Uttar Pradesh Rajarshi Tandon Open University

Bachelor of Science

DCEPHS-109

Solid State Physics and Advanced Electronics

Uttar Pradesh Rajarshi Tandon Open University

Bachelor of Science

DCEPHS-109

Solid State Physics and Advanced Electronics

Block

1 Basic Concepts of Solids

UNIT - 1	**Crystal and its Structure**
UNIT - 2	**Band theory of Solids**
UNIT - 3	**Lattice Vibration**
UNIT - 4	**Magnetism and Superconductivity**

Course Design Committee

Prof. Ashutosh Gupta **Chairman**
Director, School of Science, UPRTOU, Prayagraj
Prof. A. K. Rai **Member**
Professor, Dept. of Physics, University of Allahabad, Prayagraj
Gorakhpur
Prof. Ram Kripal **Member**
Professor, Dept. of Physics, University of Allahabad, Prayagraj
Dr. Anjani Kumar Singh **Member**
Associate Professor (Retd.), E.C.C University of Allahabad, Prayagraj

Course Preparation Committee

Mr. Ashish Sankhwar **Author**
Asst Prof. Dept. Of Electrical Engg. **Unit- 01-10**
FET,MJPRU,Bareilly

Prof. Nitya Kaushal Kishore **Editor**
Ex. HOD, Dept. of Physics, **Block -I,II & III**
S.K.M.U. University,
Dumka, Jharkhand.

Dr. Arvind Kumar Mishra **Coordinator**
Assistant Professor , Physics,
School of Science, UPRTOU, Prayagraj

DCEPHS – 109 : Plant Pathology and Microbiology
ISBN-

Unit 1- Crystal and its structure

Structure:

1.1 Introduction:

Almost all solids fall in the category of crystalline solids including metallic elements (iron, silver, and copper) and non-metallic elements (Phosphorus, Sulphur, and iodine). Also several compounds like sodium chloride, zinc sulphide and naphthalene build crystalline solids. We can define amorphous solids as materials which don't have certain organized arrangement of atoms and molecules. Most solids are amorphous in nature and are utilized in many sectors as well. One of the most common examples of amorphous solids is glass, which is used widely in the manufacturing sector.

Liquid crystal materials are unique in their properties and uses. As research into this field continues and as new applications are developed, liquid crystals will play an important role in modern technology. In a simple cubic structure, the spheres are not packed as closely as they could be, and they only "fill" about 52% of the volume of the container. This is a relatively inefficient arrangement, and only one metal (polonium, Po) crystallizes in a simple cubic structure. Many other metals, such as aluminum, copper, and lead, crystallize in an arrangement that has a cubic unit cell with atoms at all of the corners and at the centers of each face. This arrangement is called a face-centered cubic (FCC) solid. Some metals crystallize in an arrangement that has a cubic unit cell with atoms at all of the corners and an atom in the center. This is called a body-centered cubic (BCC) solid.

Atoms in the corners of a BCC unit cell do not contact each other but contact the atom in the center. A BCC unit cell contains two atoms: one-eighth of an atom at each of the eight corners (8×18=18×18=1 atom from the corners) plus one atom from the center. The smallest possible portion or part of the crystal lattice which repeats itself in different directions of the lattice is called the unit cell. Many unit cells combine to geometrically form the crystal lattice. Thus, a Bravais lattice can refer to one of the 14 different types of unit cells that a crystal structure can be made up of. These lattices are named after the French physicist Auguste Bravais. The direct lattice represents the triple periodicity of the ideal infinite perfect periodic structure that can be associated to the structure of a finite real crystal.

Thus, the Miller indices are 3,6,2. If a plane is parallel to an axis, its intercept is at infinity and its Miller index is zero.

Bragg's Law was introduced by Sir W.H. Bragg and his son Sir W.L. Bragg. The law states that when the x-ray is incident onto a crystal surface, its angle of incidence, θ, will reflect back with a same angle of scattering, θ. And, when the path difference, d is equal to a whole number, n, of wavelength, a constructive interference will occur.

The generalized Hooke's Law also reveals that strain can exist without stress. For example, if the member is experiencing a load in the y-direction (which in turn causes a stress in the y-direction), the Hooke's Law shows that strain in the x-direction does not equal to zero. In particular, the speed of body waves passing through a material is entirely dependent upon the ratio of the elastic modulus of that material to its density. Whenever any external force is applied to a system, there is a resultant strain; similarly, whenever a system is strained in some way, there is then some stress upon the system. For example, squashing a jelly will deform it; deforming a jelly will result in a restoring force or stress eager to return the jelly to its original shape.

1.2 Objectives:

After studying this unit you should be able to

- Explain and identify Crystalline and amorphous state of solids, liquid crystal and its characteristics.
- Study and identify Simple crystal structure (SC, FCC, BCC).
- Explain Unit cell and Bravais lattice.
- Study and identify Classification of lattices and types of crystals on the basis of Bravais lattice.
- Explain and identify Direct and reciprocal lattice, Miller indices and planes.
- Study and identify X-ray diffraction, Bragg's law.
- Explain and identify Generalized Hooke's law for Anisotropic body, elastic constants of cubic crystals.

1.3 Crystalline and amorphous state of solids:

Types of Solids:

On the basis of the arrangement of constituent particles, the solids are classified into two categories, namely:

- Crystalline Solids
- Amorphous Solids

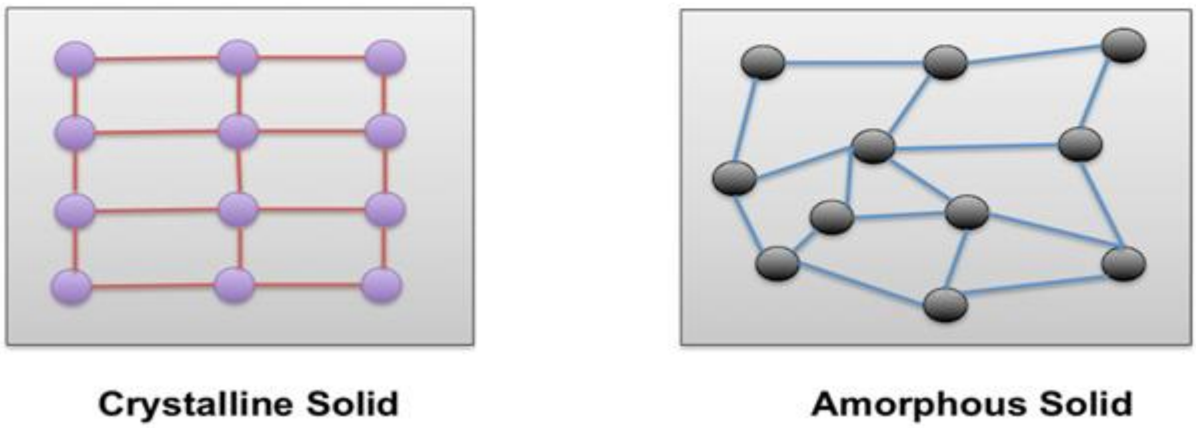


Fig.1.1 Arrangement of Atoms

Crystalline Solids:

The solids in which the constituent particles of matter are arranged and organized in a specific manner are called Crystalline Solids. These solids contain crystals in their structure and each crystal has definite geometry. Adding further, as crystalline solids have low potential energy, they are the most stable form of solids. Almost all solids fall in the category of crystalline solids including metallic elements (iron, silver, and copper) and non-metallic elements (Phosphorus, Sulphur, and iodine). Also several compounds like sodium chloride, zinc sulphide and naphthalene build crystalline solids.

Fig.1.2 Example of Crystalline Solids

Characteristics of Crystalline Solids:

The main characteristics of crystalline solids are mentioned as below:

- Crystalline solids show regular structure and have definite geometrical shape.
- The sharp freezing point is found in crystalline solids. This is because the distance between same atoms/molecules or ions is same and remains constant, unlikely from amorphous solids.
- The heat of fusion is definite and fixed as the regularity in crystal lattice remains same and is ideal.
- Crystalline Solids are also known as True Solids as they don't tend to flow like pseudo solids.
- When we cut a crystal solids with a knife, we obtain a flat and smooth surface.

- The nature of crystalline solid is anisotropic; that is, the properties turn out to be different in different direction.
- Crystalline solids depict both long range and short range order.
- Examples: Quartz, Calcite, Sugar, Mica, Diamonds etc.

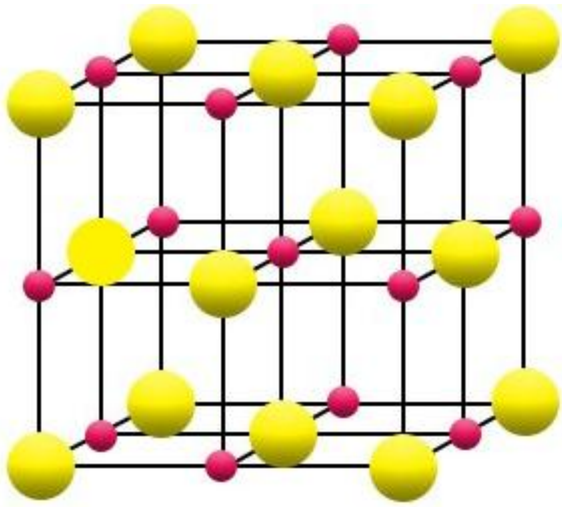


Fig.1.3 Lattice Structure of Crystalline Solids

Uses of Crystalline Solids:

There are many applications of crystalline solids, some are:

- Diamond is the most decent example of crystalline solids and is widely used in making beautiful jewelry items.
- Quartz is extensively used in manufacturing of watches and clocks.
- Many crystalline solids are used as a raw material in many industries

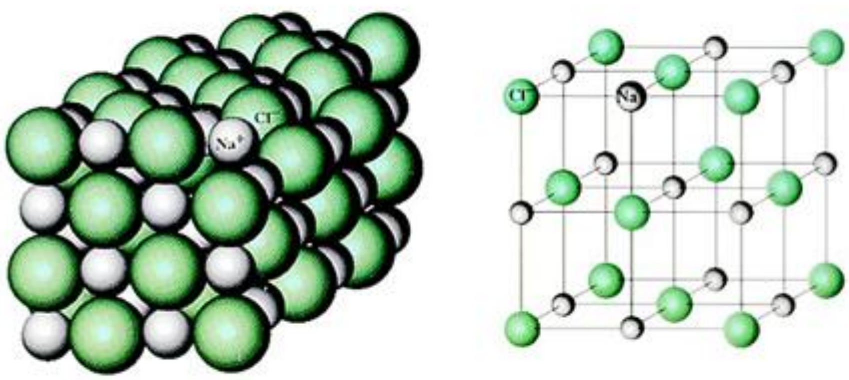


Fig.1.4 Structure of NaCl

Crystalline Solids are further classified into four categories on the basis of intermolecular interactions between molecules, they are:

- Molecular Solids
- Covalent or Network Solids
- Ionic Solids
- Metallic Solids

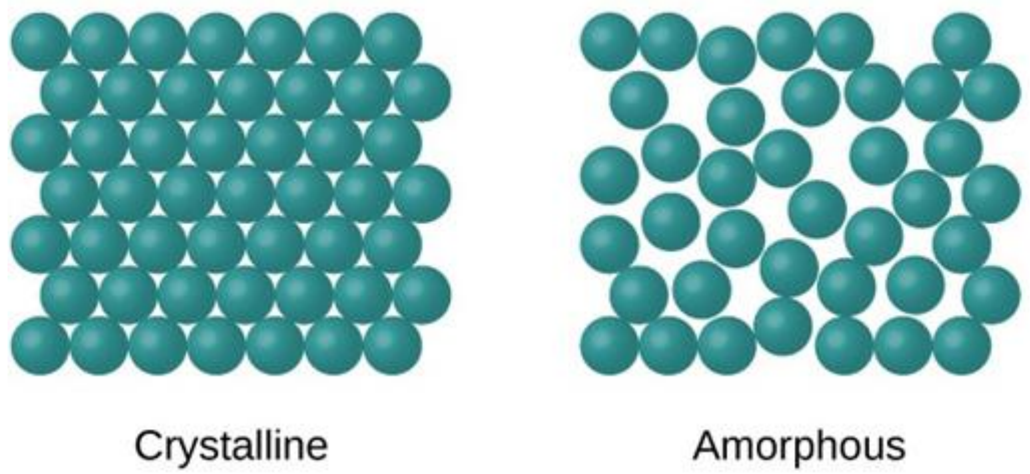


Fig.1.5 Shape of Crystalline and Amorphous Solids

Molecular Solids: In molecular solids, the constituent particles are molecules. Molecular solids are generally insulators and are soft in nature. The density of molecular solids is quite low. Based on nature of molecules molecular solids are further classified into three forms:

- Non-polar Molecular Solids
- Polar Molecular Solids
- Hydrogen-bonded Molecular Solids

Covalent Solids:

In covalent solids, the constituent atoms of molecules held together by strong covalent bonds. They form giant structures and are generally hard in nature. They are also poor conductors of electricity except graphite, which is a good conductor of electricity.

Ionic Solids: The constituent particles in ionic solids are cations and anions. There is strong electrostatic force of attraction between particles. Ionic solids are hard and brittle. They have very high melting point. Ionic solids are poor conductor of electricity in solid state, but in dissolved state, they act as good conductor of electricity. For Example: NaCl is a good conductor of electricity in dissolved state.

Metallic Solids: The constituent particles in metallic solids are metal atoms and valence electrons. Metallic Solids have high melting and boiling point. They are also good conductor of electricity due to presence of valence electrons. For Example: Copper, Gold etc.

Amorphous Solids:

The solids in which the constituent particles of matter are arranged in a random manner are called amorphous solids. It is a non-crystalline solid with no proper arrangement of atoms in the solid lattice. In other words, we can define amorphous solids as materials which don't have certain organized arrangement of atoms and molecules. Most solids are amorphous in nature and are utilized in many sectors as well. One of the most common examples of amorphous solids is glass, which is used widely in the manufacturing sector.

Characteristics of Amorphous Solids:

An Amorphous Solid depicts following properties, which are as follows:

- The constituent particles of matter inside solid are arranged in a random manner, that is, the position of atoms and molecules is not fixed and varies from one solid to another.
- Amorphous Solids don't have definite shape or geometry due to random arrangement of atoms and molecules inside the solid lattice.
- Short-range order is found in amorphous solids.
- Amorphous Solids are also called Pseudo-solids or Super-cooled Liquids because they don't form crystalline structure and has the ability to flow.
- The nature of amorphous solids is isotropic in nature that is, the properties measured in all directions come out to be same, and example refractive index of amorphous solids is same.
- Amorphous solids don't show sharp melting point, this is because of irregular packing of amorphous solids.
- When we cut an amorphous solid, we find the broken constituent particles to be irregular in shape and geometry.

- Amorphous solids are unsymmetrical in nature, due to irregular packing of atoms and molecules inside the solid lattice.
- Amorphous solids don't have fixed heat of fusion because of absence of sharp melting point.
- Examples: Plastics, Glass, Rubber, Metallic Glass, Polymers, Gel etc.

Uses of Amorphous Solids:

There are many applications of amorphous solids, some of them are:

- The glass is widely used in packaging (food jars, cosmetics box, and soft-drink bottles), making tableware (utensils), in the construction of buildings (windows, lighting, and shelves) etc.
- Rubber is mainly used in manufacturing of tires, footwear, ropes, camp cloth and as a raw material for several industries.
- Use of polymer can be seen in manufacturing of pipes, medicines and as a raw ingredient for many factories.
- Amorphous silicon is considered as the best photovoltaic material to convert sunlight into electricity.

Difference between Crystalline and Amorphous Solids:

The difference between crystalline and amorphous solids can be laid out in the table below:

Characteristic	**Crystalline Solids**	**Amorphous Solids**
Melting Point	Melt at fixed temperature	Melts steadily over range of temperatures
Arrangement of Constituent Particles	Regular	Irregular
Shape	Regular and Definite Shape	Irregular Shape in Nature
Cleavage	When cut, two smooth and	When cut, two surfaces of

	plain pieces are obtained	irregular shape is obtained
Heat of Fusion	Definite	Indefinite
Anisotropy	Anisotropic	Isotropic
Nature	True Solids	Pseudo Solids

1.4 Liquid Crystals:

Introduction to Liquid Crystals: The study of liquid crystals began in 1888 when an Austrian botanist named Friedrich Reinitzer observed that a material known as cholesteryl benzoate had two distinct melting points. In his experiments, Reinitzer increased the temperature of a solid sample and watched the crystal change into a hazy liquid. As he increased the temperature further, the material changed again into a clear, transparent liquid. Because of this early work, Reinitzer is often credited with discovering a new phase of matter - the liquid crystal phase.

Liquid crystal materials are unique in their properties and uses. As research into this field continues and as new applications are developed, liquid crystals will play an important role in modern technology. This tutorial provides an introduction to the science and applications of these materials.

What are Liquid Crystals?

Liquid crystal materials generally have several common characteristics. Among these is a rod-like molecular structure, rigidness of the long axis, and strong dipoles and/or easily polarizable substituent.

The distinguishing characteristic of the liquid crystalline state is the tendency of the molecules (mesogens) to point along a common axis, called the director. This is in contrast to molecules in the liquid phase, which have no intrinsic order. In the solid state, molecules are highly ordered and have little translational freedom. The characteristic orientational order of the liquid crystal state is between the traditional solid and liquid phases and this is the origin of the term mesogenic state, used synonymously with liquid crystal state. Note the average alignment of the molecules for each phase in the following diagram.

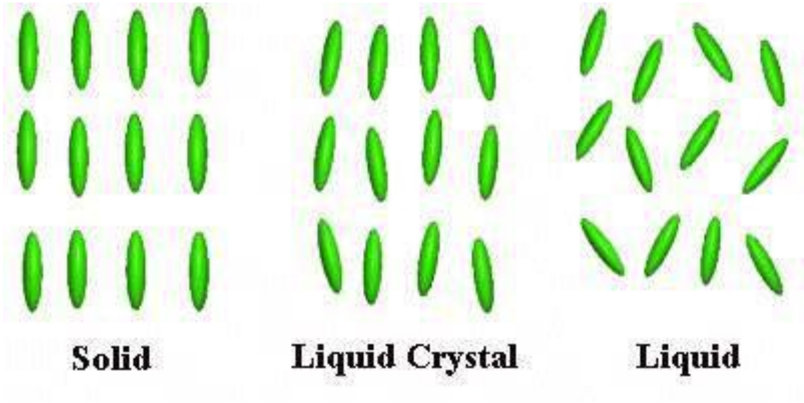


Fig.1.6 Molecular structure of Liquid Crystals

It is sometimes difficult to determine whether a material is in a crystal or liquid crystal state. Crystalline materials demonstrate long range periodic order in three dimensions. By definition, an isotropic liquid has no orientation order. Substances that aren't as ordered as a solid, yet have some degree of alignment are properly called liquid crystals.

To quantify just how much order is present in a material, an order parameter (S) is defined. Traditionally, the order parameter is given as follows:

$$S = (1/2) < 3\cos^2 q - 1>$$

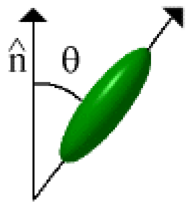


Fig.1.7 Angle between the director and the long axis of each molecule

where theta is the angle between the director and the long axis of each molecule. The brackets denote an average over all of the molecules in the sample. In an isotropic liquid, the average of the cosine terms is zero, and therefore the order parameter is equal to zero. For a perfect crystal, the order parameter evaluates to one. Typical values for the order parameter of a liquid crystal range between 0.3 and 0.9, with the exact value a function of temperature, as a result of kinetic molecular motion. This is illustrated below for a nematic liquid crystal material (to be discussed in the next section).

The tendency of the liquid crystal molecules to point along the director leads to a condition known as anisotropy. This term means that the properties of a material depend on the direction in which they are measured. For example, it is easier to cut a piece of wood along the grain than against it. The anisotropic nature of liquid crystals is responsible for the unique optical properties exploited by scientists and engineers in a variety of applications.

Characterizing Liquid Crystals:

The following parameters describe the liquid crystalline structure:

- Positional Order
- Orientational Order
- Bond Orientational Order

Each of these parameters describes the extent to which the liquid crystal sample is ordered. Positional order refers to the extent to which an average molecule or group of molecules shows translational symmetry (as crystalline material shows). Orientational order, as discussed above, represents a measure of the tendency of the molecules to align along the director on a long-range basis. Bond Orientational Order describes a line joining the centers of nearest-neighbor molecules without requiring a regular spacing along that line. Thus, a relatively long-range order with respect to the line of centers but only short range positional order along that line. (See discussion of hexatic phases in a text such as Chandrasekhar, Liquid Crystals)

Most liquid crystal compounds exhibit polymorphism, or a condition where more than one phase is observed in the liquid crystalline state. The term mesophase is used to describe the "subphases" of liquid crystal materials. Mesophases are formed by changing the amount of order in the sample, either by imposing order in only one or two dimensions, or by allowing the molecules to have a degree of translational motion. The following section describes the mesophases of liquid crystals in greater detail.

Liquid Crystal Phases: The liquid crystal state is a distinct phase of matter observed between the crystalline (solid) and isotropic (liquid) states. There are many types of liquid crystal states, depending upon the amount of order in the material. This section will explain the phase behavior of liquid crystal materials.

Nematic Phases: The nematic liquid crystal phase is characterized by molecules that have no positional order but tend to point in the same direction (along the director). In the following diagram, notice that the molecules point vertically but are arranged with no particular order.

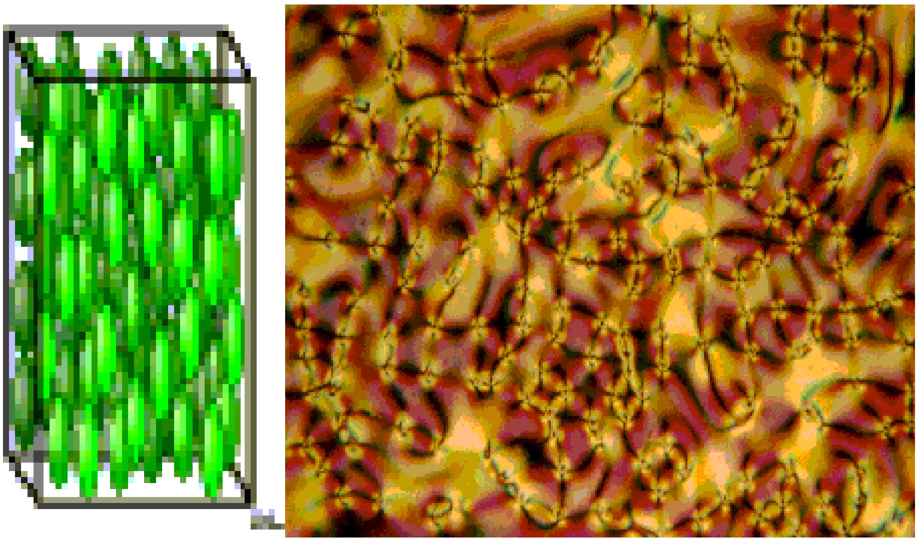


Fig.1.8 Nematic liquid crystal phase

Liquid crystals are anisotropic materials, and the physical properties of the system vary with the average alignment with the director. If the alignment is large, the material is very anisotropic. Similarly, if the alignment is small, the material is almost isotropic. The phase transition of a nematic liquid crystal is demonstrated in the following movie provided by Dr. Mary Neubert, LCI-KSU. The nematic phase is seen as the marbled texture. Watch as the temperature of the material is raised, causing a transition to the black, isotropic liquid.

A special class of nematic liquid crystals is called chiral nematic. Chiral refers to the unique ability to selectively reflect one component of circularly polarized light. The term chiral nematic is used interchangeably with cholesteric. Refer to the section on cholesteric liquid crystals for more information about this mesophase.

Smectic Phases: The word "smectic" is derived from the Greek word for soap. This seemingly ambiguous origin is explained by the fact that the thick, slippery substance often found at the bottom of a soap dish is actually a type of smectic liquid crystal. The smectic state is another distinct mesophase of liquid crystal substances. Molecules in this phase show a degree of translational order not present in the nematic. In the smectic state, the molecules maintain the general orientational order of nematics, but also tend to align themselves in layers or planes. Motion is restricted to within these planes, and separate planes are observed to flow past each other. The increased order means that the smectic state is more "solid-like" than the nematic.

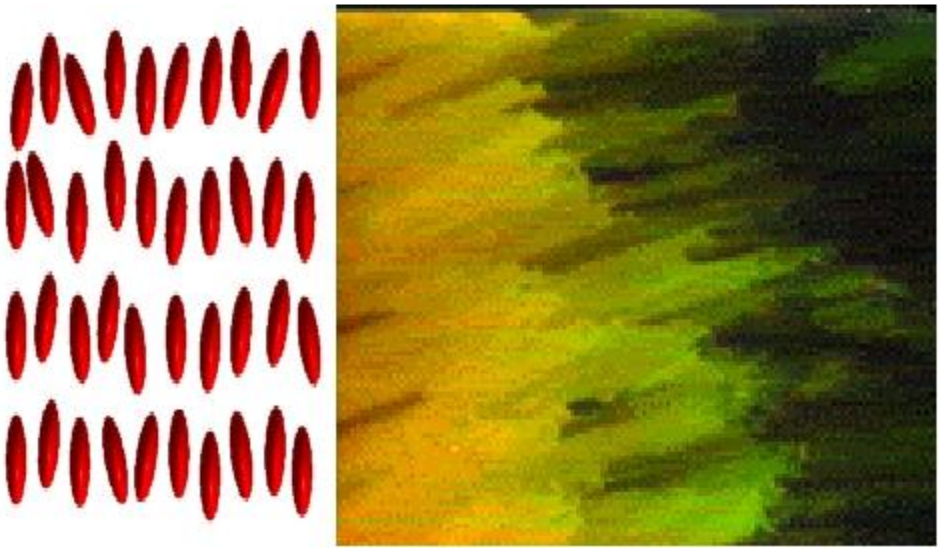


Fig.1.9 Picture of the smectic phase   Photo of the smectic phase (using polarizing microscope)

Many compounds are observed to form more than one type of smectic phase. As many as 12 of these variations have been identified, however only the most distinct phases are discussed here.

In the smectic-A mesophase, the director is perpendicular to the smectic plane, and there is no particular positional order in the layer. Similarly, the smectic-B mesophase orients with the director perpendicular to the smectic plane, but the molecules are arranged into a network of hexagons within the layer. In the smectic-C mesophase, molecules are arranged as in the smectic-A mesophase, but the director is at a constant tilt angle measured normally to the smectic plane.

As in the nematic, the smectic-C mesophase has a chiral state designated C*. Consistent with the smectic-C, the director makes a tilt angle with respect to the smectic layer. The difference is that this angle rotates from layer to layer forming a helix. In other words, the director of the smectic-C* mesophase is not parallel or perpendicular to the layers, and it rotates from one layer to the next. Notice the twist of the director, represented by the green arrows, in each layer in the following diagram.

In some smectic mesophases, the molecules are affected by the various layers above and below them. Therefore, a small amount of three dimensional order is observed. Smectic-G is an example demonstrating this type of arrangement.

Cholesteric Phases: The cholesteric (or chiral nematic) liquid crystal phase is typically composed of nematic mesogenic molecules containing a chiral center which produces intermolecular forces that favor alignment between molecules at a slight angle to one another. This leads to the formation of a structure which can be visualized as a stack of very thin 2-D nematic-like layers with the director in each layer twisted with respect to those above and below. In this structure, the directors actually form in a continuous helical pattern about the layer normal as illustrated by the black arrow in the following figure and animation. The black arrow in the animation represents director orientation in the succession of layers along the stack.

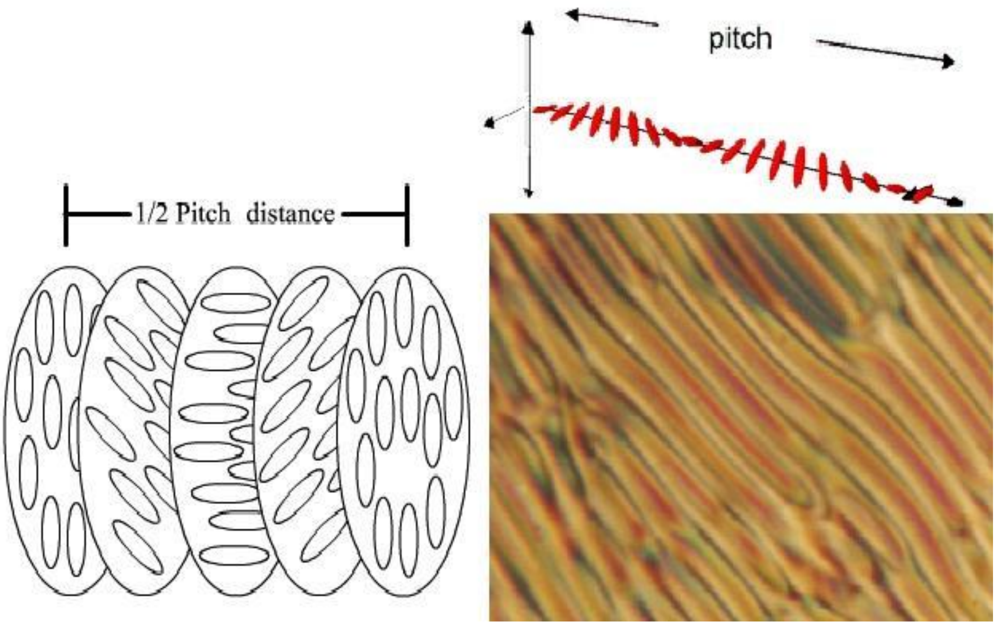


Fig.1.10 Cholesteric (or chiral nematic) liquid crystal phase

The molecules shown are merely representations of the many chiral nematic mesogens lying in the slabs of infinitesimal thickness with a distribution of orientation around the director. This is not to be confused with the planar arrangement found in smectic mesophases. An important characteristic of the cholesteric mesophase is the pitch. The pitch, p, is defined as the distance it takes for the director to rotate one full turn in the helix as illustrated in the above animation. A byproduct of the helical structure of the chiral nematic phase, is its ability to selectively reflect light of wavelengths equal to the pitch length, so that a color will be reflected when the pitch is equal to the corresponding wavelength of light in the visible spectrum. The effect is based on the temperature dependence of the gradual change in director orientation between successive layers (illustrated above), which modifies the pitch length resulting in an alteration of the wavelength of reflected light according to the temperature. The angle at which the director changes can be made larger, and thus tighten the pitch, by increasing the temperature of the molecules, hence giving them more thermal energy. Similarly, decreasing the temperature of the molecules increases the pitch length of the chiral nematic liquid crystal. This makes it possible to build a liquid crystal thermometer that displays the temperature of its environment by the reflected color. Mixtures of various types of these liquid crystals are often used to create sensors with a wide variety of responses to temperature change. Such sensors are used for thermometers often in the form of heat sensitive films to detect flaws in circuit board connections, fluid flow patterns, condition of batteries, the presence of radiation, or in novelties such as "mood" rings.

In the fabrication of films, since putting chiral nematic liquid crystals directly on a black background would lead to degradation and perhaps contamination, the crystals are micro-encapsulated into particles of very small dimensions. The particles are then treated with a binding material that will contract upon curing so as to flatten the microcapsules and produce the best alignment for brighter colors. An application of a class of chiral nematic liquid crystals which are less temperature sensitive is to create materials such as clothing, dolls, inks and paints.

The wavelength of the reflected light can also be controlled by adjusting the chemical composition, since cholesterics can either consist of exclusively chiral molecules or of nematic molecules with a chiral dopant dispersed throughout. In this case, the dopant concentration is used to adjust the chirality and thus the pitch.

Columnar Phases:

Columnar liquid crystals are different from the previous types because they are shaped like disks instead of long rods. This mesophase is characterized by stacked columns of molecules. The columns are packed together to form a two-dimensional crystalline array. The arrangement of the molecules within the columns and the arrangement of the columns themselves leads to new mesophases.

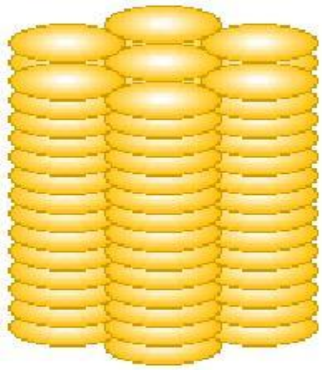


Fig.1.11 Columnar liquid crystals phases

1.4 Simple Crystal Structure (SC, FCC, BCC):

Simple Cubic structure (SC): In a simple cubic structure, the spheres are not packed as closely as they could be, and they only "fill" about 52% of the volume of the container. This is a relatively inefficient arrangement, and only one metal (polonium, Po) crystallizes in a simple cubic structure. As shown in Figure, a solid with this type of arrangement consists of planes (or layers) in which each atom contacts only the four nearest neighbors in its layer; one atom directly above it in the layer above; and one atom directly below it in the layer below. The number of other particles that each particle in a crystalline solid contacts is known as its coordination number. For a polonium atom in a simple cubic array, the coordination number is, therefore, six.

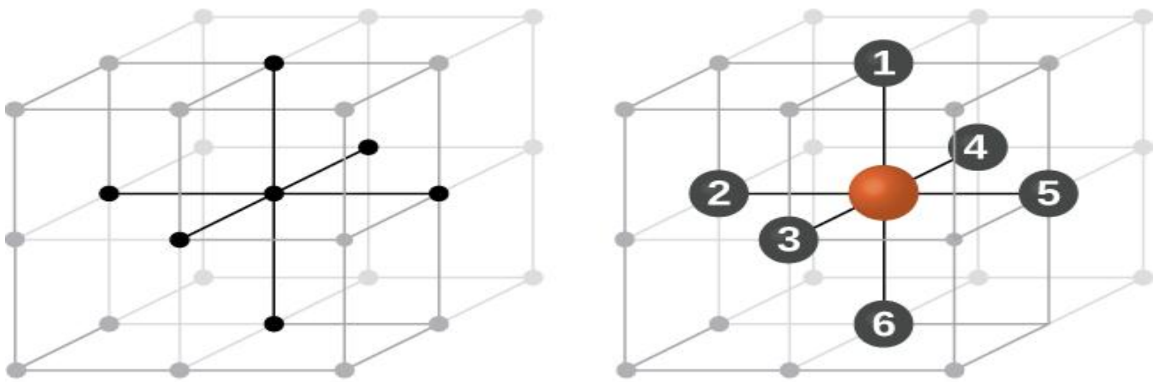


Fig.1.12 An atom in a simple cubic lattice structure contacts six other atoms, so it has a coordination number of six.

In a simple cubic lattice, the unit cell that repeats in all directions is a cube defined by the centers of eight atoms, as shown in Figure. Atoms at adjacent corners of this unit cell contact each other, so the edge length of this cell is equal to two atomic radii, or one atomic diameter. A cubic unit cell contains only the parts of these atoms that are within it. Since an atom at a corner of a simple cubic unit cell is contained by a total of eight unit cells, only one-eighth of that atom is within a specific unit cell. And since each simple cubic unit cell has one atom at each of its eight "corners," there is 8×18=18×18=1 atom within one simple cubic unit cell.

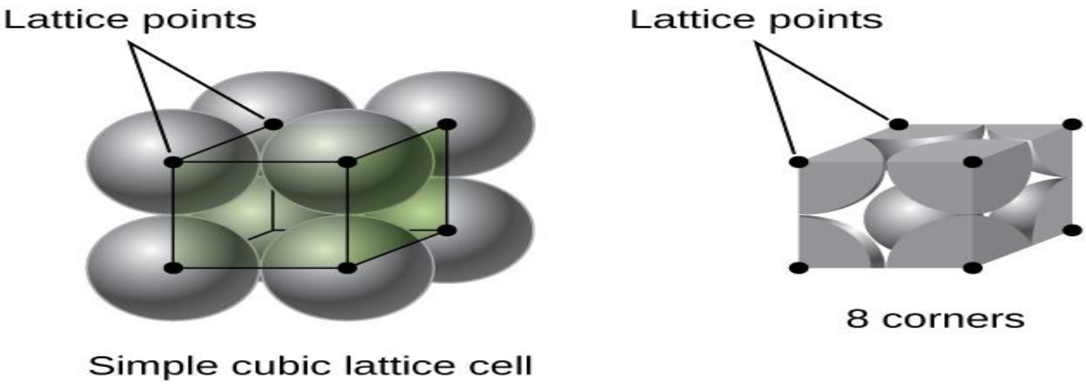


Fig.1.13 A simple cubic lattice unit cell contains one-eighth of an atom at each of its eight corners, so it contains one atom total.

Most metal crystals are one of the four major types of unit cells. For now, we will focus on the three cubic unit cells: simple cubic (which we have already seen), body-centered cubic unit cell, and face-centered cubic unit cell—all of which are illustrated in Figure. (Note that there are actually seven different lattice systems, some of which have more than one type of lattice, for a total of 14 different types of unit cells. We leave the more complicated geometries for later in this module.)

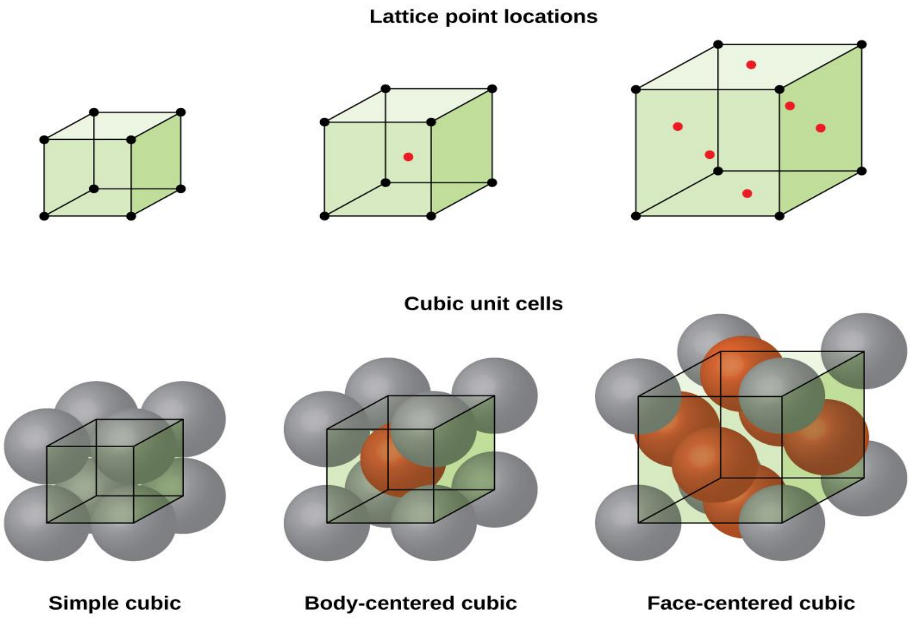
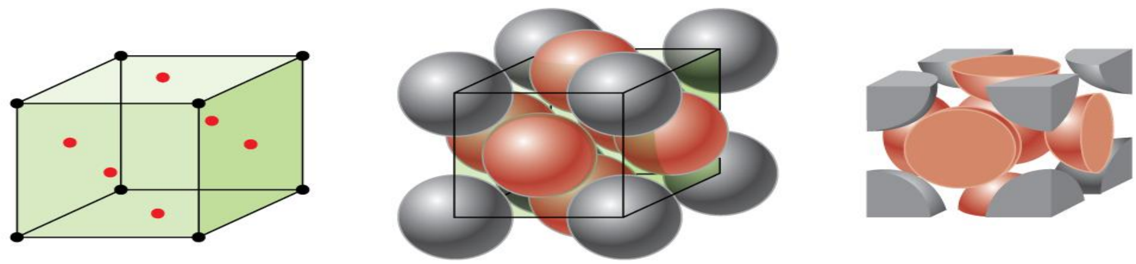


Fig.1.14. Cubic unit cells of metals show (in the upper figures) the locations of lattice points and (in the lower figures) metal atoms located in the unit cell.

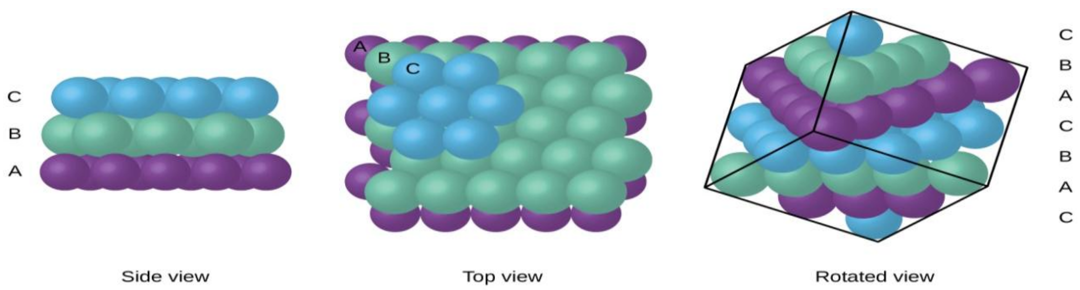
Face-Centered Cubic Structure (FCC): Many other metals, such as aluminum, copper, and lead, crystallize in an arrangement that has a cubic unit cell with atoms at all of the corners and at the centers of each face, as illustrated in Figure. This arrangement is called a face-centered cubic (FCC) solid. A FCC unit cell contains four atoms: one-eighth of an atom at each of the eight corners (8×18=18×18=1 atom from the corners) and one-half of an atom on each of the six faces (6×12=36×12=3 atoms from the faces). The atoms at the corners touch the atoms in the centers of the adjacent faces along the face diagonals of the cube. Because the atoms are on identical lattice points, they have identical environments.



Face-centered cubic structure

Fig.1.15 A face-centered cubic solid has atoms at the corners and, as the name implies, at the centers of the faces of its unit cells.

Atoms in an FCC arrangement are packed as closely together as possible, with atoms occupying 74% of the volume. This structure is also called cubic closest packing (CCP). In CCP, there are three repeating layers of hexagonally arranged atoms. Each atom contacts six atoms in its own layer, three in the layer above, and three in the layer below. In this arrangement, each atom touches 12 near neighbors, and therefore has a coordination number of 12. The fact that FCC and CCP arrangements are equivalent may not be immediately obvious, but why they are actually the same structure is illustrated in Figure 8.



Cubic closest packed structure

Fig.1.16 A CCP arrangement consists of three repeating layers (ABCABC…) of hexagonally arranged atoms.

Atoms in a CCP structure have a coordination number of 12 because they contact six atoms in their layer, plus three atoms in the layer above and three atoms in the layer below. By rotating our perspective, we can see that a CCP structure has a unit cell with a face containing an atom from layer A at one corner, atoms from layer B across a diagonal (at two corners and in the middle of the face), and an atom from layer C at the remaining corner. This is the same as a face-centered cubic arrangement.

Because closer packing maximizes the overall attractions between atoms and minimizes the total intermolecular energy, the atoms in most metals pack in this manner. We find two types of closest packing in simple metallic crystalline structures: CCP, which we have already encountered, and hexagonal closest packing (HCP) shown in Figure 9. Both consist of repeating layers of hexagonally arranged atoms. In both types, a second layer (B) is placed on the first layer (A) so that each atom in the second layer is in contact with three atoms in the first layer. The third layer is positioned in one of two ways. In HCP, atoms in the third layer are directly above atoms in the first layer (i.e., the third layer is also type A), and the stacking consists of alternating type A and type B close-packed layers (i.e., ABABAB⋯). In CCP, atoms in the third layer are not above atoms in either of the first two layers (i.e., the third layer is type C), and the stacking consists of alternating type A, type B, and type C close-packed layers (i.e., ABCABCABC⋯). About two–thirds of all metals crystallize in closest-packed arrays with coordination numbers of 12. Metals that crystallize in an HCP structure include Cd, Co, Li, Mg, Na, and Zn, and metals that crystallize in a CCP structure include Ag, Al, Ca, Cu, Ni, Pb, and Pt.

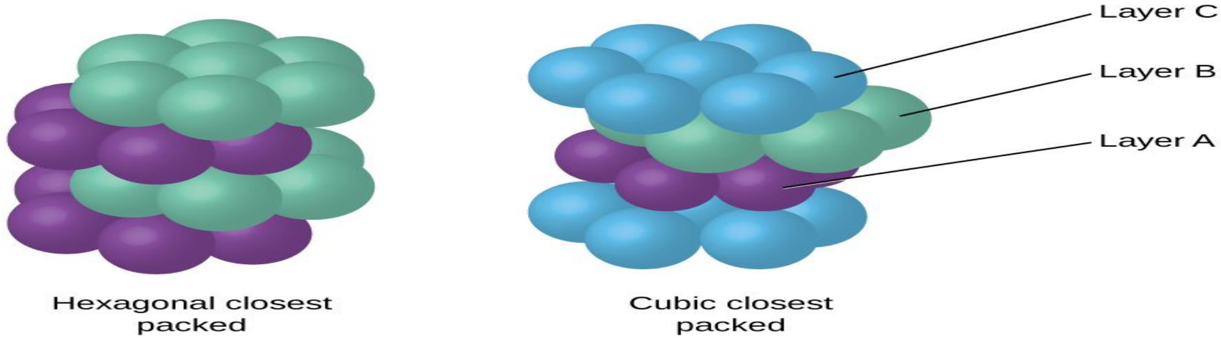
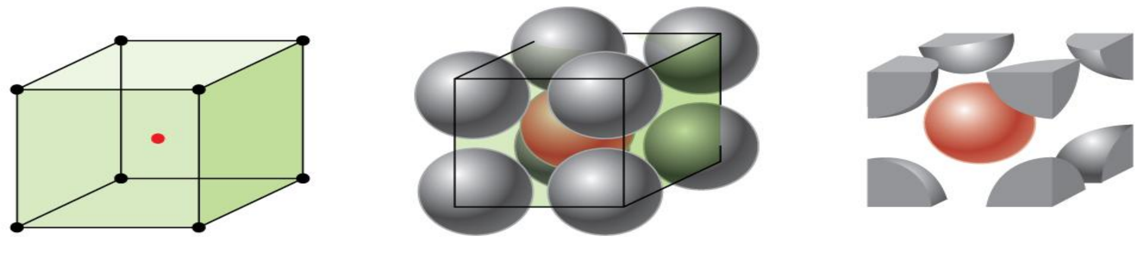


Fig.1.17 In both types of closest packing, atoms are packed as compactly as possible. Hexagonal closest packing consists of two alternating layers (ABABAB…). Cubic closest packing consists of three alternating layers (ABCABCABC…).

Body-Centered Cubic Structure (BCC): Some metals crystallize in an arrangement that has a cubic unit cell with atoms at all of the corners and an atom in the center, as shown in Figure. This is called a body-centered cubic (BCC) solid. Atoms in the corners of a BCC unit cell do not contact each other but contact the atom in the center. A BCC unit cell contains two atoms: one-eighth of an atom at each of the eight corners (8×18=18×18=1 atom from the corners) plus one atom from the center. Any atom in this structure touches four atoms in the layer above it and four atoms in the layer below it. Thus, an atom in a BCC structure has a coordination number of eight.



Body-centered cubic structure

Fig.1.18 In a body-centered cubic structure, atoms in a specific layer do not touch each other. Each atom touches four atoms in the layer above it and four atoms in the layer below it.

Atoms in BCC arrangements are much more efficiently packed than in a simple cubic structure, occupying about 68% of the total volume. Isomorphous metals with a BCC structure include K, Ba, Cr, Mo, W, and Fe at room temperature. (Elements or compounds that crystallize with the same structure are said to be isomorphous.)

SAQ.1

a) What do you mean by crystalline state of solids?
b) Give the characteristics of liquid crystal.
c) Define the Simple crystal structure SC and FCC.
d) The radius of a calcium ion is 94 pm and of an oxide ion is 146 pm. Predict the crystal structure of calcium oxide.

e) A metal crystallizes into two cubic system face centred cubic (fcc) and body centred cubic (bcc) whose unit cell lengths are 3.5 and 3.0Å respectively. Calculate the ratio of densities of fcc and bcc.

1.5 Unit Cell:

The smallest possible portion or part of the crystal lattice which repeats itself in different directions of the lattice is called the unit cell. Many unit cells combine to geometrically form the crystal lattice.

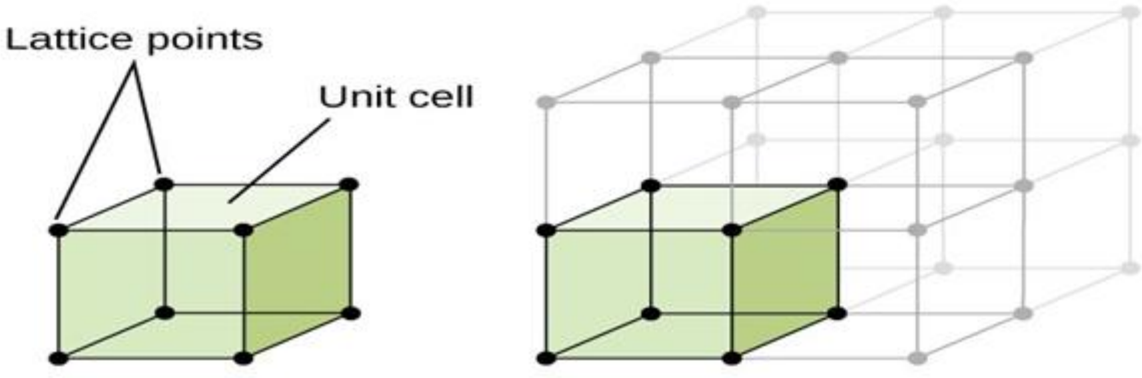


Fig.1.19 A unit cell is the shortest portion of a lattice

Characteristics of Unit Cell

The following characteristics define a unit cell:

- A unit cell has three edges a, b and c and three angles α, β and γ between the respective edges.
- The a, b and c may or may not be mutually perpendicular.
- The angle between edge b and c is α, a and c is β and that of between a and b is γ.

Unit cells are of two types namely:

- Primitive Unit Cells
- Centred Unit Cells

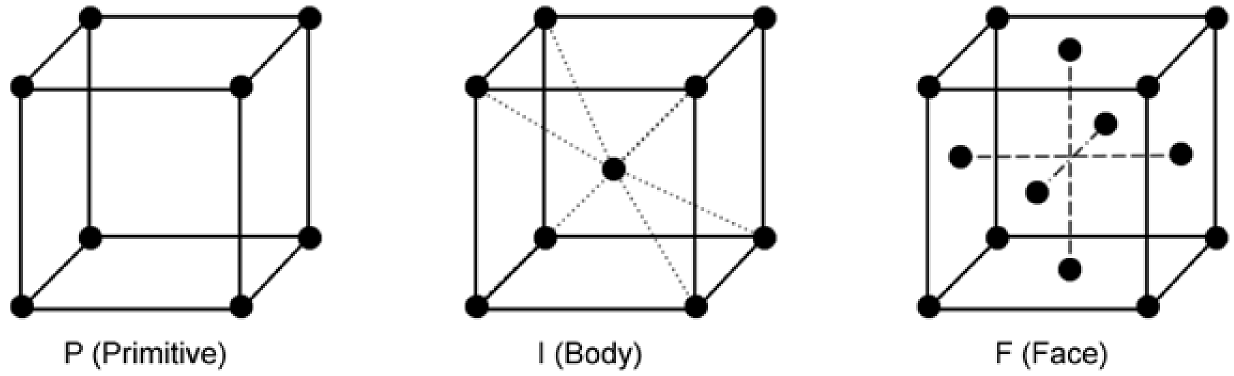


Fig.1.20 Types of Unit Cells

Primitive Unit Cells:

The unit cell in which the constituent particles (atoms, ions or molecules) are located only on the corners of the lattice is called A Primitive Unit Cell.

Centred Unit Cells:

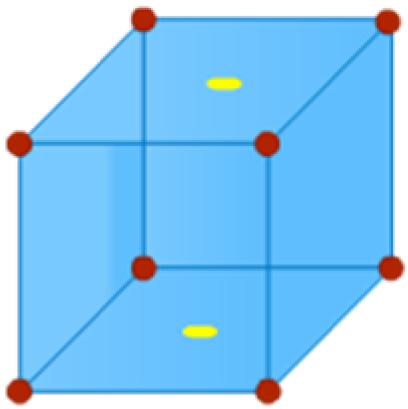


Fig.1.21 End Centred Unit Cell

The unit cell in which the constituent particles (atoms, ions or molecules) are located on the corners, as well as other positions of the lattice, is known as Centred Unit Cells. A centred unit cell is further divided into three types:

- Body Centred Unit Cells.
- Face Centred Unit Cells.
- End Centred Unit Cells.

Body Centred Unit Cells: The unit cell which contains one constituent particle (atom, molecule or ion) at its body centre and other constituent particles are located on the corners is called Body Centred Unit Cells.

Face Centred Unit Cells: The unit cell which contains constituent particles (atoms, molecules or ions) on each face of the unit cell and other constituent particles on the corners is called the Face Centred Unit Cell.

End Centred Unit Cells: In an end centred unit cell, one constituent particle (atom, molecule or ion) is present at the centre of opposite faces besides the ones located on the corners.

Bravais lattices: The Bravais lattice are the distinct lattice types which when repeated can fill the whole space. The lattice can therefore be generated by three unit vectors, a_1, a_2 and a_3 and a set of integers k, l and m so that each lattice point, identified by a vector r, can be obtained from:

$$r = k\ a_1 + l\ a_2 + m\ a_3$$

In two dimensions there are five distinct Bravais lattices, while in three dimensions there are fourteen. These fourteen lattices are further classified as shown in the table below where a_1, a_2 and a_3 are the magnitudes of the unit vectors and a, b and g are the angles between the unit vectors.

Name	Number of Bravais lattices	Conditions
Triclinic	1	$a_1 \neq a_2 \neq a_3$ $\alpha \neq \beta \neq \gamma$
Monoclinic	2	$a_1 \neq a_2 \neq a_3$ $\alpha = \beta = 90° \neq \gamma$
Orthorhombic	4	$a_1 \neq a_2 \neq a_3$ $\alpha = \beta = \gamma = 90°$
Tetragonal	2	$a_1 = a_2 \neq a_3$ $\alpha = \beta = \gamma = 90°$
Cubic	3	$a_1 = a_2 = a_3$ $\alpha = \beta = \gamma = 90°$
Trigonal	1	$a_1 = a_2 = a_3$ $\alpha = \beta = \gamma < 120° \neq 90°$
Hexagonal	1	$a_1 = a_2 \neq a_3$ $\alpha = \beta = 90°$

		$\gamma = 120°$

1.6 Classification of lattices:

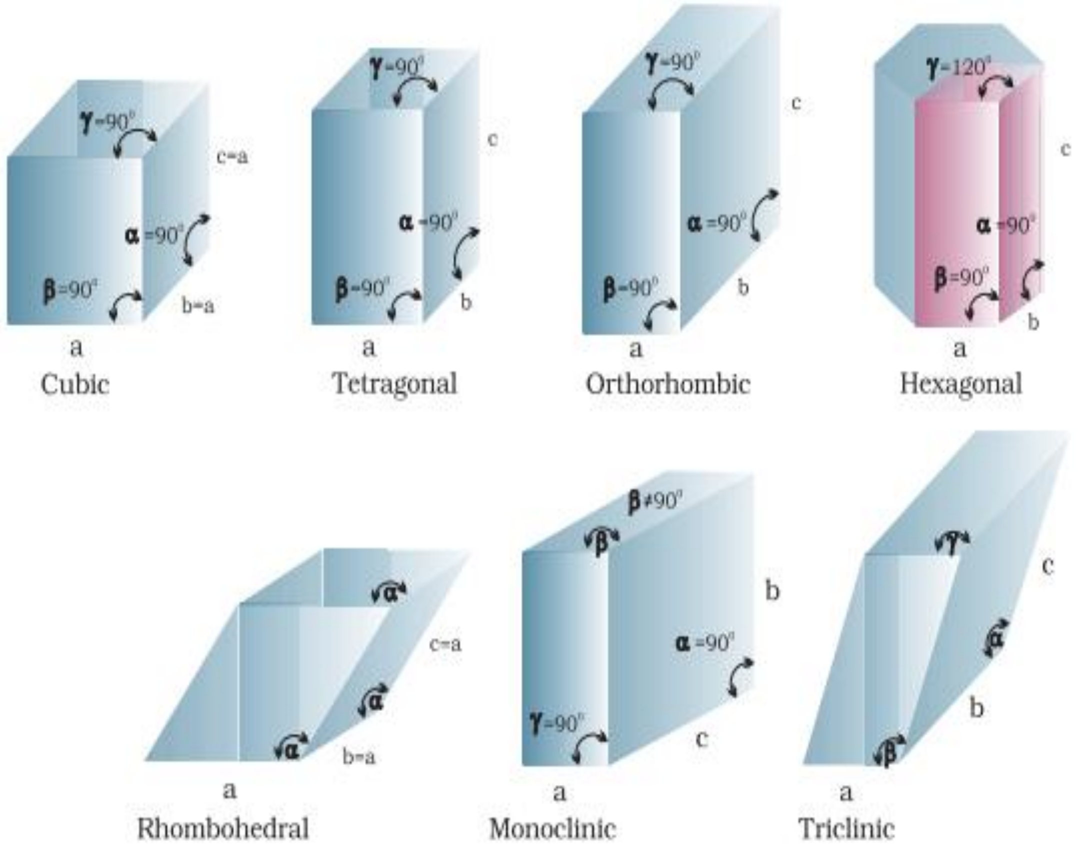


Fig.1.22 Types of unit cells which form crystal lattices

Formation of unit cells takes place in seven forms, namely:

- Cubic Lattice
- Tetragonal Lattice
- Orthorhombic Lattice
- Monoclinic Lattice
- Hexagonal Lattice
- Rhombohedral Lattice
- Triclinic Lattice

Cubic Lattice: Cubic lattice is formed into three possible geometries of unit cells: primitive, body-centred and face centred unit cells. In a cubic lattice, all the edges are equal and the angle between their faces is 90° that is, mutually perpendicular.

Tetragonal Lattice: The formation of tetragonal lattice takes place in two geometries of unit cells: primitive and body centred unit cells. In a tetragonal lattice, only one edge has different length and angle between respective edges is 90° that is, mutually perpendicular

Orthorhombic Lattice: There are four types of orthorhombic lattice mainly: primitive, end-centred, body-centred and face centred. In orthorhombic lattice, the edge lengths are unequal in nature and the angle between respective edges is 90° that is, mutually perpendicular
Monoclinic Lattice: Monoclinic lattice is formed from two types of unit cells namely: primitive and end centred. Monoclinic lattice has unequal sides and two angles between the faces of the lattice are 90°.

Hexagonal Lattice: Hexagonal lattice is formed from only one type of unit cell that is, primitive. In hexagonal lattice, only one side and two angles are 90° and one angle is 120°.
Rhombohedral Lattice: Rhombohedral Lattice is also formed from one type of unit cell that is, primitive. In Rhombohedral Lattice all the sides are equal and two angles between the faces of the rhombohedral lattice are less than 90°.

Triclinic Lattice: The formation of triclinic lattice also takes place from one type of unit cell that is, primitive. In triclinic lattice all the sides are unequal and none of the angles between the faces of the triclinic lattice are 90°.

The table given below can be used to summarize types of lattice formation.

Lattice	Types	Edge Length	Angles between faces	Examples
Cubic	Primitive, Body-centred, Face-centred	$a = b = c$	$\alpha = \beta = \gamma = 90^{\circ}$	NaCl, Copper and ZnS

Tetragonal	Primitive, Body-centred	a = b ≠ c	$\alpha = \beta = \gamma = 90°$	White tin, SnO_2, TiO_2 and $CaSO_4$
Orthorhombic	Primitive, Body-centred, Face-centred , End-centred	a ≠ b ≠ c	$\alpha = \beta = \gamma = 90°$	Rhombic Sulphur, $BaSO_4$ and KNO_3
Hexagonal	Primitive	a = b ≠ c	$\alpha = \beta = 90^{\circ}$ and $\gamma = 120°$	Graphite, ZnO and CdS
Rhombohedral	Primitive	a = b = c	$\alpha = \beta = \gamma \neq 90°$	$CaCO_3$ (Calcite) and HgS (cinnabar)
Monoclinic	Primitive, End-centred	a ≠ b ≠ c	$\alpha = \gamma = 90°$ $\beta \neq 90°$	Sulphur
Triclinic	Primitive	a ≠ b ≠ c	$\alpha \neq \beta \neq \gamma \neq 90^0$	H_3PO_3 , $CuSO_4.5H_2O$

Types of crystals on the basis of Bravais lattice:

Crystal Lattice is a three-dimensional representation of atoms and molecules arranged in a specific order/pattern. In other words, a crystal lattice can be defined as a geometrical arrangement of constituent particles of matter (atoms, ions or molecules) as points in space. There are total 14 possible three-dimensional lattices. Crystal lattices are also known by Bravais Lattices, named after the scientist Auguste Bravais.

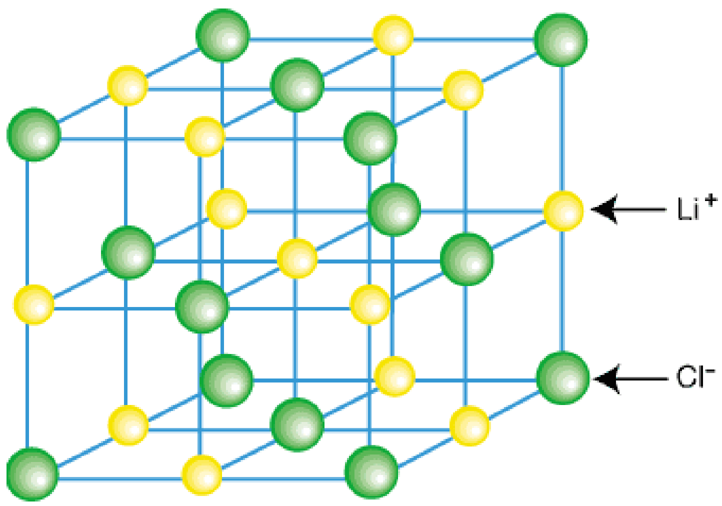


Fig.1.23 Example of a crystal lattice

Characteristics of Crystal Lattices:

The following characteristics are depicted by Bravais lattices:

- Each point in a lattice represents a lattice site or we can say lattice point.
- Each point denotes a particular type of constituent particles of matter be it an atom, molecule or an ion.
- By joining the lattices points inside the lattice we can define geometry of the lattice.

Bravais Lattice refers to the 14 different 3-dimensional configurations into which atoms can be arranged in crystals. The smallest group of symmetrically aligned atoms which can be repeated in an array to make up the entire crystal is called a unit cell.

There are several ways to describe a lattice. The most fundamental description is known as the Bravais lattice. In words, a Bravais lattice is an array of discrete points with an arrangement and orientation that look exactly the same from any of the discrete points, that is the lattice points are indistinguishable from one another.

Thus, a Bravais lattice can refer to one of the 14 different types of unit cells that a crystal structure can be made up of. These lattices are named after the French physicist Auguste Bravais.

Types of Bravais Lattices:

Out of 14 types of Bravais lattices some 7 types of Bravais lattices in three-dimensional space are listed in this subsection. Note that the letters a, b, and c have been used to denote the dimensions of the unit cells whereas the letters α, β, *and* γ denote the corresponding angles in the unit cells.

1. Cubic Systems:

In Bravais lattices with cubic systems, the following relationships can be observed.

$a = b = c$

$\alpha = \beta = \gamma = 90^o$

The 3 possible types of cubic cells have been illustrated below.

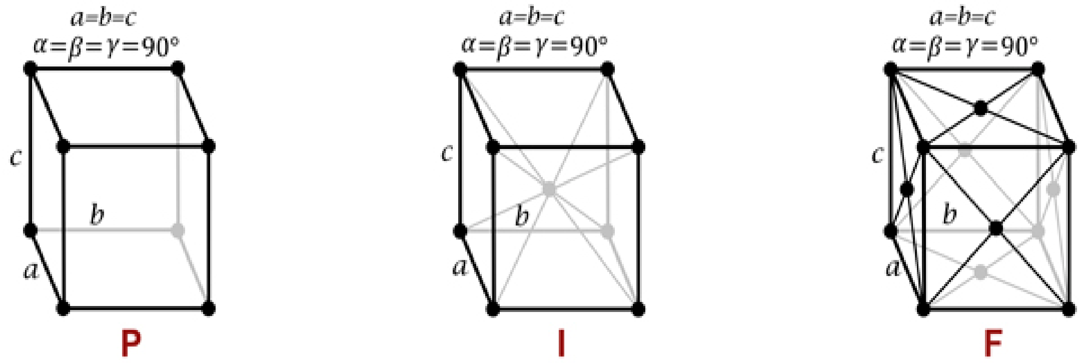


Fig.1.24 Structure of Cubic Bravais Lattice

These three possible cubic Bravais lattices are –

- Primitive (or Simple) Cubic Cell (P)
- Body-Centered Cubic Cell (I)
- Face-Centered Cubic Cell (F)

Examples: Polonium has a simple cubic structure, Iron has a body-centered cubic structure, and copper has a face-centered cubic structure.

2. Orthorhombic Systems:

The Bravais lattices with orthorhombic systems obey the following equations:

$a \neq b \neq c$

$\alpha = \beta = \gamma = 90^o$

The four types of orthorhombic systems (simple, base centered, face-centered, and body-centered orthorhombic cells) are illustrated below.

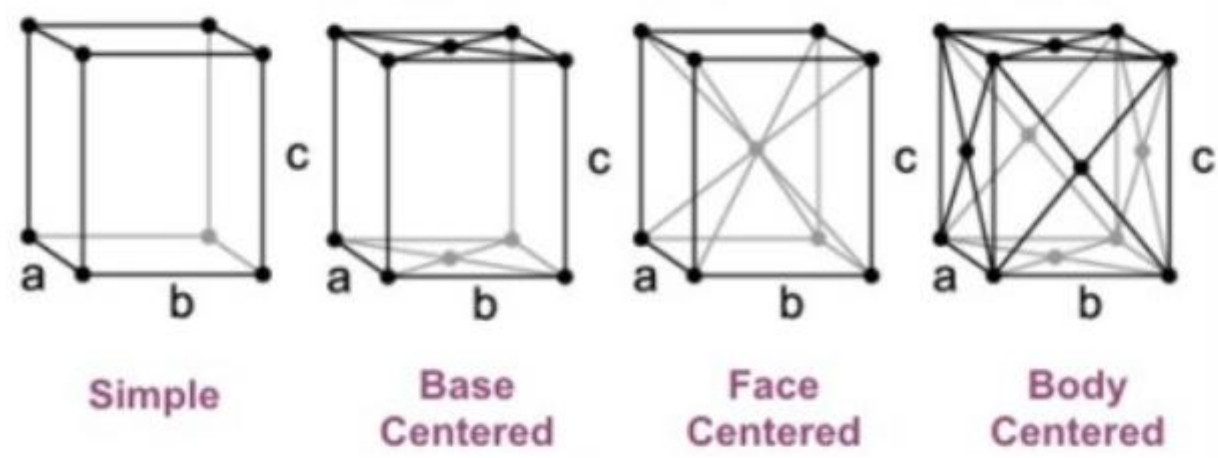


Fig.1.25 Structure of Orthorhombic Bravais Lattice

Examples of Orthorhombic Systems:

- Rhombic Sulfur has a simple orthorhombic structure
- Magnesium sulfate heptahydrate ($MgSO_4.7H_2O$) is made up of a base centered orthorhombic structure.
- Potassium Nitrate has a structure which is body-centered orthorhombic.
- An example of a substance with a face-centered orthorhombic structure is barium sulfate.

3. Tetragonal Systems:

In tetragonal Bravais lattices, the following relations are observed:

$a = b \neq c$

$\boldsymbol{\alpha} = \boldsymbol{\beta} = \boldsymbol{\gamma} = 90^o$

The two types of tetragonal systems are simple tetragonal cells and body-centered tetragonal cells, as illustrated below.

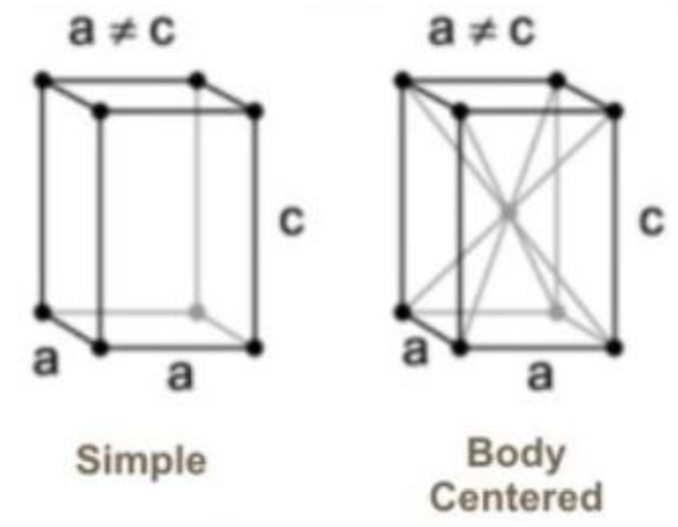


Fig.1.26 Structure of Tetragonal Bravais Lattice

Examples of tetragonal Bravais lattices are – stannic oxide (simple tetragonal) and titanium dioxide (body-centered tetragonal)

4. Monoclinic Systems:

Bravais lattices having monoclinic systems obey the following relations:

$a \neq b \neq c$

$\boldsymbol{\beta} = \boldsymbol{\gamma} = 90^o$ and $\boldsymbol{\alpha} \neq 90^o$

The two possible types of monoclinic systems are primitive and base centered monoclinic cells, as illustrated below.

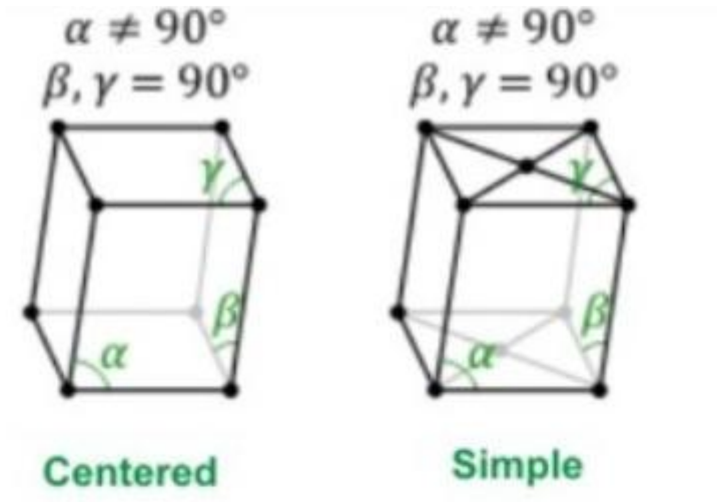


Fig.1.27 Structure of Monoclinic Bravais

Lattice cubic cells are – Monoclinic sulfur (simple monoclinic) and sodium sulfate decahydrate (base centered monoclinic)

5. Triclinic System:

There exists only one type of triclinic Bravais lattice, which is a *primitive cell*. It obeys the following relationship.

$a \neq b \neq c$

$\boldsymbol{\alpha} \neq \boldsymbol{\beta} \neq \gamma \neq 90^o$

An illustration of a simple triclinic cell is given below.

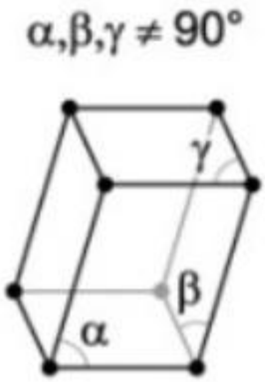


Fig.1.28 Structure of Triclinic Bravais

Lattice Such unit cells are found in the structure of potassium dichromate (Chemical formula $K_2Cr_2O_7$).

6. Rhombohedral System:

Only the primitive unit cell for a rhombohedral system exists. Its cell relation is given by:

$a = b = c$

$\boldsymbol{\alpha} = \boldsymbol{\beta} = \boldsymbol{\gamma} \neq 90^o$

An illustration of the primitive rhombohedral cell is provided below.

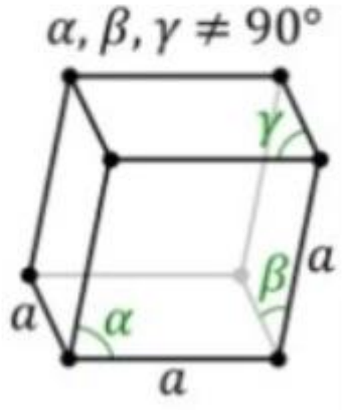


Fig.1.29 Structure of Rhombohedral Bravais Lattice

Calcite and sodium nitrate are made up of simple rhombohedral unit cells.

7. Hexagonal System:

The only type of hexagonal Bravais lattice is the simple hexagonal cell. It has the following relations between cell sides and angles.

$a = b \neq c$

$\boldsymbol{\alpha} = \boldsymbol{\beta} = 90^o$ and $\boldsymbol{\gamma} = 120^o$

An illustration of a simple hexagonal cell is provided below.

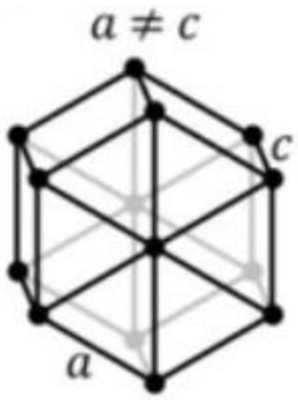


Fig.1.30 Structure of Hexagonal Bravais Lattice

Zinc oxide and beryllium oxide are made up of simple hexagonal unit cells.

Thus, it can be noted that all 14 possible Bravais lattices differ in their cell length and angle relationships. It is important to keep in mind that the Bravais lattice is not always the same as the crystal lattice.

1.7 Direct and reciprocal lattice:

Direct lattice:

The direct lattice represents the triple periodicity of the ideal infinite perfect periodic structure that can be associated to the structure of a finite real crystal. To express this periodicity one calls crystal pattern an object in point space E^n (direct space) that is invariant with respect to three linearly independent translations, t_1, t_2 and t_3. One distinguishes two kinds of lattices, the vector lattices and the point lattices.

Any translation $t = u^i t_i$ (u^i arbitrary integers) is also a translation of the pattern and the infinite set of all translation vectors of a crystal pattern is the vector lattice L of this crystal pattern.

Given an arbitrary point P in point space, the set of all the points P_i deduced from one of them by a translation $PP_i = t_i$ of the vector lattice L is called the point lattice.

A basis a, b, c of the vector space V^n is a crystallographic basis of the vector lattice L if every integral linear combination $t = u\,a + v\,b + w\,c$ is a lattice vector of L. It is called a primitive basis if every lattice vector t of L may be obtained as an integral linear combination of the basis vectors, a, b, c. Referred to any crystallographic basis the coefficients of each lattice vector are either integral or rational, while in the case of a primitive basis they are integral. Non-primitive bases are used conventionally to describe centered lattices.

The parallelepiped built on the basis vectors is the unit cell. Its volume is given by the triple scalar product, V = (a, b, c).

If the basis is primitive, the unit cell is called the primitive cell. It contains only one lattice point. If the basis is non-primitive, the unit cell is a multiple cell and it contains more than one lattice point. The multiplicity of the cell is given by the ratio of its volume to the volume of a primitive cell.

Reciprocal lattice:

The reciprocal lattice is constituted of the set of all possible linear combinations of the basis vectors a*, b*, c* of the reciprocal space. A point (*node*), *H*, of the reciprocal lattice is defined by its position vector:

$OH = r^*_{hkl} = h\, a^* + k\, b^* + l\, c^*$.

If *H* is the *n*th node on the row *OH*, one has:

$OH = n\, OH_1 = n\, (h_1\, a^* + k_1\, b^* + l_1\, c^*)$,

where H_1 is the first node on the row *OH* and h_1, k_1, l_1 are relatively prime.

Geometrical applications:

Each vector $OH = r^*_{hkl} = h\, a^* + k\, b^* + l\, c^*$ of the reciprocal lattice is associated with a family of direct lattice planes. It is normal to the planes of the family, and the lattice spacing of the family is $d = 1/OH_1 = n/OH$ if *H* is the *n*th node on the reciprocal lattice row *OH*. One usually sets $d_{hkl} = d/n = 1/OH$. If $OP = x\, a + y\, b + z\, c$ is the position vector of a point of a lattice plane, the equation of the plane is given by $OH_1 \cdot OP = K$ where *K* is a constant integer. Using the properties of the scalar product of a reciprocal space vector and a direct space vector, this equation is $OH_1 \cdot OP = h_1 x + k_1 y + l_1 z = K$. The Miller indices of the family are h_1, k_1, l_1. The subscripts of the Miller indices will be dropped hereafter.

The Miller indices of the family of lattice planes parallel to two direct space vectors, $r_1 = u_1\, a + v_1\, b + w_1\, c$ and $r_2 = u_2\, a + v_2\, b + w_2\, c$ are proportional to the coordinates in reciprocal space, *h*, *k*, *l*, of the vector product of these two vectors:

$h/(v_1 w_2 - v_2 w_1) = k/(w_1 u_2 - w_2 u_1) = l/(u_1 v_2 - u_2 v_1)$.

The coordinates *u*, *v*, *w* in direct space of the zone axis intersection of two families of lattice planes of Miller indices h_1, k_1, l_1 and h_2, k_2, l_2, respectively, are proportional to the coordinates of the vector product of the reciprocal lattice vectors associated with these two families:

$u/(k_1 l_2 - k_2 l_1) = v/(l_1 h_2 - l_2 h_1) = w/(h_1 k_2 - h_2 k_1)$.

Centred lattices:

Direct lattice	Reciprocal lattice

Bravais letter	Centring vectors	Unit-cell volume V_c	Bravais letter	Multiple unit cell	Unit cell volume V^*_c
P	0	*V*	*P*	a^*_c, b^*_c, c^*_c	V^*
A	$\tfrac{1}{2}b_c+\tfrac{1}{2}c_c$	2*V*	*A*	$a^*_c, 2b^*_c, 2c^*_c$	$\tfrac{1}{2}V^*$
B	$\tfrac{1}{2}c_c+\tfrac{1}{2}a_c$	2*V*	*B*	$2a^*_c, b^*_c, 2c^*_c$	$\tfrac{1}{2}V^*$
C	$\tfrac{1}{2}a_c+\tfrac{1}{2}b_c$	2*V*	*C*	$2a^*_c, 2b^*_c, c^*_c$	$\tfrac{1}{2}V^*$
I	$\tfrac{1}{2}a_c+\tfrac{1}{2}b_c+\tfrac{1}{2}c_c$	2*V*	*F*	$2a^*_c, 2b^*_c, 2c^*_c$	$\tfrac{1}{2}V^*$
F	$\tfrac{1}{2}a_c+\tfrac{1}{2}b_c$	4*V*	*I*	$2a^*_c, 2b^*_c, 2c^*_c$	$\tfrac{1}{4}V^*$
	$\tfrac{1}{2}b_c+\tfrac{1}{2}c_c$				
	$\tfrac{1}{2}c_c+\tfrac{1}{2}a_c$				
R	0	*V*	*R*	a^*_c, b^*_c, c^*_c	V^*
(rhombohedral axes)					
R	$\tfrac{2}{3}a^*_c+\tfrac{1}{3}b^*_c+\tfrac{1}{3}c^*_c$	3*V*	*R*	$3a^*_c, 3b^*_c, 3c^*_c$	$\tfrac{1}{3}V^*$
(hexagonal axes)	$\tfrac{1}{3}a^*_c+\tfrac{2}{3}b^*_c+\tfrac{2}{3}c^*_c$				

where a_c, b_c, c_c are the basis vectors of the conventional multiple cell and a^*_c, b^*_c, c^*_c the corresponding reciprocal lattice vectors.

An elementary proof that the reciprocal lattice of a face-centred lattice *F* is a body-centred lattice *I* and, reciprocally, is given in <u>*The Reciprocal Lattice*</u> (Teaching Pamphlet No. 4 of the International Union of Crystallography).

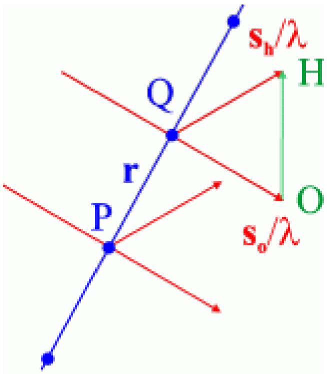


Fig.1.31 Diffraction condition in reciprocal space

The condition that the waves outgoing from two point scatterers separated by a lattice vector r = u a + v b + w c (u, v, w integers) be in phase is that the scalar product (s_h/λ - s_o/λ) · r, where s_h and s_o are unit vectors in the scattered and incident directions, respectively, be an integer, n. This condition is satisfied whatever r if the diffraction vector (OH = s_h/λ - s_o/λ) is of the form:

$(s_h/\lambda - s_o/\lambda) = h\ a^* + k\ b^* + l\ c^*$,

where h, k, l are integers, namely the diffraction vector OH is a vector of the reciprocal lattice (Fig.).

A node of the reciprocal lattice is therefore associated with each Bragg reflection on the lattice planes of Miller indices (h/K, k/K, l/K). It is called the *hkl* reflection.

The relation s_h/λ - s_o/λ = OH generalizes the Laue equations. It is equivalent to Bragg's law, as can be seen in Fig.

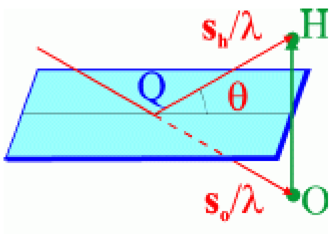


Fig.1.32

Consider the lattice plane passing through lattice point Q and perpendicular to reciprocal-lattice vector OH and let θ be the angle between the incident, s_o, or the reflected, s_h, directions and the lattice plane. It can be seen from the figure that

$$OH/2 = \sin\theta/\lambda,$$

and, since $OH = n/d$ (d lattice spacing of the family of lattice planes associated with OH) and $d_{hkl} = d/n$:

$2\ d \sin\theta = n\ \lambda$, or $2\ d_{hkl} \sin\theta = \lambda$,

which is Bragg's law. n is the order of the reflection.

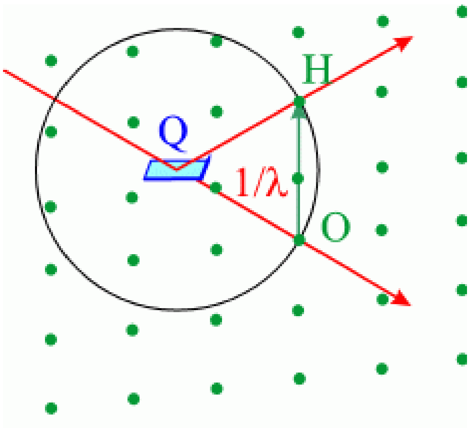


Fig.1.33

Another way to express the diffraction condition in reciprocal space is to consider a sphere centred at a node Q of the direct lattice, of radius $1/\lambda$ and passing through the origin O of the reciprocal lattice (Fig.). If it passes through another node, H, of the reciprocal lattice, Bragg's law is satisfied for the family of direct lattice planes associated with that node and of lattice spacing $d_{hkl} = n/OH$ if H is the nth node on the row OH (n = 2 in the example of Fig.). This sphere is called the Ewald sphere.

Miller indices and planes:

Miller Indices are a symbolic vector representation for the orientation of an atomic plane in a crystal lattice & are defined as the reciprocals of the fractional intercepts which the plane makes with the crystallographic axes. To find the Miller indices of a plane, take the following steps.

Rules for Miller Indices:

- Determine the intercepts of the face along the crystallographic axes, in terms of unit cell dimensions.
- Take the reciprocals
- Clear fractions
- Reduce to lowest terms

For example, if the x-, y-, and z- intercepts are 2, 1, and 3, the Miller indices are calculated as:

- Take reciprocals: 1/2, 1/1, 1/3
- Clear fractions (multiply by 6): 3, 6, 2
- Reduce to lowest terms (already there)

Thus, the Miller indices are 3,6,2. If a plane is parallel to an axis, its intercept is at infinity and its Miller index is zero. A generic miller index is denoted by (hkl).

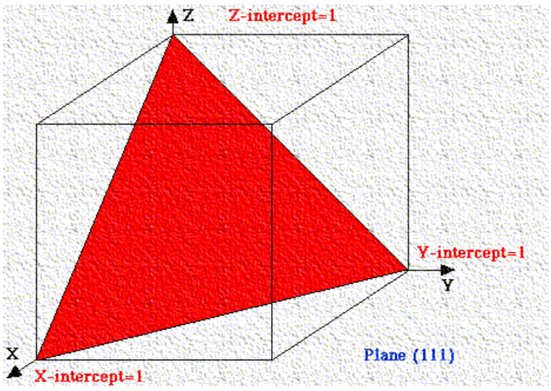


Fig.1.34 Miller indices

If a plane has negative intercept, the negative number is denoted by a bar above the number. Never alter negative numbers. For example, do not divide -1, -1, -1 by -1 to get 1,1,1. This implies symmetry that the crystal may not have!

Some General Principles

- If a Miller index is zero, the plane is parallel to that axis.
- The smaller a Miller index, the more nearly parallel the plane is to the axis.
- The larger a Miller index, the more nearly perpendicular a plane is to that axis.
- Multiplying or dividing a Miller index by a constant has no effect on the orientation of the plane
- Miller indices are almost always small.

Why Miller Indices?

- Using reciprocals spares us the complication of infinite intercepts.

- Formulas involving Miller indices are very similar to related formulas from analytical geometry.
- Specifying dimensions in unit cell terms means that the same label can be applied to any face with a similar stacking pattern, regardless of the crystal class of the crystal. Face 111 always steps the same way regardless of crystal system.

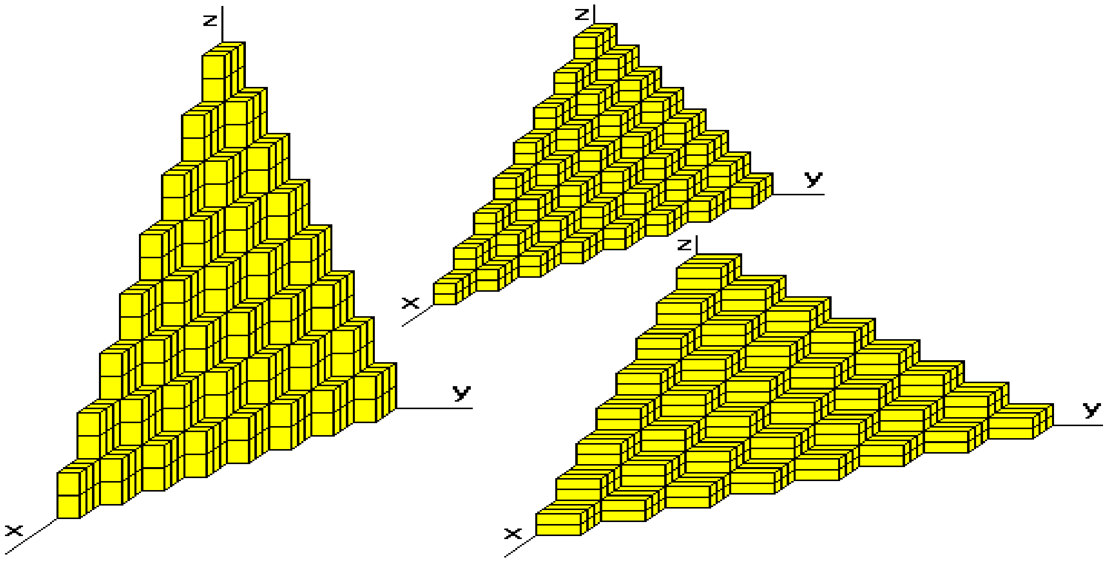


Fig.1.35 Planes of Miller indices

SAQ.2

a) What do mean by Unit cell and its characteristics?

b) Define the types of crystals on the basis of Bravais lattice.

c) What do you mean by Miller indices and its planes?

d) The density of a face-centred unit cell is 6.23 g cm^{-3}. Given the atomic mass of a single atom is 60, evaluate the edge length of the unit cell. (Take value of $N_A = 6.022 \times 10^{23}$)

e) Determine the Miller indices of a plane which is parallel to the x-axis and cuts intercepts of 2 and ½, respectively along y and z axes.

1.8 X-ray diffraction:

X-ray diffraction is a phenomenon in which the atoms of a crystal, by virtue of their uniform spacing, cause an interference pattern of the waves present in an incident beam of X rays. The atomic planes of the crystal act on the X rays in exactly the same manner as does a uniformly ruled grating on a beam of light. See also Bragg law; Laue diffraction pattern.

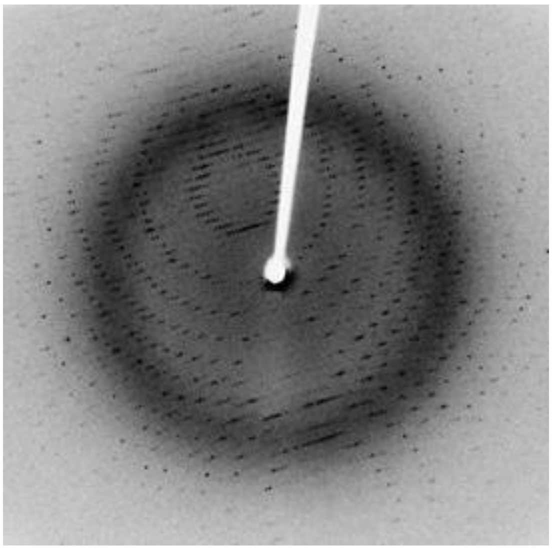


Fig.1.36 X-ray diffraction pattern of a crystallized enzyme

X-rays have photon energies in the range 100ev - 100Kev. In structural characterization, mainly short wave length x-rays are used, with a range of 10>0.1 angstroms. X-rays are typically produced by synchrotron radiation or by x-ray tubes. X-ray tubes are the main source used in laboratory x-ray instruments. Synchrotron facilities are far more expensive to use, but give improved results due to being more than a thousand times more intense than x-ray tubes.

When a x-ray beam hits the sample, it interacts with electrons in the atoms. Some photons in the incident beam will be deflected away from their original path. The deflected photons with unchanged wavelength have been elastically scattered (photons not lost energy), it is these photons that are measured in diffraction experiments since they hold information about the electron distribution in the material.

The peaks generated in x-ray diffraction are related to the atomic distances by Bragg's law.

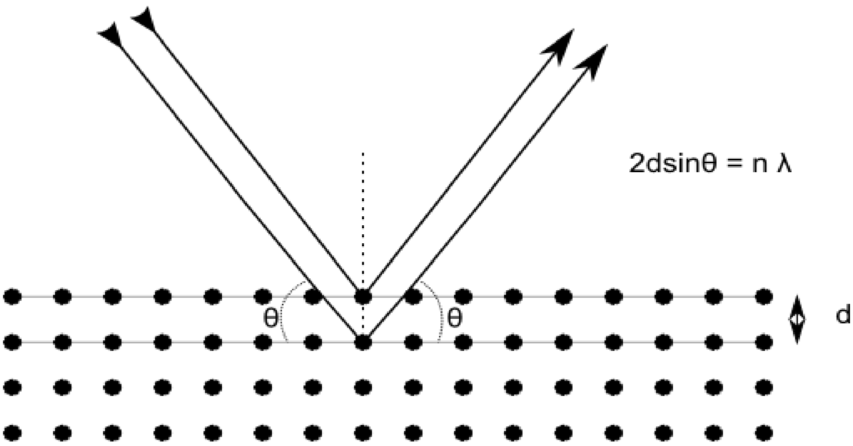


Fig.1.37 X-ray diffraction vector diagram

Ewalds Sphere:

The relationship between the crystals reciprocal lattice, the x-ray wavelength and the angle of diffraction for a given reflection can be demonstrated by a construct called Ewalds Sphere. The center of the sphere is placed on the crystal and is given a radius of one reciprocal lattice unit. An x-ray reflection is produced each time a reciprocal lattice point cuts the sphere of reflection.

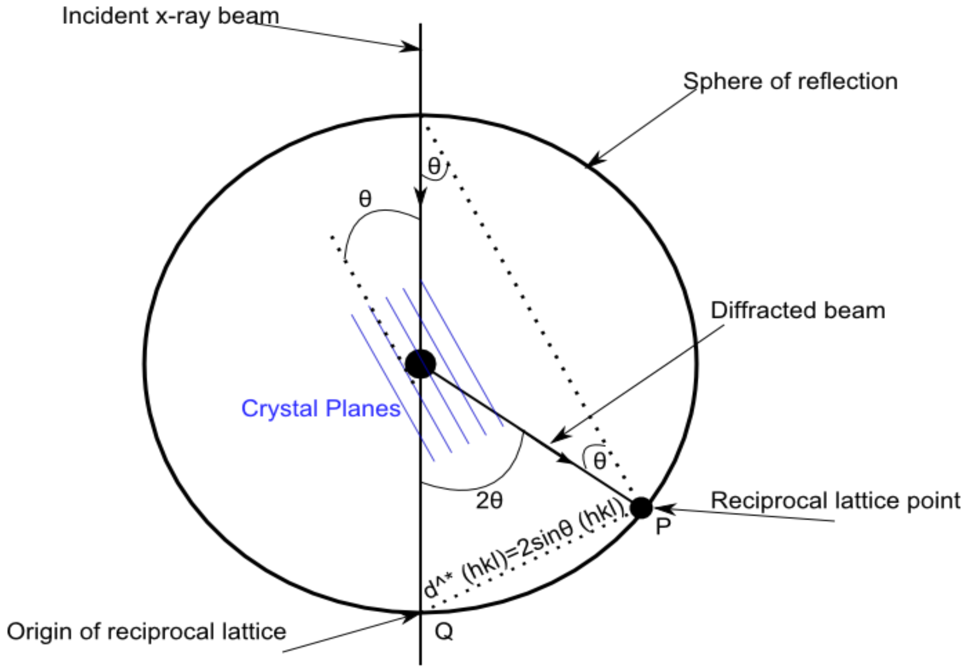


Fig.1.38 Ewalds Sphere

X-ray diffractometer:

The picture below illustrates a modern x-ray diffractometer used in the department for structural characterisation of semiconductors.

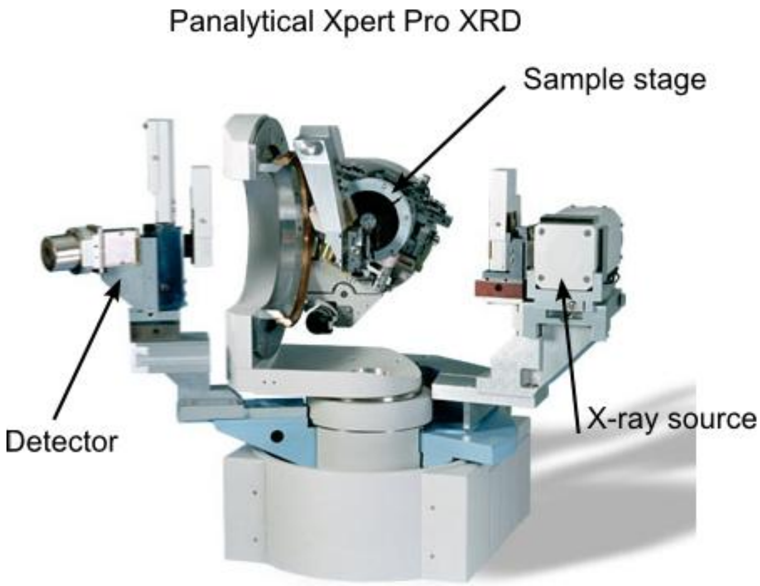


Fig.1.39 X-ray diffractometer

Structural characterisation-rocking Curves:

Much information on semiconductor structure can be inferred from a "rocking curve", including layer thickness, relaxation and composition. In a rocking curve the detector is fixed on the centroid position of a bragg peak and the detector is rocked over the bragg angle θ.

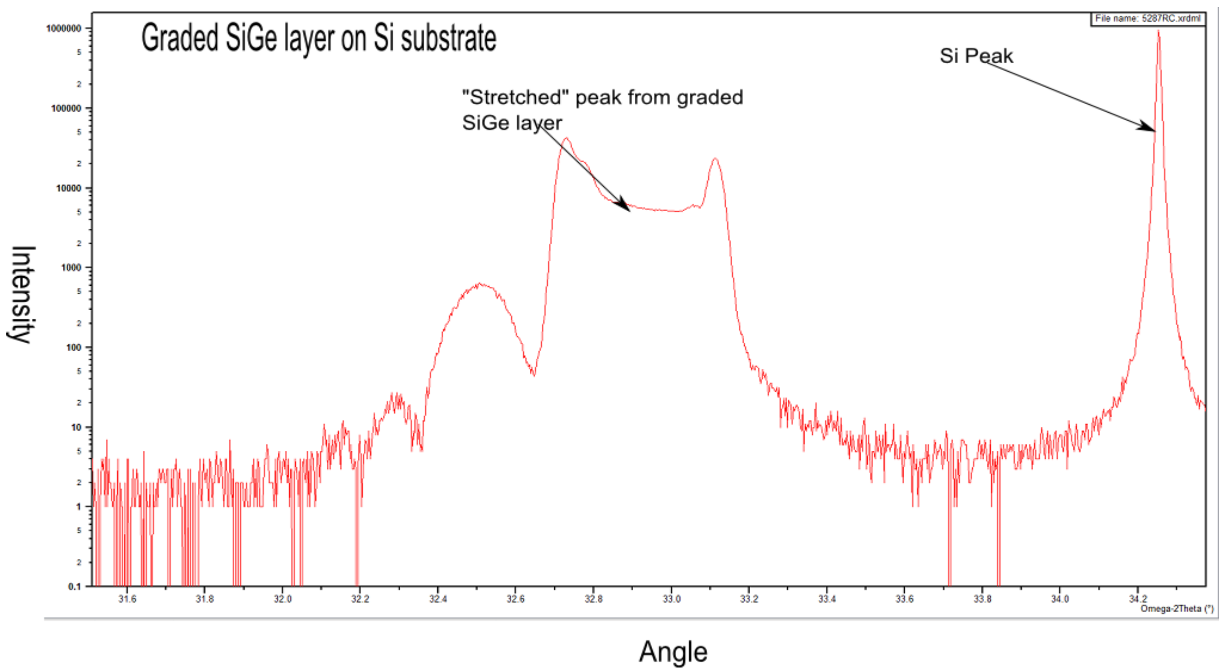


Fig.1.40 Structural characterization-rocking Curves

Bragg's Law using in X-ray reflection:

The structures of crystals and molecules are often being identified using x-ray diffraction studies, which are explained by Bragg's Law. The law explains the relationship between an x-ray light shooting into and its reflection off from crystal surface.

Bragg's Law was introduced by Sir W.H. Bragg and his son Sir W.L. Bragg. The law states that when the x-ray is incident onto a crystal surface, its angle of incidence, θ, will reflect back with a same angle of scattering, θ. And, when the path difference, d is equal to a whole number, n, of wavelength, a constructive interference will occur.

Consider a single crystal with aligned planes of lattice points separated by a distance *d*. Monochromatic X-rays A, B, and C are incident upon the crystal at an angle *θ*. They reflect off atoms X, Y, or Z.

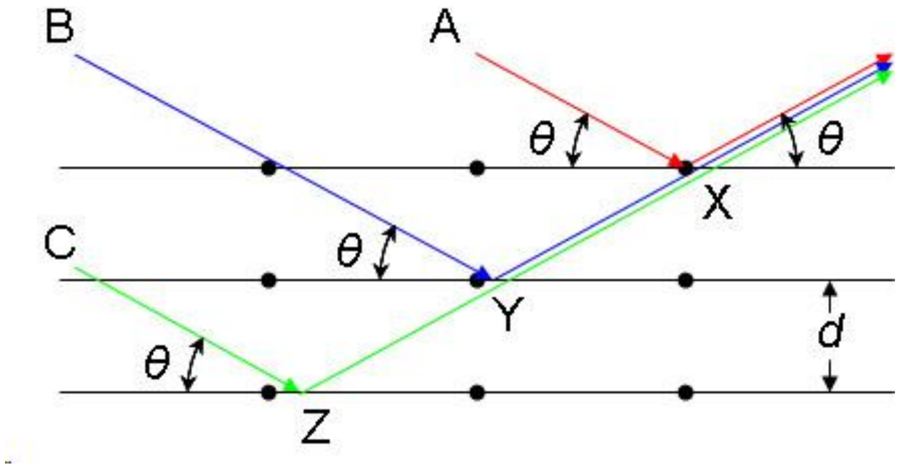


Fig.1.41 Bragg's Law using in X-ray reflection

The path difference between the ray reflected at atom X and the ray reflected at atom Y can be seen to be 2YX. From the Law of Sines we can express this distance YX in terms of the lattice distance and the X-ray incident angle:

If the path difference is equal to an integer multiple of the wavelength, then X-rays A and B (and by extension C) will arrive at atom X in the same phase. In other words, given the following conditions:

then the scattered radiation will undergo constructive interference and thus the crystal will appear to have reflected the X-radiation. If, however, this condition is not satisfied, then destructive interference will occur.

Bragg's Law

$$n\lambda=2d\sin\theta$$

where:

- λ is the wavelength of the x-ray,
- d is the spacing of the crystal layers (path difference),
- θ is the incident angle (the angle between incident ray and the scatter plane), and
- n is an integer

The principle of Bragg's law is applied in the construction of instruments such as Bragg spectrometer, which is often used to study the structure of crystals and molecules.

