators is 0 to 10 volts. Obviously, since a tachogenerator cannot produce voltage when its not turning, the zero cannot be "live" in this signal standard. Tachogenerators can be purchased with different "full-scale" (10 volt) speeds for different applications. Although a voltage divider could theoretically be used with a tachogenerator to extend the measurable speed range in the 0-10 volt scale, it is not advisable to significantly overspeed a precision instrument like this, or its life will be shortened.

Tachogenerators can also indicate the direction of rotation by the polarity of the output voltage. When a permanent-magnet style DC generator's rotational direction is reversed, the polarity of its output voltage will switch. In measurement and control systems where directional indication is needed, tachogenerators provide an easy way to determine that.

Tachogenerators are frequently used to measure the speeds of electric motors, engines, and the equipment they power: conveyor belts, machine tools, mixers, fans, etc.