Applications of Bragg's Law:

- In X-ray diffraction (XRD) the interplanar spacing (d-spacing) of a crystal is used for identification and characterization purposes. In this case, the wavelength of the incident X-ray is known and measurement is made of the incident angle (Θ) at which constructive interference occurs. Solving Bragg's Equation gives the d-spacing between the crystal lattice planes of atoms that produce the constructive interference. A given unknown crystal is expected to have many rational planes of atoms in its structure; therefore, the collection of "reflections" of all the planes can be used to uniquely identify an unknown crystal. In general, crystals with high symmetry (e.g. isometric system) tend to have relatively few atomic planes, whereas crystals with low symmetry (in the triclinic or monoclinic systems) tend to have a large number of possible atomic planes in their structures.

- In the case of wavelength dispersive spectrometry (WDS) or X-ray fluorescence spectroscopy (XRF), crystals of known d-spacings are used as analyzing crystals in the spectrometer. Because the position of the sample and the detector is fixed in these applications, the angular position of the reflecting crystal is changed in accordance with Bragg's Law so that a particular wavelength of interest can be directed to a detector for quantitative analysis. Every element in the Periodic Table has a discrete energy difference between the orbital "shells" (e.g. K, L, M), such that every element will produce X-rays of a fixed wavelength. Therefore, by using a spectrometer crystal (with fixed d-spacing of the crystal) and positioning the crystal at a unique and fixed angle (Θ), it is possible to detect and quantify elements of interest based on the characteristic X-ray wavelengths produced by each element.

1.9 Generalized Hooke's Law:

The generalized Hooke's Law can be used to predict the deformations caused in a given material by an arbitrary combination of stresses. The linear relationship between stress and strain applies for $0 \le \sigma \le \sigma_{Yield}$

$$\varepsilon_x = \frac{\sigma_x}{E} - \nu\frac{\sigma_y}{E} - \nu\frac{\sigma_z}{E}$$

$$\varepsilon_y = -\nu\frac{\sigma_x}{E} + \frac{\sigma_y}{E} - \nu\frac{\sigma_z}{E}$$

$$\varepsilon_z = -\nu\frac{\sigma_x}{E} - \nu\frac{\sigma_y}{E} + \frac{\sigma_z}{E}$$

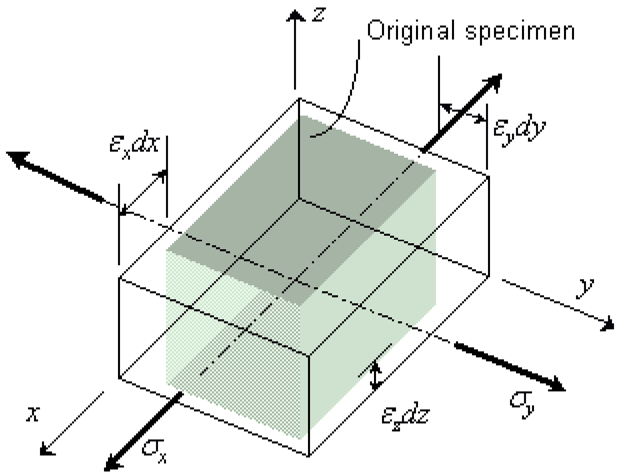


Fig.1.42 The generalized Hooke's Law

Where:

E is the Young's Modulus

n is the Poisson Ratio

The generalized Hooke's Law also reveals that strain can exist without stress. For example, if the member is experiencing a load in the y-direction (which in turn causes a stress in the y-direction), the Hooke's Law shows that strain in the x-direction does not equal to zero. This is because as material is being pulled outward by the y-plane, the material in the x-plane moves inward to fill in the space once occupied, just like an elastic band becomes thinner as you try to pull it apart. In this situation, the x-plane does not have any external force acting on them but they experience a change in length. Therefore, it is valid to say that strain exist without stress in the x-plane.

Elastic constants of cubic crystals:

The way planetary interiors deform depends primarily upon the elastic properties of the constituent polycrystalline materials. In particular, the speed of body waves passing through a material is entirely dependent upon the ratio of the elastic modulus of that material to its density. Whenever any external force is applied to a system, there is a resultant strain; similarly, whenever a system is strained in some way, there is then some stress upon the system. For example, squashing a jelly will deform it; deforming a jelly will result in a restoring force or stress eager to return the jelly to its original shape.

Stress, Strain and Elastic Moduli: An elastic modulus is just the ratio of stress to the associated strain. We wish to understand only the basic elastic equations and the physical meaning of these equations.

For a stress, s (hydrostatic, shear, axial...), resulting in an elastic deformation strain, e:

$$\sigma = M\,\varepsilon$$

where M is an elastic modulus (bulk, shear, Young's...).

For a hydrostatic stress (i.e., equally applied forces in all directions), which is often assumed within deep planetary interiors, the stress is the hydrostatic pressure:

$$\sigma = \Delta P$$

and the strain is the relative change in volume of the system:

$$\varepsilon = -\frac{\Delta V}{V}$$

Therefore:

$$\Delta P = -M\frac{\Delta V}{V}$$

and the elastic modulus in this case is the incompressibility or bulk modulus:

$$M = -V\frac{\Delta P}{\Delta V} \equiv K$$

Stress, Strain, and Tensors: The stress, s, does not have to be hydrostatic; there may be unequally applied stresses in all directions, and therefore the stress is tensorial:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$

where s_{ij} is the stress acting in the x_i direction on the plane perpendicular to the x_j direction; $s_{i=j}$ are the axial stresses and $s_{i\neq j}$ are the shear stresses.

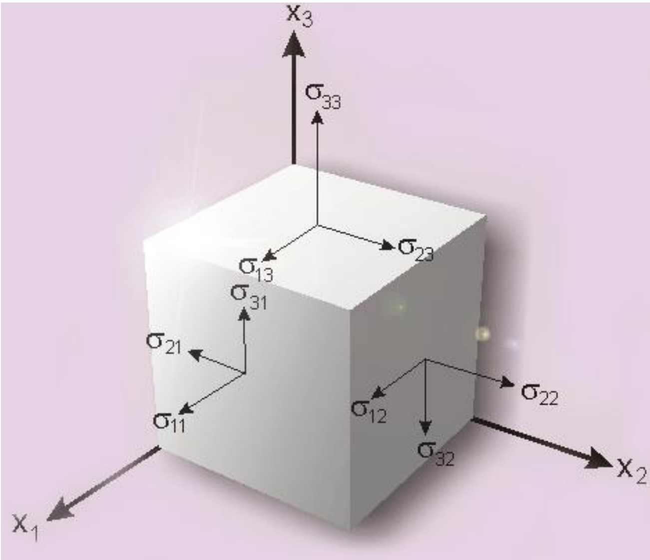


Fig.1.43 Elastic constants of cubic crystals

The strain e_{ij} is also a second order tensor. Therefore the elastic moduli, or elastic constants, are fourth order tensors:

$$\sigma_{ij} = \sum c_{ijkl}\, \varepsilon_{kl}$$

The stress, s, and the strain, e, must be symmetric, and the nature of c_{ijkl} depends on the symmetry of the crystal. It is customary to use a contracted notation thus:

$c_{1111} \neq c_{11}$ elastic constant relations s_{11} to e_{11}

$c_{1122} \neq c_{12}$ elastic constant relations s_{11} to e_{22}

$c_{2323} \neq c_{44}$ elastic constant relations s_{23} to e_{23}

In general, 11≠1; 22≠2; 33≠3; 23=32≠4; 13=31≠5; 12=21≠6.

Single, low symmetry crystals:

There are a maximum of 21 elastic constants for a crystalline body:

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ & & c_{33} & c_{34} & c_{35} & c_{36} \\ & & & c_{44} & c_{45} & c_{46} \\ & & & & c_{55} & c_{56} \\ & & & & & c_{66} \end{pmatrix}$$

N.B. symmetry allows $c_{12} \neq c_{21}$, etc..

However, all c_{ij} are rarely used, and it is therefore convenient to simplify matters by reducing the number of elastic constants to as small a number as possible.

Crystals, Rocks and Elasticity:

Cubic crystals: For cubic crystals the elastic constants, c_{ij}, may be reduced to just three independent elastic constants:

$c_{11} = c_{22} = c_{33} \neq$ modulus for axial compression, i.e., a stress s_{11} results in a strain e_{11} along an axis;

$c_{44} = c_{55} = c_{66} \neq$ shear modulus, i.e., a shear stress s_{23} results in a shear strain e_{23} across a face;

$c_{12} = c_{13} = c_{23} \neq$ modulus for dilation on compression, i.e., an axial stress s_{11} results in a strain e_{22} along a perpendicular axis.

All other $c_{ij} = 0$.

For single crystals, the elastic constants can be related to common elastic moduli such as:

Shear modulus:

$$\mu = c_{44} \quad \text{for shear on (100)}$$

$$\mu = (c_{11} - c_{12}) / 2 \quad \text{for shear on (110)}$$

Bulk modulus:

$$K = (c_{11} + 2c_{12}) / 3$$

Polycrystalline aggregates:

In the simplest case, we can consider a polycrystalline aggregate of crystals in random orientations, which is therefore isotropic. For such an isotropic system, the elastic constants may be reduced to just *two*, called the Lam Constants, l and m, and the stress-strain relation then becomes:

$$\sigma_{ij} = \lambda\delta_{ij} \sum_{k} \varepsilon_{kk} + 2\mu\mu_{ij}$$

where d is equal to 1 for i=j, and to zero for i≠j; the strain tensor $e_{kk} = \Delta V / V$.

The Lam Constants are defined by:

$$\mu = c_{44}$$

i.e., the shear modulus, and:

$$\lambda + 2\mu = c_{11}$$

and:

$$\lambda = c_{12} = c_{11} - 2c_{44}$$

So, for uniaxially applied stress:

$$\sigma_{11} = \lambda(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}) + 2\mu\varepsilon_{11}$$

The estimation of the bulk properties from the elastic constants is fairly straightforward; however, when dealing with real materials, e.g., rocks, which are made up of polycrystalline aggregates, the elastic properties have to be evaluated by averaging the elastic constants over all the crytalline structures within the aggregate. For polycrystalline materials made up of non-cubic crystals with lower symmetry, appropriate substitutions have to be made in the elastic constants to account for the asymmetry, e.g., $<c_{11}>= (c_{11}+c_{22}+c_{33})$, etc.. Therefore the bulk modulus becomes:

$$K = \frac{<c_{11}> + 2<c_{12}>}{3}$$

Similarly, two expressions may be obtained for the effective shear modulus, one under the assumption of constant stress, the other under the assumption of constant strain.

Under uniform stress:

$$\mu_R = \frac{15}{\left(\frac{12}{<c_{11}> - <c_{12}>} + \frac{9}{<c_{44}>}\right)}$$

Under uniform strain :

$$\mu_V = \frac{<c_{11}> - <c_{12}> + 3<c_{44}>}{5}$$

Normally, an average m is taken called the Voigt-Reuss-Hill (VRH) average.

Elasticity and Seismic Velocity: From an analysis of the passage of waves through a solid medium, the speed of seismic waves are given by:

$$V_P = \left(\frac{K + \frac{4}{3}\mu}{\rho}\right)^{\frac{1}{2}}$$

$$V_S = \left(\frac{\mu}{\rho}\right)^{\frac{1}{2}}$$

Therefore, knowledge of V_P and V_S is all that is required to obtain quantitative values for many elastic properties, some of which are outlined below.

a) Poisson's Ratio: For uniaxial dilation ($s_{11} \neq 0$; $s_{22} = s_{33} = 0$), Poisson's ratio is defined:

$$\nu = -\frac{\varepsilon_{22}}{\varepsilon_{11}}$$

i.e., the ratio of thinning to elongation along perpendicular axes.

Analysis of the elastic constants gives Poisson's ratio in terms of more readily available parameters:

$$\nu = \frac{3\left(\frac{K}{\mu}\right) - 2}{2\left(\frac{3K}{\mu} + 1\right)}$$

From this we can see that for an incompressible solid (K=0) or liquid (m = 0), n = 0.5; for an infinitely compressible solid (K = 0), n = -1; thus we always have -1 < n < 0.5, and generally n ~ 0.25.

From the ratio seismic velocities given above, we can get:

$$\left(\frac{V_P}{V_S}\right)^2 = \frac{K}{\mu} + \frac{4}{3}$$

therefore

$$\frac{K}{\mu} = \left(\frac{V_P}{V_S}\right)^2 - \frac{4}{3}$$

So Poisson's ratio may be given in terms of seismic velocities thus:

$$\nu = \frac{\left(\frac{V_P}{V_S}\right)^2 - 2}{2\left(\left(\frac{V_P}{V_S}\right)^2 - 1\right)}$$

Therefore, Poisson's ratio for the Earth as a function of depth is obtainable directly from PREM and other seismic models.

b) The seismic parameter: In addition to n, seismologists often use the seismic parameter, f:

$$\phi = \frac{K}{\rho} = V_P^2 - \frac{4}{3} V_S^2$$

c) The bulk velocity, V_f: The propagation velocity of the hydrostatic part of the strain (dilation), often called bulk velocity is given by:

$$V_\phi = \left(\frac{K}{\rho}\right)^{\frac{1}{2}}$$

d) The Adams-Williamson equation: If we recall that K = -V*d*P / *d*V = r*d*P / *d* r (since V / dV = - r / *d*r), then:

$$\phi = \frac{dP}{d\rho}$$

i.e. the seismic parameter gives a direct measure of density variation with depth.

However, as one descends into the Earth, the pressure increases via:

$$\Delta P = \rho g \Delta r$$

so in the limit of $\Delta P, \Delta r \neq 0$:

$$\frac{dP}{dr} = \rho g$$

When combined with f = *d*P / *d*r this gives:

$$\frac{d\rho}{dr} = \frac{\rho g}{\phi}$$

so the variation of density with depth can be inferred from the seismic parameter, and therefore from seismic velocities. This is the Adams-Williamson Equation.

Thermoelastic Coupling*:* thermodynamics-elasticity:

Having discussed the essential thermodynamic and elastic properties of solid systems, we can now put them together. The combination of the effect of temperature and the effect of pressures (stress) through thermal expansion, a, and incompressibility, K, is called *thermoelastic coupling*, and is the most important cross-term in geophysics.

a) The Gr neisen Parameter: The inter-relation between stress and temperature is dealt with via the Gr neisen parameter, an approximately constant, pressure and temperature independent parameter of the order of 1. Thermodynamically, the thermal Gr neisen parameter is defined by:

$$\gamma_{th} = \frac{\alpha V K_T}{C_V} = \frac{\alpha V K_S}{C_P}$$

and after some unpleasant thermodynamics (see problem sheet!), this leads to:

$$\frac{C_P}{C_V} = \frac{K_S}{K_T} = 1 + \alpha\gamma T$$

Therefore knowledge of g gives a good handle on many of the important thermo-elastic properties of minerals, which shall be used later in the course.

b) The Mie-Gr neisen Equation of State: When a solid is heated at constant volume, the atoms within it vibrate more vigourously, and this results in a thermal pressure acting from within the system outwards, which, if unresisted, will give rise to thermal expansion. From thermodynamics:

$$\left(\frac{\partial P}{\partial T}\right)_V = \alpha K_T = \gamma_{th}\frac{C_V}{V}$$

Integrating at constant volume and assuming g is a constant:

$$\Delta P = \frac{\gamma}{V}\int_{T_2}^{T_1} C_V dT = \frac{\gamma}{V}(\Delta U)$$

Where U is the internal energy. Therefore:

$$\Delta P = \gamma\frac{\Delta U}{V}$$

And this is the Mie-Gr neisen equation of state

c) The adiabatic temperature gradient: The adiabatic temperature gradient (no heat escapes, S is constant) is that caused by adiabatic compression; i.e., the change in temperature throughout the Earth as a result of pressure loading from above.

From thermodynamics we can show:

$$\left(\frac{\partial T}{\partial V}\right)_S = \frac{\alpha K_S T}{C_P} = -\frac{\gamma T}{V}$$

or

$$\left(\frac{\partial \ln T}{\partial \ln \rho}\right) = \gamma$$

which on integration (for constant g) gives:

$$\left(\frac{T_1}{T_2}\right)_S = \left(\frac{\rho_2}{\rho_1}\right)^{\gamma}$$

Therefore if we know g we can estimate how the temperature varies with density in a planetary interior, assuming it is adiabatically generated.

Another approach may be made by considering another relationship from thermodynamics:

$$\left(\frac{\partial T}{\partial P}\right)_S = \frac{\gamma T}{K_S}$$

but, the adiabatic temperature gradient is found to be:

$$\left(\frac{\partial T}{\partial r}\right)_S = \frac{\gamma \rho g T}{K_S} = \frac{\gamma g T}{\varphi}$$

Therefore, the adiabatic temeprature gradient may be obtained from a knowledge of the seismic parameter.

SAQ.3

a) What do you mean by X-ray diffraction?

b) Define the Bragg's law.

c) What is the Generalized Hooke's law for Anisotropic body?

d) A beam of X-rays of wavelength 0.085 nm is diffracted by (101) plane of rock salt with lattice constant of 0.24 nm. Find the glancing angle for the second-order diffraction.

e) Using Hooke's Law, how much force is needed to pull a spring with a spring constant of 20 N/m a distance of 25 cm?

Examples:

Q.1 Copper crystal has a face centred cubic structure. Atomic radius of copper atom is 128 pm. What is the density of copper metal? Atomic mass of copper is 63.5.

Solution:

In face centred cubic arrangement face diagonal is four times the radius of atoms face diagonal = 4×128 = 512 pm

Face diagonal = √2 × edge length

Edge length = 512 / √2 = 362×10^{-10} cm

Volume of the unit cell = $(362 \times 10^{-10})^3$ cm^3 = 47.4×10^{-24} cm^3

In a face centred cubic unit cell, there are four atoms per unit cell

Mass of unit cell = $4 \times 63.5 / 6.023 \times 10^{23}$ g = 4.22×10^{-22} g

Density = mass of unit cell / volume of unit cell = $4.22 \times 10^{-22} / 47.4 \times 10^{-24}$ = 8.9 g cm^{-3}

Q.2 The density of CaO is 3.35 gm/cm^3. The oxide crystallises in one of the cubic systems with an edge length of 4.80 Å. How many Ca^{++} ions and O^{-2} ions belong to each unit cell, and which type of cubic system is present?

Solution**:**

From equation

r(density) = 3.35 gm/cm^3

a = 4.80 Å

M_m of CaO = (40 + 16) gm = 56 gm CaO

Q r = where n = no. of molecules per unit cell

$p = n \times M_m / a^3 \times N_A$

∴ $n = 3.35 \times (4.8 \times 10^{-8})^3 \times 6.023 \times 10^{23} / 56 = 3.98$

or n ≈ 4

So, 4-molecules of CaO are present in 1 unit cell

So, no. of Ca^{++} ion = 4

No. of O^{--} ion = 4

So, cubic system is fcc type.

Q.3 Copper has the fcc crystal structure. Assuming an atomic radius of 130pm for copper atom (Cu = 63.54):

(a) What is the length of unit cell of Cu?

(b) What is the volume of the unit cell?

(c) How many atoms belong to the unit cell?

(d) Find the density of Cu.

Solution:

As we know

$p = n \times M_m / N_A \times a^3$

(a) for fcc structure

$4r = \sqrt{2}\, a$

$a = 2\sqrt{2}r = 2\sqrt{2} \times 130\ pm = 367.64\ pm$

(b) volume of unit cell = $a^3 = (367.64 \times 10^{-10}\ cm)^3 = 4.968 \times 10^{-23}\ cm^3$

(c) n = 4

(d) $p = 4 \times 63.54 / 6.023 \times 10^{23} \times (3.67 \times 10^{-8}\ cm^3)^3 = 8.54\ gm / cm^3$

Q.4 KBr or potassium bromide has density 2.75 g cm^{-3}. The edge length of its unit cell is 654 pm. Prove that KBr depicts face-centred cubic structure.

Solution: We have edge length of unit cell = 654 pm = 6.54×10^{-8} cm

Therefore volume of the cell = $(6.54 \times 10^{-8})^3\ cm^3$

Molar Mass of Potassium Bromide (KBr) is 119 g/mol

Density of KBr = 2.75 g cm^{-3}

We know that

$$\text{Density of Unit Cell} = \frac{\text{Mass of unit cell}}{\text{Volume of unit cell}}$$

$$2.75 = \frac{n \times M}{a^3 \times N_A}$$

$$2.75 = \frac{n \times 119}{(6.54 \times 10^{-8})^3 \times 6.022 \times 10^{23}}$$

Evaluating we get value of n to be 3.09 ≈ 4

Since a number of atoms is 4 we can clearly say that KBr is a face-centred cubic structure.

Q.5 In a crystal, a plane cuts intercepts of 2a 3b and 6c along the three crystallographic axes, Determine the miller indices of the plane.

Solution:

Intercept	2a	3b	6c
Division by lattice constants	2a/a=2	3b/b=3	6c/c=6
Reciprocal	½	1/3	1/6
After clearing fractions (multiply by 6)	3	2	1
Miller indices	(321)		

Q.6 The distance between the consecutive (111) plane in a cubic crystal is 2A determine the lattice parameter.

Solution: For the cubic crystals, we have

$$d = \frac{a}{\sqrt{h^2+k^2+l^2}} = \frac{2}{\sqrt{1+1+1}} = \frac{2}{\sqrt{3}}\text{A}$$

Q.7 X- rays of wavelength equal to 0.134 nm give a first order diffraction from the surface of a crystal when the value of θ is 10.5°. Calculate the distance between the planes in the crystal parallel to the surface examined.

Solution

Given λ = 0.134 nm, θ = 10.5°

n = 1

Applying Bragg's equation

2d Sinθ = nλ

d = nλ / 2Sinθ = 1 × 0.134 / 2 × Sin 10.5^{o} = 0.134 / 2 × 0.1822 = 3.68 A

Q.8 The first order reflections of a beam of X – rays of wavelength of 1.54A° from the (100) face of a crystal of the simple cubic type occurs at an angle 11.29°. Calculate the length of the unit cell.

Solution

Applying Bragg's equation

2 d Sinθ = nλ

Given θ = 11.29°, λ = 1.54A° = 1.54×10^{-8} cm

n = 1

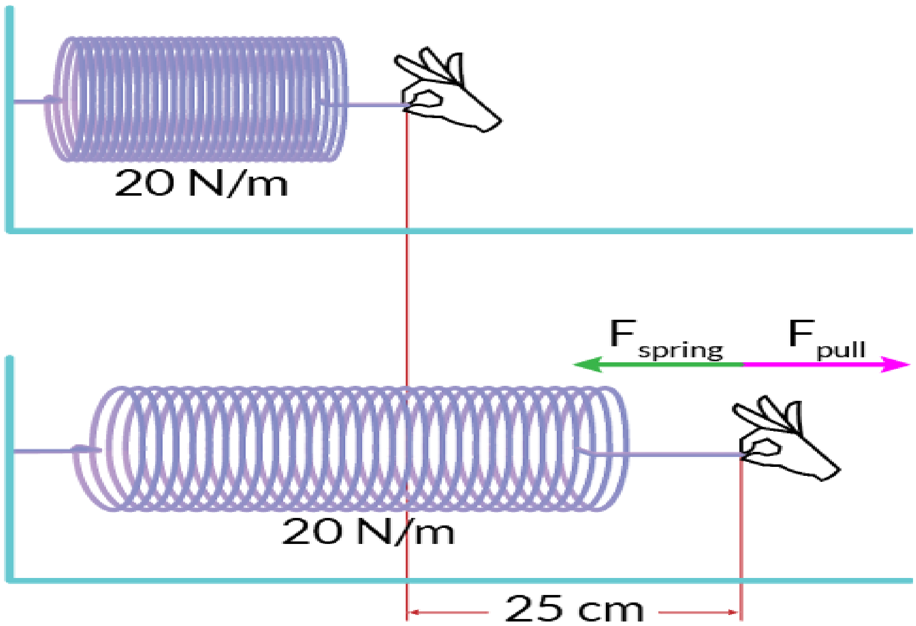
d = 1.54×10^{-8} / 2 × Sin 11.29^{o} = 1.54×10^{-8} / 2 × 0.1957 = 3.93×10^{-8} cm

dhkl = a $\sqrt{h^2 + k^2 + l^2}$ = a

a = 3.93×10^{-8} cm (length of the unit cell)

Q.9 Using Hooke's law, much force is needed to pull a spring with a spring constant of 20 N/m a distance of 25 cm?

Solution:



The k of the spring is 20 N/m.

Δx is 25 cm

We need this unit to match the unit in the spring constant, so convert the distance to meters.

Δx = 25 cm = 0.25 m.

Plug these values into the Hooke's Law formula. Since we're looking for the force required to pull the spring apart, we don't need the minus sign.

F = k·Δx

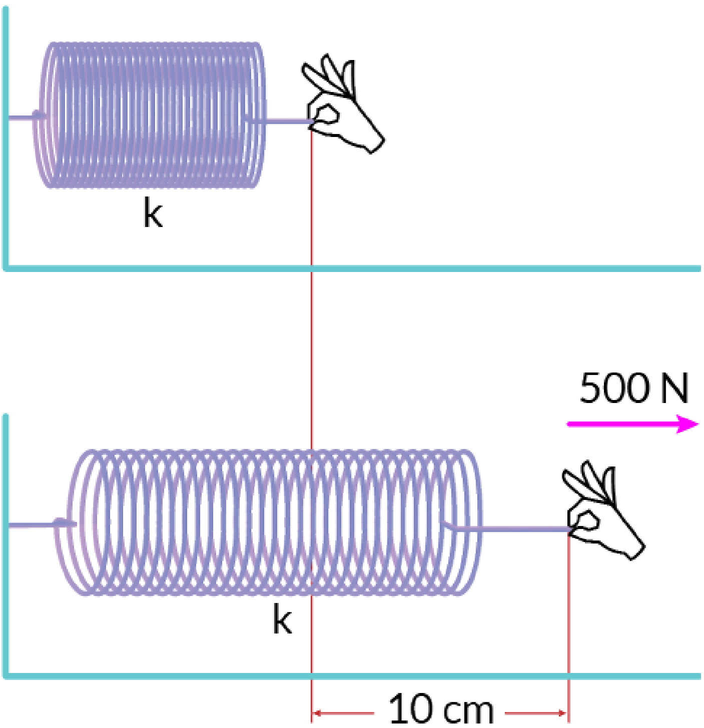
$F = 20\ N/m \cdot 0.25\ m$

$F = 5N$

A force of 5 Newtons is needed to pull this spring a distance of 25 cm.

Q.10 A spring is pulled to 10 cm and held in place with a force of 500 N. What is the spring constant of the spring?

Solution:



The change in position is 10 cm. Since the units on the spring constant are Newtons per meter, we need to change the distance to meters.

$\Delta x = 10\ cm = 0.10\ m$

$F = k \cdot \Delta x$

Solve this for k by dividing both sides by Δx

$F/\Delta x = k$

Since the force is 500 N, we get

$500\ N / 0.10\ m = k$

k = 5000 N/m

The spring constant of this spring is 5000 N/m.

Q.11 A beam of X-rays of wavelength 0.071 nm is diffracted by (110) plane of rock salt with lattice constant of 0.28 *nm*. Find the glancing angle for the second-order diffraction.

Solution:

Given data are:

Wavelength (λ) of X-rays = 0.071 nm

Lattice constant (a) = 0.28 nm

Plane (hkl) = (110)

Order of diffraction = 2

Glancing angle θ = ?

Bragg's law is $2d \sin \theta = n\lambda$

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}, \text{ because rock salt is FCC}$$

$$= \frac{0.28 \times 10^{-9}}{\sqrt{1^2 + 1^2 + 0^2}} \text{ m} = \frac{0.28 \times 10^{-9}}{\sqrt{2}} \text{ m}$$

Substitute in Bragg's equation

$$2 \times \frac{0.28 \times 10^{-9}}{\sqrt{2}} \sin \theta = 2 \times 0.071 \times 10^{-9}$$

$$\sin \theta = \sqrt{2} \times \frac{0.071}{0.28} = 0.3586$$

$$\theta = \sin^{-1}(0.3586) = 21.01^{\circ} \approx 21^{\circ}$$

1.10 Summary:

1. The solids in which the constituent particles of matter are arranged and organized in a specific manner are called Crystalline Solids**.**
2. The solids in which the constituent particles of matter are arranged in a random manner are called amorphous solids.
3. Liquid crystal materials generally have several common characteristics. Among these is a rod-like molecular structure, rigidness of the long axis, and strong dipoles and/or easily polarizable substituent.
4. In a simple cubic structure, the spheres are not packed as closely as they could be, and they only "fill" about 52% of the volume of the container.
5. Such as aluminum, copper, and lead, crystallize in an arrangement that has a cubic unit cell with atoms at all of the corners and at the centers of each face. This arrangement is called a face-centered cubic (FCC) solid.
6. Some metals crystallize in an arrangement that has a cubic unit cell with atoms at all of the corners and an atom in the center. This is called a body-centered cubic (BCC) solid.
7. The smallest possible portion or part of the crystal lattice which repeats itself in different directions of the lattice is called the unit cell.
8. The Bravais lattice are the distinct lattice types which when repeated can fill the whole space.
9. A crystal lattice can be defined as a geometrical arrangement of constituent particles of matter (atoms, ions or molecules) as points in space.

10. Bravais Lattice refers to the 14 different 3-dimensional configurations into which atoms can be arranged in crystals.
11. The direct lattice represents the triple periodicity of the ideal infinite perfect periodic structure that can be associated to the structure of a finite real crystal.
12. The reciprocal lattice is constituted of the set of all possible linear combinations of the basis vectors a*, b*, c* of the reciprocal space.
13. The Miller indices are 3,6,2. If a plane is parallel to an axis, its intercept is at infinity and its Miller index is zero. A generic miller index is denoted by (hkl).
14. X-ray diffraction is a phenomenon in which the atoms of a crystal, by virtue of their uniform spacing, cause an interference pattern of the waves present in an incident beam of X rays.
15. The principle of Bragg's law is applied in the construction of instruments such as Bragg spectrometer, which is often used to study the structure of crystals and molecules.
16. The generalized Hooke's Law also reveals that strain can exist without stress. For example, if the member is experiencing a load in the y-direction (which in turn causes a stress in the y-direction), the Hooke's Law shows that strain in the x-direction does not equal to zero.
17. The way planetary interiors deform depends primarily upon the elastic properties of the constituent polycrystalline materials.

1.11 Terminal Questions:

1) Explain the Crystalline and amorphous state of solids. Also give the characteristics.
2) Explain the liquid crystal and its characteristics.
3) What do you mean by Simple crystal structure of SC, FCC and BCC?
4) Explain the Unit cell and Bravais lattice.
5) What are types of Classification of lattices?
6) Explain types of crystals on the basis of Bravais lattice.
7) What do you mean by Direct and reciprocal lattice?
8) Explain the Miller indices and its planes.
9) Define the X-ray diffraction.
10) Explain the working of Bragg's law and also give its advantages.
11) Explain the Generalized Hooke's law for anisotropic body
12) What do mean by elastic constants of cubic crystals?

13) Copper crystal has a face centred cubic structure. Atomic radius of copper atom is 118 pm. What is the density of copper metal? Atomic mass of copper is 63.5.

14) A metal crystallizes into two cubic system face centred cubic (fcc) and body centred cubic (bcc) whose unit cell lengths are 4.5 and 4.0Å respectively. Calculate the ratio of densities of fcc and bcc.

15) An orthorhombic Crystal whose primitive translations are a= 1.21A, b=1.84A and C=1.97A respectively. If plane of miller indices $(\overline{23}\ 1)$ cuts an intercept of 1.21A along 'x' axis find the length of intercept at 'y' and 'z' axes.

16) The density of CaO is 2.35 gm/cm^3. The oxide crystallises in one of the cubic systems with an edge length of 3.80 Å. How many Ca^{++} ions and O^{-2} ions belong to each unit cell, and which type of cubic system is present?

17) Copper has the fcc crystal structure. Assuming an atomic radius of 125pm for copper atom (Cu = 63.54):

(a) What is the length of unit cell of Cu?

(b) What is the volume of the unit cell?

(c) How many atoms belong to the unit cell?

(d) Find the density of Cu.

18) X- rays of wavelength equal to 0.142 nm give a first order diffraction from the surface of a crystal when the value of θ is 11.5°. Calculate the distance between the planes in the crystal parallel to the surface examined.

.

Unit 2- Band Theory of Solids

Structure

2.1 Introduction

2.2 Objectives

2.3 Need of free electron quantum theory

2.4 Summerfeld Fermi model band theory

2.5 One dimensional motion of electron in periodic potential (Bloch theorem)

2.6 Kronning-Penny model (features and its importance).

2.7 Fermi surface, effective mass of charge carriers (electron and holes).

2.8 Concentration in semiconductors.

2.9 Hall effect (qualitative).

2.10 Summary

2.11 Terminal Question

1.1 Introduction:

Quantum free electron theory introduced by Summerfeld in 1928. This theory is based on making small concepts. This theory was proposed by making small changes in the classical free electron theory and by retaining most of the postulates of the classical free electron theory.

The electron theory of solids explains the structures and properties of solids through their electronic structure. This theory is applicable to all solids both metals and non metals.

Arnold Summerfeld succeeded in overcoming many of drawbacks of the classical free electron theory, while retaining all the essential features of the classical free electron theory. His approach is based upon quantization of electrical energy levels. He realised the role played by Pauli exclusion principle in restricting the energy values of electron.

One may appeal to Bloch's theorem in order to make headway in obviating this latter problem. Instead of being required to consider an infinite number of electrons, it is only necessary to consider the number of electrons within the unit cell (or half of this number if the electrons are spin degenerate).

The Kronning-Penney model is a simplified model for an electron in a one-dimensional periodic potential. The possible states that the electron can occupy are determined by the Schrödinger equation

The Fermi surface separates the unfilled orbitals from the filled orbitals, at absolute zero. The electrical properties of the metal are determined by the shape of the Fermi surface, because the current is due to changes in the occupancy of states near Fermi surface. The free electron Fermi surfaces were developed from spheres of radius k_F determined by the valence electron concentration.

Intrinsic carriers are the electrons and holes that participate in conduction. The concentration of these carriers is contingent upon the temperature and band gap of the material, thus affecting a material's conductivity. Knowledge of intrinsic carrier concentration is linked to our understanding of solar cell efficiency, and how to maximize it.

Mainly Lorentz force is responsible for **Hall Effect.** All of we know that when we place a current carrying conductor inside a magnetic field, the conductor experiences a mechanical force to a direction depending upon the direction of magnetic field and the direction of current in the conductor. The electric current means a flow of charge. In metal it is entirely due to the flow of electrons, in semiconductor, it is due to flow of free electrons as well as holes.

1.2 Objectives:

After studying this unit you should be able to

a) Explain and identify Need of free electron quantum theory.
b) Study and identify Summerfeld Fermi model band theory.
c) Explain and identify One dimensional motion of electron in periodic potential (Bloch theorem).
d) Study and identify Kronning-Penny model (features and its importance).
e) Explain Fermi surface, effective mass of charge carriers (electron and holes).
f) Study and identify Concentration in semiconductors.
- Explain and identify Hall effect (qualitative).

2.3 Need of free electron quantum theory:

QUANTUM FREE ELECTRON THEORY:

The failure of classical free electron theory paved this way for Quantum free electron theory. It as introduced by Summerfeld in 1928. This theory is based on making small concepts. This theory was proposed by making small changes in the classical free electron theory and by retaining most of the postulates of the classical free electron theory.

Assumptions (Postulates) of Quantum free electron theory:

1. In a metal the available free electrons are fully responsible for electrical conduction.
2. The electrons move in a constant potential inside the metal. They cannot come out from the metal surface have very high potential barrier.
3. Electrons have wave nature, the velocity and energy distribution for the electrons given by Fermi-Dirac distribution function.
4. The loss of energy due to interaction of the free electron with the other free electron
5. Electron's distributed various energy levels according to Pauli Exclusion Principle

Advantages of Quantum free electron theory:

1. This theory fails to distinguish between metal, semiconductor and Insulator.
2. It also fails to explain the positive value of Hall Co-efficient.
3. According to this theory, only two electrons are present in the Fermi level and they are responsible for conduction which is not true.

Electron Theory of Metals: The electron theory of solids explains the structures and properties of solids through their electronic structure. This theory is applicable to all solids both metals and non metals. This theory also explains the bending in solids behavior of conductors and insulators, electrical and thermal conductivities of solids, elasticity and repulsive forces in solids etc. The theory has been developed in three main stages.

Classical free electron theory: This theory was developed by Drude and Lorentz. According to this theory, a metal consists of electrons which are from to move about in the crystal molecules of a gas it contains mutual repulsion between electrons is ignored and hence potential energy is taken as zero Therefore the total energy of the electron is equal to its kinetic energy.

1. A Solid metal has nucleus with revolving electrons. The electrons have freely like molecules in a gas.
2. In the absence of electric field (E=0), the free electrons move in random directions and collide with each other. During this collision no loss of energy is observes since the collisions are elastic.
3. When the presence of electric field the free electrons are accelerated on the direction opposite to the applied electric field.
4. Since the electrons are assumed to be perfect gas, they obey the laws of classical theory of gases.

5. Classical free electrons in the metal obey Maxwell-Boltzmann statistics.

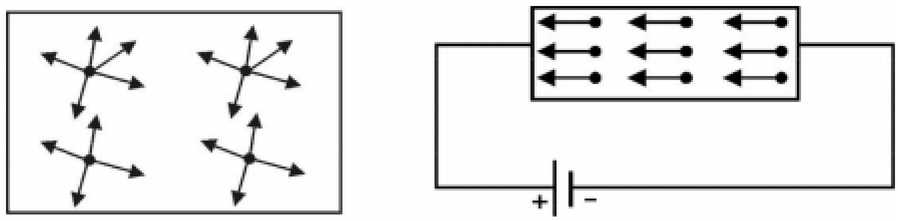


Fig.2.1 Free electron theory

Comparison Classical Free Electron Theory and Quantum Free Electron Theory

Similarities between the two theories

The following assumptions apply to both the theories:

1. The valence electrons are treated as though they constitute an ideal gas.
2. The valence electrons can move freely throughout the body of the solid.
3. The mutual repulsion between the electrons, and the force of attraction between the electrons and ions are considered insignificant.

Difference between the two theories:

Classical Free Electron Theory	Quantum Free Electron Theory
1. The free electrons, which constitute the electron gas can have continuous energy values.	1. The energy values of the free electrons are discontinuous because of which the energy levels are discrete.
2. It is possible that many electron may possess same energy.	2. The free-electrons obey the Pauli exclusion principle. Hence no two electrons can possess same energy.
3. The pattern of distribution of energy among the free electron obey Maxwell-Boltzmann statistics.	3. The distribution of energy among the free electrons is according to Fermi Dirac statistics which imposes a severe restriction on the possible way in which the electrons absorb energy from an external source.

2.4 Summerfeld Fermi model band theory:

In 1928, by applying quantum mechanical principles, **Arnold Summerfeld** succeeded in overcoming many of drawbacks of the classical free electron theory, while retaining all the essential features of the classical free electron theory. His approach is based upon quantization of electrical energy levels. He realised the role played by Pauli exclusion principle in restricting the energy values of electron. The theory proposed known as Quantum Free Electron Theory

The main assumptions of quantum free electron theory are,

1. The energy levels of the conduction electrons are quantised.
2. The distribution of electrons in various allowed energy levels occur as per Pauli Exclusion Principle.
3. However, the following assumptions of classical electron theory continue to be applicable in quantum free electron theory also.
4. The electrons travel in a constant potential inside the metal but stay confined within its boundaries.
5. Both the attraction between the electrons and the lattice ions, and the repulsion between the electrons themselves are ignored.

Fermi Energy:

Due to the quantization rules, a material in solid state possess a set of allowed energy levels. For a metal containing N free electrons, there will be N such allowed energy levels, which are separated by energy differences that are characteristic of the material. The allowed energy values for the conduction electrons in the metal are also quantised, are related to the energy levels of the metal. As per the Pauli exclusion principle each energy level can accommodate a maximum of two electrons, one with spin up and the other one with spin down (Figure).

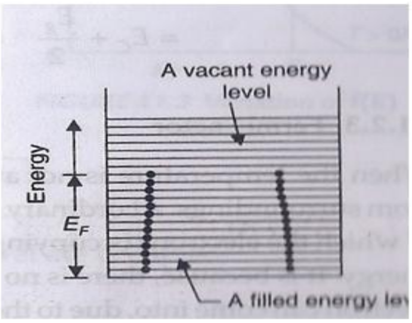


Fig.2.2 Fermi Energy

The allowed energy values of the electron are realised in terms of occupation of the allowed energy levels, under certain rules. When the filling up the energy levels is undertaken the universal rules that-any system which is free tends to go state of lowest energy-comes into picture.

Thus a pair of electrons, one with spin up, and other with spin down occupy the lowers level. The next pair of electrons occupy the next higher level, and so on, till all the electrons in metal are accommodated.

However, there will be many more allowed energy levels available for occupation by the electrons. The energy of the highest occupied level at zero degree absolute is called Fermi energy, and the energy level is referred to as the Fermi level. The Fermi energy is denoted as E_F.

When there is no external energy supply for the electrons, such as thermal energy or electrical energy, the electrons are free, and thus settle in the lowest allowed energy state available. Thus at a temperature of absolute zero, and when the metal is not in the influence of any external field, all the energy levels lying above the Fermi level are empty, and those lying below are completely filled. Since there are two electrons in each energy level, out of the N allowed energy levels N/2 of them will be occupied.

Fermi level:

The Fermi level is an energy level mid way between top surface of valence band and bottom surface of conduction-band as shown in Figure.

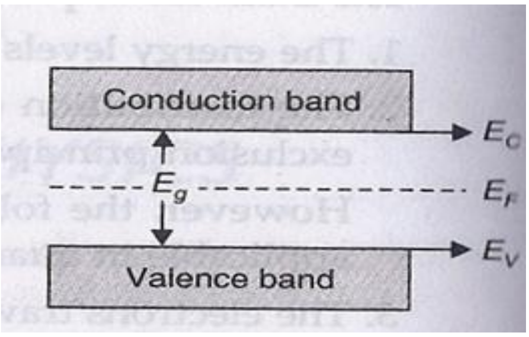


Fig.2.3 Fermi level

If E_V is the energy corresponding to top surface valence band, E_C is the energy corresponding to bottom surface of conduction band and E_F is the Fermi energy corresponding to Fermi level, then

$$E_F = \frac{E_C + E_V}{2} = E_C + \frac{E_V - E_C}{2}$$

$$= E_C + \frac{E_g}{2} \quad \Rightarrow \quad \therefore \; E_F - E_C = \frac{E_g}{2} \quad [where\ E_V - E_C = E_g\ (Energy\ gap)]$$

Fermi Factor:

When the temperature is not at absolute zero, the material will be receiving thermal energy from surroundings. At ordinary temperature, the amount of energy will be quite small", because of which the electrons occupying energy levels below the Fermi level cannot absorb the thermal energy. It is because, there is no availability of unoccupied higher energy levels into which the electron can come into, due to the increase in energy by the *absorption of thermal energy*. However, there are unoccupied higher energy levels which are above the occupied energy levels at small energy differences. They are located near Fermi level. Those are the energy levels into which, the electrons in the energy levels near Fermi level, are capable of being excited.

During thermal excitation *(i.e., T* >0), the electrons which absorb the thermal energy move into higher energy levels which were unoccupied at zero degree absolute *(i.e., T* = 0). Though such excitations seem to be random, the occupation of various energy level obey a statistical distribution called Fermi-Dirac distribution, once the system is in thermal equilibrium *(i.e.,* at a steady temperature state).

The probability *f (E)* that a given energy state with energy *E* is occupied at a steady temperature *T,* is given by

The quantity of thermal energy is given by k_bT, where k_B is Boltzmann constant. A room temperature k_BT=0.0025eV. Compare this with Fermi energy which is of the order of 5 eV.

Since such electrons constitute only a small fraction of the total number of electrons in the metal as a whole, the specific heat of the material becomes low. Thus, reason for failure of classical free electron theory to account for low specific heat of metals is now understood.

The statistical distribution deals with the question as to what is the probability occupation of a particular energy level.

$$f(E) = \frac{1}{e^{(E-E_F)/k_BT} - 1} = \frac{1}{\exp\left(\frac{\mathrm{E}-\mathrm{E_F}}{\mathrm{k_B T}}\right) - 1}$$

f (E) is called the Fermi factor and is defined as follows:

Fermi factor is the distribution function which gives the probability of occupation of a given energy state for material in thermal equilibrium in terms of the Fermi energy, Boltzmann constant and the temperature.

The dependence of Fermi factor on temperature and the effect on occupancy of energy level is shown in Figure.

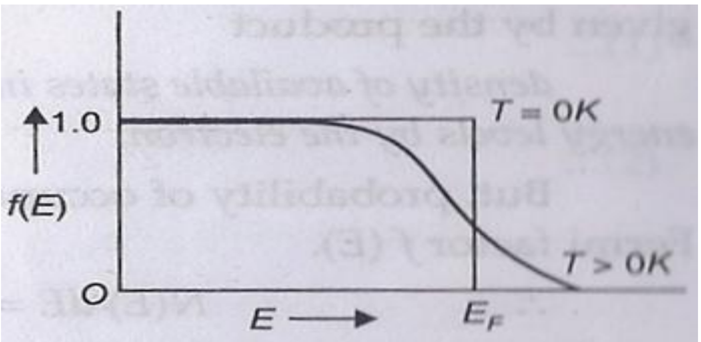


Fig.2.4 Variation of f(E)with E

Let us consider the different cases of distribution

(i) **Probability of occupation for E < E_F at T = 0**

When T = 0 and E < E_F

$$f(E) = \frac{1}{e^{-\infty}+1} = \frac{1}{0+1} \qquad \therefore f(E) = 1 \; for \; E < E_F$$

At T = 0 K, all the energy levels below the Fermi level are occupied.

(ii) **Probability of occupation for E > E_F at T= 0 K**

When T = 0, and E > E_F

$$f(E) = \frac{1}{e^{\infty}+1} = \frac{1}{\infty} \qquad \therefore f(E) = 0 \; for \; E > E_F$$

At T=0K, all the energy levels above Fermi levels are unoccupied.

In view of the above two cases, at T=0K, the variation of *f* (E) for different energy values, becomes a step function, as shown in Figure.

(iii) Probability of occupation at ordinary temperatures

At ordinary temperatures, the value of probability starts reducing from 1 for values of E close to but lesser than E_F

$$\text{The value of f(E) becomes } \frac{1}{2} \text{ at } E = E_F$$

$$\text{Since for} \quad E = E_F, e^{(E-E_F)/k_BT} = e^0 = 1$$

$$f(E) = \frac{1}{e^{\frac{E-E_F}{k_BT}}+1} = \frac{1}{1+1} = \frac{1}{2}$$

Further, for $E>E_F$, the probability value falls off to zero rapidly (Figure). The above discussion on the variation f(E) with E, brings forth the point that.

The Fermi energy is the most probable, or the average energy of the electron across which the energy transitions occurs at temperature above zero degree absolute, which result in the conductivity of the metal. This may be considered as the physical basis for the concept of Fermi energy.

Calculation of Fermi Energy at T = 0 K

Let the value of E_F of 0 K be denoted as E_{F0} . Also, if we denote the number of electrons per unit volume, which possess energy only in the range E and (E + *dE)*, by *N(E) dE*, then, *N(E) dE* is given by the product density of available states in the energy range E and E + dE × probability of occupation of those energy levels by the electron.

But, probability of occupation of any given energy state by the electron is given by the Fermi factor *f (E).*

$$N(E)\,dE = g(E)\,dE \times f(E)$$

The number of electrons per unit volume of the material *n* can be evaluated by integrating the above expression from E = 0 to E = E_{max}, where E_{max} is the maximum energy possessed by the electrons

$$\therefore \quad n = \int N(E)dE$$

$$= \int_{E=0}^{E=E_{max}} g(E)f(E)dE$$

But at T=0K, the maximum energy that any electron of the material can have is E_{F0},

Hence $\quad E_{max} = E_{F0}$

Also $\quad f(E) = 1$, and T = 0 K

$$\therefore \quad n = \int_{E=0}^{E=E_{max}} g(E)dE \times 1$$

g(E) dE is given by

$$g(E)dE = \frac{8\sqrt{2}\pi m^{3/2}}{h^3} E^{1/2} dE$$

Where m is the mass of electron and h is the Planck's constant.

$$\therefore \qquad n = \frac{8\sqrt{2}\pi m^{3/2}}{h^3}\int_{E=0}^{E=E_{max}} E^{1/2}dE$$

$$= \frac{8\sqrt{2}\pi m^{3/2}}{h^3} \times \frac{2}{3}(E_{F_0})^{3/2}$$

$$= \left[\frac{16\sqrt{2}m^{3/2}}{h^3}\right]\left[\frac{\pi}{3}\right](E_{F_0})^{3/2}$$

$$\therefore \qquad (E_{F_0})^{3/2} = \frac{h^3}{(8m)^{3/2}}\left(\frac{3n}{\pi}\right)$$

$$or \qquad (E_{F_0}) = \left(\frac{h^2}{8m}\right)\left(\frac{3n}{\pi}\right)^{2/3}$$

$$\therefore \qquad E_{F_0} = B\, n^{2/3}$$

$= 5.85 \times 10^{-38}$ J.

Fermi energy at T >0K

The Fermi energy E_F, at any temperature T in general can be expressed in terms of through the relation

$$E_F \doteq E_{F_0}\left[1 - \frac{\pi^2}{12}\left(\frac{k_B T}{E_{F_0}}\right)^2\right]$$

The Fermi energy at extremely high temperatures, the second term within the bracket is compared to unity.

$$\therefore \quad E_F \doteq E_{F_0}$$

Fermi Temperature (T_F):

Fermi temperature (T_F) is te temperature at which the average thermal energy of the free electron in a solid becomes equal to the Fermi energy at 0K. But the thermal energy possessed by electrons is given by the product $k_B T$.

When $T=T_F$, the equation, $E_{F0}= k_B\, T_F$, is satisfied. But for practical purpose, $E_{F0} = E_F$

$$\therefore \quad k_B T_F = E_F$$

$$\text{or} \quad T_F = \frac{E_F}{k_B}$$

The Fermi temperature is only a theoretical concept, since, at ordinary temperature, it is not possible for the electrons to receive thermal energy in a magnitude of E_F. For example, let us consider the case of metals. For metals, we know that E_F will be of the order of a few eV.

Let us consider the case in which EF= 3 eV.

$$= 3 \times 1.602 \times 10^{-19}\ \text{J} = 4.8 \times 10^{-19}\ \text{J}$$

$$\therefore \quad T_F = \frac{4.8 \times 10^{-19}}{1.381 \times 10^{-23}} = 34800\ \text{K}$$

This is quite an exaggerated temperature to be realized in practice.

Fermi Velocity (V_F)

The energy of the *electron*, which are at the Fermi *level* is E_F. The velocity of the electrons which occupy the Fermi level is called the Fermi Velocity v_F.

$$\therefore \quad E_F = \frac{1}{2} m\, v_F^2$$

$$\text{or} \quad v_F = \sqrt{\left(\frac{2\, E_F}{m}\right)}$$

SAQ.1

a) Discuss about the need of free electron quantum theory.
b) What do mean by Fermi energy, Fermi level and Fermi factor?
c) *Find the temperature at which there is 1% probability that a state with energy 0.5 eV above Fermi energy.*

2.5 One dimensional motion of electron in periodic potential (Bloch theorem):

Thus far, the quantum mechanical approaches to solving the many-body problem have been discussed. However, the correlated nature of the electrons within a solid is not the only obstacle to solving the Schrödinger equation for a condensed matter system: for solids, one must also bear in mind the effectively infinite number of electrons within the solid.

One may appeal to Bloch's theorem in order to make headway in obviating this latter problem. Instead of being required to consider an infinite number of electrons, it is only necessary to consider the number of electrons within the unit cell (or half of this number if the electrons are spin degenerate).

Bloch's theorem states that the wave function of an electron within a perfectly periodic potential may be written as

$$\psi_{j,\mathbf{k}}(\mathbf{r}) = u_j(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$$

Where u_i (r) is a function that possesses the periodicity of the potential, *i.e.* $u_i(r+1) = u_i$ (r), where 1 is the length of the unit cell. *i* is the band index, and k is a wave vector confined to the first Brillouin Zone. Since u_i (r), is a periodic function, we may expand it in terms of a Fourier series:

$$u_j(\mathbf{r}) = \sum_{\mathbf{G}} c_{j,\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}$$

where the $\mathbf{G}$ are reciprocal lattice vectors defined through $\mathbf{G}\cdot\mathbf{R} = 2\pi m$, where *m* is an integer, $\mathbf{R}$ is a real space lattice vector and the C_i,G are plane wave expansion coefficients. The electron wave functions may therefore be written as a linear combination of plane waves:

$$\psi_{j,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{j,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}.$$

Given that each electron occupies a state of definite $\mathbf{k}$, the infinite number of electrons within the solid gives rise to an infinite number of k-points. At each k-point, only a finite number of the available energy levels will be occupied. Thus one only needs to consider a finite number of electrons at an infinite number of k-points. This may seem to be replacing one infinity (number of electrons) with another one (number of k-points) to little discernible advantage. However, one does not need to consider all of these k-points; rather, since the electron wave functions will be almost identical for values of $\mathbf{k}$ that are sufficiently close, one can represent the wave functions over a region of reciprocal space by considering the wave function at a single k-point. It is therefore sufficient to consider the electronic states at a finite number of k-points in order to determine the ground state density of the solid. The net effect of Bloch's Theorem therefore has been to change the problem of an infinite number of electrons to one of considering only the number of electrons in the unit cell (or half that number, depending on whether the states are spin-degenerate or not) at a finite number of k-points chosen so as to appropriately sample the Brillouin Zone.

Suppose an electron passes along X-direction in a one-dimensional crystal having periodic potentials:

$$V(x) = V(x + a)$$

where '*a*' is the periodicity of the potential. The Schrödinger wave equation for the moving electron is:

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}\left[E - V(x)\right]\psi = 0$$

The solution for Equation is of the form:

$$\psi(x) = e^{iKx}\, u_k(x)$$

where $u_k(x) = u_k(x + a)$

Equation represents periodic function and e^{ikx} represents plane wave. The above statement is known as Bloch theorem and Equation is called Bloch function. The Bloch function has the property:

$$\psi(x + a) = \exp\,[ik\,(x + a)]\; u_k(x + a) = \psi(x)\exp ika$$

$$\text{or } \psi(x + a) = Q\psi$$

2.6 Kronning-Penny model (features and its importance):

The Kronning-Penney model is a simplified model for an electron in a one-dimensional periodic potential. The possible states that the electron can occupy are determined by the Schrödinger equation,

$$\frac{-\hbar^2}{2m}\frac{d^2\psi}{dx^2}+V(x)\psi=E\psi.$$

In the case of the Kronning-Penney model, the potential $V(x)$ is a periodic square wave.

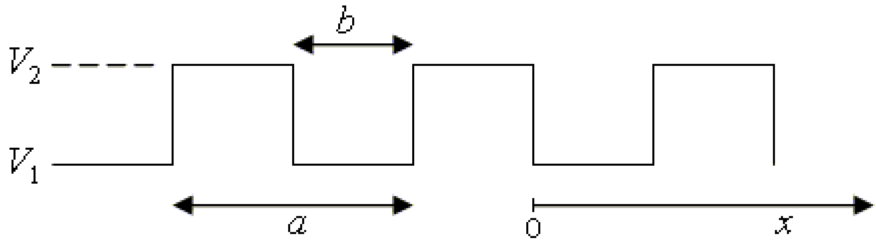


Fig.2.5 Periodic square wave for Kronning-Penney model

A virtue of this model is that it is possible to analytically determine the energy eigen values and eigen functions. It is also possible to find analytic expressions for the dispersion relation (E vs. k) and the electron density of states. Derivations are given below.

Solution of the Schrödinger equation for the Kronning-Penney potential

Since the Kronig-Penney potential exhibits translational symmetry, the energy eigen functions of the Schrödinger equation will simultaneously be eigen functions of the translation operator. As we often do in solid state physics, we proceed by seeking the eigen functions of the translation operator. The translation operator T shifts the solutions by one period, $T\psi(x) = \psi(x + a)$. Notice that any function of the form,

$$\psi_k(x)=e^{ikx}u_k(x),$$

is an eigen function of the translation operator with eigen value e^{ika}.

$$T\psi(x)=e^{ik(x+a)}u_k(x+a)=e^{ika}\psi(x).$$

In solid state physics, functions like this are said to have Bloch form. The convention for describing electron waves in a periodic medium is to express them in terms of the eigen functions of the translation operator. These eigen functions have a well defined frequency and form a complete set that can be used to describe any wave. An eigen function is specified by the k that appears in the expression for the eigen value. It turns out that the physical interpretation of k is nearly the same as the wave number of harmonic waves (usually also called k).

The eigen functions of the translation operator can be readily constructed from any two linearly independent solutions of the one-dimensional Schrödinger equation. A convenient choice is,

$$\psi_1(0)=1, \qquad \frac{d\psi_1}{dx}(0)=0, \qquad \psi_2(0)=0, \qquad \frac{d\psi_2}{dx}(0)=1.$$

The solutions in region 1 ($0 < x < b$) are,

$$\psi_1(x)=\cos(k_1 x), \qquad\qquad \psi_2(x)=\frac{\sin(k_1 x)}{k_1},$$

while the solutions in region 2 ($b < x < a$) are,

$$\psi_1(x)=\cos(k_2(x-b))\cos(k_1 b)-\frac{k_1\sin(k_2(x-b))\sin(k_1 b)}{k_2},$$
$$\psi_2(x)=\frac{\cos(k_2(x-b))\sin(k_1 b)}{k_1}+\frac{\sin(k_2(x-b))\cos(k_1 b)}{k_2}.$$

Here

$$k_1=\sqrt{2m\left(E-V_1\right)/\hbar^2} \quad \text{and} \quad k_2=\sqrt{2m\left(E-V_2\right)/\hbar^2}$$

For energies where k_1 or k_2 are imaginary, the solutions are still real since $\cos(i\theta) = \cosh(\theta)$ and $\sin(i\theta) = i\sinh(\theta)$.

Any other solution can be written as a linear combination of $\psi_1(x)$ and $\psi_2(x)$. In particular, $\psi_1(x + a)$ and $\psi_2(x + a)$ can be written in terms of $\psi_1(x)$ and $\psi_2(x)$. These solutions are related to each other by the matrix representation of the translation operator. [3]

$$\begin{bmatrix}\psi_1\left(x+a\right)\\ \psi_2\left(x+a\right)\end{bmatrix}=\begin{bmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{bmatrix}\begin{bmatrix}\psi_1\left(x\right)\\ \psi_2\left(x\right)\end{bmatrix}.$$

The elements of the translation matrix can be determined by evaluating the equation above and its derivative at $x = 0$.

$$\begin{bmatrix} \psi_1(x+a) \\ \psi_2(x+a) \end{bmatrix} = \begin{bmatrix} \psi_1(a) & \frac{d\psi_1}{dx}(a) \\ \psi_2(a) & \frac{d\psi_2}{dx}(a) \end{bmatrix} \begin{bmatrix} \psi_1(x) \\ \psi_2(x) \end{bmatrix}.$$

The eigen functions and eigen values λ of this 2 × 2 matrix are easily determined to be,

$$\psi_\pm(x) = \frac{2\psi_2(a)}{\frac{d\psi_2(a)}{dx} - \psi_1(a) \pm \delta}\psi_1(x) + \psi_2(x), \qquad \lambda_\pm = \frac{1}{2}(\alpha \pm \delta),$$

where

$$\delta = \sqrt{\alpha^2 - 4}$$

and

$$\alpha = \psi_1(a) + \frac{d\psi_2(a)}{dx} = 2\cos\left(k_2(a-b)\right)\cos(k_1 b) - \left(\frac{k_2}{k_1} + \frac{k_1}{k_2}\right)\sin\left(k_2(a-b)\right)\sin(k_1 b).$$

If periodic boundary conditions are used for a potential with N unit cells, then applying the translation operator N times brings the function back to its original position,

$$T^N\psi(x) = \psi(x+Na) = \lambda^N\psi(x) = \psi(x).$$

The eigen values of the translation operator are therefore the solutions to the equation $\lambda^N = 1$. These solutions are,

$$\lambda_j = \exp(i2\pi j/N) = \exp(i2\pi aj/L) = \exp(ik_j a)$$

where j is an integer between $-N/2$ and $N/2$, $L = Na$ is the length of the crystal, and $k_j = 2\pi j/L$ are the allowed k values in the first Brillouin zone. The dispersion relation can be determined by first calculating α for a specific energy, solving for the eigen values λ and then solving the equation above for the wave number *k*,

$$k = \pm\frac{1}{a}\tan^{-1}\left(\frac{\sqrt{4-\alpha^2}}{\alpha}\right).$$

Whether the eigen values are real or imaginary depends on the magnitude of α. If $\alpha^2 > 4$, the eigen values will be real and the solutions fall in a forbidden energy gap. If $\alpha^2 < 4$, the eigen values will be a complex conjugate pair $\lambda_+ = e^{ika}$ and $\lambda_- = e^{-ika}$.

The dispersion relation can be used to determine the density of states which is needed to calculate the thermodynamic properties of a system of non-interacting electrons. The one-dimensional density of states in *k* is $D(k) = 2/\pi$ and thus the density of states in energy is,

$$\begin{aligned} D(E) &= \frac{2}{\pi}\frac{dk}{dE} = \frac{2}{\pi}\frac{dk}{d\alpha}\frac{d\alpha}{dE} && |\alpha| < 2, \\ D(E) &= 0 && |\alpha| > 2. \end{aligned}$$

Where

$$\frac{dk}{d\alpha} = \frac{1}{a\sqrt{4-\alpha^2}},$$

and

$$\begin{aligned} \frac{d\alpha}{dE} = &-\frac{m}{\hbar^2}\left(\frac{2b}{k_1} + \left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right)\frac{(a-b)}{k_2}\right)\cos\left(k_2(a-b)\right)\sin\left(k_1 b\right) \\ &-\frac{m}{\hbar^2}\left(2\frac{(a-b)}{k_2} + \left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right)\frac{b}{k_1}\right)\sin\left(k_2(a-b)\right)\cos\left(k_1 b\right) \\ &+\frac{m}{\hbar^2}\left(\left(\frac{k_2}{k_1^2}-\frac{1}{k_2}\right)\frac{1}{k_1} + \left(\frac{k_1}{k_2^2}-\frac{1}{k_1}\right)\frac{1}{k_2}\right)\sin\left(k_2(a-b)\right)\sin\left(k_1 b\right). \end{aligned}$$

2.7 Fermi surface:

The Fermi surface is the surface of constant energy in **k** space. The Fermi surface separates the unfilled orbital's from the filled orbital's, at absolute zero. The electrical properties of the metal are determined by the shape of the Fermi surface, because the current is due to changes in the occupancy of states near Fermi surface. The free electron Fermi surfaces were developed from spheres of radius k_F determined by the valence electron concentration.

Construction of free-electron Fermi surfaces: The free electron Fermi surface for the an arbitrary electron concentration is shown in Fig.2.6.

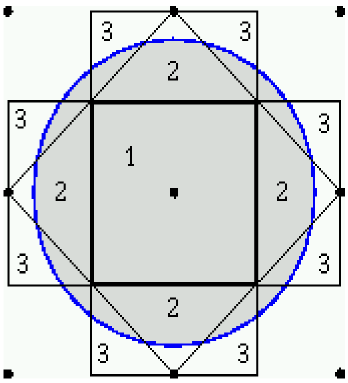


Fig.2.6

These are Brillouin zones of a square lattice in two dimensions. The blue circle shown is a surface of constant energy for free electrons; it will be the Fermi surface for some particular value of the electron concentration.

It is inconvenient to have sections of the Fermi surface that belong to the same Brillouin zone appear detached one from another. The detachment can be repaired by a transformation to the first Brillouin zone. The procedure is known as mapping the Fermi surface in the reduced zone scheme.

There is also another way to represent the Fermi surface in the reduced and periodic zone scheme. Fermi surfaces for free electrons are constructed by a procedure credited to Harrison, Fig.2.7.

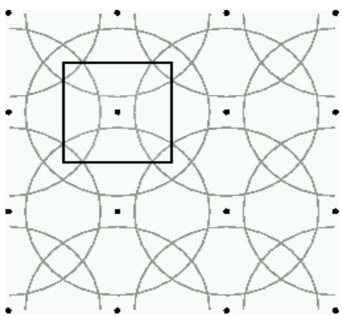


Fig.2.7

The reciprocal lattice points of a square lattice are determined and free-electron sphere of radius appropriate to the electron concentration is drawn around each point. Any point in **k** space that lies within at least one sphere corresponds to an occupied state in the first zone. Points within at least two spheres correspond to occupied states in the second zone, and similarly for points in three or more spheres. In Fig.2.8.

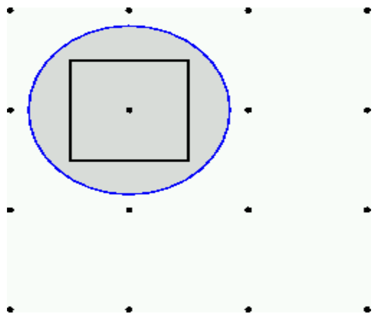


Fig.2.8

The black square shown is the first Brillouin zone, the blue circle is the surface of constant energy for free electrons, and the shaded area represents occupied electron states. As we can see, the first zone is entirely occupied. In Fig.2.9.

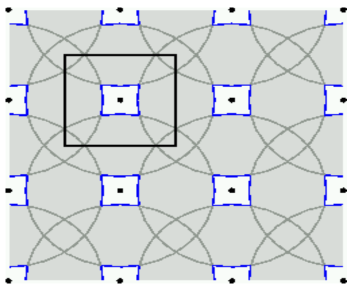


Fig.2.9

The black square shown is the first Brillouin zone, the blue lines are the Fermi surfaces for free electrons on the second zone, and the shaded area represents occupied electron states. In Fig.2.10.

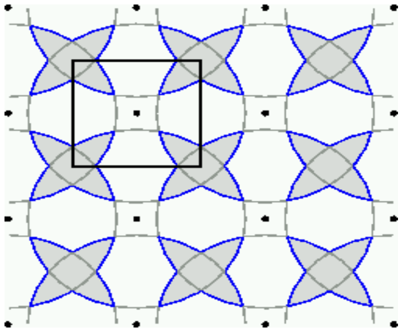


Fig.2.10

The black square shown is the first Brillouin zone, the blue lines are the Fermi surfaces for free electrons on the third zone, and the shaded area represents occupied electron states. In Fig.2.11.

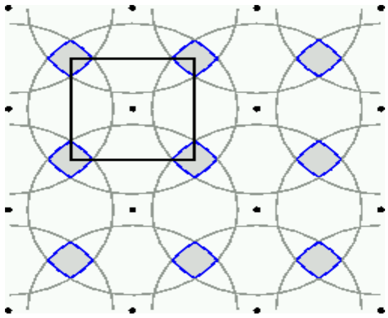


Fig.2.11

The black square shown is the first Brillouin zone, the blue lines are the Fermi surfaces for free electrons on the fourth zone, and the shaded area represents occupied electron states. Thus, in Fig.2.12.

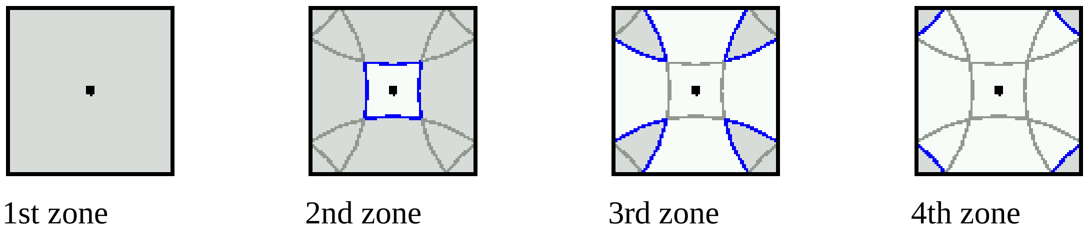


Fig.2.12

We show the free electron Fermi surface, as viewed in the reduced zone scheme. The shaded areas represent occupied electron states. Parts of Fermi surface (blue lines) fall in the second, third, and fourth zones. The first zone is entirely occupied. In Fig.2.13.

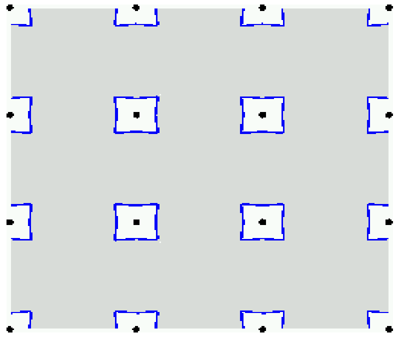


Fig.2.13

We show the Fermi surface for free electrons in the second zone as drawn in the periodic scheme. The figure can be constructed by repeating the second zone of Fig. Fig.2.12 or directly from Harrison construction. In Fig.2.14.

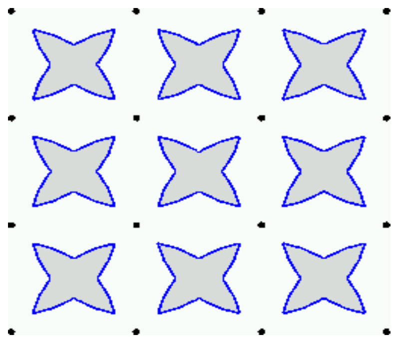


Fig.2.14

We show the Fermi surface for free electrons in the third zone as drawn in the periodic scheme. The figure can be constructed by repeating the third zone of Fig.2.12 or directly from Harrison construction. In Fig.2.15.

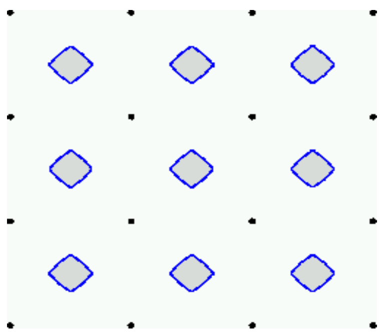


Fig.2.15

We show the Fermi surface for free electrons in the fourth zone as drawn in the periodic scheme. The figure can be constructed by repeating the fourth zone of Fig. Fig.2.12 or directly from the Harrison construction.

Effective mass of charge carriers (electron and holes):

Electrons in solids can be localized or delocalized: Wave packet formed by superposition of plain waves

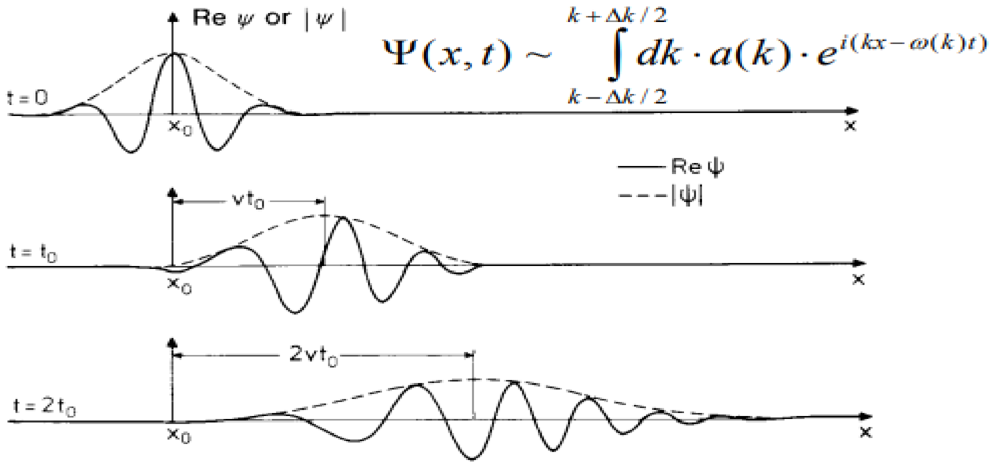


Fig.2.16 Wave packet formed by superposition of plain waves

Velocity of wave packet:

$$\vec{v} = \vec{\nabla}_k \omega(\vec{k}) = \frac{1}{\hbar} \vec{\nabla}_k \ \varepsilon(\vec{k})$$

External force on a crystal electron:

$$\vec{F} = \hbar \frac{d\vec{k}}{dt} = m^* \frac{d\vec{v}}{dt}$$

This yields equation of motion:

$$\frac{dv_i}{dt} = \frac{1}{\hbar} \frac{d}{dt} \left[\vec{\nabla}_k \ \varepsilon(\vec{k}) \right]_i$$

With the effective electron mass:

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2}\frac{\partial^2 \varepsilon(\vec{k})}{\partial k_i \partial k_j}$$

Determine by the curvature of the band structure.

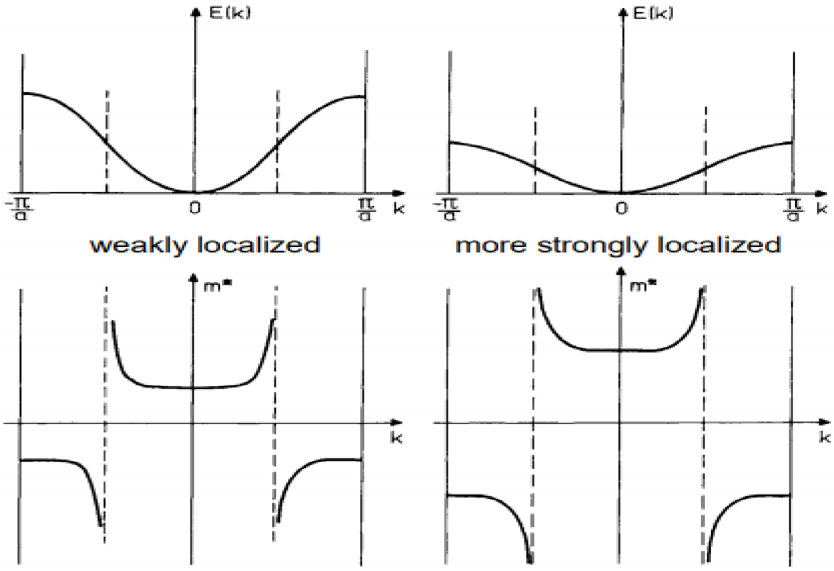


Fig.2.17 Curvature of the band structure

Isotropic case:

$$m^* = \frac{\hbar^2}{d^2\varepsilon / d^2k}$$

Approximate near center of parabolic band (points of high symmetry):

$$\varepsilon(\vec{k}) = \varepsilon_0 \pm \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2 + k_z^2)$$

How do electrons at ε(k) in electric field contribute to the electric current?

(a) Fully occupied band: vanishing current density

$$\vec{j} = \frac{-e}{8\pi^3} \int_{1st\ BZ} \vec{v}(\vec{k}) d\vec{k}$$

integration covers $-k_{max}...k_{max}$, $\sum \vec{k} = 0$

since $\vec{v}(-\vec{k}) = -\vec{v}(\vec{k})$ from $\vec{v}(\vec{k}) = \frac{1}{\hbar} \vec{\nabla} \varepsilon(\vec{k})$

$$\Rightarrow \vec{j} = 0$$

(b) Partially occupied band: finite current density direction force on electrons by electric field, symmetry around k=0 is lost:

$$\vec{j} \neq 0$$

$$\vec{j} = \frac{-e}{8\pi^3} \int_{occ.\,k} \vec{v}(\vec{k}) d\vec{k} = \frac{e}{8\pi^3} \int_{unocc.\,k} \vec{v}(\vec{k}) d\vec{k}$$

Electric current due to electrons at occupied states or holes in unoccupied states

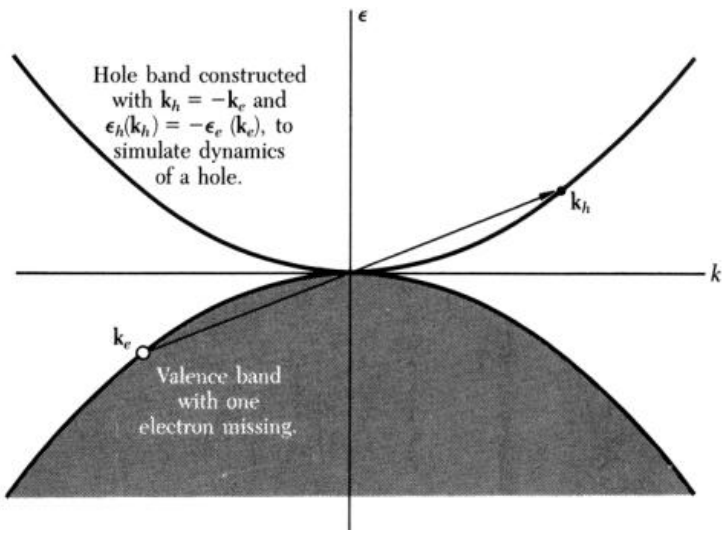


Fig.2.18

Consider a filled band with a missing electron: hole with vector $\vec{k}_h = -\vec{k}_e$ or the band has total wave vector $-\vec{k}_e$.

$\varepsilon(\vec{k}_h) = -\varepsilon(\vec{k}_e)$ as it takes more work to remove an electron at lower energy (→create a hole)

m^*_h =-m^*_e since it is determine by the band curvature $\partial^2\varepsilon / k^2$

SAQ.2

a) What do you mean by Bloch theorem?
b) Discuss about the Kronning-Penny model.
c) Define the Fermi surface.
d) What do you mean by effective mass of charge carriers?

2.8 Concentration in semiconductors:

Based on electrical conductivity, materials are divided into conductors, insulators and semiconductors. Usually, metals are good conductors of electricity and all dielectrics are insulators. The electrical conductivity of semiconductors lie in between metals and dielectrics. Good examples for semiconductor are germanium and silicon. These elements belong to IV group in the periodic table. At 0 K, these elements are insulators, whereas at room temperatures they possess certain amount of conductivity. Pure germanium and silicon are called intrinsic semiconductors. By adding a small quantity of either III group or V group element atoms as impurity into pure Ge or Si, the electrical conductivity of the material increases. This impure semiconductor is called an extrinsic semiconductor.

Intrinsic semiconductors—carrier concentration

Pure germanium or silicon crystal is called an intrinsic semiconductor. Each semiconductor atom possesses four valence electrons in the outermost orbit. To get stability, each of these atoms has to get eight electrons in the outermost orbit, so that each atom makes four covalent bonds with the surrounding four other atoms in the crystal. A two-dimensional representation of the crystal structure of silicon (or germanium) at 0 K is shown in Fig. 2.19(a). The band diagram of this material is shown in Fig. 2.19(b).

At 0 K, all the valence electrons of Si atoms are participating in covalent bonds and their energies constitute a band of energies called valence band. So, at 0 K, valence band is completely filled and conduction band is empty of electrons. The allowed band of energies above valence band is called conduction band. Suppose, if we raise the temperature of the semiconductor to some room temperature T K, at this temperature some of the electrons which are participating in covalent bonds and present in the top energy levels of valence band will take thermal energies. If the increase in thermal energy of electrons present in top energy levels of valence band is equal to or greater than energy gap of the semiconductor, then electrons come away from bonding and move freely inside the crystal as shown in Fig. 2.20(a). Now these electrons possess energies equal to the lower energy levels of conduction band. These free electrons participate in electrical conduction; hence the band in which these electrons present is named as conduction band. If an electron comes away from bonding, then that atom acquires one unit positive charge, then it participates in electrical conduction. This electron vacancy or electron deficiency of an intrinsic semiconductor is called hole. The electron vacancies in valence band will exist as holes in the valence band as shown in Fig. 2.20(b).

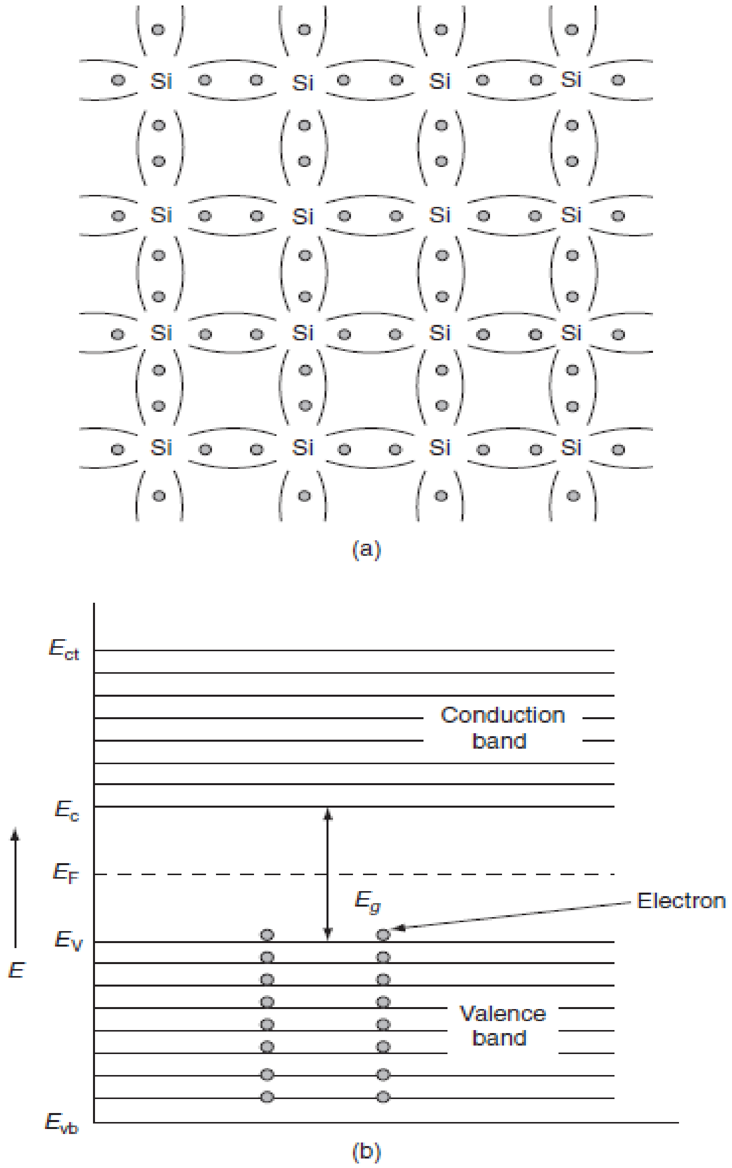
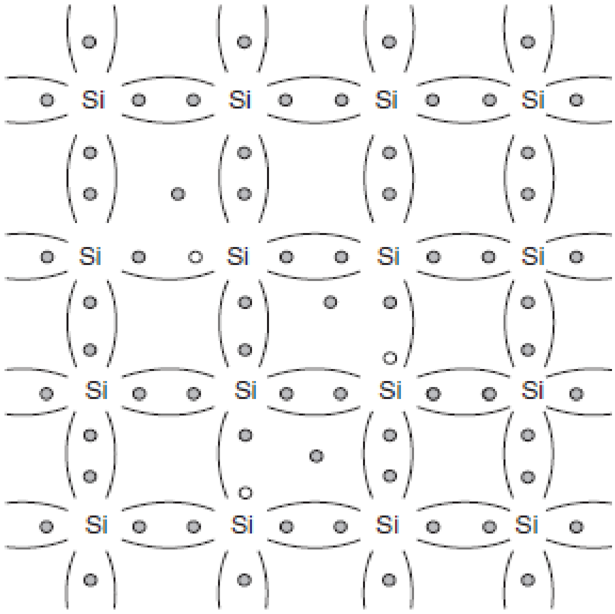


Fig.2.19 (a) Crystal structure of Si at 0 K; (b) Band diagram of Si at 0 K

Thus, at temperature T K, in an intrinsic semiconductor, if n covalent bonds are broken per unit volume of the material, then there will be n electrons in the conduction band and the same number of holes in the valence band. Usually, the number of free (or conduction) electrons present per unit volume of material, whose energies lie in the conduction band is called electron concentration and is represented as ‘n’. Similarly, the number of holes present per unit volume of the semiconductor and in the valence band is called hole concentration represented as ‘p’. Both the free electrons and holes present in the material participate in electrical conduction. The free electrons and holes present per unit volume of the material is called carrier concentration



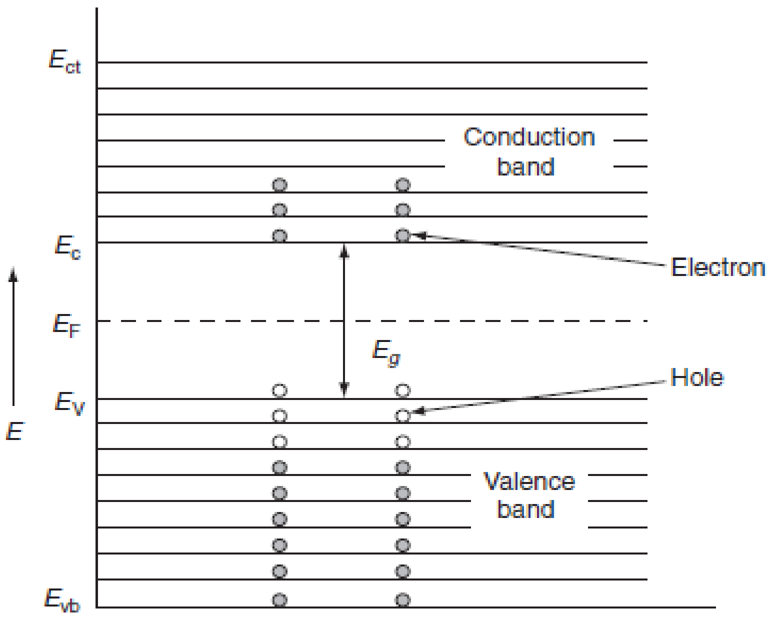


Fig.2.20 (a) Crystal structure of Si at *T* K; (b) Band diagram of Si at *T* K

At some temperature T K, the free electron and hole concentration in an intrinsic semiconductor can be extracted in the following way:

Electron concentration

The number of free electrons per unit volume of semiconductor having energies in between E and $E + \mathrm{d}E$ is represented as $N(E)\ \mathrm{d}E$ [i.e., in a width of dE]. $N(E)\ \mathrm{d}E$ can be obtained by multiplying the number of available electron states between E and $E + \mathrm{d}E$ per unit volume of the material, $g_e(E)\ \mathrm{d}E$ with the probability that an electron occupies an electron state at energy E [i.e., $f_e(E)$].

Therefore, we have:

$$N(E)\ \mathrm{d}E = g_e(E)\ \mathrm{d}E\ f_e(E) \quad \ldots\ldots\ldots\ldots\ldots(1)$$

The number of electrons present in the conduction band per unit volume of material 'n' is obtained by integrating $N(E)\ \mathrm{d}E$ between the limits E_c and E_{ct} where E_c and E_{ct} are the bottom and top energy levels of conduction band, respectively.

i.e.

$$n = \int_{E_c}^{E_{ct}} N(E)\mathrm{d}E = \int_{E_c}^{E_{ct}} g_e(E) f_e(E)\mathrm{d}E \quad \ldots\ldots\ldots\ldots\ldots(2)$$

Equation (2) can be represented as:

$$n = \int_{E_c}^{\infty} g_e(E) f_e(E)\mathrm{d}E - \int_{E_{ct}}^{\infty} g_e(E) f_e(E)\mathrm{d}E \quad \ldots\ldots\ldots\ldots\ldots(3)$$

Above E_{ct}, electrons will not be present. Hence, Equation (8.3) becomes:

$$n = \int_{E_c}^{\infty} g_e(E) f_e(E)\mathrm{d}E \quad \ldots\ldots\ldots(4)$$

The Fermi-Dirac distribution function $f_e(E)$ can be represented as:

$$f_e(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{K_B T}\right)} \qquad \text{…………………..(5)}$$

Compared to the exponential value, the '1' in the denominator is negligible.

i.e.

$$\exp\left(\frac{E - E_F}{K_B T}\right) >> 1$$

Hence

$$f_e(E) = \frac{1}{\exp\left(\frac{E - E_F}{K_B T}\right)} = \exp\left(\frac{E_F - E}{K_B T}\right) \qquad \text{………………(6)}$$

The density of electron states $g_e(E)$ in the energy space from $E = 0$ to E can be written as:

$$g_e(E) = \frac{\pi}{2}\left(\frac{8m_e^*}{h^2}\right)^{3/2} (E - 0)^{1/2} \qquad \text{……………(7)}$$

where m_e^* is the effective mass of an electron and h is Planck's constant.

$$g_e(E)\mathrm{d}E = \frac{\pi}{2}\left(\frac{8m_e^*}{h^2}\right)^{3/2} (E - 0)^{1/2}\, \mathrm{d}E \qquad \text{…………..(8)}$$

To evaluate n, the density of states is counted from E_c, since the minimum energy state in conduction band is E_c. So Equation (8) becomes:

$$g_e(E)\mathrm{d}E = \frac{\pi}{2}\left(\frac{8m_e^*}{h^2}\right)^{3/2}(E - E_c)^{1/2}\mathrm{d}E$$ ……………………(9)

Substituting Equations (8.6) and (8.9) in (8.4) gives:

$$\begin{aligned} n &= \int_{E_c}^{\infty}\frac{\pi}{2}\left(\frac{8m_e^*}{h^2}\right)^{3/2}(E - E_c)^{1/2}\exp\left(\frac{E_F - E}{K_B T}\right)\mathrm{d}E \\ &= \frac{\pi}{2}\left(\frac{8m_e^*}{h^2}\right)^{3/2}\int_{E_c}^{\infty}\left(E - E_c\right)^{1/2}\exp\left(\frac{E_F - E}{K_B T}\right)\mathrm{d}E \end{aligned}$$ …………(10)

The above equation can be simplified by the following substitution:

Put

$$\varepsilon = E - E_c$$

So

$$\mathrm{d}\varepsilon = \mathrm{d}E$$ ……………..(11)

In Equation (11), E_c is constant, as we change the variable E to ε in Equation (10), the integral limits also change.

In Equation (11), as $E \rightarrow E_c$ then $\varepsilon \rightarrow 0$ and $E \rightarrow \infty$, then ε also $\rightarrow \infty$.

With reference to Fig. 8.2(b), the exponential term in Equation (10) becomes:

$$\exp\left(\frac{E_F - E}{K_B T}\right) = \exp\left[\frac{(E_F - E_c) + (E_c - E)}{K_B T}\right] = \exp\left[\frac{(E_F - E_c) - \varepsilon}{K_B T}\right]$$

$$= \exp\left[\frac{(E_F - E_c)}{K_B T}\right]\exp\left(\frac{-\varepsilon}{K_B T}\right)$$
…………….(12)

Substituting Equations (11) and (12) in (10), we get:

$$n = \frac{\pi}{2}\left[\frac{8m_e^*}{h^2}\right]^{3/2}\int_0^\infty \varepsilon^{1/2}\exp\left[\frac{E_F - E_C}{K_B T}\right]\exp\left(\frac{-\varepsilon}{K_B T}\right)d\varepsilon$$

$$= \frac{\pi}{2}\left[\frac{8m_e^*}{h^2}\right]^{3/2}\exp\left[\frac{E_F - E_C}{K_B T}\right]\int_0^\infty \varepsilon^{1/2}\exp\left(\frac{-\varepsilon}{K_B T}\right)d\varepsilon$$
…..(13)

The integral (I) in the above equation can be simplified by substitution. Put $\varepsilon = x^2$

so that $d\varepsilon = 2x\,dx$

Then

$$I = \int_0^\infty x\exp\left(\frac{-x^2}{K_B T}\right)2x\,dx$$

$$= \int_0^\infty 2x^2\exp\left(\frac{-x^2}{K_B T}\right)dx$$

$$= \frac{\sqrt{\pi}}{2(1/K_B T)^{3/2}} = \frac{\sqrt{\pi}}{2}(K_B T)^{3/2}$$
……(14)

Substituting Equation (14) in (13) gives:

$$n = \frac{\pi}{2}\left(\frac{8m_e^*}{h^2}\right)^{3/2}\exp\left(\frac{E_F - E_C}{K_B T}\right)\frac{\sqrt{\pi}}{2}(K_B T)^{3/2}$$

$$= \frac{1}{4}\left(\frac{8m_e^*\pi K_B T}{h^2}\right)^{3/2}\exp\left(\frac{E_F - E_C}{K_B T}\right)$$

$$n = \frac{8}{4}\left[\frac{2m_e^*\pi K_B T}{h^2}\right]^{3/2} \exp\left(\frac{E_F - E_C}{K_B T}\right)$$

$$n = 2\left[\frac{2m_e^*\pi K_B T}{h^2}\right]^{3/2} \exp\left[-\left(\frac{E_C - E_F}{K_B T}\right)\right]$$

…………..(15)

The term

$$2\left[\frac{2m_e^*\pi K_B T}{h^2}\right]^{3/2}$$

is almost a constant compared with the exponential term as the temperature changes. So, it is a pseudo constant and is given by the symbol N_c. So, we have:

$$n = N_c \exp\left[-\frac{(E_C - E_F)}{K_B T}\right]$$

………(16)

For hole concentration

The number of holes per unit volume of semiconductor in the energy range E and E + dE in valence band is represented as P(E) dE. P(E) dE can be obtained by multiplying the number of available hole states between $E$and E + dE per unit volume of the material [i.e., $g_h(E)$ dE] with the hole probability in a hole state at energy E[i.e., $f_h(E)$].

Therefore, P(E) dE = $g_h(E)$ dE $f_h(E)$ _________ (8.17)

The number of holes present in the valence band per unit volume of material 'p' is obtained by integrating P(E) dE between the limits E_{vb} and E_V where E_V and E_{vb} are the top and bottom energy levels of valence band, respectively.

i.e.,

$$p = \int_{E_{vb}}^{E_V} \mathrm{P(E)}\, \mathrm{d}E = \int_{E_{vb}}^{E_V} \mathrm{g_h}(E) f_h(E) \mathrm{d}E$$

…………(18)

Equation (18) can be represented as:

$$p = \int_{-\infty}^{E_V} g_h(E)\, f_h(E) \mathrm{d}E - \int_{-\infty}^{E_{vb}} g_h(E) f_h(E) \mathrm{d}E$$

…………..(19)

below E_{vb} holes will not exist. Hence, Equation (19) becomes:

$$p = \int_{-\infty}^{E_V} g_h(E)\, f_h(E)\, \mathrm{d}E$$

…………….(20)

The presence of a hole can be represented as the absence of an electron. Hence, the Fermi-Dirac function of holes $f_h(E)$ in the valence band is:

$$f_h(E) = 1 - f_e(E) = 1 - \frac{1}{1 + \exp\left(\frac{E - E_F}{K_B T}\right)}$$

$$= \frac{\exp\left(\frac{E - E_F}{K_B T}\right)}{1 + \exp\left(\frac{E - E_F}{K_B T}\right)} = \frac{1}{1 + \frac{1}{\exp\left[\frac{E - E_F}{K_B T}\right]}} = \frac{1}{1 + \exp\left[\frac{E_F - E}{K_B T}\right]}$$

…(21)

Compared to exponential, the ‘1’ in the denominator is negligible, i.e.,

$$\exp\left(\frac{E_F - E}{K_B T}\right) >> 1$$

Hence

$$f_h(E) = \exp\left(\frac{E - E_F}{K_B T}\right) \quad \text{.............(22)}$$

The density of hole states between E and $E +$ dE in valence band can be written similar to Equation (9) for electrons.

$$g_h(E)\,dE = \frac{\pi}{2}\left[\frac{8m_h^*}{h^2}\right]^{3/2} (E_V - E)^{1/2}\,dE$$

where m_h^* is the effective mass of hole.

Substituting Equations (8.21) and (8.22) in (8.20),

We get

$$p = \int_{-\infty}^{E_V} \frac{\pi}{2}\left[\frac{8m_h^*}{h^2}\right]^{3/2} (E_V - E)^{1/2} \exp\left[\frac{E - E_F}{K_B T}\right] dE$$
$$= \frac{\pi}{2}\left[\frac{8m_h^*}{h^2}\right]^{3/2} \int_{-\infty}^{E_V} (E_V - E)^{1/2} \exp\left[\frac{E - E_F}{K_B T}\right] dE \quad \text{..........(23)}$$

The above equation can be simplified by the substitution:

Put $\varepsilon = E_V - E$ __________ (24)

so $d\varepsilon = - dE$

In Equation (24), E_V is constant, as we change the variable E to ε in Equation (23), the integral limits also change.

In Equation (24), as $E \rightarrow E_V$ then $\varepsilon \rightarrow 0$ and $E \rightarrow -\infty$, then $\varepsilon \rightarrow \infty$ with reference to Fig. 8.2(b), the exponential term in Equation (23) becomes:

$$\exp\left[\frac{E-E_F}{K_BT}\right]=\exp\left[\frac{(E-E_V)+(E_V-E_F)}{K_BT}\right]$$

$$=\exp\left[\frac{-\varepsilon+E_V-E_F}{K_BT}\right]=\exp\left[\frac{-\varepsilon}{K_BT}\right]\exp\left[\frac{E_V-E_F}{K_BT}\right] \quad \text{………(25)}$$

Substituting Equations (24) and (25) in (23), we get:

$$p=\frac{\pi}{2}\left[\frac{8m_h^*}{h^2}\right]^{3/2}\exp\left[\frac{E_V-E_F}{K_BT}\right]\int_0^{\infty}\varepsilon^{1/2}\exp\left(\frac{-\varepsilon}{K_BT}\right)d\varepsilon \quad \text{……………(26)}$$

From Equation (14), we know the integral value:

So,

$$p=\frac{\pi}{2}\left[\frac{8m_h^*}{h^2}\right]^{3/2}\exp\left[\frac{E_V-E_F}{K_BT}\right]\frac{\sqrt{\pi}}{2}(K_BT)^{3/2}$$

$$=\frac{1}{4}\left[\frac{8m_h^*\pi K_BT}{h^2}\right]^{3/2}\exp\left[\frac{E_V-E_F}{K_BT}\right]$$

$$=2\left[\frac{2m_h^*\pi K_BT}{h^2}\right]^{3/2}\exp\left[\frac{-(E_F-E_V)}{K_BT}\right] \quad \text{…………..(27)}$$

The term

$$2\left[\frac{2m_h^*\pi K_BT}{h^2}\right]^{3/2} \quad \text{…………..(28)}$$

is almost constant compared with the exponential term as the temperature changes. So, it is a pseudo constant and is given by the symbol N_V. So, we have:

$$p = N_V \exp\left[\frac{-(E_F - E_V)}{K_B T}\right] \quad \ldots\ldots\ldots\ldots(29)$$

2.9 Hall Effect (qualitative):

Mainly Lorentz force is responsible for **Hall Effect.** All of we know that when we place a current carrying conductor inside a magnetic field, the conductor experiences a mechanical force to a direction depending upon the direction of magnetic field and the direction of current in the conductor. The electric current means a flow of charge. In metal it is entirely due to the flow of electrons, in semiconductor, it is due to flow of free electrons as well as holes. In semiconductor, holes move in the direction of conventional current and free electrons move in the opposite of the direction of conventional current. As the electrons have charge, they experience a force while flowing through a conductor placed inside a magnetic field. Due to this force, the electrons get diverted towards one side of the conducted during flowing. As the following charges get shifted to one side of the conductor, there may be a tiny potential difference appeared across the cross-section of the conductor. We call this entire phenomenon as Hall Effect.

The original, classical Hall Effect was discovered in 1879 by Edwin Hall. It is a simple consequence of the motion of charged particles in a magnetic field. We'll start these lectures by reviewing the underlying physics of the Hall Effect. This will provide a useful background for our discussion of the quantum Hall effect.

Here's the set-up. We turn on a constant magnetic field, B pointing in the z-direction. Meanwhile, the electrons are restricted to move only in the (x, y)- plane. A constant current I is made to flow in the x-direction. The Hall Effect is the statement that this induces a voltage VH (H is for "Hall") in the y-direction. This is shown in the figure to the right.

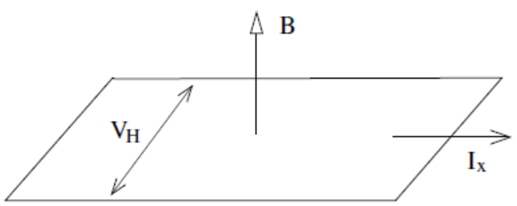


Fig.2.21 The classical Hall effect

Let us take a block of metal as shown below. We show here it as transparent for better understanding.

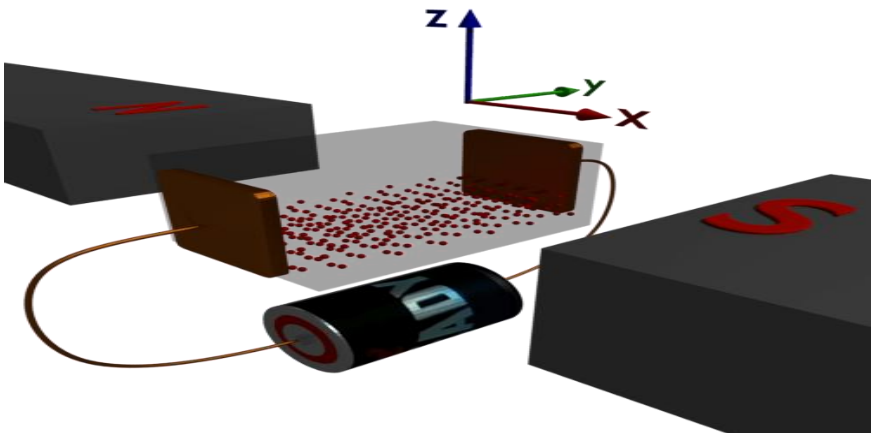


Fig.2.22 Connection for **Hall Effect**

After applying a voltage across it current, “I” starts flowing from left to right that is along the x-direction. We place the metal block in a magnetic field of density “B”. The direction of the magnetic field along the y-axis. Now by Fleming’s Left-Hand Rule, the charge carriers will experience a force depending on both – direction of current and direction of magnetic field. As per Fleming’s rule, the conduction electrons will deflect towards the bottom of the block. As a result, there will be a change in concentration of electrons at upper and lower portion of the block. Consequently, a tiny potential difference appears across the block along the z-axis. The electric field created due to this shifting of electrons also opposes the shifting of electrons towards the surface of the metal block. Hence there may be two forces acting on the charge carriers.

1. The force due to Hall effect
2. The force due to created electric field

These forces are opposite to each other. After the establishment of the certain electric field due to **Hall Effect** the system becomes in equilibrium. At that condition the force acting on the charge carriers (conduction electrons) due to the established electric field and due to Hall Effect become same, and opposite. Hence there would not be any further shifting of electrons towards the surface of the block and the system become in an equilibrium condition.

Let us concentrate on a single charge carrier moving through a conductor placed inside a magnetic field.

The charge of the charge carrier is e (say),

The magnetic flux density of the field is B (say).

Now we can write the magnetic force acting on the current carrying conductor of active length L and current I as

$$F = BIL$$

When the direction of B and I are perpendicular to each other.

We can rewrite the above equation as

$$F = B\frac{Q}{t}L = BQ\frac{L}{t} = BQV$$

Where V is the velocity of the charge passing through the conductor.

From the above equation, we can write the force acting on a single charge carrier as,

$$F = Bev$$

Where v is the drift velocity of the charge carriers.

Now we can write the field created due to **Hall Effect** as, E_H. Hence, we can write the force acting on the charge carrier due to the field as

$$F' = eE_H$$

Now at equilibrium

$$F = F' \;\Rightarrow Bev = eE_H \;\ldots\ldots (i)$$

Now we consider N is charge carrier concentration.

$$\therefore\; Current\ density\;\; J = Nev \ldots\ldots (ii)$$

From equation (i) and (ii), we can write

$$E_H = \frac{BJ}{Ne}$$
$$\Rightarrow \frac{E_H}{JB} = \frac{1}{Ne}$$

We call the term $\frac{E_H}{JB}$ as the Coefficient of Hall Effect or simply Hall Coefficient. We define Hall Coefficient as the Hall field per unit magnetic field density per unit current density.

Application of Hall Effect:

Hall Effect is a very useful phenomenon and helps to-

1. **Determine the Type of Semiconductor:**

By knowing the direction of the Hall Voltage, one can determine that the given sample is whether n-type semiconductor or p-type semiconductor. This is because Hall coefficient is negative for n-type semiconductor while the same is positive in the case of p-type semiconductor.

2. **Calculate the Carrier Concentration:**

The expressions for the carrier concentrations of electrons (n) and holes (p) in terms of Hall coefficient are given by

$$n = \frac{1}{qR_H} \quad and \quad p = \frac{1}{qR_H}$$

3. **Determine the Mobility (Hall Mobility):**

Mobility expression for the electrons (μ_n) and the holes (μ_p), expressed in terms of Hall coefficient is given by

$$\mu_n = \sigma_n R_H \quad and \quad \mu_p = \sigma_p R_H$$

Where, σ_n and σ_p represent the conductivity due to the electrons and the holes, respectively.

4. **Measure Magnetic Flux Density:**

This equation can be readily deduced from the equation of Hall voltage and is given by

$$B = \frac{V_H d}{R_H I}$$

Further, there are many commercially available types of equipment based on the principle of Hall Effect including Hall-effect sensors and Hall-effect probes.

SAQ.3

a) What do you mean by Concentration in semiconductors?
b) Discuss about the Hall Effect and Hall effect co-efficient.
c) Hall coefficient of a specimen of depend silicon found to be 3.66×10^{-4} m^3 C^{-1}. The resistivity of the specimen is 8.93×10^{-3} m. Find the mobility and density of the charge carriers.

Examples:

Q.1 *Fermi energy of copper is 7 eV at room temperature. What is the total number of free electrons/unit volume at the same temperature?*

Sol: Fermi energy, $E_F = 7\ eV = 7 \times 1.602 \times 10^{-19}\ J = 11.214 \times 10^{-19}\ J$

$$E_F = \left(\frac{h^2}{8m}\right)\left(\frac{3}{\pi}\right)^{2/3} n^{2/3}$$

$$11.214 \times 10^{-19} = \frac{\left[6.63 \times 10^{-34}\right]^2}{8 \times 9.11 \times 10^{-31}} \times \left[\frac{3 \times 7}{22}\right]^{3/2} n^{2/3}$$

$$11.214 \times 10^{-19} = \frac{43.9569 \times 10^{-68}}{72.88 \times 10^{-31}} \times 0.9326 \times n^{2/3}$$

$$n^{2/3} = \frac{11.214 \times 72.88}{43.9569 \times 0.9326 \times 10^{-18}} = 19.9364 \times 10^{18}$$

$$n = [19.9364 \times 10^{18}]^{3/2} \text{ electrons/m}^3 = 8.9106 \times 10^{28} \text{ electrons/m}^3$$

Q.2 *Calculate the free electron concentration, mobility and drift velocity of electrons in aluminium wire of length of 5 m and resistance of 0.06 Ω carrying a current of 15 A, assuming that each aluminium atom contributes 3 free electrons for conduction.*

Sol. Number of conduction electrons per m^3,

$$n = \frac{\text{no. of electrons per atom} \times N_A \times D}{\text{atomic weight}}$$

$$= \frac{3 \times 6.025 \times 10^{26} \times 2.7 \times 10^3}{26.98} = 1.8088 \times 10^{29}/\text{m}^3$$

We know

$$\rho = \frac{1}{ne\mu}$$

or

$$\mu = \frac{1}{ne\rho}$$

Mobility,

$$\mu = \frac{1}{1.8088 \times 10^{29} \times 1.6 \times 10^{-19} \times 2.7 \times 10^{-8}} = 0.00128 \ \text{m}^2/\text{VS}$$

Drift velocity,

$$v_d = \left(\frac{eE}{m}\right) \times \tau$$

$$E = \frac{V}{L} = \frac{IR}{L}$$

and $\sigma = \dfrac{ne^2\tau}{m}$ or $\tau = \dfrac{\sigma m}{ne^2} = \dfrac{m}{\rho ne^2}$

$$\therefore \ v_d = \frac{e}{m} \times \frac{IR}{L} \times \frac{m}{\rho ne^2} = \frac{IR}{L\rho ne}$$

$$= \frac{15 \times 0.06}{5 \times 2.7 \times 10^{-8} \times 1.8088 \times 10^{29} \times 1.6 \times 10^{-19}}$$

$$= \frac{0.9 \times 10^{-2}}{39.07} = 2.3 \times 10^{-4} \ \text{m/s.}$$

Q.3 Calculate the intrinsic concentration of charge carriers at 300 K given that $m^*_e = 0.12m_o$, $m^*_h = 0.28m_o$ and the value of brand gap = 0.67 eV.

Solution:

Given:

$$m_e^* = 0.12m_o = 0.12 \times 9.1 \times 10^{-31} = 1.092 \times 10^{-31} \text{Kgm}^{-3}$$
$$m_h^* = 0.28m_o = 0.28 \times 9.1 \times 10^{-31} = 2.548 \times 10^{-31} \text{Kgm}^{-3}$$
$$T = 300\text{K}.$$

Intrinsic carrier concentration is given by,

$$n_i = 2\left[\frac{2\pi kT}{h^2}\right]^{3/2} (m_e^* m_h^*)^{3/4} \exp\left[\frac{-E_g}{2K_B T}\right]$$

$$2\left[\frac{2\pi kT}{h^2}\right]^{3/2} = 2\left[\frac{2\pi \times 1.38 \times 10^{-23} \times 300}{6.626\times 10^{-34}}\right]^{3/2}$$
$$= 2\ (1.4421 \times 10^{70})$$
$$= 2.884 \times 10^{70}$$

$$(m_e^* m_h^*)^{3/4} = (1.092 \times 10^{31} \times 2.548 \times 10^{-31})^{3/4}$$
$$= 6.813 \times 10^{-47}$$

$$\exp\left[\frac{-E_g}{2K_B T}\right] = \exp\left[-\left(\frac{0.67 \times 1.6 \times 10^{-19}}{2 \times 1.38 \times 10^{-23} \times 300}\right)\right]$$
$$= \exp\ (-12.9468)$$
$$= 2.3838 \times 10^{-6}$$

$$n_i = [1.442 \times 10^{70}] \times 6.813 \times 10^{-47} \times 2.3836 \times 10^{-6}$$
$$= 2.3407 \times 10^{18}$$

$$\left.\begin{array}{l}\text{Intrinsic carrier}\\ \text{concentration}\end{array}\right\} n_i = 2.3407 \times 10^{18}\ \text{m}^{-3}$$

Q.4 The intrinsic carrier density is 1.5×10^{16} m^{-3}. If the mobility of electron and hole are 0.13 and 0.05 m^2 V^{-1} s^{-1}, calculate the conductivity.

Solution :

Given :

$$n_i = 1.5 \times 10^{16}\ m^{-3}$$

$$\mu_e = 0.13\ m^2\ V^{-1}\ s^{-1}$$

$$\mu_h = 0.05\ m^2\ V^{-1}\ s^{-1}$$

Conductivity $\sigma = n_i\, e\, (\mu_e + \mu_h)$

$$\sigma = 1.5 \times 10^{16} \times 1.6 \times 10^{-19}\ (0.13 + 0.05)$$

Conductivity $\sigma = 4.32 \times 10^{-4}\ \Omega^{-1} m^{-1}$

Q.5 The Intrinsic carrier density at room temperature in Ge is $2.37 \times 10^{19}\ m^3$ if the electron and hole mobilities are 0.38 and 0.18 $m^2\ V^{-1}\ s^{-1}$ respectively, calculate the resistivity.

Solution:

Given:

$$n_i = 2.37 \times 10^{19}\ m^3$$

$$\mu_e = 0.38\ m^2\ V^{-1}\ s^{-1}$$

$$\mu_h = 0.18\ m^2\ V^{-1}\ s^{-1}$$

Conductivity $\sigma = n_i\, e\, (\mu_e + \mu_h)$

$$= 2.37 \times 10^{19} \times 1.6 \times 10^{-19}\ (0.38 + 0.18)$$

$$= 2.1235\ \Omega^{-1}\ m^{-1}$$

Resistivity $\rho = \dfrac{1}{\sigma}$

$$\rho = \frac{1}{2.1235}$$

Resistivity $\rho = 0.4709\ \Omega\ m$

Q.6 The Hall coefficient of certain silicon specimen was found to be $-7.35 \times 10^{-5}\ m^3\ C^{-1}$ from 100 to 400 K. Determine the nature of the semiconductor. If the conductivity was found to be $200\ ^{-1}\ m^{-1}$. Calculate the density and mobility of the charge carrier.

Solution:

$$\text{Conductivity } \sigma = 200\ \Omega^{-1}\text{m}^{-1}$$

$$\text{Hall co-efficient } R_H = -7.35 \times 10^{-5}\text{m}^3\text{C}^{-1} \quad \ldots(1)$$

a) Density of electrons

$$n = \frac{-1}{R_H e} \text{ (from equation (1))}$$

$$n = \frac{1}{(7.35 \times 10^{-5} \times 1.609 \times 10^{-19})}$$

$$\text{(i.e)} = 8.455 \times 10^{22}\text{m}^{-3}$$

We know Conductivity

$$\sigma = n\, e\, \mu_e$$

b) Mobility

$$\mu = \frac{\sigma}{ne} = \frac{200}{8.455 \times 10^{22} \times 1.6 \times 10^{-19}}$$

$$= 0.0147$$

$$\text{Mobility } \mu = 0.0147\text{m}^2\text{v}^{-1}\text{s}^{-1}$$

$$\text{Density of electrons (n)} = 8.053 \times 10^{22}\ \text{m}^{-3}$$

$$\text{Mobility } (\mu) = 0.0147\text{m}^2\text{v}^{-1}\text{s}^{-1}$$

Q.7 In a P-type germanium, $n_i = 2.1 \times 10^{19}$ m^{-3}density of boran 4.5×10^{23} atoms /m^3. The electron and hole mobility are 0.4 and 0.2 $m^2\ v^{-1}\ s^{-1}$ respectively. What is its conductivity before and after addition of boron atoms?

Solution:

Given:

Intrinsic carrier concentration $n_i = 2.1 \times 10^{19}\ m^{-3}$

Mobility of electrons $\mu_e = 0.4\ m^2\ v^{-1}\ s^{-1}$

Mobility of holes $\mu_h = 0.2\ m^2\ v^{-1}\ s^{-1}$

a) Conductivity before the addition of boron atoms

$$\sigma = n_i e(\mu_e + \mu_h)$$

$$= 2.1 \times 10^{19} \times 1.6 \times 10^{-19}(0.4 + 0.2)$$

$$= 2.016\ \Omega^{-1}\ m^{-1}$$

b) Conductivity after the addition of boron atoms, Boron is a P-type impurity atom

$$\sigma = p e \mu_h$$
$$= 4.5 \times 10^{23} \times 1.6 \times 10^{-19} \times 0.2$$
$$\sigma = 14400\ \Omega^{-1}\ m^{-1}$$

Q.8 An N-type semiconductor has hall coefficient = 4.16 × 10^{-4} m^3 C^{-1}. The conductivity is 108 $^{-1}$ m^{-1}. Calculate its charge carrier density 'n_e'and electron mobility at room temperature.

Solution:

Given:

Hall Co-efficient $R_H = 4.16 \times 10^{-4}\ m^3\ C^1$

Conductivity $\sigma = 108\ \Omega m^{-1}$

1. For 'n' type the charge carriers density $n_e = \frac{-1}{R_H e}$ Here the negative signindicates the field direction alone.

$$n_e = \frac{3\pi}{8}\frac{-1}{R_H e}$$

$$n_e = \left[\frac{3 \times 3.14}{8}\right]\left[\frac{1}{1.6 \times 10^{-19} \times 4.6 \times 10^{-4}}\right]$$

$$n_e = 1.7690 \times 10^{22} m^{-3}$$

2. Electron mobility $\mu_e = \frac{\sigma_e}{n_e e}$

$$= \frac{108}{\left(1.7690 \times 10^{22} \times 1.6 \times 10^{-19}\right)}$$

$$\mu_e = 0.0381\ m^2 v^{-1} s^{-1}$$

Q.9 In an N-type semiconductor, the concentration of electron is 2 × 10^{22} m^{-3} . Its electrical conductivity is 112 $^{-1}$ m^{-1}. Calculate the mobility of electrons.

Solution:

Given:

Conductivity $\sigma = 112\ \Omega^{-1}\ m^{-1}$

Carrier conerntration of electron

$$n_i = 2 \times 10^{22}\ m^{-3}$$

Hall coefficient $R_H = \dfrac{1}{ne}$

$$= \frac{1}{2 \times 10^{22} \times 1.6 \times 10^{-19}}$$

$$= 3.125 \times 10^{-4}\ m^{-3}\ C^{-1}$$

Mobility $\mu = \sigma R_H = 112 \times 3.125 \times 10^{-4}$

$$= 0.035\ m^2\ v^{-1}\ s^{-1}$$

Q.10 For an intrinsic Semiconductor with a band gap of 0.7 eV, determine the position of EF at T = 300 K if m*h = 6m*e.

Solution:

Bandgap $Eg = 0.7\ eV = 0.7 \times 1.6 \times 10^{-19} V$

$T = 300\ K$

Fermi energy for an intrinsic semiconductor

$$E_F = \frac{E_g}{2} + \frac{3KT}{4} \log\left[\frac{m_h^*}{m_e^*}\right]$$

$$E_F = \left[\frac{0.7 \times 1.6 \times 10^{-19}}{2}\right] + \left[\frac{3 \times 1.38 \times 10^{-23} \times 300}{4}\right] \log_e 6$$

$$= 6.1563 \times 10^{-20}\ \text{Joules}$$

$$E_F = \frac{6.1563 \times 10^{-20}}{1.6 \times 10^{-19}}$$

Fermi energy level $E_F = 0.3847\ eV$

Q.11 *The electron and hole mobilites in a silicon sample are 0.135 and 0.048 m^2/V-s, respectively. Determine the conductivity of intrinsic Si at 300 K if the intrinsic carrier concentration is 1.5 × 10^{16} atoms/m^3.The sample is doped with 10^{23} phosphorous atoms/m^3. Determine the hole concentration and conductivity.*

Solution:

Mobility of electrons (μ_e) = 0.135 m^2/V-s

Mobility of holes (μ_h) = 0.048 m^2/V-s

Instrinsic carrier concentration (n_i) = 1.5 × 10^{16}/m^3

Conductivity (σ) = $n_i e$ ($\mu_e + \mu_h$) = 1.5 × 10^{16} × 1.6 × 10^{-19} [0.135 + 0.048]

= 1.5 × 1.6 × 0.183 × 10−3 = 0.439 × 10^{-3}/Ω-m.

Doping concentration, N_D = 10^{23} phosphorous atoms/m^3

hole concentration, p = ?

conductivity (σ_n) = ?

$$p = \frac{n_i^2}{N_D} = \frac{\left(1.5 \times 10^{16}\right)^2}{10^{23}} = 2.25 \times 10^9/\text{m}^3$$

$$\sigma_n = N_D e \mu_e = 10^{23} \times 1.6 \times 10^{-19} \times 0.135 = 2.16 \times 10^3/\Omega\text{-m.}$$

2.10 Summary:

1. Quantum free electron theory introduced by Summerfield in 1928. This theory is based on making small concepts. This theory was proposed by making small changes in the classical free electron theory and by retaining most of the postulates of the classical free electron theory.
2. Due to the quantization rules, a material in solid state possesses a set of allowed energy levels. For a metal containing N free electrons, there will be N such allowed energy levels, which are separated by energy differences that are characteristic of the material.
3. Bloch's theorem states that the wave function of an electron within a perfectly periodic potential.
4. The Kronning-Penney model is a simplified model for an electron in a one-dimensional periodic potential.
5. The Fermi surface is the surface of constant energy in k space. The Fermi surface separates the unfilled orbital’s from the filled orbital’s, at absolute zero.
6. Intrinsic carriers are the electrons and holes that participate in conduction. The concentration of these carriers is contingent upon the temperature and band gap of the material, thus affecting a material's conductivity.

7. As the electrons have charge, they experience a force while flowing through a conductor placed inside a magnetic field.
8. Due to this force, the electrons get diverted towards one side of the conducted during flowing. As the following charges get shifted to one side of the conductor, there may be a tiny potential difference appeared across the cross-section of the conductor. We call this entire phenomenon as Hall Effect.

2.11 Terminal Questions:

1) Explain in detail about the Need of free electron quantum theory.
2) Explain the Summerfeld Fermi model using band theory.
3) State Bloch theorem using One dimensional motion of electron in periodic potential
4) Explain in detail Kronning-Penny model. Also explain its features and importance.
5) Explain in detail of the Fermi surface.
6) What do you mean by effective mass of charge carriers and explain using electron and holes?
7) Explain in detail Concentration in semiconductors.
8) Define the Hall Effect and derive the equation of Hall Coefficient. Also write its applications.
9) *A uniform silver wire has a resistivity of* 1.54×10^{-8} *Ω–m at a temperature 300 K. For an electric field along the wire of 1* V/cm. *Calculate:*

   (i) The drift velocity

   (ii) The mobility and relaxation time of electrons assuming that there are 5.8×10^{28} conduction electrons per m^3 of the material,

   (iii) Calculate the thermal velocity of conduction electrons.
10) *The Fermi energy of silver is 5.5* eV, *and the relaxation time of electrons is* 3.97×10^{-14} s. *Calculate the Fermi velocity and the mean free path for the electrons in silver*.
11) *Mobilities of electrons and holes in a sample of intrinsic germanium at 300 K are 0.36* m^2*/V-s and 0.17* m^2*/ V-s, respectively. If the resistivity of the specimen is 2.12 Ω-m, compute the intrinsic concentration.*
12) *An intrinsic Ge at room temperature with a carrier concentration of* $2.4 \times 10^9\ m^{-3}$ *is doped with one Sb atom in* 10^6 *Ge atoms. What would be the concentration of holes if the Ge atom concentration is* $4 \times 10^{28}\ m^{-3}$*?*

13) Find the resistance of an intrinsic Ge rod 1 mm long, 1 mm wide and 1 mm thick at 300 K. the intrinsic carrier density 2.5 ×1019 m^{-3} is at 300 K and the mobility of electron and hole are 0.39 and 0.19 $m^2\ v^{-1}\ s^{-1}$.

14) Hall coefficient of a specimen of depend silicon found to be $3.66 \times 10^{-4}\ m^3\ C^{-1}$. The resistivity of the specimen is 8.93×10^{-3} m. Find the mobility and density of the charge carriers.

15) A silicon plate of thickness 1mm, breadth 10mm, and length 100mm is placed magnetic field of 0.5 wb/m^2 acting perpendicular to its thickness. If A 10^{-2} current flows along its length, calculate the Hall voltage developed if the Hall coefficient is 3.66 × 10–4 m3 / coulomb.

Unit 3-Lattice Vibration

Structure

3.1 Introduction:

When an external force is applied on a solid, this distance between its atoms changes and inter-atomic force works to restore the original dimension. The ratio of inter-atomic force to that of change in inter-atomic distance is defined as the inter-atomic force constant. Solids can be classified on the basis of the bonds that hold the atoms or molecules together. This approach categorizes solids as molecular, covalent, ionic, or metallic. There are four types of crystalline solids: ionic solids, molecular solids, network covalent solids and metallic solids.

Lattice energy is an estimate of the bond strength in ionic compounds. It is defined as the heat of formation for ions of opposite charge in the gas phase to combine into an ionic solid. Lattice Vibration is the oscillations of atoms in a solid about the equilibrium position. For a crystal, the equilibrium positions form a regular lattice, due to the fact that the atoms are bound to neighboring atoms. The vibration of these neighboring atoms is not independent of each other.

The monatomic chain is a one-dimensional model representing the situation in a crystal with a primitive lattice, i.e. with only a single atom in the unit cell. ... In the same way, we can use a one-dimensional diatomic chain model to represent centered lattices, where more than one atom is present in the unit cell. The wave vector which characterizes the Brillouin zones restrict the number of vibrational modes in the crystal. A unit cell consisting of two atoms therefore leads to the formation of 3 branches; 2 acoustic and one optical.

Near room temperature, the heat capacity of most solids is around 3k per atom (the molar heat capacity for a solid consisting of n-atom molecules is ~3nR). This is the well-known Dulong and Petit law. It is the amount of heat required to change, The temperature of unit mass of substance by unit degree temperature. In Classical Theory of Specific heat of a solid; Solid molecules have 6 degrees of freedom (3 translational and 3 vibrational). Dulong–Petit law, statement that the gram-atomic heat capacity (specific heat times atomic weight) of an element is a constant; that is, it is the same for all solid elements, about six calories per gram atom.

Einstein developed the quantum theory of heat capacity and proposed that atoms in a solid should be treated as Planck oscillators characterised by a constant frequency. This implies that energy (E,) of a quantum oscillator is nhv, where n is a positive integer, h is Planck constant and v is frequency of oscillation. A theory of the specific heat of solids proposed by Albert Einstein in 1906; In this theory, Einstein attributed the specific heat of solids to the vibrations of the solid and made the simplifying assumption that all the vibrations have the same frequency. A theory of the specific heat capacity of solids put forward by Peter Debye in 1912, in which it was assumed that the specific heat is a consequence of the vibrations of the atoms of the lattice of the solid. This result is in very good agreement with experiment at low temperatures.

A temperature, characteristic of a substance, that appears in Einstein's equation for specific heat; it is equal to the product of Planck's constant and the Einstein frequency divided by Boltzmann's constant. A theory of the specific heat of solids proposed by Albert Einstein in 1906; This theory was partially successful since it was able to derive Dulong and Petit's law at high temperatures and showed that the specific heat capacity goes to zero as the absolute temperature also goes to zero. In Debye theory, the Debye temperature is the temperature of a crystal's highest normal mode of vibration, i.e., the highest temperature that can be achieved due to a single normal vibration. The Debye temperature is given by. Where h is Planck's constant, k is Boltzmann's constant, and is the Debye frequency.

3.2 Objectives:

After studying this unit you should be able to

- Explain and identify Inter atomic force and classification of solids
- Study and identify Lattice energy of ionic crystals
- Explain Vibration of mono atomic and diatomic linear chain, acoustic and optical modes, phonon.
- Study and identify Thermal capacity of solids, classical theory of specific heats (Dulong and Petit's law).
- Explain and identify Experimental results and need of quantum theory of specific heat of solids.
- Study and identify Einstein's theory of specific heats (need, statement, assumptions, derivations and limitations)
- Explain and identify Debye theory of specific heats (need, statement, assumptions, derivations and limitations).
- Study and identify Concept of Einstein's temperature and Debye temperature

3.3 Inter-atomic force :

The inter-atomic forces are electrical in nature. When two atoms are brought near to each other, then the charges on each atom are distributed. As a result of this, attractive inter-atomic force is produced between them. This inter-atomic force increases as the distance between the atoms decreases. When the distance between two atoms is very small, then the electron clouds of two start overlapping and the inter-atomic becomes repulsive.

An object is an arrangement of atoms or molecules in three dimensions. The variation of Potential energy (U) with intermolecular distance (r) is as shown in the figure. At equilibrium distance 'r_o' the potential energy is minimum. So, it is the stable state. The variation of inter-atomic force with the distance between atoms is shown in the figure. The inter-atomic force is equal to the negative gradient of the corresponding potential energy function

$$F = -\frac{dU(r)}{dr}$$

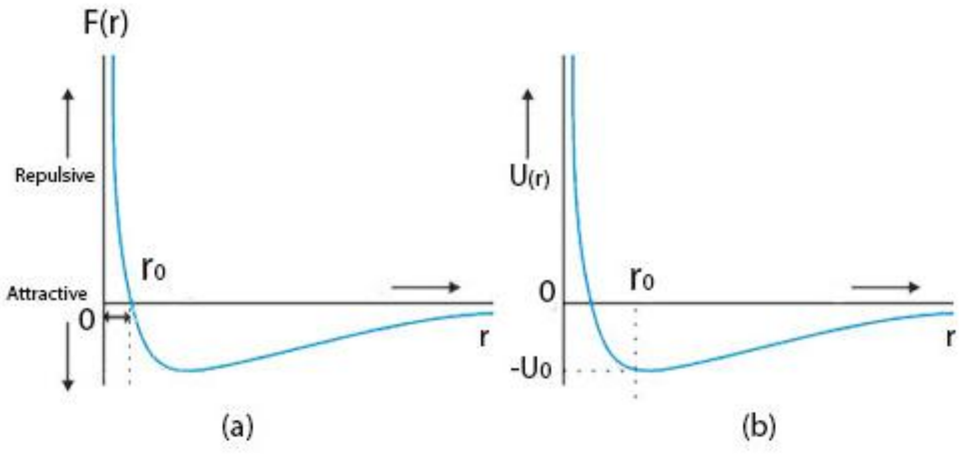


Fig.3.1 the variation of inter-atomic forces

Intermolecular forces:

The force which is responsible for holding together the atoms or molecules of a matter is called intermolecular force. When two molecules are far from each other, the forces between them are attractive in nature and negligible. As these molecules are brought closer to each other, the forces of attraction between them increase. It has been found that the intermolecular force of attraction is proportional to the seventh power of the distance between the molecules

$$\text{i.e.} \quad \boldsymbol{F_a} \propto \frac{\boldsymbol{1}}{\boldsymbol{r^7}}$$

$$\text{or} \quad \boldsymbol{F_a} = -\frac{\boldsymbol{A}}{\boldsymbol{r^7}} \quad \ldots\ldots\ldots\ldots.(\text{i})$$

Where A constant related to the nature of the molecules. The negative sign indicates that the force is attractive in nature.

As the distance between molecules is decreased and made equal to the order of the dimension of the molecule, they begin to repeat each other. The force changes rapidly than the attractive force. The force varies inversely as the ninth power of the intermolecular separation.

$$\text{i.e.} \quad \boldsymbol{F_r} = \frac{\boldsymbol{1}}{\boldsymbol{r^9}}$$

$$\text{or,} \quad \boldsymbol{F_r} = -\frac{\boldsymbol{B}}{\boldsymbol{r^9}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(\text{ii})$$

Where B is constant related to the natural of the molecule;

Thus, the net forces acting on a molecule is given by

$$\mathbf{F} = -\frac{\mathbf{A}}{\mathbf{r}^7} + \frac{\mathbf{B}}{\mathbf{r}^9}$$

There is a definite distance r_o between the molecules at which the force of attraction and repulsion balance each other and hence the resultant force acting is zero. That is F =0 at r = r_o.

$$\text{Or, } \frac{A}{r_0^7} = \frac{B}{r_0^9}$$

$$\text{Or, } \boldsymbol{r_0} = \sqrt{\frac{B}{A}}$$

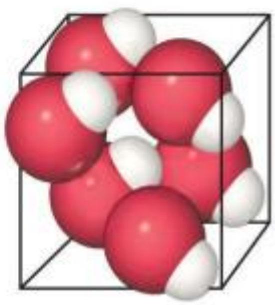
At this state, the molecules are in the state of stable equilibrium. As the intermolecular separation, r is decreased (i.e. $r<r_o$) the repulsion component given by equation (ii) dominates the attractive component of the force gives by equation (i).

Classification of solids:

Solids are classified into four types; based on the intermolecular forces operating in them.

1. Molecular Solids
2. Ionic Solids
3. Metallic Solids
4. Covalent solids

1 - Molecular Solids – Solids having molecules as their constituent particles are called Molecular solids. For, example, Hydrogen, Chlorine, Water, HCl, solid carbon dioxide, sucrose, etc.



Molecular solids

Discrete molecules held together by intermolecular forces (HBr, H_2O)

Fig.3.2

Molecular solids are classified into three types on the basis of their bond:

a. Non-Polar Molecular solids
b. Polar Molecular Solids
c. Hydrogen Bonded Molecular Solids

(a) Non Polar Molecular Solids – Solids which are comprised of atoms only, such as helium and argon or molecules; formed because of the non polar covalent bonds are known as Non-Polar Molecular Solids. For example – H_2, Cl_2, I_2, etc.

Characteristic of Non-Polar Molecular Solids –

- The molecules of non-polar molecular solids are held together by weak dispersion forces or London forces.
- Non-Polar Molecular Solids are soft.
- Non-polar molecular solids are non-conductor of electricity.
- Non-polar molecular solids have low melting points.
- Non-polar molecular solids are usually in liquid or gaseous state at the room temperature and pressure.

(b) Polar Molecular Solids – The solids which are formed by polar covalent bonds are known as Polar Molecular solids. For example – HCl, SO_2, NH_3, etc.

Characteristic of Polar Molecular Solids –

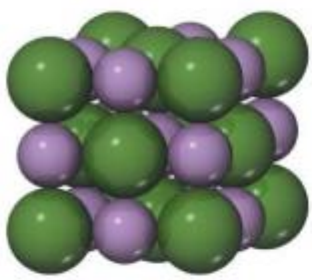
- The molecules in polar molecular solids are held together with dipole-dipole interactions.
- Polar molecular solids are generally soft in nature.
- Polar molecular solids are non-conductor of electricity.
- Polar molecular solids have higher melting points in comparison to non-polar molecular solids.
- Most of the polar molecular solids are gases or liquids at room temperature and pressure.
- Solid SO_2 and solid NH_3 are some examples of polar molecular solids.

(c) Hydrogen bonded Molecular Solids – The molecules of hydrogen bonded molecular solids contain polar covalent bond between H and O, F or N. In solids such as H2O (ice) molecules are bound together strongly with hydrogen bond. HF, H2O (ice), etc are the examples of hydrogen bound molecular solids.

Characteristics of Hydrogen bonded Molecular Solids –

- Hydrogen bound molecular solids are generally volatile liquid or soft solids at room temperature and pressure.
- Hydrogen bound molecular solids are non-conductor of electricity.

2 - Ionic Solids – Solids, in which ions are the constituent particles, are called ionic solids. These solids are formed because of three dimensional arrangements of cations and anions bound together with strong electrostatic forces (coulombic forces). For example NaCl.



Ionic solids

Extended networks of ions held together
by ion–ion interactions (NaCl, MgO)

Fig.3.3

Characteristics of Ionic Solids –

- High melting and boiling points.
- Non-conductor of electricity in solid state.
- Conductor of electricity in molten state.
- Conducted electricity when dissolved in water.

3 - Metallic Solids – All metals are referred as Metallic solids. Their constituent particles are positive ions. These positive ions are surrounded by free moving electrons. For example – iron, aluminium, etc.

Metallic solids

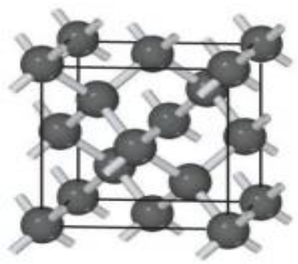
Extended networks of atoms held together by metallic bonding (Cu, Fe)

Fig.3.4

Characteristics of Metallic Solids –

- High melting points.
- Good conductors of electricity and heat.
- Lustrous, and are of specific colors.
- Hard but malleable and ductile in nature

4 - Covalent Solids – Crystalline solids are formed by non metals because of formation of covalent bonds between the adjacent molecules throughout the crystal. These are also known as Network Solids. These are also called giant molecules. For example – diamond, graphite, silicon carbide, etc.



Covalent-network solids

Extended networks of atoms held together by covalent bonds (C, Si)

Fig.3.5

3.4 Lattice energy of ionic crystals:

Lattice Energies and the Strength of the Ionic Bond:

The force of attraction between oppositely charged particles is directly proportional to the product of the charges on the two objects (q_1 and q_2) and inversely proportional to the square of the distance between the objects (r^2).

$$F = \frac{q_1 \times q_2}{r^2}$$

The strength of the bond between the ions of opposite charge in an ionic compound therefore depends on the charges on the ions and the distance between the centers of the ions when they pack to form a crystal.

An estimate of the strength of the bonds in an ionic compound can be obtained by measuring the **lattice energy** of the compound, which is the energy given off when oppositely charged ions in the gas phase come together to form a solid.

Example: The lattice energy of NaCl is the energy given off when Na^+ and Cl^- ions in the gas phase come together to form the lattice of alternating Na^+ and Cl^- ions in the NaCl crystal shown in the figure below.

$$Na^+(g) + Cl^-(g) \rightarrow NaCl(s) \qquad \Delta H^o = -787.3 \text{ kJ/mol}$$

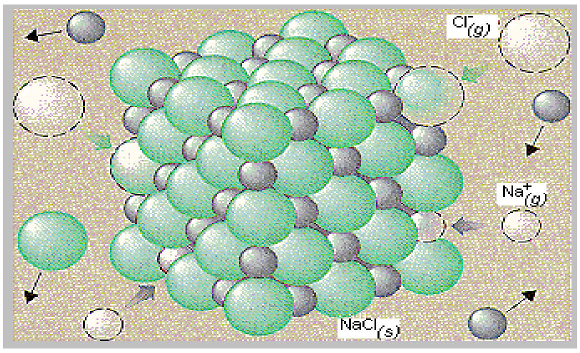


Fig.3.6 The lattice energy of NaCl crystal

The lattice energies of ionic compounds are relatively large. The lattice energy of NaCl, for example, is 787.3 kJ/mol, which is only slightly less than the energy given off when natural gas burns.

The bond between ions of opposite charge is strongest when the ions are small.

The lattice energies for the alkali metal halides is therefore largest for LiF and smallest for CsI, as shown in the table below.

Lattice Energies of Alkali Metals Halides (kJ/mol)

	F^-	Cl^-	Br^-	I^-
Li^+	1036	853	807	757
Na^+	923	787	747	704
K^+	821	715	682	649
Rb^+	785	689	660	630
Cs^+	740	659	631	604

The ionic bond should also become stronger as the charge on the ions becomes larger. The data in the table below show that the lattice energies for salts of the OH^- and O^{2-} ions increase rapidly as the charge on the ion becomes larger.

Lattice Energies of Salts of the OH^- and O^{2-} Ions (kJ/mol)

	OH^-	O^{2-}
Na^+	900	2481
Mg^{2+}	3006	3791
Al^{3+}	5627	15,916

Lattice Energies and Solubility

When a salt, such as NaCl dissolves in water, the crystals disappear on the macroscopic scale. On the atomic scale, the Na^+ and Cl^- ions in the crystal are released into solution.

$$NaCl(s) \xrightarrow{H_2O} Na^+(aq) + Cl^-(aq)$$

The lattice energy of a salt therefore gives a rough indication of the solubility of the salt in water because it reflects the energy needed to separate the positive and negative ions in a salt.

Sodium and potassium salts are soluble in water because they have relatively small lattice energies. Magnesium and aluminum salts are often much less soluble because it takes more energy to separate the positive and negative ions in these salts. NaOH, for example, is very soluble in water (420 g/L), but $Mg(OH)_2$ dissolves in water only to the extent of 0.009 g/L, and $Al(OH)_3$ is essentially insoluble in water.

SAQ.1

a) What do you mean by Inter-atomic force?
b) Define the Intermolecular forces.
c) What is the classification of solids?
d) What do you mean by Lattice energy of ionic crystals?
e) Use your knowledge of different types of intermolecular forces to explain the following statement: The boiling point of F_2 is much lower than the boiling point of NH_3.

3.5 Vibration of mono-atomic and diatomic linear chain, acoustic and optical modes, phonon:

Lattice vibration:

To get into the deeper knowledge of Lattice Vibrations first we should understand what do the "vibrations in a lattice" means. It is well known that the Heisenberg's uncertainty principle accounts for the vibration of atoms (in a real crystal) around their equilibrium position even at absolute zero temperature. At this temperature the energy of each atom is known as zero point energy with the amplitude of vibrations known as zero point amplitude. The amplitude of vibrations of the atoms around their equilibrium position starts increasing when the temperature of the crystal is increased. The atoms gain more thermal energy at higher temperatures and thus start oscillating with greater amplitude as shown in Fig.1.

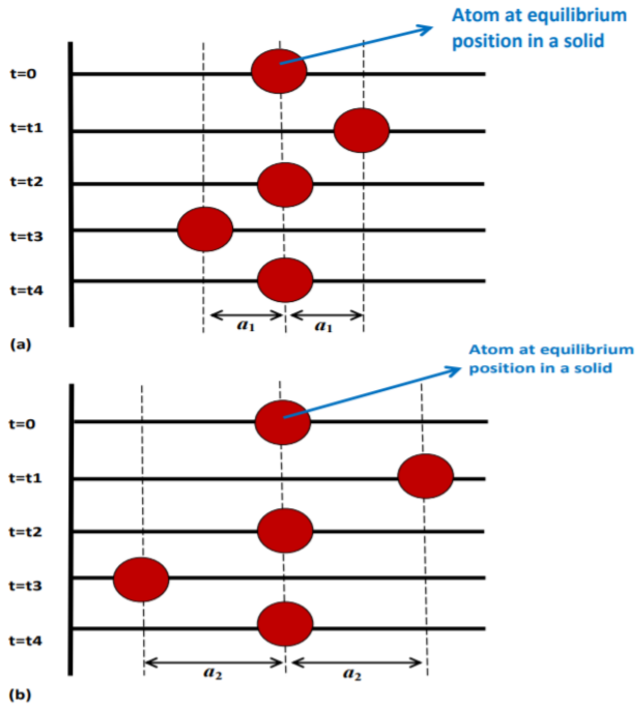


Fig.7 (a) The atoms in a crystal at absolute zero temperature with zero point motion of atoms with displacement from the equilibrium position as a_1 (b) The atoms in a crystal at room temperature with increased amplitude of vibrations of atoms with displacement from the equilibrium position as a_2.

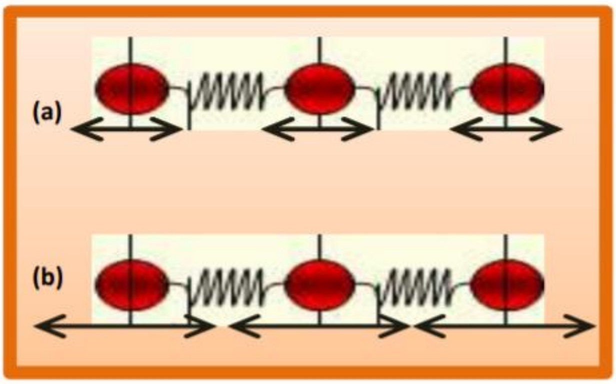


Fig.8 (a) The atoms in a crystal at absolute zero temperature with zero point motion of atoms with displacement from the equilibrium position as a_1 (b) The atoms in a crystal at room temperature with increased amplitude of vibrations of atoms with displacement from the equilibrium position as a_2.

In a real crystal atoms are not bounded to their equilibrium positions only, but the motion of one atom also affects the motion of neighbouring atom. Thus when one atom of a crystal vibrated about the equilibrium position the neighbouring atom also start vibrating and so the next neighbouring atom. In this way when an entire group of atoms vibrate in a coordinated way it is referred to as a Lattice vibration. The forces which lock the atoms in a crystal to their equilibrium position are directly proportional to their displacements from equilibrium position in the elastic limit and therefore we assume that atoms are being bound by elastic springs between them as shown in Fig.3.9. This assumption is known as Harmonic approximation where we have assumed the particles (atoms) of a crystal being coupled by an ideal elastic spring. In this approximation atoms vibrate about the equilibrium position under a simple harmonic oscillation (like a simple harmonic oscillator).

One important point to mention is that, the lattice vibration character is highly dependent on the–

(a) Number of atoms in one unit cell of crystal (Monoatomic, Diatomic, Triatomic etc.),

(b) Symmetry of the crystal,

(c) Type of chemical bond between the atoms,

(d) Crystal defect concentration.

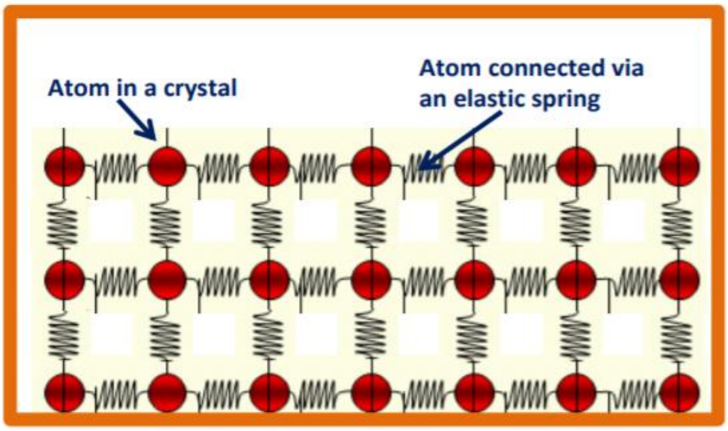


Fig.3.9. The model of a simplest lattice showing the elastic coupling between the neighbouring atoms.

Lattice vibrations as explained above accounts for the thermal properties of the crystals and contribute to the heat capacity of metals. In our further discussion we want to study the traits of elastic vibrational motion of the crystal lattices by considering the case of one dimensional monoatomic and di-atomic chain of atoms first. In reality the crystal is a 3D structure but to simplify the problem we have reduced the system to lower dimension (1D) and if required then we can generalize the results to 2D and 3D.

Dynamics of one dimensional infinite monoatomic chain of atoms:

To investigate the dynamics of the vibrational motion in an infinite 1D-chain of identical atoms (each having the same mass m), we assume that the distance between the equilibrium position of nearest-neighbouring atoms is a such that the total number of atoms N in the chain is very large . The system we have considered is non-homogeneous as atoms are separated from one another, and being bound by the ideal elastic springs between them. The 1D-chain of atoms is assumed to be lying along x-axis (Fig.3.10).

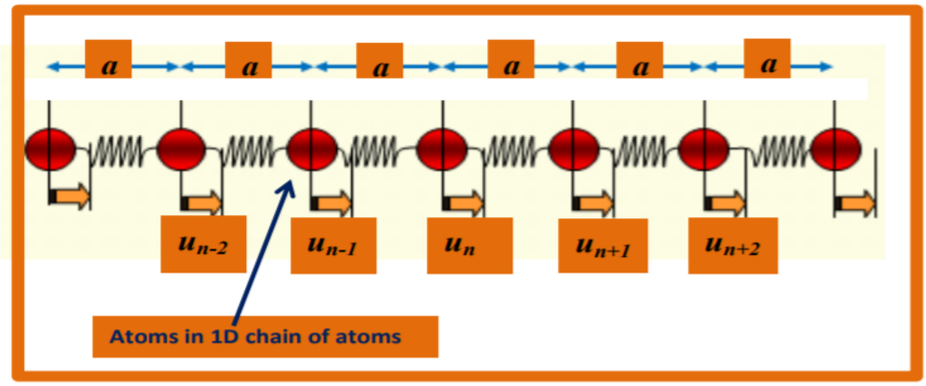


Fig.3.10 The model of a linear (1D) monoatomic lattice

The x-coordinates of the atoms present in the 1D-chain at $(n+2)^{th}$, $(n+1)^{th}$, n^{th}, $(n-1)^{th}$, $(n-2)^{th}$, ………sites are at $x_{n+2}=(n+2)a$, $x_{n+1}=(n+1)a$, $x_n=na$, $x_{n-1}= (n-1)a$, $x_{n-2}=(n-2)a$ ……… Similarly the symbols u_{n+2}, u_{n+1}, u_n, u_{n-1}, u_{n-2}, …… represents the displacements of the atoms present at $(n+2)^{th}$, $(n+1)^{th}$, n^{th}, $(n-1)^{th}$, $(n-2)^{th}$, ………sites in the 1D-chain within the elastic limits. These displacements arises due to excitation in vibrational motion of the atoms, otherwise the atoms execute only zero-point oscillations around their mean positions (or more or less stay at the equilibrium positions). In the elastic limit it is assumed that the restoring forces acting between the nearest-neighbour atoms are linear. Under these assumptions we can write the force equation for the n^{th} atom as –

$$\begin{aligned} F_n &= f_1 - f_2 \\ &= \eta\left(u_{n+1} - u_n\right) - \eta\left(u_n - u_{n-1}\right) \\ &= \eta\left(u_{n+1} + u_{n-1} - 2u_n\right) \end{aligned} \qquad \text{……………(a)}$$

In the above equation we have assumed that η is the spring constant or force constant (or force of interaction per unit displacement), F_n is the net force acting on the n^{th} atom, and $(u_{n+1}-u_n)$, (u_n-u_{n-1}) are the extensions produced in the springs connected with the n th atom. The force $f_1 = \eta(u_{n+1}-u_n)$, acts in the right direction, while the force $f_2 = \eta(u_n-u_{n-1})$ acts in the left direction. In writing the above force equation we have strictly neglected the effect of atoms other than nearest ones on n^{th} atom. According to Newton's second law of motion Eqn.(1) can be rewritten as –

$$m\frac{d^2x_n}{dt^2} = m\ddot{x}_n = \eta\left(u_{n+1} + u_{n-1} - 2u_n\right) \qquad \text{…………………………(b)}$$

The solution of above equation of motion can be assumed as travelling waves (i.e. the displacements produced in atoms are in the form of travelling waves) as –

$$u_n = u_0 e^{i(\omega t - Kna)} \qquad \text{……………………………….(1)}$$

$$u_{n+1} = u_0 e^{i(\omega t - K(n+1)a)} \qquad \text{…………………………….(2)}$$

$$u_{n-1} = u_0 e^{i(\omega t - K(n-1)a)} \qquad \text{…………………………..(3)}$$

Where, K is the wave-vector, u_0 is amplitude of oscillation of an atom, ω is the frequency of oscillations which is same for each atom (i.e. all atoms in lattice vibration oscillate with the same frequency). Substituting the Eqns.(1), (2), (3) in Eqn.(b) we arrive at the following equation –

$$\begin{aligned} -m\omega^2 &= \eta\left(e^{iKa} - 2 + e^{-iKa}\right) \\ &= \eta\left(e^{iKa/2} - e^{-iKa/2}\right)^2 \end{aligned} \qquad \text{……………….(4)}$$

Substituting the,

$$\sin(x) = \left(\frac{e^{ix} - e^{-ix}}{2i}\right)$$

and

$$\sin^2(x) = -\frac{1}{4}\left(e^{ix} - e^{-ix}\right)^2$$

in the above equation we can again rewrite it as –

$$-m\omega^2 = -4\eta \sin^2\left(\frac{Ka}{2}\right) \qquad \text{………………….(5)}$$

$$\begin{aligned} \Rightarrow \omega &= \left|\sqrt{\frac{4\eta}{m}} \sin\left(\frac{Ka}{2}\right)\right| = +\frac{2}{a}\sqrt{\frac{c}{\rho}}\left|\sin\left(\frac{Ka}{2}\right)\right| \\ &= \omega_{\max}\left|\sin\left(\frac{Ka}{2}\right)\right| \end{aligned} \qquad \text{………………….(6)}$$

Where the maximum value or cut-off value of frequency is $\omega_{\max} = 2v_s / a$ with $v_s = \sqrt{c/\rho}$, $\rho = m/a$ being the mass per unit length and η = c /a is the longitudinal stiffness per unit length. As frequency is a positive quantity therefore we have neglected the negative solution in Eqn.(6) and have considered only magnitude. Above relation in Eqn.(6) is known as Dispersion relation. This dispersion curve is shown in Fig.3.11, which is periodic in nature with a period of 2π and symmetric about the origin at K=0.

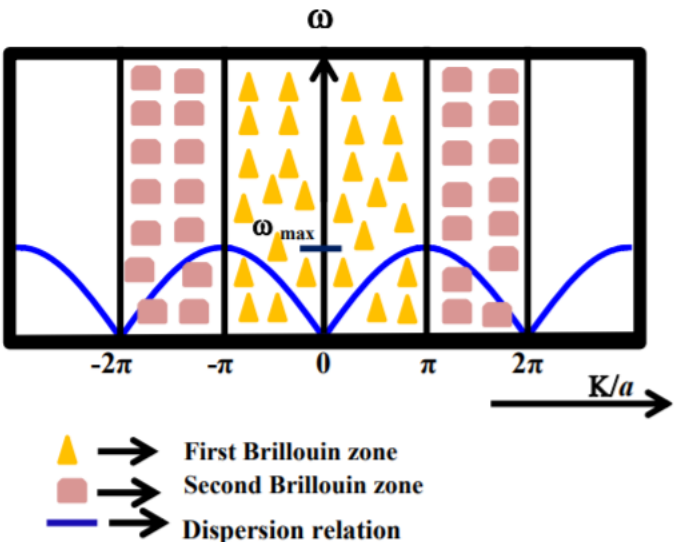


Fig.3.11 Dispersion relation for a 1D mono-atomic chain of atoms, showing periodic curve

Above dispersion curve clearly shows that for one value of ω there are several values of wave-vector K. Therefore we have defined the brillouin zones as –

First Brillouin zone:

$$-\pi < \frac{K}{a} < \pi$$

Second Brillouin Zone:

$$-2\pi < \frac{K}{a} < -\pi \quad \text{and} \quad \pi < \frac{K}{a} < 2\pi$$

Now in first brillouin zone each value of frequency corresponds to a unique value of wave vector. The above dispersion curve also indicates the mirror symmetry as $\omega(-K) = \omega(K)$. The mirror symmetry implies that +K,-K represents a plane wave propagating in the positive, negative direction through the mono-atomic lattice. Now we will examine some simple cases of this Dispersion relation, of which first is –

If the Frequency of oscillations is very low – This case is often regarded as Long wavelength limit also. In this limit K→0 which implies that,

$$\sin\left(\frac{Ka}{2}\right) \to \frac{Ka}{2}.$$

This reduces the Dispersion relation in Eqn.(6) to –

$$\omega = \omega_{\max}\left|\frac{Ka}{2}\right| \qquad \text{..................(7)}$$

$$\Rightarrow \qquad \omega = \frac{2v_s}{a}\frac{|K|a}{2} = v_s|K| \qquad \text{...................(8)}$$

Using the dispersion relation in above equation we can calculate the, Phase velocity v_p as,

$$v_p = \frac{\omega}{K} = v_s \quad \text{……………………..(9)}$$

Group velocity v_g as,

$$v_g = \frac{d\omega}{dK} = v_s \quad \text{……………(10)}$$

Above shows that in this limit, phase velocity and group velocity are equal to vs and the dispersion relation is linear. This is the case where discrete chain of atoms behaves as if it is continuous and approaches continuum. In fact, long wavelengths in the system do not respond to discreteness of the system. Here a very large number of atoms contribute to the displacements, which is very much similar to the case of homogeneous line. The system here follows the dynamical behavior, neglecting the effect of atomic nature of chain.

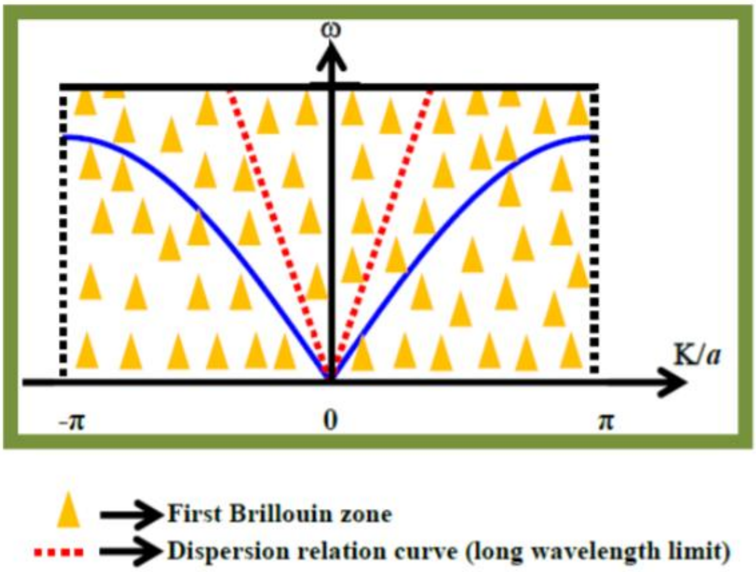


Fig.3.12. Dispersion relation for a 1D mono-atomic chain of atoms (low frequency case)

If the Frequency of oscillations is high – In the high frequency limit, the phase velocity and group velocity are no longer equal as compared to the previous case. Using the Eqn.(6) we obtain,

Phase velocity v_p as,

$$v_p = \frac{\omega}{K} = \frac{\omega_{\max}}{K}\left|\sin\left(\frac{Ka}{2}\right)\right| \quad \text{………………(11)}$$

Group velocity v_g as,

$$v_g = \frac{d\omega}{dK} = \omega_{\max}\frac{a}{2}\left|\cos\left(\frac{Ka}{2}\right)\right| \quad \text{……………………..(12)}$$

It is clear from the above equations that both group and phase velocities are a function of frequency ω_{max}. This kind of medium is highly dispersive in nature. In the previous case the medium was not dispersive and it follows the characteristics of a homogenous continuous medium. (Note: Dispersion refers to a phenomenon in which group/phase velocity of a wave travelling through a medium is dependent on its own frequency and such a medium is known as Dispersive medium.)

If the Frequency of oscillations is maximum– In this limit the atoms in monoatomic chain of vibrate with the maximum frequency which is –

$$\omega = \omega_{max} = 2\frac{v_s}{a} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(13)$$

Using above the group and phase velocities are calculated as,

Phase velocity v_p as,

$$v_p = \frac{\omega}{K} = 2\frac{v_s}{aK} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(14)$$

Group velocity v_g as,

$$v_g = \frac{d\omega}{dK} = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(15)$$

Zero group velocity refers to no propagation of energy or signal. In this condition only standing wave is produced as shown in Fig.(3.13).

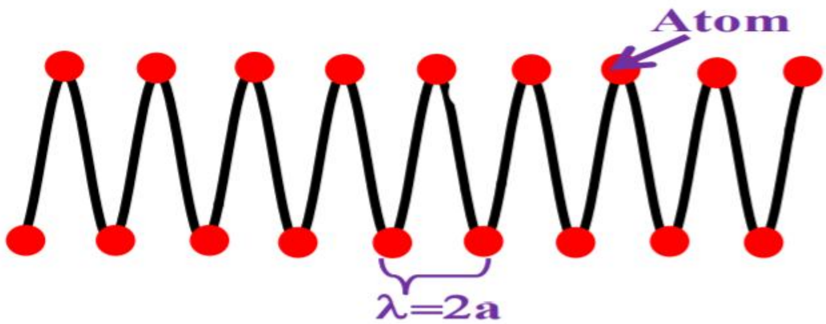


Fig.3.13. Standing wave produced in lattice when each atom vibrates with the same maximum frequency.

At $K=\pi/a$, wavelength is equal to 2a. This situation is parallel to Bragg's reflection (According to Bragg reflection of X-rays by the atomic planes in a crystal $n\lambda=2d\sin(\theta)$, where n is the order of reflection. Here for the 1st order reflection with normal incidence n=1, which implies that $\lambda=2d\sin(90)=2d$).

Dynamics of one dimensional finite monoatomic chain of atoms:

(a) In the previous case we considered the infinite length of the chain but now we want to examine the normal modes of vibration when the length of chain is finite. We assume that the length of chain is L and is fixed at both the ends as shown in Fig.(3.14) with zero displacement at the two fixed atoms.

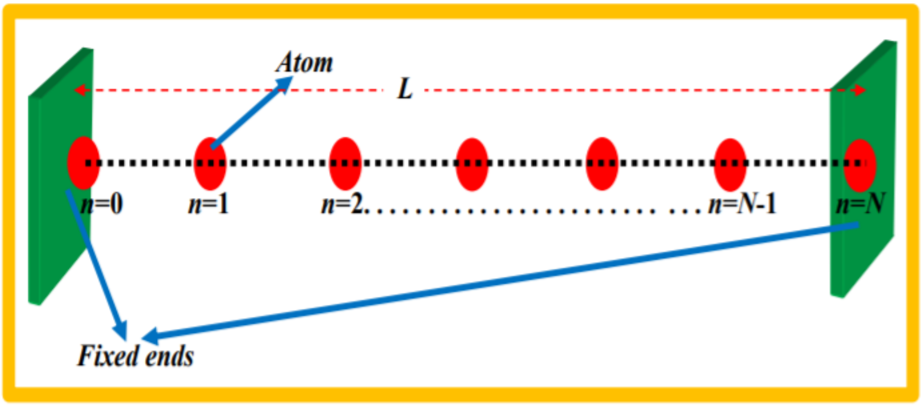


Fig.3.14 One dimensional finite mono-atomic chain of length L having N+1 atoms fixed at both the ends

We number the atoms in the chain in such a way that 0^{th} atom is fixed at the left end and N^{th} atom is fixed at right end of the chain. Therefore total number of atoms present in the lattice is N+1 with L=Na. We assume that (due to fixed boundary conditions a standing wave is produced in the lattice) the symbol un is given as –

$$u_n = u_0 \sin(Kna)\sin(\omega t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(16)$$

such that, $u_o = u_n = 0$ which ensures the normal mode of vibration of the lattice. For this we must have sin(Kna)=0 giving,

$$K = \frac{m\pi}{L}$$

Where m = 1,2,3,4,…….N-1

We have not included m=0, N values as they correspond to u_n=0 (zero displacement of the atom). This clearly indicates that in the N+1 number of atoms we have fixed the two atoms and thus there are N-1 normal modes of vibration where each mode corresponds to one atom which is free to move.

(b) Now we consider the case where the one dimensional mono-atomic chain is bent in form of a circular ring and 0th atom is joined with N th atom. In this situation when a vibrational mode is excited then both joined atoms suffer the same displacement and are free to move (shown in Fig.3.15), which is unlike the previous case where the two end atoms were fixed.

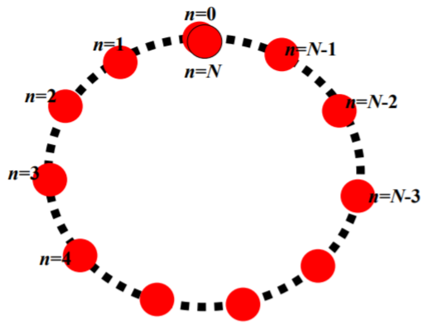


Fig.3.15 One dimensional finite mono-atomic chain of length L having N+1 atoms aligned along a circular ring with first and last atoms being superimposed on each other

The running wave solution for this case can be written as –

$$u_n = u_0 e^{i(\omega t - Kna)} \quad \text{……………………..(17)}$$

The boundary condition for this case is $u_n = u_{n+N}$, as the amplitude is same after every N number of atoms. To satisfy this boundary condition (Born and Von Karman cyclic boundary conditions) we must have –

$$\exp(-iKNa) = 1 \quad \text{……………….(18)}$$

Above implies that

$$K = 0, \pm\frac{2\pi}{Na}, \pm\frac{4\pi}{Na}, \pm\frac{6\pi}{Na}, ... \pm\frac{N\pi}{Na} \quad \text{…………………..(19)}$$

It means that there is total number of N independent K values which makes the frequency spectrum discrete. The dispersion relation curve is no longer a continuous curve for this case.

Dynamics of one dimensional diatomic linear chain of atoms:

Next we consider a linear diatomic chain of atoms as shown in Fig.9. We assume that it is a one dimensional lattice with two different kinds of atoms having masses M_1 and M_2 in one unit cell of the lattice. We also assume that the separation between any two consecutive M_1 and M_2 atoms is a and symbol u_r represents the displacement of r^{th} atom from its equilibrium position. In this case also we restrict the interactions between the atoms to nearest neighbors only within the consecutive atoms M_1 and M_2. In the Fig.9 all M_1 atoms are present at the even sites (…2n-4, 2n-2, 2n, 2n+2, 2n+4…..) and all M_2 atoms are present at the odd sites (…2n-5, 2n-3, 2n-1, 2n+1, 2n+3…..).

We can write the equations of motion for both the kind of atoms as –

$$F_{2n,M_1} = M_1\ddot{u}_{2n} = \eta\left(u_{2n+1} - 2u_{2n} + u_{2n-1}\right) \quad \ldots\ldots\ldots\ldots\ldots(20)$$

$$F_{2n+1,M_2} = M_2\ddot{u}_{2n+1} = \eta\left(u_{2n+2} - 2u_{2n+1} + u_{2n}\right) \quad \ldots\ldots\ldots\ldots\ldots(21)$$

In the above equations of motion, η is again the spring constant or force of interaction per unit displacement which is equal between any two consecutive atoms M_1 and M_2. If we assume that all the atoms vibrate with same frequency ω, then the solutions of Eqns.(20) and (21) can be written as –

$$u_{2n} = A_1 e^{i(\omega t - 2Kna)} \quad \ldots\ldots\ldots\ldots(22)$$

$$u_{2n+1} = A_2 e^{i(\omega t - (2n+1)Ka)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(23)$$

In the above running wave type solutions K represents a particular vibrational mode. An important point to note here is that we have taken the same frequency of oscillations for both the kinds of atoms, irrespective of their different masses. This makes them oscillate with different amplitudes. On substituting Eqns.(22) and (23) in Eqns.(20) and (21) we get –

$$-M_1\omega^2 A_1 = \eta A_2\left(e^{-iKa} + e^{iKa}\right) - 2\eta A_1 \quad \ldots\ldots\ldots\ldots\ldots\ldots(24)$$

$$-M_2\omega^2 A_2 = \eta A_1\left(e^{-iKa} + e^{iKa}\right) - 2\eta A_2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(25)$$

$$\Rightarrow \quad \left(2\eta - \omega^2 M_1\right)A_1 - \left(2\eta\cos(Ka)\right)A_2 = 0 \quad \ldots\ldots\ldots\ldots\ldots(26)$$

$$\left(-2\eta\cos(Ka)\right)A_1 + \left(2\eta - \omega^2 M_2\right)A_2 = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(27)$$

The above system of equations is known as "constant coefficient linear homogeneous system".

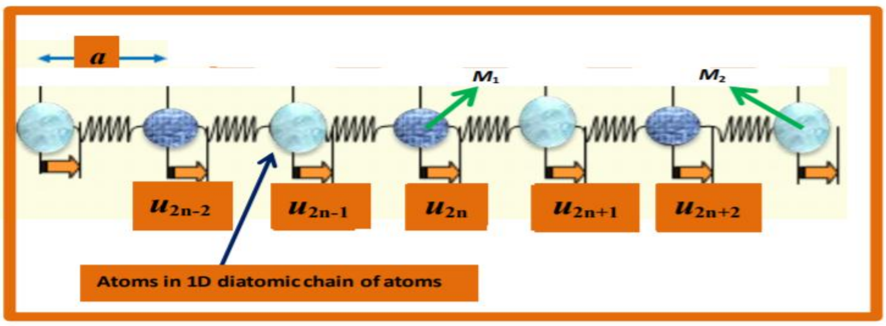


Fig.3.16. The model of a diatomic linear lattice showing the elastic coupling between the neighboring atoms having two kinds of atoms with masses M_1 and M_2.

Now we put the determinant of the system in equations (26) and (27) equal to zero to get the solution of this system as –

$$\begin{vmatrix} 2\beta-\omega^2 M_1 & -2\beta\cos(Ka) \\ -2\beta\cos(Ka) & 2\beta-\omega^2 M_2 \end{vmatrix} = 0 \quad \text{……………………..(28)}$$

On solving the above determinant the dispersion relation for the diatomic linear chain of atoms is obtained as follows –

$$\omega^2 = \eta\left(\frac{1}{M_1}+\frac{1}{M_2}\right) \pm \eta\sqrt{\left(\frac{1}{M_1}+\frac{1}{M_2}\right)^2 - \frac{4\sin^2(Ka)}{M_1 M_2}} \quad \text{……………….(29)}$$

It is clear from the above dispersion relation that even if we have single value of wave-vector, it corresponds to two different values of ω as –

$$\omega_+ = \left(\eta\left(\frac{1}{M_1}+\frac{1}{M_2}\right) + \eta\sqrt{\left(\frac{1}{M_1}+\frac{1}{M_2}\right)^2 - \frac{4\sin^2(Ka)}{M_1 M_2}}\right)^{1/2} \Rightarrow \textbf{Optical Branch}$$

$$\omega_- = \left(\eta\left(\frac{1}{M_1}+\frac{1}{M_2}\right) - \eta\sqrt{\left(\frac{1}{M_1}+\frac{1}{M_2}\right)^2 - \frac{4\sin^2(Ka)}{M_1 M_2}}\right)^{1/2} \Rightarrow \textbf{Acoustical Branch}$$

Using the above frequencies we can plot the Dispersion relation for the diatomic linear chain of atoms as shown in Fig.3.17 below. It is clear from the Fig.3.17 above that the only range of frequencies that can be excited in a diatomic linear chain of atoms gets broken up into two branches, where top branch is called optical and lower one is called acoustical branch. It means that for each value of K there are two types of vibration in which a lattice can go or we can say that there are two different modes for each K. Also in between these two branches there exists a band gap of frequencies corresponding to no lattice vibrations. It means that this range of frequencies cannot be excited in a diatomic linear chain of atoms.

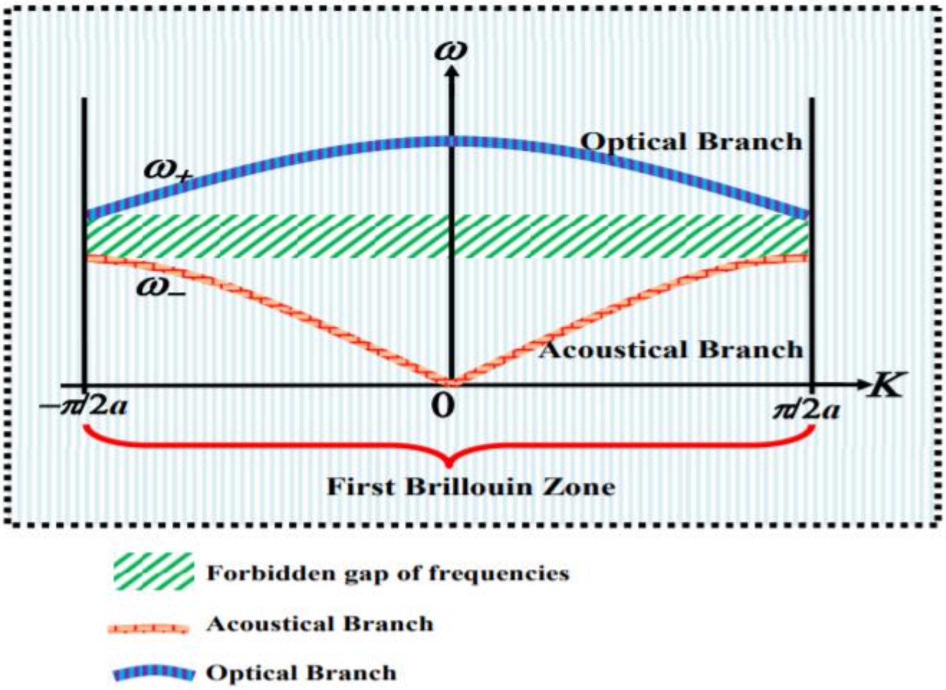


Fig.3.17 The optical and acoustical branches and forbidden frequency band gap in the frequency spectrum for a linear diatomic chain of atoms (for the first brillouin zone only).

Optical Branch:

In this case, the atoms undergo a lattice vibration such that both the kinds of atoms move in the opposite directions as shown in Fig.3.18.

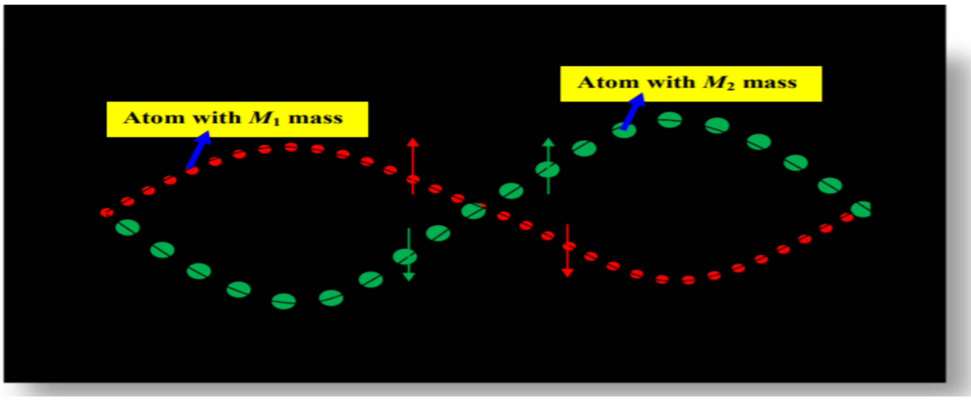


Fig.3.18. Motion of atoms (neighbouring atoms out of phase) in a diatomic linear chain of atoms when the optical branch vibrations are excited.

To excite these kinds of vibrations a force is needed such that, it incorporates the opposite motions on two kinds of atoms such that the center of mass (of the unit cell) is at rest with the amplitude of vibration of atoms being inversely proportional to their masses. Using the expression of ω_+ we get the optical branch frequency as –

$$\lim_{K\to 0}\omega_+ = \left(2\eta\left(\frac{1}{M_1}+\frac{1}{M_2}\right)\right)^{1/2} \quad \text{……………………(30)}$$

$$\lim_{K\to \pm\frac{\pi}{2a}}\omega_+ = \left(\frac{2\eta}{M_1}\right)^{1/2} \quad \text{………………………(31)}$$

where we have assumed M_1 to be less than M_2. Hence in the Fig. above the maximum value of ω_+ is given at the point where K=0 and minimum at the K=±π/2a at the first brillouin zone boundary. Using the Eqns.(26) and (27) at K=0, we get –

$$\left(2\eta-\omega^2 M_1\right)A_1-\left(2\eta\right)A_2=0 \quad \text{……………………(32)}$$

$$\left(-2\eta\right)A_1+\left(2\eta-\omega^2 M_2\right)A_2=0 \quad \text{………………..(33)}$$

On solving the above we get,

$$\frac{A_1}{A_2}=-\frac{M_2}{M_1}$$

which also shows that atoms move in opposite directions.

Acoustical Branch: In this case, the atoms undergo a lattice vibration such that both the kinds of atoms move in the same direction with the same amplitude irrespective of their different masses as shown in Fig. 3.19.

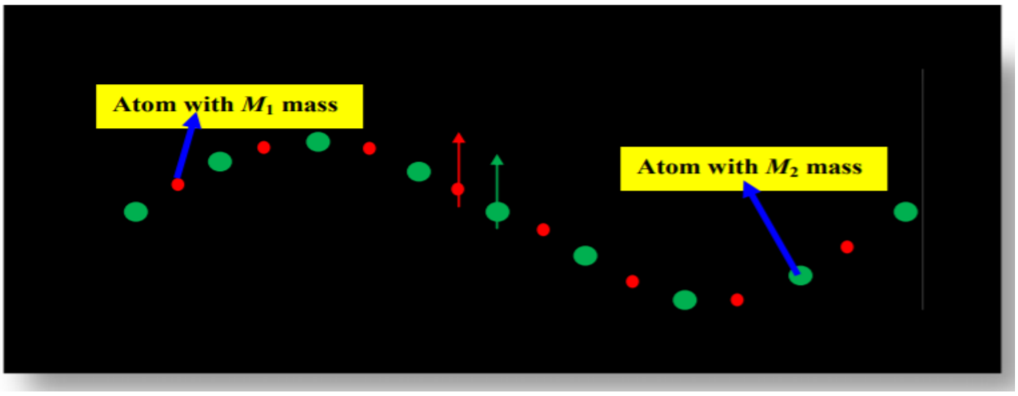


Fig.3.19 Motion of atoms (neighbouring atoms in phase) in a diatomic linear chain of atoms when the acoustical branch vibrations are excited

The motion of the center of mass is also in the same direction. To excite these kinds of vibrations a force is needed such that, it incorporates the motions in two kinds of atoms in the same direction with equal amplitudes such that the center of mass (of the unit cell) also moves in same direction. Using the expression of ω- we get the optical branch frequency as –

$$\lim_{K \to 0} \omega_- = 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(34)$$

$$\lim_{K \to \pm\frac{\pi}{2a}} \omega_- = \left(\frac{2\eta}{M_2}\right)^{1/2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots..(35)$$

Hence in the Fig. above the minimum value of ω- is given at the point where K=0 and maximum at the K=±π/2a at the first brillouin zone boundary. Using the Eqns.(31) and (35) we can calculate the width of the forbidden band gap of frequencies in Fig.[]. This gap depends on the ratio of two masses as –

$$\begin{array}{llll} M_2 / M_1 & \rightarrow \text{ Increases} & \Rightarrow & \text{Forbidden frequency bandgap also increases} \\ & \rightarrow \text{ Decreases} & \Rightarrow & \text{Forbidden frequency bandgap also decreases} \\ & \rightarrow \text{ 1} & \Rightarrow & \text{Forbidden frequency bandgap disappear} \\ & & & \text{(The two spilt branches join at } K = \pm\frac{\pi}{2a}\text{)} \end{array}$$

At θ =0, we can replace sin(θ) by θ only, therefore in the limit of K→0, we can also substitute,

$$\omega_- = \left(\eta\left(\frac{1}{M_1}+\frac{1}{M_2}\right) - \eta\sqrt{\left(\frac{1}{M_1}+\frac{1}{M_2}\right)^2 - \frac{4}{M_1 M_2}K^2a^2}\right)^{1/2} \qquad \ldots\ldots\ldots\ldots\ldots..(36)$$

$$\Rightarrow \qquad \omega_- = Ka\sqrt{\frac{2\eta}{M_1+M_2}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(37)$$

Substituting the

$$\cos(Ka) \simeq \left(1-\sin^2(Ka)\right)^{1/2} \simeq \left(1-(Ka)^2\right)^{1/2} \simeq 1-\frac{(Ka)^2}{2}$$

and using Eqns.(37), (26) and (27) we get –

$$\left(2\eta-\left[Ka\sqrt{\frac{2\eta}{M_1+M_2}}\right]^2 M_1\right)A_1-2\eta\left(1-\frac{(Ka)^2}{2}\right)A_2=0 \qquad \text{……………………….(38)}$$

$$-2\eta\left(1-\frac{(Ka)^2}{2}\right)A_1+\left(2\eta-\left[Ka\sqrt{\frac{2\eta}{M_1+M_2}}\right]^2 M_2\right)A_2=0 \qquad \text{………………..(39)}$$

On solving the above we get,

$$\frac{A_1}{A_2}=+1$$

which also shows that atoms move in same directions with equal amplitudes.

- If $M_2 \rightarrow \infty$, then acoustical branch frequency becomes single valued at zero and optical branch frequency at $\sqrt{2\eta/M_1}$

- If $M_1 \rightarrow 0$, the diatomic case reduces to mono-atomic case with lattice constant being 2a. Here optical branch does not appear without affecting the acoustical branch.
- If $M_1=M_2$ then lattice vibration frequency range lies between 0 and $\sqrt{4\eta/M_1}$. The only difference then between monoatomic and diatomic case is that, for monoatomic case whole range of this frequency accounts for acoustical branch and for diatomic case this range of frequency splits into two branches acoustical one corresponding to frequency from 0 to $\sqrt{2\eta/M_1}$ and $\sqrt{2\eta/M_1}$ to $\sqrt{4\eta/M_1}$ corresponding to optical branch.
- For a crystal with N number of atoms per unit cell, the frequency range will split up into N number of bands.
- For a fixed length L the periodic boundary condition is, $u_{2n}(x) = u_{2n}(x+L)$ which ensures

$$K=\pm\frac{\pi}{L},\pm\frac{2\pi}{L},\pm\frac{3\pi}{L},\ldots.\pm\frac{N\pi}{2L}$$

Here N is the total number of allowed K values representing N normal modes for vibration.

Concept of Phonons:

The concept of photons is well known. A photon is basically a quantum of light (or all forms of EM radiations). The energy of a single photon is given by ħω where ω is the photon frequency. If we have an EM radiation with a particular mode (having photons of single frequency) then its energy can be expressed as n× ħω where n is the number of photons. Similar is the concept of phonons. Basically a phonon is a quantum of vibrational motion. Whenever a lattice (all atoms in a lattice) vibrates with single frequency, then the energy of this particular mode is given by

$$E_{LatticeVibration} = n \times \hbar\omega \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (40)$$

Here n is the number of Phonons and ω is a single frequency at which all the atoms in lattice vibrate. This is known as the quantization of lattice vibrations. A particular mode of lattice vibrations represents an average number of phonons given by –

$$n_{AVERAGE} = \left(e^{\hbar\omega / K_B T} - 1\right)^{-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots..(41)$$

A phonon is sometimes referred to as a quasi particle also. Phonons also follow the Bose Einstein distribution function. There could be different types of phonons as shown in Fig.3.20 below.

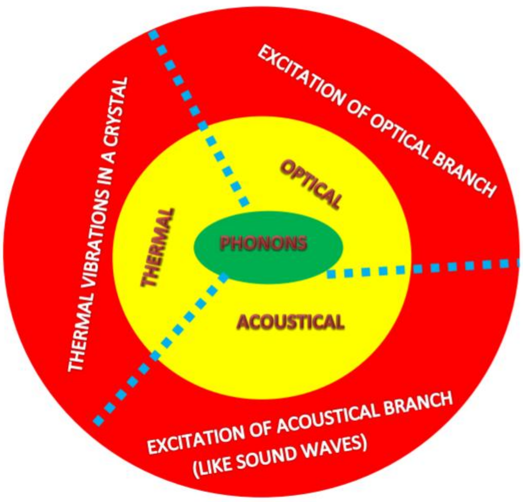


Fig.3.20 Three types of phonons optical, acoustical and thermal

The change of energy in the lattice vibrations is also quantized, for example whenever the energy of a lattice vibration is increased or decreased it has to be –

$$\Delta E_{LatticeVibration} = \pm\hbar\omega \ , \quad \Delta n = \pm 1$$
………………(42)

Because we assume that up to the first order approximation only, the energy of lattice vibration changes by the gain or loss of one phonon only. Similar to the photons, phonons also have momentum related to them known as the Phonon Momentum. For the phonons also we can use the De-Broglie relation as –

$$p = \hbar K$$
……………………(43)

Where, K is the wave vector for a particular phonon.

3.6 Thermal capacity of solids:

Heat capacity or thermal capacity is a physical property of matter, defined as the amount of heat to be supplied to an object to produce a unit change in its temperature. The SI unit of heat capacity is joule per Kelvin (J/K).

Heat capacity is an extensive property. The corresponding intensive property is the specific heat capacity, found by dividing the heat capacity of an object by its mass. Dividing the heat capacity by the amount of substance in moles yields its molar heat capacity. The volumetric heat capacity measures the heat capacity per volume. In architecture and civil engineering, the heat capacity of a building is often referred to as its thermal mass.

Basic definition:

The heat capacity of an object, denoted by C, is the limit

$$C = \lim_{\Delta T \to 0} \frac{\Delta Q}{\Delta T}$$

Where ΔQ is the amount of heat that must be added to the object (of mass M) in order to raise its temperature by ΔT.

The value of this parameter usually varies considerably depending on the starting temperature T of the object and the pressure P applied to it. In particular, it typically varies dramatically with phase transitions such as melting or vaporization (see enthalpy of fusion and enthalpy of vaporization). Therefore, it should be considered a function C(P,T) of those two variables.

Classical theory of specific heats:

This theory is based on the assumption that atoms in a periodic lattice vibrate about their equilibrium sites and obey Hookes' law. That is, each atom is bound to its equilibrium site by a harmonic force and when thermal energy is supplied, it vibrates like an harmonic oscillator. It is further assumed that each atom is equivalent to a classical three-dimensional (3-D) harmonic oscillator and vibrates independently. This means that a solid consisting of N atoms can be treated as a collection of as many three-dimensional, independent harmonic oscillators. Equivalently, N vibrating atoms can be considered as a collection of 3N, one-dimensional (I -D) independent harmonic oscillators. And; to obtain an expression for the heat capacity of a solid, we need to know the average energy of a 1-D harmonic oscillator and multiply it by 3N. Therefore, our problem reduces to obtaining an expression for the average energy of a I -D harmonic oscillator. The total energy of a 1-D harmonic oscillator is written as

$$\varepsilon = \text{Kinetic energy} + \text{Potential energy}$$

$$= \frac{p^2}{2m} + \frac{1}{2} m\omega^2 x^2$$

Where m, p, ω and x respectively denote mass, momentum, angular frequency and displacement from equilibrium position of an atom. According to classical mechanics, energy of an oscillator can take any value between zero and infinity. To calculate the average energy: we note that energy distribution of a classical system is governed by M-B distribution. The M-B distribution function, which gives the probability that an oscillator will have energy ε, is given by

$$f(\varepsilon) = A \exp\left[-\frac{\varepsilon}{k_B T}\right]$$

Where A is an arbitrary constant and k_B is Boltzmann constant. Thus, the average energy of a 1-D oscillator will be given by

$$<\varepsilon> = \frac{\int_0^\infty \varepsilon\, f(\varepsilon)\, d\varepsilon}{\int_0^\infty f(\varepsilon)\, d\varepsilon} = k_B T$$

Since a solid having N atoms is equivalent to 3N one-dimensional harmonic oscillators, its total energy is given by

$$E = 3\, N\, k_B T$$

and the heat capacity is given by

$$C_v = \left(\frac{\partial E}{\partial T}\right)_v = 3Nk_B$$

If we consider one mole of a substance, the total number of atoms will be equal to the Average number, N_A, and the heat capacity per mole or the molar heat capacity will be equal to

$$C_v = 3\, N_A k_B = 3R \text{ mole}^{-1}$$

Where R = $N_A k_B$ is the universal gas constant. Let us pause for a while and try to understand the significance of this result. It shows that heat capacity of solids is constant, equal to 3R and independent of temperature. This explains Dulong-Petit law, which holds rather well at room temperature and above. However, as it became possible to carry out measurements at lower temperatures, it was observed that heat capacity decreases with temperature for almost all solids. Physically, it implies that as temperature decreases, the atomic motion is constrained and at absolute zero, all motion should cease. The classical theory, which predicts temperature independent heat capacity, fails to explain why heat capacity tends to zero as temperature approaches absolute zero and is not consistent with the third Law of thermodynamics.

Dulong and Petit's law:

In early 19th century, Dulong and Petit conducted a series of classical experiments to measure heat capacities of solids at different temperatures. Fig.3.21a shows a typical temperature versus heat capacity curve obtained by Dulong and Petit. Fig.3.21b shows the temperature dependence of heat capacity of silver and germanium, respectively a metal and a semiconductor.

Note that:

At high temperatures, heat capacity is independent of temperature and is equal to 3 R for all solids, where R is the universal gas constant. Its value is 8.3 JK^{-1}. This is known as Dulong-Petit law.

As temperature decreases, heat capacity decreases almost exponentially and approaches zero as $T \rightarrow 0K$

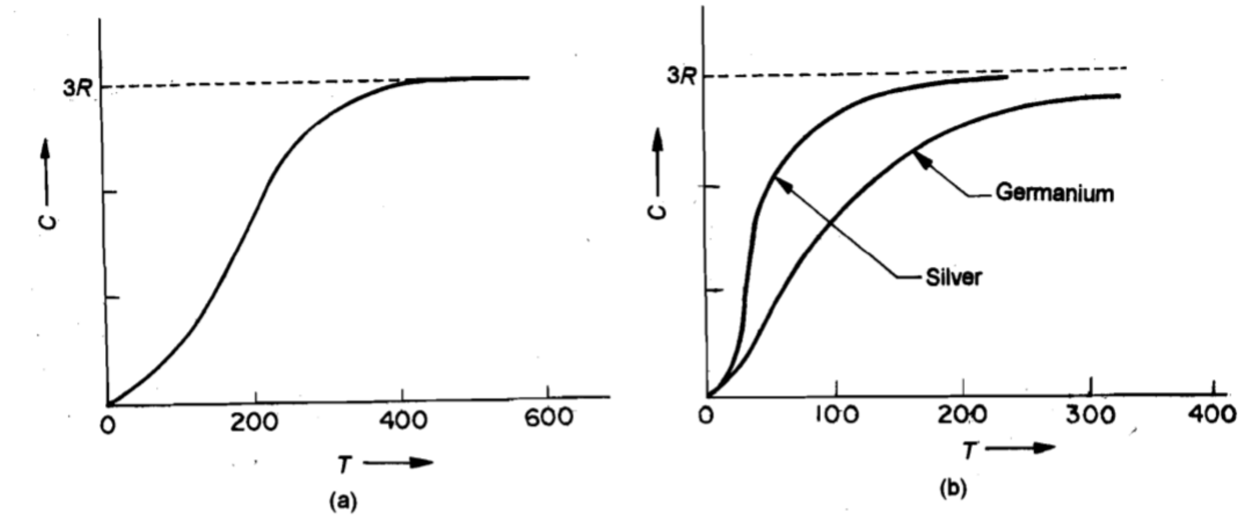


Fig.3.21 a) Typical temperature dependence of heat capacity of solids; and b) heat capacity versus temperature curves for silver and germanium

Now, you may like to know: What is the theoretical basis for the observed temperature dependence of heat capacity? Let us first understand the observations qualitatively. You know that a solid is a collection of atoms arranged along a periodic lattice. Since we confine our discussion to heat capacity at constant volume, it is logical to think that the (heat) energy supplied to a solid will be completely used in increasing its internal energy. This can happen via two mechanisms. Firstly, atoms can oscillate about their equilibrium positions much more vigorously by absorbing heat energy. Secondly, in metals and semiconductors, the heat energy can also be absorbed by electrons in outer orbits and be excited to higher energy states. Corresponding to these mechanisms of absorption of heat energy, we have lattice and electronic contributions to the heat capacity. In this unit, however, we limit our discussion to lattice heat capacity.

3.7 Need of quantum theory of specific heat of solids:

Einstein developed the quantum theory of heat capacity and proposed that atoms in a solid should be treated as Planck oscillators characterised by a constant frequency. This implies that energy (E,) of a quantum oscillator is nhv, where n is a positive integer, h is Planck constant and v is frequency of oscillation. Note that Einstein assumed that all oscillators vibrate with the same frequency.

According to Einstien's theory a solid made of N atoms is equivalent to 3N, 1-D independent harmonic oscillators. To obtain an expression for the heat capacity, Einstien assumed that the probability of an oscillator possessing energy nhv, when the system is at temperature T, is proportional to

$$\exp\left[-\frac{nh\nu}{k_B T}\right]$$

That is, Einstein also used M-B distribution law probably because the oscillators are distinguishable from heat capacity as temperature each other due to their location at different lattice sites. Thus, in Einstein's theory, we decrease. Analyse the statistical behaviour of quantum oscillators using classical (M-B) distribution law. Since the energy is quantised, we obtain the average energy of an oscillator by summation, rather than by integration (as we did in classical theory). Under these conditions, we may write the average energy of a quantum harmonic oscillator as:

$$<\varepsilon> = \frac{\sum_{n=0}^{\infty} \varepsilon_n \exp\left[-\frac{\varepsilon_n}{k_B T}\right]}{\sum_{n=0}^{\infty} \exp\left[-\frac{\varepsilon_n}{k_B T}\right]}$$

Substituting ε_n = nhv in this expression, we get

$$<\varepsilon> = \frac{h\nu \sum_{n=0}^{\infty} n \exp\left[-\frac{nh\nu}{k_B T}\right]}{\sum_{n=0}^{\infty} \exp\left[-\frac{nh\nu}{k_B T}\right]} \qquad \text{………………(1)}$$

To evaluate the RHS of Eq. **(1),** we introduce a change of variable and write

$$\frac{h\nu}{k_B T} = x.$$

Then, expression for average energy simplifies to

$$<\varepsilon> = \frac{h\nu \sum_{n=0}^{\infty} n e^{-nx}}{\sum_{n=0}^{\infty} e^{-nx}} = \frac{h\nu\left[e^{-x} + 2e^{-2x} + 3e^{-3x} + \dots\right]}{\left[1 + e^{-x} + e^{-2x} + e^{-3x} + \dots\right]}$$

$$= \frac{h\nu N}{D} \qquad \text{……………….(2)}$$

Now let us first consider the expression in the numerator:

$$N = e^{-x} + 2e^{-2x} + 3e^{-3x} + \dots \qquad \text{………………..(3a)}$$

Multiply by e^x on both sides. This gives

$$e^x N = 1 + 2e^{-x} + 3e^{-2x} + \ldots$$
………………..(3b)

On subtracting the expression in Eq. (3a) from that in Eq. (3b), we get

$$\left(e^x - 1\right)N = 1 + e^{-x} + e^{-2x} + \ldots$$
…………(3c)

The right hand side in this expression is an infinite geometric series with common ratio e^{-x} and is simply identical to the denominator in Eq. (2). Hence

$$\frac{N}{D} = \frac{1}{e^x - 1}$$
…………….(4)

and the expression for average energy of an Einstenian harmonic oscillator simplifies to

$$< \varepsilon > = \frac{h\nu}{e^x - 1} = \frac{h\nu}{\exp\left[\frac{h\nu}{k_B T}\right] - 1}$$
………………..(5)

If a solid consists of 3N one-dimensional oscillators, the expression for its total energy is obtained by multiplying Eq. (5) by 3N

$$E = \frac{3Nh\nu}{\exp\left[\frac{h\nu}{k_B T}\right] - 1}$$
……………..(6)

so that the heat capacity is given by

$$C_v = 3Nk_B \left(\frac{h\nu}{k_B T}\right)^2 \frac{\exp\left[\frac{h\nu}{k_B T}\right]}{\left(\exp\left[\frac{h\nu}{k_B T}\right] - 1\right)^2}$$
………………….(7)

If we take one mole of the solid, N is replaced by N_A in this expression to obtain molar heat capacity:

$$C_v = 3R \left(\frac{h\nu}{k_B T}\right)^2 \frac{\exp\left[\frac{h\nu}{k_B T}\right]}{\left(\exp\left[\frac{h\nu}{k_B T}\right] - 1\right)^2}$$
……………….(8)

This result shows that heat capacity is a function of temperature - a feature missing in the classical theory. Note that the key to success of Einstein's theory is the concept of quantisation of energy.

To compare the predictions of this theory with experimental results, we need to know the frequency of the oscillators so that the theoretical value of C_v at a given temperature can be obtained using Eq. (8). The question is: How to estimate the value of frequency? This difficulty is overcome by choosing such a value of frequency for which Eq. (8) fits the experimental curve. That is, it is assumed that each solid has a characteristic frequency, called Einstein frequency, VE. It is customary to write

$$h\nu_E = k_B T_E \qquad \text{..................(9)}$$

Where T_E denotes the Einstein temperature. Like the characteristic frequency, T_E will have a fixed value for a given solid. In terms of Einstein temperature, the expression in Eq. (8) for heat capacity takes the form

$$C_v = 3R\left(\frac{T_E}{T}\right)^2 \frac{\exp\left[\frac{T_E}{T}\right]}{\left(\exp\left[\frac{T_E}{T}\right]-1\right)^2} \qquad \text{..............................(10)}$$

Note that C_v is a function of (T_E/T). You may now like to investigate the nature of temperature dependence. For copper, a plot of Eq. (10) is shown in Fig.3.22. You will note that this relation reproduces all the general features of the observed curve. Now, let us consider the nature of temperature dependence in two limiting cases: at high temperatures ($T >> T_E$) and at low temperatures ($T << T_E$).

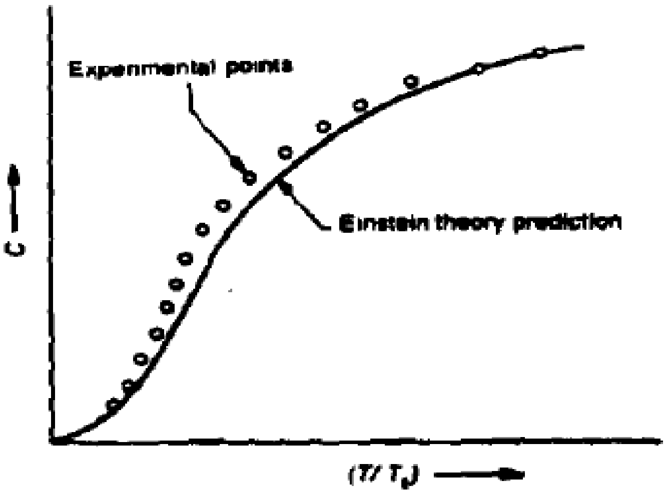


Fig.3.22 Heat capacity versus temperature plot for copper at low temperatures as predicted by Einstein's theory and as obtained experimentally.

High temperature limit ($T>>T_E$):

In this case, (T_E/ T) << 1 and we can write the exponential term as

$$\exp\left[\frac{T_E}{T}\right] = 1 + \frac{T_E}{T} + \frac{1}{2!}\left(\frac{T_E}{T}\right)^2 + \ \ldots$$

Using this expansion in Eq. (10), we get

$$C_v = 3R\left(\frac{T_E}{T}\right)^2 \frac{\left(1 + \frac{T_E}{T} + \frac{1}{2!}\left(\frac{T_E}{T}\right)^2 + \ldots\right)}{\left[\frac{T_E}{T} + \frac{1}{2!}\left(\frac{T_E}{T}\right)^2 + \frac{1}{3!}\left(\frac{T_E}{T}\right)^3 + \ldots\right]^2}$$

You can readily simplify this expression to obtain

$$C_v = 3R\left[1 - \frac{1}{12}\left(\frac{T_E}{T}\right)^2 + \ldots\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

For many materials, molar heat capacity at $T_E \approx 200$ K arid high temperatures will approach 3R. From this we can conclude that Einstein's theory provides sound basis for validity of Dulong-Petit law.

Low temperature limit (T << T_E):

For **(T_E/** T) >> I, we can ignore unity in comparison to the exponential term in the denominator of Eq. (10) and write:

$$C_v = 3R\left(\frac{T_E}{T}\right)^2 \exp\left[-\frac{T_E}{T}\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots(12)$$

You may now ask: What does Eq. (12) signify physically in relation to temperature dependence of heat capacity? We know that exponential function changes more rapidly than any power of x. Therefore, Einstein theory predicts that heat capacity at low temperatures will decay exponentially and approach zero as temperature tends to absolute zero. This prediction of Einstein led to a lot of experimental activity on measurement of heat capacity at low temperatures and results for several solids were reported. Again refer to Fig. which depicts the temperature variation of heat capacity of copper as predicted by Einstein's theory along with experimentally observed curve. Note that, at low temperatures, the exponential decrease predicted by Einstein is much faster compared to experimental result. We may, therefore, conclude that Einstein's theory is qualitative, particularly at low temperatures, as it fails to explain experimental observations near absolute zero for most solids.

SAQ.2

a) What do you mean by acoustic, optical modes and phonon?
b) Define the Thermal capacity of solids.
c) Explain in brief Dulong and Petit’s law.
d) What is need of quantum theory of specific heat of solids?
e) A body with mass 3 kg absorbs heat 150 calories when its temperature raises from 30^{o}C to 80^{o}C. What is the specific heat of the body?

3.8 Einstein’s theory of specific heats:

- The crystal lattice structure of solid comprising N atoms can be treated as an assembly of 3N distinguishable one-dimensional oscillators!
- The assumption is based on that each atom is free to move in three dimensions!
- the internal energy for N linear oscillators is U= Nkθ(1/2 + 1/(e$^{\theta/T}$ -1)) with θ = hv/k
- The internal energy of a solid is thus

$$U = 3Nk\theta_E\left(\frac{1}{2} + \frac{1}{e^{\frac{\theta_E}{T}} - 1}\right)$$

- Here θ is the Einstein temperature and can be replaced by θ_E.
- The heat capacity:

$$C_v = \left(\frac{\partial U}{\partial T}\right)_v = \frac{\partial(3Nk\theta_E \frac{1}{2} + \frac{3Nk\theta_E}{e^{\frac{\theta_E}{T}} - 1})}{\partial T}$$

$$= 0 + \left(\frac{-3Nk\theta}{(e^{\frac{\theta}{T}} - 1)^2} \cdot e^{\frac{\theta}{T}} \cdot -\frac{\theta}{T^2} \right)$$

$$C_v = \frac{3Nk\theta^2}{T^2} \cdot \frac{e^{\frac{\theta}{T}}}{(e^{\frac{\theta}{T}} - 1)^2} = 3Nk(\frac{\theta}{T})^2 \cdot \frac{e^{\frac{\theta}{T}}}{(e^{\frac{\theta}{T}} - 1)^2}$$

Case 1: when $T >> \theta_E$

$$\frac{\theta}{T} \ll 1$$

$$\therefore e^{\frac{\theta}{T}} \rightarrow 1$$

$$e^{\frac{\theta}{T}} - 1 \approx 1 + \frac{\theta}{T} + \frac{1}{2!}\left(\frac{\theta}{T}\right)^2 + \cdots - 1 \approx \frac{\theta}{T}$$

$$C_v = 3Nk \left(\frac{\theta}{T}\right)^2 \frac{1}{(\frac{\theta}{T})^2} = 3Nk$$

This result is the same as Dulong & Petit's

Case 2: $T << \theta_E$

$$\therefore \frac{\theta}{T} \gg 1$$

$$\therefore e^{\frac{\theta}{T}} - 1 \approx e^{\frac{\theta}{T}}$$

$$\therefore C_v = 3Nk \left(\frac{\theta}{T}\right)^2 \frac{1}{e^{\frac{\theta}{T}}}$$

As discussed earlier, the increase of θ/T is out powered by the increase of $e^{(\theta/T)}$

As a result, when $T \rightarrow 0$, $C_v \rightarrow 0$

If an element has a large θ_E , the ratio θ/T will be large even for temperatures well above absolute zero

When θ/T is large, C_v is small

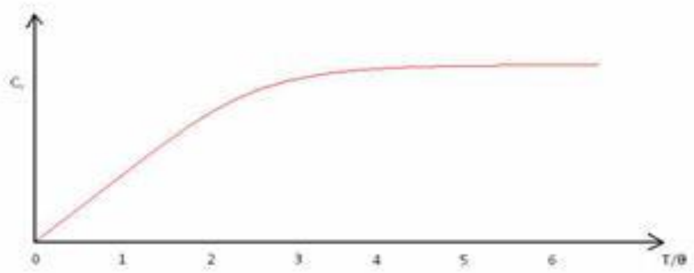


Fig.3.23 Curve of Temperature v/s Heat capacity

Since

A large θ_E value means a bigger v

On the other hand

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{u}}$$

$$\theta_E = \frac{hv}{k}$$

To achieve a larger v, we need a large k or a small *u* (reduced mass), which corresponds to lighter element and elements that produce very hard crystals.

- The essential behavior of the specific heat capacity of solid is incorporated in the ratio of θ_E/T.

- For example, the heat capacity of diamond approaches 3Nk only at extremely high temperatures as θ_E = 1450 k for diamond.

- Different elements at different temperatures will pose the same specific heat capacity if the ratio θ_E/T is the same.

- Careful measurements of heat capacity show that Einstein's model gives results which are slightly below experimental values in the transition range of θ/T

Einstein's derivation of specific heat formula is based upon the following assumptions:

- All atoms of a monatomic solid vibrate with the same frequency *v*.
- The frequency depends on the mass of the atom and the restoring force.
- These atoms, like Planck's oscillators, are in equilibrium with ether vibrations of the same frequency.
- Instead of classical law of equipartition, that is ½KT for each degree of freedom, Einstein uses the mean energy of Planck's oscillator for each degree of freedom, that is,

$$\overline{U} = \frac{hv}{e^{\frac{hv}{kT}} - 1}$$

Limitations of Einstein theory of specific heat:

- Einstein's theory of specific heat of Solid couldn't explain the experimental results obtained at very low temperatures

- From the experiment it is observed that the specific heat of solids has a T^3 dependence on the absolute temperature of the solid. But according to Einstein's theory there is an Exponential dependence on the absolute temperature.

3.9 Debye theory of specific heats:

In thermodynamics and solid state physics, the Debye model is a method developed by Peter Debye in 1912 for estimating the phonon contribution to the specific heat (heat capacity) in a solid. This model correctly explains the low temperature dependence of the heat capacity, which is proportional to T^3. It also recovers the Dulong-Petit law at high temperatures. But due to simplifying assumptions, its accuracy suffers at intermediate temperatures.

In 1912 Debye realized that, inconsistent with the Einstein model, low-energetic excitations of a solid material were not oscillations of a single atom, but collective modes propagating through the material. Such vibrations can be considered to be sound waves, and their propagation speed is the speed of sound in the material. Moreover, these modes only accept energy in discreet amounts.

Quantum theory uses the concepts of phonons, which are "quasi-particles" with definite energies and directions of motion, to treat the vibrations. The concept of phonon is analogous with photons of the electromagnetic wave. The relations between the energy of a phonon ε, the angular frequency ω and the wave vector $\vec{q}$ are:

$$\varepsilon = \hbar\omega \qquad \text{...............(1)}$$

$$\omega = v_s|\vec{q}| \qquad \text{...........(2)}$$

where v_s is the velocity of the sound wave.

As a kind of Bosons, phonons obey Bose–Einstein statistics. The expectation number of bosons in a state with energy E is

$$n_{(E)} = \frac{1}{e^{E/k_BT} - 1} = \frac{1}{e^{\hbar\omega/k_BT} - 1} \qquad \text{...............(3)}$$

where k_B= 1.380 6504(24)×10^{-23}J/K is the Boltzmann constant.

Debye frequency and Debye Temperature:

Unlike electromagnetic radiation in a box, a phonon cannot have infinite frequency. Its frequency is bound by the medium of its propagation — the atomic lattice of the solid. If there are N primitive cells in the specimen, the total number of phonon modes are N. A cut-off frequency ω_D, known as Debye frequency, is determined by the following

In the 3 dimensional reciprocal space, the volume for each allowed wave vector $\vec{q}$ is

$$(\frac{2\pi}{L})^3 = \frac{8\pi^3}{V} \quad \text{………………………….(4)}$$

Since there is a cut-off wave vector $qD=\omega D/\upsilon s$, all the modes are confined within a sphere with radius q_D. Thus number of modes (not number of phonons) should be

$$N = (\frac{4}{3}\pi {q_D}^3)/(\frac{8\pi^3}{V}) \quad \text{……………..(5)}$$

Or

$$q_D = (6\pi^2 \frac{N}{V})^{\frac{1}{3}} \quad \text{……………(6)}$$

$$\omega_D = \upsilon_s(6\pi^2 \frac{N}{V})^{\frac{1}{3}} \quad \text{………………..(7)}$$

Debye temperature T_D is defined as

$$T_D = \frac{\hbar\omega_D}{k_B} = \frac{\hbar\upsilon_s}{k_B}(6\pi^2 \frac{N}{V})^{\frac{1}{3}} \quad \text{…………..(8)}$$

The significance of this physical term will be discussed in the 4th section. For elements in the same group, heavier atoms have lower Debye temperatures, simply because the velocity of sound decreases as the density increases. The Debye temperatures of several substances are listed in Table 1.

Table1: Debye Temperatures of several substances.

Aluminum	428K	Iron	470K	Silicon	645K	Tungsten	400K
Cadmium	209K	Lead	105K	Silver	225K	Zinc	327K
Chromium	630K	Manganese	410K	Tantalum	240K	Carbon	2230K
Copper	343.5K	Nickel	450K	Tin(white)	200K	Ice	192K
Gold	165K	Platinum	240K	Titanium	420K		

Derivation for Specific Heat:

In the Debye approximation, the velocity of sound υs is taken as constant for each polarization type, as it would be for a classical elastic continuum. According to equation (7), the density of states is-

$$D_{(\omega)} = \frac{dN}{d\omega} = \frac{V\omega^2}{2\pi^2 \upsilon_s^3} \quad \text{..........(9)}$$

Thus, thermal energy for each polarization type is given by

$$U = \int d\omega D(\omega) n(\omega) \hbar\omega = \int_0^{\omega_D} d\omega \frac{V\omega^2}{2\pi^2 \upsilon_s^3} \frac{\hbar\omega}{e^{\hbar\omega/k_B T} - 1} \quad \text{.......................(10)}$$

There are 3 polarization types, 1 longitude and 2 transverse. For brevity, we assume that phonon velocity is independent of polarization. Thus we multiply a factor 3 and use equation (7) to obtain the total energy of phonon:

$$U = \frac{3V\hbar}{2\pi^2 \upsilon_s^3} \int_0^{\omega_D} d\omega \, \frac{\omega^3}{e^{\hbar\omega/k_B T} - 1}$$

$$= \frac{3V k_B^4 T^4}{2\pi^2 \upsilon_s^3 \hbar^3} \int_0^{x_D} dx \, \frac{x^3}{e^x - 1}$$

$$= 9Nk_B T (\frac{T}{T_D})^3 \int_0^{x_D} dx \, \frac{x^3}{e^x - 1} \quad \text{....(11)}$$

where $x \equiv \hbar\omega/k_B T$ and $x_D \equiv T_D/T$.

The heat capacity is:

$$C_V = \frac{\partial U}{\partial T} = 9Nk_B (\frac{T}{T_D})^3 \int_0^{x_D} dx \, \frac{x^4 e^x}{(e^x - 1)^2} \quad \text{................(12)}$$

At the left of figure below, the experimental results of specific heats of four substances are plotted as a function of temperature and they look very different. But if they are scaled to T/T_D, they look very similar and are very close to the Debye theory.

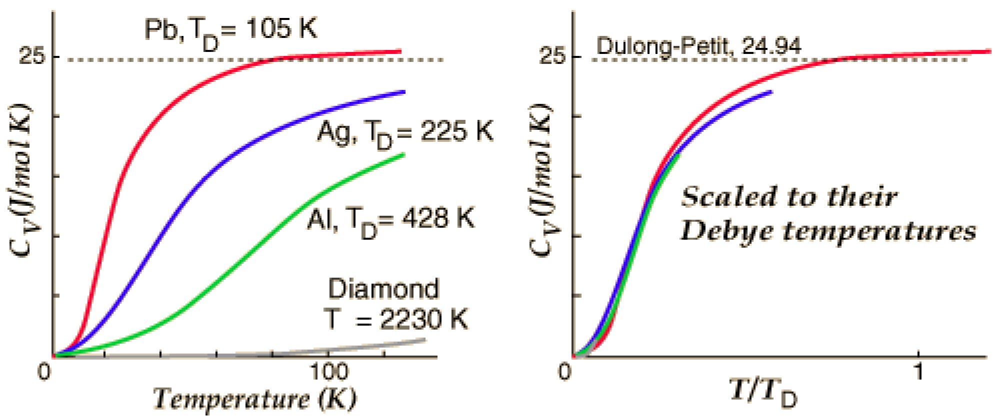


Fig.3.24 Specific Heats of Lead, Silver, Aluminum and Diamond

High and Low Temperature Limits:

The integral in equation (12) cannot be evaluated in closed form. But the high and low temperature limits can be assessed.

1. High Temperature Limit:

For the high temperature case where $T \gg T_D$, the value of x is very small throughout the range of the integral. This justifies using the approximation to the exponential $e^x = 1 + x$ and reduces equation (11) and (12) to

$$U = 9Nk_BT(\frac{T}{T_D})^3 \int_0^{x_D} x^2 dx = 3Nk_BT \quad \text{................(13)}$$

$$C_V = 3Nk_B \quad \text{..............(14)}$$

which is the classical Dulong-Petit result.

When the temperature is above the Debye temperature, the heat capacity is very close to the classical value $3NkBT$. For temperatures below the Debye temperature, quantum effects become important and C_V decreases to zero. Note that diamond, with a Debye temperature of 1860K, is a “quantum solid” at room temperature.[8]

2. Low Temperature Limit:

At very low temperature where $T \ll T_D$, only long wavelength acoustic modes are thermally excited. These are just the modes that can be treated as elastic continuum with macroscopic elastic constants. The energy of those short wavelength modes are too high to be populated significantly at low temperatures. We may approximate $x_D \equiv T_D/T$ to infinity and make use of the standard integral

$$\int_0^\infty dx\,\frac{x^3}{e^x-1}=\frac{\pi^4}{15}$$ ……………………….(15)

To obtain

$$U=\frac{3\pi^4 Nk_BT^4}{5T_D^3}$$ ………………………………..(16)

$$C_V=\frac{12\pi^4 Nk_BT^3}{5T_D^3}\cong 324Nk_B\frac{T^3}{T_D^3}$$ ………………….(17)

For actual crystals, the temperatures at which the T^3 approximation holds are quite low, even may be below $T_D/50$. It is easy to understand $(T/T_D)^3$ with a simple argument. Only the modes with $\hbar\omega<k_BT$ will be excited to any appreciable extent at a low temperature T. Define $\omega_T \equiv kT/\hbar$ as the largest frequency excited at this temperature. In the reciprocal space, the fraction occupied by the excited states is $(q_T/q_D)^3$ or $(w_T/w_D)^3 = (T/T_D)^3$.

Extension: Einstein-Debye Specific Heat:

This T^3 dependence of the specific heat at very low temperatures agrees with experiment for nonmetals. For metals the specific heat of highly mobile conduction electrons is approximated by Einstein Model, which is composed of single-frequency quantum harmonic oscillators. The temperature dependence of Einstein model is just T. It becomes significant at low temperatures and is combined with the above lattice specific heat in the Einstein-Debye specific heat

$$C_{metal}=C_{electron}+C_{phonon}=\frac{\pi^2Nk^2}{2E_f}T+\frac{12\pi^4Nk_B}{5T_D^3}T^3$$ …………(18)

Finally, experiments suggest that amorphous materials do not follow the Debye T^3 law even at temperatures below $0.01T_D$. There is more yet to be learned.

3.10 Concept of Einstein's temperature:

Einstein assumed three things when he investigated the heat capacity of solids. First, he assumed that each solid was composed of a lattice structure consisting of N atoms. Each atom was treated as moving independently in three dimensions within the lattice (3 degrees of freedom). This meant that the entire lattice's vibrational motion could be described by a total of 3N motions, or degrees of freedom. Secondly, he assumed that the atoms inside the solid lattice did not interact with each other and thirdly, all of the atoms inside the solid vibrated at the same frequency. The third point highlights the main difference in Einstein's and Debye's two models.

Einstein's first point is accurate because the experimental data supported his hypothesis, however his second point is not because if atoms inside a solid could not interact sound could not propagate through it. For example, a tuning fork's atoms, when struck, interact with one another to create sound which travels through air to the listener's ear. Atoms also interact in a solid when they are heated. Take for example a frying pan. If the pan is heated on one side, the heat transfers throughout the metal effectively warming the entire pan. Molecules that make up the frying pan interact to transfer heat. Much in the same way the oscillators in a solid interact when energy is added to the system. The extent of these interactions leads to the physically observed heat capacity.

The heat capacity of a solid at a constant volume is

$$c_v = \left(\frac{\partial U}{\partial T}\right)_v$$

$$= 3Nk_B\left(\frac{\theta_E}{T}\right)^2$$

$$= \frac{\exp\left(\frac{\theta_E}{T}\right)}{\left(\exp\frac{\theta_E}{T} - 1\right)^2}$$

Where

- $\theta_E = \frac{hv}{k_B}$ is the Einstein temperature,
- h is Planck's constant,
- k_B is Boltzmann's constant, and
- ν is the oscillator frequency of the atoms inside the solid.

The Einstein temperature's accessibility of the vibrational energy inside of a solid molecule determines the heat capacity of that solid. The greater the accessibility the greater the heat capacity. If the vibrational energy is easily accessible the collisions in the molecule have a greater probability of exciting the atom into an upper vibrational level. This is displayed below.

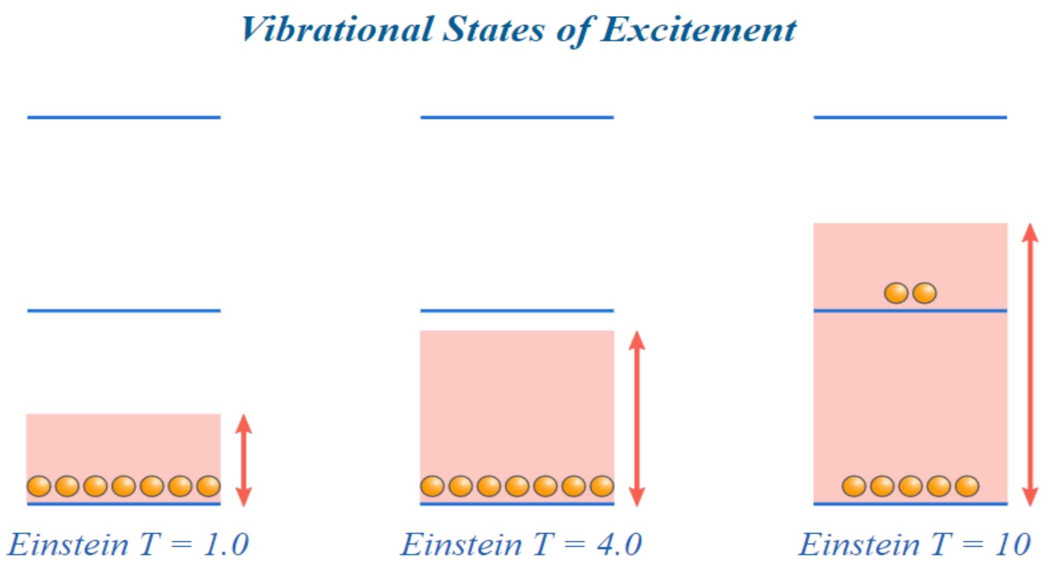


Fig.3.25. This Figure illustrates the effect of the Einstein temperature on the likelihood of an oscillator absorbing energy from a collision and transferring that energy into stored heat or heat capacity. As the Einstein temperature increases, the greater the probability of an oscillator being excited to the next vibrational state.

So the Einstein temperature specifically indicates the probability that a molecule has in its degrees of freedom to store energy in its atomic oscillators (or bonds). Comparing the Einstein temperature to the traditional classical values of Heat capacity will illustrate the differences the specific strengths (high temperature) and weaknesses (low temperature) of the Einstein model.

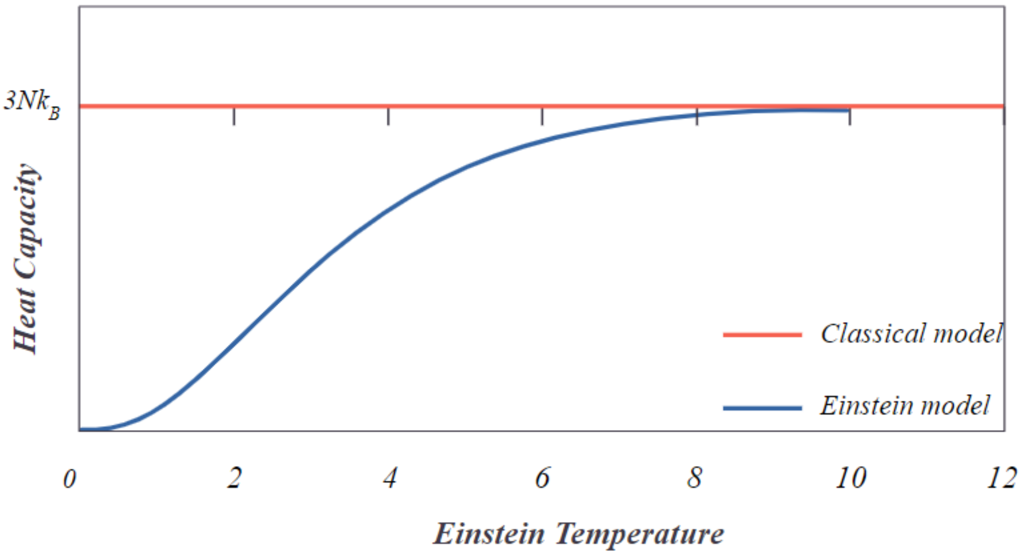


Fig.3.26. This graph shows the accuracy of the of the Einstein temperature at low temperatures. The classical model is effective at temperatures above 10 Einstein Temperatures.

Thus, it can be seen when the temperature is very large compared to the Einstein temperature

$$c_v \approx 3Nk_B = 3nR.$$

Einstein's model reveals the accuracy of the Petit and Dulong model and models high temperatures accurately. However, just as Petit and Dulong's model decreased in accuracy as the temperature decreased, so followed Einstein's.

When examining the extremely low temperature limit:

$$\frac{\theta_E}{T} \gg 1$$

it can be seen:

$$c_v = 3Nk_B\left(\frac{\theta_E}{T}\right)^2 e^{\frac{-\theta_E}{T}}$$

As temperature (T) goes to zero, the exponential portion of the above equation goes to zero and therefore c_v also approaches zero. This supports the experimental values as temperature approaches zero the heat capacity of the solid likewise decreases to zero.

Einstein's theory also explains solids that exhibit a low heat capacity even at relatively high temperatures. An example of such a solid is diamond. The heat capacity of diamond approaches 3Nk_B as temperature greatly increases. Einstein's model supports this through the definition of an Einstein temperature. As the Einstein temperature increases v must increase likewise. This is the equivalent of each atom possessing more energy and therefore vibrating more rapidly within the solid itself. The oscillator frequency can be written as:

$$v = \frac{1}{2\pi}\sqrt{\frac{\kappa}{\mu}}$$

Where κ is the force constant and μ is the reduced mass. This formula better predicts solids with high force constants or low reduced masses. This corrects deviations from the Petit and Dulong model.

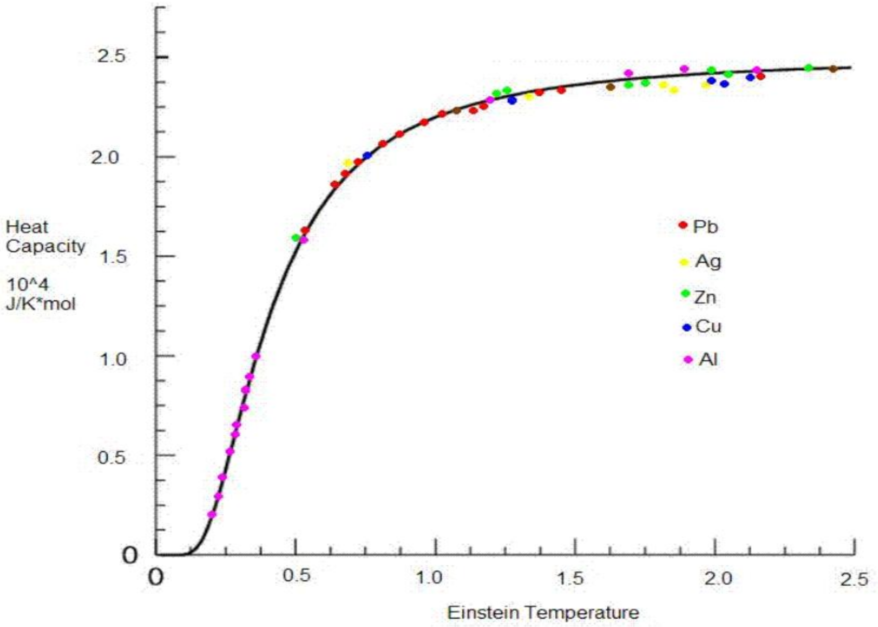


Fig.3.27. This graph indicates the heat capacity of several solid metals as a function of the Einstein temperature: $\theta_E = \frac{hv}{k_B}$

Essentially the Einstein temperature allows for the heat capacity equation and the vibrational frequencies in the solid to change as the temperature changes. This effectively adjusts for the deviations seen in the Petit/Dulong model. As the temperature increases or decreases, the Einstein temperature increases or decreases likewise to mirror the actual physical activity within the solid.

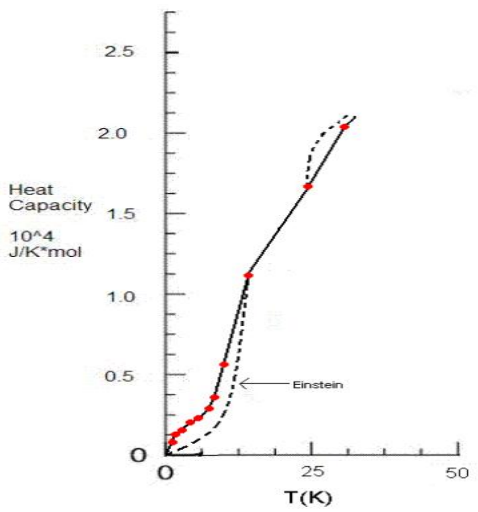


Fig.3.28. This plot shows the heat capacity of lead compared to the expected values calculated from Einstein's model.

Einstein's model predicts relatively low temperatures well. However, when decreasing from approximately 15 K, Einstein's model deviates from experimental values. Also, this can be observed, although not as dramatically, for temperatures from 25 K to 30 K. Clearly a term or correction is still missing from Einstein's model to increase its accuracy.

Debye Temperature:

In thermodynamics and solid state physics, the Debye model is a method developed by Peter Debye in 1912 for estimating the phonon contribution to the specific heat (heat capacity) in a solid.[1] It treats the vibrations of the atomic lattice (heat) as phonons in a box, in contrast to the Einstein model, which treats the solid as many individual, non-interacting quantum harmonic oscillators. The Debye model correctly predicts the low temperature dependence of the heat capacity, which is proportional to T^3 – the Debye T^3 law. Just like the Einstein model, it also recovers the Dulong–Petit law at high temperatures. But due to simplifying assumptions, its accuracy suffers at intermediate temperatures.

Derivation:

The Debye model is a solid-state equivalent of Planck's law of black body radiation, where one treats electromagnetic radiation as a photon gas. The Debye model treats atomic vibrations as phonons in a box (the box being the solid). Most of the calculation steps are identical as both are examples of a mass less Bose gas with linear dispersion relation

Consider a cube of side L. From the particle in a box article, the resonating modes of the sonic disturbances inside the box (considering for now only those aligned with one axis) have wavelengths given by

$$\lambda_n = \frac{2L}{n}$$

Where n is an integer. The energy of a phonon is

$$E_n = h\nu_n$$

Where h is Planck's constant and v_n is the frequency of the phonon. Making the approximation that the frequency is inversely proportional to the wavelength, we have

$$E_n = h\nu_n = \frac{hc_s}{\lambda_n} = \frac{hc_s n}{2L}$$

in which C_s is the speed of sound inside the solid. In three dimensions we will use

$$E_n^2 = p_n^2 c_s^2 = \left(\frac{hc_s}{2L}\right)^2 \left(n_x^2 + n_y^2 + n_z^2\right)$$

in which p_n is the magnitude of the three-dimensional momentum of the phonon.

The approximation that the frequency is inversely proportional to the wavelength (giving a constant speed of sound) is good for low-energy phonons but not for high-energy phonons (see the article on phonons.) This disagreement is one of the limitations of the Debye model, and produces incorrect results at intermediate temperatures, whereas at the low-temperature and high-temperature limits the results are exact.

Let's now compute the total energy in the box,

$$E = \sum_n E_n \, \bar{N}(E_n)$$

where $\bar{N}(E_n)$ is the number of phonons in the box with energy E_n. In other words, the total energy is equal to the sum of energy multiplied by the number of phonons with that energy (in one dimension). In 3 dimensions we have:

$$U = \sum_{n_x} \sum_{n_y} \sum_{n_z} E_n \, \bar{N}(E_n)$$

Here, the Debye model and Planck's law of black body radiation differ. Unlike electromagnetic radiation in a box, there is a finite number of phonon energy states because a phonon cannot have arbitrarily high frequencies. Its frequency is bounded by the medium of its propagation—the atomic lattice of the solid. Consider an illustration of a transverse phonon below.

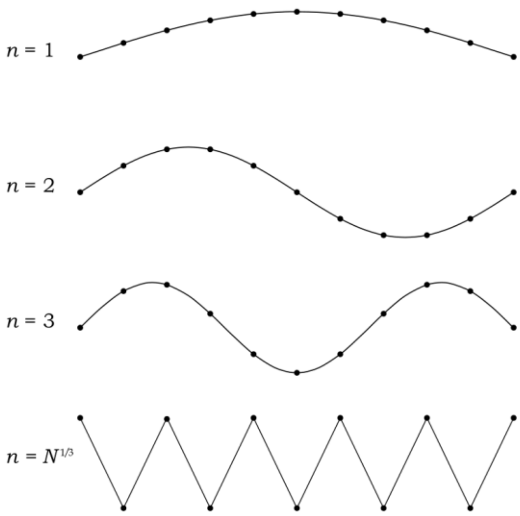


Fig.3.29 Transverse phonon

It is reasonable to assume that the minimum wavelength of a phonon is twice the atom separation, as shown in the lower figure. There are N atoms in a solid. Our solid is a cube, which means there are $\sqrt[3]{N}$ atoms per edge. Atom separation is then given by $L/\sqrt[3]{N}$, and the minimum wavelength is

$$\lambda_{\min} = \frac{2L}{\sqrt[3]{N}}$$

Making the maximum mode number n (infinite for photons)

$$n_{\max} = \sqrt[3]{N}$$

This number bounds the upper limit of the triple energy sum

$$U = \sum_{n_x}^{\sqrt[3]{N}} \sum_{n_y}^{\sqrt[3]{N}} \sum_{n_z}^{\sqrt[3]{N}} E_n \, \bar{N}(E_n)$$

For slowly varying, well-behaved functions, a sum can be replaced with an integral (also known as Thomas–Fermi approximation)

$$U \approx \int_0^{\sqrt[3]{N}} \int_0^{\sqrt[3]{N}} \int_0^{\sqrt[3]{N}} E(n) \, \bar{N}\left(E(n)\right) \, dn_x \, dn_y \, dn_z$$

So far, there has been no mention $\bar{N}(E_n)$ of , the number of phonons with energy E Phonons obey Bose–Einstein statistics. Their distribution is given by the famous Bose–Einstein formula

$$\langle N\rangle_{BE} = \frac{1}{e^{E/kT}-1}$$

Because a phonon has three possible polarization states (one longitudinal, and two transverse which approximately do not affect its energy) the formula above must be multiplied by 3,

$$\bar{N}(E) = \frac{3}{e^{E/kT}-1}$$

(Actually one uses an effective sonic velocity C_s=C_{eff}, i.e. the Debye temperature T_D (see below) is proportional to C_{eff}, more precisely

$$T_{\mathrm{D}}^{-3} \propto c_{\mathrm{eff}}^{-3} := (1/3)c_{\mathrm{long}}^{-3} + (2/3)c_{\mathrm{trans}}^{-3},$$

where one distinguishes longitudinal and transversal sound-wave velocities (contributions 1/3 and 2/3, respectively). The Debye temperature or the effective sonic velocity is a measure of the hardness of the crystal.)

Substituting into the energy integral yields

$$U = \int_0^{\sqrt[3]{N}}\int_0^{\sqrt[3]{N}}\int_0^{\sqrt[3]{N}} E(n)\,\frac{3}{e^{E(n)/kT}-1}\,dn_x\,dn_y\,dn_z$$

The ease with which these integrals are evaluated for photons is due to the fact that light's frequency, at least semi-classically, is unbound. As the figure above illustrates, this is not true for phonons. In order to approximate this triple integral, Debye used spherical coordinates

$$(n_x, n_y, n_z) = (n\sin\theta\cos\phi, n\sin\theta\sin\phi, n\cos\theta)$$

and approximated the cube by an eighth of a sphere

$$U \approx \int_0^{\pi/2}\int_0^{\pi/2}\int_0^{R} E(n)\,\frac{3}{e^{E(n)/kT}-1} n^2\sin\theta\,dn\,d\theta\,d\phi$$

where R is the radius of this sphere, which is found by conserving the number of particles in the cube and in the eighth of a sphere. The volume of the cube is N unit-cell volumes,

$$N = \frac{1}{8}\frac{4}{3}\pi R^3$$

so we get

$$R = \sqrt[3]{\frac{6N}{\pi}}$$

The substitution of integration over a sphere for the correct integral introduces another source of inaccuracy into the model.

The energy integral becomes

$$U = \frac{3\pi}{2} \int_0^R \frac{hc_s n}{2L} \frac{n^2}{e^{hc_s n/2LkT} - 1} \, dn$$

Changing the integration variable to

$$x = \frac{hc_s n}{2LkT}$$

$$U = \frac{3\pi}{2} kT \left(\frac{2LkT}{hc_s} \right)^3 \int_0^{hc_s R/2LkT} \frac{x^3}{e^x - 1} \, dx$$

To simplify the appearance of this expression, define the Debye temperature T_D

$$T_D \stackrel{\text{def}}{=} \frac{hc_s R}{2Lk} = \frac{hc_s}{2Lk} \sqrt[3]{\frac{6N}{\pi}} = \frac{hc_s}{2k} \sqrt[3]{\frac{6}{\pi} \frac{N}{V}}$$

Where V is the volume of the cubic box of side L.

Debye temperature table:

It gives a good approximation for the low temperature heat capacity of insulating, crystalline solids where other contributions (such as highly mobile conduction electrons) are negligible. For metals, the electron contribution to the heat is proportional to T, which at low temperatures dominates the Debye T^3 result for lattice vibrations. In this case, the Debye model can only be said to approximate for the lattice contribution to the specific heat. The following table lists Debye temperatures for several pure elements and sapphire:

Aluminium	428 K	Copper	343 K	Nickel	450 K	Silver	215 K
Beryllium	1440 K	Germanium	374 K	Platinum	240 K	Tantalum	240 K
Cadmium	209 K	Gold	170 K	Rubidium	56 K	Tin (white)	200 K

Caesium	38 K	Iron	470 K	Sapphire	1047 K	Titanium	420 K
Carbon	2230 K	Lead	105 K	Selenium	90 K	Tungsten	400 K
Chromium	630 K	Manganese	410 K	Silicon	645 K		

SAQ.3

a) What is the need and statement of Einstein's theory of specific heats?

b) Write the assumptions and limitations of Debye theory of specific heats.

c) Define the Concept of Einstein's temperature.

d) What do you mean by Debye temperature?

Examples:

Q.1. A body with mass 2 kg absorbs heat 100 calories when its temperature raises from 20^oC to 70^oC. What is the specific heat of the body?

Solution:

$c = Q / m\ \Delta T$

$c = 100\ cal / (2000\ gr)(50^oC)$

$c = 100\ cal / 100{,}000\ gr\ ^oC$

$c = 10^2\ cal / 10^5\ gr\ ^oC$

$c = (10^2\ cal)(10^{-5}\ gr^{-1}\ {}^oC^{-1})$

$c = 10^{-3}\ cal\ gr^{-1}\ {}^oC^{-1}$

$c = 10^{-3}\ cal/gr\ ^oC$

Q.2. The specific heat of water is 4180 J/kg C^o. How much the heat capacity of 2 kg water heat capacity?

Solution:

$C = m\ c$

$C = (2\ kg)(4180\ J/kg\ C^o)$

$C = (2)(4180\ J/C^o)$

$C = 8360\ J/C^o$

Q.3. The specific heat of aluminum is 900 J/kg Co. How much the heat capacity of 2 gram aluminum heat capacity?

Solution:

C = m c

$C = (2 \times 10^{-3} \text{ kg})(9 \times 10^{2} \text{ J/kg C}^{o})$

$C = 18 \times 10^{-3} \times 10^{2} \text{ J/C}^{o}$

$C = 18 \times 10^{-1} \text{ J/C}^{o}$

$C = 1.8 \text{ J/C}^{o}$

Q.4. A 500 gram cube of lead is heated from 25 °C to 75 °C. How much energy was required to heat the lead? The specific heat of lead is 0.129 J/g°C.

Solution: First, let’s the variables we know.

m = 500 grams

c = 0.129 J/g°C

$\Delta T = (T_{final} - T_{initial}) = (75\ °C - 25\ °C) = 50\ °C$

Plug these values into the specific heat equation from above.

$Q = mc\Delta T$

Q = (500 grams)·(0.129 J/g°C)·(50 °C)

Q = 3225 J

Q.5. A 25-gram metal ball is heated 200 °C with 2330 Joules of energy. What is the specific heat of the metal?

Solution: List the information we know.

m = 25 grams

$\Delta T = 200\ °C$

Q = 2330 J

Place these into the specific heat equation.

$Q = mc\Delta T$

2330 J = (25 g)c(200 °C)

2330 J = (5000 g°C)c

Divide both sides by 5000 g°C

$$\frac{2330\, J}{5000\ g^{o}\ C} = c$$

c = 0.466 J/g°C

Q.6. A hot 1 kg chunk of copper is allowed to cool to 100°C. If the copper gave off 231 kJ of energy, what was the initial temperature of the copper? The specific heat of copper is 0.385 J/g°C.

Solution: List our given variables:

m = 1 kg

T_{final} = 100 °C

Q = -231 kJ (The negative sign is because the copper is cooling and losing energy.)

c = 0.385 J/g°C

We need to make our units consistent with the specific heat units, so let's convert the mass and energy units.

m = 1 kg = 1000 grams

1 kJ = 1000 J

Q = -231 kJ · (1000 J/kJ) = -231000 J

Plug these values into the specific heat formula.

$Q = mc\Delta T$

-231000 J = 1000 g · (0.385 J/g°C) · ΔT

-231000 J = 385 J/°C · ΔT

$$\frac{-231000\,J}{385\ \mathrm{J/°C}} = \Delta T$$

ΔT = -600 °C

$\Delta T = (T_{final} - T_{initial})$

Plug in the values for ΔT and T_{final}.

-600 °C = (100 °C – $T_{initial}$)

Subtract 100 °C from both sides of the equation.

-600 °C – 100 °C = – $T_{initial}$

-700 °C = – $T_{initial}$

$T_{initial}$ = 700 °C

3.11 Summary:

1. The inter-atomic forces between the crystal atoms (ions) are assumed to undergo simple harmonic motion, where the force acting on a displaced nucleus, is proportional to the displacement distance of the nucleus from its equilibrium lattice position.

2. Solids can be classified on the basis of the bonds that hold the atoms or molecules together. This approach categorizes solids as either molecular, covalent, ionic, or metallic.
3. Lattice energy is an estimate of the bond strength in ionic compounds. It is defined as the heat of formation for ions of opposite charge in the gas phase to combine into an ionic solid.
4. The monatomic chain is a one-dimensional model representing the situation in a crystal with a primitive lattice, i.e. with only a single atom in the unit cell. In the same way, we can use a one-dimensional diatomic chain model to represent centred lattices, where more than one atom is present in the unit cell.
5. he allowed frequencies of propagation wave are split into an upper branch known as the optical branch, and a lower branch called the acoustical branch. For acoustical branch (in the long wavelength limit) the displacement of both atoms has the same amplitude, direction and phase.
6. Phonon, in condensed-matter physics, a unit of vibrational energy that arises from oscillating atoms within a crystal. A packet of these waves can travel throughout the crystal with a definite energy and momentum, so in quantum mechanical terms the waves can be treated as a particle, called a phonon.
7. Near room temperature, the heat capacity of most solids is around 3k per atom (the molar heat capacity for a solid consisting of n-atom molecules is ~3nR). This is the well-known Dulong and Petit law. At low temperatures, C_v decreases, becoming zero at T=0.
8. It is the amount of heat required to change. the temperature of unit mass of substance by unit degree temperature. Classical Theory of Specific heat of a solid. Solid molecules have 6 degrees of freedom (3 translational and 3 vibrational).
9. A statement of the Dulong–Petit law in modern terms is that, regardless of the nature of the substance, the specific heat capacity c of a solid element (measured in joule per kelvin per kilogram) is equal to 3R/M, where R is the gas constant (measured in joule per kelvin per mole) and M is the molar mass.
10. In quantum theory of specific heat of solids: A theory of the specific heat of solids proposed by Albert Einstein in 1906. In this theory, Einstein attributed the specific heat of solids to the vibrations of the solid and made the simplifying assumption that all the vibrations have the same frequency.

11. Einstein's theory of specific heat of Solid couldn't explain the experimental results obtained at very low temperatures… From the experiment it is observed that the specific heat of solids has a T^3 dependence on the absolute temperature of the solid.
12. A theory of the specific heat capacity of solids put forward by Peter Debye in 1912, in which it was assumed that the specific heat is a consequence of the vibrations of the atoms of the lattice of the solid. This result is in very good agreement with experiment at low temperatures.
13. θ_E is the 'Einstein temperature', which is different for each solid, and reflects the rigidity of the lattice. At the high temperature limit, when $T >> \theta_E$ (and $x << 1$), the Einstein heat capacity reduces to $C_v = 3Nk$, the Dulong and Petit law [prove by setting ex ~ 1+x in the denominator].
14. The Debye temperature to simplify the integration of the heat capacity. The Debye cut off frequency or temperature separates the collective thermal lattice vibration from the independent thermal lattice vibration.

3.12 Terminal Questions:

1) Explain the inter-atomic force and inter molecule force in detail.
2) Explain in detail classification of solids.
3) Discuss about the Lattice energy of ionic crystals.
4) Explain the Vibration of mono-atomic and diatomic linear chain. Also discuss about acoustic, optical modes and phonon.
5) What do mean by Thermal capacity of solids?
6) Explain the classical theory of specific heats.
7) Explain the Dulong and Petit's law.
8) Explain the need of quantum theory of specific heat of solids.
9) Explain the need of Einstein's theory of specific heats. Also derive its equations.
10) What do mean by the need of Debye theory of specific heats? derive its equations.
11) Explain the Concept of Einstein's temperature.
12) What do mean by the Debye temperature?
13) A body with mass 1.5 kg absorbs heat 200 calories when its temperature raises from 15°C to 65°C. What is the specific heat of the body?
14) The specific heat of water is 5260 J/kg C°. How much the heat capacity of 3 kg water heat capacity?

15) The specific heat of aluminum is 800 J/kg Co. How much the heat capacity of 3 gram aluminum heat capacity?

16) A 600 gram cube of lead is heated from 35 °C to 85 °C. How much energy was required to heat the lead? The specific heat of lead is 0.145 J/g°C.

17) A 30-gram metal ball is heated 300 °C with 2530 Joules of energy. What is the specific heat of the metal?

18) A hot 2 kg chunk of copper is allowed to cool to 150°C. If the copper gave off 255 kJ of energy, what was the initial temperature of the copper? The specific heat of copper is 0.485 J/g°C.

Unit 04 Magnetism and superconductivity

Structure:

4.1 Introduction:

Diamagnetic materials, like water, or water-based materials, have a relative magnetic permeability that is less than or equal to 1, and therefore a magnetic susceptibility less than or equal to 0, since susceptibility is defined as $\chi_v = \mu_v - 1$. This means that diamagnetic materials are repelled by magnetic fields. Paramagnetism is a form of magnetism whereby some materials are weakly attracted by an externally applied magnetic field, and form internal, induced magnetic fields in the direction of the applied magnetic field. ... Paramagnetic materials include aluminium, oxygen, titanium, and iron oxide (FeO).

According to the Curie's Law, the magnetization which is present in a paramagnetic material is said to be directly proportional to the applied field of magnetic. If the object which we have used is heated then the magnetization is viewed to be temperature which is inversely proportional. The Curie–Weiss law describes the magnetic susceptibility χ of a ferromagnet in the paramagnetic region above the Curie point: where C is a material-specific Curie constant, T is the absolute temperature, and T_C is the Curie temperature, both measured in Kelvin. Relationship of the Curie Law with the Curie-Weiss Law, According to the Curie Law, the magnetization of any paramagnetic element is directly proportional to the applied magnetic field. Often represented as: M = C x (BT) here M = Magnetization, B = Magnetic Field, T = absolute temperature, C = Curie Constant.

In Classical theory of diamagnetism, we will obtain an expression for the change in magnetic moment of an orbiting electron in a diamagnetic atom and the induced magnetic moment per unit volume of diamagnetic material in the applied magnetic field B_0. The classical theory of paramagnetism is based on the assumption that the permanent. magnetic moments of a given atom or ion can rotate freely and possess any orientation with. respect to the applied magnetic field. According to quantum theory the magnetic moments are. quantized.

Diamagnetism is a quantum mechanical effect that occurs in all materials; when it is the only contribution to the magnetism, the material is called diamagnetic. In paramagnetic and ferromagnetic substances, the weak diamagnetic force is overcome by the attractive force of magnetic dipoles in the material. Quantum theory of Paramagnetism. According to classical theory the atoms of the paramagnetic gas are assumed to be small permanent magnets due to circulating electrons. In the absence of the external magnetic field , the magnetic axes of the atoms are uniformly distributed in all directions.

Ferromagnetism is the basic mechanism by which certain materials (such as iron) form permanent magnets, or are attracted to magnets. In physics, several different types of magnetism are distinguished. The common ones are iron, cobalt, nickel and most of their alloys, and some compounds of rare earth metals. Anti-ferromagnets are magnetic materials that have no net macroscopic magnetization and, therefore, are almost insensitive to external magnetic fields. Néel first assumed that the ordered magnetic arrangements could be described in terms of “spin sublattices,” each having all spins aligned in the same direction. Ferri-magnetism, type of permanent magnetism that occurs in solids in which the magnetic fields associated with individual atoms spontaneously align themselves, some parallel, or in the same direction (as in ferromagnetism), and others generally antiparallel, or paired off in opposite directions.

Superconductivity, complete disappearance of electrical resistance in various solids when they are cooled below a characteristic temperature. This temperature, called the transition temperature, varies for different materials but generally is below 20 K (−253 °C). While many materials exhibit some small amount of diamagnetism, superconductors are strongly diamagnetic. Since diamagnetics have a magnetization that opposes any applied magnetic field, the superconductor is repelled by the magnetic field.

Superconductors also have interesting magnetic properties; they are perfect diamagnets: when an applied magnetic field is applied, eddy currents in the superconductor induce a magnetic field which exactly cancels the applied magnetic field.

Meissner effect, the expulsion of a magnetic field from the interior of a material that is in the process of becoming a superconductor, that is, losing its resistance to the flow of electrical currents when cooled below a certain temperature, called the transition temperature, usually close to absolute zero. A BCS theory of superconductivity formulated by John Bardeen, Leon Cooper, and Robert Schrieffer. It explains the phenomenon in which a current of electron pairs flows without resistance in certain materials at low temperatures. The main point of the BCS theory is that the attractive electron-electron interaction mediated by the phonons gives rise to Cooper pairs, i.e. bound states formed by two electrons of opposite spins and moment.

A superconductor is a perfect conductor of electricity; it carries direct current with 100% efficiency because no energy is dissipated by resistive heating. Once induced in a superconducting loop, direct current can flow undiminished forever. There are two types of superconductors. There are 30 pure metals that exhibit zero resistivity below their critical temperature and exhibit the Meissner effect, the property of excluding magnetic fields from the interior of the superconductor while the superconductor is at a temperature below the critical temperature.

Josephson Effect, flow of electric current between two pieces of superconducting material separated by a thin layer of insulating material. Superconductors are materials that lose all electrical resistance when cooled below a certain temperature near absolute zero. The Quantum Hall effect is phenomena exhibited by 2D materials, and can also be found in graphene. When electrons in a 2D material at very low temperature are subjected to a magnetic field, they follow cyclotron orbits with a radius inversely proportional to the magnetic field intensity. The quantum Hall effect (or integer quantum Hall effect) is a quantized version of the Hall Effect and which is observed in two-dimensional electron systems subjected to low temperatures and strong magnetic fields, in which the Hall resistance R_{xy} exhibits steps that take on the quantized values.

4.2 Objectives:

After studying this unit you should be able to

- Explain Comparison of features of diamagnetic and paramagnetic materials with examples, Curie law and Curie Wiess law.

- Study and identify Classical and quantum theory of diamagnetism and paramagnetism.
- Explain and identify Qualitative discussion of ferromagnetism, anti-ferromagetism and ferrimagnetism.
- Study and identify Superconductivity and its characteristics, magnetic behavior of superconductor.
- Explain Meisener's effect, BCS theory.
- Explain and identify Types of superconductors with examples, properties and applications.
- Study and identify Josephson Effect, quantum Hall effect.

4.3 Comparison of features of diamagnetic and paramagnetic materials with examples:

Diamagnetic Materials:

Diamagnetic materials are materials that do not attract to an external magnetic field. That is because the atoms or ions present in these materials do not have unpaired electrons. Therefore, diamagnetic materials are repelled by magnetic fields. That happens because an induced magnetic field is created in these materials to the opposite direction to that of the external magnetic field. This induced magnetic field causes the creation of a repulsive force. Diamagnetism can be observed in materials that have symmetry of electronic structure and no permanent magnetic moment. And also, the diamagnetism is not dependent on the temperature.

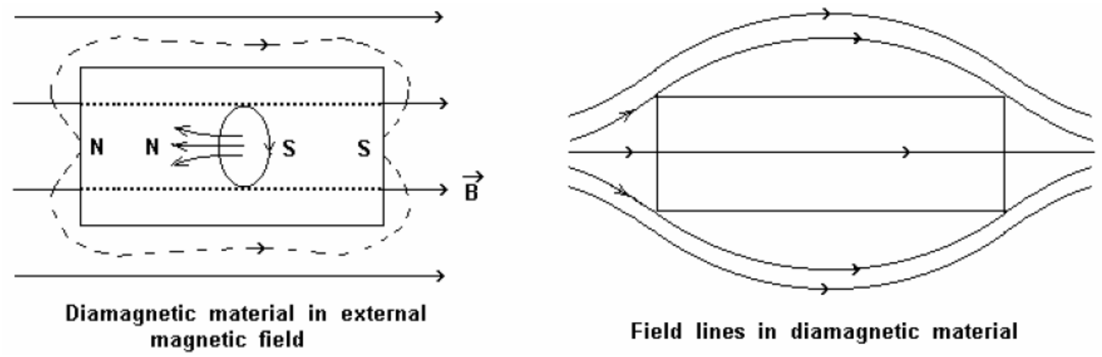


Fig.4.1 The effect of an External Magnetic Field on Diamagnetic Materials

Some examples of diamagnetic materials include;

1. Quartz (silicon dioxide)
2. Calcite (calcium carbonate)

3. Water

For example in quartz, there are silicone atoms and oxygen atoms in the form of SiO_2. The oxidation state of Si atom is +4, and the oxidation state of O atom is -2. Therefore, there are no unpaired electrons in both these atoms.

Paramagnetic Materials:

Paramagnetic materials are materials that are attracted to an external magnetic field. This happens because these materials have unpaired electrons in the atoms or ions present in these materials. These unpaired electrons can create magnetic attraction.

Fig.4.2 Garnet

Paramagnetic materials can be separated from other materials using high-intensity magnetic separators. These separators use a magnetic field with strength of 0.2-0.4 Tesla.

Some examples of paramagnetic materials include;

1. Limonite ($FeTiO_3$)
2. Hematite (Fe_2O_3)
3. Chalcopyrite ($CuFeS_2$)
4. Garnet (Fe-silicates)

Comparison between paramagnetic materials and diamagnetic materials:

Properties	Paramagnetic Materials	Diamagnetic Materials
State	They can be solid, liquid or gas.	They can be solid, liquid or gas.
Effect of Magnet	Weakly attracted by a magnet.	Weakly repelled by a magnet.
Behavior under non-uniform field	Tend to move from low to high field region.	Tend to move from high to low region.
Behavior under external field	They do not preserve the magnetic properties once the external field is removed.	They do not preserve the magnetic properties once the external field is removed.
Effect of Temperature	With the rise of temperature, it becomes a diamagnetic.	No effect.
Permeability	Little greater than unity	Little less than unity
Susceptibility	Little greater than unity and positive	Little less than unity and negative
Examples	Lithium, Tantalum, Magnesium	Copper, Silver, Gold

Curie Law:

The curie law states that in a paramagnetic material, the material's magnetization is directly proportional to an applied magnetic field. But the case is not the same when the material is heated. When it is heated, the relation is reversed i.e. the magnetization becomes inversely proportional to temperature.

Mathematically, it is written as

$M = C \times (B / T)$,

Where; M is the magnetization

B is the magnetic field, measured in teslas

T is absolute temperature, measured in kelvins

C is a material-specific Curie constant

This concept was initially proposed by French physicist, Pierre Curie and the concept holds good for high temperatures and weak magnetic fields. Various experiments by Pierre Curie showed that for many substances the susceptibility is inversely proportional to the absolute temperature T

$$\chi = C/T.$$

This relationship is defined as the Curie's law. The constant 'C' is called the curie constant. The above equation may also be modified to $\chi = C/(T - \theta)$, where θ is a constant. This equation is called the Curie –Weiss Law which will be discussed later.

Now we shall discuss some of the important concepts related with this law:

- On increasing temperature, the magnetic susceptibility of paramagnetic materials decreases and vice versa.
- The magnetic susceptibility of ferromagnetic substances does not change according to curie law.

Curie temperature (T_C): The temperature above which a ferromagnetic material behaves like a paramagnetic material is defined as Curie temperature (T_C).
The minimum temperature at which a ferromagnetic substance is converted into paramagnetic substance is defined as Curie temperature.
For various ferromagnetic materials its values are different. e.g. for Ni.
At this temperature the ferromagnetism of the substances suddenly vanishes.

Below the Curie point, say 770 degree Celsius, for iron, atoms that act as small magnets align themselves in various magnetic materials. In antiferromagnetic materials, the magnetic fields cancel each other as the atomic magnets alternate in opposite directions. The scenario is quite different in ferromagnetic materials. A fractional reinforcement of magnetic field occurs in them as the impulsive arrangement which is a collection of both patterns, includes two different magnetic atoms.

Curie-Weiss law:
Above a critical temperature T_c, the Curie temperature, all ferromagnetic materials become paramagnetic. This is because thermal energy is large enough to overcome the cooperative ordering of the magnetic moments.

The susceptibility of a material, χ, indicates how dramatically a material responds to an applied magnetic field, and is defined as the ratio of the magnetisation of the material, M, and the applied magnetic field, H.

$$\chi = \frac{M}{H}$$

The magnetisation of a material, M, is defined as the magnetic moment per unit volume or per unit mass of a material and is dependent on the individual magnetic dipole moments of the atoms in the material and on the interactions of these dipoles with each other.

Above the Curie temperature there will be a change in the susceptibility as the material becomes paramagnetic, therefore giving the equation:

$$\chi = \frac{C}{T - T_c} = \frac{M}{H}$$

where C is a constant.

The graph below shows the saturation magnetisation (i.e. that obtained in a high magnetic field) of a ferromagnetic element, nickel, as a function of temperature. We see that the saturation magnetisation decreases with increasing temperature until it falls to zero at the Curie temperature where the material becomes paramagnetic:

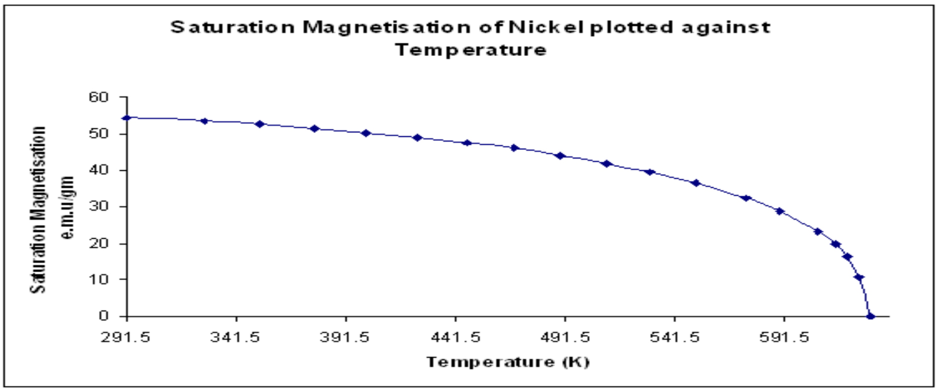


Fig.4.3 Variation of saturation magnetisation with temperature for Nickel

Differentiation of Equation with respect to temperature shows that the susceptibility is a maximum at the Curie temperature. This is no surprise; it is easiest to increase the magnetic moment of the material by applying a magnetic field when the material is undergoing a transition between magnetic order and disorder. The graph below for nickel shows the susceptibility tending to infinity as the temperature moves closer to the Curie temperature:

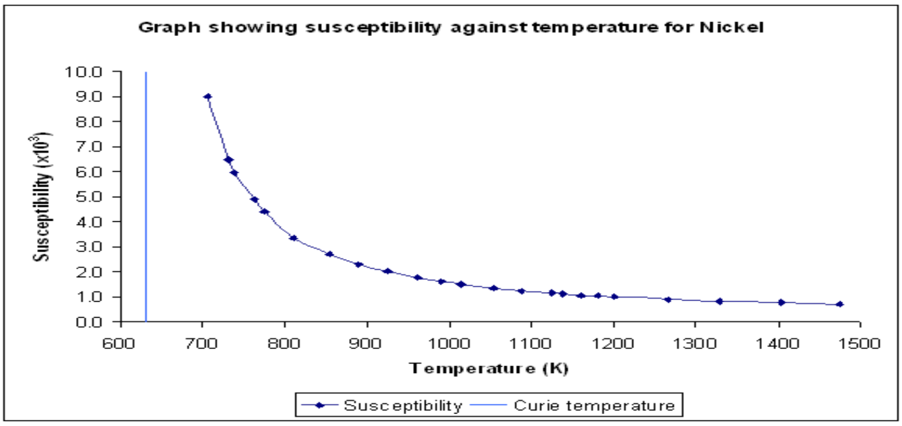


Fig.4.4 Variation of susceptibility with temperature for Nickel

4.4 Classical (Langevin) Theory of Diamagnetism:

Paul Langevin proposed a classical-based model of diamagnetism.

Considering an electron moving in a loop, the induced magnetic moment is the product of the current in the loop and the area it encloses, or

$$\mu_m = IA$$

Taking this as a model for a single orbiting electron, if exposed to the presence of an external magnetic field, the resulting change in the electron acceleration would induce a change in the magnetic moment. The acceleration can be quantified as

$$\frac{dv}{dt} = \frac{F}{m} = \frac{\mathcal{E}e}{m}$$

Where the acceleration is equal to force per unit mass, which is the electric field strength times an electric charge per electron mass. Now apply Lenz's law and see that an emf is created to counteract the change in flux of the loop per unit length

$$\frac{\mathcal{E}e}{m} = \frac{-1}{2\pi r}\frac{d}{dt}(\mu_0 HA)$$

We can now find the acceleration by writing the magnetic moment as

$$\mu_m = e\frac{v}{2\pi r}\pi r^2$$

and equating this to the previous acceleration equation giving the differential equation

$$dv = -\frac{er\mu_0}{2m}dH$$

Which can be integrated to give:

$$\Delta v = -\frac{er\mu_0 H}{2m}$$

Substitute back into the magnetic moment equation and find

$$\Delta\mu_m = -\frac{e^2 r^2 \mu_0 H}{4m}$$

Taking the average of the change in magnetic moment (use polar coordinates, it is zero when the external magnetic field is parallel to the plane of the current loop) and the average distance ($\bar{r}$) from all electrons (Z) belonging to the atom

$$\bar{\Delta}\mu_m = -\frac{e^2 Z\bar{r}^2 \mu_0}{6m}$$

The average change in magnetic moment per atom becomes the diamagnetic susceptibility by simply taking into account the volume, V, over which it acts.

$$\boxed{\chi = -\frac{e^2 Z\bar{r}^2 \mu_0}{6mV}}$$

Diamagnetic Materials:

The diamagnetic response of a material has a measurable contribution to the materials' magnetization only if there are no other magnetic effects present, such as Ferrimagnetism whose susceptibility is much larger in most cases. For this reason, we classify only materials whose net magnetization is diamagnetic, as a diamagnet. This requires that compound to have empty or closed valence shell. The inert noble gases have filled valence shells and thus respond diamagnetically. Substances like silicon, germanium, most covalent solids and polymers also exhibit diamagnetic behavior.

Diamagnetism arises in metals when the paramagentic behavior is sufficiently small. For example, examine beryllium. It has no contribution from ferro, ferri, or antiferromagnetism, so we check its paramagnetic contribution. A single atom of beryllium has paired 1s and 2s electrons.

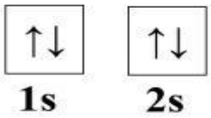


Fig.4.5 Paired 1s and 2s electrons

However, in a crystal lattice, the 2s electrons populate the bottom of the empty 2p band because of band overlap (see: Band Theory of Metals and Insulators). This makes the density of states at the Fermi level very low, thus the paramagnetic susceptibility is much smaller than any diamagnetic contribution. Landau set the framework for diamagnetic calculations of atoms in a lattice.

Applications: Because diamagnetism is essentially the expelling of magnetic fields within a material, strong diamagnetic materials can be levitated, or if they are sufficiently strong and enough area, can levitate magnets. Figure shows a levitating piece of graphite.

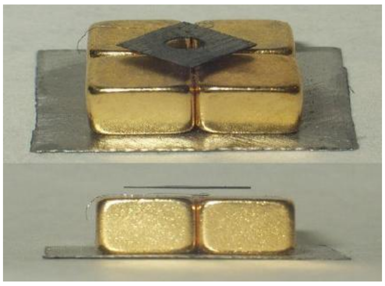


Fig.4.6 Diamagnetic levitation of pyrolytic graphite over permanent neodymium magnets

In the case of superconductors, the diamagnetic response leaves no internal magnetic field. These materials can be easily levitated in the presence of a strong permanent magnet as seen in Figure; this is called the Meissner effect. However, high temperature superconductors (~100 K) are made from exotic materials with expensive processing routes and require cryogenic fluids to accomplish the superconducting state.

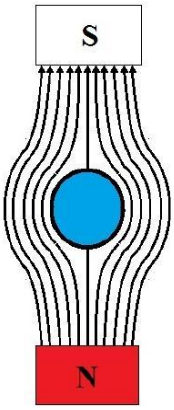


Fig.4.7 A superconductor (blue) expels all of its external magnetic field. They are perfect diamagnets.

Classical Theory of Paramagnetism:

Langevin considered a paramagnetic gas containing N atoms per unit volume each having a permanent magnetic moment μ. The mutual interaction between the magnetic dipoles was assumed to be negligible. In the presence of a magnetic field H, these dipoles tend to orient themselves in the direction of the field in order to minimize their energy. However, the thermal energy at ordinary temperature resists such an alignment of dipoles. In thermal equilibrium, let assume that the dipoles orient themselves at an angle θ with the direction of the applied field as shown in the figure below.

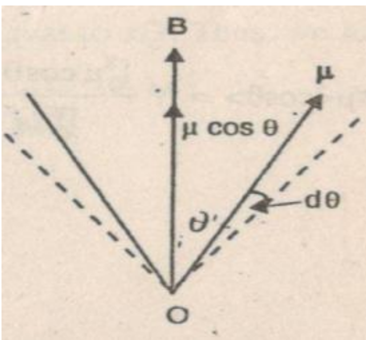


Fig.4.8 Direction of the applied field at an angle θ

The potential energy of each dipole in this position is given by

$$E = -\mu . H = -\mu H \cos\theta$$

Using Maxwell-Boltzmann distribution law, the number of magnetic dipoles having this particular orientation is proportional to

$$exp\left(-\frac{E}{k_B T}\right) \text{ or } exp\left(\frac{\mu H\cos\theta}{k_B T}\right)$$

Also, according to statistical mechanics, the probability for a magnetic dipole to make an angle between θ and (θ + dθ) with the magnetic field, or the number of dipoles, dn, having axes within the solid angle dΩ lying between two hollow cones of semi-angles θ and (θ + dθ) (see the figure above) is given by

$$dn \; \alpha \; exp\left(\frac{\mu H\cos\theta}{k_B T}\right) d\Omega = A \exp\left(\frac{\mu H\cos\theta}{k_B T}\right) 2\pi \sin\theta \, d\theta$$

where A is a constant. Each one of these dipoles contributes a component of magnetic moment μ cosθ to the magnetization, whereas the components perpendicular to the field direction cancel each other. Hence the average component of magnetic moment of each atom along the field direction is given by

$$\langle\mu\rangle = \frac{\int_0^\pi \mu\cos\theta \, dn}{\int_0^\pi dn} = \frac{A\int_0^\pi \mu\cos\theta \, exp\left({\mu H\cos\theta}/{k_B T}\right) 2\pi\sin\theta \, d\theta}{A\int_0^\pi exp\left({\mu H\cos\theta}/{k_B T}\right) 2\pi\sin\theta \, d\theta}$$

$$= \frac{\int_0^\pi \mu\cos\theta \, exp\left({\mu H\cos\theta}/{k_B T}\right) \sin\theta d\theta}{\int_0^\pi exp\left({\mu H\cos\theta}/{k_B T}\right) \sin\theta d\theta}$$

Let $y = \left(\frac{\mu H}{k_B T}\right)$ and x = cos θ, then dx = - sin θ dθ

When θ→0, then x→1, and if θ→π, then x→-1

Therefore,

$$\langle\mu\rangle = \frac{\mu\int_{-1}^{1} x e^{yx}dx}{\int_{-1}^{1} e^{yx}dx} = \frac{\mu\left[\frac{x}{y}e^{yx} - \frac{1}{y^2}e^{yx}\right]_{-1}^{1}}{\left[\frac{e^{yx}}{y}\right]_{-1}^{1}}$$

$$= \frac{\mu\left[\frac{e^{y}}{y} + \frac{e^{-y}}{y} - \frac{e^{y}}{y^2} + \frac{e^{-y}}{y^2}\right]}{\frac{e^{y}}{y} - \frac{e^{-y}}{y}} = \mu\left\{\left(\frac{e^{y}+e^{-y}}{e^{y}-e^{-y}}\right) - \frac{1}{y}\right\}$$

$$= \mu\left[\coth y - \frac{1}{y}\right] = \mu L(y)$$

Where L (y) is called the Langevin function and variation of L (y) with y $(= \frac{\mu H}{k_B T})$

Now, the average magnetic moment of each atom along the field direction multiplied by the number of atoms per unit volume, N, gives the magnetization M = NμL(y).

Case I: For large values of y, i.e. when y >> 1

i.e. (μH >> $k_B T$) when applied magnetic field strength is high and the specimen is kept at very low temperature; then L(y)→1and magnetization becomes maximum M = Nμ = M_S. This is the saturation condition which corresponds to the complete alignment of the magnetic dipoles in the field direction. Ms is called saturation magnetization. Therefore, susceptibility per unit volume $\chi = \frac{M}{H} = \frac{N\mu}{H}$ is independent of temperature.

Case II: For small values of y, i.e. when i.e. at normal magnetic field strengths and ordinary temperatures, the curve is almost linear and coincides with the tangent to the curve at the origin which is equal to y/3.

$L(y) = \left\{\left(\frac{e^{y}+e^{-y}}{e^{y}-e^{-y}}\right) - \frac{1}{y}\right\}$ and if $y << 1$ then

$$L(y) = \left\{\left(\frac{\left[1+y+{}^{y^2}/_{2!}+{}^{y^3}/_{3!}+\ldots\ldots\right]+\left[1-y+{}^{y^2}/_{2!}-{}^{y^3}/_{3!}+\ldots\ldots\right]}{\left[1+y+{}^{y^2}/_{2!}+{}^{y^3}/_{3!}+\ldots\ldots\right]-\left[1-y+{}^{y^2}/_{2!}-{}^{y^3}/_{3!}+\ldots\ldots\right]}\right)-\frac{1}{y}\right\}$$

$$\cong \frac{2\left[1+{y^2}/{2!}+\ldots\right]}{2\left[y+{y^3}/{3!}+\ldots\right]} - \frac{1}{y} \cong \frac{1+{y^2}/{2}}{y\left(1+{y^2}/{6}\right)} - \frac{1}{y}$$ (Neglecting higher order terms of y)

$$\cong \frac{1}{y}\left(1+{y^2}/{2}\right)\left(1-{y^2}/{6}\right) - \frac{1}{y}$$

$$\cong \frac{1}{y}\left(1+\frac{y^2}{2}-\frac{y^2}{6}\right) - \frac{1}{y} = \frac{1}{y}\left(1+\frac{y^2}{3}\right) - \frac{1}{y} = \frac{y}{3}$$

Thus magnetization per unit volume

$$M = N\mu L(y) = N\mu \frac{y}{3} = N\mu \frac{\mu H}{3k_B T} = \frac{N\mu^2 H}{3k_B T}$$

Hence magnetic susceptibility per unit volume becomes

$$\chi = \frac{M}{H} = \frac{N\mu^2}{3k_B T} = \frac{C}{T}$$

Where $C = \frac{N\mu^2}{3k_B}$ is known as Curie constant.

This expression is known as Curie law, which shows that paramagnetic susceptibility is inversely proportional to temperature [for the condition ($\mu H \ll k_B T$)].

Quantum theory of diamagnetism:

Hamiltonian for an electron in a magnetic field→Starting from Lorentz equation for the force on an electron moving in a combined electric and magnetic field

$$F = -e\boldsymbol{E} - \left(\frac{e}{c}\right)\boldsymbol{v} \times \boldsymbol{H}$$

Introducing the vector potential, **A** by means of the relation H = ∇ x A

Thus the Hamiltonian is given by –

$$\mathcal{H} = \text{K.E.} + \text{P.E.} = \frac{1}{2}mv^2 + V$$

Here, the spin of the electron will be neglected.

Let

$$\boldsymbol{p} = -i\hbar\boldsymbol{\nabla}$$

And

$$\nabla = \hat{\imath}\,\frac{\partial}{\partial x} + \hat{\jmath}\,\frac{\partial}{\partial y} + \hat{k}\,\frac{\partial}{\partial z}$$

$$\mathcal{H} = \frac{(-\,i\,\hbar\nabla)^2}{2m} + V = -\frac{(\hbar\nabla)^2}{2m} + \mathrm{V}$$

In the absence of the magnetic field moment p is given by (p = mv) and in the presence

$$\boldsymbol{p} = m\boldsymbol{v} + \frac{Q}{C}\boldsymbol{A}$$

$$m\boldsymbol{v} = \boldsymbol{p} - \frac{Q}{C}\boldsymbol{A} = [-\,i\,\hbar\boldsymbol{\nabla} + \frac{e}{C}\boldsymbol{A}]$$

$$\mathcal{H} = \frac{1}{2m}\{\left[\frac{e}{C}\boldsymbol{A}\right]^2 + h^2\boldsymbol{\nabla}^2 - i\,\hbar\frac{e}{C}[\boldsymbol{A}.\boldsymbol{\nabla} + \boldsymbol{\nabla}.\mathbf{A}]\} + V$$

$$\mathcal{H} = -\frac{h^2\boldsymbol{\nabla}^2}{\mathbf{2m}} - \underbrace{\frac{i\hbar e}{2mC}[\boldsymbol{A}.\boldsymbol{\nabla} + \boldsymbol{\nabla}.\mathbf{A}]}_{\text{Paramagnetic Contribution}} + \underbrace{\frac{1}{2m}\left[\frac{e}{C}\boldsymbol{A}\right]^2}_{\text{Diamagnetic Contribution}} + V$$

Where V is the potential energy, thus if we take:

$$A_x = -\frac{1}{2}yH,\ A_y = \frac{1}{2}xH\ and\ A_z = 0$$

Then

$$H_x = H_y = 0\ and\ H = H_z$$

$$\mathrm{H} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{1}{2}yH & \frac{1}{2}xH & 0 \end{vmatrix} = H_z$$

Now solving for –

$$\boldsymbol{A}.\boldsymbol{\nabla} + \boldsymbol{\nabla}.\mathbf{A} = -\frac{1}{2}yH\frac{\partial}{\partial x} + \frac{1}{2}xH\frac{\partial}{\partial y}$$

Since

$$\boldsymbol{p} = -\,i\,\hbar\boldsymbol{\nabla}$$

And

$$\nabla = \hat{\imath}\,\frac{\partial}{\partial x} + \hat{\jmath}\,\frac{\partial}{\partial y} + \hat{k}\,\frac{\partial}{\partial z}$$

Therefore,

$$\boldsymbol{p_x} = -\,i\,\hbar\frac{\partial}{\partial x} \quad \text{and} \quad \boldsymbol{p_y} = -\,i\,\hbar\frac{\partial}{\partial y}$$

$$\boldsymbol{A}.\boldsymbol{\nabla} + \boldsymbol{\nabla}.\mathbf{A} = \frac{\mathrm{Hi}}{2\hbar}(xp_y - yp_x)$$

Therefore, for a magnetic field in z direction Hamiltonian becomes

$$\mathcal{H} = -\frac{h^2\nabla^2}{2\mathbf{m}} + \frac{eH}{4mC}(xp_y - yp_x) + \frac{1}{2m}\left[\frac{e}{C}\boldsymbol{A}\right]^2 + V$$

Rom the definition of angular momentum –

L = r x p

$$L_z = xp_y - yp_x$$

And

$$\mu_z = -\frac{e}{2mc}L_z = -\frac{e}{2mc}(xp_y - yp_x)$$

$$\mathcal{H} = -\frac{h^2}{2\mathrm{m}}\nabla^2 - \frac{1}{2}H\mu_z + \frac{e^2}{2mc^2}A^2 + V$$

$$\mathcal{H} = -\frac{h^2}{2\mathrm{m}}\nabla^2 - \frac{1}{2}\boldsymbol{\mu}.\boldsymbol{H} + \frac{e^2}{2mc^2}A^2 + V$$

The second term contributes to the permanent dipole i.e. paramagnetic contribution.

Now, considering the third term –

$$\frac{e^2}{2mc^2}A^2 = \frac{e^2}{2mc^2}({A_x}^2 + {A_y}^2)$$

Substitution the values of Ax and Ay in this from equation

$$\frac{e^2}{2mc^2}A^2 = \frac{e^2}{2mc^2}(< x^2 > + < y^2 >)\frac{H^2}{4}$$

$$\frac{e^2}{2mc^2}A^2 = \frac{e^2}{2mc^2}(< \rho^2 >)\frac{H^2}{4}$$

$$< \rho^2 > = \frac{2}{3} < r^2 >$$

Suppose we had written the H for the electrons associated with unit volume of substance containing N atoms and each atom containing Z electrons.

$$\frac{e^2}{2mc^2}A^2 NZ = \frac{e^2H^2}{12mc^2} < r^2 > NZ$$

Now if the magnetic field induces a dipole moment in the material, the corresponding energy term should be quadratic in H. This equation (23) may be considered the energy term associated with the diamagnetism of solid.

$$-\frac{1}{2}\chi H^2 = \frac{e^2H^2}{12mc^2} < r^2 > NZ$$

$$\chi_{dia} = -\frac{e^2H^2}{6mc^2} < r^2 > NZ$$

Where $< r^2 >$ represents the mean square distance of the electrons relative to the nucleus.

SAQ.1

a) What do you mean diamagnetic and paramagnetic materials with examples?
b) Define the Curie law.
c) What do you mean by Curie Wiess law?
d) Discuss and explain in brief classical theory of diamagnetism.
e) A paramagnetic material has FCC structure with a cubic edge of 3.5 A°. If the saturation value of magnetization is 2.8×10^{6} A m^{-1}, Calculate the magnetization contributed per atom in Bohr magnetrons.

4.5 Ferromagnetism:

When you think of magnetic materials, you probably think of iron, nickel or magnetite. Unlike paramagnetic materials, the atomic moments in these materials exhibit very strong interactions. These interactions are produced by electronic exchange forces and result in a parallel or antiparallel alignment of atomic moments. Exchange forces are very large, equivalent to a field on the order of 1000 Tesla, or approximately a 100 million times the strength of the earth's field.

The exchange force is a quantum mechanical phenomenon due to the relative orientation of the spins of two electrons.

Ferromagnetic materials exhibit parallel alignment of moments resulting in large net magnetization even in the absence of a magnetic field.

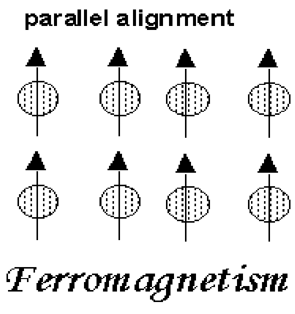


Fig.4.9 Ferromagnetism

The elements Fe, Ni, and Co and many of their alloys are typical ferromagnetic materials.
Two distinct characteristics of ferromagnetic materials are their
(1) Spontaneous magnetization and the existence of

(2) Magnetic ordering temperature

Spontaneous Magnetization:

The spontaneous magnetization is the net magnetization that exists inside a uniformly magnetized microscopic volume in the absence of a field. The magnitude of this magnetization, at 0 K, is dependent on the spin magnetic moments of electrons.

A related term is the saturation magnetization which we can measure in the laboratory. The saturation magnetization is the maximum induced magnetic moment that can be obtained in a magnetic field (H_{sat}); beyond this field no further increase in magnetization occurs.

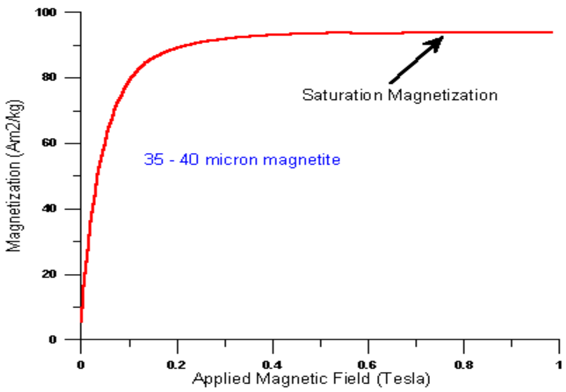


Fig.4.10 Curve for increase in magnetization

The difference between spontaneous magnetization and the saturation magnetization has to do with magnetic domains (more about domains later). Saturation magnetization is an intrinsic property, independent of particle size but dependent on temperature.

There is a big difference between paramagnetic and ferromagnetic susceptibility. As compared to paramagnetic materials, the magnetization in ferromagnetic materials is saturated in moderate magnetic fields and at high (room-temperature) temperatures:

	H_{sat} Tesla	T range (K)	$10^{-8}m^3/kg$
paramagnets	>10	<<100	~50
ferromagnets	~1	~300	1000-10000

Curie temperature:

Even though electronic exchange forces in ferromagnets are very large, thermal energy eventually overcomes the exchange and produces a randomizing effect. This occurs at a particular temperature called the Curie temperature (T_C). Below the Curie temperature, the ferromagnet is ordered and above it, disordered. The saturation magnetization goes to zero at the Curie temperature. A typical plot of magnetization vs temperature for magnetite is shown below.

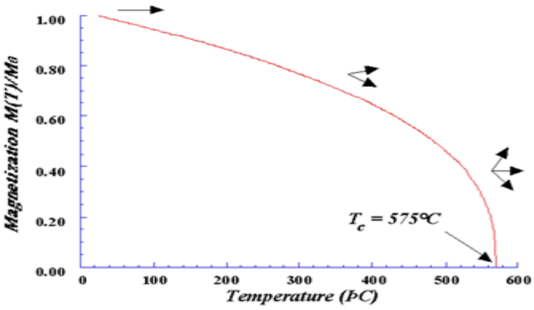


Fig.4.11 A typical plot of magnetization vs temperature for magnetite

The Curie temperature is also an intrinsic property and is a diagnostic parameter that can be used for mineral identification. However, it is not foolproof because different magnetic minerals, in principle, can have the same Curie temperature.

Hysteresis:

In addition to the Curie temperature and saturation magnetization, ferromagnets can retain a memory of an applied field once it is removed. This behavior is called hysteresis and a plot of the variation of magnetization with magnetic field is called a hysteresis loop.

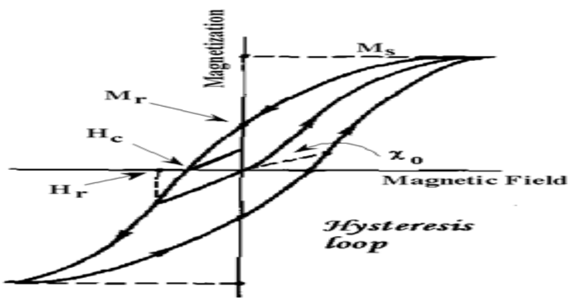


Fig.4.12 Plot of the variation of magnetization with magnetic field

Another hysteresis property is the coercivity of remanence (Hr). This is the reverse field which, when applied and then removed, reduces the saturation remanence to zero. It is always larger than the coercive force.

The initial susceptibility is the magnetization observed in low fields, on the order of the earth's field (50-100 KT).

The various hysteresis parameters are not solely intrinsic properties but are dependent on grain size, domain state, stresses, and temperature. Because hysteresis parameters are dependent on grain size, they are useful for magnetic grain sizing of natural samples.

Ferrimagnetism:

In ionic compounds, such as oxides, more complex forms of magnetic ordering can occur as a result of the crystal structure. One type of magnetic ordering is call ferrimagnetism. A simple representation of the magnetic spins in a ferrimagnetic oxide is shown here.

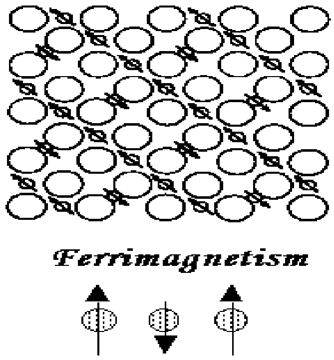


Fig.4.13 Ferrimagnetic oxide

The magnetic structure is composed of two magnetic sublattices (called A and B) separated by oxygens. The exchange interactions are mediated by the oxygen anions. When this happens, the interactions are called indirect or superexchange interactions. The strongest superexchange interactions result in an antiparallel alignment of spins between the A and B sublattice.

In ferrimagnets, the magnetic moments of the A and B sublattices are not equal and result in a net magnetic moment. Ferrimagnetism is therefore similar to ferromagnetism. It exhibits all the hallmarks of ferromagnetic behavior- spontaneous magnetization, Curie temperatures, hysteresis, and remanence. However, ferro- and ferrimagnets have very different magnetic ordering.

Magnetite is a well known ferrimagnetic material. Indeed, magnetite was considered a ferromagnet until Néel in the 1940's, provided the theoretical framework for understanding ferrimagnetism.

Crystal Structure of Magnetite:

Magnetite, Fe3O4 crystallizes with the spinel structure. The large oxygen ions are close packed
in a cubic arrangement and the smaller Fe ions fill in the gaps.

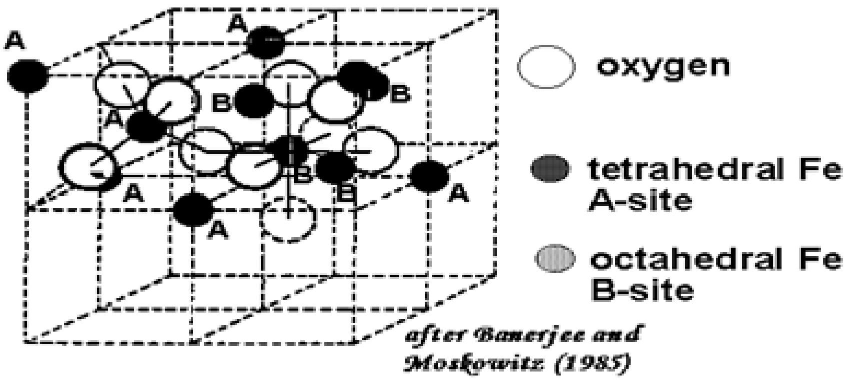


Fig.4.14 Crystal Structure of Magnetite

The gaps come in two flavors:

Tetrahedral site: Fe ion is surrounded by four oxygens

Octahedral site: Fe ion is surrounded by six oxygens

The tetrahedral and octahedral sites form the two magnetic sublattices, A and B respectively. The spins on the A sublattice are antiparallel to those on the B sublattice. The two crystal sites are very different and result in complex forms of exchange interactions of the iron ions between and within the two types of sites.

The structural formula for magnetite is

$$[Fe^{3+}]A\ [Fe^{3+},Fe^{2+}]B\ O_4$$

This particular arrangement of cations on the A and B sublattice is called an inverse spinel structure. With negative AB exchange interactions, the net magnetic moment of magnetite is due to the B-site Fe^{2+}.

Anti-ferromagnetism:

If the A and B sub-lattice moments are exactly equal but opposite, the net moment is zero. This type of magnetic ordering is called anti-ferromagnetism.

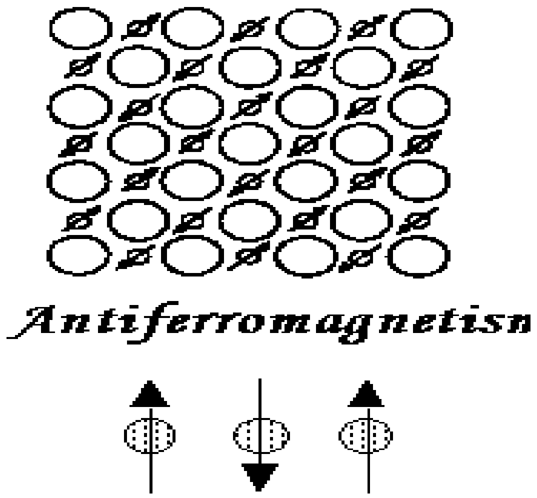


Fig.4.15 Anti-ferromagnetism

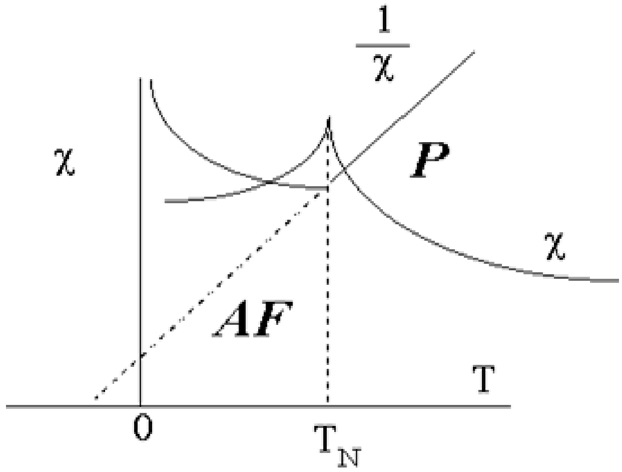


Fig.4.16 Curve for Anti-ferromagnetism

The clue to anti-ferromagnetism is the behavior of susceptibility above a critical temperature, called the Neel temperature (T_N). Above T_N, the susceptibility obeys the Curie-Weiss law for paramagnets but with a negative intercept indicating negative exchange interactions.

Crystal Structure of Hematite:

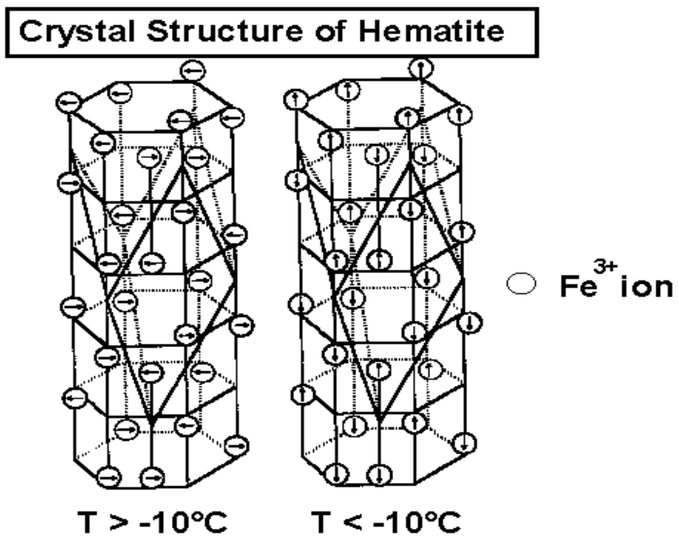


Fig.4.17 Crystal Structure of Hematite

Hematite crystallizes in the corundum structure with oxygen ions in an hexagonal close packed framework. The magnetic moments of the Fe^{3+} ions are ferromagnetically coupled within specific c-planes, but antiferromagnetically coupled between the planes.

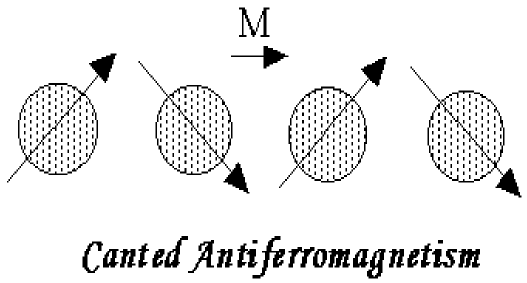


Fig.4.18 Canters Anti-ferromagnetism

Above -10°C, the spin moments lie in the c-plan but are slightly canted. This produces a weak spontaneous magnetization within the c-plan (0.4 Am^2/kg).

Below -10°C, the direction of the antiferromagnetism changes and becomes parallel to the c-axis; there is no spin canting and hematite becomes a perfect antiferromagnet.

This spin-flop transition is called the Morin transition.

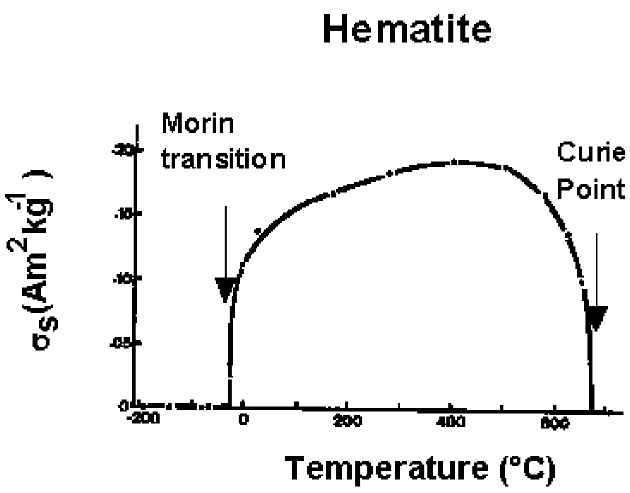


Fig.4.19 Curve for Hematite

Magnetic Properties of Minerals:

Mineral	Composition	Magnetic Order	T_c(°C)	□$_s$ (Am^2/kg)
Oxides				
Magnetite	Fe_3O_4	ferrimagnetic	575-585	90-92
Ulvospinel	Fe_2TiO_2	AFM	-153	
Hematite	αFe_2O_3	canted AFM	675	0.4
Ilmenite	$FeTiO_2$	AFM	-233	
Maghemite	γFe_2O_3	ferrimagnetic	~600	~80
Jacobsite	$MnFe_2O_4$	ferrimagnetic	300	77
Trevorite	$NiFe_2O_4$	ferrimagnetic	585	51
Magnesioferrite	$MgFe_2O_4$	ferrimagnetic	440	21
Sulfides				
Pyrrhotite	Fe_7S_8	ferrimagnetic	320	~20
Greigite	Fe_3S_4	ferrimagnetic	~333	~25
Troilite	FeS	AFM	305	
Oxyhydroxides				
Goethite	αFeOOH	AFM, weak FM	~120	<1

Lepidocrocite	$\gamma FeOOH$	AFM(?)	-196	
Feroxyhyte	$FeOOH$	ferrimagnetic	~180	<10
Metals & Alloys				
Iron	Fe	FM	770	
Nickel	Ni	FM	358	55
Cobalt	Co	FM	1131	161
Awaruite	Ni_3Fe	FM	620	120
Wairauite	$CoFe$	FM	986	235

FM = ferromagnetic order

AFM = antiferromagnetic order

T_c = Curie or Néel Temperature

σ_s = saturation magnetization at room-temperature,

Difference between Ferromagnetism and Antiferromagnetism:

Ferromagnetism vs Antiferromagnetism	
Ferromagnetism is the presence of magnetic domains that are aligned in the same direction in magnetic materials.	Antiferromagnetism is the presence of magnetic domains that are aligned in opposite directions in magnetic materials.
Alignment of Magnetic Domains	
The magnetic domains of ferromagnetic materials are aligned in the same direction.	The magnetic domains of antiferromagnetic materials are aligned in opposite directions.
Net Magnetic Moment	
Ferromagnetic materials have a value for net magnetic moment.	Antiferromagnetic materials have a zero net magnetic moment.
Examples	
Examples of ferromagnetic materials include metals such as iron, nickel, cobalt and their metal alloys.	Examples of antiferromagnetic materials include transition metal oxides.

4.6 Superconductivity:

Superconductivity was discovered by Dutch Physicist Heike Kamerlingh Onnes in 1911 in Leiden. He was awarded the Nobel Prize in Physics in 1913 for his low-temperature research. Some materials when they are cooled, below certain temperature their resistivity get abolished means they exhibit the infinite conductivity.

The property/ phenomenon of infinite conductivity in materials is called superconductivity. The temperature, at which the metals change from normal conducting state to superconducting state, is called critical temperature/transition temperature. An example of superconductors is Mercury. It becomes superconductor at 4k. In superconducting state the materials expel the magnetic field. A transition curve for mercury is shown in figure below-

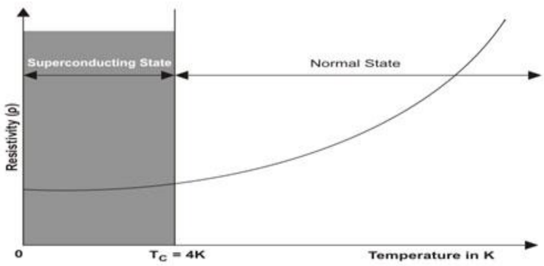


Fig.4.20 A transition curve for mercury

The transition from normal conducting state to superconducting state is reversible. Moreover, below critical temperature the superconductivity can be abolished either by passing sufficient large current through conductor itself or by applying sufficient strong external magnetic field. Below critical temperature/transition temperature, the value of current through conductor itself at which the superconducting state abolished is called critical current. As the temperature (below the critical temperature) reduces the value of critical current increase. The value of critical current increase with decrease in temperature. The value of critical magnetic field also depends on temperature. As the temperature (below the critical temperature) reduces the value of critical magnetic field increase.

Characteristics of Superconductivity:

1. Superconductivity is found to occur in metallic elements in which the number of valence electron lies between 2 and 4.
2. Materials having high normal resistivities exhibit superconductivity.
3. Superconductivity is also favored by small atomic volume accompanied by a small atomic mass.
4. The transition temperature (Tc) is different for different substances.

5. Ferromagnetic and antiferromagnetic materials are not superconductors.
6. The electrical resistivity drops to zero.
7. The magnetic flux lines are expelled from the material.
8. There is a discontinuous change in the specific heat.
9. Further, there is some small changes in the thermal conductivity and the volume of the
   materials.

Properties of Superconductivity:

The superconducting material shows some extraordinary properties which make them very important for modern technology. The research is still going on to understand and utilize these extraordinary properties of superconductors in various fields of technology. Such properties of superconductors are listed below-

- Zero Electric Resistance (Infinite Conductivity)
- Meissner Effect: Expulsion of magnetic field
- Critical Temperature/Transition Temperature
- Critical Magnetic Field
- Persistent Currents
- Josephson Currents
- Critical Current

Zero Electric Resistance or Infinite Conductivity:

In Superconducting state, the superconducting material shows the zero electric resistance (infinite conductivity). When the sample of a superconducting material is cooled below its critical temperature/transition temperature, its resistance reduces suddenly to zero. For example Mercury shows zero resistance below 4k.

Meissner Effect (Expulsion of Magnetic Field):

A Superconductor, when it is cooled below the critical temperature Tc), expel the magnetic field and doesn't allow the magnetic field to penetrate inside it. This phenomenon in superconductors is called Meissner effect.

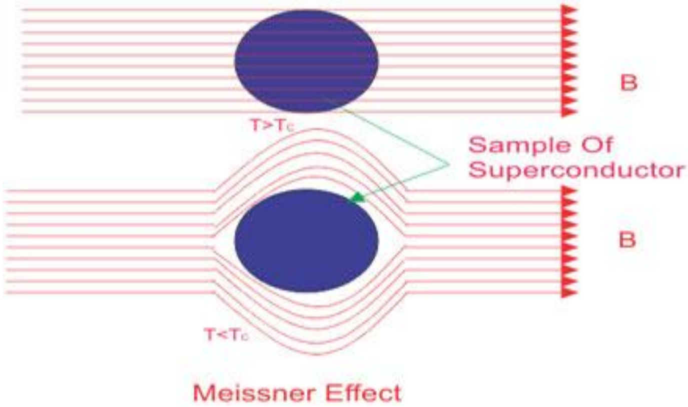


Fig.4.21 Meissner effect (Expulsion of Magnetic Field)

Critical Temperature/Transition Temperature:

Critical temperature of a superconducting material is the temperature at which the materials changes from normal conducting state to superconducting state. This transition from normal conducting state (phase) to superconducting state (phase) is sudden / sharp and complete. The transition of mercury from normal conducting state to superconducting state is shown in figure.

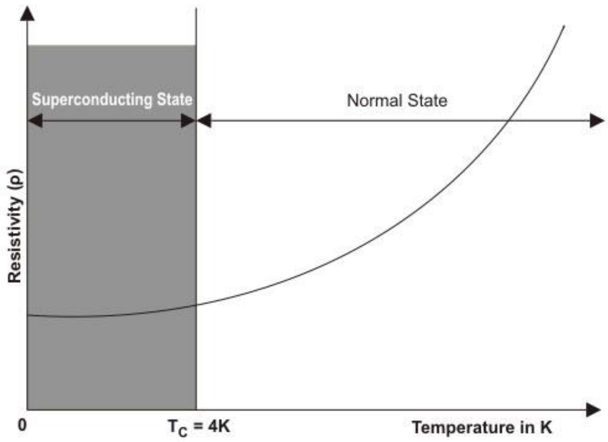


Fig4.22 Transition curve of mercury from normal conducting state to superconducting state

Critical Magnetic Field:

The superconducting state / phase, of a superconducting material, breaks when the magnetic field (either external or produced by current flowing superconductor itself) increases beyond a certain value and sample starts behaving like an ordinary conductor. This certain value of magnetic field beyond which superconductor returns back to ordinary state, is called Critical magnetic field. The value of critical magnetic field depends on temperature. As the temperature (below the critical temperature) reduces the value of critical magnetic field increase. The variation in critical magnetic field with temperature is shown in figure below-

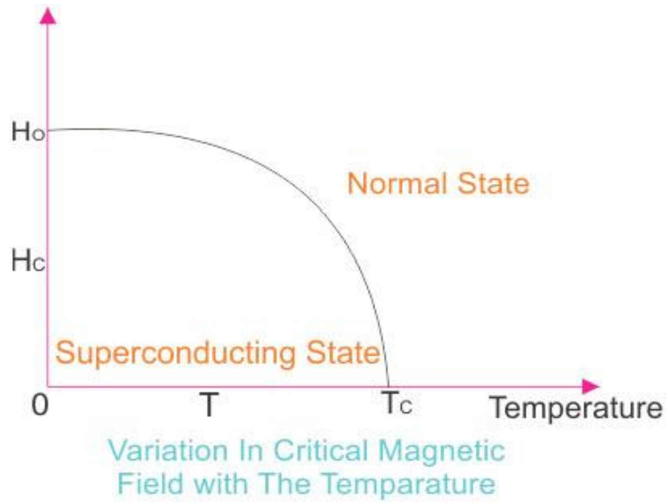


Fig.4.23 Variation in critical magnetic field with temperature

Persistent Current:

If a ring made of a superconductor is placed in a magnetic field above its critical temperature, now cool the ring of superconductor below its critical temperature and now if we remove the magnetic field a current is induced in ring due to its self-inductance. By Lenz law the direction of this induced current is such that it opposes the change in flux passing through the ring. As the ring is in superconducting state (zero resistance), the current induced is ring will be continue to flow this current is called the persistent current. This persistent current produces a magnetic flux which makes the magnetic flux passing through the ring constant.

Josephson Current:

If two superconductors are separated by a thin film of insulating material, which forms a low resistance junction, it is found that the cooper pairs (formed by phonon interaction) of electrons, can tunnel from one side of junction to the other side. The current, due to flow of such cooper pairs, is called Josephson Current.

Critical Current:

When a current is passed through a conductor under superconducting state, a magnetic field is developed. If the current increase beyond certain value the magnetic field increased up to critical value at which conductor returns to its normal state. This value of current is called critical current.

Applications of Superconductivity:

In modern field of technology the superconductivity is widely used in different fields of technology. Some of these applications are listed below-

1. Medical: MRI (Magnetic Resonance Imaging), Ultra-Low Field Magnetic Resonance Imaging (ULF-MRI), Magneto-encephalography (MEG) and Magnetic Source Imaging (MSI), Magneto- cardiography (MCG) etc.
2. Electric field: Generators, motors, transformers, relays, magnetic energy storages (SMES), superconducting magnets, HTS Induction Heater, Fusion etc.
3. Electronics: SQUIDS (superconducting quantum interference device), High Speed computing, Quantum computing, Sensors, filters, circuitry, radar etc.
4. Transportation: Magnetically levitated trains, Marine Propulsion (magneto-hydrodynamic), Marine Propulsion (motor) etc.
5. Physics: Particle Accelerators, Magnets, Plasma / fusion research etc.

Magnetic behavior of superconductor:

An extreme example of a diamagnet is a superconductor. These materials are known primarily through their electrical properties - at some relatively low temperature their electrical resistance is exactly zero. Thus, one can establish a current in a superconductor and it will never die away due to resistance, even when the source of potential difference that started the current is removed. Superconductors also have interesting magnetic properties; they are perfect diamagnets: when an applied magnetic field is applied, eddy currents in the superconductor induce a magnetic field which exactly cancels the applied magnetic field. This is the Meissner effect.

This effect is responsible for the magnetic levitation of a magnet when placed above a superconductor. In this device, a magnetic field is generated from an electromagnet, which causes eddy currents to be produced. The magnetic fields from the induced currents are in turn picked up by the detector in the form of small currents being produced. Most diamagnetic materials are metals, which have good electrical conductivity properties and so the eddy currents can be relatively easily established. This is the reason these detectors can readily sense metallic objects but not plastics or other poor conductors.

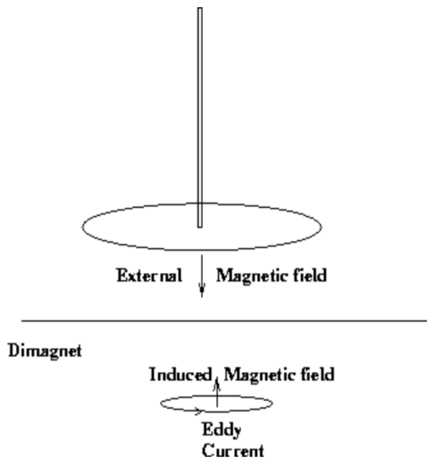


Fig.4.24 Metal detector / magnetic levitation

Suppose, as in Fig. 4.24, we place a magnet above a superconductor. The induced magnetic field inside the superconductor is exactly equal and opposite in direction to the applied magnetic field, so that they cancel within the superconductor. What we then have are two magnets equal in strength with poles of the same type facing each other. These poles will repel each other, and the force of repulsion is enough to float the magnet. Such magnetic levitation devices are being tried on train tracks in Japan; if successful, this would make train travel much faster, smoother, and more efficient due to the lack of friction between the tracks and train (in some cases, rather than superconductors, strong electromagnets are used to provide the magnetic levitation).

Despite these interesting properties, superconductors are not widely used in today's world, outside of as electromagnets to generate strong magnetic fields in certain medical diagnostic devices and in particle accelerators. The reason for this is that superconductors exist only below a certain critical temperature, and above that temperature they behave like normal materials. When first discovered these critical temperatures were of the order of 10 K (about -260° C), which was (and still is) fairly difficult to reach (this is about the temperature at which helium liquefies). However, recently high temperature superconductors have been discovered which have critical temperatures of the order of 100 K and above (about -170° C). This is about the temperature that nitrogen liquefies, and is relatively easy to reach with today's technology - ``dry ice'' is liquid carbon dioxide at this temperature. These developments has spurred research into other uses of superconductors such as in magnetic levitation devices and as circuit elements in computers to increase speed by cutting down on resistance.

4.7 Meissner effect:

The Meissner effect (or Meissner–Ochsenfeld effect) is the expulsion of a magnetic field from a superconductor during its transition to the superconducting state when it is cooled below the critical temperature. The German physicists Walther Meissner and Robert Ochsenfeld discovered this phenomenon in 1933 by measuring the magnetic field distribution outside superconducting tin and lead samples. The samples, in the presence of an applied magnetic field, were cooled below their superconducting transition temperature, whereupon the samples cancelled nearly all interior magnetic fields. They detected this effect only indirectly because the magnetic flux is conserved by a superconductor: when the interior field decreases, the exterior field increases. The experiment demonstrated for the first time that superconductors were more than just perfect conductors and provided a uniquely defining property of the superconductor state. The ability for the expulsion effect is determined by the nature of equilibrium formed by the neutralization within the unit cell of a superconductor.

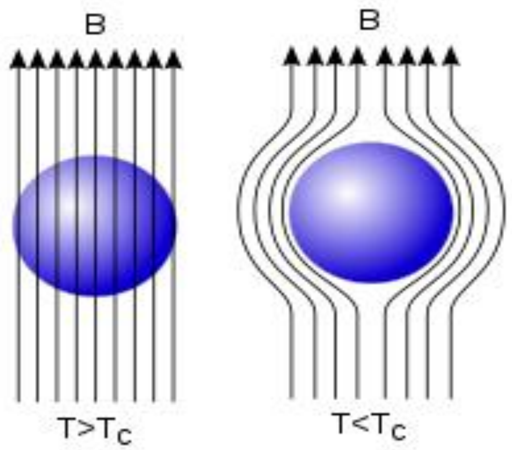


Fig.4.25 Diagram of the Meissner effect

Magnetic field lines, represented as arrows, are excluded from a superconductor when it is below its critical temperature.

A superconductor with little or no magnetic field within it is said to be in the Meissner state. The Meissner state breaks down when the applied magnetic field is too strong. Superconductors can be divided into two classes according to how this breakdown occurs.

In type-I superconductors, superconductivity is abruptly destroyed when the strength of the applied field rises above a critical value H_c. Depending on the geometry of the sample, one may obtain an intermediate state consisting of a baroque pattern[4] of regions of normal material carrying a magnetic field mixed with regions of superconducting material containing no field.

In Type-II Superconductors, raising the applied field past a critical value H_{c1} leads to a mixed state (also known as the vortex state) in which an increasing amount of magnetic flux penetrates the material, but there remains no resistance to the electric current as long as the current is not too large. At second critical field strength H_{c2}, superconductivity is destroyed. The mixed state is caused by vortices in the electronic superfluid, sometimes called fluxons because the flux carried by these vortices is quantized. Most pure elemental superconductors, except niobium and carbon nanotubes, are type I, while almost all impure and compound superconductors are type II.

The Meissner effect was given a phenomenological explanation by the brothers Fritz and Heinz London, who showed that the electromagnetic free energy in a superconductor is minimized provided

$$\nabla^2\mathbf{H} = \lambda^{-2}\mathbf{H}$$

where H is the magnetic field and λ is the London penetration depth.

This equation, known as the London equation, predicts that the magnetic field in a superconductor decays exponentially from whatever value it possesses at the surface. This exclusion of magnetic field is a manifestation of the super diamagnetism emerged during the phase transition from conductor to superconductor, for example by reducing the temperature below critical temperature.

In a weak applied field (less than the critical field that breaks down the superconducting phase), a superconductor expels nearly all magnetic flux by setting up electric currents near its surface, as the magnetic field H induces magnetization M within the London penetration depth from the surface. These surface currents shields the internal bulk of the superconductor from the external applied field. As the field expulsion, or cancellation, does not change with time, the currents producing this effect (called persistent currents or screening currents) do not decay with time.

Near the surface, within the London penetration depth, the magnetic field is not completely canceled. Each superconducting material has its own characteristic penetration depth.

Any perfect conductor will prevent any change to magnetic flux passing through its surface due to ordinary electromagnetic induction at zero resistance. However, the Meissner effect is distinct from this: when an ordinary conductor is cooled so that it makes the transition to a superconducting state in the presence of a constant applied magnetic field, the magnetic flux is expelled during the transition. This effect cannot be explained by infinite conductivity, but only by the London equation. The placement and subsequent levitation of a magnet above an already superconducting material do not demonstrate the Meissner effect, while an initially stationary magnet later being repelled by a superconductor as it is cooled below its critical temperature does.

The persisting currents that exist in the superconductor to expel the magnetic field is commonly misconceived as a result of Lenz's Law or Faraday's Law. A reason this is not the case is that no change in flux was made to induce the current. Another explanation is that since the superconductor experiences zero resistance, there cannot be an induced emf in the superconductor.

BCS theory (qualitative):

The microscopic theory of superconductivity developed by J. Bardeen, L.N. Cooper and J.R. Scriffer in 1957, successfully explained the effect like zero resistivity, Meissner effect etc. this theory is known as BCS theory.

Principle:

This theory states that the electrons experience a special kind of attractive interaction, overcoming the coulomb forces of repulsion between them; as a result cooper pairs (i.e) electro pair is formed. At low temperature, these pairs move without any restriction through the lattice points and the material becomes superconductor. Here the electron-lattice-electrons interaction should be stronger than electrons-electros interaction.

Important features of BCS theory:

1. Electrons form pairs (called cooper pair) which propagate throughout the lattice.
2. The propagation of cooper pairs is without resistance because the electrons move in resonance with phonons.

Electron-lattice-electron interaction:

When an electron (1st) moves through the lattice, it will be attracted by the core (+ve charge) of the lattice. Due to this attraction, ion core is disturbed and it is called as lattice distortion. The lattice vibrations are quantized in terms of phonons.

The deformation produces a region of increased positive charge. Thus if another electron (2nd) moves through this region as shown in fig. it will be attracted by the greater concentration of positive charge and hence the energy of the 2nd electron is lowered.

Hence two electrons interact through the lattice or the phonons field resulting in lowering the energy of electrons. This lowering of energy implies that the force between the two electrons is attractive. This type of interaction is called electrons-lattice electron interaction. The interaction is strong only when the two electrons have equal and opposite momenta and spins.

Explanation:

Consider the 1st electron with wave vector k distorts the lattice, here by emitting phonons of wave vector q. This results in the wave vector k-q for the 1st electron. now if the 2nd electron with wave vector k', seeks the lattice it takes up the energy from the lattice and its wave vector changes k'+q as shown in fig. two electrons with wave vectors k-q and k'+q form a pair known as cooper pair.

Cooper pair:

The pair of electrons formed due to electron-lattice (phonons)-electron interaction (force of attraction) by overcoming the electron-electron interaction (force of repulsion), with equal and opposite momentum and spins (i.e.) with wave vector k-q and k'+q, are called cooper pair.

Coherence length:

In the electron-lattice-electron interaction, the electrons will not be fixed, they move in opposite directions and their co-relations may persist over lengths of maximum 10-6m. This length is called coherence length.

→BCS theory hold good only for low temperature superconductivity.

SAQ.2

a) What do you mean by ferromagnetism, anti-ferromagetism and ferrimagnetism?
b) Define the Superconductivity and write its characteristics.
c) Discuss about the magnetic behavior of superconductor.
d) What do you mean by Meisener's effect?
e) Define the BCS theory.
f) A superconducting tin has a critical temperature of 4.2K at zero magnetic field and a critical field of 0.0306 Tesla at 0 K. Find the critical field at 3K.

4.8 Types of superconductors:

There are two types of super conductors based on their variation in magnetization, due to external magnetic field applied.

- Type I superconductor or soft superconductor
- Type II superconductor or hard superconductor

Type I Superconductor:

When the super conductor is kept in the magnetic field and if the field is increased the superconductor becomes normal conductor abruptly at critical magnetic field as shown in fig. These types of materials are termed as Type I superconductors.

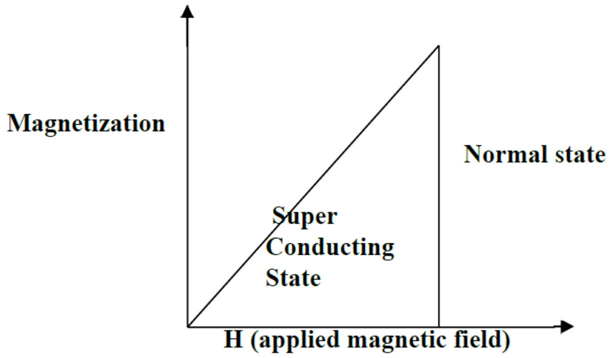


Fig.4.26 Waveform of Type I superconductors

Below critical field, the specimen excludes all the magnetic lines of force and exhibit perfect Meissner effect. Hence, Type I superconductors are perfect diamagnet, represented by negative sign in magnetization.

Type II Superconductors:

When the super conductor kept in the magnetic field and if the field is increased, below the lower critical field Hc1, the material exhibit perfect diamagnetism (i.e.) it behaves as a super conductor and above Hc1, the magnetization decreases and hence the magnetic flux starts penetrating through the material. The specimen is said to be in a mixed state between Hc1 and Hc2. above Hc2 (upper critical field) it becomes normal conductor as shown in figure below:

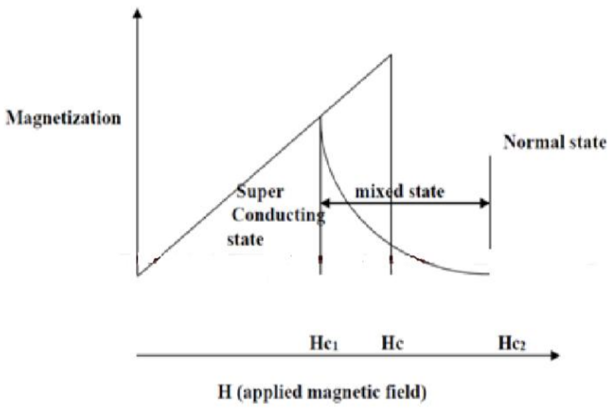


Fig.4.27 Waveform of Type II superconductor

The materials which lose its superconducting property gradually due to increase on the magnetic field are called Type II superconductor**.**

Difference between Type – I and Type – II Superconductors:

S.No	Type – I Superconductors	Type – II Superconductors
1.	The material loses magnetization suddenly.	The material loses magnetization gradually.
2.	They exhibit complete Meissner effect i.e., they are completely diamagnetic.	They do not exhibit complete Meissner effect.
3.	There is only one critical magnetic field (HC).	There are two critical magnetic fields i.e., lower critical field (HC_1) and upper critical field (HC2).
4.	No mixed state exists.	Mixed state is present.

Difference between Semiconductor and Superconductor:

Semiconductor	**Superconductor**
The resistivity of semiconductor is finite	The resistivity of a superconductor is zero electrical resistivity
In this, electron repulsion leads to finite resistivity.	In this, electron attraction leads to the loss of resistivity
Superconductors do not show perfect diamagnetism	Superconductors show perfect diamagnetism
The energy gap of a superconductor is the order of a few eV.	The energy gap of superconductors is of the order of 10^-4 eV.
Flux quantization in superconductors is 2e units	The unit of a superconductor is e.

Properties of Superconductor:

The superconducting materials show some amazing properties which are essential for current technology. The research on these properties is still going on to recognize and utilize these properties in various fields which are listed below.

1. **Infinite Conductivity/ Zero Electric Resistance:** In the Superconducting condition, the superconducting material illustrates the zero electric resistance. When the material is cooled under its transition temperature, then its resistance will be reduced to zero suddenly. For instance, Mercury shows zero resistance under 4k.

2. **Meissner Effect:** When a superconductor is cooled under the critical temperature, then it doesn't permit the magnetic field to go through in it. This occurrence in superconductors is known as the Meissner effect.
3. **Transition Temperature:** This temperature is also known as critical temperature. When the critical temperature of a superconducting material is changing the conducting state from normal to superconducting.
4. **Josephson Current:** If the two superconductors are divided with the help of thin-film in insulating material, then it forms a junction of low resistance to found the electrons with copper pair. It can tunnel from one surface of the junction to the other surface. So the current because of the flow of cooper pairs is known as Josephson Current.
5. **Critical Current:** When the current supplied through a conductor under the condition of superconducting, then a magnetic field can be developed. If the current flow increases beyond a certain rate then the magnetic field can be enhanced, this is equivalent to the critical value of the conductor at which this returns to its usual condition. The flow of current value is known as the critical current.
6. **Persistent Currents:** If a superconductor ring is arranged in a magnetic field above its critical temperature, at the present cool the superconductor ring under its critical temperature. If we eliminate this field, then the flow of current can be induced within the ring because of its self-inductance. From Lenz law, the induced current opposes the change within flux that flows through the ring. When the ring is placed in a superconducting condition, then the flow of current will be induced to continue the flow of current is named as the persistent current. This current generates a magnetic flux to make the flux flowing throughout the constant ring.

Applications of Super-Conductor:

The applications of superconductors include the following:

- These are used in generators, particle accelerators, transportation, electric motors, computing, medical, power transmission, etc.
- Superconductors mainly used for creating powerful electromagnets in MRI scanners. So these are used to divide. They can also be used to separate magnetic and non-magnetic materials
- This conductor is used to transmit power for long distances
- Used in memory or storage elements.

4.9 Josephson Effect:

The Josephson Effect is the phenomenon of super current, a current that flows continuously without any voltage applied, across a device known as a Josephson junction (JJ), which consists of two or more superconductors coupled by a weak link. The weak link can consist of a thin insulating barrier (known as a superconductor–insulator–superconductor junction, or S-I-S), a short section of non-superconducting metal (S-N-S), or a physical constriction that weakens the superconductivity at the point of contact (S-s-S).

The Josephson Equations:

A diagram of a single Josephson junction is shown at right. Assume that superconductor A has Ginzburg–Landau order parameter

$$\psi_A = \sqrt{n_A}e^{i\phi_A}$$

and superconductor

$$\text{B}\ \psi_B = \sqrt{n_B}e^{i\phi_B}$$

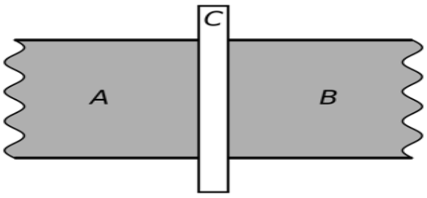


Fig.4.28 Diagram of a single Josephson junction. A and B represent superconductors, and C the weak link between them

which can be interpreted as the wave functions of Cooper pairs in the two superconductors. If the electric potential difference across the junction is V, then the energy difference between the two superconductors is 2eV, since each Cooper pair has twice the charge of one electron. The Schrödinger equation for this two-state quantum system is therefore:

$$i\hbar\frac{\partial}{\partial t}\begin{pmatrix}\sqrt{n_A}e^{i\phi_A}\\ \sqrt{n_B}e^{i\phi_B}\end{pmatrix} = \begin{pmatrix}eV & K\\ K & -eV\end{pmatrix}\begin{pmatrix}\sqrt{n_A}e^{i\phi_A}\\ \sqrt{n_B}e^{i\phi_B}\end{pmatrix}$$

where the constant K is a characteristic of the junction. To solve the above equation, first calculate the time derivative of the order parameter in superconductor A.

$$\frac{\partial}{\partial t}(\sqrt{n_A}e^{i\phi_A}) = \dot{\sqrt{n_A}}e^{i\phi_A} + \sqrt{n_A}(i\dot{\phi}_A e^{i\phi_A}) = (\dot{\sqrt{n_A}} + i\sqrt{n_A}\dot{\phi}_A)e^{i\phi_A}$$

and therefore the Schrödinger equation gives:

$$(\dot{\sqrt{n_A}} + i\sqrt{n_A}\dot{\phi}_A)e^{i\phi_A} = \frac{1}{i\hbar}(eV\sqrt{n_A}e^{i\phi_A} + K\sqrt{n_B}e^{i\phi_B})$$

The phase difference of Ginzburg-Landau order parameters across the junction is called the Josephson phase:

$$\varphi = \phi_B - \phi_A$$

The Schrödinger equation can therefore be rewritten as:

$$\dot{\sqrt{n_A}} + i\sqrt{n_A}\dot{\phi}_A = \frac{1}{i\hbar}(eV\sqrt{n_A} + K\sqrt{n_B}e^{i\varphi})$$

and its complex conjugate equation is:

$$\dot{\sqrt{n_A}} - i\sqrt{n_A}\dot{\phi}_A = \frac{1}{-i\hbar}(eV\sqrt{n_A} + K\sqrt{n_B}e^{-i\varphi})$$

Add the two conjugate equations together to eliminate $\dot{\phi}A$:

$$2\dot{\sqrt{n_A}} = \frac{1}{i\hbar}(K\sqrt{n_B}e^{i\varphi} - K\sqrt{n_B}e^{-i\varphi}) = \frac{K\sqrt{n_B}}{\hbar} \cdot 2\sin\varphi.$$

Since

$$\dot{\sqrt{n_A}} = \frac{\dot{n}_A}{2\sqrt{n_A}}$$

We have:

$$\dot{n}_A = \frac{2K\sqrt{n_A n_B}}{\hbar}\sin\varphi$$

Now, subtract the two conjugate equations to eliminate $\dot{\sqrt{nA}}$

$$2i\sqrt{n_A}\dot{\phi}_A = \frac{1}{i\hbar}(2eV\sqrt{n_A} + K\sqrt{n_B}e^{i\varphi} + K\sqrt{n_B}e^{-i\varphi})$$

Which gives:

$$\dot{\phi}_A = -\frac{1}{\hbar}(eV + K\sqrt{\frac{n_B}{n_A}}\cos\varphi)$$

Similarly, for superconductor B we can derive that:

$$\dot{n}_B = -\frac{2K\sqrt{n_A n_B}}{\hbar}\sin\varphi$$

$$\dot{\phi}_B = \frac{1}{\hbar}(eV - K\sqrt{\frac{n_A}{n_B}}\cos\varphi)$$

Noting that the evolution of Josephson phase is $\dot{\varphi} = \dot{\phi}_B - \dot{\phi}_A$ and the time derivative of charge carrier density $\dot{n}_A$ is proportional to current I, the above solution yields the Josephson equations:

$I(t) = I_c \sin(\varphi(t))$ (1st Josephson relation or weak-link current-phase relation)

$\frac{\partial\varphi}{\partial t} = \frac{2eV(t)}{\hbar}$ (2nd Josephson relation or superconducting phase evolution equation)

Where V(t) and I(t) are the voltage across and the current through the Josephson junction, and I_c is a parameter of the junction named the critical current. The critical current of the Josephson junction depends on the properties of the superconductors, and can also be affected by environmental factors like temperature and externally applied magnetic field.

The Josephson constant is defined as:

$$K_J = \frac{2e}{h}$$

and its inverse is the magnetic flux quantum:

$$\Phi_0 = \frac{h}{2e} = 2\pi\frac{\hbar}{2e}$$

The superconducting phase evolution equation can be re-expressed as:

$$\frac{\partial\varphi}{\partial t} = 2\pi[K_J V(t)] = \frac{2\pi}{\Phi_0}V(t)$$

If we define:

$$\Phi = \Phi_0\frac{\varphi}{2\pi}$$

Then the voltage across the junction is:

$$V = \frac{\Phi_0}{2\pi}\frac{\partial\varphi}{\partial t} = \frac{d\Phi}{dt}$$

Quantum Hall effect:

The quantum Hall effect (or integer quantum Hall effect) is a quantized version of the Hall effect, observed in two-dimensional electron systems subjected to low temperatures and strong magnetic fields, in which the Hall resistance R_{xy} exhibits steps that take on the quantized values at certain level

$$R_{xy} = \frac{V_{\text{Hall}}}{I_{\text{channel}}} = \frac{h}{e^2\nu}$$

Where V_{Hall} is the Hall voltage, $I_{channel}$ is the channel current, e is the elementary charge and h is Planck's constant. The divisor ν can take on either integer (ν = 1, 2, 3,...) or fractional values.

$$(\nu = \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{2}{3}, \frac{3}{5}, \frac{1}{5}, \frac{2}{9}, \frac{3}{13}, \frac{5}{2}, \frac{12}{5}, \ldots)$$

Here, ν is roughly but not exactly equal to the filling factor of Landau levels. The quantum Hall effect is referred to as the integer or fractional quantum Hall effect depending on whether ν is an integer or fraction, respectively.

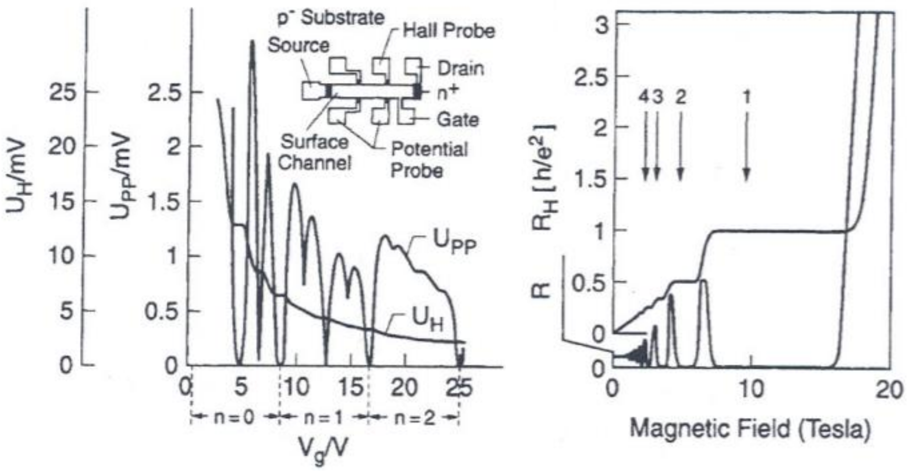
The striking feature of the integer quantum Hall effect is the persistence of the quantization (i.e. the Hall plateau) as the electron density is varied. Since the electron density remains constant when the Fermi level is in a clean spectral gap, this situation corresponds to one where the Fermi level is energy with a finite density of states, though these states are localized.

The fractional quantum Hall effect is more complicated, its existence relies fundamentally on electron–electron interactions. The fractional quantum Hall effect is also understood as an integer quantum Hall effect, although not of electrons but of charge-flux composites known as composite fermions. In 1988, it was proposed that there was quantum Hall effect without Landau levels. This quantum Hall effect is referred to as the quantum anomalous Hall (QAH) effect. There is also a new concept of the quantum spin Hall Effect which is an analogue of the quantum Hall effect, where spin currents flow instead of charge currents.

The integral quantum Hall effect was discovered in 1980 by Klaus von Klitzing, Michael Pepper, and Gerhardt Dordda. Truly remarkably, at low temperatures (~ 4 K), the Hall resistance of a two dimensional electron system is found to have plateaus at the exact values of 2 h ie . In the above expression, i is an integer, h is Planck's constant and e is the electron charge. At the same time, in the applied magnetic field range where the Hall resistance shows the plateaus, the magnetoresistance (i.e., the resistance measured along the direction of the current flow) drops to negligible values.

iscovered by Horst L. Störmer, Daniel C. Tsui, and Arthur C. Gossard. When cooled down below ~2 K, the Hall resistance of the 2D electron systems shows plateaus at the values of 2 h ν e , where ν is a fraction such as 1/3, 1/7, 2/3, 4/5 and so on. The value RH ≈ 25.812 kΩ for i = 1, the quantum of resistance, became the new world's resistance standard in 1990.

We are thankful to Professor Horst Stromer, who kindly agreed to direct the setting up of the first integral quantum Hall effect experiment designed especially for the education of undergraduate students. We also wish to thank Alexander Elias, the remarkable undergraduate student who was in charge of this project. In the future, we hope we can expand the range of applications of this experiment to observe also the fractional quantum Hall effect.



(a) (b)

Fig.4.29 (a) original data of the discovery of the integral quantum Hall effect. (b) New data

In the Fig.4.29 (a) is shown the original data of the discovery of the integral quantum Hall effect. The data was taken using a Metal Oxide Semiconductor Field Effect Transistor (MOSFET). In the experimental setup shown in the inset, by adjusting the gate voltage one can change the career density in the sample. The plateaus of the Hall voltage and the minima of the magneto resistance are distinguishable as the gate voltage is varied. Newer data taken using a GaAs/AlGaAs hetero-junction confirms the quantization of the Hall resistance, as shown on the right hand side of Fig.4.30.

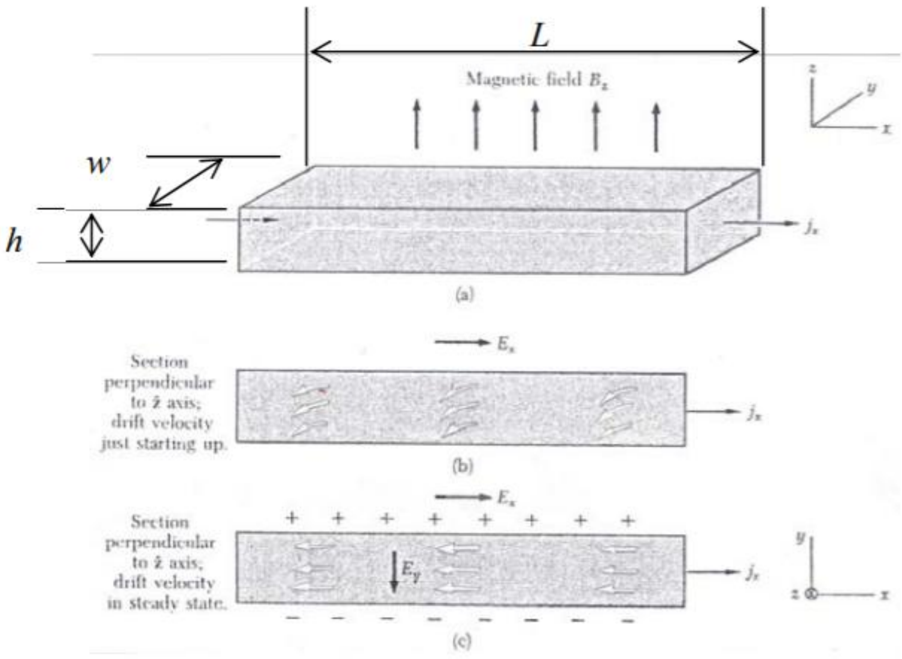


Fig.4.30 The standard Hall Effect

Fig.4.30 shows the standard geometry for observing the Hall Effect, with axes labeling used throughout the experiment. A specimen of rectangular cross – section is placed in a magnetic field, B_z. An electric field E_x is applied between the end electrodes (the source and the drain) and makes an electric current of current density j_x flow through the sample. The magnetic Lorentz force acting on the moving electric charges causes the electrons to accumulate on one face of the specimen, leaving an excess of positive charges on the other face. These charges produce the transverse electric field E_y (Hall field). E_y increases as the electric charge builds up on the sides of the sample, until the electrostatic force with which it acts on the electrons balances the Lorentz force created by the magnetic field. Disregarding the signs, $eE_y = evB_z$, or $E_y = vB_z$., with v the drift velocity of electrons along the x direction. The Hall resistance is defined as $R_H=V_H/ I_x$. V_H is the Hall voltage, $V_H = wE_y$. The magnitude of the current flowing in the x direction is $I_x = Aj_x$. A = wh is the area of the cross-section of the sample perpendicular to the direction of the current. $j_x = n_{3D}ev$, where n_{3D} is the volume density of free electrons in the sample. Therefore

$$R_H = \frac{wvB}{whn_{3D}ev} = \frac{B}{ne}.\ n = n_{3D}h$$

is the density of electrons per unit area (n_{3D} = N/whL and n = N/wL, with N the number of electrons in the sample)

SAQ.3

a) Discuss about the Types of superconductors.
b) Write the properties and applications of superconductors.
c) What do you mean by the Josephson Effect?
d) Explain in brief quantum Hall effect.

Examples:

Q.1 A toroid has a mean radius R equal to 20/π cm, and a total of 400 turns of wire carrying a current of 2.0A. An aluminium ring at temperature 280 K inside the toroid provides the core.

(a) If the magnetization I is 4.8×10-2 Am-1, find the susceptibility of aluminium at 280 K.

(b) If the temperature of the aluminium ring is raised to 320 K, what will be the magnetization?

Solution:

(a)The number of turns per unit length of the toroid is

$$n = \frac{400}{2\pi R}$$

The magnetic intensity H in the core is

$$H = ni = \frac{400*2.0A}{2\pi*\frac{20}{\pi}*10^{-2}m} = 2000Am^{-1}$$

The susceptibility is

$$\chi = I/H$$

$$\therefore \chi = \frac{4.8*10^{-2}Am^{-1}}{2000Am^{-1}} = 2.4*10^{-5}$$

(b) The susceptibility χ of a paramagnetic substance varies with absolute temperature as χ=c/T

Thus,

$\chi_2/\chi_1=T_1/T_2$

The susceptibility of aluminium at temperature 320 K is, therefore,

$$\chi = \frac{280}{320}*2.4*10^{-5} = 2.1*10^{-5}$$

Thus, the magnetization at 320 K is

$$I = \chi H = 2.1*10^{-5}*2000Am^{-1}$$

$$\therefore I = 4.2 * 10^{-5} Am^{-1}$$

Q.2 A paramagnetic material has a magnetic field intensity of 10^4 Am^{-1}. If the susceptibility of the material at room temperature is 3.7×10^{-5}. Calculate the magnetization and flux density in the material.

Solution:

Given data:

Magnetic field indensity $H = 10^4\ Am^{-1}$

Susceptibility $\chi = 3.7 \times 10^{-4}$

1. Susceptibility $\chi = \frac{M}{H}$

$\therefore$ Magnetization $M = \chi H$

$$= 3.7 \times 10^{-3} \times 10^4$$

$$= 3.7 \times 10$$

Magnetization $M = 37 A\ m^{-1}$

2. Flux density $B = m_0 (M + H)$

$$= 4\pi \times 10^{-7} \times (37 + 10^4)$$

$$= 126179.4 \times 10^{-7}$$

$$= 0.0126\ Wb\ m^{-2}$$

Magnetization $M = 37 A\ m^{-1}$

Flux density $B = 0.0126\ Wb\ m^{-2}$.

Q.3 A magnetic material has a magnetization of 2300 A m^{-1} and produces a flux density of 0.00314 Wb m^{-2}. Calculate the magnetizing force and the relative permeability of the material.

Solution:

Given data:

Magnetization M = 2300 A m^{-1}

Flux density B = 0.00314 Web m^{-2}.

i) The magnetic flux density

$$B = \mu_0 (M + H)$$

The magnetic force $H = \left[\frac{B}{\mu_0} - M\right]$

$$= \frac{0.00314}{4\pi \times 10^{-7}} - 2300$$

$$H = 198.7326 \text{ A m}^{-1}$$

ii) Susceptibility $\chi = \frac{M}{H} = (\mu_r - 1)$

$\therefore$ Relative permeability $\mu_r = \frac{M}{H} + 1$

$$= \frac{2300}{198.7326} + 1$$

$$\mu_r = 12.573$$

Magnetic force $H = 198.7326 \text{ A m}^{-1}$

Relative permeability $\mu_r = 12.573$

Q.4 A paramagnetic material has FCC structure with a cubic edge of 2.5 A°. If the saturation value of magnetization is 1.8×10^6 A m^{-1}, Calculate the magnetization contributed per atom in Bohr magnetrons.

Solution:

Given data:

The interatomic spacing $a = 2.5 \times 10^{-10}$ m

The magnetization $M = 1.8 \times 10^6 \text{ A m}^{-1}$

The number of atoms present per unit volume

$$N = \frac{\text{Number of atomas present in an unit cell}}{\text{Volume of the unit cell}}$$

$$= \frac{2}{\left(2.5 \times 10^{-10}\right)}$$

$$N = 1.28 \times 10^{29} \text{ m}^3$$

Total magnetization $M = 1.8 \times 10^6 \text{ A m}^{-1}$

The magnetization produced per atom

$$= \frac{M}{N}$$

$$= \frac{1.8 \times 10^{6}}{1.28 \times 10^{29}}$$

$$= 1.4062 \times 10^{-23} \text{ A m}^{-2}$$

Bohr magneton $\mu_B = \dfrac{eh}{4\pi m}$

$$= \frac{1.6 \times 10^{-19} \times 6.625 \times 10^{-34}}{4 \times 3.14 \times 9.1 \times 10^{-31}}$$

$$\mu_B = 9.27 \times 10^{-24} \text{ A m}^{-2}$$

$\therefore$ Magnetization produced per atom

$$M = \frac{1.40625 \times 10^{-23}}{9.27 \times 10^{-24}}$$

$$M = 1.519 \text{ Bohr magnetons}$$

The average magentization per atom=1.517 Bohr magnetons

Q.5 In a magnetic material the field strength is found to be 10^6 A m^{-1}. If the magnetic susceptibility of the material is 0.5×10^{-5}, calculate the intensity of magnetization and flux density in the material.

Solution:

Given data:

Magnetic field strenght $H = 10^6 \text{ A m}^{-1}$

Susceptibility $\chi = 0.5 \times 10^{-5}$

i) Magnetization $M = \chi H$

$= 10^6 \times 0.5 \times 10^{-5}$

$M = 5 \text{ A m}^{-1}$

ii) Flux density $B = \mu_0 (M + H)$

$= 4 \times 3.14 \times 10^{-7} (5 + 10^6)$

$B = 1.257 \text{ Wb m}^{-2}$

Magnetization $M = 5 \text{ A m}^{-1}$

Fluxdensity $B = 1.257 \text{ Wb m}^{-2}$

Q.6 A superconducting tin has a critical temperature of 3.7 K at zero magnetic field and a critical field of 0.0306 Tesla at 0 K. Find the critical field at 2 K.

Solution:

Given data:

Critical temperature	$T_C = 3.7$ K
Critical field	$H_C = 0.0306$ Tesla
Temperature	$T = 2$ K

The critical magnetic field

$$H_C = H_0\left[1-\left[\frac{T^2}{T_c^2}\right]\right]$$

$$= 0.0306\left[1-\left[\frac{2}{3.7}\right]^2\right] = 0.0216 \text{ Tesla}$$

The critical magnetic field $H_c = 0.0216$ Tesla.

Q.7 Calculate the critical current and current density for a wire of a lead having a diameter of 1 mm at 4.2 K. The critical temperature for lead is 7.18 K and $H = 6.5 \times 10^4$ A m^{-1}.

Solution:

Given data:

Critical temperature	$T_c = 7.18$ K 7.18 K
Critical field	$H_0 = 6.5 \times 10^4$ A m^{-1}
Temperature	$T = 4.2$ K
Radius of the wire	$r = 0.5 \times 10^{-3}$ m

The critical magnetic field $H_c = H_0\left[1-\left[\frac{T^2}{T_c^2}\right]\right]$

$$= 6.5 \times 10^4\left[1-\left[\frac{4.2}{7.18}\right]^2\right]$$

$$= 4.276 \times 10^4 \text{ A m}^{-1}$$

i) Critical current $\quad I_c = 2\pi r H_c$

$$= 2 \times 3.14 \times 0.5 \times 10^{-3} \times 4.276 \times 10^4$$

$$= 134.39 \text{ A}$$

ii) Critical density $\quad J_c = \frac{I_C}{\pi r^2}$

$$= \frac{134.39}{3.15 \times (0.5 \times 10^{-3})^2}$$

$$= 1.71 \times 10^8 \text{ A m}^{-2}$$

Critical current $\quad I_c = 134.\ 39$ A

Critical density $\quad J_c = 1.71 \times 10^8$ A m^{-2}

Q.8 Prove that susceptibility of superconductor is -1 and relative permeability is zero.

Solution:

Given data:

We know, the induced magnetic field $\quad B = \mu_0(M+H) \quad \ldots(1)$

In superconductor, $\quad B = 0$

Therefore, $\quad 0 = \mu_0(M+H)$

Since μ_0 # 0, $\quad M = -H$

$$\frac{M}{H} = \chi = -1 \quad \ldots(2)$$

Also, $\quad \chi = \mu_r - 1$

$$-1 + 1 = \mu_r = 0$$

Therefore the susceptibility $\quad x = -1$

and Relative permeability $\quad \mu_r = 0$

Q.9 Find the critical current which can pass through a long thin superconducting wire of aluminum of diameter 2 mm, the critical magnetic field for aluminum is 7.9×10^3 A m^{-1}.

Solution:

Given data:

The critical magnetic field $H_c = 7.9 \times 10^3 \text{ A m}^{-1}$

$$\text{Radius } r = \frac{\text{Diameter}}{2}$$

$$= \frac{2}{2} = 1 \times 10^{-3}$$

$$\text{Critical current } I_c = 2\pi r H_c$$

$$= 2 \times 3.14 \times 1 \times 10^{-3} \times 7.9 \times 10^3 \text{ A m}^{-1}$$

$$I_c = 49.65 \text{ A}$$

$$\text{Critical current } I_c = 49.65 \text{ A}$$

Q.10 The superconducting transition temperature of lead of 7.26 K. The initial field at 0 K is 64×10^3 Amp m^{-1}. Calculate the critical field at 5 K.

Solution:

Given data:

$$\text{Critical temperature } T_c = 7.26 \text{ K}$$

$$\text{Critical field } H_0 = 64 \times 10^3 \text{ Amp m}^{-1}$$

$$\text{Temperature } T = 5\text{K}$$

$$\text{The critical field } H_c = H_0 [1 - (T/T_C)^2]$$

$$= 64 \times 10^3 [1 - (5/7.26)^2]$$

$$= 64 \times 10^3 \times 0.5257$$

$$H_c = 33.644 \times 10^3 \text{ Amp m}^{-1}$$

Q.11 Calculate the critical current which can flow through a long thin superconducting wire of diameter 1 mm. The critical magnetic field is 7.9×10^3 Amp m^{-1}.

Solution:

Given data:

$$\text{Diameter of the wire d} = 1\text{ mm} = 1 \times 10^{-3}\text{m}$$

$$\text{radius of the wire r} = \frac{d}{2} = \frac{1 \times 10^{-3}}{2}\text{ m}$$

$$\text{The critical magnetic field } H_c = 7.9 \times 10^{3}\text{ Amp m}^{-1}$$

Critical current flowing through the wire

$$I_c = 2\pi r H_c$$

$$= 2 \times 3.14 \left(\frac{1 \times 10^{-3}}{2}\right)(7.9 \times 10^{3})$$

$$\boxed{I_c = 24.81\text{ Amp}}$$

4.10 Summary:

1. Diamagnetic materials are repelled by a magnetic field; an applied magnetic field creates an induced magnetic field in them in the opposite direction, causing a repulsive force. In contrast, paramagnetic and ferromagnetic materials are attracted by a magnetic field.
2. Paramagnetic materials, similar to ferromagnetic materials, have a positive response to external magnetic fields, i.e., it becomes a magnet. As long as the strong magnetic field is present, it will attract and repel other magnets in the usual way.
3. According to Curie's Law, the magnetization in a paramagnetic material is directly proportional to the applied magnetic field. If the object is heated, the magnetization is viewed to be inversely proportional to the temperature. The law was discovered by the French physicist, Pierre Curie.
4. The Curie-Weiss law is one of the important laws in electromagnetism that says that the magnetic susceptibility is above the Curie temperature point of a ferromagnet in the paramagnetic region. The magnetic moment is a quantity of a magnet that determines its torque in an external magnetic field.
5. In Classical theory of diamagnetism, we will obtain an expression for the change in magnetic moment of an orbiting electron in a diamagnetic atom and the induced magnetic moment per unit volume of diamagnetic material in the applied magnetic field B_0.

6. Diamagnetism is a quantum mechanical effect that occurs in all materials; when it is the only contribution to the magnetism, the material is called diamagnetic. In paramagnetic and ferromagnetic substances, the weak diamagnetic force is overcome by the attractive force of magnetic dipoles in the material.
7. Quantum theory of Paramagnetism. According to classical theory the atoms of the paramagnetic gas are assumed to be small permanent magnets due to circulating electrons. In the absence of the external magnetic field , the magnetic axes of the atoms are uniformly distributed in all directions .
8. Ferromagnetism, physical phenomenon in which certain electrically uncharged materials strongly attract others. Two materials found in nature, lodestone (or magnetite, an oxide of iron, Fe_3O_4) and iron, have the ability to acquire such attractive powers, and they are often called natural ferromagnets.
9. Antiferromagnetic materials occur commonly among transition metal compounds, especially oxides. Examples include hematite, metals such as chromium, alloys such as iron manganese (FeMn), and oxides such as nickel oxide (NiO). There are also numerous examples among high nuclearity metal clusters.
10. Ferrimagnetism, type of permanent magnetism that occurs in solids in which the magnetic fields associated with individual atoms spontaneously align themselves, some parallel, or in the same direction (as in ferromagnetism), and others generally antiparallel, or paired off in opposite directions.
11. Superconductivity is a phenomenon in which the electrical resistivity suddenly drops to zero at its transition temperature Tc.
12. While many materials exhibit some small amount of diamagnetism, superconductors are strongly diamagnetic. Since diamagnetics have a magnetization that opposes any applied magnetic field, the superconductor is repelled by the magnetic field.
13. Meissner effect, the expulsion of a magnetic field from the interior of a material that is in the process of becoming a superconductor, that is, losing its resistance to the flow of electrical currents when cooled below a certain temperature, called the transition temperature, usually close to absolute zero.
14. A theory of superconductivity formulated by John Bardeen, Leon Cooper, and Robert Schrieffer. It explains the phenomenon in which a current of electron pairs flows without resistance in certain materials at low temperatures.

15. Superconductors—special metals that can conduct electrical current with no loss of energy—could one day have a monumental impact on the efficient transmission of power in the United States and around the world. They could also lead to great innovations in medical imaging, drug analysis, and even telecommunications.
16. Josephson Effect, flow of electric current between two pieces of superconducting material separated by a thin layer of insulating material. Superconductors are materials that lose all electrical resistance when cooled below a certain temperature near absolute zero.
17. The quantum Hall effect (or integer quantum Hall effect) is a quantized version of the Hall Effect and which is observed in two-dimensional electron systems subjected to low temperatures and strong magnetic fields, in which the Hall resistance R_{xy} exhibits steps that take on the quantized values.

4.11 Terminal Questions:

1) What are the Comparisons of diamagnetic and paramagnetic materials with examples?
2) Explain in detail of Curie law.
3) Discuss about the concept of Curie Wiess law.
4) Explain the Classical theory of diamagnetism and paramagnetism.
5) Explain the quantum theory of diamagnetism and paramagnetism.
6) What do you mean by ferromagnetism, antiferromagetism and ferrimagnetism?
7) Explain the Superconductivity and also write its characteristics.
8) What do you mean by magnetic behavior of superconductor?
9) Explain the concept of Meisener's effect.
10) What do you mean by BCS theory**?**
11) What are the Types of superconductors and also write its properties and applications?
12) Explain the working of Josephson Effect.
13) Explain the concept of quantum Hall effect.
14) A paramagnetic material has a magnetic field intensity of 10^5 Am^{-1}. If the susceptibility of the material at room temperature is 4.7×10^{-5}. Calculate the magnetization and flux density in the material.
15) A magnetic material has a magnetization of 3200 A m^{-1} and produces a flux density of 0.00314 Wb m^{-2}. Calculate the magnetizing force and the relative permeability of the material.
16) A superconducting tin has a critical temperature of 3.9 K at zero magnetic fields and a critical field of 0.0306 Tesla at 0 K. Find the critical field at 3 K.

17) Calculate the critical current and current density for a wire of a lead having a diameter of 2 mm at 5.2 K. The critical temperature for lead is 8.18 K and H = 7.5×10^4 A m^{-1}.
18) Find the critical current which can pass through a long thin superconducting wire of aluminum of diameter 3 mm, the critical magnetic field for aluminum is 8.9×10^3 A m^{-1}.
19) The superconducting transition temperature of Lead is 8.26 K. The initial field at 0 K is 74×10^3 Amp m^{-1}. Calculate the critical field at 6 K.
20) Calculate the critical current which can flow through a long thin superconducting wire of diameter 2 mm. The critical magnetic field is 8.9×10^3 Amp m^{-1}.

Uttar Pradesh Rajarshi Tandon

Open University

Bachelor of Science

DCEPHS-109

Solid State Physics and Advanced Electronics

Block

2 Advanced Analog Electronics

Course Design Committee

Prof. Ashutosh Gupta **Chairman**
Director, School of Science, UPRTOU, Prayagraj
Prof. A. K. Rai **Member**
Professor, Dept. of Physics, University of Allahabad, Prayagraj
Gorakhpur
Prof. Ram Kripal **Member**
Professor, Dept. of Physics, University of Allahabad, Prayagraj
Dr. Anjani Kumar Singh **Member**
Associate Professor (Retd.), E.C.C University of Allahabad, Prayagraj

Course Preparation Committee

Mr. Ashish Sankhwar **Author**
Asst Prof. Dept. Of Electrical Engg. **Unit- 01-10**
FET,MJPRU,Bareilly

Prof. Nitya Kaushal Kishore **Editor**
Ex. HOD, Dept. of Physics, **Block -I,II & III**
S.K.M.U. University,
Dumka, Jharkhand.

Dr. Arvind Kumar Mishra **Coordinator**
Assistant Professor , Physics,
School of Science, UPRTOU, Prayagraj

DCEPHS – 109 : Plant Pathology and Microbiology
ISBN-

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Nehru Road, Prayagraj.

Unit 5 Different modes of operations

Structure:

5.1 Introduction:

The Ebers-Moll Model is an electronic representation of a transistor, either NPN or PNP, in any of the four fundamental configurations. For the ideal transistor, these parameters are the saturation currents of the emitter and collector junctions and the normal and inverted alphas.

The Ebers-Moll equations are based on two exponential diodes plus two current-controlled current sources. The NPN Bipolar Transistor block provides the following enhancements to that model: Early voltage effect, optional base, collector, and emitter resistances.

The Ebers-Moll BJT Model is a good large-signal, steady-state model of the transistor and allows the state of conduction of the device to be easily determined for different modes of operation of the device. The different modes of operation are determined by the manner in which the junctions are biased. Ebers-Moll model also known as "Coupled Diode Model".

The Ebers-Moll model provides an alternative view or representation of the voltage-current equation model.

Ebers–Moll model for pnp transistor involves two ideal diodes placed back to back with saturation current.

Transistor datasheets will define this voltage as CE saturation voltage V_{CE}(sat) -- a voltage from collector to emitter required for saturation. This value is usually around 0.05-0.2V. This value means that V_C must be slightly greater than V_E (but both still less than V_B) to get the transistor in saturation mode.

In a nutshell, "saturation" for a BJT is the point where a further increase in base current will not result in a corresponding increase in collector current. That's because Vce is as low as it's going to go for that particular Ic.

The four transistor operation modes are: Cut-off: The transistor acts like an open circuit. Saturation: The transistor acts like a short circuit. Current freely flows from collector to emitter. In the inverse active mode of transistor operation, the base-emitter junction is reverse biased and the base-collector junction is forward biased.

A transistor is said to be in its active mode if it is operating somewhere between fully on (saturated) and fully off (cutoff). Base current regulates collector current. By regulate, we mean that no more collector current can exist than what is allowed by the base current.

5.2 Objectives:

After studying this unit you should be able to

- Explain and identify Eber's moll model for PNP and NPN transistors.
- Study and identify Expressions for various currents and voltage.
- Explain the Saturation parameters and its importance.
- Explain and identify Conditions for cut off mode, saturation mode, inverse mode and active mode.
- Study and identify Comparison among all modes of operations.

5.3 Eber's moll model for PNP and NPN transistors: -

If transistor circuits are to be of any use or amenable to diagnostic procedures, we must be able to model them. Even the best electronic test equipment is useless if we don't know what to look for in the circuits under investigation.

Transistors characteristically have multiple modes of conduction. We can view these phenomena in the two-diode model of a bipolar junction transistor (BJT). Two diodes whose anodes join to form a center tap are analogous to an NPN transistor insofar as ohmmeter readings accurately represent the real device. Two diodes with cathodes connected to a common node are analogous to a PNP transistor. (NPN transistors are preferred due to increased mobility of electrons compared to holes and also because they are compatible with a negative ground system.) Because two diodes are separate components and cannot share in common a semiconducting layer, they do not function as an amplifier, go into oscillation or perform switching action in the manner of actual transistors.

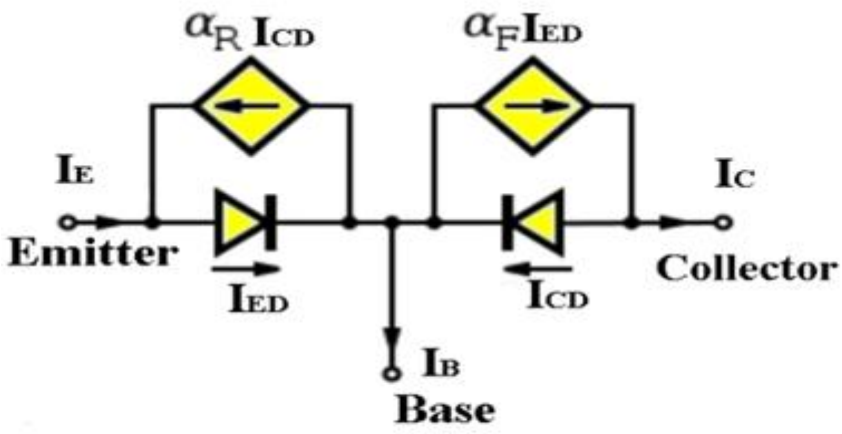


Fig.5.1 Ebers-Moll for PNP

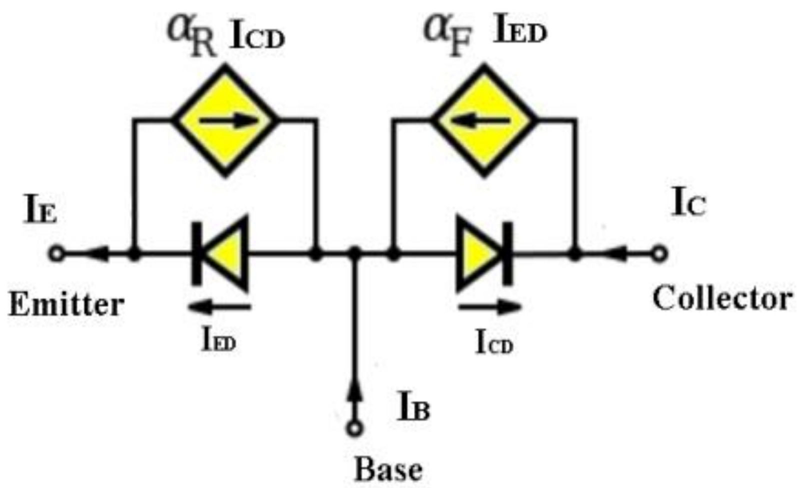


Fig.5.2 Ebers-Moll for NPN

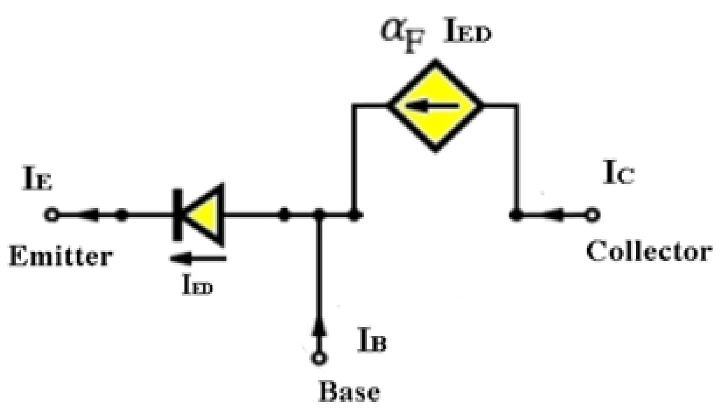


Fig.5.3 Ebers-Moll for NPN forward active

Ebers-Moll models for PNP and NPN devices, and an approximated Ebers–Moll model for an NPN transistor in the forward active mode. When in forward-active mode, the collector diode is reverse-biased so I_{CD} is virtually zero. Most of the emitter diode current (αF is nearly 1) is drawn from the collector, providing the amplification of the base current.

To accurately model a BJT, we must look beyond the simple diode hookup, although that remains relevant. The Ebers-Moll Model is an electronic representation of a transistor, either NPN or PNP, in any of the four fundamental configurations. In addition to the diode model, which is a physical simulation, Ebers-Moll is a paper construct, having its existence in part as a schematic diagram and also a set of equations, either of these deploying conventional symbols.

Jewel James Ebers and John L. Moll introduced this mathematical model of transistor currents in 1954. The model is described in a paper titled Large Signal Behavior of Junction Transistors, which appeared in Proceedings of the Institute of Radio Engineers. In the paper's Abstract, the authors say: In the consideration of the junction transistor as a switch there are three characteristics of primary interest, the open impedance, the closed impedance, and the switching-time. A generalized two-terminal-pair theory of junction transistors is presented which is applicable, on a dc basis, in all regions of operation. Using this theory, the open and closed impedances of the transistor switch are expressible in terms of easily measurable transistor parameters. For the ideal transistor, these parameters are the saturation currents of the emitter and collector junctions and the normal and inverted alphas. The transition of the transistor switch from open to closed, or vice versa, is discussed, including the effects of minority carrier storage. This transition can be expressed in analytic form in terms of the alphas and the normal and inverted alpha cut-off frequencies.

5.4 Expressions for various currents and voltage for Eber's moll model for PNP and NPN transistors:

The behavior of the bipolar transistor can be described in both conceptual and quantitative terms by ob- serving that this device consists of two coupled pn junctions. The base region is common to both junctions and forms the coupling between them. Bipo- lar transistors are constructed with very narrow base regions (considerably smaller than one diffusion length). The current components which comprise the terminal currents I_E and I_C are shown for a pnp transistor in figure-4(a).

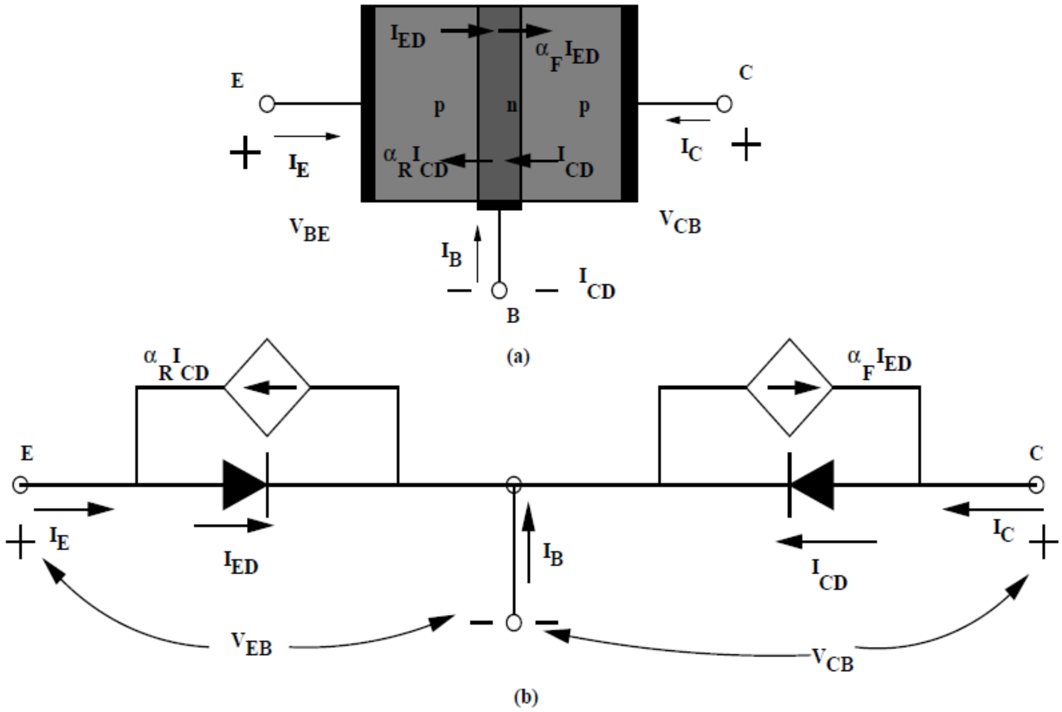


Fig.5.4 (a) Current components in pnp transistor (b) Large-signal (Ebers-Moll) representation of a pnp transistor

The voltages V_{EB} and V_{CB} are the voltage drops from emitter to base and collector to base, respectively. Assuming that there is no voltage drop across the bulk semiconductors forming the emitter, base, and collector regions, these voltages appear across the respective junctions. With both voltages measured with respect to the base, this connection is called the common-base(CB) config- uration. The emitter current has two components. The current associated with the emitter-base diode is designated I_{ED} and that with the collector-base diode, I_{CD}. The component $\alpha_R I_{CD}$ is the portion of I_{CD} that is coupled through the base to the emitter. Similarly, α_F I_{ED} is the fraction of I_{ED} coupled into the collector. On the basis of the considerations in the preceding paragraph we can construct the Ebers-Moll model in fig.4(b). The two back-to-back diodes (whose cathodes are connected) represent the junctions of the bipolar transistor, whereas the two controlled sources indicate the coupling between junctions. The currents I_{ED} and I_{CD} are related to V_{EB} and V_{CB} by the diode volt-ampere relation given. Thus we have

$$I_{ED} = I_{E0}(e^{V_{EB}/V_T} - 1) \ldots\ldots\ldots\ldots(1)$$

$$I_{CD} = I_{C0}(e^{V_{CB}/V_T} - 1) \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

where I_{E0} I_{C0} are the reverse saturation currents. By applying KCL we have,

$$I_E = I_{ED} - \alpha_R I_{CD} \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$I_C = -\alpha_F\ I_{ED} + I_{CD} \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

substituting Eqn.(1) and (2) in (3) and (4) we get,

$$I_E = I_{E0}(e^{V_{EB}/V_T} - 1) - \alpha_R I_{C0}(e^{V_{CB}/V_T} - 1) \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

$$I_C = -\alpha_F I_{E0}(e^{V_{EB}/V_T} - 1) + I_{C0}(e^{V_{CB}/V_T} - 1) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

The Eqns. (5) and (6) are called “Ebers-Moll Equations” for the pnp transistor. The values αF and αR are each less than unity, since not all the current from one diode is coupled to the other junction. The subscripts refers to forward transmission (F) from the emitter to collector and reverse transmission (R) from collector to emitter. I_{E0}, I_{C0}, α_F and α_R are found from the theory and are related by the equation,

$$\alpha_F\ I_{E0} = \alpha_R I_{C0} \ldots\ldots\ldots\ldots (7)$$

This condition is often called reciprocity condition for BJT. The base current can be calculated by applying KCL at the transistor (i.e., treating the tran- sistor as node, so sum of all currents entering the node is zero.) Hence we have

$$I_B + I_C + I_E = 0 \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

$$\Rightarrow I_B = -(I_C + I_E) \ldots\ldots\ldots\ldots\ldots (9)$$

For transistor with small dimensions used we have,

$$0.98 \leq \alpha_F \leq 0.998 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

$$0.4 \leq \alpha_R \leq 0.8 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

and the currents I_{E0}, I_{C0}, are of the order of 10−15 A. Similarly the Ebers-Moll equations can be obtained for npn transistor as shown in fig.5(a) and (b) as follows,

$$I_E = -I_{E0}(e^{-V_{EB}/V_T} - 1) + \alpha_R I_{C0}(e^{-V_{CB}/V_T} - 1) \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

$$I_C = \alpha_F I_{E0}(e^{-V_{EB}/V_T} - 1) - I_{C0}(e^{-V_{CB}/V_T} - 1) \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

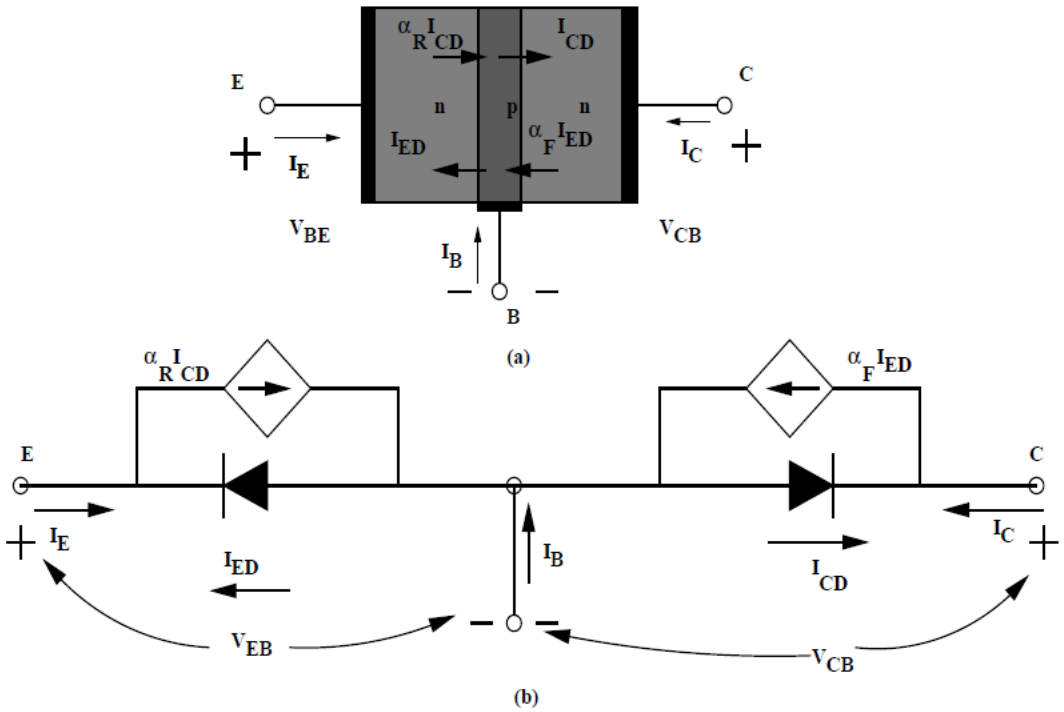


Fig.5.5 (a) Current components in npn transistor (b) Large-signal (Ebers-Moll) representation of a npn transistor

Large Signal Current Gains: Let us consider an npn transistor for the situation that emitter-base junction is forward biased (V_{EB}) and the collector- base are short circuited (V_{CB} = 0). Substituting these values in the Eqns.(12) and (13)

$$-I_E = I_{E0}(e^{-V_{EB}/V_T} - 1)$$

and

$$\alpha_F\, I_{E0}(e^{-V_{EB}/V_T} - 1)$$

Hence eliminating $I_{E0}(e^{-V_{EB}/V_T} - 1)$ weget $I_C = \alpha_F I_E$.
Thus we define,

$$\alpha_F = -\frac{I_C}{I_E}\bigg|\, V_{CB} = 0 \qquad \text{...............(14)}$$

The quantity αF is called “common-base forward short circuit current gain”. Similarly when $V_{CB} < 0$, and $V_{EB} = 0$ we have “common-base reverse short circuit current gain”.

$$\alpha_R = -\frac{I_E}{I_C}\bigg|\, V_{EB} = 0 \qquad \text{....................(15)}$$

For npn devices I_C is positive, while I_E is negative. $I_B = -(I_C + I_E)$ is positive and hence it have the same sign of I_C ; similarly for pnp transistors I_C is negative, I_E is positive which implies I_B is negative. Thus for any BJT, I_B and I_C have same sign and opposite to that of I_E. Substituting Eqn.(14) in (9) we have,

$$I_B = -(1 - \alpha_F)I_E \text{……………(16)}$$

If α_F is almost unity then I_B is quite small compared with I_E and hence magnitudes of I_C and I_E are almost same.

It is often convenient to express the collector and emitter currents in terms of much smaller base currents.

$$\begin{aligned} I_C &= -\alpha_F I_E \\ &= -\alpha_F \left(-\frac{1}{(1-\alpha_F)}\right) I_B \\ &= \frac{\alpha_F}{1-\alpha_F} I_B \\ \Rightarrow I_C &= \beta_F I_B \end{aligned} \qquad \text{……………..(17)}$$

where β_F is called "common-emitter forward short circuit current gain", given by

$$\beta_F = \frac{\alpha_F}{1-\alpha_F} \qquad \text{…………..(18)}$$

Similarly reverse condition yields the "common-emitter reverse short circuit current gain" as

$$\beta_R = \frac{\alpha_R}{1-\alpha_R} \qquad \text{…………..(19)}$$

β_F lies between 50 to 250 and β_R is between 1 to 5.

In the BJT, typically there are two two-wire circuits, input and output. The device in its fundamental form has three rather than four terminals because one of them, which can be base emitter or collector, is common to both circuits. The output circuit can convey to the next stage an amplified or an attenuated version of the signal at the input. When the input and output at each point in time are in the same ratio, the device is said to be linear and when that ratio varies, the device is non-linear.

$X = 2Y$ is a linear equation.

$X = Y^2$ is a non-linear equation.

Both linear and non-linear devices possess some finite gain. Or gain can be infinite, theoretically but not in actuality, where a big-bang condition would exist. Gain is denoted by Greek alpha (α) and Greek beta (β). To clarify, α is I_C/I_E.β is I_C/I_B. The common-emitter voltage gain is always in low negative territory.

The current gain in a common base transistor circuit is by definition the change in collector current over the change in emitter current when the voltage difference between base and collector does not vary. A typical common base current gain is one.
In a BJT, each of the two junctions can be forward biased or reverse biased. Accordingly, there are four possible modes in which the transistor can operate. It is cut off when the emitter-base junction and the collector-base junction are reverse biased. When both of these junctions are forward biased, the device is in saturation. When the emitter-base junction is forward biased and the collector-base junction is reverse biased, the transistor is in the forward-active mode. Finally, the transistor is in the reverse-active mode when the emitter-base is reverse biased and the collector-base is forward biased.
Between cutoff and saturation, the device acts as a switch, which may be open with high impedance or closed with a low impedance. In these biasing conditions, there is no intermediate state. In the forward-active mode, the transistor operates as an amplifier, and in the reverse active mode, it may be used in digital and analog switching operations.
Surprisingly, under certain biasing and signal input conditions, the physical dimensions of a BJT will actually change. This phenomenon was first noted by James Early in 1952 and is known as the early effect. It manifests as a shrinking base width due to a widening of the base-collector depletion region. The result is a rise in collector current and voltage. Metal-oxide semiconductor field-effect transistors (MOSFETs) also exhibit this strange behavior.

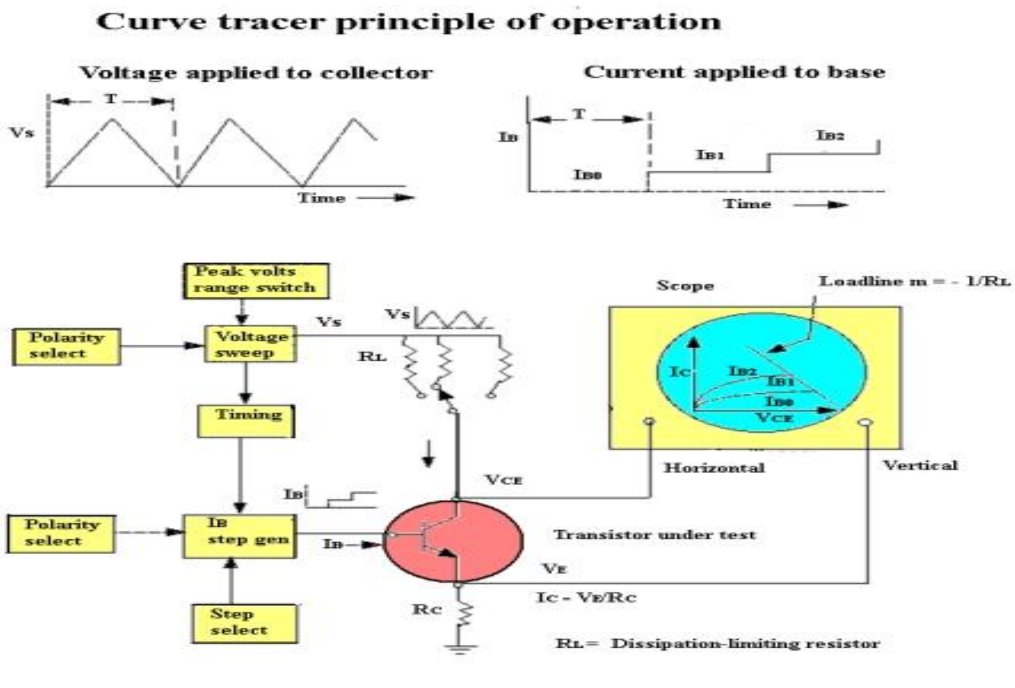


Fig 5.6 The wave-shape polarities shown here apply to a npn transistor.

When it becomes necessary to physically measure transistor parameters such as current gain, breakdown voltages, and impedance, a transistor curve tracer is usually the instrument of choice. The curve tracer can generate and display a family of curves of collector current, I_c, versus collector-to-emitter voltage, V_{CE}, for various values of base current, I_B. From this display, the current gain, β, can be directly determined.

A curve tracer uses three basic circuits to generate this display: a sweep-voltage-generator for control of the collector voltage; a base current source which can be controlled to provide a number of equal increments of base currents with each sweep of the voltage generator; and a timing source to change the base current at the start of each voltage sweep.

The waveform of the sweep-voltage generator, V_s consists of repetitive sweeps occurring with a time period T. This is the collector supply voltage which is repetitively applied to the transistor. The collector voltage, V_{ce}, will provide the horizontal (x-axis) sweep.

A view of the output of the base current source shows that for each consecutive voltage sweep the base current, I_B, is incremented in equal steps with each step synchronized to the beginning of each collector voltage sweep. As the last increment period ends, the base current generator repeats the step sequence. In the U.S., the 60 Hz power line frequency is generally used as the synchronizing signal for the collector sweep voltage and base current steps.

The collector-to-emitter voltage, V_{ce}, provides the horizontal sweep, while the voltage across the current sensing resistor, R_c, which is proportional to collector current, provides the vertical sweep, resulting in a family of curves of I_c versus V_{ce} for a series of equal increment changes in base current.

The displays shown here are for an npn transistor. The current gain of the transistor is determined from: β = current gain = $\Delta I_c/\Delta I_B$ where ΔI_B is set by the curve-tracer Step Selector switch.

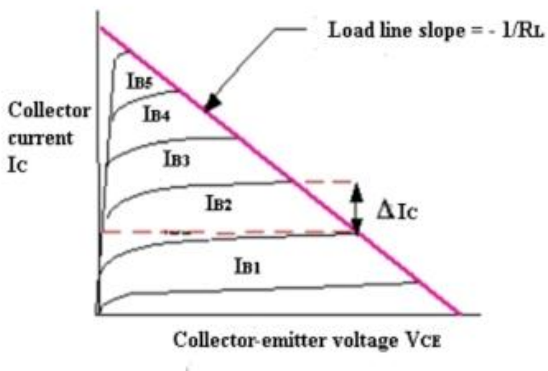


Fig 5.7 Characteristic curve for NPN

The current gain of the transistor is determined from this curve tracer display using the relationship $\beta = \Delta I_c/\Delta I_B$ where ΔI_B is the setting of the tracer's Step Selector switch.

The slope of the load line is determined by a dissipation-limiting resistor, R_L, selected in the collector sweep control section. This resistor is selected so the maximum allowable collector current, I_c, for the transistor is not exceeded for $V_{ce} = 0$ V.

When put on a curve tracer, small transistors don't have much heat dissipation capability so they should be limited to about 50 mA and 40 V; higher power transistors usually have a case that permits attachment to a heat sink. It's usually safe to assume they can handle 1 or 2 A at 40 V.

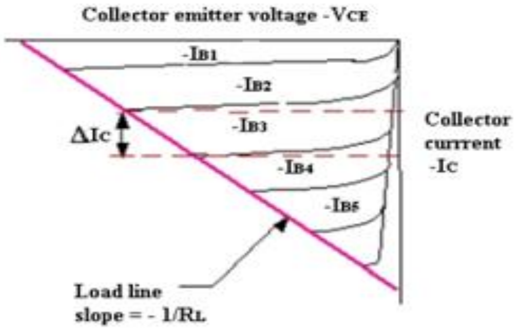


Fig 5.8 Characterstic curve for PNP

The direction of current flow in a pnp transistor is the reverse of that for an npn transistor, so the V_{CE} and the direction of the base current must be reversed, resulting in the characteristic curve visible here.

β will vary depending on the collector current drawn, dropping as the current rises. The gain is generally measured in the voltage/current region in which the transistor is expected to operate.

Finally, transistors on a curve tracer can heat up, so use caution in handling them.

SAQ.1

a) What do you mean by Eber's moll model for PNP transistors?
b) Discuss about the Eber's moll model for NPN transistors.
c) Write the Expressions of various currents and voltages for Eber's moll model for PNP and NPN transistors.

d) Use the Ebers-Moll model to determine I_C for an NPN BJT biased at I_B=200 nA and V_{CE} = 0.25 V. Assume that the Ebers-Moll model parameters are $I_{F0} = 1.5 \times 10^{-16}$A, I_{R0}=2.25×10^{-16}A, and $\alpha_F = 0.996$.

e) What is β_R for a bipolar transistor described by an Ebers-Moll model with I_{F0}=1.3×10^{-16}A , $I_{R0} = 2.45 \times 10^{-16}$A and $\alpha_F = 0.996$?

5.5 Saturation parameters and its importance:

Current Gain (β): In any circuit, the current gain of a transistor is an important parameter. Current gain is usually referred to as a β or h_{fe}. Current is the ratio of the base current to the collector current and a measure of the amplifying ability of the transistor. If you want to use the transistor as an amplifier, then choose a transistor with higher current gain.

Collector-Emitter Voltage (V_{CEO}): V_{CEO} is the maximum voltage that the collector-emitter junction of a transistor can handle. For most of the transistors, V_{CEO} is usually 30V or more and measured with the base open circuit. Applying a higher voltage than V_{CEO} can damage your transistor. So before using the transistor, check the maximum V_{CEO} from the datasheet.

Emitter-Base Voltage (V_{EBO}): V_{EBO} is the maximum voltage that can be applied across the emitter-base junction. Higher voltage than the V_{EBO} can damage or destroy your transistor. V_{EBO} is relatively smaller than the V_{CEO}. Maximum V_{EBO} is usually 6V or more for most of the transistors and measured with the collector open circuit.

Collector-Base Voltage (V_{CBO}): V_{CBO} is the maximum voltage that can be applied across the collector-base junction, and it is measured with the emitter open circuit. V_{CBO} is usually 50V or more. V_{CBO} is relatively higher than V_{CEO} because the collector to base voltage is often higher than the collector to emitter voltage.

Collector current (I_C): Collector current is the maximum current that can flow through the collector. It is generally defined in milliamps, but for high power transistors, it is defined in amps. Collector current should not exceed its maximum value otherwise, it can damage the transistor. You can use a resistor to limit the collector current.

Total Power Dissipation (P_{tot}): It is the total power dissipated by the transistor. Power dissipation varies by a transistor to a transistor. For small transistors, the power rating is on the order of a few hundred milliwatts, but for high power transistors, it is defined in watts.

The power dissipation across the device can be calculated by multiplying the collector current to the voltage across the device itself.

So these are some basic parameters to select the right transistor for your application. If you are using a PCB, then you should also check the package type of the transistor.

Transistor biasing: The supply of suitable external dc voltage is called as biasing. Either forward or reverse biasing is done to the emitter and collector junctions of the transistor. These biasing methods make the transistor circuit to work in four kinds of regions such as Active region, Saturation region, Cutoff region and Inverse active region. This is understood by having a look at the following table.

Emitter junction	**Collector junction**	**Region of operation**
Forward biased	Forward biased	Saturation region
Forward biased	Reverse biased	Active region
Reverse biased	Forward biased	Inverse active region
Reverse biased	Reverse biased	Cutoff region

Among these regions, Inverse active region, which is just the inverse of active region, is not suitable for any applications and hence not used.

5.6 Conditions for cut off mode, saturation mode, inverse mode and active mode:

Operation Modes:

The four transistor operation modes are:

- **Cut-off**: The transistor acts like an open circuit**.** No current flows from collector to emitter.

- **Saturation**: The transistor acts like a short circuit**.** Current freely flows from collector to emitter
- **Reverse-Active**: Like active mode, the current is proportional to the base current, but it flows in reverse. Current flows from emitter to collector (not, exactly, the purpose transistors were designed for).
- **Active**: The current from collector to emitter is proportional to the current flowing into the base.

To determine which mode a transistor is in, we need to look at the voltages on each of the three pins, and how they relate to each other. The voltages from base to emitter (V_{BE}), and the from base to collector (V_{BC}) set the transistor's mode:

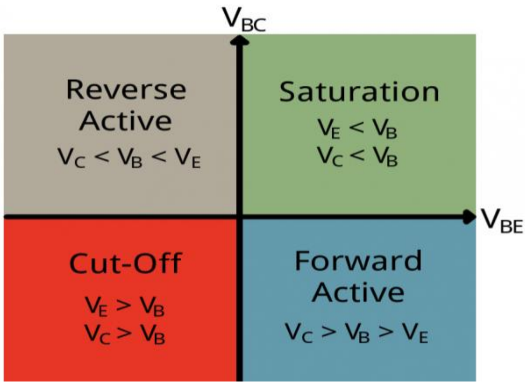


Fig.5.9 Operation modes of transistor

The simplified quadrant graph above shows how positive and negative voltages at those terminals affect the mode. In reality it's a bit more complicated than that.

Let's look at all four transistor modes individually; we'll investigate how to put the device into that mode, and what effect it has on current flow.

Note: The majority of this page focuses on **NPN transistors**. To understand how a PNP transistor works, simply flip the polarity or > and < signs.

Cutoff Mode: Cutoff mode is the opposite of saturation. A transistor in cutoff mode is off**,** there is no collector current, and therefore no emitter current. It almost looks like an open circuit.

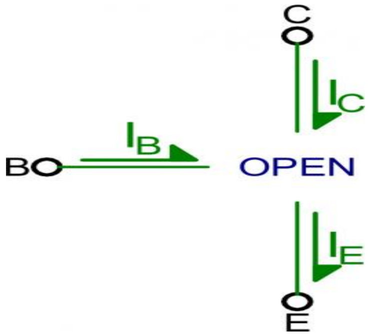


Fig.5.10 Transistor in cutoff mode

To get a transistor into cutoff mode, the base voltage must be less than both the emitter and collector voltages. V_{BC} and V_{BE} must both be negative.

$$V_C > V_B$$

$$V_E > V_B$$

In reality, V_{BE} can be anywhere between 0V and V_{th} (~0.6V) to achieve cutoff mode.

Saturation Mode: Saturation is the **on mode** of a transistor. A transistor in saturation mode acts like a short circuit between collector and emitter.

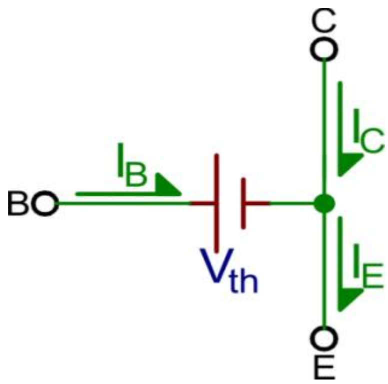


Fig.5.11 Transistor in saturation mode

In saturation mode both of the "diodes" in the transistor are forward biased. That means V_{BE} must be greater than 0, and so must V_{BC}. In other words, V_B must be higher than both V_E and V_C.

$$V_B > V_C$$

$$V_B > V_E$$

Because the junction from base to emitter looks just like a diode, in reality, V_{BE} must be greater than a threshold voltage to enter saturation. There are many abbreviations for this voltage drop V_{th}, V_{γ}, and V_d are a few -- and the actual value varies between transistors (and even further by temperature). For a lot of transistors (at room temperature) we can estimate this drop to be about 0.6V.

Another reality bummer: there won't be perfect conduction between emitter and collector. A small voltage drop will form between those nodes. Transistor datasheets will define this voltage as CE saturation voltage $\mathbf{V_{CE(sat)}}$, a voltage from collector to emitter required for saturation. This value is usually around 0.05-0.2V. This value means that V_C must be slightly greater than V_E (but both still less than V_B) to get the transistor in saturation mode.

Reverse Active: Just as saturation is the opposite of cutoff, reverse active mode is the opposite of active mode. A transistor in reverse active mode conducts, even amplifies, but current flows in the opposite direction, from emitter to collector. The downside to reverse active mode is the β (β_R in this case) is *much* smaller.

To put a transistor in reverse active mode, the emitter voltage must be greater than the base, which must be greater than the collector ($V_{BE}<0$ and $V_{BC}>0$).

$$V_C < V_B < V_E$$

Reverse active mode isn't usually a state in which you want to drive a transistor. It's good to know it's there, but it's rarely designed into an application.

Active Mode: To operate in active mode, a transistor's V_{BE} must be greater than zero and V_{BC} must be negative. Thus, the base voltage must be less than the collector, but greater than the emitter. That also means the collector must be greater than the emitter.

$$V_C > V_B > V_E$$

In reality, we need a non-zero forward voltage drop (abbreviated either V_{th}, V_{γ}, or V_d) from base to emitter (V_{BE}) to "turn on" the transistor. Usually this voltage is usually around 0.6V.

Amplifying in Active Mode: Active mode is the most powerful mode of the transistor because it turns the device into an amplifier. Current going into the base pin amplifies current going into the collector and out the emitter.

Our shorthand notation for the gain (amplification factor) of a transistor is β (you may also see it as β_F, or h_{FE}). β linearly relates the collector current (I_C) to the base current (I_B):

$$I_C = \beta I_B$$

The actual value of β varies by transistor. It's usually around 100, but can range from 50 to 200 even 2000, depending on which transistor you're using and how much current is running through it. If your transistor had a β of 100, for example, that'd mean an input current of 1mA into the base could produce 100mA current through the collector.

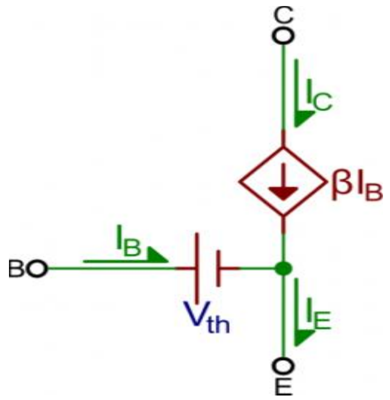


Fig.5.12 Transistor for Amplifying in Active Mode

Active mode model, $V_{BE} = V_{th}$, and $I_C = \beta I_B$.

What about the emitter current, I_E=? In active mode, the collector and base currents go into the device, and the I_E comes out. To relate the emitter current to collector current, we have another constant value: **α**. α is the common-base current gain, it relates those currents as such:

$$I_C = \alpha\, I_E$$

α is usually very close to, but less than, 1. That means I_C is very close to, but less than I_E in active mode.

You can use β to calculate α, or vice-versa:

$$\beta = \frac{\alpha}{(1-\alpha)}$$

$$\alpha = \frac{\beta}{(\beta+1)}$$

If β is 100, for example, that means α is 0.99. So, if I_C is 100mA, for example, then I_E is 101mA.

Relating to the PNP: After everything we've talked about on this page, we have still only covered half of the BJT spectrum. What about PNP transistors? PNP's work a lot like the NPN's, they have the same four modes -- but everything is turned around. To find out which mode a PNP transistor is in, reverse all of the < and > signs.

For example, to put a PNP into saturation V_C and V_E must be higher than V_B. You pull the base low to turn the PNP on, and make it higher than the collector and emitter to turn it off. And, to put a PNP into active mode, V_E must be at a higher voltage than V_B, which must be higher than V_C.

In summary:

Voltage relations	NPN Mode	PNP Mode
$V_E < V_B < V_C$	Active	Reverse
$V_E < V_B > V_C$	Saturation	Cutoff
$V_E > V_B < V_C$	Cutoff	Saturation
$V_E > V_B > V_C$	Reverse	Active

Another opposing characteristic of the NPNs and PNPs is the direction of current flow. In active and saturation modes, current in a PNP flows from emitter to collector**.** This means the emitter must generally be at a higher voltage than the collector.

5.7 Comparison among all modes of operations:

The table below gives a summary of the major properties of the different transistor configurations. Not only is gain a major aspect when designing a transistor circuit, but so too are parameters like input and output impedance.

Transistor Configuration Summary Table			
Transistor configuration	**Common base**	**Common collector**	**Common emitter**
Voltage gain	High	Low	Medium
Current gain	Low	High	Medium
Power gain	Low	Medium	High
Input / output phase relationship	0°	0°	180°
Input resistance	Low	High	Medium

Output resistance	High	Low	Medium

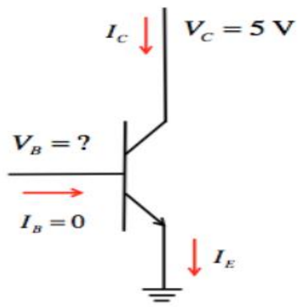
SAQ.2

a) What do you mean by Saturation parameters of transistor and its importance?

b) Discuss the Conditions for cut off mode, saturation mode, inverse mode and active mode of transistor.

c) Discuss the comparisons of transistor for all modes of operations.

d) The transistor show below has a voltage, V_C=5 V, applied to the collector, the emitter is grounded, and the base terminal is open-circuited. Use the Ebers-Moll model to answer the following questions. Assume that $I_{F0}=1.35\times10^{-16}$A, $I_{R0}=2.45\times10^{-16}$A and α_F=0.996 and α_R=0.498. What is the base voltage?



Examples:

Q.1 Use the Ebers-Moll model to determine I_C for an NPN BJT biased at I_B = 100 nA and V_{CE} = 0.10 V. Assume that the Ebers-Moll model parameters are $I_{F0} = 1.25\times10^{-16}$A , $I_{R0} = 2.50\times10^{-16}$A, and α_F = 0.996 .

Solution:

The Ebers-Moll equations are:

$$I_C(V_{BE},V_{BC}) = \alpha_F I_{F0}\left(e^{qV_{BE}/k_BT}-1\right) - I_{R0}\left(e^{qV_{BC}/k_BT}-1\right) \quad \ldots\ldots\ldots\ldots(1)$$

$$I_E(V_{BE},V_{BC}) = I_{F0}\left(e^{qV_{BE}/k_BT}-1\right) - \alpha_R I_{R0}\left(e^{qV_{BC}/k_BT}-1\right) \quad \ldots\ldots\ldots\ldots\ldots(2)$$

Use KVL to write:

$$V_{CE} = V_{BE} + V_{CB}$$
$$V_{CE} = V_{BE} - V_{BC}$$

We are given V_{CE}, but we need V_{BE} and V_{BC} (the voltages across the two PN junctions). If we knew V_{BE}, then we could determine V_{BC} from KVL.

Use the given base current and V_{CE} to determine V_{BE}.

KCL: $I_B = I_E - I_C$

$$I_B=(1-\alpha_F)I_{F0}\left(e^{qV_{BE}/k_BT}-1\right)+(1-\alpha_R)I_{R0}\left(e^{qV_{BC}/k_BT}-1\right)=100\times10^{-9}$$
$$V_{BC}=V_{BE}-V_{CE}$$
$$I_B=(1-\alpha_F)I_{F0}\left(e^{qV_{BE}/k_BT}-1\right)+(1-\alpha_R)I_{R0}\left(e^{q(V_{BE}-V_{CE})/k_BT}-1\right)=100\times10^{-9}$$
$$I_B=(1-\alpha_F)I_{F0}e^{qV_{BE}/k_BT}-(1-\alpha_F)I_{F0}+(1-\alpha_R)I_{R0}e^{q(V_{BE}-V_{CE})/k_BT}-(1-\alpha_R)I_{R0}$$
$$I_B=\left\{(1-\alpha_F)I_{F0}+(1-\alpha_R)I_{R0}e^{-qV_{CE}/k_BT}\right\}e^{qV_{BE}/k_BT}-\left\{(1-\alpha_F)I_{F0}+(1-\alpha_R)I_{R0}\right\}$$
$$e^{qV_{BE}/k_BT}=\frac{I_B+\left\{(1-\alpha_F)I_{F0}+(1-\alpha_R)I_{R0}\right\}}{(1-\alpha_F)I_{F0}+(1-\alpha_R)I_{R0}e^{-qV_{CE}/k_BT}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

From the given base current and saturation currents, it is clear that

$$\left\{(1-\alpha_F)I_{F0}+(1-\alpha_R)I_{R0}\right\}<<I_B$$

So the numerator in (3) can be simplified.
The result is that (3) becomes (to a good approximation)

$$e^{qV_{BE}/k_BT}=\frac{I_B}{(1-\alpha_F)I_{F0}+I_{R0}e^{-qV_{CE}/k_BT}-\alpha_F I_{F0}e^{-qV_{CE}/k_BT}}$$

Where we have used the "reciprocity relation" $\alpha_F I_{F0}=\alpha_R I_{R0}$.
Also, from the given V_{CE}

$$e^{-qV_{CE}/k_BT}=e^{-0.1/0.026}=2.14\times10^{-2}$$

Putting in numbers:

$$e^{qV_{BE}/k_BT}=\frac{10^{-7}}{(1-0.996)\left(1.25\times10^{-16}\right)+\left(2.50\times10^{-16}\right)\left(2.14\times10^{-2}\right)-0.996\left(1.25\times10^{-16}\right)\left(2.14\times10^{-2}\right)}$$

$$e^{qV_{BE}/k_BT}=3.11\times10^{10}$$

$$V_{BE}=\frac{k_BT}{q}\ln\left(3.11\times10^{10}\right)=0.026\times24.2=0.628\text{ V}$$

Now we can go back to (1) and compute the collector current:

$$I_C=\alpha_F I_{F0}\left(e^{qV_{BE}/k_BT}-1\right)-I_{R0}\left(e^{qV_{BC}/k_BT}-1\right)$$
$$V_{BC}=V_{BE}-V_{CE}=0.628-0.10=0.528$$

Both junctions are forward biased, this transistor is operating in the saturation region.

$$I_C=\alpha_F I_{F0}\left(e^{0.628/0.026}-1\right)-I_{R0}\left(e^{0.528/0.026}-1\right)$$
$$I_C=0.996\left(1.25\times10^{-16}\right)\left(3.09\times10^{10}\right)-\left(2.50\times10^{-16}\right)\left(6.60\times10^{8}\right)$$
$$I_C=3.85\times10^{-6}-1.65\times10^{-7}$$

$$\boxed{I_C=3.68\times10^{-6}\text{ A}}$$

Q.2 What is β_R for a bipolar transistor described by an Ebers-Moll model with I_{F0}=1.25×10^{-16}A , I_{R0} = 2.50 ×10^{-16}A and α_F = 0.996 ?

Solution:

$$\beta_R = \frac{\alpha_R}{1-\alpha_R} \text{ but we are not given } \alpha_R$$

According to the reciprocity relation: $\alpha_F I_{F0} = \alpha_R I_{R0}$, so

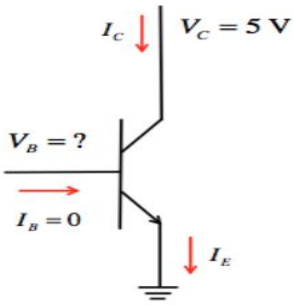
$$\alpha_R = \alpha_F \frac{I_{F0}}{I_{R0}} = 0.996 \frac{1.25\times10^{-16}}{2.50\times10^{-16}} = 0.498$$

$$\beta_R = \frac{0.498}{1-0.498} = 0.992$$

$$\boxed{\beta_R = 0.992}$$

for this transistor, $\beta_F = \alpha_F/(1-\alpha_F) = 249$.

Q.3 The transistor show below has a voltage, V_C=5 V, applied to the collector, the emitter is grounded, and the base terminal is open-circuited. Use the Ebers-Moll model to answer the following questions. Assume that I_{F0}=1.25×10^{-16}A, I_{R0}=2.50×10^{-16}A and α_F=0.996 and α_R=0.498. What is the base voltage?



Solution:

The Ebers-Moll equations are:

$$I_C(V_{BE},V_{BC}) = \alpha_F I_{F0}\left(e^{qV_{BE}/k_BT}-1\right) - I_{R0}\left(e^{qV_{BC}/k_BT}-1\right) \quad \text{…………(1)}$$

$$I_E(V_{BE},V_{BC}) = I_{F0}\left(e^{qV_{BE}/k_BT}-1\right) - \alpha_R I_{R0}\left(e^{qV_{BC}/k_BT}-1\right) \quad \text{………….(2)}$$

Use KCL: $I_B = I_E - I_C$

$$I_B = (1-\alpha_F) I_{F0}\left(e^{qV_{BE}/k_BT}-1\right) + (1-\alpha_R) I_{R0}\left(e^{qV_{BC}/k_BT}-1\right) = 0$$
$$(1-\alpha_F) I_{F0}\left(e^{qV_{BE}/k_BT}-1\right) = -(1-\alpha_R) I_{R0}\left(e^{qV_{BC}/k_BT}-1\right) \quad \text{………..(3)}$$

$$V_{BE} = V_B - V_E = V_B$$
$$V_{BC} = V_B - V_C = V_B - 5$$

Equation (3) becomes:

$$\left(e^{qV_B/k_BT}-1\right)=-\frac{(1-\alpha_R)}{(1-\alpha_F)}\left(\frac{I_{R0}}{I_{F0}}\right)\left(e^{qV_B/k_BT}e^{-q5/k_BT}-1\right)$$

Solve for the base voltage:

$$e^{qV_B/k_BT}\left(1+\frac{(1-\alpha_R)}{(1-\alpha_F)}\left(\frac{I_{R0}}{I_{F0}}\right)e^{-q5/k_BT}\right)=1+\frac{(1-\alpha_R)}{(1-\alpha_F)}\left(\frac{I_{R0}}{I_{F0}}\right)$$

$$e^{qV_B/k_BT}=\frac{1+\frac{(1-\alpha_R)}{(1-\alpha_F)}\left(\frac{I_{R0}}{I_{F0}}\right)}{\left(1+\frac{(1-\alpha_R)}{(1-\alpha_F)}\left(\frac{I_{R0}}{I_{F0}}\right)e^{-q5/k_BT}\right)}$$

The quantity, e^{-q5/k_BT}, is really, reallysmall.

$e^{-q5/k_BT}=e^{-5/0.026}=3.30\times10^{-84}$,so the denominator simplifies

$$e^{qV_B/k_BT}=\frac{1+\frac{(1-\alpha_R)}{(1-\alpha_F)}\left(\frac{I_{R0}}{I_{F0}}\right)}{1}$$

Putting in numbers:

$$e^{qV_B/k_BT}=1+\frac{(1-0.498)}{(1-0.996)}\left(\frac{2.50}{1.25}\right)=252$$

$$V_B=0.026\ln(252)=0.144\text{ V}$$

$$\boxed{V_B=0.144\text{ V}}$$

This makes sense. Most of the voltage drops across the reverse-biased (high resistance junction) –not the forward-biased (low resistance junction).

Q.4 For a transistor, β = 45 and voltage drop across 1kΩ which is connected in the collector circuit is 1 volt. Find the base current for common emitter connection.

Solution:

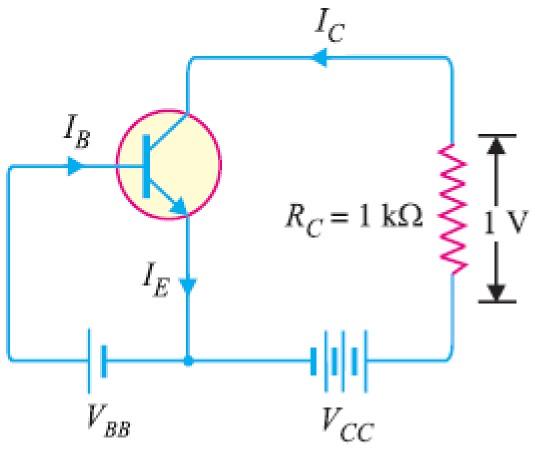


Fig. shows the required common emitter connection. The voltage drop across RC (= 1 kΩ) is 1 volt.

$$\therefore \qquad I_C = \frac{1\,V}{1\,k\Omega} = 1\text{ mA}$$

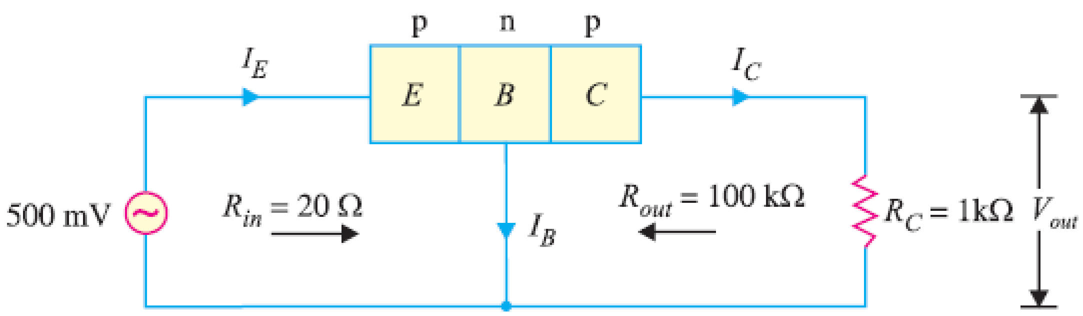
$$\text{Now} \qquad \beta = \frac{I_C}{I_B}$$

$$\therefore \qquad I_B = \frac{I_C}{\beta} = \frac{1}{45} = \mathbf{0.022\ mA}$$

Q.5. A common base transistor amplifier has an input resistance of 20 Ω and output resistance of 100 kΩ. The collector load is 1 kΩ. If a signal of 500 mV is applied between emitter and base, find the voltage amplification. Assume α_{ac} to be nearly one.

Solution:

Fig. shows the conditions of the problem. Here the output resistance is very high as compared to input resistance, since the input junction (base to emitter) of the transistor is forward biased while the output junction (base to collector) is reverse biased.



5.8 Summary:

1) The Ebers-Moll BJT Model is a good large-signal, steady-state model of the transistor and allows the state of conduction of the device to be easily determined for different modes of operation of the device. The different modes of operation are determined by the manner in which the junctions are biased.
2) The Ebers-Moll equations are based on two exponential diodes plus two current-controlled current sources. The NPN Bipolar Transistor block provides the following enhancements to that model: Early voltage effect. Optional base, collector, and emitter resistances.
3) Since the physical construction of the transistor determines the electrical relationship between these three currents, (Ib), (Ic) and (Ie), any small change in the base current (Ib), will result in a much larger change in the collector current (Ic).
4) 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.
5) If the transistor is made up of a silicon material, the base-emitter voltage (VBE) will be 0.7 V. If the transistor is made up of a germanium material, the base-emitter voltage (VBE) will be 0.3 V.
6) A typical value for a small signal transistor is 60 to 80 V. In power transistors, this could range to 1000 V, for example, a horizontal deflection transistor in a cathode ray tube display.
7) Determine the voltage drop between the collector and emitter junctions (Vce) of the transistor using the formula Vce = Vcc - IcRc, where "Vce" is the collector emitter voltage; "Vcc" is the supply voltage; and "IcRc" is the voltage drop across the base resistor (Rb).
8) Transistor datasheets will define this voltage as CE saturation voltage $V_{CE(sat)}$ a voltage from collector to emitter required for saturation. This value is usually around 0.05-0.2V. This value means that V_C must be slightly greater than V_E (but both still less than V_B) to get the transistor in saturation mode.
9) To get a transistor into cutoff mode, the base voltage must be less than both the emitter and collector voltages. V_{BC} and V_{BE} must both be negative. In reality, V_{BE} can be anywhere between 0V and V_{th} (~0.6V) to achieve cutoff mode.
10) Saturation is the on mode of a transistor. A transistor in saturation mode acts like a short circuit between collector and emitter. In saturation mode both of the

"diodes" in the transistor are forward biased. That means V_{BE} must be greater than 0, and so must V_{BC}.

11) In the inverse active mode of transistor operation, the base-emitter junction is reverse biased and the base-collector junction is forward biased.

12) A transistor is said to be in its active mode if it is operating somewhere between fully on (saturated) and fully off (cutoff). Base current regulates collector current. By regulate, we mean that no more collector current can exist than what is allowed by the base current.

13) The four transistor operation modes are:

    Saturation: The transistor acts like a short circuit. Current freely flows from collector to emitter.

    Cut-off: The transistor acts like an open circuit. No current flows from collector to emitter.

    Active: The current from collector to emitter is proportional to the current flowing into the base.

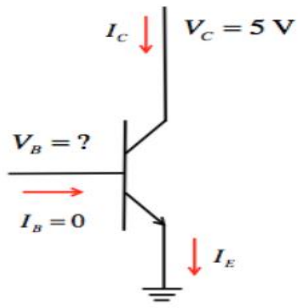
    Reverse-Active: Like active mode, the current is proportional to the base current, but it flows in reverse. Current flows from emitter to collector (not, exactly, the purpose transistors were designed for).

5.9 Terminal Questions:

1) Explain the working principle of Eber's moll model for PNP transistors.
2) Explain the working principle of Eber's moll model for NPN transistors.

f) Discuss and derive the expressions of various currents and voltage for Eber's moll model for PNP and NPN transistors.

3) Explain in detail for saturation parameters of transistor and its importance.
4) Discuss in detail conditions for cut off mode, saturation mode, inverse mode and active mode for transistor.
5) What are the comparisons between all modes of operations for transistor?
6) Use the Ebers-Moll model to determine I_C for an NPN BJT biased at I_B=100 nA and V_{CE} = 0.10 V. Assume that the Ebers-Moll model parameters are $I_{F0}=1.25\times10^{-16}$A, $I_{R0}=2.50\times10^{-16}$A, and $\alpha_F=0.996$.
7) What is β_R for a bipolar transistor described by an Ebers-Moll model with $I_{F0}=1.25\times10^{-16}$A , $I_{R0} = 2.50\times10^{-16}$A and $\alpha_F = 0.996$?
8) The transistor show below has a voltage, V_C=5 V, applied to the collector, the emitter is grounded, and the base terminal is open-circuited. Use the Ebers-Moll model to answer the following questions. Assume that $I_{F0}=1.25\times10^{-16}$A, $I_{R0}=2.50\times10^{-16}$A and $\alpha_F=0.996$ and $\alpha_R=0.498$. What is the base voltage?



9) A common base transistor amplifier has an input resistance of 10 Ω and output resistance of 500 kΩ. The collector load is 1 kΩ. If a signal of 300 mV is applied between emitter and base, find the voltage amplification. Assume α_{ac} to be nearly one.
10) For a transistor, β = 45 and voltage drop across 2kΩ which is connected in the collector circuit is 5 volt. Find the base current for common emitter connection.

Unit 6 Transmission and reception

Structure:

6.1 Introduction:

In radio communication systems, information is carried across space using radio waves. At the sending end, the information to be sent is converted by some type of transducer to a time-varying electrical signal called the modulation signal.

Radio links are used to transmit telephone messages, telegraph messages, digital data, facsimile images, and television programs. Waves in the very high frequency range or higher are generally used for television programs.

A radio communication station is a set of equipment necessary to carry on communication via radio waves. Generally, it is a receiver or transmitter or transceiver, an antenna, and some smaller additional equipment necessary to operate them.

Based on physical infrastructure there are two types of communication systems: Line communication systems: Uses the existing infrastructure of power lines to transfer data from one point to another point.

A radio transmitter design has to meet certain requirements. These include the frequency of operation, the type of modulation, the stability and purity of the resulting signal, the efficiency of power use, and the power level required to meet the system design objectives.

The four major process variables measured and represented by a transmitter are Pressure, Level, Temperature, and Flow.

The mouth (and vocal cords) is the transmitter, ears are the receivers, and air is the transmission medium over which sound travels between mouth and ear. The transmitter and receiver elements of a data modem (such as the type used in a traffic signal system controller box) may not be readily visible.

Modulation is the process of converting data into radio waves by adding information to an electronic or optical carrier signal. A carrier signal is one with a steady waveform constant height, or amplitude, and frequency.

Amplitude modulation (AM) is a modulation technique used in electronic communication, most commonly for transmitting messages with a radio wave. In amplitude modulation, the amplitude (signal strength) of the wave is varied in proportion to that of the message signal, such as an audio signal.

Frequency Modulation (FM) is the encoding of information in a carrier wave by changing the instantaneous frequency of the wave. FM technology is widely used in the fields of computing, telecommunications, and signal processing.

Phase modulation (PM) is a modulation pattern for conditioning communication signals for transmission. It encodes a message signal as variations in the instantaneous phase of a carrier wave. Phase modulation is one of the two principal forms of angle modulation, together with frequency modulation.

Modulation index describes the extent to which modulation is done on a carrier signal. In an amplitude modulation, it is defined as the ratio of the amplitude of modulating signal to that of the carrier signal.

A frequency spectrum in mobile communications is the range of radio frequencies allocated to each mobile network operator (MNO) in their country of operation for transmitting and receiving their RF signals. An MNO can add more cells with more spectrums to improve network capacity and coverage.

If the modulation index $\mu=1$ then the power of AM wave is equal to 1.5 times the carrier power. So, the power required for transmitting an AM wave is 1.5 times the carrier power for a perfect modulation.

A modulator is a circuit that combines two different signals in such a way that they can be pulled apart later and the information obtained.

The process of separating the original information or signal from the modulated carrier. In the case of amplitude or frequency modulation it involves a device, called a demodulator or detector, which produces a signal corresponding to the instantaneous changes in amplitude or frequency, respectively.

Demodulation is extracting the original information-bearing signal from a carrier wave. A demodulator is an electronic circuit (or computer program in a software-defined radio) that is used to recover the information content from the modulated carrier wave.

The FM demodulator is done with the help of a circuit called Phase Locked Loop (PLL). A PLL should have basic functional blocks like Voltage Controlled Oscillator (VCO), Phase comparator, Low Pass Filter (LPF) and Source follower.

Demodulation is extracting the original information-bearing signal from a carrier wave. A demodulator is an electronic circuit (or computer program in a software-defined radio) that is used to recover the information content from the modulated carrier wave.

6.2 Objectives:

After studying this unit you should be able to

- Explain and identify Basic elements of radio communication systems.
- Study and identify Requirements of transmitter, medium and receiver.
- Explain Modulation (need, types and statements).
- Explain and identify Analysis of AM, FM and PM, modulation index.
- Study and identify Frequency spectrum and power in modulations.
- Explain Circuit of modulator.
- Explain and identify Demodulation (need and statements).
- Study and identify Circuit for demodulator.

6.3 Basic elements of radio communication systems:

Fig.6.1 shows the elements of a radio communication system. Unlike the telephone transmitter, the microphone used in radio communication has an electromagnet as main component. The microphone transforms sound waves into electrical signals. These signals are fed to the modulator. At the same time, an oscillator circuit generates the fixed frequency carrier waves, which are also fed to the modulator. The modulator puts out the modulated carrier waves.

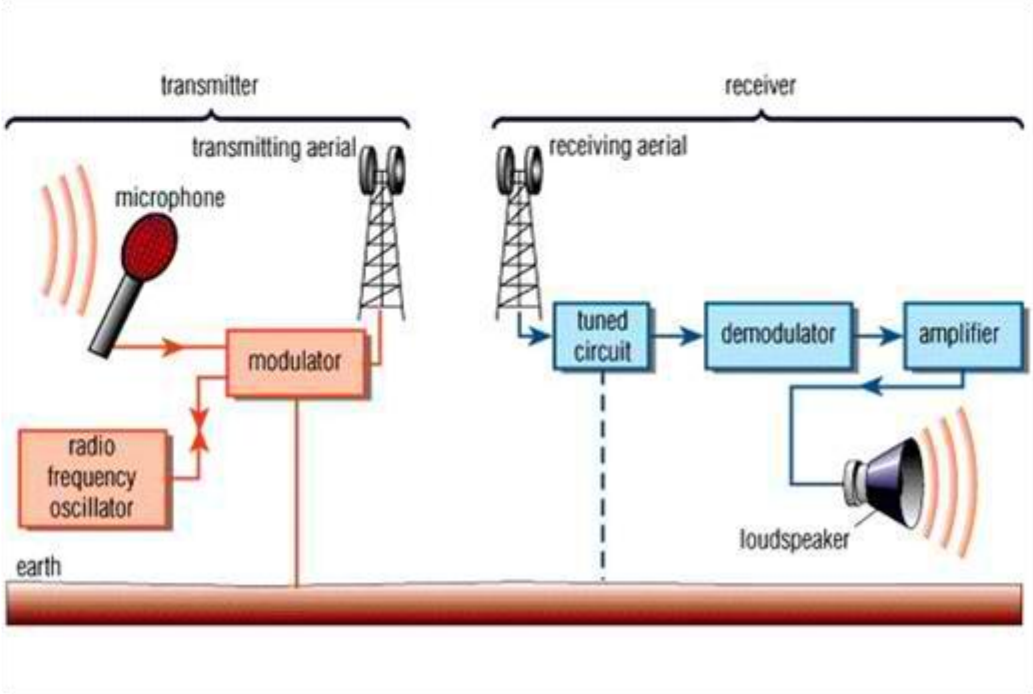


Fig.6.1 Basic elements of a radio communication system

Radio transmitters: A radio transmitter consists of several elements that work together to generate radio waves that contain useful information such as audio, video, or digital data.

Power supply: Provides the necessary electrical power to operate the transmitter.

Oscillator: Creates alternating current at the frequency on which the transmitter will transmit. The oscillator usually generates a sine wave, which is referred to as a carrier wave.

Modulator: Adds useful information to the carrier wave. There are two main ways to add this information. The first, called amplitude modulation or AM, makes slight increases or decreases to the intensity of the carrier wave. The second, called frequency modulation or FM, makes slight increases or decreases the frequency of the carrier wave.

Amplifier: Amplifies the modulated carrier wave to increase its power. The more powerful the amplifier, the more powerful the broadcast.

Antenna: Converts the amplified signal to radio waves.

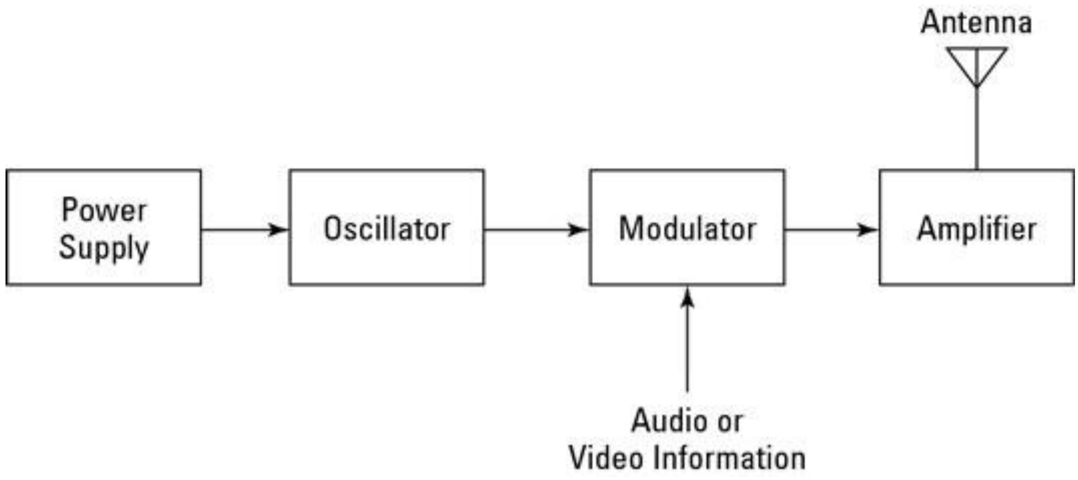


Fig.6.2 Main components of Radio transmitter

Radio receivers: A radio receiver is the opposite of a radio transmitter. It uses an antenna to capture radio waves, processes those waves to extract only those waves that are vibrating at the desired frequency, extracts the audio signals that were added to those waves, amplifies the audio signals, and finally plays them on a speaker.

Antenna: Captures the radio waves. Typically, the antenna is simply a length of wire. When this wire is exposed to radio waves, the waves induce a very small alternating current in the antenna.

RF amplifier: A sensitive amplifier that amplifies the very weak radio frequency (RF) signal from the antenna so that the signal can be processed by the tuner.

Tuner: A circuit that can extract signals of a particular frequency from a mix of signals of different frequencies. On its own, the antenna captures radio waves of all frequencies and sends them to the RF amplifier, which dutifully amplifies them all.

Unless you want to listen to every radio channel at the same time, you need a circuit that can pick out just the signals for the channel you want to hear. That's the role of the tuner.

The tuner usually employs the combination of an inductor (for example, a coil) and a capacitor to form a circuit that resonates at a particular frequency. This frequency, called the resonant frequency, is determined by the values chosen for the coil and the capacitor. This type of circuit tends to block any AC signals at a frequency above or below the resonant frequency.

You can adjust the resonant frequency by varying the amount of inductance in the coil or the capacitance of the capacitor. In simple radio receiver circuits, the tuning is adjusted by varying the number of turns of wire in the coil. More sophisticated tuners use a variable capacitor (also called a tuning capacitor) to vary the frequency.

Detector: Responsible for separating the audio information from the carrier wave. For AM signals, this can be done with a diode that just rectifies the alternating current signal. What's left after the diode has its way with the alternating current signal is a direct current signal that can be fed to an audio amplifier circuit. For FM signals, the detector circuit is a little more complicated.

Audio amplifier: This component's job is to amplify the weak signal that comes from the detector so that it can be heard. This can be done using a simple transistor amplifier circuit.

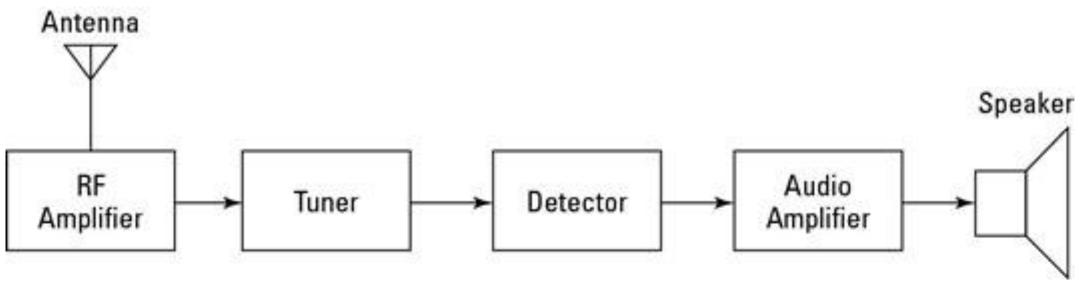


Fig.6.3 Main components of Radio receiver

Of course, there are many variations on this basic radio receiver design. Many receivers include additional filtering and tuning circuits to better lock on to the intended frequency or to produce better-quality audio output and exclude other signals. Still, these basic elements are found in most receiver circuits.

6.4 Requirements of transmitter, medium and receiver:

Requirements of transmitter: A transmitter is an electronic device used in telecommunications to produce radio waves in order to transmit or send data with the aid of an antenna. The transmitter is able to generate a radio frequency alternating current that is then applied to the antenna, which, in turn, radiates this as radio waves. A radio transmitter design has to meet certain requirements. These include the frequency of operation, the type of modulation, the stability and purity of the resulting signal, the efficiency of power use, and the power level required to meet the system design objectives.

Requirements of medium: A medium in communication is a system or channel through which a speaker or writer addresses their audience. It's an outlet that a sender uses to express meaning to their audience, and it can include written, verbal or nonverbal elements. A communication medium can either be virtual or physical.

Requirements of receiver: The ability of a receiver to select a signal of a desired frequency while rejecting those on closely adjacent frequencies. With good selectivity, the receiver can select the desired signal and eliminate all other RF signals. The main criteria are gain, selectivity, sensitivity, and stability. The receiver must contain a detector to recover the information initially impressed on the radio carrier signal, a process called modulation.

6.5 Modulation (need, types and statements):

What is Modulation and It's Types?

Modulation is one of the crucial branches of electronics science that is widely used in communication systems. It includes the different fundamental properties of the signal to transpose it from one location to another.

Types of Signals used in the Modulation

Modulating Signal: This is the signal that contains the message to be transmitted from the sender to the receiver and is called a message signal. Generally, the message signals are the band of low or high frequencies and are often called baseband signals. The message signals are the signals to be transmitted from the sender to the receiver. The frequency of the message signals to be sent is generally low. Thus, these signals undergo modulation to get correctly transmitted from one location to another.

Carrier Signal: The other signal used in the process of modulation is the carrier signal that has high-frequency sinusoidal waves. The high-frequency carrier wave can travel much quicker as compared to the baseband signal. These signals have a specific frequency, amplitude, and phase, but no information. After modulation, carrier signals are used to transmit the signal to the receiver.

Modulated Signal: After the modulation is done, the resultant signal refers to the modulated signal. This signal is the mixture of the carrier signal and message signal.

What is the Need for Modulation?

Increase the Signal Strength: The baseband signals transmitted by the sender are not capable of direct transmission. The strength of the message signal should be increased so that it can travel longer distances. This is where modulation is essential. The most vital need of modulation is to enhance the strength of the signal without affecting the parameters of the carrier signal.

Wireless Communication System: Modulation has removed the necessity for using wires in the communication systems. It is because modulation is widely used in transmitting signals from one location to another with faster speed. Thus, the modulation technique has helped in enhancing wireless communication systems.

Prevention of Message Signal From Mixing: Modulation and its types prevent the interference of the message signal from other signals. It is because a person sending a message signal through the phone cannot tell such signals apart. As a result, they will interfere with each other. However, by using carrier signals having a high frequency, the mixing of the signals can be prevented. Thus, modulation ensures that the signals received by the receiver are entirely perfect.

Size of the Antenna: The signals within 20 Hz to 20 kHz frequency range can travel only a few distances. To send the message signal, the length of the antenna should be a quarter wavelength of the used frequency. Thus, modulation is required to increase the frequency of the message signal and to enhance its strength to reach the receiver.

Length of the antenna can be easily calculated using this formula:

$$L = \lambda = u/\nu$$

$$= (3 \times 10^8) / \nu$$

Here, L = length of antenna

λ = wavelength of the transmitted signal

ν = carrier wave frequency

What Are The Types Of Modulation?

There are generally three types of modulation:

Amplitude Modulation: By superimposing the base signal with the carrier signal having a different amplitude, but the same frequency, if the amplitude of the base signal modifies or modulates, then it is said to be amplitude modulation.

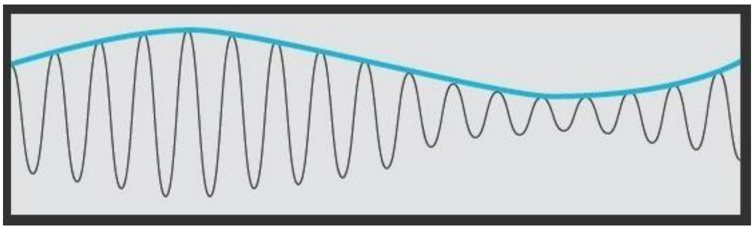


Fig.6.4 Amplitude modulation

The diagram shows the amplified modulated wave after superimposing the message signal with the carrier signal.

Frequency Modulation: By superimposing the base signal with the carrier signal having a different frequency, but the same amplitude, if the frequency of the base signal modifies or modulates, then it is said to be frequency modulation.

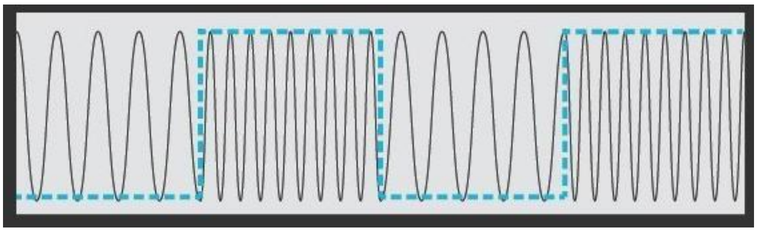


Fig.6.5 Frequency Modulation

Phase Modulation: It is the type of modulation in which the phase of the base signal changes while superimposing it with a carrier signal.

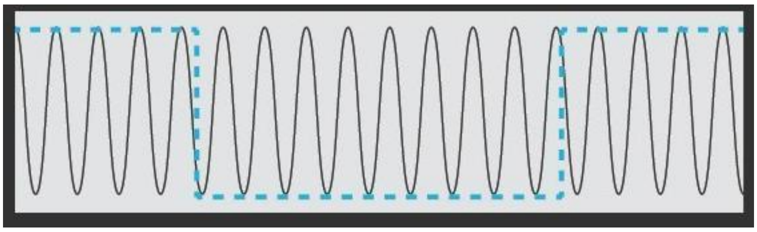


Fig.6.6 Phase Modulation

Polarization modulation: The angle of rotation of an optical carrier signal is varied to reflect transmitted data.

Pulse-code modulation: An analog signal is sampled to derive a data stream that is used to modulate a digital carrier signal.

Quadrature amplitude modulation (QAM): Uses two AM carriers to encode two or more bits in a single transmission.

What are the Uses of Modulation?

One of the most common uses of different types of modulation is the inter-conversion of signals from its existing to another form.

Digital Modulation is used for the transmissions of the digital signals over analog baseband.

Analog Modulation is used to transfer the low bandwidth signals such as TV or radio signals over a higher bandwidth.

Modern modulation techniques are widely used to carry out FDM, that is, Frequency Division Multiplexing.

SAQ.1

a) What do you mean by Basic elements of radio communication systems?
b) Discuss the requirements of transmitter, medium and receiver in radio communication systems.
c) What do you mean by the need of Modulation?
d) Discuss the types of Modulation.
e) A sinusoidal modulating waveform of amplitude 10V and a frequency of 5 KHz is applied to FM generator, which has a frequency sensitivity of 30 Hz/volt. Calculate the frequency deviation, modulation index, and bandwidth.

6.6 Analysis of AM, FM and PM, modulation index:

What is AM, FM, and PM?

In FM, a radio wave known as the "carrier" or "carrier wave" is modulated in frequency by the signal that is to be transmitted. However, the amplitude and the phase constant remain constant in the case of frequency modulation. While amplitude modulation or AM is a

technique that we use in electronic communication, most commonly for transmitting information through a radio carrier wave.
Additionally, there is a term that is one of the two principal forms of angle modulation alongside frequency modulation, and it is called phase modulation. PM is a modulation pattern that we use for conditioning communication signals for transmission.

This help for distinguish between amplitude modulation and frequency modulation. Also, you will get to learn the difference between FM, AM, and PM modulation in tabular form.

How does FM Work?

Frequency modulation works by continuously changing the strength of the transmitted signal in relation to the information being sent.
FM radio uses frequency modulation. However, to understand frequency modulation, assume a signal with a low frequency and amplitude. As the signal passes, its frequency remains unchanged or un-modulated. We find that a signal carries very little information.
So, when you introduce information to this signal, there's a variation to the frequency that varies directly with the information. Also, when the frequency is modulated between low and high, the carrier frequency transmits music or voice. Therefore, we find that the frequency changes as a result, but the amplitude remains constant the entire time.
For example, the changes in the strength of the signal may be utilized to specify the sounds to be reproduced by the speaker, or the light intensity of television pixels.

How does an AM Work?

While receiving input signals, an AM receiver discovers amplitude variations in the radio waves at a specific frequency. It amplifies changes in the signal voltage to operate a loudspeaker or earphone.

Working principle of an AM receiver in reality:

In this, a radio receiver is present in the opposite of a radio transmitter that uses an antenna to capture radio waves. Further, it processes those waves to elicit only those waves that are vibrating at the needed frequency and filters the audio signals that were added to those waves, following, amplifying the audio signals, and finally plays them on a speaker.

How does PM Work?

Phase modulation encodes a message or an input signal as variations in the instantaneous phase of a carrier wave.

In phase modulation, the working principle is that the instantaneous amplitude of the baseband signal modifies the phase of the carrier signal thereby keeping the constant amplitude and frequency. The phase of a carrier signal is modulated by phase modulation to follow the altering signal level or amplitude of the message signal.

Furthermore, the peak amplitude and the frequency of the carrier signal are also kept constant. However, as the amplitude of the message signal alters, the phase of the carrier also changes.

Modulation Index: The relationship between the information signal amplitude, Vm , and the unmodulated carrier amplitude, Vc, is expressed as a ratio called the modulation index (m), defined as:

$$m = Vm/Vc$$

Sometimes m is expressed as a percentage: percent modulation = m x 100%. The following figure shows the AM signal at three different values of percent modulation: 20%, 50% and 90%. Overall, the greater the value of m, the closer the envelope gets to the horizontal (time) axis.

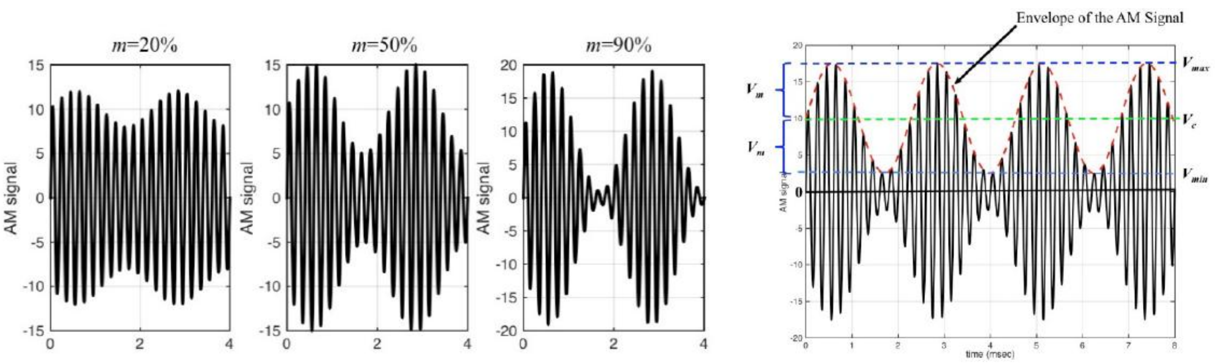


Fig.6.7 AM signal at three different values of percent modulation: 20%, 50% and 90%

We can also mathematically determine the modulation index m from the maximum and minimum values of the envelope of $v_{AM}(t)$ as follows, where Vmax is the maximum value of the envelope and Vmin is the minimum value:

$$V_m = \frac{V_{max} - V_{min}}{2}$$
$$V_c = \frac{V_{max} + V_{min}}{2}$$
$$m = \frac{V_m}{V_c} = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$$

In order for the AM signal to convey the original signal accurately and prevent distortion, the information signal amplitude (Vm) must be less than the un-modulated carrier signal amplitude (Vc). Here again, the un-modulated carrier refers to the AM signal if the information signal amplitude is equal to 0 (Vm = 0), in which case, $v_m(t) = V_m \cos(2\pi f_m t)$. The maximum usable modulation index is m = 1.0, corresponding to 100% modulation, when Vm is equal to Vc. When Vm is greater than Vc (that is, m > 1), over modulation occurs. Over modulation, depicted below, results in distortion of the AM signal's envelope, and since the envelope holds the information, the recovered information signal is also distorted.

Difference between AM and FM:

Below is the table of AM versus FM:

S.No.	Parameters	AM	FM
1.	Full-form	Amplitude modulation	Frequency modulation
2.	Origin	The AM method of audio transmission was successfully carried out in the mid-1870s.	FM radio was developed in the United States in the 1930s by Edwin Armstrong.
3.	Modulating differences	In AM, a radio wave known as the "carrier" or "carrier wave" is modulated in amplitude by the signal that is to be transmitted.	In FM, a radio wave known as the "carrier" or "carrier wave" is modulated in frequency by the signal that is to be transmitted.
4.	Constant parameters	The frequency and phase remain the same.	The amplitude and phase remain the same.
5.	Quality	AM has poorer sound quality, and a lower bandwidth but is cheaper. It can be transmitted over long distances as it has a lower bandwidth, which is why it can hold more stations available in any frequency range.	FM is less affected by interference, but FM signals are impacted by physical barriers. They have a better sound quality due to higher bandwidth.

6.	Frequency range	AM radio ranges from 535 to 1700 kHz or up to 1200 bits per second.	FM radio ranges in a higher spectrum from 88.1 to 108.1MHz or up to 1200 to 2400 bits per second.
7.	Bandwidth BW	BW is much less than FM. B.W. = 2 fm	BW is large. Hence a wide channel is required. B.W. = 2 x (δ + fm)
8.	Bandwidth requirements	Bandwidth is less than FM or PM and doesn't depend upon the modulation index. The bandwidth requirement is twice the highest modulating frequency.	Bandwidth requirement is greater and depends upon the modulating. The bandwidth requirement is twice the sum of the modulating signal frequency and the frequency deviation.
9.	The frequency required for broadcasting	In AM radio broadcasting, if the modulating signal has a bandwidth of 15 kHz, then the bandwidth of an amplitude- The modulated signal is 30 kHz.	Let's say, if the frequency deviation is 75kHz and the modulating signal frequency is 15kHz, the bandwidth required is 180kHz.
10.	No of Sidebands	The number of sidebands is constant and equal to 2.	The number of sidebands having significant amplitude depends upon the modulation index
11.	Zero crossings in modulating signal	Equidistant	Not equidistant
12.	Complexity	AM transmitters and receivers are less complex than FM and PM, but synchronization is needed in the case of SSBSC carriers.	FM (or PM) transmitters are more complex than AM because the variation of modulating signal has to be converted and detected from the corresponding variation in frequencies.
13.	Noise	AM receivers are very less susceptible to noise because noise affects the amplitude, which is where information is stored in AM signals.	FM receivers are better immune to noise and it is possible to decrease noise by further deviation.
14.	Efficiency	Power is wasted in transmitting the carrier.	All transmitted power is useful so that's why FM is very efficient.

15.	Application	MW (Medium wave), SW (short wave) band broadcasting, video transmission in T.V.	Broadcasting FM, audio transmission on T.V.

Difference between AM, FM, and PM:

S.No.	Parameters	FM	AM	PM
1.	Definition	Frequency modulation is a technique of modulation, in which the frequency of the carrier varies in accordance with the amplitude of the modulating signal. The amplitude and phase are constant.	Amplitude modulation is a technique of modulation in which the amplitude of the carrier wave varies in accordance with the amplitude of the modulating signal. The frequency and phase are constant.	Phase modulation is a technique of modulation in which the phase of the carrier wave varies in accordance with the amplitude of the modulating signal. The amplitude and frequency are constant.
2.	Noise	Noise immunity of FM is superior to AM and PM.	AM receivers are very susceptible to noise.	Noise immunity is better than AM but not FM.
3.	Function	The frequency of the carrier wave deviates as per the voltage of the modulating signal input.	The amplitude of a carrier wave in AN diverges as per amplitude or voltage of modulating signal input.	A phase of the carrier wave varies as per the voltage of modulating signal input.
4.	Constant parameter	The amplitude of the carrier wave is kept changeless.	The frequency of the carrier wave is kept invariable.	The amplitude of the carrier wave is kept changeless.
5.	Types	Digital FM types: FSK, GFSK, offset PSK, etc.	AM types: DSB-SC, SSB, VSB, etc.	Digital PM types: QPSK, BPSK, QAM (the combination of amplitude and phase, modulation).

For a radio signal to carry audio or other information for broadcasting or for two-way radio communication, signals must be modulated or changed in some way. Though we have several ways in which a radio signal may be modulated, one of the easiest is to change its amplitude in line with variations of the sound. Here, we discussed three types of modulation, viz: FM, AM, and PM, which will help you understand the basics of the modulation along with the difference between each.

6.7 Frequency spectrum and power in modulations:

Frequency spectrum of FM signals: Unlike amplitude modulation, frequency modulation produces (theoretically) an infinite number of sidebands. It is not possible to evaluate the Fourier transform of a general FM signal, therefore, for the sake of simplicity, the case of a sinusoidal modulating signal is considered. In this case

$$f(t) = A\cos(\omega_m t)$$

and the instantaneous frequency is

$$\omega_i(t) = \omega_c + Ak_f\cos(\omega_m t)$$

which may be expressed

$$\omega_i(t) = \omega_c + \Delta\omega\cos(\omega_m t)$$

where $\Delta\omega$ is called the peak frequency deviation. The phase angle of the FM signal may be expressed

$$\theta(t) = \int_0^t \omega_i(\tau)\,d\tau$$

which may be expressed

$$\theta(t) = \omega_c t + \beta\sin(\omega_m t)$$

where $\beta = \Delta\omega/\omega_c$ is the modulation index of the FM signal.
The resulting FM signal may be expressed in phasor notation as

$$y(t) = Re\left\{Ae^{j\theta(t)}\right\}$$

or

$$y(t) = Re\left\{Ae^{j\omega_c t}e^{j\beta\sin\omega_m t}\right\}$$

The second exponential term in the expression above can be expanded in a Fourier series

$$e^{j\beta\sin\omega_m t} = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_m t}$$

Where

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} e^{j\beta\sin\omega_m t} e^{-jn\omega_m t}\, dt$$

Making a change of variable $\xi = \omega_m t = (2\pi/T)t$ gives

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\beta\sin\xi - n\xi)}\, d\xi = J_n(\beta)$$

Where $J_n\beta$ is the Bessel function of the first kind of order *n*. Using this result gives

$$e^{j\beta\sin\omega_m t} = \sum_{n=-\infty}^{\infty} J_n(\beta)\, e^{jn\omega_m t}$$

Therefore

$$y(t) = Re\left\{Ae^{j\omega_c t}\sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t}\right\} = A\sum_{n=-\infty}^{\infty} J_n(\beta)\cos(\omega_c + n\omega_m)t$$

From this it is evident that an FM waveform with sinusoidal modulation has an infinite number of sidebands. However, the magnitudes of the spectral components of the higher-order sidebands are negligible. The number of sidebands which are significant depend on the order of the Bessel

function *n*, and the value of β.

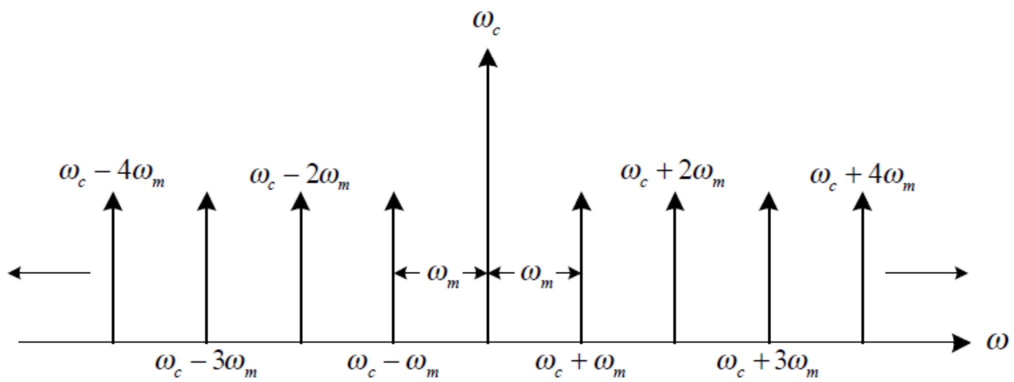


Fig.6.8 Spectrum of an FM waveform.

Power in modulations:

Power Calculations of AM Wave

Consider the following equation of amplitude modulated wave.

$$s(t) = A_c\cos(2\pi f_c t) + \frac{A_c\mu}{2}\cos[2\pi(f_c + f_m)t] + \frac{A_c\mu}{2}\cos[2\pi(f_c - f_m)t]$$

Power of AM wave is equal to the sum of powers of carrier, upper sideband, and lower sideband frequency components.

$$P_t = P_c + P_{USB} + P_{LSB}$$

We know that the standard formula for power of cos signal is

$$P = \frac{{v_{rms}}^2}{R} = \frac{\left(v_m/\sqrt{2}\right)^2}{2}$$

Where,

v_{rms} is the rms value of cos signal.

v_m is the peak value of cos signal.

First, let us find the powers of the carrier, the upper and lower sideband one by one.

Carrier power

$$P_c = \frac{\left(A_c/\sqrt{2}\right)^2}{R} = \frac{{A_c}^2}{2R}$$

Upper sideband power

$$P_{USB} = \frac{\left(A_c\mu/2\sqrt{2}\right)^2}{R} = \frac{{A_c}^2{\mu}^2}{8R}$$

Similarly, we will get the lower sideband power same as that of the upper side band power.

$$P_{LSB} = \frac{{A_c}^2{\mu}^2}{8R}$$

Now, let us add these three powers in order to get the power of AM wave.

$$P_t = \frac{{A_c}^2}{2R} + \frac{{A_c}^2{\mu}^2}{8R} + \frac{{A_c}^2{\mu}^2}{8R}$$

$$\Rightarrow P_t = \left(\frac{{A_c}^2}{2R}\right)\left(1 + \frac{\mu^2}{4} + \frac{\mu^2}{4}\right)$$

$$\Rightarrow P_t = P_c\left(1 + \frac{\mu^2}{2}\right)$$

We can use the above formula to calculate the power of AM wave, when the carrier power and the modulation index are known.

If the modulation index μ=1 then the power of AM wave is equal to 1.5 times the carrier power. So, the power required for transmitting an AM wave is 1.5 times the carrier power for a perfect modulation.

SAQ.2

a) What do you mean by Analysis of AM and FM for radio communication system?
b) Discuss in detail of PM and modulation index for radio communication system.
c) What do you mean by Frequency spectrum?
d) What is equation of power in modulations?
e) An FM wave is given by: $s(t) = 40 \cos (9\pi \times 10^6 t + 15\sin (6\pi \times 10^3 t)$. Calculate the frequency deviation, bandwidth, and power of FM wave.

6.8 Circuit of modulator:

Circuit Design: How to make amplitude modulated wave:

The AM modulation is a kind of modulation technique which is in use since the very early days of wireless data transmission. In a radio transmission system there is a relation between the ranges of frequencies which can be transmitted wirelessly with the length of the transmitting antenna. The relation is inversely proportional to one another, means as the frequency of the signal to be transmitted increases the length of the antenna can be reduced and as the frequency of the signal to be transmitted decreases the length of the transmitting antenna should be increased accordingly.

Using an antenna of few meters the frequencies in the range of Mhz can be easily transmitted to a distance. The basic purpose of the wireless transmitting system in early days was to transmit the audio signals, but to transmit audio signals which fall in the range of few Khz an antenna of more than a kilometer height would have been required. Since it was practically impossible to construct such a long antenna, the high frequency signals are transmitted after they are modulated with the low frequency audio signals.

The amplitude modulation is the simplest modulation technique among the wide verity of modulation techniques in use. The amplitude modulation of a high frequency signal is easy to achieve and the demodulation is also simple compared to other techniques. The high frequency signal which is modulated to carry the low frequency audio signals are called 'carrier frequency' and the audio signals used for modulation is called 'modulating signal' or ‘message signal’ or ‘base band signal’. This article demonstrates how to generate an Amplitude Modulation (AM) using the simplest possible circuit.

Description:

To demonstrate the AM modulation of a carrier signal with a message signal, both the carrier signal and message signal generating circuits are also made and the details of them will be discussed in the subsequent section. The carrier signal and message signal used in this project are pure sine waves. Hence the entire circuit can be divided into three blocks:

1. Carrier frequency generator (High frequency sine wave)
2. Message frequency generator (Low frequency sine wave)
3. AM Modulator.

The block diagram of the AM modulation used in this project is shown in the following diagram;

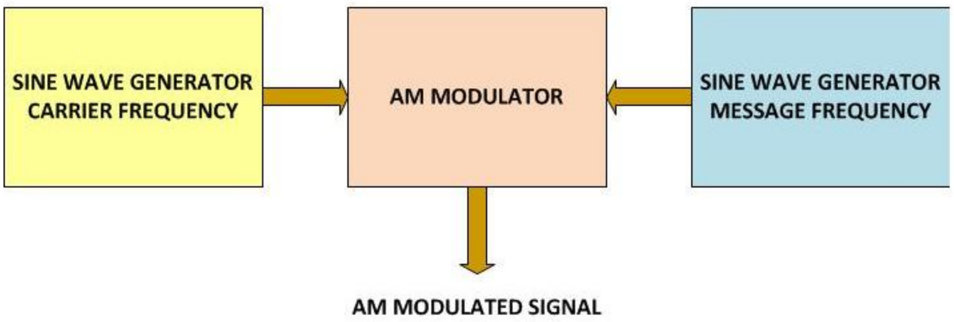


Fig.6.9 Block diagram of AM modulation

For both the carrier signal of high frequency and the message signal of low frequency, exactly same sine wave generating circuits are designed but the frequencies are set to high and low respectively with the help of their variable components. Hence this project has two similar variable frequency sine wave generator circuits and an AM modulator circuit.

Variable frequency sine wave generator: The sine wave generation circuit used in this project is the Wien bridge oscillator circuit. This is the only circuit which can generate the pure sine wave without any distortion. The amplifier component used in the Wien bridge circuit is an op-amp with dual-power supply. Both the circuits are built around the versatile op-amp IC, 741. The circuit of the sine wave generator is shown in the following figure.

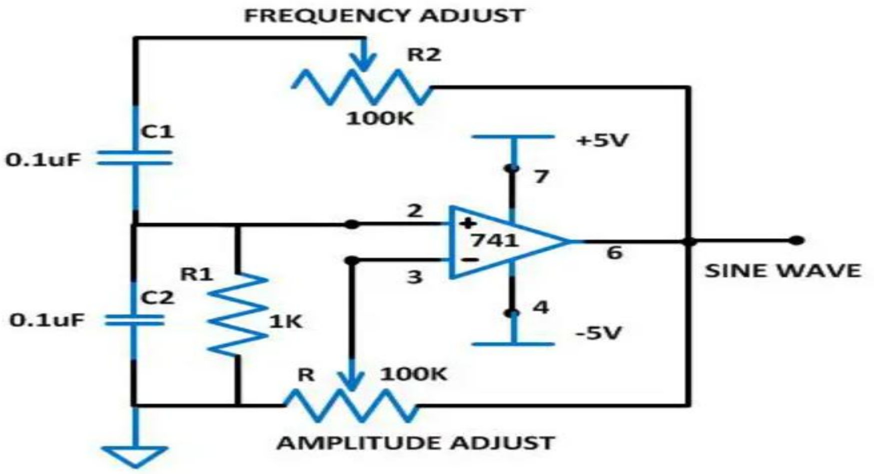


Fig.6.10 Circuit Diagram of Sine Wave Frequency Generator

The value of the resistance R, R1 and R2 are kept same as explained in the article AM modulation. Now the frequency can be calculated using the equation:

$$F = \frac{1}{2\pi\sqrt{R1R2C1C2}}$$

The value of R1 is kept the same as 1K but the value of the resistance R2 can be varied. As the value of the R2 decreases the output frequency of the circuit increases. As it is mentioned in the article AM modulation that the minimum value of the R2 which produces the highest stable frequency is around 130 ohms. In this project to increase the carrier frequency at least 10 times that achieved in the previous project the value of C1 and C2 is reduced to 10 times than that used in the previous AM modulation project. Hence the maximum carrier frequency can be calculated using the frequency equation as shown below:

$$Fmax = \frac{1}{2\pi\sqrt{1000 \ast 130 \ast 0.01 \ast 10^{-6} \ast 0.01 \ast 10\wedge{-6}}} = 44141.6\ Hz$$

The message frequency or modulating frequency generator circuit is kept the same as discussed in the article AM modulation.

AM Modulator: There are different kinds of circuits which can produce AM modulation. The most common among them are transistor based circuits. The transistor based circuits requires proper biasing and a single transistor in most of the cases is not enough to handle both the positive and negative cycles of large amplitude signals. There are inductors or coil based circuits which can also produce AM modulation but they also require proper tuning and they are vulnerable to noise in the surroundings.

The simplest and the stable AM modulator circuit can be designed with the help of an FET. The carrier wave can be allowed to flow through the channel of the FET and the message signal can be used to modify the width of the channel and hence one can achieve the simplest AM modulation.

The component used here as an AM modulator is the N-channel FET, BFW10. The carrier signal is fed through the N-channel from source to drain of the FET which is then modulated by applying the message signal on the gate of the FET. The circuit built around BFW10 which can act as an AM modulator is shown in the following fig.

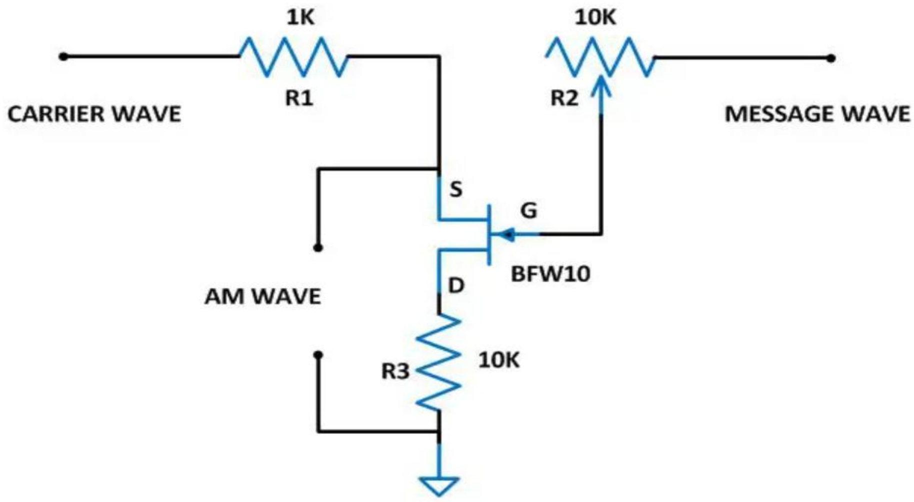


Fig.6.11 Circuit Diagram Of AM Modulator Built From N-channel FET, BFW10

Compared to other circuits there is no complex calculations involved in the design of the components used in this circuits. R1 is used as the current limiting resistor and the resistor R3 is used to generate a reasonable voltage drop when the AM signal current flow through it so that one can get the AM signal voltage across it. R3 is again the current limiting resistor for the base of the FET and it is selected as a potentiometer so that by varying it the depth of the modulation can be demonstrated as varying. The depth of the modulation simply means the amount of message signal amplitude that is required to be present in given amplitude of the carrier signal. The carrier wave is applied through the resistor R1 to the FET and the message wave is applied to the gate of the FET through the potentiometer R2. The potentiometer R2 can be varied to adjust the depth of the modulation.

The carrier signal flows through the N-channel of the FET and as they flows the message signal voltage at the gate of the FET continuously increases and decreases the width of the N-channel. Thus the carrier signal flowing through the channel experiences an increase and decrease in resistance corresponding to the increase or decrease of the amplitude of the message signal. Henceforth the amplitude of the carrier signal varies according to the message signal as it flows through the N-channel. This modulated amplitude carrier signal appears across the source of the FET and ground as AM wave.

The complete circuit for the AM generation

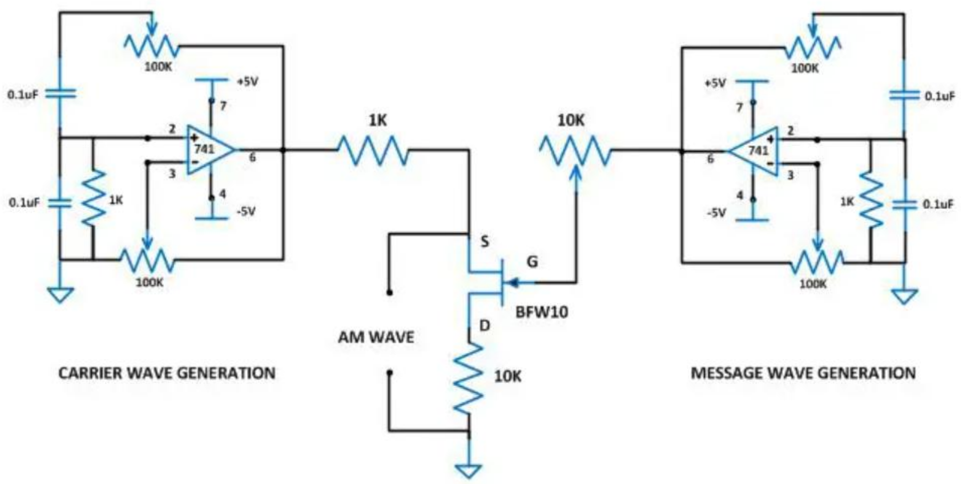


Fig.6.12 Complete circuit for the AM generation

6.9 Demodulation (need and statements):

What is Demodulation?

The process of recovering the original signal from the modulated wave is known as demodulation or detection. At the broadcasting station, modulation is done to transmit the audio signal over larger distances to a receiver. When the modulated wave is picked up by the radio receiver, it is necessary to recover the audio signal from it. This process is accomplished in the radio receiver and is called demodulation.

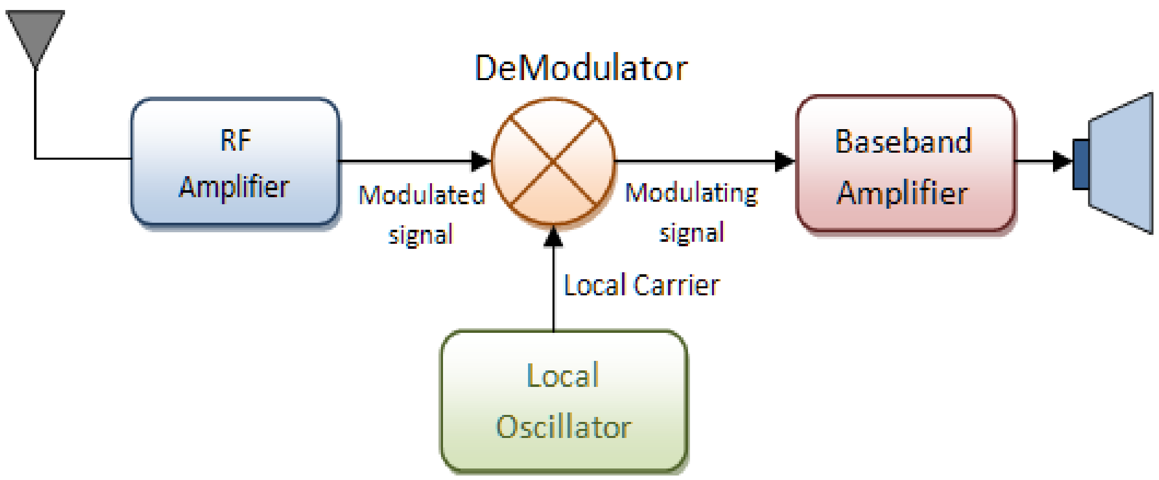


Fig 6.13 Demodulation Block diagram

Need of demodulation:

The wireless signal consists of radio frequency (high frequency) carrier wave modulated by audio frequency (low frequency). The diaphragm of a telephone receiver or a loud speaker cannot vibrate with high frequency. Moreover, this frequency is beyond the audible range of human ear. So, it is necessary to separate the audio frequencies from radio- frequency carrier waves.

Difference between Modulation and Demodulation:

Modulation is the process of imposing data information on the carrier, while demodulation is the recovery of original information at the distant end from the carrier.

Modem is the device that performs both modulation and demodulation.

Both processes try to achieve transfer information with the minimum distortion, minimum loss and efficient utilization of spectrum.

Even though there are different methods for modulation and demodulation processes, each has its own advantages and disadvantages. For example, AM is used in shortwave and radio wave broadcasting; FM is mostly used in high-frequency radio broadcasting, and pulse modulation is known for digital signal modulation.

6.10 Circuit for demodulator:

Circuit Design: How to Demodulate AM Signal: The amplitude modulation is the simplest modulation technique among the wide verity of modulation techniques in use. In this technique the amplitude of a high frequency signal is varied corresponding to the variation in the amplitude of the low frequency modulating signal. The amplitude modulation of a high frequency signal is easy to achieve and the demodulation is also less complex compared to other techniques. The high frequency signal which is modulated to carry the low frequency audio signals are called 'carrier frequency' since they are used to carry the message signal to distant places with the help of wireless transmission devices. The audio signals used for modulation is called 'modulating signal' or 'message signal' or 'base band signal'.

The demodulation of an AM wave can be done with only few components and unlike most of the demodulation technique there is no synchronization required between the modulator and demodulator circuits. The message signal appears as an envelope over the amplitude of the carrier wave and the demodulator make use of this to extract the modulating signal from the carrier and hence the technique of AM modulation is called envelope detection.

This article demonstrates how to generate an Amplitude Modulation (AM) and demodulate the same wave to get the original modulating wave. The AM wave is generated based on the circuits explained in article on AM modulation.

The AM demodulation is done using a low pass filter which can filter out the high frequency carrier from the AM wave in such a way that only the envelope of the carrier wave appears at the output of the filter. The amplitude of the filtered wave has variations corresponding to the amplitude of the modulating low frequency signal.

To get a better filtering using the Low pass filter the carrier frequency must be as large as possible and hence the carrier frequency generator circuit explained in the article AM modulation has to be modified for a very high frequency carrier signal. The only change that is required is the value of the capacitors C1 and C2 which determines the carrier frequency generation. The circuit used for generating the sine wave frequency is given below:

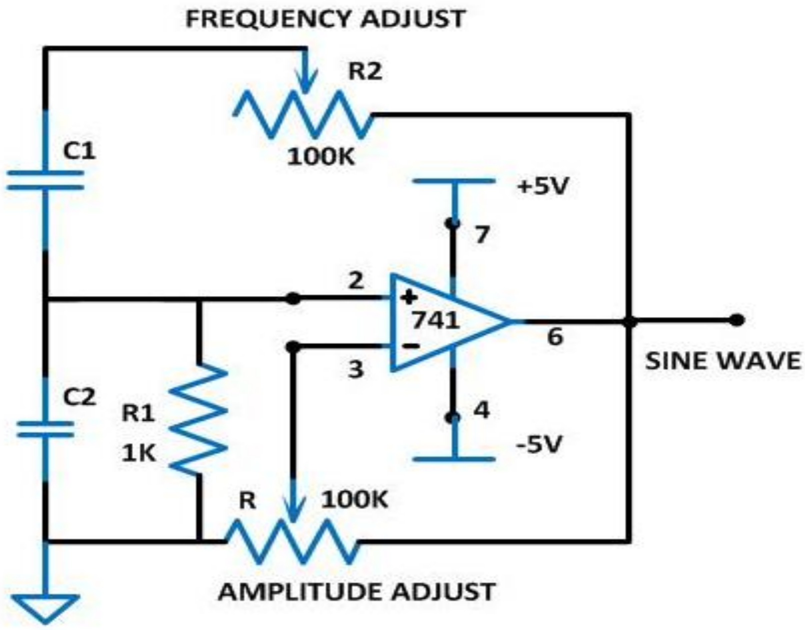


Fig.6.14 Circuit Diagram of Sine Wave Frequency Generator

The value of the resistance R, R1 and R2 are kept same as explained in the article AM modulation. Now the frequency can be calculated using the equation:

$$F = \frac{1}{2\pi\sqrt{R1R2C1C2}}$$

The value of R1 is kept the same as 1K but the value of the resistance R2 can be varied. As the value of the R2 decreases the output frequency of the circuit increases. As it is mentioned in the article AM modulation that the minimum value of the R2 which produces the highest stable frequency is around 130 ohms. In this project to increase the carrier frequency at least 10 times that achieved in the previous project the value of C1 and C2 is reduced to 10 times than that used in the previous AM modulation project. Hence the maximum carrier frequency can be calculated using the frequency equation as shown below:

$$Fmax = \frac{1}{2\pi\sqrt{1000*130*0.01*10^{-6}*0.01*10^{\wedge}-6}} = 44141.6\ Hz$$

The message frequency or modulating frequency generator circuit is kept the same as discussed in the article AM modulation.

The modulation is controlled in this project by some adjustment on the input potentiometer and also with the introduction of a 100K ohm resistor. The carrier signal is now fed through the N-channel from drain to source other than from the source to drain as in the previous project. The output can now be taken from the drain end of the FET where the internal channel modulation by the gate signal is more pronounced. The modified AM modulator circuit and the image of the circuit wired in the breadboard are shown below:

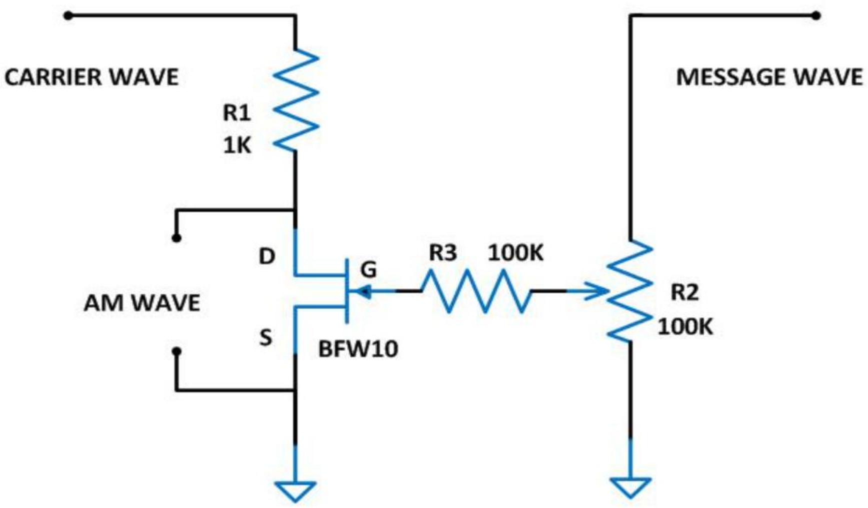


Fig.6.15 Modified AM modulator circuit

Demodulation Steps:

The demodulation of an AM wave is done using three basic steps which are listed below:

1) Rectification
2) Low pass filtering
3) Amplification

The AM wave in most of the cases will be a sine wave with both the half cycles and the Rectification is a process through which either one of the half cycles is eliminated so as to make the AC wave to a DC voltage with high frequency ripples and varying amplitude. The amplitude of the DC voltage still contains the variations which have been there in it before the rectification.

The high frequency ripples are eliminated using a filter circuit making the DC voltage smooth and continuous however maintaining the low frequency variations in the amplitude. The output of the filter circuit resembles the original modulating message wave. The demodulation is almost complete with the filter circuit but to use the demodulated signal for some useful purpose one more step has to be performed.

The output of the filter circuit is very small in amplitude and the noise will be so much that the Signal to Noise Ratio (SNR) is very low. To increase the amplitude of the demodulated signal and to improve the SNR, the amplification of the demodulated signal is necessary. The following section discusses the demodulation steps in detail.

Step:1 Rectification:

In this project a simplest diode rectifier is used to rectify the AM wave. The circuit diagram used in this project and the image of the circuit wired in the bread board is shown in the following figure:

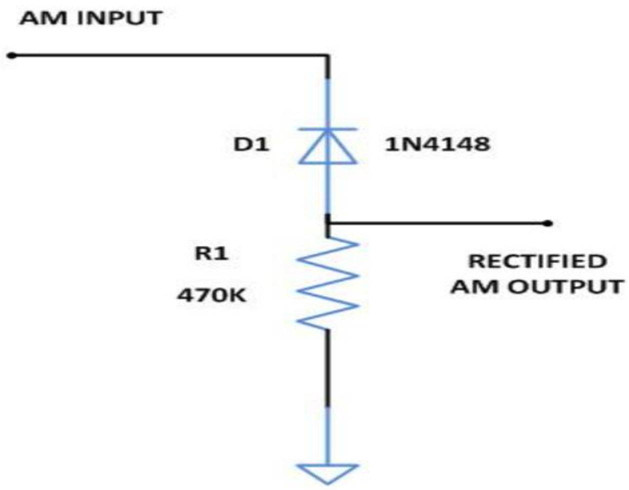


Fig.6.16 Circuit Diagram of Rectifier

The circuit uses a germanium diode which has a cut in voltage of 0.3 volts which is less than that of the silicon diodes. The germanium diode 1N4148 has a very good high frequency response also when compared to the silicon diodes and hence perfect for the rectification of high frequency AM waves.

A high valued resistor of 470K is used to connect the diode to the ground of the circuit to complete the conduction path for the negative half cycles of the AM wave to ground, but without affecting the amplitude due to loading effect.

Simply the negative half cycles only flows through the diode and resistor and they can be taken out across the resistor. Thus the output of the rectifier circuit will be a DC voltage with amplitude varying high frequency ripples.

Step:2 Low pass filtering:

A simple capacitor filter is used to filter out the low frequency amplitude variations from the rectified AM wave in the previous step. The capacitor allows only high frequency ripples to pass through it towards the ground and hence literally shorting them to the ground. The voltage appears across the capacitor will be the low frequency component of the rectified AM wave. Since this circuit filters out the low frequency component which appears like the envelope of the AM wave, this technique is called ‘envelope detection’.

The circuit diagram and the image of the circuit wired in the breadboard are shown in the following figure:

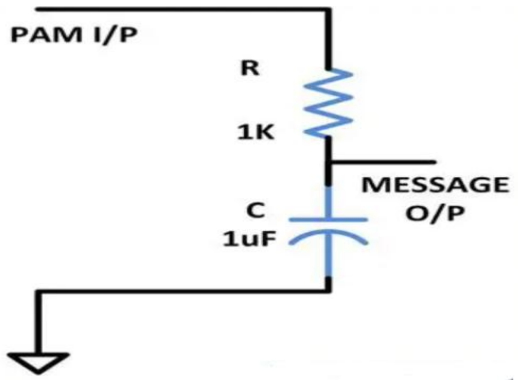


Fig.6.17 Circuit Diagram of Low Pass Filter

The capacitor used to short the high frequency carrier component through the ground is of the same value which has been used to generate the same frequency in the Wien bridge oscillator for the carrier frequency generation circuit. Any other capacitor will develop considerable impedance while the high frequency carrier component flows through it towards ground, and hence makes it impossible to eliminate the high frequency by shorting them to the ground.

The rectifier and the filter circuits form the actual demodulation circuitry for the AM modulated wave. So far the detection of the message wave is done but it needs to be amplified for better SNR and for an increased voltage so that it can be used with other circuits.

Step: 3 Amplification:

In this project the amplification is done using a simple transistor based amplifier circuit. The NPN transistor in the circuit is wired in fixed bias common emitter configuration. The circuit used in this project along with the image of the circuit wired in the breadboard is shown below:

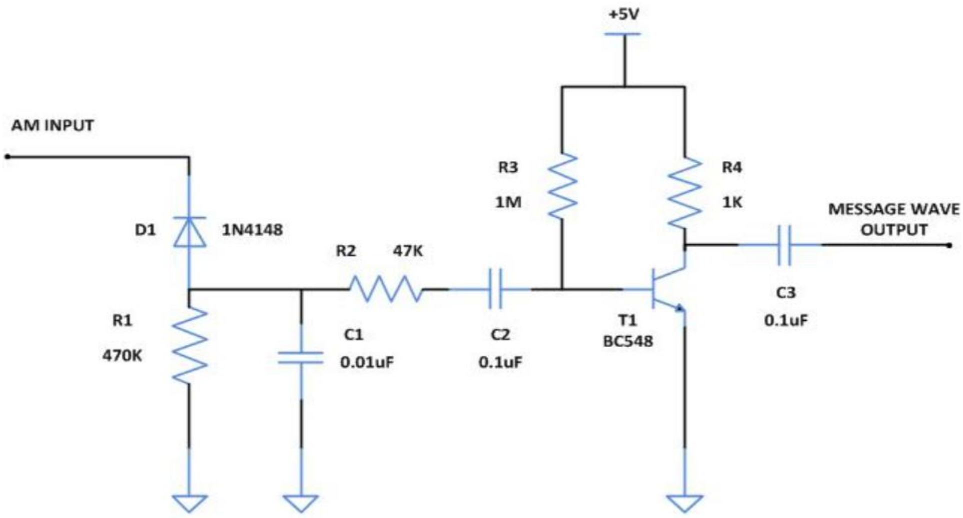


Fig.6.18 Circuit Diagram Of Amplification With NPN Transistor

The Demodulator circuit is connected to an amplifier through a 47K ohm resistance so as to avoid the loading effect of the weak signal by the input of the amplifier circuit. This circuit uses a NPN transistor BC548 which has a very good current gain with a base current limiting resistor and a collector resistor. The base current limiting resistor is selected as high valued as possible since the transistor already has a higher current amplification factor (hfe). The amplified current flows through the collector resistor and develops a voltage drop which can then be coupled out using an output coupling capacitor. The input and the output coupling capacitors are used to allow only the varying component to flow in and out of the amplifier circuit. Note that the coupling capacitors are used here to couple only the message signal to the input and output of the transistor amplifier and hence their values are selected in same as those used in the message signal generating Wien bridge circuit discussed in the article AM modulation . One can use any other amplifier circuit or an external amplifier device to amplify the already detected message wave and hence this particular circuit dose not forms a part of standard AM demodulator circuit.

SAQ.3

a) What do you mean by the Circuit of modulator?
b) Discuss in detail need of Demodulation.
c) What do you mean by the Circuit for demodulator?

Examples:

Q.1 A sinusoidal modulating waveform of amplitude 5V and a frequency of 2 KHz is applied to FM generator, which has a frequency sensitivity of 40 Hz/volt. Calculate the frequency deviation, modulation index, and bandwidth.

Solution

Given, the amplitude of modulating signal, A_m=5V

Frequency of modulating signal, f_m=2KHz

Frequency sensitivity, k_f=40Hz/volt

We know the formula for Frequency deviation as $\Delta f = k_f / A_m$

Substitute k_f and A_m values in the above formula.

$$\Delta f = 40 \times 5 = 200Hz$$

Therefore, frequency deviation, Δf is 200Hz

The formula for modulation index is

$$\beta = \Delta f / f_m$$

Substitute Δf and fm values in the above formula.

$$\beta = 200/(2 \times 1000) = 0.1$$

Here, the value of modulation index, β is 0.1, which is less than one. Hence, it is Narrow Band FM.

The formula for Bandwidth of Narrow Band FM is the same as that of AM wave.

$$BW = 2f_m$$

Substitute fm value in the above formula.

$$BW = 2 \times 2K = 4KHz$$

Therefore, the bandwidth of Narrow Band FM wave is 4KHz.

Q.2. An FM wave is given by: $s(t) = 20 \cos (8\pi \text{ x } 10^6 t + 9 \sin (2\pi \text{ x } 10^3 t)$. Calculate the frequency deviation, bandwidth, and power of FM wave.

Solution:

Given, the equation of an FM wave as

$$s(t) = 20 \cos (8\pi \text{ x } 10^6 t + 9 \sin (2\pi \text{ x } 10^3 t)$$

We know the standard equation of an FM wave as:

$$s(t) = A_c \cos (2\pi f_c t + \beta \sin(2\pi f_m t))$$

We will get the following values by comparing the above two equations.

Amplitude of the carrier signal, A_c=20V

Frequency of the carrier signal, $f_c = 4 \times 10^6$Hz = 4MHz

Frequency of the message signal, $fm = 1 \times 10^3$Hz = 1KHz

Modulation index, β=9

Here, the value of modulation index is greater than one. Hence, it is Wide Band FM.

We know the formula for modulation index as

$$\beta = \Delta f / f_m$$

Rearrange the above equation as follows.

$$\Delta=\beta f_m$$

Substitute β and f_m values in the above equation.

$$\Delta=9\times 1K = 9KHz$$

Therefore, frequency deviation, Δf is 9KHz.

The formula for Bandwidth of Wide Band FM wave is

$$BW=2(\beta+1)f_m$$

Substitute β and fm values in the above formula.

$$BW=2(9+1)1K = 20KHz$$

Therefore, the bandwidth of Wide Band FM wave is 20KHz

Formula for power of FM wave is

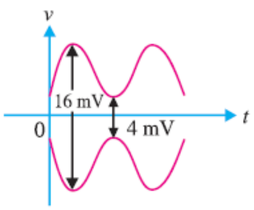
$$Pc=A_c^2/2R$$

Assume, R=1Ω and substitute A_c value in the above equation.

$$P= (20)^2/ 2(1)=200W$$

Therefore, the power of FM wave is 200 watts.

Q.3 The maximum peak-to-peak voltage of an AM wave is 16 mV and the minimum peak-to-peak voltage is 4 mV. Calculate the modulation factor.



Solution:

Maximum voltage of AM wave is

$$V_{max} = \frac{16}{2} = 8 \text{ mV}$$

Minimum voltage of AM wave is

$$V_{min} = \frac{4}{2} = 2 \text{ mV}$$

$$\therefore \quad \text{Modulation factor, } m = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$$

$$= \frac{8-2}{8+2} = \frac{6}{10} = \mathbf{0.6}$$

Q.4 A carrier of 100V and 1200 kHz is modulated by a 50 V, 1000 Hz sine wave signal. Find the modulation factor.

Solution:

$$\text{Modulation factor, } m = \frac{E_S}{E_C} = \frac{50 \text{ V}}{100 \text{ V}} = \mathbf{0.5}$$

Q.5 The load current in the transmitting antenna of an unmodulated AM transmitter is 8A. What will be the antenna current when modulation is 40%?
Solution:

$$P_S = \frac{1}{2}m^2 P_C$$

$$P_T = P_C + P_S = P_C\left(1 + \frac{m^2}{2}\right)$$

$$\therefore \quad \frac{P_T}{P_C} = 1 + \frac{m^2}{2}$$

$$\text{or} \quad \left(\frac{I_T}{I_C}\right)^2 = 1 + \frac{m^2}{2}$$

$$\text{Given that } I_C = 8\text{A};\ m = 0.4$$

$$\therefore \quad \left(\frac{I_T}{8}\right)^2 = 1 + \frac{(0.4)^2}{2}$$

$$\text{or} \quad (I_T/8)^2 = 1.08$$

$$\text{or} \quad I_T = 8\sqrt{1.08} = \mathbf{8.31\ A}$$

Q.6 A 25 MHz carrier is modulated by a 400 Hz audio sine wave. If the carrier voltage is 4V and the maximum frequency deviation is 10 kHz, write down the voltage equation of the FM wave.
Solution: The voltage equations of the FM wave is

$$e = E_c \cos(\omega_c t + m_f \sin \omega_s t)$$

Here

$$\omega_c = 2\pi f_c = 2\pi \times 25 \times 10^6 = 1.57 \times 10^8 \text{ rad/s}$$

$$\omega_s = 2\pi f_s = 2\pi \times 400 = 2513 \text{ rad/s}$$

$$m_f = \frac{\Delta f}{f_s} = \frac{10\text{ kHz}}{400\text{ Hz}} = \frac{10 \times 10^3\text{ Hz}}{400\text{Hz}} = 25$$

$$\therefore \quad e = 4\cos(1.57 \times 10^8\, t + 25 \sin 2513t) \quad \textbf{Ans.}$$

Q.7 Calculate the modulation index for an FM wave where the maximum frequency deviation is 50 kHz and the modulating frequency is 5 kHz.
Solution:

$$\text{Max. frequency deviation, } \Delta f = 50\text{ kHz}$$

$$\text{Modulating frequency, } f_s = 5\text{ kHz}$$

$$\text{Modulation index, } m_f = \frac{\Delta f}{f_s} = \frac{50\text{ kHz}}{5\text{ kHz}} = \mathbf{10}$$

6.11 Summary:

1) In radio communication systems, information is carried across space using radio waves. At the sending end, the information to be sent is converted by some type of transducer to a time-varying electrical signal called the modulation signal.
2) A radio transmitter design has to meet certain requirements. These include the frequency of operation, the type of modulation, the stability and purity of the resulting signal, the efficiency of power use, and the power level required to meet the system design objectives.
3) The mouth (and vocal cords) is the transmitter, ears are the receivers, and air is the transmission medium over which sound travels between mouth and ear. The transmitter and receiver elements of a data modem (such as the type used in a traffic signal system controller box) may not be readily visible.
4) Modulation, in electronics, technique for impressing information (voice, music, picture, or data) on a radio-frequency carrier wave by varying one or more characteristics of the wave in accordance with the intelligence signal.
5) Modulation is the process of converting data into radio waves by adding information to an electronic or optical carrier signal. A carrier signal is one with a steady waveform constant height, or amplitude, and frequency.
6) In AM, a radio wave known as the "carrier" or "carrier wave" is modulated in amplitude by the signal that is to be transmitted. In FM, a radio wave known as the "carrier" or "carrier wave" is modulated in frequency by the signal that is to be transmitted. 4. The frequency and phase remain the same.
7) Frequency Modulation (FM) is the encoding of information in a carrier wave by changing the instantaneous frequency of the wave. FM technology is widely used in the fields of computing, telecommunications, and signal processing.
8) Phase modulation (PM) is a modulation pattern for conditioning communication signals for transmission. It encodes a message signal as variations in the instantaneous phase of a carrier wave. Phase modulation is one of the two principal forms of angle modulation, together with frequency modulation.
9) Modulation index describes the extent to which modulation is done on a carrier signal. In an amplitude modulation, it is defined as the ratio of the amplitude of modulating signal to that of the carrier signal.
10) A frequency spectrum in mobile communications is the range of radio frequencies allocated to each mobile network operator (MNO) in their country of operation for

transmitting and receiving their RF signals. An MNO can add more cells with more spectrum to improve network capacity and coverage.

11) If the modulation index μ=1 then the power of AM wave is equal to 1.5 times the carrier power. So, the power required for transmitting an AM wave is 1.5 times the carrier power for a perfect modulation.
12) In FM, carrier amplitude is constant. Therefore transmitted power is constant. The number of significant sidebands in FM is large.
13) A modulator is a circuit that combines two different signals in such a way that they can be pulled apart later and the information obtained.
14) The process of separating the original information or signal from the modulated carrier. In the case of amplitude or frequency modulation it involves a device, called a demodulator or detector, which produces a signal corresponding to the instantaneous changes in amplitude or frequency, respectively.
15) Demodulation is extracting the original information-bearing signal from a carrier wave. A demodulator is an electronic circuit (or computer program in a software-defined radio) that is used to recover the information content from the modulated carrier wave.
16) Demodulation is extracting the original information-bearing signal from a carrier wave. A demodulator is an electronic circuit (or computer program in a software-defined radio) that is used to recover the information content from the modulated carrier wave.

6.12 Terminal Questions:

1) Explain and discuss the basic elements of radio communication systems.
2) What do you mean by the requirements of transmitter, medium and receiver?
3) Explain the working of Modulation?
4) What do you mean by the need of modulation? Explain in the detail types of modulation.
5) Explain the analysis of AM, FM and PM
6) What are differences between the AM, FM and PM?
7) Explain condition of the modulation index.
8) Explain the frequency spectrum and power in modulations in radio communication systems.
9) Discuss and explain the working of the principle of the circuit of modulator.

10) What do you mean by the need of the Demodulation? Explain the working of the demodulation.
11) Explain the working of the principle of the circuit for demodulator.
12) A modulating signal $m(t) = 10 \cos(2\pi \times 10^3 t)$ is amplitude modulated with a carrier signal
$c(t) = 50 \cos (2\pi \times 10^5 t)$. Find the modulation index, the carrier power, and the power required for transmitting AM wave.
13) A sinusoidal modulating waveform of amplitude 10 V and a frequency of 5 KHz is applied to FM generator, which has a frequency sensitivity of 50 Hz/volt. Calculate the frequency deviation, modulation index, and bandwidth.
14) An FM wave is given by: $s(t) = 30 \cos (5\pi \times 10^6 t + 8 \sin (5\pi \times 10^3 t)$. Calculate the frequency deviation, bandwidth, and power of FM wave.
15) The equation of amplitude wave is given by: $s(t) = 20 [1+ 0.8 \cos(2\pi \times 10^3 t)] \cos(4\pi \times 10^5 t)$. Find the carrier power, the total sideband power, and the band width of AM wave.
16) A carrier of 200V and 1500 kHz is modulated by a 60 V, 2000 Hz sine wave signal. Find the modulation factor.
17) The load current in the transmitting antenna of an unmodulated AM transmitter is 10A. What will be the antenna current when modulation is 50%?
18) Calculate the modulation index for an FM wave where the maximum frequency deviation is 100 kHz and the modulating frequency is 15 kHz.

Unit 7 Operational amplifier

7.1 Introduction:

An operational amplifier (op-amp) is an integrated circuit (IC) that amplifies the difference in voltage between two inputs. It is so named because it can be configured to perform arithmetic operations.

op-amps were so named because they were used to model the basic mathematical operations of addition, subtraction, integration, differentiation, etc. in electronic analog computers. In this sense a true operational amplifier is an ideal circuit element.

An op amp needs a power supply because internally it is composed of a number of transistors. , you see the enormous amount of transistors which makes up an op amp.

An ideal amplifier has infinite input impedance, zero output impedance, and a fixed gain at all frequencies. An ideal op amp has infinite input impedance and zero output impedance, but has infinite gain.

When the inputs are equal, there is no output. When the inverting input is greater, the op amp becomes saturated and output voltage is equal to the positive voltage supply. When the inverting input is greater, the output voltage is equal to the negative voltage supply.

The input offset voltage is defined as the voltage that must be applied between the two input terminals of the op amp to obtain zero volts at the output. Ideally the output of the op amp should be at zero volts when the inputs are grounded. In reality the input terminals are at slightly different dc potentials.

Output offset voltage (V_{oo})- It is the output voltage of op-amp when both inputs are zero. V_{oo} is due to dissimilarities in transistor and due to mismatch in Resistor values in the internal circuit of the op-amp.

A differential amplifier (also known as a difference amplifier or op-amp subtractor) is a type of electronic amplifier that amplifies the difference between two input voltages but suppresses any voltage common to the two inputs.

The Common-Mode Rejection Ratio (CMRR) indicates the ability of a differential amplifier to suppress signals common to the two inputs. Desired signals should appear on only one input or with opposite polarities on both inputs. These desired signals are amplified and appear on the outputs.

The output current from the op-amp is that current needed to keep the inverting input at ground potential.

The simplest method for calculating power consumption of an op amp using power dissipation in the op amp is to solve a power-balance equation, in which supply power equals the sum of power dissipated in the load and in the op amp. Thus, the op amp dissipation equals supply power minus load power.

The slew rate of an op amp or any amplifier circuit is the rate of change in the output voltage caused by a step change on the input. It is measured as a voltage change in a given time - typically V / µs or V / ms. A typical general purpose device may have a slew rate of 10 V / microsecond.

The Gain-Bandwidth of the circuit (usually amplifier) is the product of the bandwidth and the gain at which the bandwidth is measured. For an operational amplifier, the gain-bandwidth product for one configuration will always equal the gain-bandwidth product for any other configuration of the same amplifier.

Op-amp characteristics are infinite open-loop gain G = V_{out} / V, Infinite input impedance R_{in}, and so zero input current, Zero input offset voltage, Infinite output voltage range, Infinite bandwidth with zero phase shift and infinite slew rate, Zero output impedance R_{out}, and so infinite output current range, Zero noise.

The important characteristics of comparator are: Speed of operation. The output of comparator must switch rapidly between the saturation level (+Vsat or -Vsat) and also respond instantly to any change of condition at its input, Accuracy, Compatibility of output.

A detector is a device that recovers data of interest from a modulated wave. General Characteristics of Detectors are: Sensitivity, Detector Response, Energy Resolution, The Response Function, Response Time, Detector Efficiency, Dead Time

An inverting op-amp is a type of operational amplifier circuit used to generate an output that is out of phase as compared to its input through 180 degrees which means, if the input signal is positive (+), then the output signal will be opposite. The inverting op-amp is designed through an op-amp with two resistors.

A non-inverting op amp is an operational amplifier circuit with an output voltage that is in phase with the input voltage. Its complement is the inverting op amp, which produces an output signal that is 180^o out of phase.

A differentiator is a circuit that is designed such that the output of the circuit is approximately directly proportional to the rate of change (the time derivative) of the input.

The operational amplifier integrator is an electronic integration circuit. Based on the operational amplifier (op-amp), it performs the mathematical operation of integration with respect to time; that is, its output voltage is proportional to the input voltage integrated over time.

7.2 Objective:

After studying this unit you should be able to

- Explain and identify OP-amplifier symbol, number code, power supply and characteristics.
- Study and identify Input-output relationship, input-offset and output offset voltage.
- Explain Differential input and output resistance.
- Explain and identify Common mode rejection ratio, output current, power consumption, slew rate gain-band width product.
- Study and identify Characteristics of OP- amplifier, comperator and detector.
- Explain and identify Inverting and non inverting amplifier.
- Explain Differentiator and basic integrator.

7.3 OP-amplifier (symbol, number code, power supply and characteristics):

Operational Amplifier (IC 741 Op-Amp):

The most commonly used op-amp is IC741. The 741 op-amp is a voltage amplifier, it inverts the input voltage at the output, can be found almost everywhere in electronic circuits. Op amps usually have three terminals: two high-impedance inputs and a low-impedance output port. The inverting input is denoted with a minus (-) sign, and the non-inverting input uses a positive (+) sign. Operational amplifiers work to amplify the voltage differential between the inputs, which is useful for a variety of analog functions including signal chain, power, and control applications.

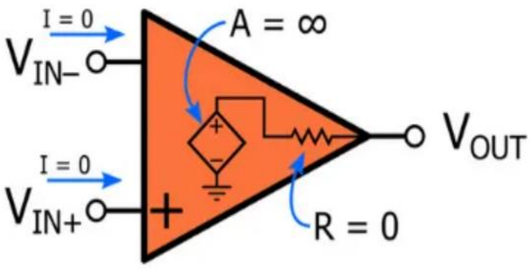


Fig. 7.1 Symbol of Operational Amplifier

Pin Configuration of IC741 Op-Amp:

Let's see the pin configuration and testing of 741 op-amps. Usually, this is a numbered counter clockwise around the chip. It is an 8 pin IC. They provide superior performance in integrator, summing amplifier and general feedback applications. These are high gain op-amp; the voltage on the inverting input can be maintained almost equal to V_{IN}.

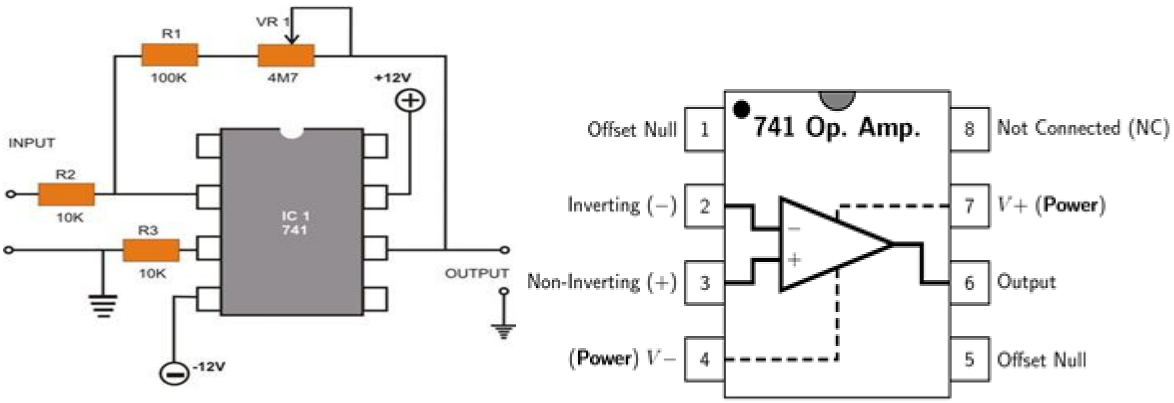


Fig.7.2 Pin Configuration of IC741 Op-Amp

It is a 8-pin dual-in-line package with a pin-out shown above.

Pin 1: Offset null.

Pin 2: Inverting input terminal.

Pin 3: Non-inverting input terminal.

Pin 4: –VCC (negative voltage supply).

Pin 5: Offset null.

Pin 6: Output voltage.

Pin 7: +VCC (positive voltage supply).

Pin 8: No Connection.

The main pins in the 741 op-amp are pin2, pin3 and pin6. In inverting amplifier, a positive voltage is applied to pin2 of the op-amp; we get output as negative voltage through pin 6. The polarity has been inverted. In a non-inverting amplifier, a positive voltage is applied to pin3 of the op-amp; we get output as positive voltage through pin 6. Polarity remains the same in non-inverting amplifier. Vcc is usually in the range from 12 to 15 volts. When two supplies (+Vcc/-Vcc) are used, they are the same voltage and of opposite sign in almost all cases. Remember that the operational amplifier is a high gain, differential voltage amplifier. For a 741 operational amplifier, the gain is at least 100,000 and can be more than a million (1,000,000). That's an important fact you'll need to remember as you put the 741 into a circuit.

There are many common application circuits using IC741 op-amp, they are adder, comparator, subtractor, integrator, differentiator and voltage follower.

Below is some example of 741 IC based circuits. However, the 741 is used as a comparator and not an amplifier. The difference between the two is small but significant. Even if used as a comparator the 741 still detects weak signals so that they can be recognized more easily. A comparator is a circuit that compares two input voltages. One voltage is called the reference voltage and the other is called the input voltage. It is a circuit which compares a signal voltage applied at one input of an op-amp with a known reference voltage at the other input. The 741 op-amp has ideal transfer characteristics (output $\pm V_{sat}$); and the output is changed by increment in the input voltage of 2mV.

The 741 Op Amp IC is a monolithic integrated circuit, comprising of a general purpose Operational Amplifier. It was first manufactured by Fairchild semiconductors in the year 1963. The number 741 indicates that this operational amplifier IC has 7 functional pins, 4 pins capable of taking input and 1 output pin.

IC 741 Op Amp can provide high voltage gain and can be operated over a wide range of voltages, which makes it the best choice for use in integrators, summing amplifiers and general feedback applications. It also features short circuit protection and internal frequency compensation circuits built in it. This Op-amp IC comes in the following form factors:

- 8 Pin DIP Package
- TO5-8 Metal can package
- 8 Pin SOIC

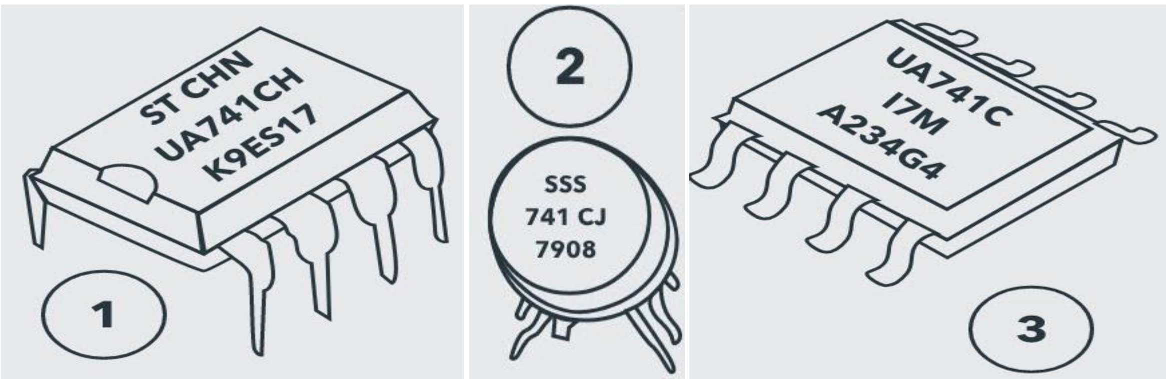


Fig.7.3 Different IC 741 package

The below figure illustrates the pin configurations and internal block diagram of IC 741 in 8 pin DIP and TO5-8 metal can package.

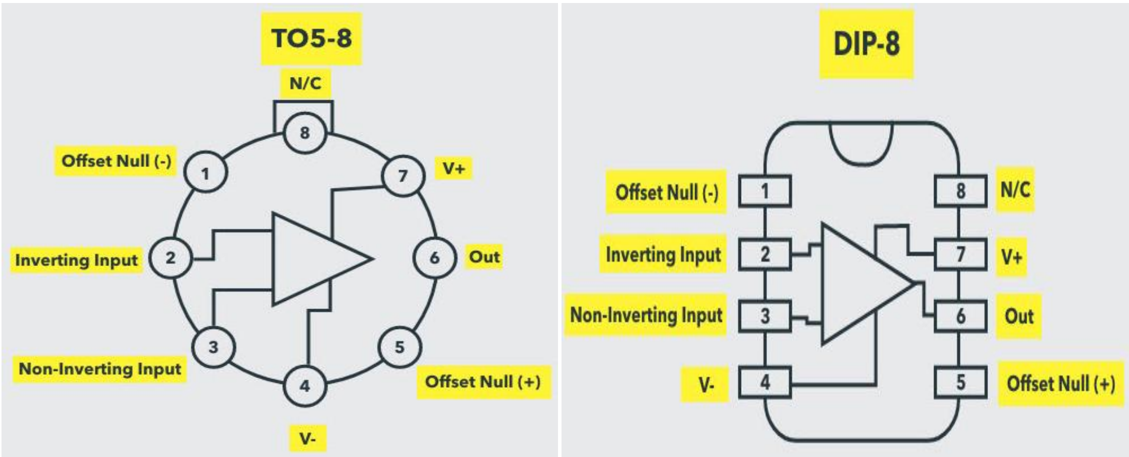


Fig.7.4 Pin-out of TO5-8 and DIP-8 Op Amp

Now let's take a look at the functions of different pins of 741 IC:

- **Pin4 & Pin7 (Power Supply):** Pin7 is the positive voltage supply terminal and Pin4 is the negative voltage supply terminal. The 741 IC draws in power for its operation from these pins. The voltage between these two pins can be anywhere between 5V and 18V.
- **Pin6 (Output):** This is the output pin of IC 741. The voltage at this pin depends on the signals at the input pins and the feedback mechanism used. If the output is said to be high, it means that voltage at the output is equal to positive supply voltage.

Similarly, if the output is said to be low, it means that voltage at the output is equal to negative supply voltage.

- **Pin2 & Pin3 (Input):** These are input pins for the IC. Pin2 is the inverting input and Pin3 is the non-inverting input. If the voltage at Pin2 is greater than the voltage at Pin3, i.e., the voltage at inverting input is higher, the output signal stays low. Similarly, if the voltage at Pin3 is greater than the voltage at Pin2, i.e., the voltage at non-inverting input is high, the output goes high.
- **Pin1 & Pin5 (Offset Null):** Because of high gain provided by 741 Op-Amp, even slight differences in voltages at the inverting and non-inverting inputs, caused due to irregularities in manufacturing process or external disturbances, can influence the output. To nullify this effect, an offset voltage can be applied at pin1 and pin5, and is usually done using a potentiometer.
- **Pin8 (N/C):** This pin is not connected to any circuit inside 741 IC. It's just a dummy lead used to fill the void space in standard 8 pin packages.

Specifications: The following are the basic specifications of IC 741:

- **Power Supply:** Requires a Minimum voltage of 5V and can withstand up to 18V
- **Input Impedance:** About 2 MΩ
- **Output impedance:** About 75 Ω
- **Voltage Gain:** 200,000 for low frequencies (200 V / mV)
- **Maximum Output Current:** 20 mA
- **Recommended Output Load:** Greater than 2 KΩ
- **Input Offset:** Ranges between 2 mV and 6 mV
- **Slew Rate:** 0.5V/µS (It is the rate at which an Op-Amp can detect voltage changes).

The high input impedance and very small output impedance makes IC 741 a near ideal voltage amplifier.

Input and Output Terminals of an Operational Amplifier: An op-amp has two input terminals and one output terminal. The op-amp also has two voltage supply terminals as seen above. Two input terminals form the differential input. We call the terminal, marked with negative (-) sign as the inverting terminal and the terminal marked with positive (+) sign as the non-inverting terminal of the operational amplifier. If we apply an input signal at the inverting terminal (-) than the amplified output signal is 180o out of phase concerning the applied input signal. If we apply an input signal to the non-inverting terminal (+) then the output signal obtained will be in phase, i.e. it will have no phase shift concerning the input signal.

Power Supply for an Operational Amplifier: As seen from the circuit symbol above it has two input power supply terminals $+V_{CC}$ and $–V_{CC}$. For the operation of an op-amp a dual polarity DC supply is essential. In the dual polarity supply, we connect the $+V_{CC}$ to the positive DC supply and the $–V_{CC}$ terminal to the negative DC supply. However few op-amps can also operate on a single polarity supply. Note that there is no common ground terminal in the op-amps hence the ground has to be established externally.

Working Principle of Op-Amp:

Open Loop Operation of an Operational Amplifier: said above an op-amp has a differential input and single ended output. So, if we apply two signals one at the inverting and another at the non-inverting terminal, an ideal op-amp will amplify the difference between the two applied input signals. We call this difference between two input signals as the differential input voltage. The equation below gives the output of an operational amplifier.

$$V_{OUT} = A_{OL} (V_1 – V_2)$$

Where, V_{OUT} is the voltage at the output terminal of the op-amp. A_{OL} is the open-loop gain for the given op-amp and is constant (ideally). For the IC 741 A_{OL} is 2×10^5.

V_1 is the voltage at the non-inverting terminal.

V_2 is the voltage at the inverting terminal.

$(V_1 – V_2)$ is the differential input voltage.

It is clear from the above equation that the output will be non-zero if and only if the differential input voltage is non-zero (V_1 and V_2 are not equal), and will be zero if both V_1 and V_2 are equal. Note that this is an ideal condition; practically there are small imbalances in the op-amp. The open-loop gain of an op-amp is very high. Hence, an open loop operational amplifier amplifies a small applied differential input voltage to a huge value.

Also, it is true that if we apply small differential input voltage, the operational amplifier amplifies it to a considerable value but this significant value at the output cannot go beyond the supply voltage of the op-amp. Hence it does not violate the law of conservation of energy.

Closed Loop Operation: The above-explained operation of the op-amp was for open-loop i.e. without a feedback. We introduce feedback in the closed loop configuration. This feedback path feeds the output signal to the input. Hence, at the inputs, two signals are simultaneously present. One of them is the original applied signal, and the other is the feedback signal. The equation below shows the output of a closed loop op-amp.

$$V_{OUT} = A_{CL} (V_1 – V_2) = A_{CL} V_D$$

Where V_{OUT} is the voltage at the output terminal of the op-amp. A_{CL} is the closed loop gain. The feedback circuit connected to the op-amp determines the closed loop gain A_{CL}. $V_D = (V_1 – V_2)$ is the differential input voltage. We say the feedback as positive if the feedback path feeds the signal from the output terminal back to the non-inverting (+) terminal. Positive feedback is used in oscillators. The feedback is negative if the feedback path feeds the part of the signal from the output terminal back to the inverting (-) terminal. We use negative feedback to the op-amps used as amplifiers. Each type of feedback, negative or positive has its advantages and disadvantages.

Positive Feedback ⇒ Oscillator

Negative Feedback ⇒ Amplifier

The above explanation is the most basic working principle of operational amplifiers.

Operational Amplifiers: Key Characteristics and Parameters:

There are many different important characteristics and parameters related to op amps (Fig.7.5)**.** These characteristics are described in greater detail below.

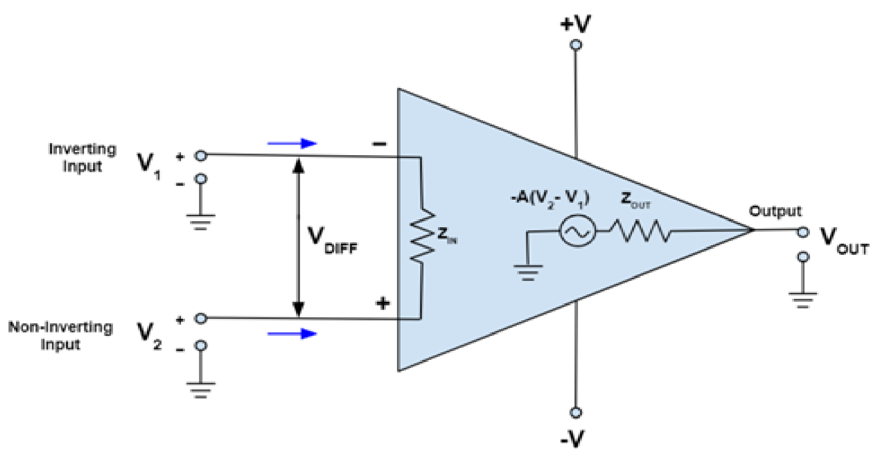


Fig.7.5 Operational Amplifier Schematic

Open-loop gain: Open-loop gain: The open-loop gain ("A" in Fig.7.5) of an operational amplifier is the measure of the gain achieved when there is no feedback implemented in the circuit. This means the feedback path, or loop, is open. An open-loop gain often must be exceedingly large (10,000+) to be useful in itself, except with voltage comparators.
Voltage comparators compare the input terminal voltages. Even with small voltage differentials, voltage comparators can drive the output to either the positive or negative rails. High open-loop gains are beneficial in closed-loop configurations, as they enable stable circuit behaviors across temperature, process, and signal variations.

Input impedance: Another important characteristic of op amps is that they generally have high input impedance ("Z_{IN}" in Fig.7.5). Input impedance is measured between the negative and positive input terminals, and its ideal value is infinity, which minimizes loading of the source. (In reality, there is a small current leakage.) Arranging the circuitry around an operational amplifier may significantly alter the effective input impedance for the source, so external components and feedback loops must be carefully configured. It is important to note that input impedance is not solely determined by the input DC resistance. Input capacitance can also influence circuit behavior, so that must be taken into consideration as well.

Output impedance: An operational amplifier ideally has zero output impedance ("Z_{OUT}" in Fig.7.5). However, the output impedance typically has a small value, which determines the amount of current it can drive, and how well it can operate as a voltage buffer.

Frequency response and bandwidth (BW): An ideal op amp would have an infinite bandwidth (BW), and would be able to maintain a high gain regardless of signal frequency. However, all operational amplifiers have a finite bandwidth, generally called the "-3dB point," where the gain begins to roll as frequency increases. The gain of the amplifier then decreases at a rate of -20dB/decade while the frequency increases. Op amps with a higher BW have improved performance because they maintain higher gains at higher frequencies; however, this higher gain results in larger power consumption or increased cost.

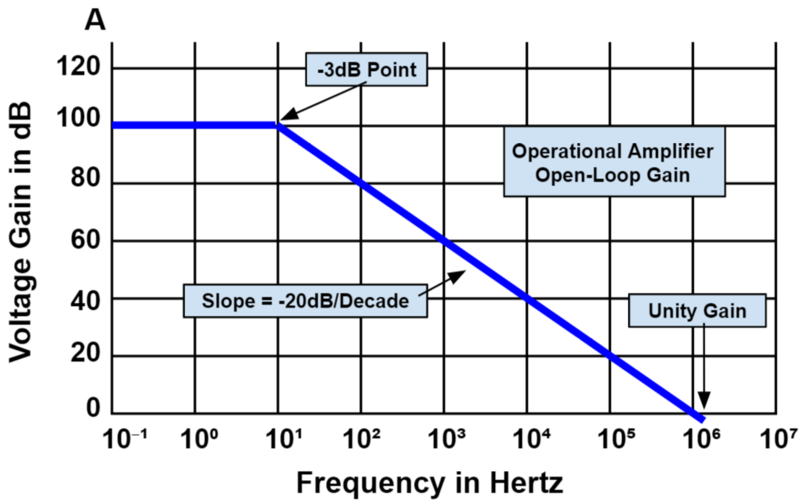


Fig.7.6 Operational Amplifier Open-Loop Frequency Response Curve

Gain bandwidth product (GBP): As the name suggests, GBP is a product of the amplifier's gain and bandwidth. GBP is a constant value across the curve, and can be calculated with Equation (1):

$$GBP = Gain \times Bandwidth = A \times BW \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

GBP is measured at the frequency point at which the operational amplifier's gain reaches unity. This is useful because it allows the user to calculate the device's open-loop gain at different frequencies. An operational amplifier's GBP is generally a measure of its usefulness and performance, as op amps with a higher GBP can be used to achieve better performance at higher frequencies.

These are the major parameters to consider when selecting an operational amplifier in your design, but there are many other considerations that may influence your design, depending on the application and performance needs. Other common parameters include input offset voltage, noise, quiescent current, and supply voltages.

Negative Feedback and Closed-Loop Gain: In an operational amplifier, negative feedback is implemented by feeding a portion of the output signal through an external feedback resistor and back to the inverting input (Fig.7.7).

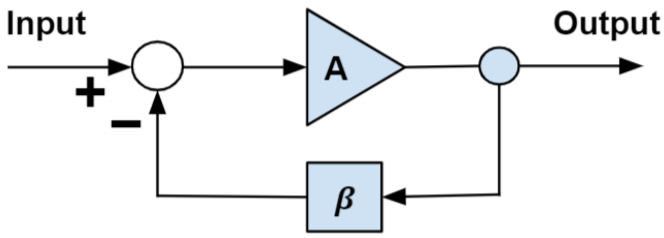


Fig.7.7 Negative Feedback with Inverting Operational Amplifier

Negative feedback is used to stabilize the gain. By using a negative feedback, the closed-loop gain can be determined via external feedback components that can have higher accuracy compared to the operational amplifier's internal components. This is because the internal op amp components may vary substantially due to process shifts, temperature changes, voltage changes, and other factors. The closed-loop gain can be calculated with Equation (2)**:**

$$\frac{V_{out}}{V_{in}} = \frac{1}{f} \dots\dots\dots\dots\dots\dots (2)$$

Advantages of op-amp:

- Increased circuit stability,
- Increased input impedance,
- Decreased output impedance,
- Increased frequency bandwidth at constant gain.

Limitations of Operational Amplifier:

- Voltage Supply Limitations.
- Finite Bandwidth Limitations.
- Input Offset Voltage Limitations.
- Input Bias Current Limitations.
- Output Offset Voltage Limits.
- Slew Rate Limitation.
- Short Circuit Output Limits.
- Limited Common Mode Rejection Ratio.

Applications of op-amp:

- Amplifiers,
- Active filters,
- Arithmetic circuits,
- Log/antilog amp,
- Voltage comparators,
- Waveform Generators,
- Precision rectifiers,
- Multipliers,
- Timers,
- Multi-vibrators,
- Regulated power supplies.

7.4 Input-output relationship, input-offset and output offset voltage:

Input-output relationship using Non-inverting Voltage Follower:

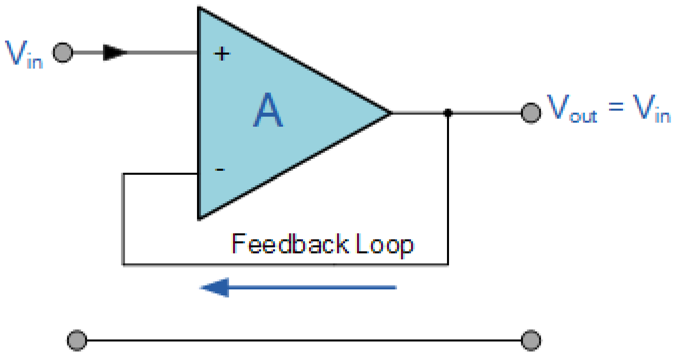


Fig.7.8 Non-inverting circuit

As the input signal is connected directly to the non-inverting input of the amplifier the output signal is not inverted resulting in the output voltage being equal to the input voltage, thus

$$V_{out} = V_{in}$$

In this non-inverting circuit configuration, the input impedance R_{in} has increased to infinity and the feedback impedance R_f reduced to zero. The output is connected directly back to the negative inverting input so the feedback is 100% and Vin is exactly equal to V_{out} giving it a fixed gain of 1 or unity. As the input voltage V_{in} is applied to the non-inverting input, the voltage gain of the amplifier is therefore given as:

$$V_{out} = A(V_{in})$$

$$(V_{in} = V+) \text{ and } (V_{out} = V-)$$

Therefore

$$\text{Gain, } (A_v) = V_{out}/ V_{in} = +1$$

Since no current flows into the non-inverting input terminal the input impedance is infinite (ideal conditions) so zero current will flow through the feedback loop. Thus any value of resistance may be placed in the feedback loop without affecting the characteristics of the circuit as no current flows through it so there is zero voltage drop across it resulting in zero power loss.

As the input impedance is extremely high, the unity gain buffer (voltage follower) can be used to provide a large power gain as the extra power comes from the op-amps supply rails and through the op-amps output to the load and not directly from the input. However in most real unity gain buffer circuits there are leakage currents and parasitic capacitances present so a low value (typically 1kΩ) resistor is required in the feedback loop to help reduce the effects of these leakage currents providing stability especially if the operational amplifier is of a current feedback type.

The voltage follower or unity gain buffer is a special and very useful type of Non-inverting amplifier circuit that is commonly used in electronics to isolated circuits from each other especially in High-order state variable or Sallen-Key type active filters to separate one filter stage from the other. Typical digital buffer IC's available are the 74LS125 Quad 3-state buffer or the more common 74LS244 Octal buffer.

One final thought, the closed loop voltage gain of a voltage follower circuit is "1" or Unity. The open loop voltage gain of an operational amplifier with no feedback is Infinite. Then by carefully selecting the feedback components we can control the amount of gain produced by a non-inverting operational amplifier anywhere from one to infinity.

Thus far we have analyzed an inverting and non-inverting amplifier circuit that has just one input signal, V_{in}. In the next tutorial about Operational Amplifiers, we will examine the effect of the output voltage, V_{out} by connecting more inputs to the amplifier. This then produces another common type of operational amplifier circuit called a Summing Amplifier which can be used to "add" together the voltages present on its inputs.

Input Offset Voltage for Op Amp: In the case of the ideal op-amp, the DC voltage of the $V_{IN(+)}$ and $V_{IN(-)}$ terminals match exactly when the input voltage (V_i) is 0 V. In reality, however, there are differences in input impedance and input bias current between the $V_{IN(+)}$ and $V_{IN(-)}$ terminals, causing a slight difference in their voltages. This difference called input offset voltage is multiplied by a gain, appearing as an output voltage deviation from the ideal value.

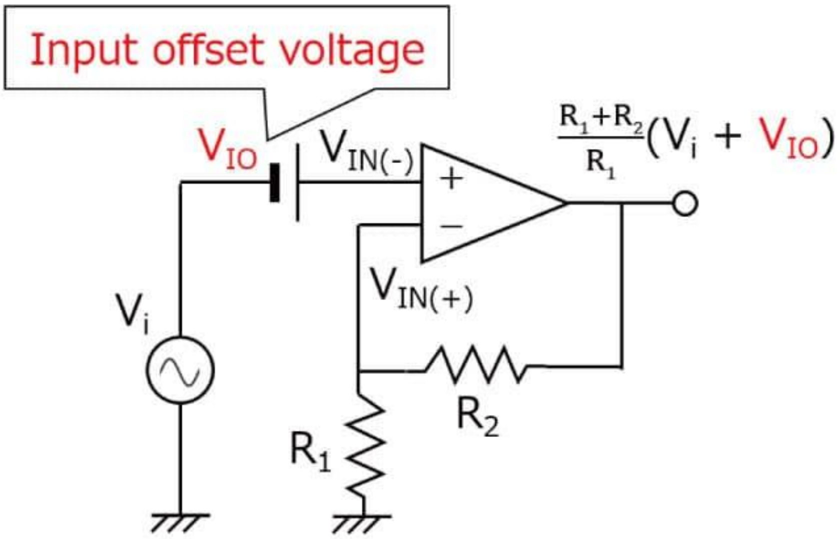


Fig.7.9 Inverting amplifier with an input offset

When used in amplifiers of sensors, etc., the input offset voltage of an op-amp result in an error of sensor detection sensitivity. To keep sensing errors below a specified tolerance level, it is necessary to select an op-amp with low input offset voltage.

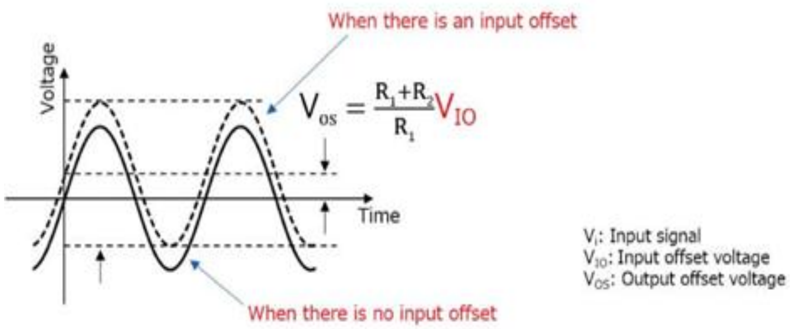


Fig.7.10 Output waveform of an inverting amplifier with an input offset

Output offset voltage for Op Amp: The output offset voltage cannot be specified, but is a result of the input bias and offset currents, the input offset voltage, and the resistors used in the feedback and bias network connected externally around the amplifier. To calculate it, we start with Fig. This represents all possible bias and d.c. feedback paths around an op amp.

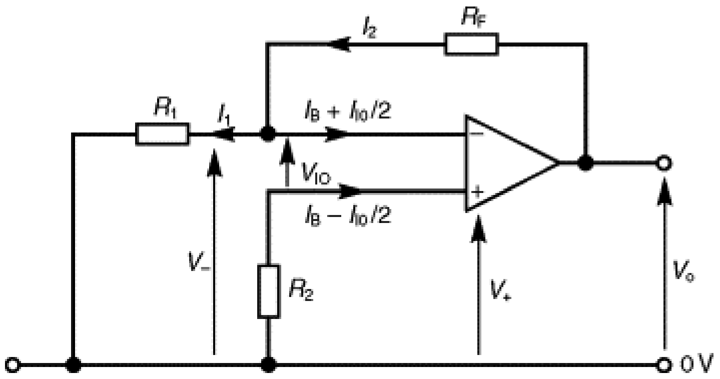


Fig.7.11 Biasing an Op-amp

If a non-inverting configuration is used, then R_2 represents the d.c. source resistance. If the source is a.c. coupled, *via* a coupling capacitor to the non-inverting input, then this resistor must be included to provide a d.c. path to 0 V for the input bias current.

If an inverting configuration is used, then R_1 represents the d.c. source resistance. In this case, the bias current will flow through both R_1 and the feedback resistor R_F.If the source is a.c. coupled, *via* a coupling capacitor to the inverting input, then the bias current will flow through the feedback resistor R_F alone, and the value of R_F is set to infinity in the calculation.

Now assume that the output offset voltage is sufficiently close to zero that one can use the values of input bias current, input offset current and input offset voltage which are defined for the condition $V_o = 0$. For instance, suppose $V_o = 1$V. Then, with a typical op amp d.c. open-loop voltage gain of 10^5, an extra input of 10^{-5} V(= 10 μV) will be needed to bring the output to 0 V, implying that the value of V_{IO} chosen is in error by 10 μV. This is far less than the spreads on typical specified values. This assumption does not mean that V_o is taken as zero; but only the actual values of input bias current, input offset current and input offset voltage in the circuit as being the same as those defined for the condition $V_o = 0$.

The two input bias currents, shown on the circuit of Fig.7.11, must be $I_B + I_{IO}/2$ and $I_B - I_{IO}/2$, so that their average is I_B and their difference is I_{IO}, in accordance with their definitions. (The choice of which is arbitrary, since the value of I_{IO} can be either positive or negative.)

The voltage at the non-inverting input is

$$V_+ = -(I_B - \frac{I_{IO}}{2})R_2$$

The voltage at the inverting input is

$$V_- = V_+ + V_{IO} = -(I_B - I_{IO})R_2 + V_{IO}$$

So,

$$I_1 = \frac{V_-}{R_1} = \frac{-\left(I_B - \frac{I_{IO}}{2}\right)R_2 + V_{IO}}{R_1}$$

Equating the currents at the non-inverting input,

$$I_2 = I_1 + I_B + \frac{I_{IO}}{2} = \frac{-\left(I_B - \frac{I_{IO}}{2}\right)R_2 + V_{IO}}{R_1} + I_B + \frac{I_{IO}}{2}$$

The output offset voltage is

$$V_O = I_2R_F + V_- = I_2R_F - \left(I_B - \frac{I_{IO}}{2}\right)R_2 + V_{IO}$$

$$= \left\{\frac{-\left(I_B - \frac{I_{IO}}{2}\right)R_2 + V_{IO}}{R_1} + I_B + \frac{I_{IO}}{2}\right\}R_F - \left(I_B - \frac{I_{IO}}{2}\right)R_2 + V_{IO}$$

$$= \left(\frac{-R_2R_F}{R_1} + R_F - R_2\right)I_B + \frac{R_F}{R_1}.V_{IO} + V_{IO} + \left(\frac{R_2R_F}{R_1} + R_F - R_2\right)\left(\frac{I_{IO}}{2}\right)$$

$$V_O = \left\{R_F - R_2\left(\frac{R_F}{R_1} + 1\right)\right\}I_B + \left(\frac{R_F}{R_1} + 1\right)V_{IO} + \left\{R_F + R_2\left(\frac{R_F}{R_1} + 1\right)\right\}\left(\frac{I_{IO}}{2}\right)$$

………………(1)

The first term shows the effect of the input bias currents. It can be made zero by making

$$R_F = R_2\left(\frac{R_F}{R_1} + 1\right) \text{ or}$$

$$R_2 = \frac{R_1R_F}{R_1+R_F} \quad \text{………… (2)}$$

In other words, R_2 should be made equal to the resistance of R_1 and R_F in parallel. It is good practice to design to this requirement, to minimize the effects of bias current on the output offset.

The third term of Eqn.(1) shows the effect of the input offset current. Note the similarity between the terms in braces (curly brackets) in the equation. With the above condition met, the third term becomes

$$R_FI_{IO}$$

So, with the condition of Eqn. (2) met, the output offset voltage reduces to

$$V_O = \left(\frac{R_F}{R_1} + 1\right) V_{IO} + R_F I_{IO} \quad \text{..........................(3)}$$

Clearly, the first approach to reducing the output offset is to choose an op amp with low values of input offset voltage and current. In the most critical applications, an FET type, or a type with an FET input stage can be used, with negligible input current and, hence, input offset current. With a chosen op amp, the next strategy is to reduce the value of R_F, and the ratio R_F / R_1.

7.5 Differential input and output resistance:

Operational Amplifier is internally a Differential Amplifier (its first stage) with other important features like High Input Impedance, Low Output Impedance etc. For more information on Op-Amp, read Operational Amplifier Basics.

The Differential Pair or Differential Amplifier configuration is one of the most widely used building blocks in analog integrated-circuit design. It is the input stage of every Operational Amplifier.

A Difference Amplifier or a Differential Amplifier amplifies the difference between the two input signals. An operational amplifier is a difference amplifier; it has an inverting input and a non-inverting input. But the open loop voltage gain of an operational amplifier is too high (ideally infinite) to be used without a feedback connection.

So, a practical differential amplifier uses a negative feedback to control the voltage gain of the amplifier.

The following Fig.7.12 shows a simple Differential Amplifier using an Op Amp. Here, V_1 is the Non-Inverting Input Voltage, V_2 is the Inverting Input Voltage and V_{OUT} is the Output Voltage.

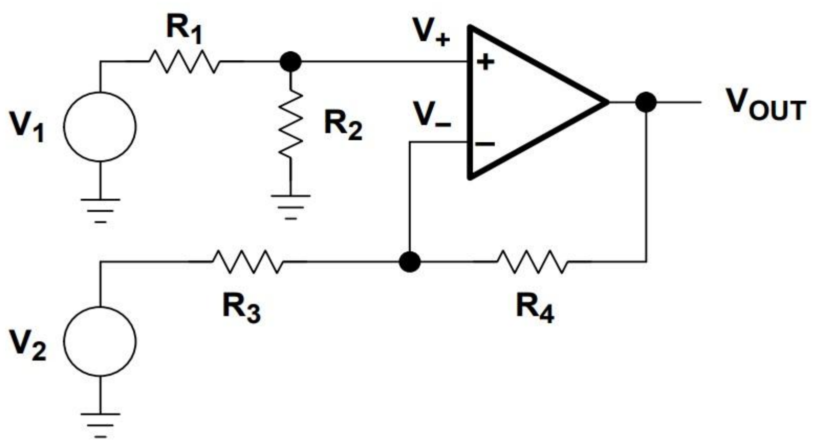


Fig.7.12 simple Differential Amplifier using an Op Amp

If you observe the above circuit of the difference amplifier, it is a combination of both the Inverting Amplifier and the Non-Inverting Amplifier. So, to calculate the output voltage of a Differential Amplifier, we will use both the Inverting and Non-Inverting outputs and add them together.

Calculating the Output Voltage:

Let V_+ be the voltage at the Non-Inverting terminal and $V_–$ be the voltage at the Inverting Terminal of the above Differential Amplifier Circuit. We can calculate the value of V_+ using the Potential Divider Rule.

Resistors R_1 and R_2 form a Voltage Divider Network with V_1 as the Input Voltage and V_+ as the output voltage and this V_+ is applied at the non-inverting terminal. So,

$$V_+ = V_1 (R_2 / R_1 + R2)$$

If V_+ is the input to the non-inverting terminal and G_+ is the gain of the Non-Inverting Amplifier, then non-inverting output V_{OUT+} is given by:

$$V_{OUT+} = V_+ G_+$$

From the above circuit, we can calculate the Non-Inverting Gain G_+ as:

$$G_+ = (R_3 + R_4) / R_3 = 1 + (R_4 / R_3)$$

Using the values of V_+ and G_+ in the equation of V_{OUT+}, we get

$$V_{OUT+} = V_1 (R_2 / R_1 + R2) (1 + (R_4 / R_3))$$

Coming to the Inverting Output $V_{OUT–}$, we have to calculate it with respect to the inverting input V_2 and the Inverting Gain $G_–$.

$$V_{OUT–} = V_2 G_–$$

From the above circuit, we can calculate the Inverting Gain $G_–$ as:

$$G_– = – R_4 / R_3$$

So, $V_{OUT–}$ is given by:

$$V_{OUT–} = V_2 (– R_4 / R_3) = – V_2 (R_4 / R_3)$$

We have both V_{OUT+} and $V_{OUT–}$ values. To get the final V_{OUT} value, we have to add these values.

$$V_{OUT} = V_{OUT+} + V_{OUT–}$$

$$V_{OUT} = V_1 (R_2 / R_1 + R2) (1 + (R_4 / R_3)) – V_2 (R_4 / R_3)$$

This is the output voltage of a Differential Amplifier. The above equation looks complex. So, to reduce the complexity and simply the equation, let us take a special case where $R_3 = R_1$ and $R_4 = R_2$.

If we apply these values in the above equation, we the output voltage as:

$$V_{OUT} = R_2 / R_1 (V_1 - V_2) = R_4 / R_3 (V_1 - V_2)$$

Now, from this equation, it is clear that the differential voltage ($V_1 - V_2$) is multiplied by the gain R_2 / R_1. Hence, it is Differential Amplifier.

Alternative way to Calculate Output Voltage:

Let us now calculate the output voltage by determining the current at the Inverting Input of the Op Amp. Let us assume the following circuit for a Differential Amplifier. This circuit is similar to the previous one, except this a special case of $R_3 = R_1$ and $R_4 = R_2$ of the previous circuit.

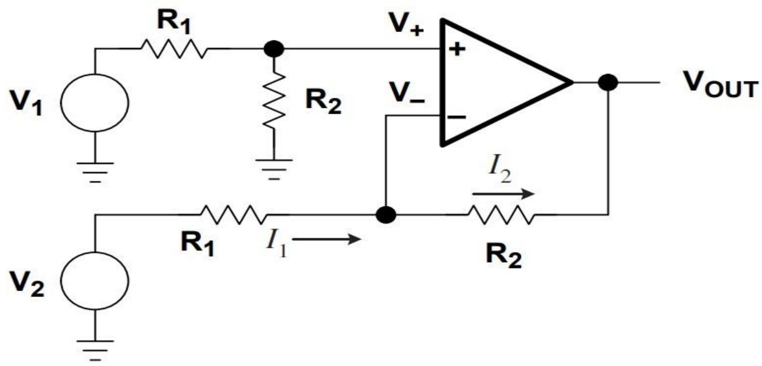


Fig.7.13 Circuit for a Differential Amplifier

First, we have to determine the voltage at the Non-Inverting terminal (V_+). We already calculated this in the previous derivation using the voltage divider rule. The value is given by:

$$V_+ = V_1 (R_2 / R_1 + R_2)$$

Now, from the basic understanding of the Operational Amplifier, we can say that no current flows in or out of the Op Amp input terminals. So, the current entering the Inverting Terminal I_1 is same as the current leaving the terminal I_2.

$$I_1 = I_2$$

Using this rule as a reference, we can apply Kirchhoff's Current Law at the Inverting Input Terminal and we get:

$$(V_2 - V_-) / R_1 = (V_- - V_{OUT}) / R_2$$

Another important rule about Operational Amplifier is that it tries to keep the Input Terminals at same voltage. So, $V_+ = V_-$. Using this rule, we can replace V_- in the above equation with the previously calculated V_+ value.

After replacing and performing some calculations, we arrive at:

$$V_{OUT} = R_2 / R_1 (V_1 - V_2)$$

NOTE: In all the previous calculations, we took a special as $R_3 = R_1$ and $R_4 = R_2$. Actually, instead of this we have to consider the ratios i.e.,

$$R_3 / R_4 = R_1 / R_2$$

If this condition is used, then the resistances are said to be in a Balanced Bridge.

Important Parameters of Differential Amplifier:

Let us now see some of the important parameters of a Difference Amplifier. They are:

- Gain
- Common Mode Input
- Common Mode Rejection Ratio (CMRR)

Differential Amplifier Gain: The gain of a difference amplifier is the ratio of the output signal and the difference of the input signals applied. From the previous calculations, we have the output voltage V_{OUT} as

$$V_{OUT} = R_2 / R_1 (V_1 - V_2)$$

So, Differential Amplifier Gain A_D is given by

$$A_D = V_{OUT} / (V_1 - V_2) = R_2 / R_1$$

Common Mode Input: In all the previous calculations, we assumed the Balanced Bridge condition i.e., $R_3 / R_4 = R_1 / R_2$. To understand a unique characteristic of the Differential Amplifier or Difference Amplifier, we have to take a look at the Differential Mode Input and Common Mode Input Components.

The Differential Mode Input V_{DM} and Common Mode Input V_{CM} are given by:

$$V_{DM} = V_1 - V_2$$

$$V_{CM} = (V_1 + V_2) / 2$$

Rearranging the above two equations, we get

$$V_1 = V_{CM} + V_{DM} / 2 \text{ and } V_2 = V_{CM} - V_{DM} / 2$$

The following circuit shows the Common Mode Input Signals.

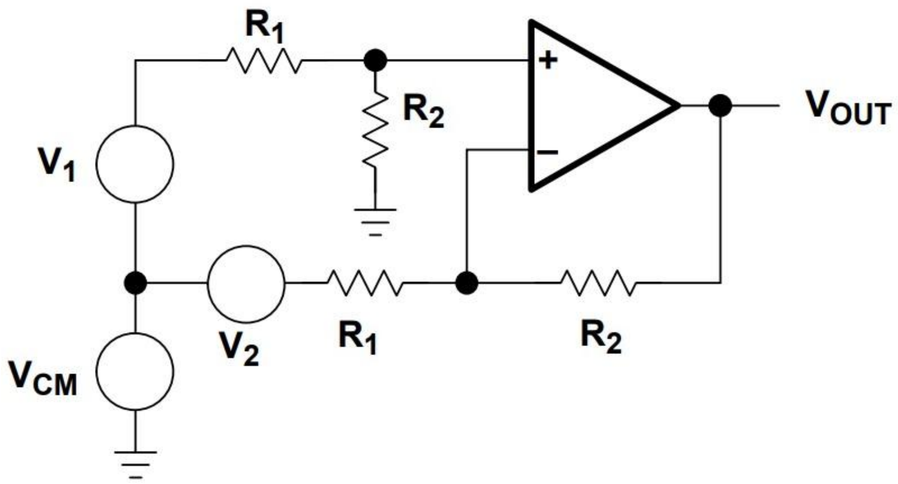


Fig.7.14 Common Mode Input Signals

As the Difference Amplifier amplifies only the Difference Mode component, it ignores the Common Mode Component. If we tie the inputs together, the V_{DM} becomes 0 and the V_{CM} is a non-zero value.

But a true Differential Amplifier will result in $V_{OUT} = 0$, as it completely ignores the Common Mode portion of the input signal. Due to this, the Differential Amplifier is often used at the input stage of a system to strip the DC or the Common-Mode noise from the input.

All these calculations are true if and only if the Resistances form the Balanced Bridge Condition. Since the output of a practical difference amplifier depends upon the ratio of the input resistances, if these resistor ratios are not exactly equal, the common mode voltage V_{CM} will not be completely cancelled. Because it is practically impossible to match resistor ratios perfectly, there is likely to be some common mode voltage.

With the common mode input voltage present, the output voltage of the differential amplifier is given as:

$$V_{OUT} = A_D\, V_{DM} + A_C\, V_{CM}$$

Where V_{DM} is the difference voltage $V_1 - V_2$

V_{CM} is the common mode voltage $(V_1 + V_2) / 2$

A_D and A_C are Differential Mode and Common Mode Gains respectively.

Common Mode Rejection Ratio (CMRR): The ability of a Differential Amplifier to reject common mode input signals is expressed in terms of Common Mode Rejection Ratio (CMRR). The Common Mode Rejection Ratio of a Differential Amplifier is mathematically given as the ratio of Differential Voltage gain (A_D) of the Differential Amplifier to its Common Mode gain (A_C).

$$CMMR = A_D / A_C$$

In terms of decibels (dB), the CMRR is expressed as

$$CMMR_{dB} = 20 \log_{10} (| A_D / A_C |)$$

For an ideal Difference Amplifier, the common mode voltage gain is zero. Hence, the CMRR is infinite.

Characteristics of a Differential Amplifier:

a) High Differential Voltage Gain
b) Low Common Mode Gain
c) High Input Impedance

d) Low Output Impedance
e) High CMRR
f) Large Bandwidth
g) Low offset voltages and currents

Differential Amplifier as Comparator:

A Differential Amplifier circuit is a very useful Op Amp circuit, since it can be configured to either “add” or “subtract” the input voltages, by suitably adding more resistors in parallel with the input resistors.

A Wheatstone Bridge Differential Amplifier circuit design is as shown in the following Fig.7.15. This circuit behaves like a Differential Voltage Comparator.

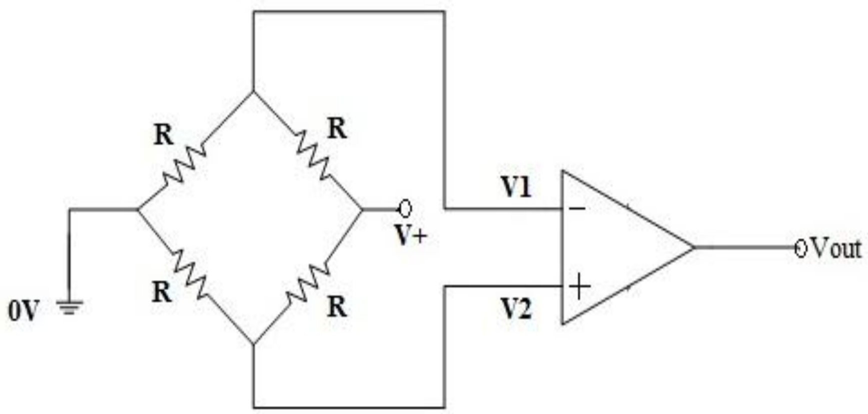


Fig.7.15 Wheatstone bridge Differential Amplifier

By connecting one input to a fixed voltage and the other to a Thermistor (or a light-dependent resistor), the differential amplifier circuit detects high or low levels of temperature (or intensity of light) as the output voltage becomes a linear function of the changes in the active leg of the resistive bridge network.

A Wheatstone Bridge Differential Amplifier can also be used to find the unknown resistance in the resistive bridge network, by comparing the input voltages across the resistors.

Light Activated Switch using Differential Amplifier: The circuit shown in the following Fig.7.16 acts as a light-dependent switch, which turns the output relay either “ON” or “OFF” as the intensity of the light falling upon the light-dependent resistor (LDR) exceeds or falls below a pre-set value at the non-inverting input terminal V_2.

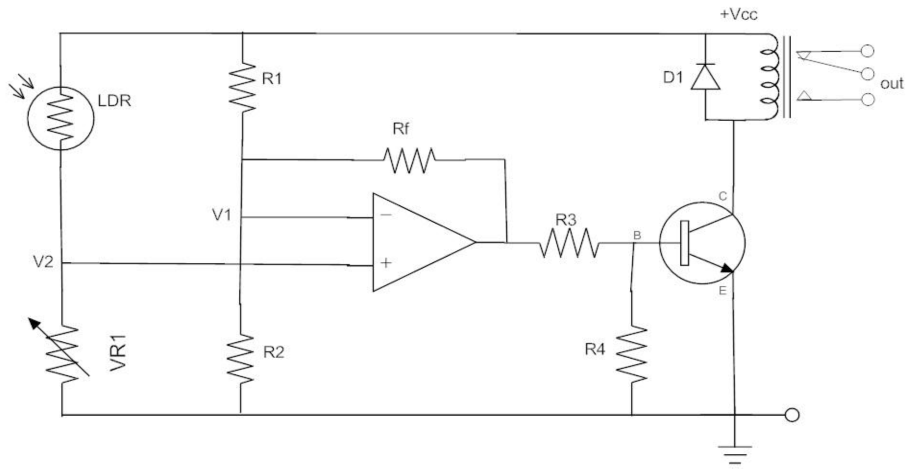


Fig.7.16 Light-dependent switch using Differential Amplifier

The voltage V_2 is determined by the variable resistor V_{R1}. The resistors R_1 and R_2 act as a potential divider network. A fixed reference voltage is applied to the inverting input, through R_1 and R_2.

Advantages: Some advantages of Differential Amplifier are**:**

- Differential Amplifier has noise cancellation property.
- Differential Amplifier can reduce external interference.
- The nature of these amplifiers is linear.
- These amplifiers help to increase CMRR(Common Mode Rejection Ratio) which further helps to avoid unwanted signal.

Disadvantages: Some disadvantages of Differential Amplifier are,

- Complexity
- Proper biasing needed

Differential Amplifier Applications: The important applications of Differential Amplifier are:

- The main application of the differential amplifier is to amplify the balanced differential signal.
- Differential Amplifier circuits are used in the audio amplifier for accurate and noiseless volume control.
- In analog and digital data transmission system differential amplifiers are used for noise cancellation.

- Differential Amplifiers are used for audio and video processing.
- They also used as an automatic gain control circuit.
- These amplifiers are used for amplitude modulation.
- Differential amplifier circuit also used as a negative feedback circuit.
- They are also used as an electronic switch.
- They are also used for motor control.
- There are huge applications of Differential amplifier in the control system.
- These amplifier circuits are also used as a high pass filter circuit.s
- Differential amplifiers are used in earlier days in analog computers.

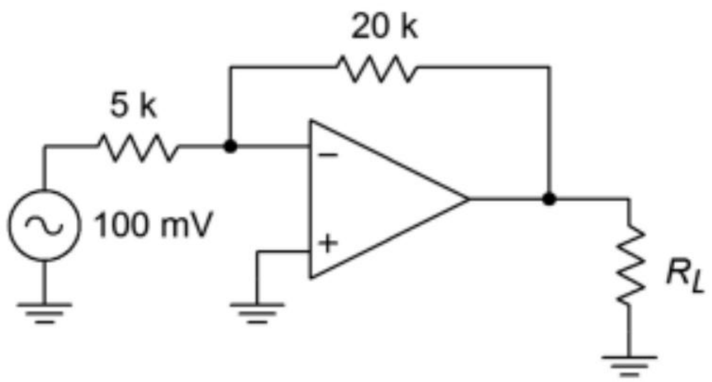
SAQ.1

a) What do you understand by OP-amplifier with symbol and number code?

b) Discuss about the input-offset and output offset voltage for op-amp.

c) What do you mean by Differential input and output resistance using differential amplifier?

d) Determine the input impedance and output voltage for the op-amp circuit shown below. R_L is the load resistance.



7.6 Common mode rejection ratio, output current, power consumption, slew rate gain-band width product:

Common Mode Rejection Ration of Op Amp:

The common mode rejection ratio is a differential amplifier and the op amps are amplified in with the differential input. Hence the CMMR ratio can be applied to the operational amplifier. By using the condition of common mode rejection ratio, i.e. when both the input of the amplifier has same voltages, then the output of the amplifier should be zero or the amplifier should be rejecting the signal. The following Fig.7.17 shows the amplifier of MCP601 of common mode rejection ratio.

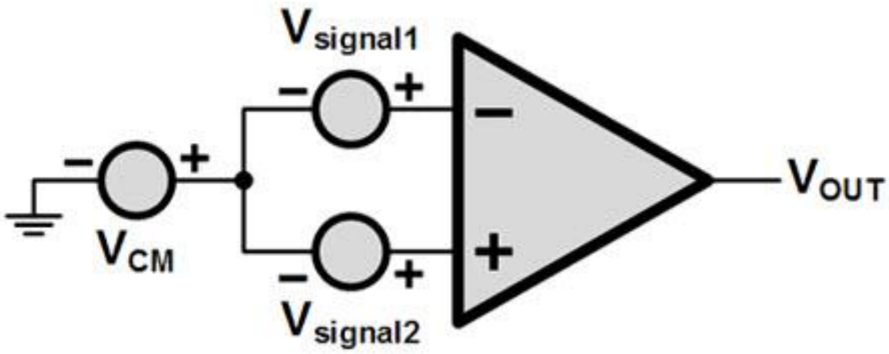


Fig.7.17 Common Mode Rejection Ration of Op Amp

Common-mode Rejection Ratio Formula: The common mode rejection ratio is formed by the two inputs which will have the same sign of DC voltage. If we assume one input voltage is 8V and the other 9V here the 8V is common and the input voltage should be calculated through the equation of V+ – V- . Hence the result will be 1V but the common DC voltage between the two inputs has a non-zero gain.

The differential gain Ad magnifies the difference between the two input voltages. But the common mode gain Ac magnifies the common mode DC voltage between the two inputs. The ratio of two gains is said to be as a common mode rejection ratio. The value of the format is in dB. The formula of a common mode rejection ratio is calculated by the following equation.

$$CMRR = 20\log|Ao/Ac| \text{ dB}$$

Power Supply Rejection Ratio: The power supply rejection ratio is defined as the changes in input offset voltage per unit changes in the DC supply voltage. The power supply is also calculated in the format of dB. The mathematical equation of the power supply rejection ratio is given below.

$$PSRR= 20\log|\Delta VDc/\Delta Vio| \text{ dB}$$

Offset Error of a CMRR of the Op-Amp: The CMRR can build parallel out offset voltage in op amps configured in the non-inverting amplifier which is shown in the below Fig.7.18. The non-inverting operating amplifier will have a small amount of CMRR error because both the inputs are connected to ground; there is no presence of CM dynamic voltage.

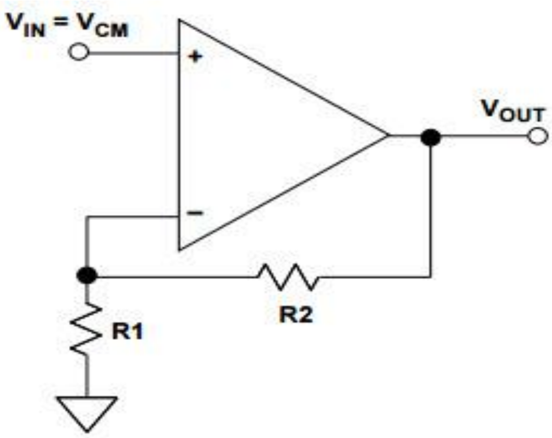


Fig.7.18 Offset Error of a CMRR of the op amp

Error (RTI) = Vcm / CMRR = Vin / CMRR

Vout = [1 + R2/R1] [Vin + Vin/ CMRR]

Error (RTO) = [1+R2/R1] [Vin/CMRR]

Measuring Common Mode Rejection Ratio: There are different ways to measure the common mode rejection ratio. In the below Fig.7.19 we will discuss the four precision resistor to configure the op amp as a differential amplifier. A signal is applied to the both inputs, changes in the output are measured and an amplifier with infinite CMRR also no changes in the output. The inherent difficulties of this circuit are that the ratio match of the resistors is important as the CMRR of the op amp. The 0.1% mismatch is between resistor pair and the result will be in CMR of 66 dB. Hence the most of the amplifiers will have a low frequency of CMR is between the 80dB to 120Db. In this circuit, it is clear that there is only marginally useful for measuring the CMRR.

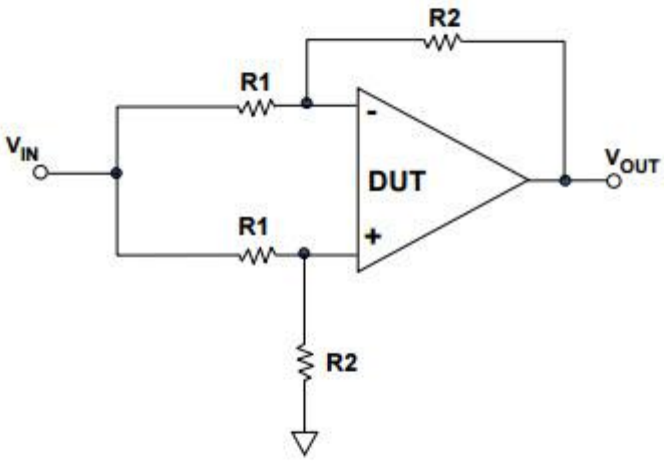


Fig.7.19 Measuring Common Mode Rejection Ratio

$\Delta Vout = \Delta Vin / CMRR (1 + R2/R1)$

CMRR without Using Precision Resistors: The following circuit is more complicated by comparing with the above circuit (Fig.7.20) and it can measure the CMRR by without using a precision resistor. By switching the power supply voltage the common mode rejection ratio is changed. Practically, the circuit can be implemented easily and by using the same circuit we can apply different power supply voltages to measure the power supply rejection ratio.

In the following circuit, the power supply is from the +-15 DUT op amp with the common mode voltage range of +-10V. From the following circuit, the integrated amplifier A1 should have high gain, low Vos and low IB and the op amp is 097 devices.

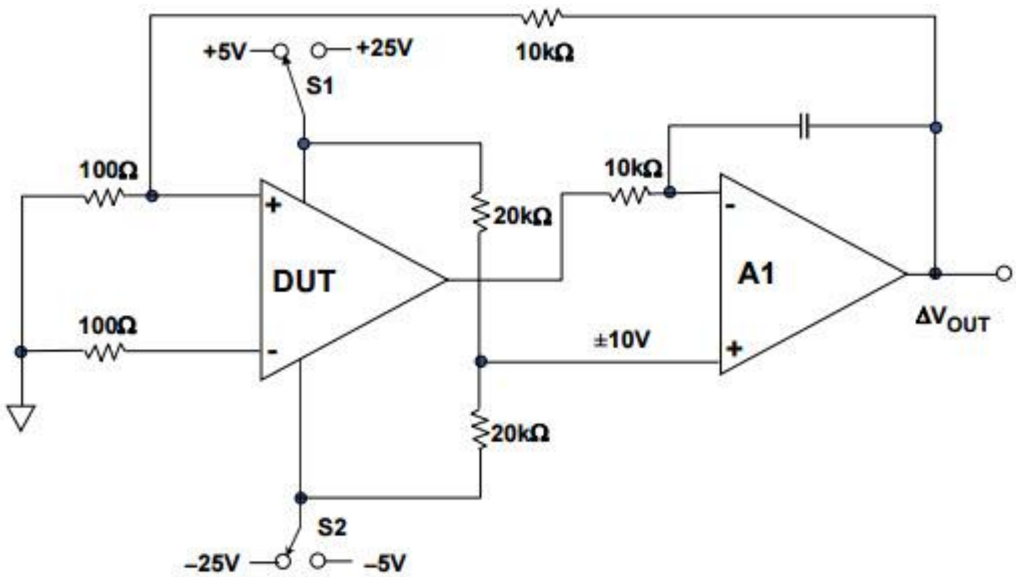


Fig.7.20 CMRR without Using Precision Resistors

Output current of op-amp: A typical op-amp can be expected to continuously sink or source not more than 30 or 40 mA, though some parts can handle closer to 100 mA, and others will struggle to give you 10 mA. There is a special category of high-output-current amplifiers, with current capability approaching or even exceeding 1000 mA.

The most basic circuit (Fig.7.21) for buffering an op-amp's output current is the following:

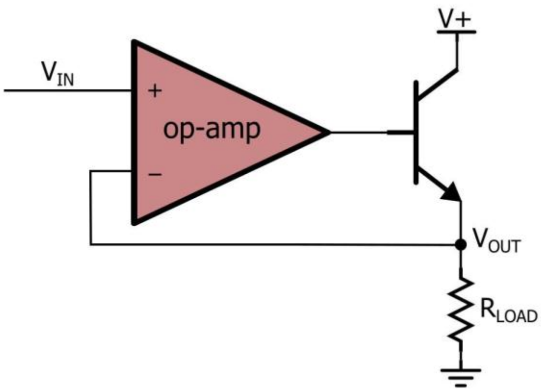


Fig.7.21 Basic circuit for buffering an op-amp's output current

And here is a corresponding LT spice schematic (Fig.7.22):

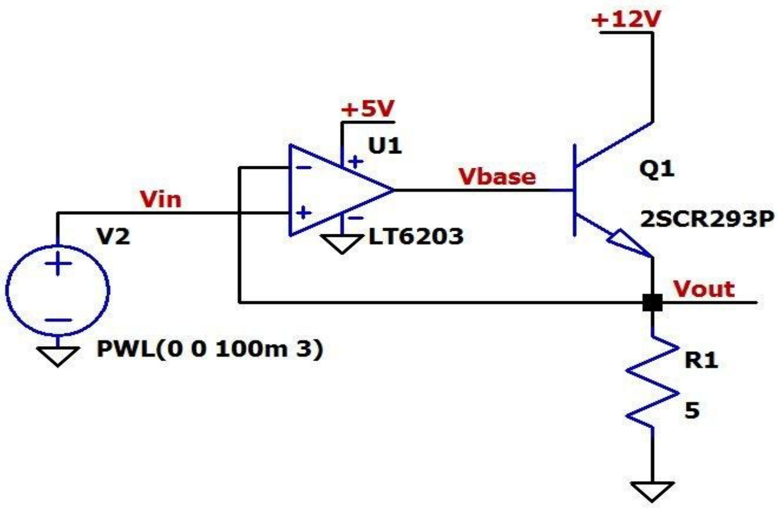


Fig.7.22 LT spice schematic

Let's get a solid conceptual understanding of this circuit before we move on. The input is applied to the non-inverting op-amp terminal, and the output is connected directly to the base of the BJT. The op-amp and the BJT could use the same positive supply, but in this case, we are assuming that two voltages are available a 5 V supply for low-power, low-noise circuitry, and 12 V for the high-power portion of the design. The value of the load resistor is very low, such that output voltages greater than about 200 mV applied directly to the load would require more output current than the LT6203 can provide. The transistor chosen in the LT spice schematic can handle about 1000 mA, which means it is good for load voltages up to 5V.

The key to this circuit is the feedback connection. Remember the “virtual short”: when analyzing an op-amp in a negative-feedback configuration, we can assume that the voltage at the non-inverting terminal equals the voltage at the inverting terminal. This alone tells us that the output voltage (i.e., the voltage across the load) will be equal to the input voltage. But let’s go a little deeper to make sure we actually understand what’s going on; the virtual short is sort of a superstition that can distract us from the reality of how an op-amp functions. The op-amp multiplies the differential input voltage by a very large gain. Thus, with negative feedback, the op-amp rapidly reaches equilibrium because the large changes in output voltage reduce the differential voltage that is causing these very output changes. In this equilibrium state, the output has stabilized at whatever voltage eliminates the difference between the voltages at the inverting and non-inverting input terminals—in other words, the op-amp automatically adjusts its output in whatever way is needed to make V_{IN-} equal to V_{IN+}.

In the context of this output-buffering circuit, the op-amp automatically generates whatever output voltage is needed to make the BJT’s emitter voltage equal to the input voltage. Think how difficult this would be in an open-loop situation—somehow the amplifier’s input-to-output relationship would have to be designed to compensate for the BJT’s base-to-emitter voltage drop, which is neither linear nor predictable. But with an op-amp plus some negative feedback, the problem becomes trivial.

Let’s reinforce this conceptual understanding with a couple simulations. The first is not very exciting; it simply confirms that the output voltage tracks the input voltage (the V_{IN} trace is hidden underneath the V_{OUT} trace):

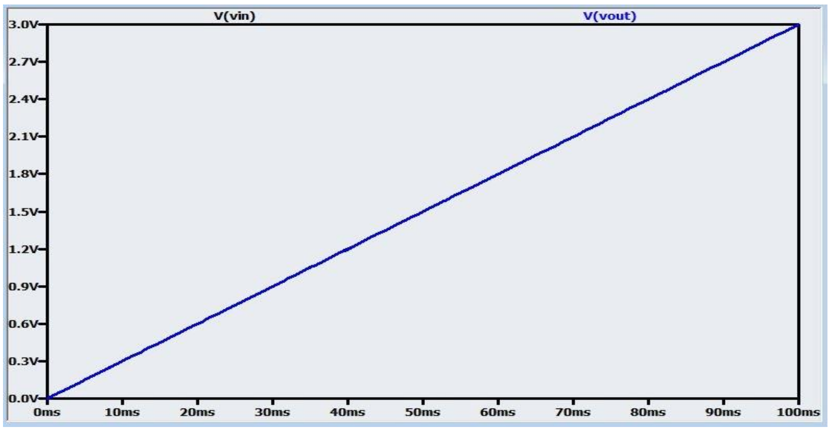


Fig.7.23 Plot of the output voltage tracks the input voltage

This next plot shows what the op-amp output terminal must do to produce the proper voltage across the load.

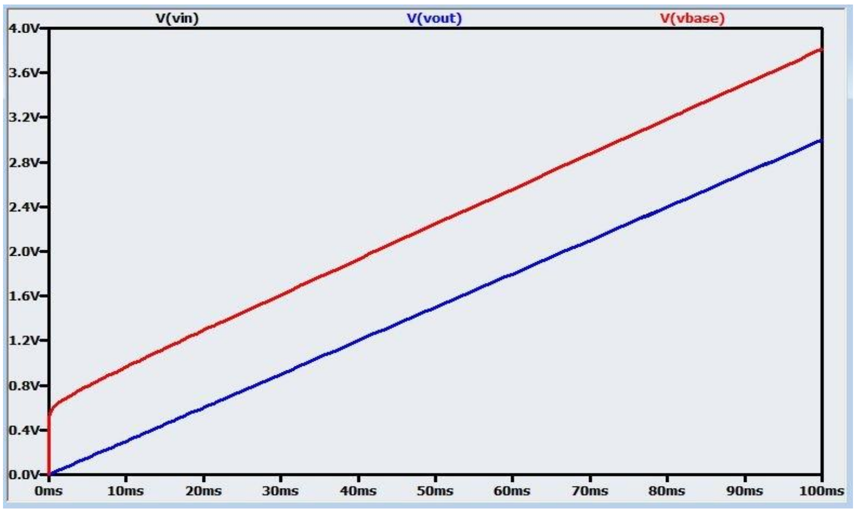


Fig.7.24 plot of the op-amp output terminal voltage across the load

Adding Gain: This basic circuit is not limited to the unity-gain configuration. As with a non-buffered op-amp, you can insert resistors into the feedback path to create overall gain from the input to the load voltage. Here is the non-unity-gain version of the circuit (Fig.7.25):

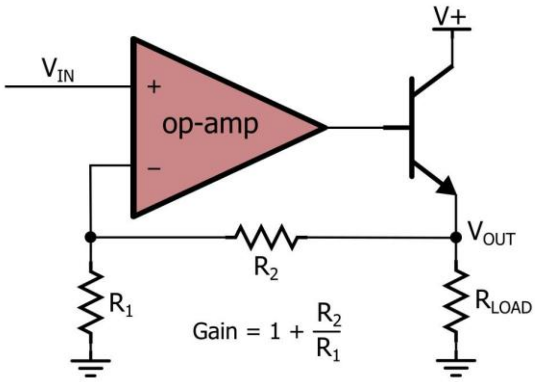


Fig.7.25 New LT spice schematic

And here is the new LT spice schematic, followed by a plot with V_{IN}, V_{OUT}, and the voltage applied to the base of the BJT (Fig.7.26).

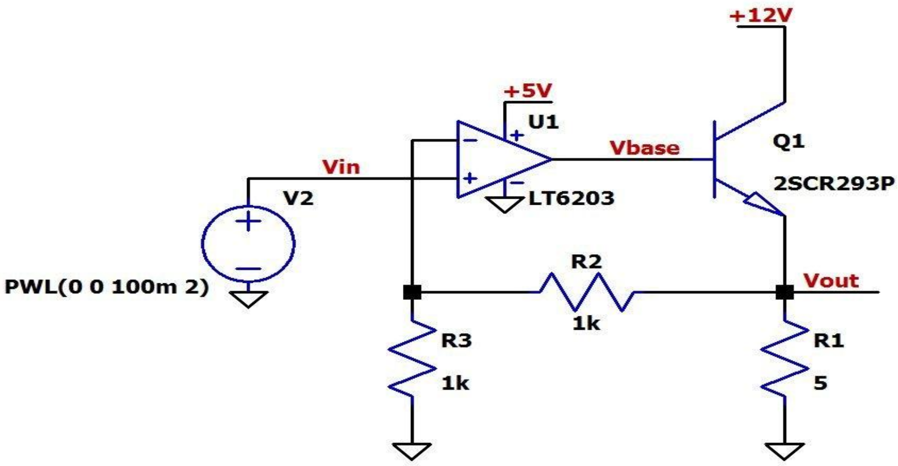


Fig.7.26 Voltage applied to the base of the BJT

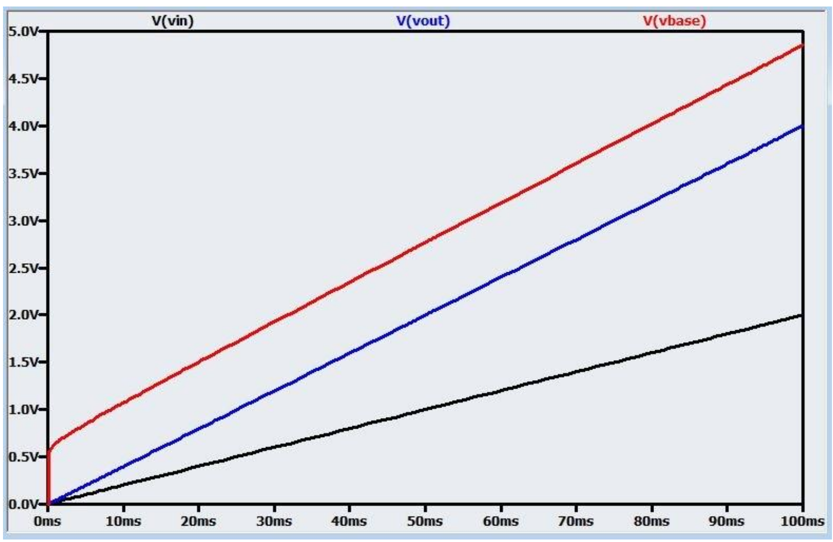


Fig.7.27 plot with V_{IN}, V_{OUT}

Power consumption: The simplest method for calculating power consumption or power dissipation in the op amp is to solve a power-balance equation, in which supply power equals the sum of power dissipated in the load and in the op amp. Thus, the op amp dissipation equals supply power minus load power.

To minimize the power dissipation in a single-supply op amp, driving a ground-referenced load, connect a pull-up resistor, with value equal to the load resistor between the output and the positive supply voltage. This addition enables the op amp to operate at higher ambient temperatures and drive lower-resistance loads, limited only by its maximum ratings for output voltage and current (rather than the package power dissipation).

To maximize signal swing, the output of a single-supply op amp is usually biased at half the supply voltage (Fig.7.28). For ground-referenced loads, however, this configuration causes maximum power dissipation in the IC.

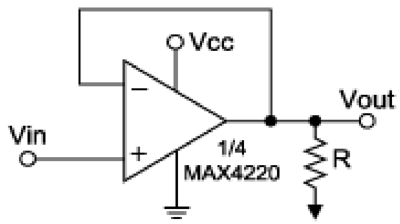


Fig.7.28 A single-supply op amp biased at mid-V_{CC}.

The solution is simple and effective: connect a pull-up resistor, with value equal to the load resistor, between the output and the positive supply voltage (Fig.7.29). This addition enables the op amp to operate at higher ambient temperatures and drive lower-resistance loads, limited only by its maximum ratings for output voltage and current (rather than the package power dissipation).

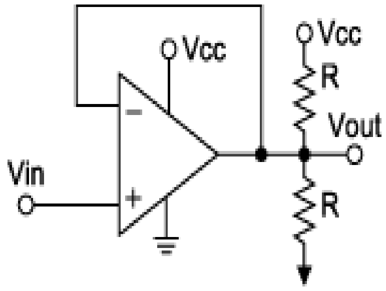


Fig.7.29 Power dissipated in Fig.7.28 is reduced by adding a load-value pull-up to V_{CC}.

For example, consider the MAX4220 quad op amp, in which each output drives a 30Ω resistor to ground. For V_{CC} = 5V, the device exceeds its package power rating. Connecting 30Ω pullups at each output, however, minimizes the IC's power dissipation because each op-amp's output current is zero. Power is now dissipated in the pull-up resistors and not in the op amps.

Calculating power dissipation for the op amp in Figure 1a is straightforward: $P_{DC} = (V_{CC} - V_{OUT}) V_{OUT}/R$.

Solving the differential equation $dP_{DC}/dV_{OUT} = 0$ for V_{OUT} shows that the op amp's maximum power dissipation ($V_{CC}^2/4R$) occurs when $V_{OUT} = V_{CC}/2$.

Slew Rate: In electronics, the slew rate is defined as the maximum rate of output voltage change per unit time. It is denoted by the letter S. The slew rate helps us to identify the amplitude and maximum input frequency suitable to an operational amplifier (OP amp) such that the output is not significantly distorted.

The slew rate should be as high as possible to ensure the maximum undistorted output voltage swing.

Slew rate is a critical factor in ensuring that an OP amp can deliver an output that is reliable to the input. Slew rate changes with the change in voltage gain. Therefore, it is generally specified at unity (+1) gain condition.

A typically general-purpose device may have a slew rate of 10 V/µs. This means that when a large step input signal is applied to the input, the electronic device can provide an output of 10 volts in 1 microsecond.

Slew Rate Formula: The equation for the slew rate is given by

$$S = \frac{dV_0}{dt}\Big|_{\text{maximum}} \text{Volts / µs}$$

Where Vo is the output produced by the amplifier as a function of time t.

Slew Rate Units: The slew rate of an electronic circuit is defined as the rate of change of the voltage per unit of time. The units for slew rate are Volts per second or V/µs.

How to Measure Slew Rate?

The slew rate is measured by applying a step signal to the input stage of the op-amp and measuring the rate of change occurs at the output from 10% to 90% of the output signal's amplitude. Generally, the applied step signal is large and it is about 1 V.

The slew rate is measured from the output voltage waveform as:

$$\text{Slew Rate (S)} = \frac{dV_0}{dt} = \frac{V_{0(90\%)} - V_{0(10\%)}}{t_{(90\%)} - t_{(10\%)}}$$

The slew rate can be measured by using an oscilloscope and a function generator.

The circuit used for slew rate measurement is shown in the Fig.7.30 below.

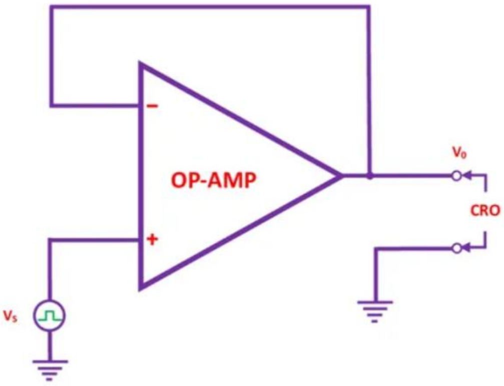


Fig.7.30 Slew Rate Measurement Circuit

The input and slew limited output voltage waveform are shown in the Fig.7.31 below.

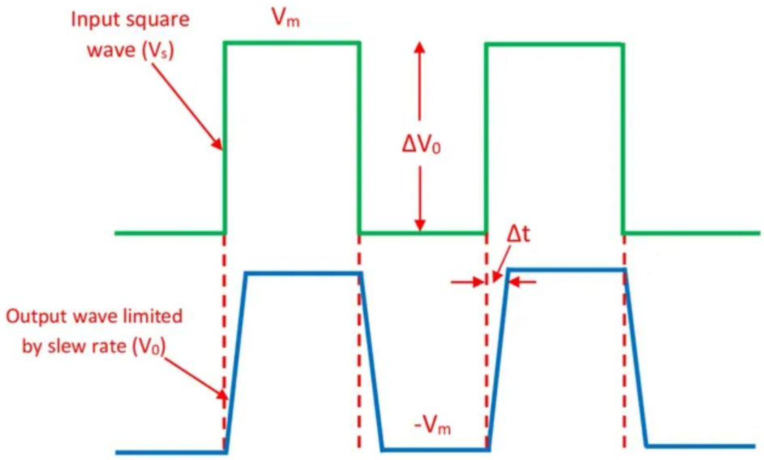


Fig.7.31 Input and Slew Limited Output Voltage Waveform

Slew Rate of OP Amp: Slew rate decides the capability of an op-amp to change its output rapidly; hence it decides the highest frequency of the operation of a given op-amp.

The Slew rate of the op-amp can limit the performance of a circuit and it can distort the output waveform if its limit is exceeded.

Op-amps may have different slew rates for positive and negative transitions because of the circuit configuration.

The Slew rate should be ideally infinite and practically as high as possible. The Slew rate of IC 741 op-amp is only about 0.5 V/μs which is its major drawback. Therefore, it cannot be used for high-frequency applications.

Slew Rate Limiting in Amplifiers

High Gain at the Input Stage of OP-Amp:

The modern operational amplifiers use high gain differential input stages with trans-conductance characteristics. This means that an amplifier takes a differential input voltage at the input stage and produces an output current at the output stage. Note that trans-conductance is nothing but the transfer conductance, also called as mutual conductance, it is the electrical characteristic and is defined as the current through the output of a device to the voltage across the input of a device. Mathematically it is expressed as $g_m = I_{out}/V_{in}$.

This trans-conductance of amplifiers is typically very high; also at this point, the large open-loop gain of the amplifier is produced. This means that the small input voltage can saturate the input stage. In this saturation condition, the stage produces approximately constant output current and acts as a constant current source. In this condition, the rate of change occurs at the output of the amplifiers is severely limited. This limits the slew rate of an OP-amp.

Frequency Compensation at the Second Stage of OP-Amp: To provide stability, frequency compensation is used in all op-amps to reduce the high-frequency response have a considerable effect on slew rate. A reduced frequency response limits the rate of change that occurs at the output of the amplifiers and hence it affects the slew rate of an op-amp.

Now, the frequency compensation at the second stage of the op-amp is the low pass characteristic and it is similar to an integrator. Hence constant current input will produce a linearly increasing output. If the second stage has an effective input capacitance C and voltage gain A2, then the slew rate can be expressed as

$$\text{Slew Rate (S)} = \frac{I_{constant}}{C} A_2$$

Where $I_{constant}$ is the constant current of the first stage in saturation.

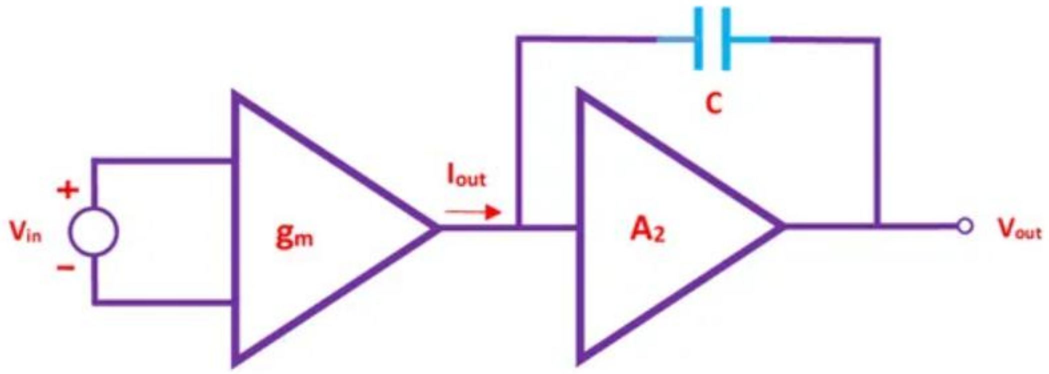


Fig.7.32 Slew Limiting Circuit

Where:

g_m = Amplifier with trans-conductance characteristics = I_{out}/V_{in},

I_{out} = Constant current in saturation, so V_{out} linearly increases across the capacitor C,

A_2 = Voltage gain,

V_{out} = Slew limited output, the output can change at a faster rate.

Temperature: The slew rate is a temperature-dependent parameter. A positive slew rate occurs when a signal is rising and the negative slew rate occurs when a signal is falling. Typically, the slew rate of an amplifier will increase with increasing temperature.

Slew Rate vs Bandwidth

Slew Rate: Slew rate is the maximum rate at which an amplifier can respond to the sudden change of input level. Slew rate can distort (or limit) any signal amplified by an op-amp.

The sinusoidal input signal multiplied by the gain of the op-amp results in a slope which is higher than the slew rate of the op-amp. Hence output waveform will be a straight line instead of a curving section of the sinusoidal. This effect is non-linear. Thus, slewing can modify or distort the shape of a signal.

Bandwidth: The Bandwidth or power bandwidth of an amplifier is the range of frequencies for which all the signal frequencies are amplified almost equally (without distortion) also poles in the transfer function of op-amp leads to low-pass-filter behavior i.e., the amplitude of the signal decreases as frequency increases, and phase shift occurs. This effect is linear and they do not produce distortion into the output signal.

The bandwidth of an op-amp should be as large as possible. It should be capable of amplifying the signal from zero frequency. Thus, the gain of an op-amp should be constant from 0 frequency to infinite frequency. Bandwidth is expressed in hertz.

$$V_S = V_m \sin\omega t \text{……………….(1)}$$

Now, for a unity gain non-inverting amplifier, the output is exactly equal to the input.

$$V_0 = V_m \sin\omega t$$

Differentiate the above equation both the sides we get,

$$\frac{dV_0}{dt} = \frac{d}{dt} V_m \sin\omega t$$

$$\frac{dV_0}{dt} = \omega V_m \cos\omega t \text{ ……………..(2)}$$

Now, dVo/dt will be maximum when cos = 1 (i.e. $\omega t = 0^o$)and maximum value of dVo/dt is nothing but the slew rate S. Put it into equation (2) we get,

$$S = \frac{dV_0}{dt}\Big|\text{maximum} = \omega V_m = 2\pi\, f_m\, V_m \text{ V/Sec……………(3)}$$

Where, f_m= the maximum signal frequency in Hz

V_m= the maximum peak voltage of the signal

Rearranging the term in equation (3), we get

$$f_m = \frac{S}{2\pi\, V_m} \text{ …………………..(4)}$$

The above equation indicates the highest frequency at which the peak-to-peak output voltage swing is equal to the DC output voltage range. In other words, it is the maximum frequency f_m for which the amplifier produces an undistorted output. It is called as full power bandwidth**.** It is also sometimes described as the slew-rate-limited-bandwidth**.**

Applications of Slew Rate: Some of the applications of the Slew Rate include:

- In musical instruments, slew circuitry is used to provide a slide from one note to another i.e. portamento (also called glide or lag).
- Slew circuitry is used where the control voltage is slowly transitioned to different values over a period of time.

- In certain electronics applications where speed is required and the output needs to change over a period of time, software-generated slew functions or slew circuitry are used.

Gain-bandwidth product:

When designing an op amp circuit, a figure known as the op amp gain bandwidth product is important.

The op amp gain bandwidth product is generally specified for a particular op amp type an open loop configuration and the output loaded:

$$GBP = Av \times f$$

Where:

GBP = op amp gain bandwidth product

Av = voltage gain

f = cutoff frequency (Hz)

The op amp gain bandwidth product is constant for voltage-feedback amplifiers. However it is not applicable for current feedback amplifiers because relationship between gain and bandwidth is not linear.

Therefore decreasing the gain by a factor of ten will increase the bandwidth by the same factor.

7.7 Characteristics Op-amp: The op-amp characteristics of an op-amp are:

a) Infinite open-loop gain G = Vout / V.
b) Infinite input impedance Rin, and so zero input current.
c) Zero input offset voltage.
d) Infinite output voltage range.
e) Infinite bandwidth with zero phase shift and infinite slew rate.
f) Zero output impedance Rout, and so infinite output current range.
g) Zero noise.

We will some explain them one by one here:

Open Loop Voltage Gain (A): The open loop voltage gain without any feedback for an ideal op amp is infinite. But typical values of open loop voltage gain for a real op amp ranges from 20,000 to 2, 00,000. Let the input voltage be V_{in}. Let A be the open loop voltage gain. Then the output voltage is $V_{out} = AV_{in}$. The value of a typically is in the range specified above but for an ideal op amp, it is infinite.

Input Impedance (Z_{in}): Input Impedance is defined as the input voltage by the input current. The input impedance of an ideal op amp is infinite. That is there no current flowing in the input circuit. However, an ideal op amp has certain current flowing in the input circuit of the magnitude of few pico-amps to a few milli-amps.

Output Impedance (Z_{out}): Output impedance is defined as the ratio of the output voltage to the input current. The output impedance of an ideal op amp is zero, however, real op amps have an output impedance of 10-20 kΩ. An ideal op amp behaves like a perfect voltage source delivering current without any internal losses. The internal resistance reduce the voltage available to the load.

Bandwidth (BW): An ideal op amp has an infinite bandwidth that is it can amplify any signal from DC to the highest AC frequencies without any losses. So therefore, an ideal op amp is said to have infinite frequency response. In real op amps, the bandwidth is generally limited. The limit depends on the gain bandwidth (GB) product. GB is defined as the frequency where the amplifier gain becomes unity.

Offset Voltage (V_{io}): The offset voltage of an ideal op amp is zero, which means that the output voltage will be zero if the difference between the inverting and non-inverting terminal is zero. If both the terminals are grounded, the output voltage will be zero. But real op amps have an offset voltage.

Common Mode Rejection Ratio (CMRR): Common mode refers to the situation when the same voltage is applied to both the inverting and non-inverting terminal of the op amp. The common mode rejection refers to the ability of the op amp to reject the common mode signal. Now we are in a position to understand the term common mode rejection ratio.

The common mode rejection ratio refers to the measure of the ability of the op amp to reject the common mode signal. Mathematically it is defined as

$$CMMR = |A_D / A_{CM}|$$

Where, A_D is the differential gain of the op amp, ∞ for an ideal op amp. A_{CM} refers to the common mode gain of the op-amp. The CMRR of an ideal op amp is ∞. That means it is able to reject all common mode signal. Also from the formula, we can see the A_D is infinite for an ideal op amp and A_{CM} is zero. Therefore the CMRR of an ideal op-amp is infinite. Therefore it will reject any signal which is common to both. However, real omp have finite CMRR, and does not reject all common mode signals.

Characteristics of Comparator: A comparator as name says, it compares a signal on one input of an op-amp with a known voltage called the reference voltage on the other input comparator is nothing but an open loop op-amp with two analog inputs (differential input) and one digital output (signal ended output).

The op-amp has very large gain used in open loop. Hence the output may be in positive or negative saturation voltage depending upon which input is larger. Op-amp is perfectly suited for comparator application because of its high input impedance and large open loop gain.

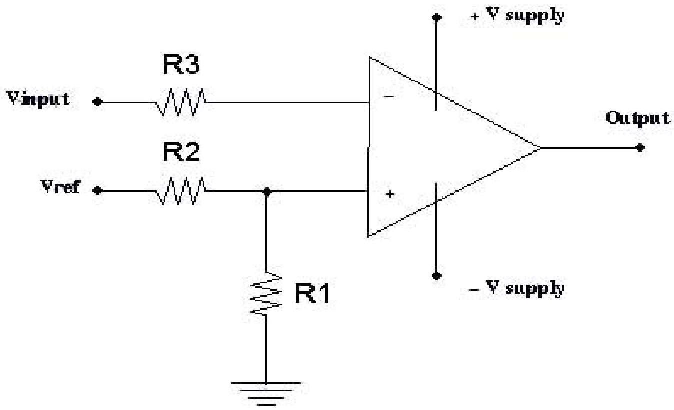


Fig.7.33 Comparator using op-amp

The important characteristics of comparator are:

1. Speed of operation
2. Accuracy
3. Compatibility of output

1. Speed of operation: The output of comparator must switch rapidly between the saturation level (+Vsat or -Vsat) and also respond instantly to any change of condition at its input. It says that bandwidth of op-amp should be very high because wider bandwidth, higher is the speed of operation.

2. Accuracy: It is smallest amount of difference voltage required at the inputs of comparator to make the output change its state .It is measured in mv. The accuracy depends on voltage gain, common -mode rejection ratio (CMRR), input offset voltage and thermal drifts.

3. Compatibility of output: The comparator is a form of analog to digital converter; its output must swing between two logic levels suitable for a certain logic family such as transistors -transistor logic (TTL).

Characteristics of Detector:

Simple detector circuit:

Rectifier circuit gives average value of input signal; but in practice we need peak value of input signal. This is achieved by peak detector circuit. The following Fig.7.34 shows a simple detector circuit using diode and capacitor.

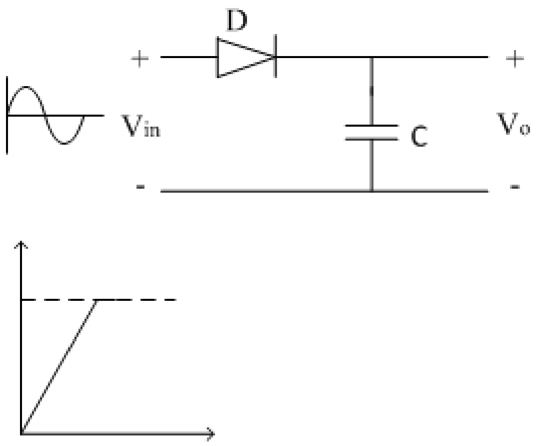


Fig.7.34 Simple detector circuit using diode and capacitor

In the positive half cycle, diode D is forward biased and capacitor C starts charging. When input reaches its peak value capacitor gets charged to positive peak value.

In negative half cycle, as input decreases, diode D is reversed biased and capacitor is isolated and holds the peak value of previous cycle. Hence it is called as a detector.

But in practice, output is taken across some load RL, so when input voltage decreases capacitor discharges through load RL. To avoid this select RL of very large value so that capacitor discharges very slowly hence almost holds the charge. Whatever charge it lost through RL is gets back in next half cycle.

Limitation:

The diode D is acting as an instant switch, so supply gets loaded.

To avoid the loading while charging capacitor, we use op-amp as follows. Op-amp is placed between input and diode D so loading is avoided as shown in circuit diagram (Fig.7.35) below,

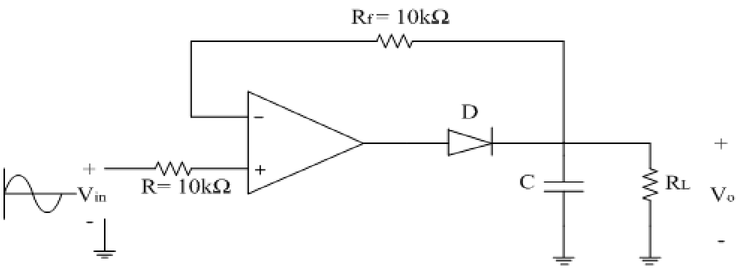


Fig.7.35 Op-amp is placed between input and diode D

In positive half cycle, output of op-amp is positive so diode D is forward biased, capacitor charges to peak value of input signal.

In negative half cycle, when input decreases diode D is reversed biased and capacitor is isolated and holds the charge of previous half cycle. Since diode is reversed biased, op-amp is in open loop condition and goes into saturation. Capacitor starts discharging through RL.

Let peak value of input is V_{in} peak = 10 V.

In the positive half cycle the capacitor holds the positive peak value i.e. +10V. In the negative half cycle negative peak input is Vinpeak = -10V. Due to negative output diode D is reverse biased and acts as open circuit isolating op-amp output and capacitor C. Capacitor C has a charge of +10V from previous positive half cycle. This voltage is appeared to be as input to inverting terminal of op-amp.

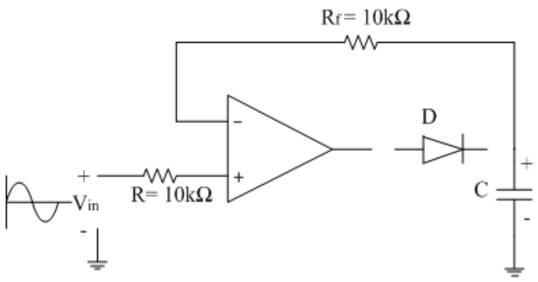


Fig.7.36 Input to inverting terminal of op-amp

Therefore differential input (V_{id}) to op-amp is,

V_{id}=-10-10=-20V=2×V(in peak)

For every op-amp there is a limit for maximum differential input voltage V_{id}. So care must be taken while selecting op-amp.

The load resistance RL is not possible to have a very large value always, so we use another op-amp as follows,

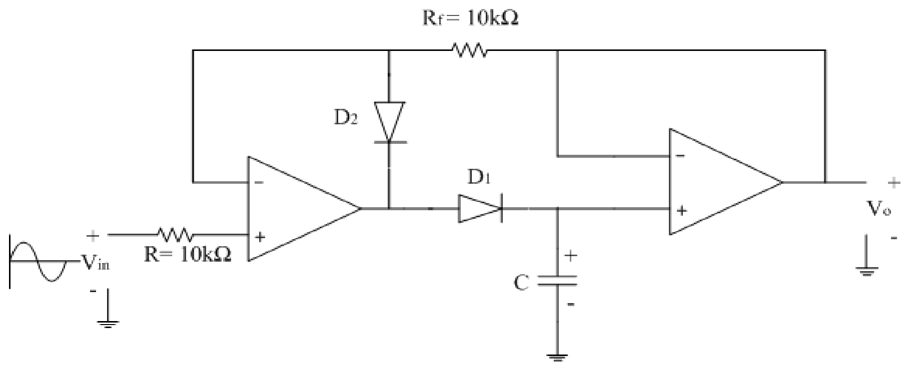


Fig.7.37 Second op-amp acts as a voltage follower

Here second op-amp acts as a voltage follower. Its input impedance is very high so capacitor discharges very slowly i.e. capacitor almost holding the charge. Therefore output voltage is nothing but voltage across capacitor (Peak value of input signal).

V_{out} = Voltage across capacitor (V_c)

As output impedance of voltage follower is very small we can connect any value of RL.

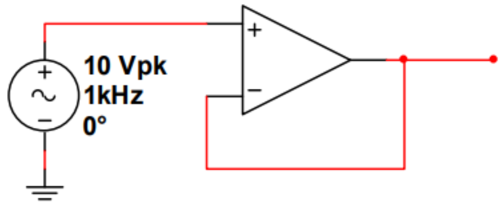
General Characteristics of Detectors are:

a) Sensitivity,
b) Detector Response,
c) Energy Resolution,
d) The Response Function,
e) Response Time,
f) Detector Efficiency,
g) Dead Time.

SAQ.2

a) What do you mean by Common mode rejection ratio?
b) Discuss about the output current and power consumption in op-amp.
c) What do you mean by slew rate in op-amp and write its equation?
d) What do you understand by the gain-band width product?
e) Write the Characteristics of OP- amplifier and comparator.
f) The differential amplifier has a common input signal of 3.20 V to both terminals. This results in an output signal of 26 mV. Determine the common-mode gain and the CMRR.

g) The circuit below utilizes a 741 op-amp with a slew-rate of 0.5V/µsec. The input signal has an amplitude of 10Vpk and a frequency of 1kHz. Will the output be slew-rate limited?



7.8 Inverting Operational Amplifier Configuration:

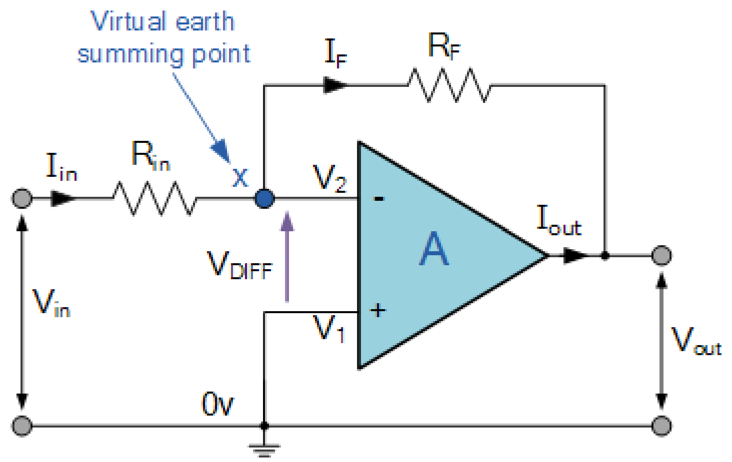


Fig.7.38 Inverting Amplifier circuit the operational amplifier

In this Inverting Amplifier circuit (Fig.7.38) the operational amplifier is connected with feedback to produce a closed loop operation. When dealing with operational amplifiers there are two very important rules to remember about inverting amplifiers, these are: "No current flows into the input terminal" and that "V_1 always equals V_2". However, in real world op-amp circuits both of these rules are slightly broken.

This is because the junction of the input and feedback signal (X) is at the same potential as the positive (+) input which is at zero volts or ground then, the junction is a "Virtual Earth". Because of this virtual earth node the input resistance of the amplifier is equal to the value of the input resistor, Rin and the closed loop gain of the inverting amplifier can be set by the ratio of the two external resistors.

We said above that there are two very important rules to remember about Inverting Amplifiers or any operational amplifier for that matter and these are.

- No Current Flows into the Input Terminals
- The Differential Input Voltage is Zero as $V_1 = V_2 = 0$ (Virtual Earth)

Then by using these two rules we can derive the equation for calculating the closed-loop gain of an inverting amplifier, using first principles.

Current (i) flows through the resistor network as shown in Fig.7.39.

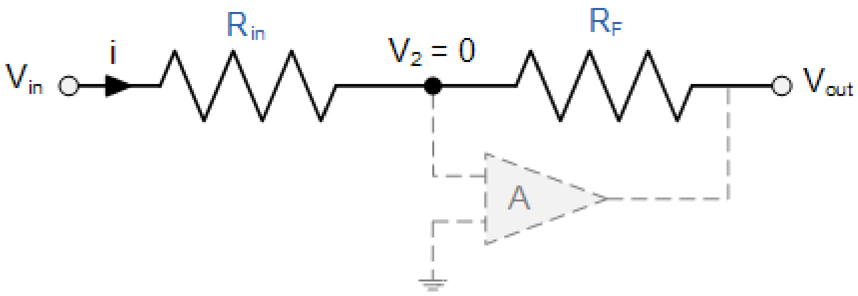


Fig.7.39 Current (i) flows through the resistor network

$$\mathrm{i} = \frac{\mathrm{Vin} - \mathrm{Vout}}{\mathrm{Rin} + \mathrm{Rout}}$$

Therefore,

$$\mathrm{i} = \frac{\mathrm{Vin} - \mathrm{V2}}{\mathrm{Rin}} = \frac{V2 - Vout}{Rf}$$

$$\mathrm{i} = \frac{\mathrm{Vin}}{\mathrm{Rin}} - \frac{\mathrm{V2}}{\mathrm{Rin}} = \frac{V2}{Rf} - \frac{Vout}{Rf}$$

So,

$$\frac{\mathrm{Vin}}{\mathrm{Rin}} = \mathrm{V2}[\frac{1}{\mathrm{Rin}} + \frac{1}{\mathrm{Rf}}] - \frac{Vout}{Rf}$$

and as,

$$\mathrm{i} = \frac{\mathrm{Vin} - 0}{\mathrm{Rin}} = \frac{0 - \mathrm{Vout}}{\mathrm{Rf}}$$

$$\frac{\mathrm{Rf}}{\mathrm{Rin}} = \frac{0 - \mathrm{Vout}}{\mathrm{Vin} - 0}$$

The Close Loop Gain (Av) is given as:

$$\frac{Vout}{Vin} = -\frac{Rf}{Rin}$$

Then, the Closed-Loop Voltage Gain of an Inverting Amplifier is given as:

$$\text{Gain } (Av) = \frac{V_{out}}{V_{in}} = -\frac{R_f}{R_{in}}$$

and this can be transposed to give V_{out} as:

$$V_{out} = -\frac{Rf}{Rin} \times V_{in}$$

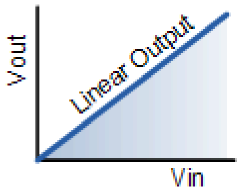


Fig.7.40 Linear Output

The negative sign in the equation indicates an inversion of the output signal with respect to the input as it is 180^o out of phase. This is due to the feedback being negative in value.

The equation for the output voltage V_{out} also shows that the circuit is linear in nature for a fixed amplifier gain as $V_{out} = V_{in}$ x Gain. This property can be very useful for converting a smaller sensor signal to a much larger voltage.

Advantages: The advantages of inverting amplifiers include the following.

- These are not expensive
- Its size is small
- Versatility
- Dependability
- Flexibility
- The two input terminals of this op-amp are zero always. In addition, simply the differential mode signal will exist.
- The device with inverting amplifier includes strong anti-interference capacity.
- It uses negative feedback.
- The gain factor is extremely high.
- The output of this op-amp will be out of phase through the input signal.

Disadvantages: The disadvantages of inverting amplifiers include the following**.**

- It has a small input impedance (equal to r1)
- It has high gain but the feedback must be maintained distortion less.
- The input signal should not include the noise because the small value will be multiplied & attained at the output.
- The signal is reversed.

Inverting Op-Amp Applications: The applications of inverting amplifiers include the following.

- An inverting amplifier can be used as a trans resistance amplifier which is also called a trans-impedance amplifier. This amplifier works as a current to voltage converter, used in less power-based applications.
- Inverting amplifier is used at the output stage when any system is designed with different types of sensors.
- This op-amp maintains the equal potential of voltage at two terminals, so it can be used in many fields.
- These op-amps are used in the mixers concept where the RF signals are present.
- It can be utilized as a phase shifter.
- These types of op-amps are used where the balancing of the signal is necessary.
- It is used practically in integration applications.
- The op-amp-based inverting circuits are more stable; distortion is fairly lower & provide a superior transitory response.
- Op-amps are used in every electronic device where linear ICs are used
- These are used in analog filtering and signal processing.
- These are used in various fields like communications, process control, displays, computers, measuring systems, power sources & signal sources.
- These are applicable in linear op-amp applications.

Non-inverting Operational Amplifier Configuration:

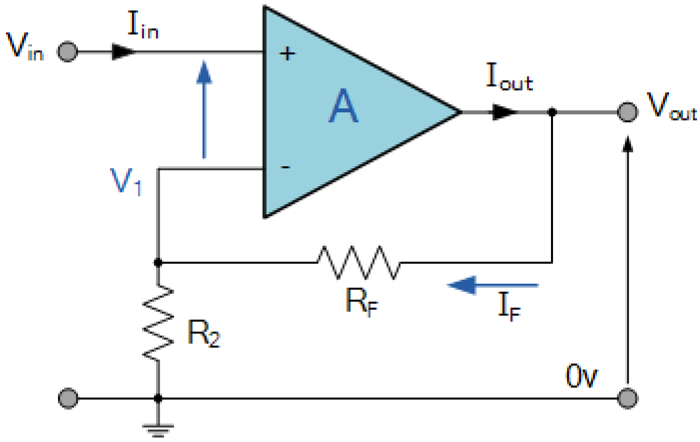


Fig.7.41 Non-inverting Operational Amplifier

In this configuration, the input voltage signal, (V_{in}) is applied directly to the non-inverting (+) input terminal which means that the output gain of the amplifier becomes "Positive" in value in contrast to the "Inverting Amplifier" circuit (Fig.7.41) we saw in the last tutorial whose output gain is negative in value. The result of this is that the output signal is "in-phase" with the input signal.

Feedback control of the non-inverting operational amplifier is achieved by applying a small part of the output voltage signal back to the inverting (–) input terminal via a Rf – R2 voltage divider network, again producing negative feedback. This closed-loop configuration produces a non-inverting amplifier circuit with very good stability, a very high input impedance, R_{in} approaching infinity, as no current flows into the positive input terminal, (ideal conditions) and a low output impedance, Rout as shown below.

we said that for an ideal op-amp "No current flows into the input terminal" of the amplifier and that "V1 always equals V2". This was because the junctions of the input and feedback signal (V1) are at the same potential.

In other words the junction is a "virtual earth" summing point. Because of this virtual earth node the resistors, Rf and R2 form a simple potential divider network across the non-inverting amplifier with the voltage gain of the circuit being determined by the ratios of R2 and Rf as shown below.

Equivalent Potential Divider Network:

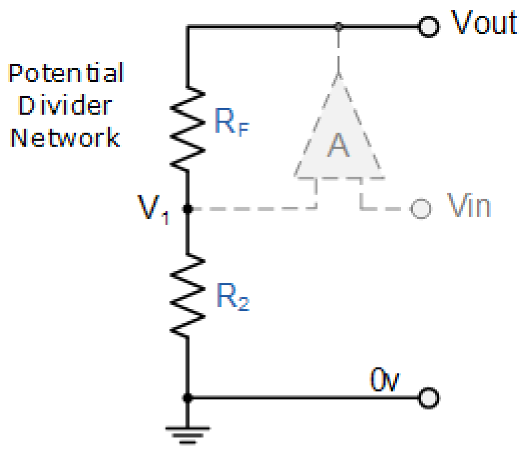


Fig.7.42 Equivalent Potential Divider Network

Then using the formula to calculate the output voltage of a potential divider network, we can calculate the closed-loop voltage gain (A_V) of the Non-inverting Amplifier as follows:

$$V_1 = \frac{R_2}{R_2 + R_F} \times V_{OUT}$$

Ideal Summing Point: $V_1 = V_{IN}$

Voltage Gain, $A_{(v)}$ is equal to: $\frac{V_{OUT}}{V_{IN}}$

Then,

$$A_{(V)} = \frac{V_{OUT}}{V_{IN}} = \frac{R_2 + R_F}{R_2}$$

Transpose to give:

$$A_{(V)} = \frac{V_{OUT}}{V_{IN}} = 1 + \frac{R_F}{R_2}$$

Then the closed loop voltage gain of a Non-inverting Operational Amplifier will be given as:

$$A_{(V)} = 1 + \frac{R_F}{R_2}$$

We can see from the equation above that the overall closed-loop gain of a non-inverting amplifier will always be greater but never less than one (unity), it is positive in nature and is determined by the ratio of the values of Rf and R2.

If the value of the feedback resistor Rf is zero, the gain of the amplifier will be exactly equal to one (unity). If resistor R2 is zero the gain will approach infinity, but in practice it will be limited to the operational amplifiers open-loop differential gain, (A_O).

We can easily convert an inverting operational amplifier configuration into a non-inverting amplifier configuration by simply changing the input connections as shown.

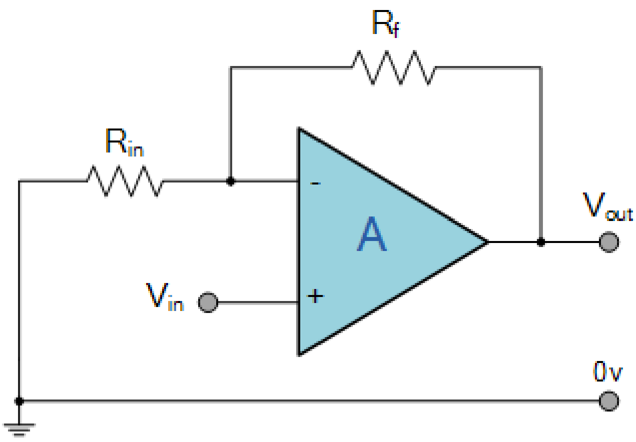


Fig.7.43 Non-inverting amplifier configuration by simply changing the input connections

Advantages: The advantages of non-inverting op-amp include the following.

- The output signal can be attained devoid of phase inversion.
- The voltage gain is changeable.
- The voltage gain is positive.
- Better matching of impedance can be obtained with the non-inverting amplifiers.

Disadvantages: The disadvantages of the non-inverting amplifier are as follows:

- More stages are utilized based on the requirement of achieving desired gain.
- Based on the respective amplifiers chosen the input and the output resistance gets varied.

Applications: The applications of non-inverting op-amp include the following.

- The non-inverting op-amp circuits are used where high input impedance is necessary.
- These circuits are used as a voltage follower by giving the output to the inverting input as an inverter.
- These are used to isolate the particular cascaded circuits.

7.9 The Differentiator Amplifier:

Here, the position of the capacitor and resistor have been reversed and now the reactance, XC is connected to the input terminal of the inverting amplifier while the resistor, Rf forms the negative feedback element across the operational amplifier as normal.

This operational amplifier circuit performs the mathematical operation of Differentiation that is it "produces a voltage output which is directly proportional to the input voltage's rate-of-change with respect to time". In other words the faster or larger the change to the input voltage signal, the greater the input current, the greater will be the output voltage change in response, becoming more of a "spike" in shape.

As with the integrator circuit, we have a resistor and capacitor forming an RC Network across the operational amplifier and the reactance (Xc) of the capacitor plays a major role in the performance of a Op-amp Differentiator.

Op-amp Differentiator Circuit:

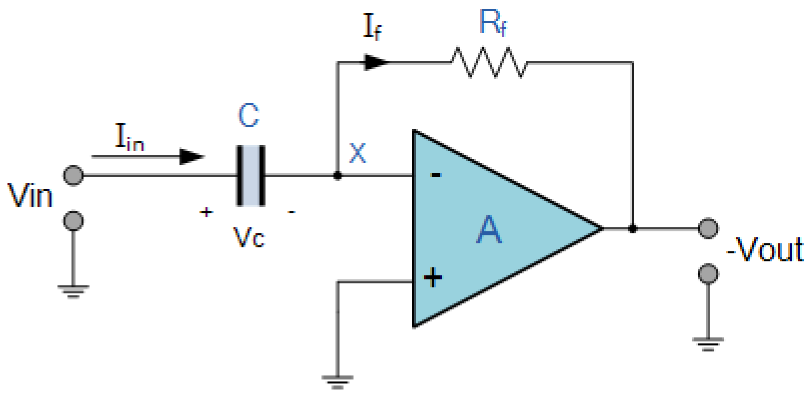


Fig.7.44 Op-amp Differentiator Circuit

The input signal to the differentiator is applied to the capacitor. The capacitor blocks any DC content so there is no current flow to the amplifier summing point, X resulting in zero output voltage. The capacitor only allows AC type input voltage changes to pass through and whose frequency is dependent on the rate of change of the input signal.

At low frequencies the reactance of the capacitor is "High" resulting in a low gain (Rf/Xc) and low output voltage from the op-amp. At higher frequencies the reactance of the capacitor is much lower resulting in a higher gain and higher output voltage from the differentiator amplifier.

However, at high frequencies an op-amp differentiator circuit becomes unstable and will start to oscillate. This is due mainly to the first-order effect, which determines the frequency response of the op-amp circuit causing a second-order response which, at high frequencies gives an output voltage far higher than what would be expected. To avoid this high frequency gain of the circuit needs to be reduced by adding an additional small value capacitor across the feedback resistor R_F.

Since the node voltage of the operational amplifier at its inverting input terminal is zero, the current, I flowing through the capacitor will be given as:

$$I_{IN} = I_F \text{ and } I_F = -\frac{V_{OUT}}{R_F}$$

The charge on the capacitor equals Capacitance times Voltage across the capacitor

$$Q = C \times V_{IN}$$

Thus the rate of change of this charge is:

$$\frac{dQ}{dt} = C\,\frac{dV_{IN}}{dt}$$

but dQ/dt is the capacitor current, I_{IN}

$$I_{IN} = C\frac{dV_{IN}}{dt} = I_F$$

$$-\frac{V_{OUT}}{R_F} = C\frac{dV_{IN}}{dt}$$

From which we have an ideal voltage output for the op-amp differentiator is given as:

$$V_{OUT} = -R_F\,C\frac{dV_{IN}}{dt}$$

Therefore, the output voltage V_{OUT} is a constant $-R_F*C$ times the derivative of the input voltage V_{IN} with respect to time. The minus sign (–) indicates a 180^o phase shift because the input signal is connected to the inverting input terminal of the operational amplifier.

One final point to mention, the Op-amp Differentiator circuit in its basic form has two main disadvantages compared to the previous operational amplifier integrator circuit. One is that it suffers from instability at high frequencies as mentioned above, and the other is that the capacitive input makes it very susceptible to random noise signals and any noise or harmonics present in the source circuit will be amplified more than the input signal itself. This is because the output is proportional to the slope of the input voltage so some means of limiting the bandwidth in order to achieve closed-loop stability is required.

Op-amp Differentiator Waveforms:

If we apply a constantly changing signal such as a Square-wave, Triangular or Sine-wave type signal to the input of a differentiator amplifier circuit the resultant output signal will be changed and whose final shape is dependent upon the RC time constant of the Resistor/Capacitor combination.

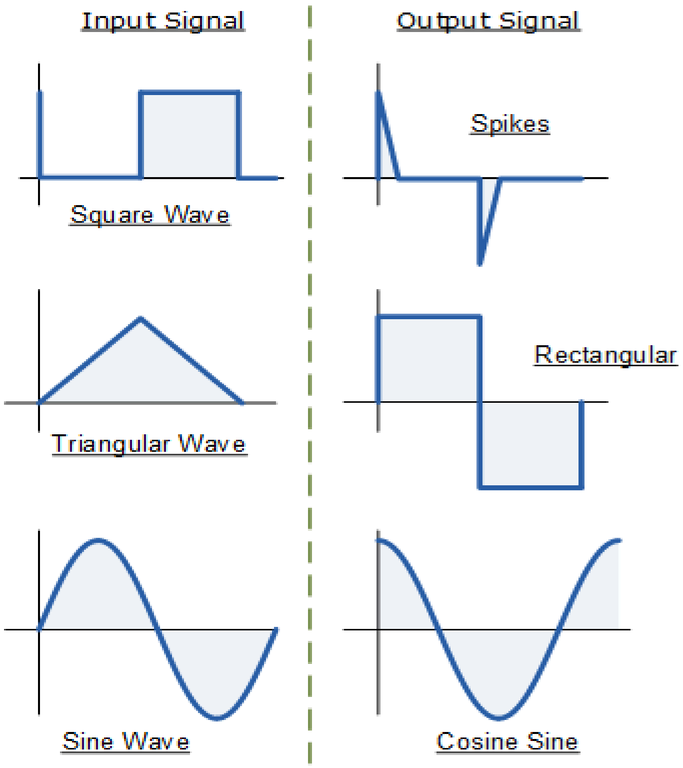


Fig.7.45 Waveforms of the Op-amp Differentiator

Improved Op-amp Differentiator Amplifier:

The basic single resistor and single capacitor op-amp differentiator circuit is not widely used to reform the mathematical function of Differentiation because of the two inherent faults mentioned above, "Instability" and "Noise". So in order to reduce the overall closed-loop gain of the circuit at high frequencies, an extra resistor, Rin is added to the input as shown below (Fig.7.46).

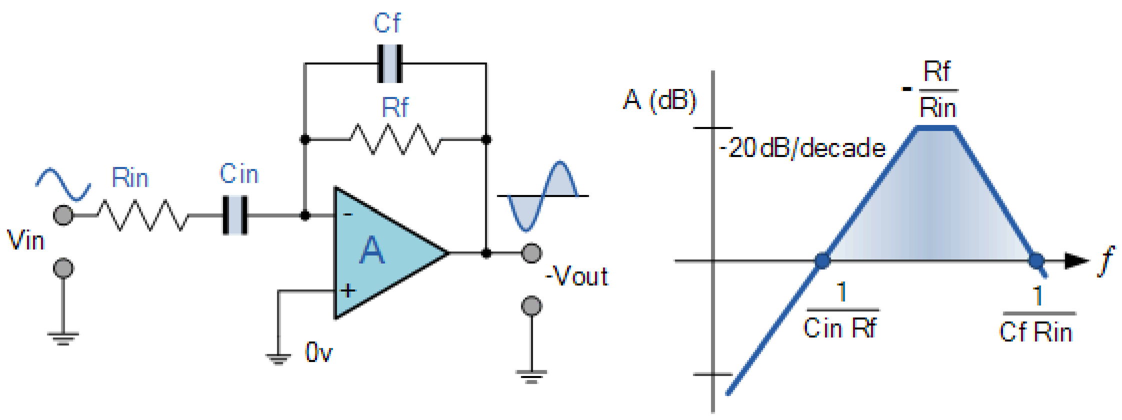


Fig.7.46 Improved Op-amp Differentiator Amplifier and output waveform at high frequency.

Adding the input resistor R_{IN} limits the differentiators increase in gain at a ratio of Rf/Rin. The circuit now acts like a differentiator amplifier at low frequencies and an amplifier with resistive feedback at high frequencies giving much better noise rejection.

Additional attenuation of higher frequencies is accomplished by connecting a capacitor Cf in parallel with the differentiator feedback resistor, Rf. This then forms the basis of a Active High Pass Filter as we have seen before in the filters section.

Advantages of differentiator circuit:

- The main advantage of such ideal differentiator is the small time constant required for differentiation.

Disadvantages of Differentiator:

- The gain of the differentiator increases as frequency increases. Thus at some high frequency, the differentiator may become unstable and break into oscillations. There is a possibility that Op-amp may go into saturation.
- Also, the input impedance decreases as frequency increases. This makes circuit very much sensitive to the noise.

Applications of Op-amp Differentiator:

- Differentiating amplifiers are most commonly designed to operate on triangular and rectangular signals.
- Differentiators also find application as wave shaping circuits, to detect high frequency components in the input signal.

The Integrator Amplifier: Operational amplifiers can be used as part of a positive or negative feedback amplifier or as an adder or subtractor type circuit using just pure resistances in both the input and the feedback loop.

But what if we where to change the purely resistive (R*f*) feedback element of an inverting amplifier with a frequency dependant complex element that has a reactance, (X), such as a Capacitor, C. What would be the effect on the op-amps voltage gain transfer function over its frequency range as a result of this complex impedance?

By replacing this feedback resistance with a capacitor we now have an RC Network connected across the operational amplifiers feedback path producing another type of operational amplifier circuit commonly called an Op-amp Integrator circuit as shown below.

Op-amp Integrator Circuit:

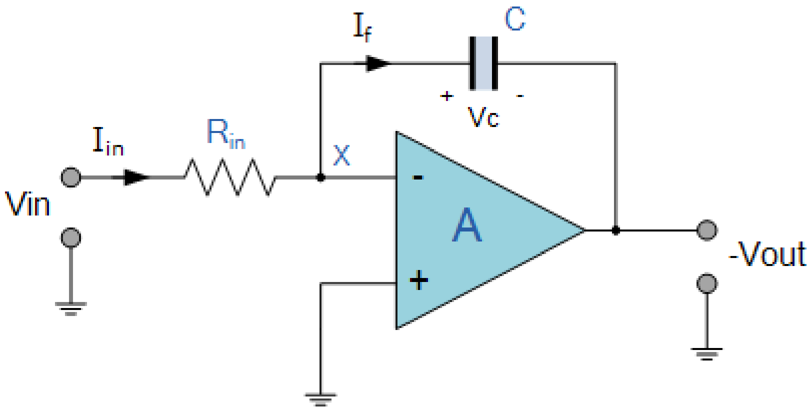


Fig.7.47 Op-amp Integrator circuit

As its name implies, the Op-amp Integrator is an operational amplifier circuit (Fig.7.47) that performs the mathematical operation of Integration that is we can cause the output to respond to changes in the input voltage over time as the op-amp integrator produces an output voltage which is proportional to the integral of the input voltage.

In other words the magnitude of the output signal is determined by the length of time a voltage is present at its input as the current through the feedback loop charges or discharges the capacitor as the required negative feedback occurs through the capacitor.

When a step voltage, V_{in} is firstly applied to the input of an integrating amplifier, the uncharged capacitor C has very little resistance and acts a bit like a short circuit allowing maximum current to flow via the input resistor, R_{in} as potential difference exists between the two plates. No current flows into the amplifiers input and point X is a virtual earth resulting in zero output. As the impedance of the capacitor at this point is very low, the gain ratio of X_C/R_{in} is also very small giving an overall voltage gain of less than one, (voltage follower circuit).

As the feedback capacitor, C begins to charge up due to the influence of the input voltage, its impedance Xc slowly increase in proportion to its rate of charge. The capacitor charges up at a rate determined by the RC time constant, (τ) of the series RC network. Negative feedback forces the op-amp to produce an output voltage that maintains a virtual earth at the op-amp's inverting input.

Since the capacitor is connected between the op-amp's inverting input (which is at virtual ground potential) and the op-amp's output (which is now negative), the potential voltage, Vc developed across the capacitor slowly increases causing the charging current to decrease as the impedance of the capacitor increases. This results in the ratio of X_c/R_{in} increasing producing a linearly increasing ramp output voltage that continues to increase until the capacitor is fully charged.

At this point the capacitor acts as an open circuit, blocking any more flow of DC current. The ratio of feedback capacitor to input resistor (X_C/R_{in}) is now infinite resulting in infinite gain. The result of this high gain (similar to the op-amps open-loop gain), is that the output of the amplifier goes into saturation as shown below. (Saturation occurs when the output voltage of the amplifier swings heavily to one voltage supply rail or the other with little or no control in between).

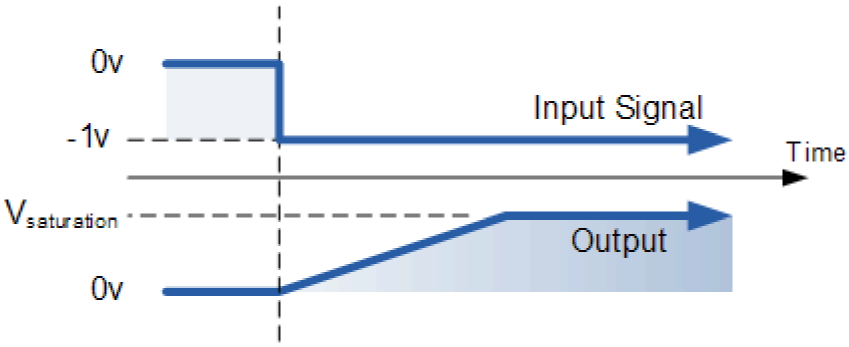


Fig.7.48 Voltage saturation output waveform

The rate at which the output voltage increases (the rate of change) is determined by the value of the resistor and the capacitor, "RC time constant". By changing this RC time constant value, either by changing the value of the Capacitor, C or the Resistor, R, the time in which it takes the output voltage to reach saturation can also be changed for example.

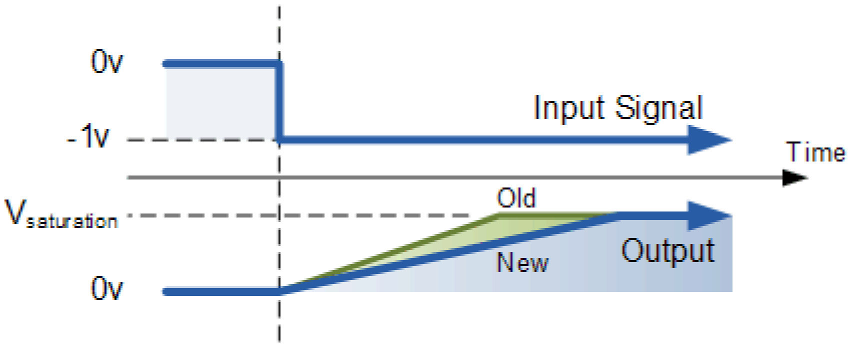


Fig.7.49 Voltage increases to reach saturation output waveform

If we apply a constantly changing input signal such as a square wave to the input of an Integrator Amplifier then the capacitor will charge and discharge in response to changes in the input signal? This results in the output signal being that of a saw tooth waveform whose output is affected by the RC time constant of the resistor/capacitor combination because at higher frequencies, the capacitor has less time to fully charge. This type of circuit is also known as a Ramp Generator and the transfer function is given below.

Op-amp Integrator Ramp Generator:

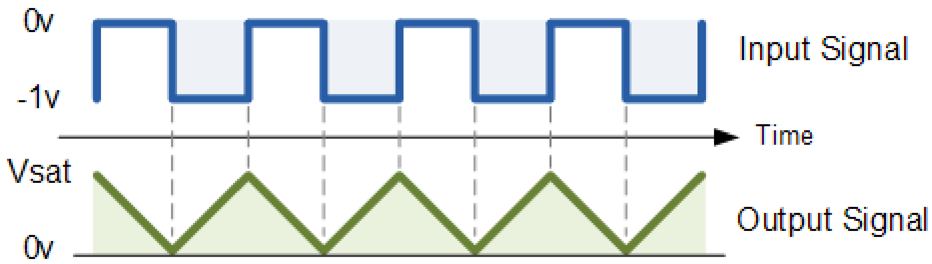


Fig.7.50 Output waveform of the Op-amp Integrator Ramp Generator

We know from first principals that the voltage on the plates of a capacitor is equal to the charge on the capacitor divided by its capacitance giving Q/C. Then the voltage across the capacitor is output V_{out} therefore: $-V_{out} = Q/C$. If the capacitor is charging and discharging, the rate of charge of voltage across the capacitor is given as:

$$V_c = \frac{Q}{C}, \quad V_c = V_x - V_{out} = 0 - V_{out}$$

$$-\frac{dV_{out}}{dt} = \frac{dQ}{Cdt} = \frac{1}{C}\frac{dQ}{dt}$$

But dQ/dt is electric current and since the node voltage of the integrating op-amp at its inverting input terminal is zero, X = 0, the input current I(in) flowing through the input resistor, R_{in} is given as:

$$I_{in} = \frac{V_{in} - 0}{R_{in}} = \frac{V_{in}}{R_{in}}$$

The current flowing through the feedback capacitor C is given as:

$$I_f = C\frac{dV_{out}}{dt} = C\frac{dQ}{Cdt} = \frac{dQ}{dt} = \frac{dV_{out}.C}{dt}$$

Assuming that the input impedance of the op-amp is infinite (ideal op-amp), no current flows into the op-amp terminal. Therefore, the nodal equation at the inverting input terminal is given as:

$$I_{in} = I_f = \frac{V_{in}}{R_{in}} = \frac{dV_{out}.C}{dt}$$

$$\frac{V_{in}}{V_{out}} \times \frac{dt}{R_{in}C} = 1$$

From which we derive an ideal voltage output for the Op-amp Integrator as:

$$V_{out} = -\frac{1}{R_{in}C}\int_0^t V_{in}\, dt = -\int_0^t V_{in}\frac{dt}{R_{in}.C}$$

To simplify the math's a little, this can also be re-written as:

$$V_{out} = -\frac{1}{j\omega RC}\, V_{in}$$

Where: $\omega = 2\pi f$ and the output voltage V_{out} is a constant 1/RC times the integral of the input voltage V_{in} with respect to time.

Thus the circuit has the transfer function of an inverting integrator with the gain constant of -1/RC. The minus sign (–) indicates a 180^o phase shift because the input signal is connected directly to the inverting input terminal of the operational amplifier.

The AC or Continuous Op-amp Integrator:

If we changed the above square wave input signal to that of a sine wave of varying frequency the Op-amp Integrator performs less like an integrator and begins to behave more like an active "Low Pass Filter", passing low frequency signals while attenuating the high frequencies.

At zero frequency (0Hz) or DC, the capacitor acts like an open circuit due to its reactance thus blocking any output voltage feedback. As a result very little negative feedback is provided from the output back to the input of the amplifier.

Therefore with just a single capacitor, C in the feedback path, at zero frequency the op-amp is effectively connected as a normal open-loop amplifier with very high open-loop gain. This results in the op-amp becoming unstable cause undesirable output voltage conditions and possible voltage rail saturation.

This circuit connects a high value resistance in parallel with a continuously charging and discharging capacitor. The addition of this feedback resistor, R_2 across the capacitor, C gives the circuit the characteristics of an inverting amplifier with finite closed-loop voltage gain given by: R_2/R_1.

The result is at high frequencies the capacitor shorts out this feedback resistor, R_2 due to the effects of capacitive reactance reducing the amplifiers gain. At normal operating frequencies the circuit acts as an standard integrator, while at very low frequencies approaching 0Hz, when C becomes open-circuited due to its reactance, the magnitude of the voltage gain is limited and controlled by the ratio of: R_2/R_1.

The AC Op-amp Integrator with DC Gain Control:

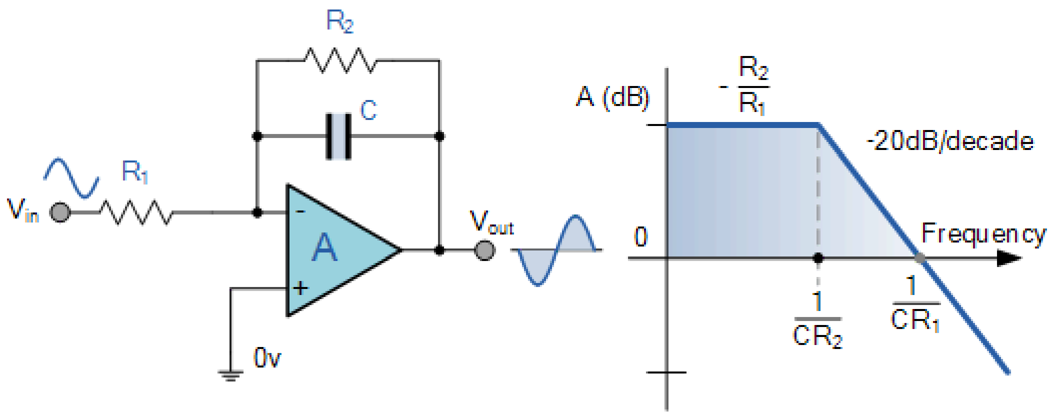


Fig.7.51 AC Op-amp Integrator with DC Gain Control and output waveform

Unlike the DC integrator amplifier above whose output voltage at any instant will be the integral of a waveform so that when the input is a square wave, the output waveform will be triangular. For an AC integrator, a sinusoidal input waveform will produce another sine wave as its output which will be 90^o out-of-phase with the input producing a cosine wave.

Furthermore, when the input is triangular, the output waveform is also sinusoidal. This then forms the basis of a Active Low Pass Filter as seen before in the filters section tutorials with a corner frequency given as.

D.C. Voltage Gain, $(Av_0) = -\frac{R_2}{R_1}$

A.C. Voltage Gain, $(Av) = -\frac{R_2}{R_1} \times \frac{1}{1+2\pi fCR_2}$

Corner Frequency, $(f_0) = \frac{1}{2\pi CR_2}$

In another type of operational amplifier circuit which is the opposite or complement of the Op-amp Integrator circuit above called the Differentiator Amplifier.

As its name implies, the differentiator amplifier produces an output signal which is the mathematical operation of differentiation that is it produces a voltage output which is proportional to the input voltage's rate-of-change and the current flowing through the input capacitor.

Advantages of op amp integrator: The advantages of op amp integrator include the following.

- Op-amp circuits can be made to perform both integration and differentiation.
- The circuit can be modified to give various applications.
- Op-amp integrators produce an output voltage that is almost 100 times greater than the input voltage.

Disadvantages of integrator circuit:

- Integrated circuits are not flexible.
- It is impossible to fabricate transformers.
- The IC will not work properly if wrongly handled or it must be exposed to excessive heat.
- The power that integrated circuits can produce is limited and calls for extension.

Op-amp Integrator Applications:

- Integrator is an important part of the instrumentation and is used in Ramp generation.
- In function generator, the integrator circuit is used to produce the triangular wave.
- Integrator is used in wave shaping circuit such as a different kind of charge amplifier.
- It is used in analog computers, where integration is needed to be done using the analog circuit.
- Integrator circuit is also widely used in analog to the digital converter.

- Different sensors also use an integrator to reproduce useful outputs.

SAQ.3

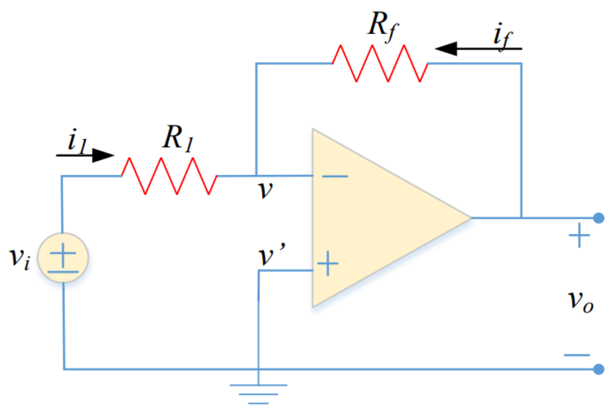
a) What do you mean by inverting amplifier and also write the expression of the output voltage?

b) Discuss about the non inverting amplifier. Also write the expression for the closed loop voltage gain.

c) What do you understand by the Differentiator amplifier? Also write the expression for the output voltage.

d) Discuss and write about the basic integrator in brief. Also write the expression for voltage output of the Op-amp Integrator.

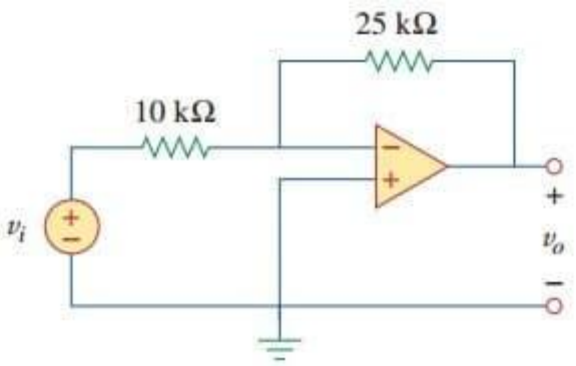
e) Design an inverting amplifier with a gain of 10 and an input impedance of 15 kΩ.



f) A differential amplifier has an open-circuit voltage gain of 100. The input signals are 3.25 and 3.15 V. Determine the output voltage.

Examples:

Q.1 Refer to the op amp in Figure. If $v_i = 0.5$ V, calculate: (a) the output voltage v_o, and (b) the current in the 10-kΩ resistor.



Solution:

(a) Using Equation.(2),

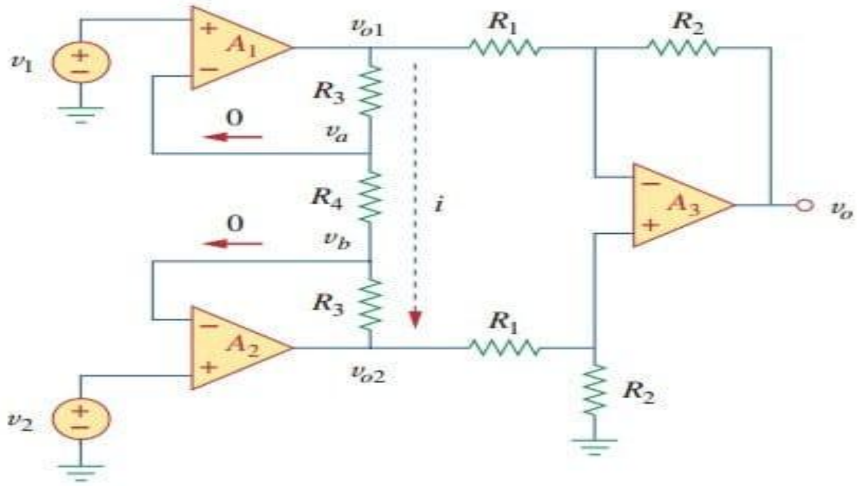
$$\frac{v_o}{v_i} = -\frac{R_f}{R_1} = -\frac{25}{10} = -2.5$$

$$v_o = -2.5v_i = -2.5(0.5) = -1.25 \text{ V}$$

(b) The current through the 10-kΩ resistor is

$$i = \frac{v_i - 0}{R_1} = \frac{0.5 - 0}{10 \times 10^3} = 50\ \mu\text{A}$$

Q.2 An instrumentation amplifier given below in Figure is an amplifier of low-level signals used in process control or measurement applications and commercially available in single-package units.



Show that

$$v_o = \frac{R_2}{R_1}\left(1 + \frac{2R_3}{R_4}\right)(v_2 - v_1)$$

Solution:

We recognize that the amplifier A_3 in Figure is a difference amplifier. Thus, from Equation.(5),

$$v_o = \frac{R_2}{R_1}(v_{o2} - v_{o1}) \qquad (2.1)$$

Since the op amps A_1 and A_2 draw no current, current i flow through the three resistors as though they were in series. Hence,

$$v_{o1} - v_{o2} = i(R_3 + R_4 + R_3) = i(2R_3 + R_4) \quad (2.2)$$

But

$$i = \frac{v_a - v_b}{R_4}$$

and $v_a = v_1$, $v_b = v_2$. Therefore,

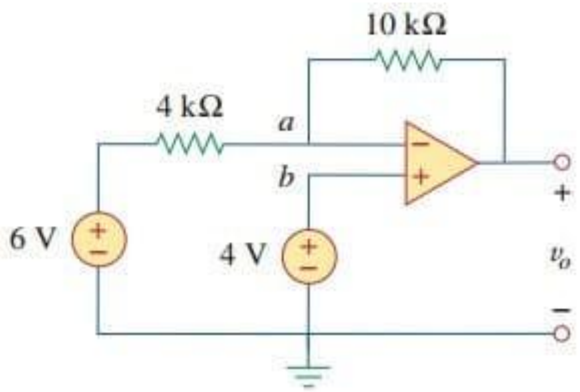
$$i = \frac{v_1 - v_2}{R_4} \quad (2.3)$$

Inserting Equations.(2.2) and (2.3) into (2.1) gives

$$v_o = \frac{R_2}{R_1}\left(1 + \frac{2R_3}{R_4}\right)(v_2 - v_1)$$

as required.

Q.3 For the op amp circuit in Figure, calculate the output voltage v_o.



Solution:

We may solve this in two ways: using superposition and using nodal analysis.

■ METHOD 1: Using superposition, we let

$$v_o = v_{o1} + v_{o2}$$

where v_{o1} is due to the 6-V voltage source, and v_{o2} is due to the 4-V input. To get v_{o1}, we set the 4-V source equal to zero. Under this condition, the circuit becomes an inverter. Hence the Equation.(2) in 'Inverting Op Amp' gives

$$v_{o1} = -\frac{10}{4}(6) = -15 \text{ V}$$

To get v_{o2}, we set the 6 -V source equal to zero. The circuit becomes a non-inverting amplifier so that Equation.(3) applies

$$v_{o2} = \left(1 + \frac{10}{4}\right)4 = 14 \text{ V}$$

Thus,

$$v_o = v_{o1} + v_{o2} = -15 + 14 = -1 \text{ V}$$

■ METHOD 2: Applying KCL at node *a*,

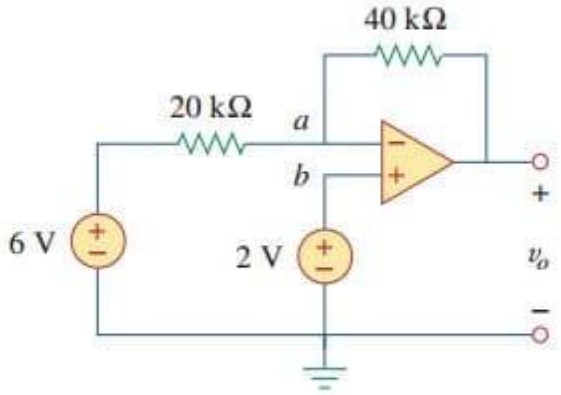
$$\frac{6 - v_a}{4} = \frac{v_a - v_o}{10}$$

But $v_a = v_b = 4$, and so

$$\frac{6 - 4}{4} = \frac{4 - v_o}{10} \quad \Rightarrow \quad 5 = 4 - v_o$$

or $v_o = -1$ V, as before.

Q.4 Determine v_o in the op amp circuit shown in Figure.



Solution:

Applying KCL at node *a*,

$$\frac{v_a - v_o}{40\text{ k}\Omega} = \frac{6 - v_a}{20\text{ k}\Omega}$$

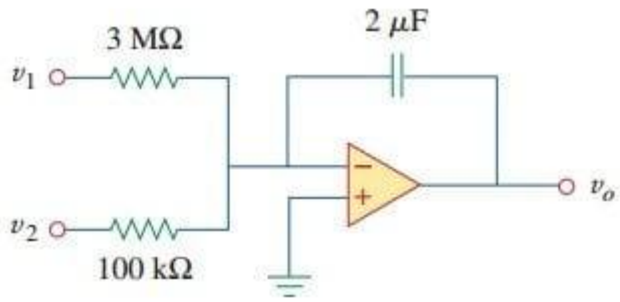
$$v_a - v_o = 12 - 2v_a \quad \Rightarrow \quad v_o = 3v_a - 12$$

But $v_a = v_b = 2$ V for an ideal op amp, because of the zero voltage drop across the input terminals of the op amp. Hence,

$$v_o = 6 - 12 = -6\text{ V}$$

Notice that if $v_b = 0 = v_a$, then $v_o = -12$, as expected from Equation (2).

Q.5 If $v_1 = 10 \cos 2t$ mV and $v_2 = 0.5t$ mV, find v_o in the op amp circuit in Figure. Assume that the voltage across the capacitor is initially zero.

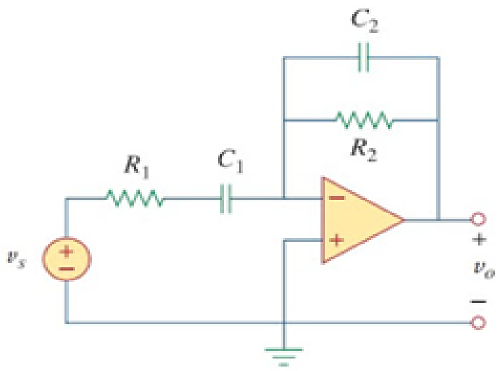


Solution:

This is a summing integrator, and

$$\begin{aligned} v_o &= -\frac{1}{R_1C}\int v_1\,dt - \frac{1}{R_2C}\int v_2\,dt \\ &= -\frac{1}{3\times 10^6 \times 2\times 10^{-6}}\int_0^t 10\cos(2\tau)d\tau \\ &\quad -\frac{1}{100\times 10^3\times 2\times 10^{-6}}\int_0^t 0.5\tau d\tau \\ &= -\frac{1}{6}\frac{10}{2}\sin 2t - \frac{1}{0.2}\frac{0.5t^2}{2} = -0.833\sin 2t - 1.25t^2\text{ mV} \end{aligned}$$

Q.6 Compute the closed-loop gain and phase shift for the circuit in Figure.(2). Assume that $R_1 = R_2 = 10$ kΩ, $C_1 = 2$ uF, $C_2 = 1$ uF, and $\omega = 200$ rad/s.



Solution:

The feedback and input impedances are calculated as

$$\mathbf{Z}_f = R_2 \left\| \frac{1}{j\omega C_2} = \frac{R_2}{1 + j\omega R_2 C_2} \right.$$

$$\mathbf{Z}_i = R_1 + \frac{1}{j\omega C_1} = \frac{1 + j\omega R_1 C_1}{j\omega C_1}$$

Since the circuit in Figure,is an inverting amplifier, the closed-loop gain is given by

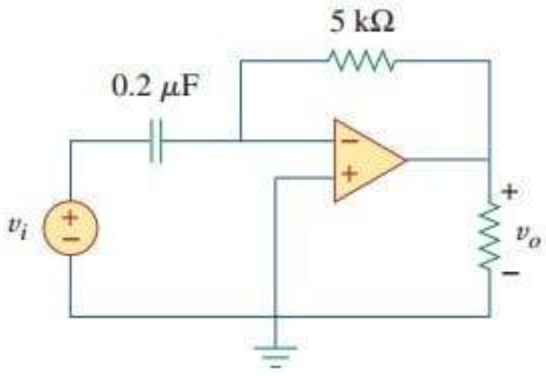
$$\mathbf{G} = \frac{\mathbf{V}_o}{\mathbf{V}_s} = -\frac{\mathbf{Z}_f}{\mathbf{Z}_i} = \frac{-j\omega C_1 R_2}{(1 + j\omega R_1 C_1)(1 + j\omega R_2 C_2)}$$

Substituting the given values of R_1, R_2, C_1, C_2, and ω, we get

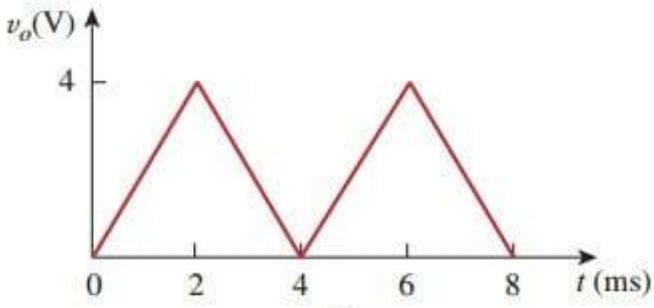
$$\mathbf{G} = \frac{-j4}{(1 + j4)(1 + j2)} = 0.434\angle 130.6^\circ$$

Hence, the closed-loop gain is 0.434 and the phase shift is 130.6^o.

Q.7 Sketch the output voltage for the circuit in Figure (a), given the input voltage in Figure (b). Take $v_o = 0$ at t = 0.



(a)



(b)

Solution:

This is a differentiator with

$$RC = 5 \times 10^3 \times 0.2 \times 10^{-6} = 10^{-3} \text{ s}$$

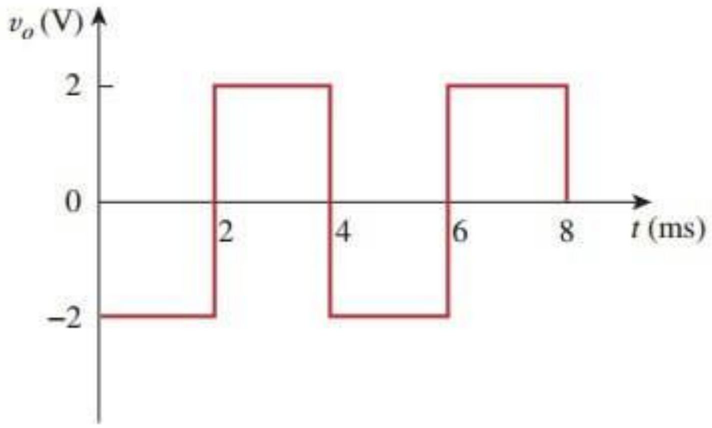
For 0 < t < 4 ms, we can express the input voltage in Figure (b) as

$$v_i = \begin{cases} 2000t & 0 < t < 2 \text{ ms} \\ 8 - 2000t & 2 < t < 4 \text{ ms} \end{cases}$$

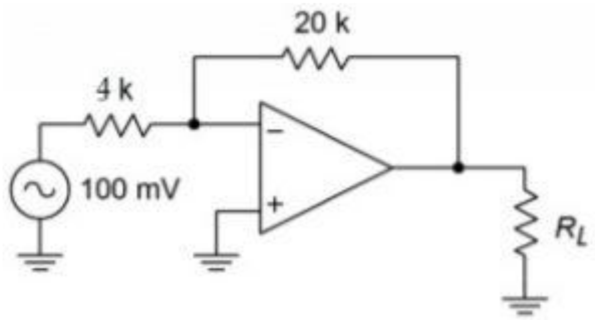
This is repeated for 4 < t < 8 ms. Using Equation, the output is obtained as

$$v_o = -RC\frac{dv_i}{dt} = \begin{cases} -2 \text{ V} & 0 < t < 2 \text{ ms} \\ 2 \text{ V} & 2 < t < 4 \text{ ms} \end{cases}$$

Thus, the output is as sketched in Figure(c).



Q.8 For the following inverting amplifier circuit, calculate the input impedance and output voltage.



Solution: The input impedance is set through input resistance Ri which is 4kΩ. So Zin=4kΩ.

Vout=Vin Av

Av = −Rf/Ri

Av=−20k/4k

Av=−5

Vout=100mV * (−5)

Vout = −500mV

Q.9 An inverting amplifier including a gain = 8 & 10 k Ω of an input impedance. The input impedance (Zi) tells us what 'Ri' must be?

Solution:

Zin = Ri

Ri=10k

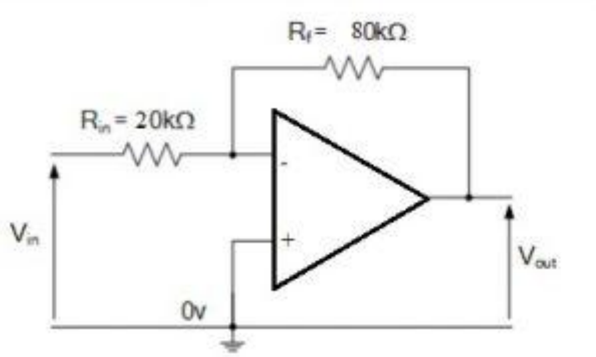
So Rf =?

We know that, Av =−Rf/Ri

$$Rf = 10\ (-8)$$

$$Rf = 80k$$

Q.10 For the following inverting amplifier circuit, please calculate the closed-loop gain.



Solution: The gain formula for the above circuit is

$$\text{Gain (Av)} = \text{Vout/Vin} = \text{-Rf/Rin}$$

Now we have to substitute the above-given values within the circuit are

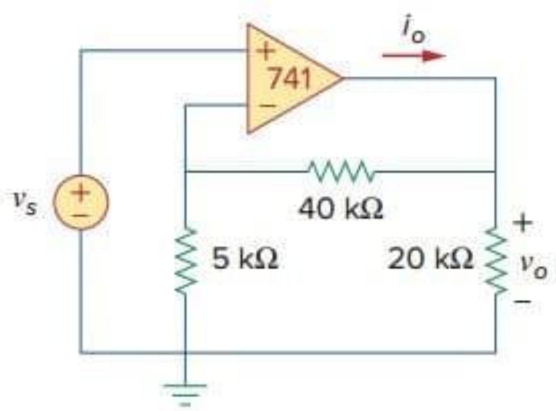
$$\text{Rin} = 20\text{k}\Omega \text{ and } \text{R}f = 80\text{k}\Omega$$

The circuit gain can be measured as Av = -Rf/Rin = -80k/20k = -4

So, for inverting amplifier circuit, the closed-loop gain is -4.

Q.11 If the same 741 op amp is used in the circuit of Figure (a) below, calculate the closed-loop gain v_o/v_s.

Find i_o when v_s = 1 V. Rework it using the ideal op amp model



Figure(a)

Solution:

We just need to keep Equations.(1) and (3) in mind as we analyze the circuit in Figure (a). Thus, the Figure (a) the circuit is presented as in Figure (b).

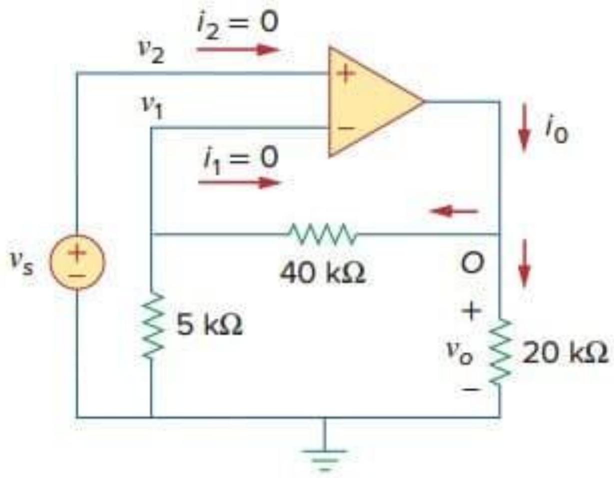


Figure (b)

Notice that

$$v_2 = v_s \quad \text{…………….(1)}$$

Since i_1 = 0, the 40 – and 5 -kΩ resistors are in series; the same current flows through them. v_1 is the voltage across the 5 -kΩ resistor. Hence, using the voltage division principle,

$$v_1 = \frac{5}{5+40}v_o = \frac{v_o}{9} \quad \text{……………..(2)}$$

According to Equation.(3),

$$v_2 = v_1 \quad \text{………………(3)}$$

Substituting Equations (1) and (2) into (3) yields the closed-loop gain,

$$v_s = \frac{v_o}{9} \quad \Rightarrow \quad \frac{v_o}{v_s} = 9 \quad \text{…………………..(4)}$$

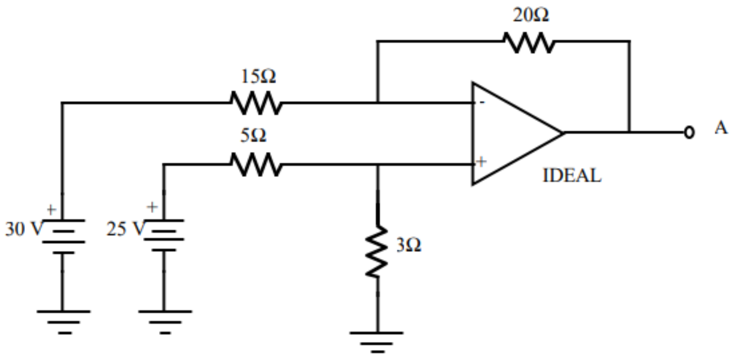
At node *O*,

$$i_o = \frac{v_o}{40+5} + \frac{v_o}{20}\text{mA} \quad \text{………………….(5)}$$

From Equation (1.4), when v_s = 1 V, v_o = 9 V. Substituting for v_o = 9 V in Equation (5) produces

$$i_o = 0.2 + 0.45 = 0.65 \text{ mA}$$

Q.12 For the difference amplifier circuit shown, determine the output voltage at terminal A.



Solution:

By voltage division,

$$v_{in+} = 25V\left(\frac{3\Omega}{5\Omega + 3\Omega}\right) = 9.375V$$

By the virtual short circuit between the input terminals, $v_{in-} = 9.375$ V
Using **Ohm's** law, the current through the 15 Ω resistor is

$$I_{15} = \left(\frac{30V - 9.375V}{15\Omega}\right) = 1.375V$$

The input impedance is infinite; therefore, $I_{in-}=0$ and $I_{15}=I_{20}$.

Use Kirchoff's voltage law to find the output voltage at A.

$v_A = v_{in-} - 20I_{20} = 9.375\text{ V} - (20\Omega)(1.375\text{ A}) = -18.125\text{ V}$

Q.13 An operational amplifier is required to amplify a signal with a peak voltage of 5 volts at a frequency of 20kHz. Find out a slew rate.

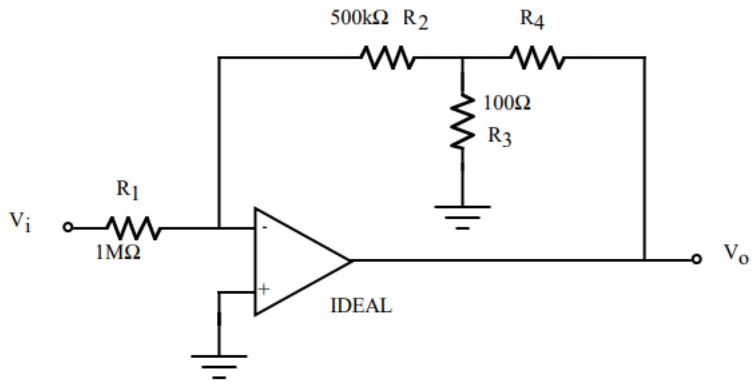
Solution:

Given data:

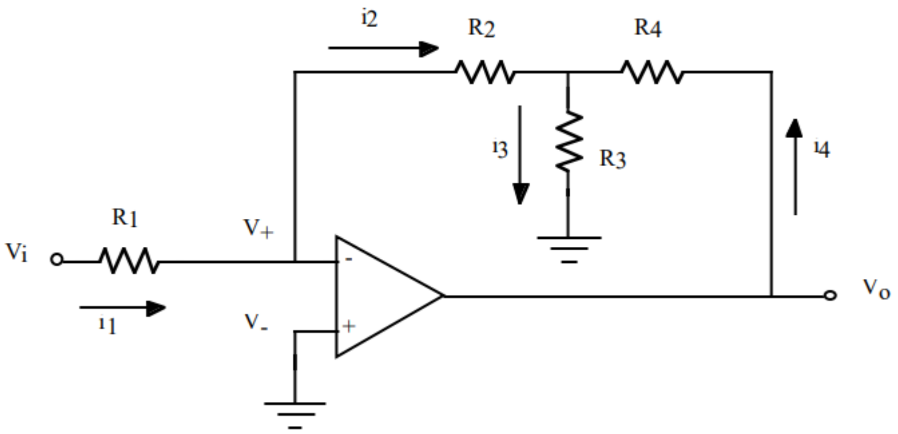
$$V_m = 5volts,\ f_m = 20kHz$$

$$\begin{aligned} S &= 2\pi f_m V_m \\ &= 2 * 3.14 * 20 * 10^3 * 5 \\ &= 628 * 10^3 \\ S &= 628000V/S\ or\ 0.628V/\mu S \end{aligned}$$

Q.14 Evaluate the following amplifier circuit to determine the value of resistor R_4 in order to obtain a voltage gain (v_o/v_i) of -120.



Solution:



v_{in+} is grounded, so v_{in-} is also a virtual ground.

$v_{in-} = 0$

Since $v_{in-} = 0$, $v_i=i_1R_1$ and $i_1=v_i/R_1$.

Since $v_{in-} = 0$, $v_x=-i_2R_2$ and $i_2=-v_x/R_2$.

Similarly,
$v_x=-i_3R_3$
$v_x-v_o =-i_4R_4$

From Kirchhoff's current law,

$i_4 = i_2 + i_3$

$$\frac{v_x - v_o}{R_4} = \frac{-v_x}{R_2} + \frac{-v_x}{R_3}$$

Now, v_o=-120v_i.

Also, i_1=i_2, so

$$\frac{v_i}{R_1} = \frac{-v_x}{R_2}$$

$$v_x = -\left(\frac{R_2}{R_1}\right)v_i$$

$$\frac{-\left(\frac{R_2}{R_1}\right)v_i - (-120v_i)}{R_4} = \frac{\left(\frac{R_2}{R_1}\right)v_i}{R_2} + \frac{\left(\frac{R_2}{R_1}\right)v_i}{R_3}$$

$$\frac{120\left(\frac{R_1}{R_2}\right) - 1}{R_4} = \frac{1}{R_2} + \frac{1}{R_3} = \frac{R_2 + R_3}{R_2 R_3}$$

$$R_4 = \frac{120\left(\frac{R_1}{R_2}\right) - 1}{\frac{R_2 + R_3}{R_2 R_3}} = \frac{120\left(\frac{1x10^6\Omega}{5x10^5\Omega}\right) - 1}{\frac{5x10^5\Omega + 100\Omega}{(5x10^5\Omega)(100\Omega)}} = 2.39x10^4\Omega \text{ (24 k}\Omega)$$

Q.15 Design an op amp circuit with inputs v_1 and v_2 such that $v_o = -5v_1 + 3v_2$.

Solution:

The circuit requires that

$$v_o = 3v_2 - 5v_1 \quad \text{……………..(1)}$$

This circuit can be realized in two ways.

<u>Design 1</u> – If we desire to use only one op amp, we can use the op amp circuit of Figure (a).

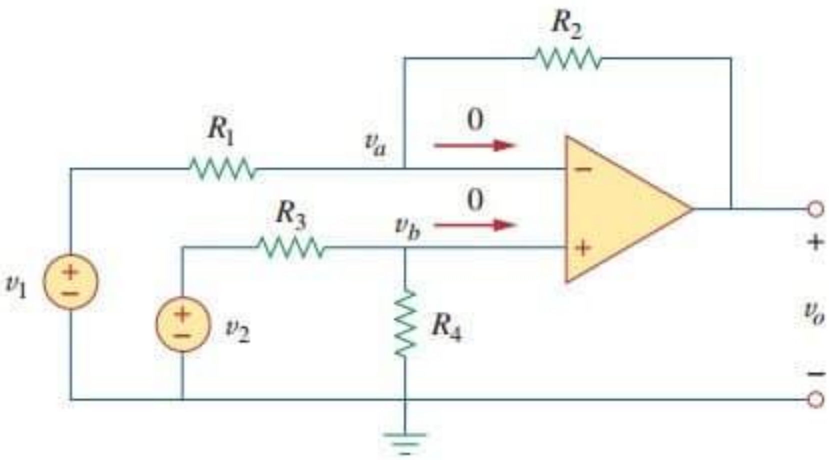


Figure (a) The difference amplifier

Comparing Equation (1) with Equation (3), we see

$$\frac{R_2}{R_1} = 5 \qquad \Rightarrow \qquad R_2 = 5R_1$$

……………(2)

Also,

$$5\frac{(1 + R_1/R_2)}{(1 + R_3/R_4)} = 3 \qquad \Rightarrow \qquad \frac{\frac{6}{5}}{1 + R_3/R_4} = \frac{3}{5}$$

$$2 = 1 + \frac{R_3}{R_4} \qquad \Rightarrow \qquad R_3 = R_4$$

……………..(3)

If we choose $R_1 = 10$ k Ω and $R_3 = 20$ k Ω, then $R_2 = 50$ k Ω and $R_4 = 20$ kΩ.

<u>Design 2</u> – If we desire to use more than one op amp, we may cascade an inverting amplifier and a two-input inverting summer, as shown in Figure (b).

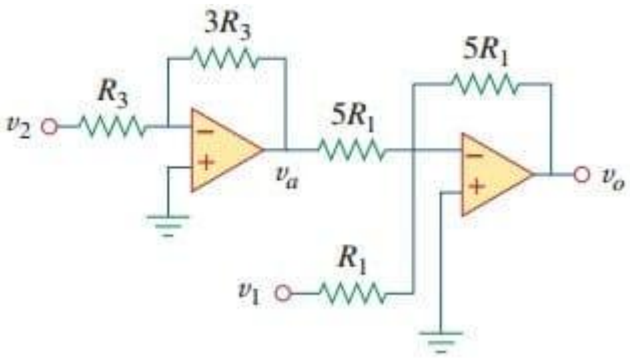


Figure (b)

For the summer,

$$v_o = -v_a - 5v_1 \quad \ldots\ldots\ldots\ldots\ldots(4)$$

and for the inverter,

$$v_a = -3v_2 \quad \ldots\ldots\ldots\ldots (5)$$

Combining Equations. (4) and (5) gives

$$v_o = 3v_2 - 5v_1$$

Which is the desired result, in Figure (b), we may select $R_1 = 10$ kΩ and $R_3 = 20$ kΩ or $R_1 = R_3 = 10$ kΩ.

7.10 Summary:

1. Operational amplifiers are linear devices that have all the properties required for nearly ideal DC amplification and are therefore used extensively in signal conditioning, filtering or to perform mathematical operations such as add, subtract, integration and differentiation.
2. Op-amps use a DC supply voltage, typically anywhere from a few volts on up to 30 V or more. If the power supply is a perfect DC voltage source (that is, it gives the same voltage no matter what happens), the op-amp's output would be solely governed by its inputs.
3. The characteristics of an ideal op-amp are as follows: Infinite bandwidth due to the ideal gain inside of the op-amp. Infinite open-loop gain A. Infinite or zero common-mode gain.
4. As the input signal is connected directly to the non-inverting input of the amplifier the output signal is not inverted resulting in the output voltage being equal to the input voltage, thus $V_{out} = V_{in}$.
5. The input offset voltage is defined as the voltage that must be applied between the two input terminals of the op amp to obtain zero volts at the output. Ideally the output of the op amp should be at zero volts when the inputs are grounded. In reality the input terminals are at slightly different dc potentials.
6. The voltage existing at the output when inputs are zero due to input offset voltage & bias current currents is called output offset voltage. Ideally it is zero when both input op-amp zero.

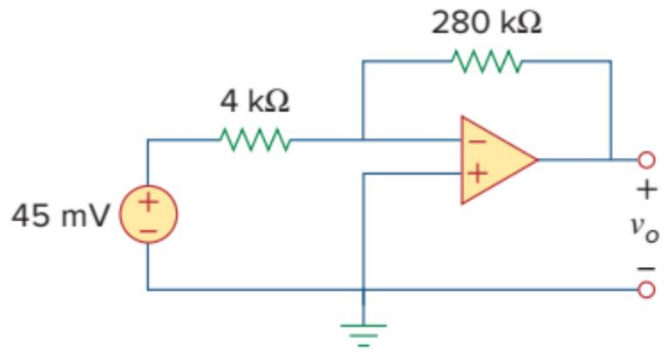
7. A differential amplifier is a type of electronic amplifier that amplifies the difference between two input voltages but suppresses any voltage common to the two inputs.
8. The op amp common-mode rejection ratio (CMRR) is the ratio of the common-mode gain to differential-mode gain.
9. The output current from the op-amp is that current needed to keep the inverting input at ground potential. So, with 1V at R_1 (left hand side), there has to be -1V at the output to make the inverting input zero volts. This means the current is -1V/100R = -10 mA.
10. If everything is same then current consumed should be low or high. Input power consumption is 15 V*1mA=15 mW , and output power given is 9.59 V*9.59 mA= 91.9 mW, how output can deviled more power than input consumed.
11. Slew rate is defined as the change of voltage or current, or any other electrical quantity, per unit of time. Expressed in SI units, the unit of measurement is volts/second or amperes/second, but is usually expressed in terms of microseconds (µs) or nanoseconds (ns).
12. The op amp gain bandwidth product is constant for voltage-feedback amplifiers. However it is not applicable for current feedback amplifiers because relationship between gain and bandwidth is not linear. Therefore decreasing the gain by a factor of ten will increase the bandwidth by the same factor.
13. The characteristics of an ideal op-amp are as follows: Infinite bandwidth due to the ideal gain inside of the op-amp, Infinite open-loop gain A, Infinite or zero common-mode gain, Input impedance of an infinite value, Output impedance of zero.
14. The important characteristic of comparator are: Speed of operation, Accuracy, Compatibility of output.
15. A detector is a device or circuit that extracts information from a modulated radio frequency current or voltage. The main Characteristics of Detectors are: Sensitivity, Detector Response, Energy Resolution, The Response Function, Response Time, Detector Efficiency, Dead Time.
16. An operational amplifier is a three-terminal device consisting of two high impedance input terminals, one is called the inverting input denoted by a negative sign and the other is the non-inverting input denoted with a positive sign. The third terminal is the output of the Op-Amp.
17. An op-amp differentiator or a differentiator amplifier is a circuit configuration which is inverse of the integrator circuit. It produces an output signal where the

instantaneous amplitude is proportional to the rate of change of the applied input voltage.

18. The basic operational amplifier integrator circuit consists of an op amp with a capacitor between the output and the inverting input, and a resistor from the inverting input to the overall circuit input.

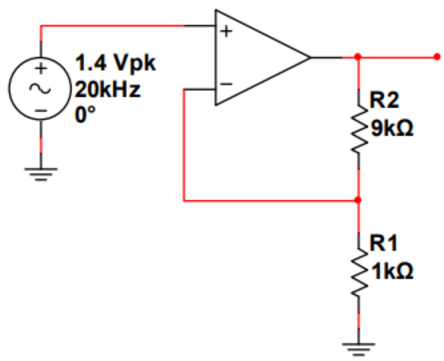
7.11 Terminal Questions:

1) Explain the working principle of an OP-amplifier with the help of symbol, number code and power supply.
2) What is the Input-output relationship of an op-amp? Also discuss the input-offset and output offset voltage in detail.
3) Explain the working principle of the Differential amplifier with the help of input and output resistance.
4) Discuss and explain of the Common mode rejection ratio and output current of the op-amp.
5) Write short notes of the followings: power consumption, slew rate and gain-band width product.
6) Explain the major characteristics of OP- amplifier.
7) What do you mean by comparator? Also write its characteristics.
8) Discuss and explain of the detector. Also write its characteristics.
9) Explain the working principle of the Inverting amplifier and also derive the expression for the output voltage.
10) Explain the working principle of the non-inverting amplifier. Also derive the expression for the closed loop voltage gain.
11) Explain the working principle of the Differentiator amplifier. Also derive the expression for the output voltage.
12) Discuss and explain working principle of the basic integrator and derive the expression for voltage output of the Op-amp Integrator.
13) Find the output of the op amp circuit. Calculate the current through the feedback resistor.



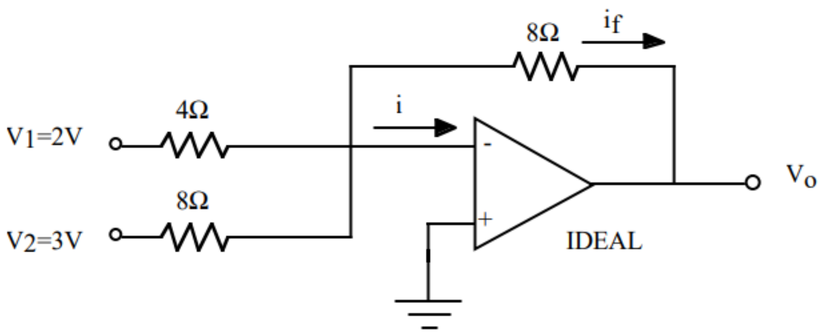
14) The differential amplifier has a common input signal of 3.20 V to both terminals. This results in an output signal of 26 mV. Determine the common-mode gain and the CMRR.

15) The following amplifier circuit is used to amplify a 1.4Vpk line-level audio signal with a frequency of 20kHz. The op-amp has a slew-rate of 1V/ µsec. Determine if the output will be slew-rate limited. If so, what is the minimum slew-rate required to produce the desired signal? (Remember, the maximum slope of the desired output signal must be considered.)

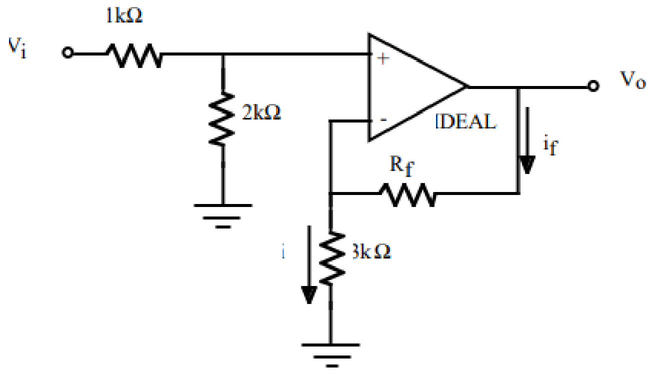


16) Refer to the following figure:

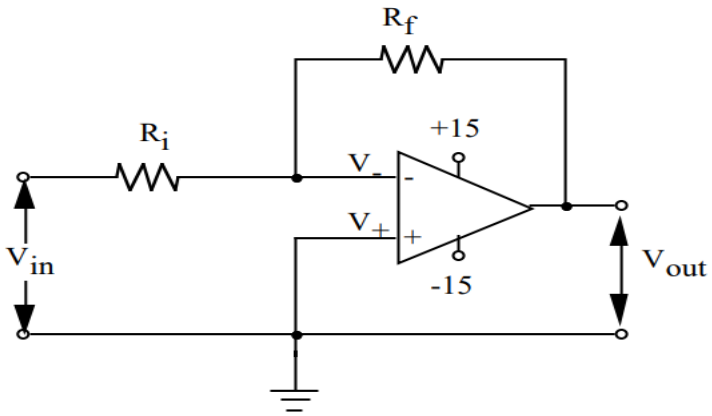
(a) What is the current, i? (b) What is the output voltage, Vo?



17) For the ideal op amp shown, what should be the value of resistor R_f to obtain a gain of 5?



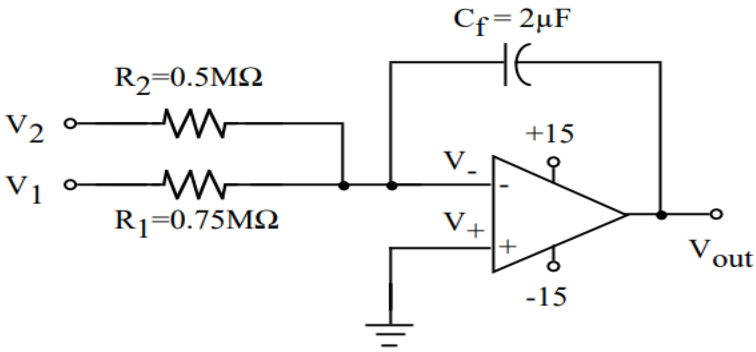
18) For the circuit shown below:



(a) If R_f = 1MW and R_i=50W, what is the voltage gain?

(b) If V_{in}=0.1 volts, what is V_{out}?

19) For the circuit shown below, V_1 = 10sin(200t) and V_2 = 15sin(200t) . What is V_{out}? The op amp is ideal with infinite gain.



20) An operational amplifier is required to amplify a signal with a peak voltage of 10 volts at a frequency of 30 kHz. Find out a slew rate.

DCEPHS-109

Solid State Physics and Advanced Electronics

Uttar Pradesh Rajarshi Tandon Open University

Bachelor of Science

Block

3 Advanced Digital Electronics

Course Preparation Committee

Mr. Ashish Sankhwar **Author**
Asst Prof. Dept. Of Electrical Engg. **Unit- 01-10**
FET,MJPRU,Bareilly

Prof. Nitya Kaushal Kishore **Editor**
Ex. HOD, Dept. of Physics, **Block -I,II & III**
S.K.M.U. University,
Dumka, Jharkhand.

Dr. Arvind Kumar Mishra **Coordinator**
Assistant Professor , Physics,
School of Science, UPRTOU, Prayagraj

DCEPHS – 109 : Plant Pathology and Microbiology
ISBN-

Printed By: Chandrakala Universal Pvt. 42/7 Jawahar Lal Nehru Road, Prayagraj.

Unit 8 Logic families

8.1 Introduction:

A logic family is a collection of different integrated circuit chips that have similar input, output, and internal circuit characteristics, but they perform different logic gate functions such as AND, OR, NOT, etc.

The logic families which fall under the first kind are called bipolar families, this include diode logic (DL), emitted coupled logic (ECL), resistor transistor logic (RTL), diode transistor logic (DTL), transistor-transistor logic (TTL).

The examples of unipolar families include PMOS, NMOS, and CMOS. Bipolar logic family. Transistors and diodes are bipolar devices, in which the current flows because of. both the charge carriers (electrons and holes). In bipolar logic families, transistors.

The main characteristics of Logic families include: Speed, Fan-in and Fan-out, Noise Immunity and Power Dissipation.

Fan-in is a term that defines the maximum number of digital inputs that a single logic gate can accept. Most transistor-transistor logic (TTL) gates have one or two inputs, although some have more than two. A typical logic gate has a fan-in of 1 or 2.

A fan-out ratio is the relationship in quantity between a single port on a storage device and the number of servers that are attached to it. It is important to know the fan-out ratio in a storage area network (SAN) design so that each server gets optimal access to storage resources.

A measure of a circuit's noise immunity is called the noise margin, which is expressed in volts. There are two values of noise margin specified for a given logic circuit: the HIGH level noise margin (V_{NH}) and the LOW-level noise margin (V_{NL}).

Typical rise and fall times for most logic devices will range from between 1nS and 4nS. Many FPGA's will have a selectable slew rate for it's outputs. The Xilinx 20KE FPGA may be configured for a Fast slew rate of 0.4nS rise and fall time, or slow slew rate of 1.1nS rise time and 2.04nS fall time.

A bipolar transistor switch is the simplest RTL gate (inverter or NOT gate) implementing logical negation. It consists of a common-emitter stage with a base resistor connected between the base and the input voltage source.

In digital circuit design, register-transfer level (RTL) is a design abstraction which models a synchronous digital circuit in terms of the flow of digital signals (data) between hardware registers, and the logical operations performed on those signals.

Diode- Transistor Logic, or DTL, refers to the technology for designing and fabricating digital circuits wherein logic gates employ diodes in the input stage and bipolar junction transistors at the output stage. The output BJT switches between its cut-off and saturation regions to create logic 1 and 0, respectively.

Transistor–transistor logic (TTL) is a logic family built from bipolar junction transistors. Its name signifies that transistors perform both the logic function (the first "transistor") and the amplifying function (the second "transistor"), as opposed to resistor–transistor logic (RTL) or diode–transistor logic (DTL).

Transistor-Transistor Logic (TTL) Applications are:

- TTL or Transistor-Transistor Logic is used in Mini Computer Processors.
- TTL is used Controller circuits, Display Driver Circuits.
- TTL is used in low-end programming devices such as remote controls, light controllers.
- TTL method is used in the microprocessor and microcontroller also.
- Smaller and low power electronic devices use TTL logic method.

In the TTL family, the Totem-pole circuit on the output is used to provide active pull-up and active pull-down. In TTL the output stage has two transistor amplifiers (one CE and another CC) connected in a push-pull configuration to act as current sink and source at the output stage.

Transistor–transistor logic (TTL) is a logic family built from bipolar junction transistors. Its name signifies that transistors perform both the logic function (the first "transistor") and the amplifying function (the second "transistor"), as opposed to resistor–transistor logic (RTL) or diode–transistor logic (DTL).

8.2 Objectives:

After studying this unit you should be able to

- Explain and identify Introduction and classification of logic families.
- Study and identify Input and output characteristics.
- Explain Fan-in and fan-out.
- Explain and identify Noise margin and noise immunity.
- Study and identify Rise and fall time.
- Explain and identify RTL (circuit, analysis and applications).
- Study and identify DTL (circuit, analysis and applications).
- Explain and identify TTL (circuit, analysis and applications), totem-pol.
- Study and identify Comparison of RTL, DTL and TTL.

8.3 Introduction and classification of logic families:

Introduction of logic families: Logic families are groups of logic circuits that are based on particular types of elements (resistors, transistors, and so forth). Families are identified by the manner in which the elements are connected, and, in some cases, by the types of elements used.

Logic circuits of a particular family can be interconnected without having to use additional circuitry. In other words, the output of one logic circuit can be used as the input to another logic circuit. This feature is known as compatibility. All circuits within a logic family will be compatible with the other circuits within that family.

Logic circuits are usually manufactured as integrated circuits and packaged in dual-inline (DIP) or other packages. Circuitry in a package is normally shown using standard logic symbols instead of individual components such as transistors, diodes, and so forth.

Several logic-circuit groups or families have been introduced. They differ primarily in the methods for carrying out the logic and the coupling to the inverter stages. For example, the transistor-transistor logic (TTL) uses a multiemitter transistor instead of the diodes found in DTL circuits. In emitter-coupled logic (ECL), the circuits are coupled by a common-emitter resistor, and complementary transistor logic (CTL) uses a combination of PNP and NPN transistors.

For each family, variations of a basic gate circuit are used to design a wide range of logic circuits with compatible input and output logical levels. In the design of a complete logic system, it is generally necessary to use logic circuits of one family only.

Before the various logic families are described, it is appropriate to discuss the general terminology used to describe logic families. These characteristics include logic assignments, logic voltage levels, supply voltage, noise margins, operating speeds, fan-in and fan-out, operating temperature, and power dissipation.

Classification of Logic Families: The two basic techniques for manufacturing ICs are:

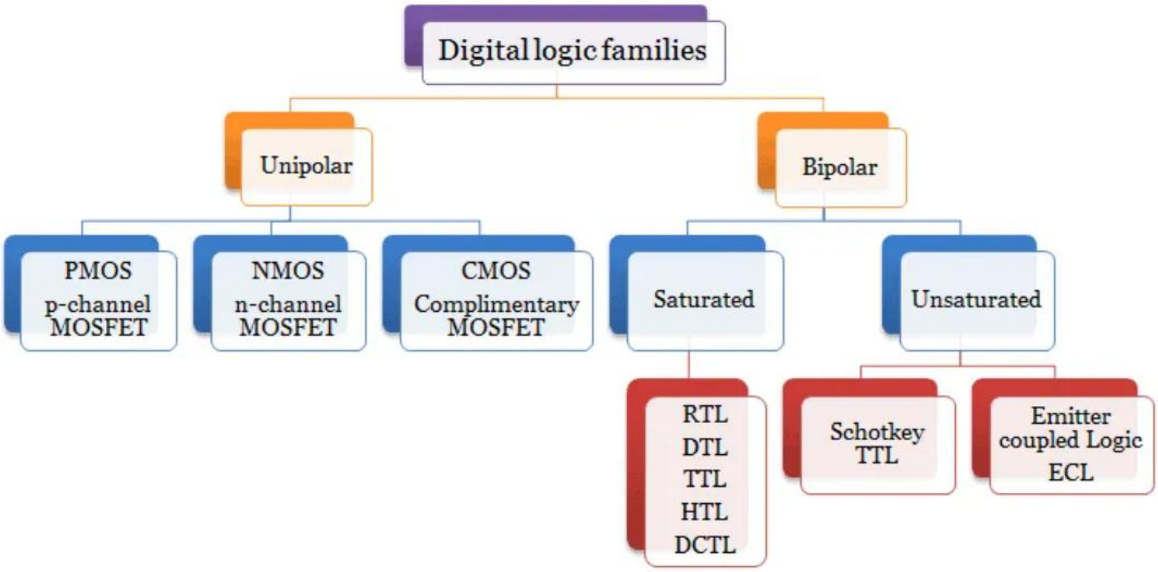


Fig.8.1 Classification of Logic Families

Logic families are mainly classified as Bipolar Logic Families and Unipolar Logic Families.

1. **Bipolar Logic Families:** It mainly uses bipolar devices like diodes, transistors in addition to passive elements like resistors and capacitors. These are sub classified as saturated bipolar logic family and unsaturated bipolar logic family.

- **Saturated Bipolar Logic Family:** In this family the transistors used in ICs are driven into saturation. For example:
   a) Transistor-Transistor Logic (TTL)
   b) Resistor-Transistor Logic (RTL)
   c) Direct Coupled Transistor Logic (DCTL)
   d) Diode Transistor Logic (DTL)

e) High Threshold Logic(HTL)
f) Integrated Injection Logic (IIL or I^2L)

- **Unsaturated bipolar logic family:** In this family the transistors used in IC is not driven into saturation. For example:
  a) Schottky TTL
  b) Emitter Coupled Logic(ECL)

2. **Unipolar Logic Families:** It mainly uses Unipolar devices like MOSFETs in addition to passive elements like resistors and capacitors. These logic families have the advantages of high speed and lower power consumption than Bipolar families. These are classified as:
   a) PMOS or P-Channel MOS Logic Family
   b) NMOS or N-Channel MOS Logic Family
   c) CMOS Logic Family.

8.4 Input and output characteristics of logic family: various logic families possess different characteristic, so one family may be best suited to one situation whereas another family in some other situation. for example in certain cases, low power consumption's may be the prime requirement where as some other cases speed.

Various characteristic of a logic family are given below in brief:

- **Operating Speed:** Speed of a logic gate depends upon the time that elapses between the application of a signal to an input terminal and the resulting change in logical state at the output terminals. It takes into consideration the transition time (rise and fall duration of a pulse) and propagation delays. Both of these times depend upon the loading and increase with increase in load. The more inputs are attached to the output of a logic gate, the more load is to be handled by that output. High operating speed is usually the main requirement of digital ICs.
- **Fan-In:** The fan-in of a logic gate is defined as the number of inputs (coming from similar circuits) that it can handle properly.
- **Fan-Out:** In general, a logic circuit is required to drive several logic inputs. The fan-out (also sometimes called the loading factor) is defined as the maximum number of standard logic inputs that an output can drive reliably. For example, a logic gate that is specified to have a fan-out of 8 can drive 8 standard logic inputs. If this number exceeds the output logic-level voltages cannot be guaranteed.

- **Power Dissipation:** This is the amount of power dissipated in an IC. It is determined by the current, Icc. that it draws from the Vcc supply and equals Vcc Icc where Icc is average value of Icc(0) and Icc(1). This power is specified in mW. Lower power dissipation is desirable feature for any IC.
- **Power Supply Requirements:** Every IC requires a certain amount of electrical power to operate. The power is supplied by one or more power-supply voltage connected to the power pin (or pins) on the chip. Usually there is only one power-supply terminal on the chip as it is marked Vcc for TTL or VDD for MOS devices. Obviously low power consumption is desirable features in any digital ICs.
- **Noise Immunity:** Stray electric and magnetic fields can induce voltages on the connecting wires between logic circuits. These unwanted, spurious signals are known as noise and can sometimes lead to false triggering of logic levels in the circuits. The noise immunity of a logic circuit refers to the circuit's ability to tolerate noise voltages on its inputs. A quantitative measure of noise margin. Higher the noise margin, better the logic circuit.
- **Operating Temperature Range:** Digital ICs should be capable of operating for temperature ranging from 0^0Cto 70^0C for consumers and from -55^0C to $+125^0$C for military applications.
- **Figure of Merit:** The figure of merit of a digital IC is defined as the product of speed and power. The speed is specified in terms of propagation delay time expressed in nano seconds.

  Figure of merit = Propagation time (ns) x power (mW) [Measured in pico joules (pJ)]
  A low value of speed-power product is desirable. In a digital circuit if high speed or low propagation delay is desired, then there will be corresponding increase in power dissipation and vice-versa.
- **Flexibilities Available:** various flexibilities are available in different IC logic families and these most be considered while selecting a logic family for a particular job. Some of the flexibilities available are:

a) **Wire-logic Capability:** Connection of gate output terminals together or using them directly to perform additional logic functions without any extra hardware.

b) **Availability of Complement Outputs:** This eliminates the need for additional inverters.

c) **Breadth of Series:** Types of different logic functions available in the series.

d) **Popularity of Series:** The cost of manufacturing depends upon the number of ICs manufactured. When the ICs are manufactured in a large number the cost per function is reduced and it will be easily available because of multiple resources.
e) **Input-Output Facilities:** The number of input terminals of a gate and its input/output impedance's in both 0 and 1 states are important. Former governs the fan-in and the later its fan-out. For high fan-out, the gates have low output impedance for both 0 and 1.

8.5 Fan-in and fan-out:

Fan-In: Fan-in refers to the maximum number of input signals that feed the input equations of a logic cell. Fan-in is a term that defines the maximum number of digital inputs that a single logic gate can accept. Most transistor-transistor logic (TTL) gates have one or two inputs, although some have more than two. A typical logic gate has a fan-in of 1 or 2.
In some digital systems, it is necessary for a single TTL logic gate to drive several devices with fan-in numbers greater than 1. If the total number of inputs a transistor-transistor logic (TTL) device must drive is greater than 10, a device called a buffer can be used between the TTL gate output and the inputs of the devices it must drive. A logical inverter (also called a NOT gate) can serve this function in most digital circuits.
In quantum logic gates the fan-in always has to be equal to the number of outputs, the Fan-out. Gates for which the numbers of inputs and outputs differ would not be reversible (unitary) and are therefore not allowed.
In the fan-in always has to be equal to the number of outputs, the fan in. Gates for which the numbers of inputs and outputs differ would not be reversible and are therefore not allowed.

Fan-out: The fan-out is defined as the maximum number of inputs (load) that can be connected to the output of a gate without degrading the normal operation. Fan Out is calculated from the amount of current available in the output of a gate and the amount of current needed in each input of the connecting gate. It is specified by manufacturer and is provided in the data sheet. Exceeding the specified maximum load may cause a malfunction because the circuit will not be able supply the demanded power.
More complex analysis than fan-in and fan-out is required when specific good judgment families are interconnected. Fan-out is in the long run determined by the maximum supply and sink currents of an output and the maximum supply and sink currents of the connected inputs; the riding tool must be able to supply or sink at its output the sum of the currents

needed or supplied (depending on whether the output is a good judgment excessive or low voltage degree) by all of the linked inputs, whilst maintaining the output voltage specifications. For every good judgment family, usually a “general” enter is defined by the producer with maximum input currents at each logic degree, and the fan-out for an output is computed as the number of those popular inputs that may be driven inside the worst case.

Fan-in and Fan-out Effects:

- In the LOW state the output voltage V_{OL}may increase above V_{OLmax}.
- In the HIGH state the output voltage V_{OH}may decrease below V_{OHmin}.
- The operating temperature of the device may increase, thereby reducing the reliability of the device and eventually causing the device to fail.
- Output rise and fall times may increase beyond specifications.

Applications of Fan-in and Fan-out:

- Fan in is also used when we need to connect two devices together, but we don’t have any electricity.
- Fan out is used when we need to connect two devices together, like when you plug an electric cord into a wall and a battery into the other side.

SAQ.1

a) What do you understand by logic families?
b) Discuss and write the classification of logic families.
c) What do you mean by Input and output characteristics of logic families.
d) Write in brief of the fan-in and fan-out in logic families.

8.6 Noise margin and noise immunity:

It is an unwanted signal that is superimposed on the normal operating signal. Noise may be due to various factors like operating environment, radiations, stray electrical and magnetic fields.

In digital logic circuits, the binary values 0 and 1 represent the LOW and HIGH voltage levels. Due to the interference of the noises, the voltage levels may increase or decrease. This may lead to the wrong operation of the device.

The noise immunity is the ability of the logic device to tolerate the noise without causing spurious change to the output voltage. Noise margin allows the logic device to function properly within the specified limits.

In logic systems, the word noise refers to any unwanted voltage (AC or DC) appearing at the input of a logic circuit. If such a noise voltage were high enough, it could cause the circuit to change state with no change in the input signal voltage.

Both DC and AC noise should be considered in the design of digital systems. The DC noise is the steady drift in the voltage levels of the logic states, and AC noise is the narrow pulses that are created, primarily, by switching transients.

The DC noise margin (immunity) of a digital circuit is the ability of that circuit to maintain a logic state in the presence of DC noise. The DC noise margin is expressed by the following equations

$$N_{\mathrm{L}} = V_{\mathrm{ILmax}} - V_{\mathrm{OLmax}}$$
$$N_{\mathrm{H}} = V_{\mathrm{OHmin}} - V_{\mathrm{IHmin}}$$

Where

N_{L} - noise immunity of the digital circuit input when the input signal is LOW,

N_{H} - noise immunity of the digital circuit input when the input signal is HIGH,

V_{ILmax} - maximum input voltage that can be read by the circuit as LOW,

V_{OLmax} - maximum output voltage that can represent LOW,

V_{OHmin} - minimum output voltage than can represent HIGH,

V_{IHmin} - minimum input voltage that can be read by the circuit as HIGH.

Integrated circuit manufacturers include the voltage level limits of their digital circuits on the data sheets. From those values, the noise immunity can be computed.

For example, a standard TTL logic gate has the following input and output level characteristics:

$V_{\mathrm{ILmax}} = 0.8$ V, $V_{\mathrm{OLmax}} = 0.4$ V

$V_{\mathrm{IHmin}} = 2$ V, $V_{\mathrm{OHmin}} = 2.4$ V

Substituting these values into the noise immunity equations:

$$N_{\mathrm{L}} = V_{\mathrm{ILmax}} - V_{\mathrm{OLmax}} = 0.8 - 0.4 = 0.4 \text{ V}$$

and

$$N_{\mathrm{H}} = V_{\mathrm{OHmin}} - V_{\mathrm{IHmin}} = 2.4 - 2 = 0.4 \text{ V}$$

These values indicate that for reliable operation, the DC noise on the signal lines should not exceed 0.4 V.

It is relatively straightforward to specify the DC noise margin of a circuit, but this is not the whole story. Noise can appear at the input to a logic circuit as short, transient voltage pulses. The ability of a circuit to withstand such transient noise voltages is called the AC noise margin or AC noise immunity. The AC value therefore depends on the shape and duration of the voltage pulses, and may be affected by stray capacitances; the AC noise margin is much more difficult to define and specify.

8.7 Input rise and fall times: The input rise and fall times of general-purpose CMOS logic ICs are specified in the operating ranges in which their functional operation is guaranteed.

Use CMOS logic ICs within the operating ranges to prevent a malfunction due to output oscillation etc.

If a slowly rising or falling signal (a low slew rate signal) is applied to an input, a current spike occurs during switching, causing V_{CC} and GND bounce, which might result in output oscillation or a malfunction.

Use ICs with a Schmitt-trigger input for slowly changing inputs. However, in the case of excessively slowly changing inputs, even ICs with a Schmitt-trigger input might not be able to suppress noise on power supply or signal lines, resulting in output oscillation or instability.

The rise time is then based on the RC time constant of the circuit, and the amount of current the IC can deliver. Typical rise and fall times for most logic devices will range from between 1nS and 4nS.

8.8 RTL (circuit, analysis and applications):

The resistor-transistor logic is the first digital circuit to construct the logic gates. It is also known as Transistor-Resistor Logic or TRL. Here, the resistor and transistor are used to make the logic gates. The resistor is used as the input component and the transistor is used as the output component. Here, you can see a basic circuit of NOT gate using RTL Logic.

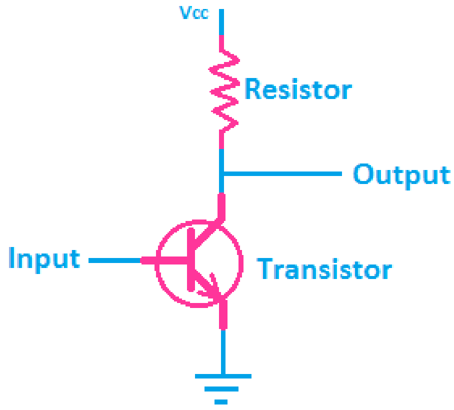


Fig.8.2 Resistor-Transistor Logic

The NOT gate uses RTL logic also known as RTL inverter. The working of this circuit is very simple. When no voltage is applied to the input, the transistor acts as an open circuit so the output will be high. When the voltage is applied to the input terminal, the transistor act as a closed circuit, so the Vcc will be grounded and the output of the circuit will be Low.

2-input RTL NOR gate: The following figure shows the circuit diagram of 2-input RTL NOR gate. Q_1 and Q_2 are the two transistors. A and B are the two inputs, given to the base of two transistors and Y is the output.

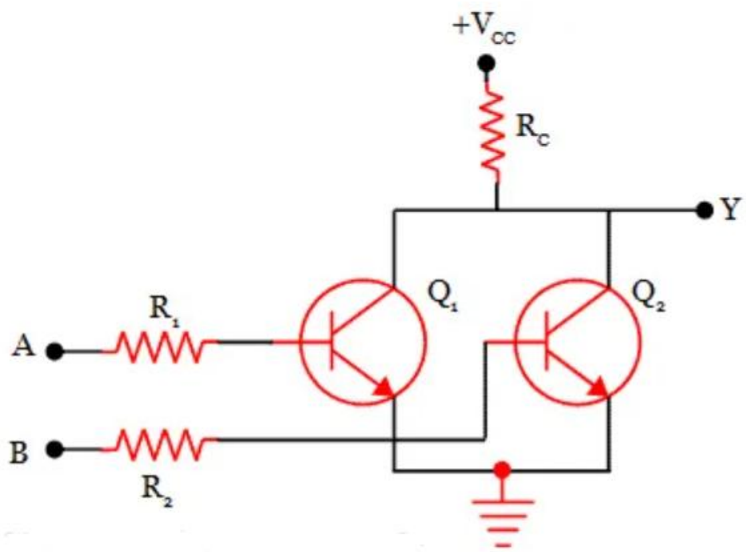


Fig.8.3 2-input RTL NOR gate

When both the inputs A and B are at 0V or logic 0, it is not enough to turn on the gates of both the transistor. So the transistors will not conduct. Due to this, the voltage +VCC will appear at the output Y. Hence the output is logic 1 or logic HIGH at terminal Y.

When any one of the inputs, either A or B is given HIGH voltage or logic 1, then the transistor with HIGH gate input will be turned on. This will make a path for the supply voltage to go to the ground through the resistor R_C and transistor. Thus there will be 0V at the output terminal Y.

When both the inputs are HIGH, it will drive both the transistor to turn on. It will make a path for the supply voltage to flow to the ground through resistor R_C and transistor. Therefore, there will be 0V at the output terminal Y.

The below table shows the truth table for NOR gate.

Inputs		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

3-input RTL NOR gate:

The above discussed 2-input RTL NOR gate is the basis for all the logic circuits built with resistors and transistors. The 3-input Resistor-Transistor Logic NOR gate can also be constructed as shown below. The operation is similar to the 2-input RTL NOR gate.

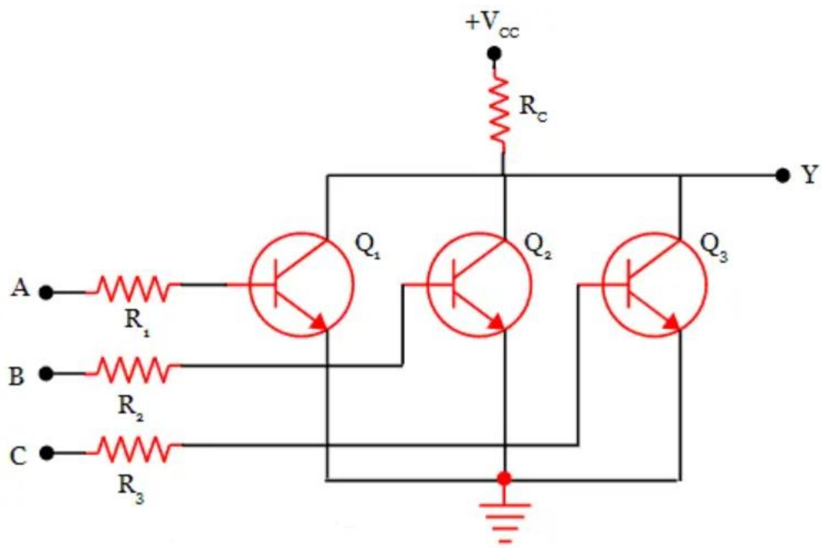


Fig.8.4 3-input RTL NOR gate

Resistor-Transistor Logic (RTL) NOT Gate:

A RTL inverter or NOT gate is shown in figure below.

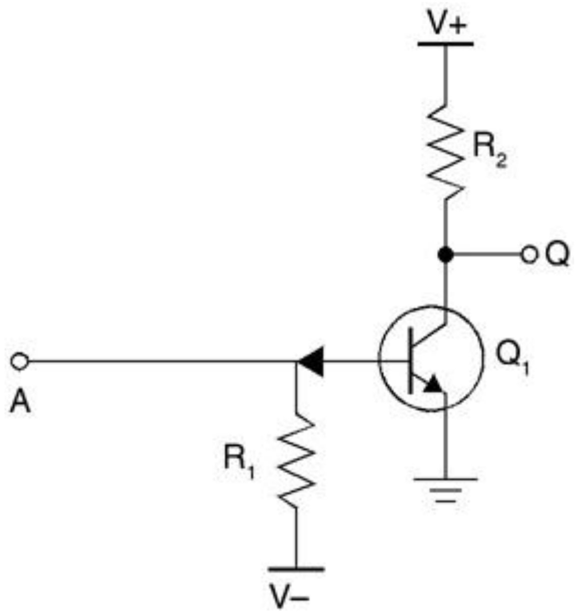


Fig.8.5 RTL NOT gate

Bipolar transistor switch is the simplest RTL gate. The resistor R_1 in the circuit is used across the base and input terminals. This resistor increases the voltage drop from 0.7 V to 1 V by converting the input voltage into current. The resistance R_1 is chosen in such a way that it saturates the transistor and obtains high input resistance. The collector resistor R_2 converts collector current into voltage. The resistance of R_2 is high to saturate the transistor and low to obtain output resistance.

RTL NOR Gate:

The RTL NOT gate can be converted into an RTL NOR gate by introducing additional base resistors R3 and R4. The circuit of an RTL NOR gate is shown in the figure below.

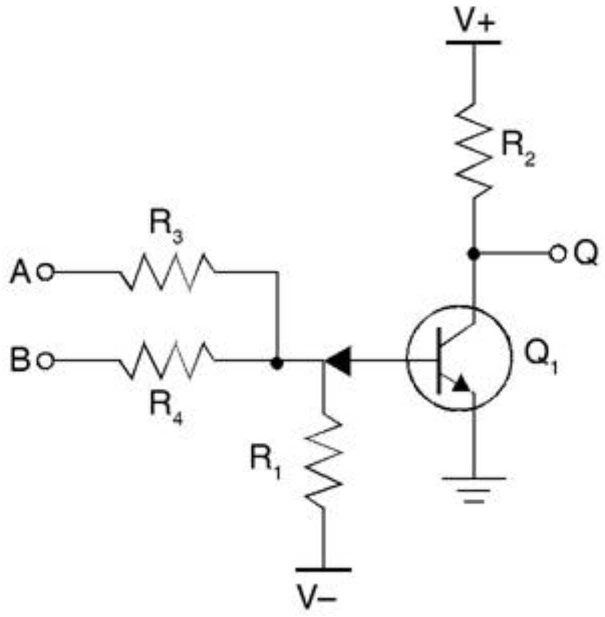


Fig.8.6 RTL NOR

When both A and B are given logic 0, then the transistor is cut-off. The output is inverted since it is complimentary. This is because the voltage drop across the collector-emitter junction of the transistor Q1 is taken at Q instead of taking it across collector resistor R2.

Multistage NOR Gate:

A multistage NOR gate circuit is shown in the figure below. It consists of parallel connections of BJTs that are controlled by input logic. The inputs are not interconnected and hence if A is high, transistor Q1 conducts and pulls Q to the ground. Similarly, if B is high then transistor Q2 conducts and pulls Q to the ground.

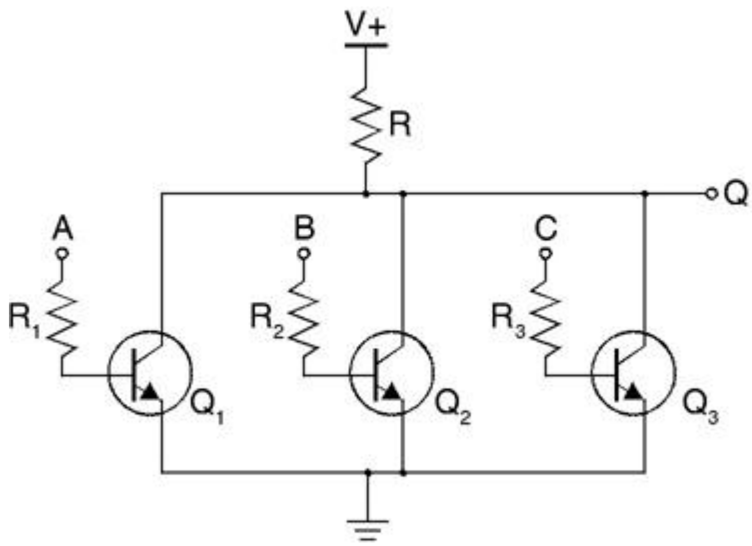


Fig.8.7 Multistage RTL NOR gate

Advantages of RTL Logic:

- It is the simplest digital circuit.
- The minimum numbers of transistors are required to build the logic circuit than other logic families.
- Operation of this circuit is very simple.

Disadvantages of RTL Logic:

- Very poor response time, high propagation delay.
- Power dissipation is high due to the resistor.

Applications of RTL Logic:

- This combination provides clean output signals and either inversion or non-inversion as needed.
- RTL gates are almost as simple as DL gates, and remain inexpensive.
- They also are handy because both normal and inverted signals are often available.

SAQ.2

a) What do you understand by the noise margin and noise immunity?
b) Discuss and write in brief Rise and fall time in logic families.
c) What do you mean by RTL and write its applications?
d) Write the advantages and disadvantages of RTL.

8.9 DTL (circuit, analysis and applications):

In the diode transistor logic, the logic gates are built with PN junction diodes and transistors. Here, diodes are used as input components and transistors are used as output components. Here, you can see the basic circuit diagram of NAND gate using diodes and transistors.

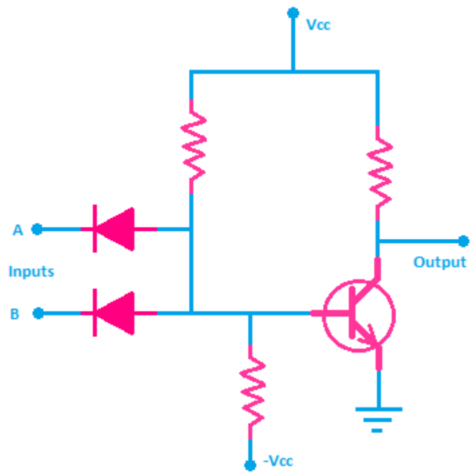


Fig.8.8 Diode Transistor Logic (DTL)

You can see it have two input terminals and one output terminal. Let's take a condition to understand its working principle. You can see in both input terminals, diodes are connected in reverse bias. So, when both input is low the Vcc at the input side of the transistor will be grounded through the diodes and resistor. So, the transistor will be in the cutoff condition and the output of the circuit will be high.

DTL NAND gate: The diode-transistor logic, also termed as DTL, replaced RTL family because of greater fan-out capability and more noise margin.

As its name suggests, DTL circuits mainly consists of diodes and transistors that comprises DTL devices.

The basic DTL device is a NAND gate, shown below.

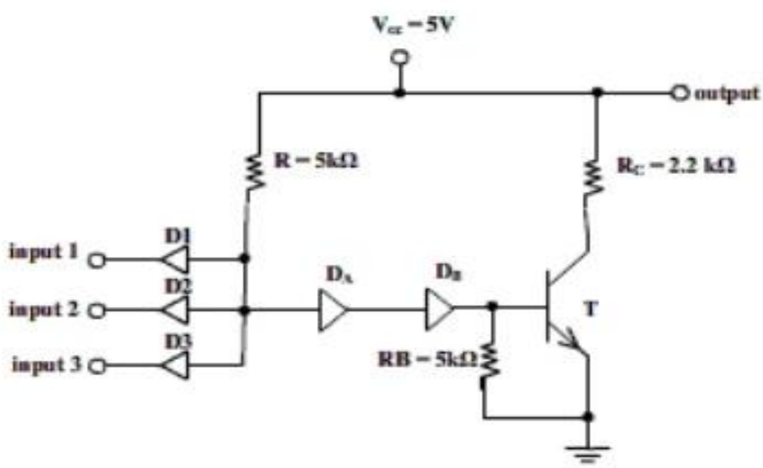


Fig.8.9 DTL NAND gate

Three inputs to the gate are applied through three diodes viz. D1, D2 and D3. The diode will conduct only when corresponding input is LOW.

If any of the diode is conducting i.e. when at least one input is LOW, the voltage at cathode of diode DA is such that it keeps transistor T in cut-off and subsequently, output of transistor is HIGH.

If all inputs are HIGH, all diodes are non-conducting, transistor T is in saturation, and its output is LOW.

Due to number of diodes used in this circuit, the speed of the circuit is significantly low. Hence this family of logic gates is modified to transistor-transistor logic i.e. TTL family.

Advantages of DTL Logic:

- Very low power dissipation.
- Response time is better than RTL logic.

Disadvantages of DTL Logic:

- Types of logic gates built by DTL logic is limited.

Applications of DTL Logic:

- Diode-Transistor Logic, or DTL, refers to the technology for designing and fabricating digital circuits where in logic gates employ both diodes and transistors.
- DTL offers better noise margins and greater fan-outs than RTL, but suffers from low speed, especially in comparison to TTL.

8.10 TTL (circuit, analysis and applications):

In transistor-transistor logic, logic gates are built with bipolar junction transistors. Here, one transistor works for amplifying while another works for switching. Almost, all types of logic gates are possible to build with transistor-transistor logic. Here, you can see the basic circuit diagram of NAND gate using TTL logic.

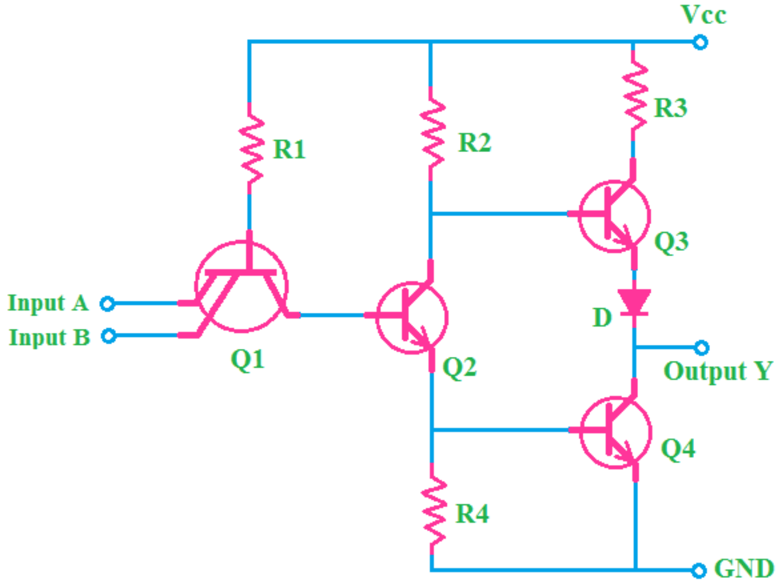


Fig.8.10 Transistor-Transistor Logic (TTL)

You can see it is a two-input circuit. The base terminal of the first transistor is connected to the Vcc and the base terminal of the second transistor is connected to the collector terminal of the first transistor. The first transistor has multiple emitter terminals. Both of the input terminals of the circuit are connected to the emitter terminal of the first transistor.

TTL logic includes several transistors that have several emitters as well as several inputs. The types of TTL or transistor-transistor logic mainly include Standard TTL, Fast TTL, Schottky TTL, High power TTL, Low power TTL & Advanced Schottky TTL.

The designing of TTL logic gates can be done with resistors and BJTs. There are several variants of TTL which are developed for different purposes such as the radiation-hardened TTL packages for space applications and Low power Schottky diodes that can provide an excellent combination of speed and lesser power consumption.

Classification of Transistor-Transistor Logic:

It is a logical family consisting completely of transistors. It employs a transistor with multiple emitters. Commercially it starts with the 74 series like the 7404, 74S86, etc. It was built in 1961 by James L Bui and commercially used in logic design in 1963. TTLs are classified based on the output.

Open Collector Output:

The main feature is that its output is 0 when low and floating when high. Usually, an external Vcc may be applied.

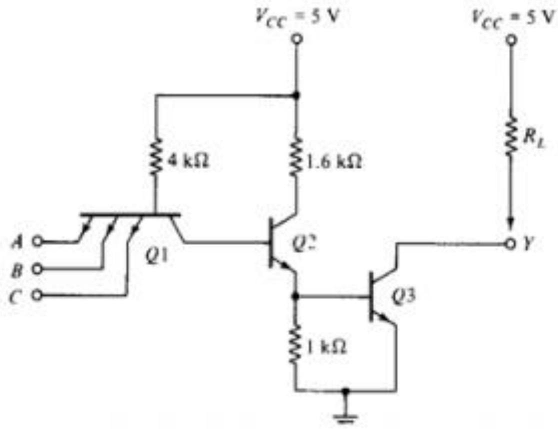


Fig.8.11 Open Collector Output of Transistor-Transistor Logic

Transistor Q1 behaves as a cluster of diodes placed back to back. With any of the input at logic low, the corresponding emitter-base junction is forward biased and the voltage drop across the base of Q1 is around 0.9V, not enough for the transistors Q2 and Q3 to conduct. Thus the output is either floating or Vcc, i.e. High level.

Similarly, when all inputs are high, all base-emitter junctions of Q1 are reverse biased and transistor Q2 and Q3 get enough base current and are in saturation mode. The output is at logic low. (For a transistor to go to saturation, collector current should be greater than β times the base current).

Applications:

The applications of open collector output include the following.

- In driving lamps or relays
- In performing wired logic
- In the construction of a common bus system

Three State Gate:

It provides 3 state output like the following

- Low-level state when a lower transistor is ON and an upper transistor is OFF.
- High-level state when the lower transistor is OFF and the upper transistor is ON.
- Third state when both transistors are OFF. It allows a direct wire connection of many outputs.

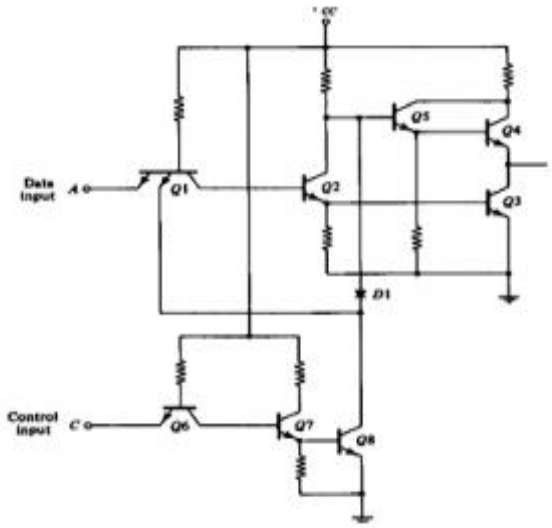


Fig.8.13 Three State Gate Transistor-Transistor Logic

TTL Family Features:

The features of the TTL family include the following.

- Logic low level is at 0 or 0.2V.
- Logic high level is at 5V.
- Typical fan out of 10. It means it can support at most 10 gates at its output.
- A basic TTL device draws a power of almost 10mW, which reduces with the use of Schottky devices.
- The average propagation delay is about 9ns.
- The noise margin is about 0.4V.

Series of TTL IC:

TTL ICs mostly start with the 7 series. It has 6 subfamilies given as:

1. Low Power device with a propagation delay of 35 ns and power dissipation of 1mW.
2. Low power Schottky device with a delay of 9ns
3. Advanced Schottky device with a delay of 1.5ns.
4. Advanced low power Schottky device with a delay of 4 ns and power dissipation of 1mW.

In any TTL device nomenclature, the first two names indicate the name of the subfamily the device belongs to. The first two digits indicate the temperature range of operation. The next two alphabets indicate the subfamily the device belongs to. The last two digits indicate the logic function performed by the chip. The examples are 74LS02- 2 neither input NOR gate, 74LS10- Triple 3 input NAND gate.

Typical TTL Circuits

Logic Gates are used in daily life in applications like a clothes dryer, computer printer, doorbell, etc.

The 3 basic Logic gates implemented using TTL logic is given below:-

TTL NOR Gate:

Suppose input A is at logic high, the corresponding transistor's emitter-base junction is reverse biased, and base-collector junction is forward biased. Transistor Q3 gets base current from supply voltage Vcc and goes to saturation. As a result of the low collector voltage from Q3, transistor Q5 goes to cut off and on the other hand, if another input is low, Q4 is cut off and correspondingly Q5 is cut off and output is connected directly to the ground through transistor Q3. Similarly, when both inputs are logic low, the output will be at logic high.

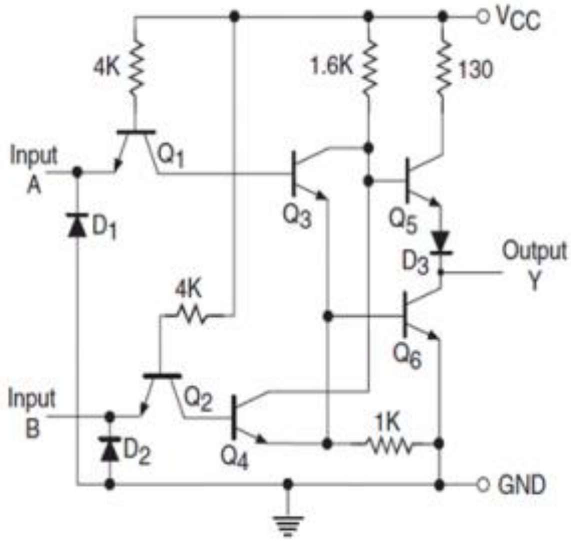


Fig.8.14 NOR Gate TTL

TTL NOT Gate:

When the input is low, the corresponding base-emitter junction is forward biased, and the base-collector junction is reverse biased. As a result transistor Q2 is cut off and also transistor Q4 is cut off. Transistor Q3 goes to saturation and diode D2 starts conducting and output is connected to Vcc and goes to logic high. Similarly, when input is at logic high, the output is at logic low.

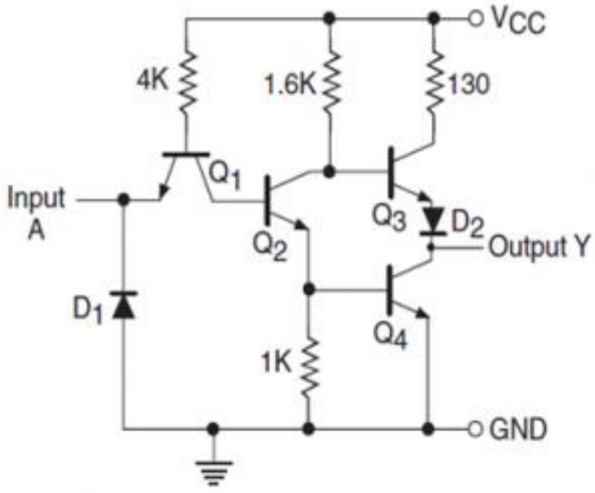


Fig.8.15 NOT Gate TTL

TTL NAND gate:

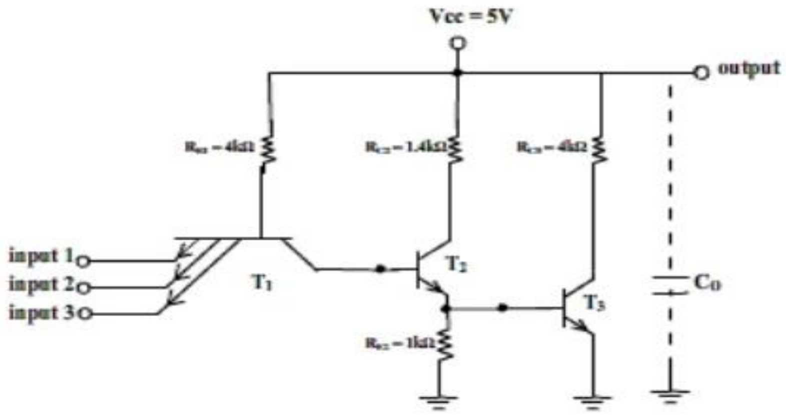


Fig.8.16 TTL NAND gate

TTL family is a modification to the DTL. It has come to existence so as to overcome the speed limitations of DTL family. The basic gate of this family is TTL NAND gate.

Modifications to DTL NAND:

- The diodes D1, D2 and D3 are replaced by emitter-base junctions of a multiple-emitter transistor labeled T1.
- Diode DA is replaced by collector-base junction of T1.
- Diode DB is replaced by emitter-base junction of transistor labeled T2.

The working of this circuit is identical to that of DTL circuit.

Case1- When at least one input is logic LOW, transistor T2 and T3 are in cut-off and hence, output of T3 is HIGH.

Case2- When all inputs are HIGH, T1 operates in active inverse mode, driving T2 & T3 in saturation. Since T3 is ON, the output is LOW.

Case3- While all inputs are HIGH, if any of the inputs suddenly goes LOW, then T2 and T3 will be turned off only when stored base charge is removed. The collector-base junction of T1 is back-biased and T1 operates in normal active region. A large collector current of T1 is in such direction that it helps removing base charge of T2 and T3. In this way, the circuit speed is increased in TTL over speed of DTL.

Advantages and Disadvantages:

The advantages of disadvantages of TTL include the following.

- The main benefit of TTL is we can easily interface with other circuits & the ability to generate difficult logic functions because of certain voltage levels as well as good noise margins TTL has good features like fan-in which means the number of i/p signals that can be accepted through an input.
- TTL is mainly immune to harm from stationary electricity discharges not like CMOS & as compared to CMOS these are economical. The main drawback of TTL is high current utilization. The TTL's high current demands can lead to offensive functioning because o/p states will be turned off. Even with different TTL versions that have low current consumption will be competitive to CMOS.
- With the arrival of CMOS, TTL applications have been replaced through CMOS. But, TTL is still utilized in applications because they are quite robust & the logic gates are fairly cheap.

TTL Applications:

The applications of TTL include the following.

- Used in controller application for providing 0 to 5Vs.
- Used as a switching device in driving lamps and relays.
- Used in processors of mini computers like DEC VAX.
- Used in printers and video display terminals.

Totem Pole Output: Totem Pole means the addition of an active pull up the circuit in the output of the Gate which results in a reduction of propagation delay.

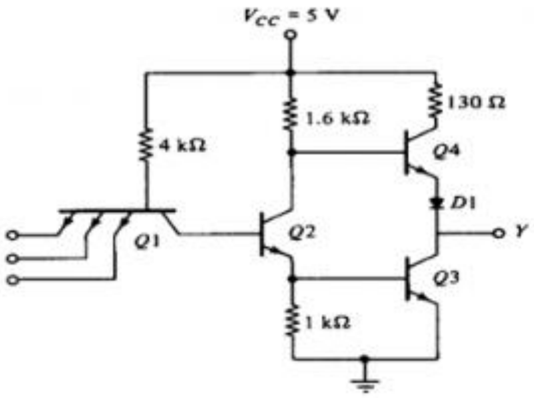


Fig.8.12 Totem Pole Output TTL

Logic operation is the same as the open collector output. The use of transistors Q4 and diode is to provide quick charging and discharging of parasitic capacitance across Q3. The resistor is used to keep the output current to a safe value.

8.11 Comparison of RTL, DTL and TTL:

S. No.	RTL Logic	DTL Logic	TTL Logic
1	Built with Resistor and Transistor	Built with Diode and Transistor	Built with Transistors
2	Slow Response	Better than RTL Logic	Much better than RTL and DTL
3	High Power Loss	Low Power Loss	Low Power Loss
4	Very simple in construction and operation	Simple in construction and operation	Complex in construction and operation
5	RTL Logic used in old computers	DTL logic used in basic digital circuits, switching circuits	All the modern digital circuits, Integrated Circuits are mostly built with TTL Logic

SAQ.3

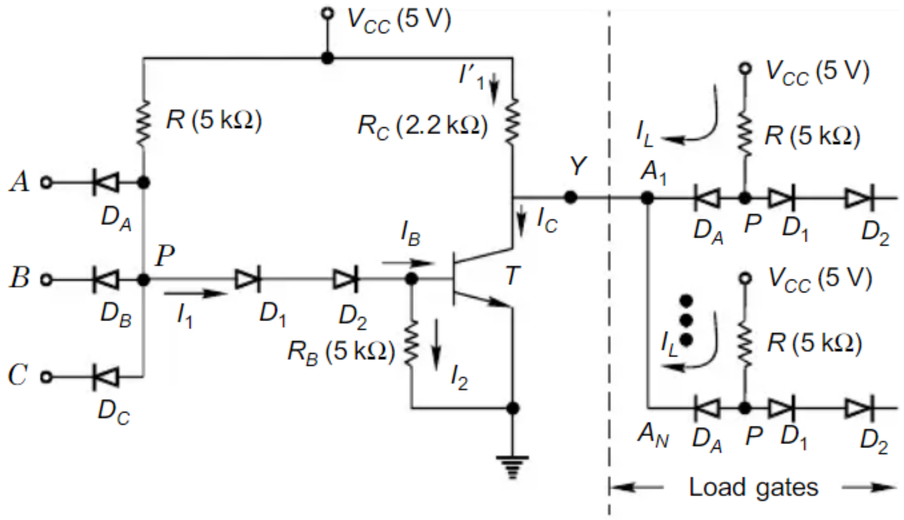
a) What do you mean by DTL circuit and its applications?

b) Discuss in brief TTL circuit and its applications.

c) What do you mean by totem-pol?

d) Write the comparison between of RTL, DTL and TTL.

e) Calculate the value of h FE required for a fan-out of 10 in the DTL gate of Fig. given below.



8.12 Summary:

1. A logic family of monolithic digital integrated circuit devices is a group of electronic logic gates constructed using one of several different designs, usually with compatible logic levels and power supply characteristics within a family.
2. Digital logic family is a group of logic gates constructed using passive devices like a resistor, transistor, diodes, etc. Such devices have compatible logic levels and supply voltage with certain characteristics.
3. The two basic techniques for manufacturing ICs are: (a) Bipolar Technology (b) Metal oxide semiconductor (MOS) technology.
4. Fan In: The fan-in defined as the maximum number of inputs that a logic gate can accept. If number of input exceeds, the output will be undefined or incorrect. It is specified by manufacturer and is provided in the data sheet.
5. Fan Out: The fan-out is defined as the maximum number of inputs (load) that can be connected to the output of a gate without degrading the normal operation. Fan Out is calculated from the amount of current available in the output of a gate and the amount of current needed in each input of the connecting gate. It is specified by manufacturer and is provided in the data sheet. Exceeding the specified maximum load may cause a malfunction because the circuit will not be able supply the demanded power.
6. The noise immunity is the ability of the logic device to tolerate the noise without causing spurious change to the output voltage. Noise margin allows the logic device to function properly within the specified limits.

7. The rise time (tr) is the length of time it takes for a pulse to rise from its 10% point up to its 90% point.
8. The fall time (tf) is the length of time it takes for a pulse to fall from its 90% point to its 10% point.
9. The resistor-transistor Logic(RTL) circuit is one of the basic logic circuits in digital logic families. It is a bipolar saturated device. The RTL logic is popular because of its simplicity. The RTL circuit consists of resistors at inputs and transistors at the output side.
10. Diode transistor logic (DTL) belongs to the digital logic family. This logic circuit has diodes at the input side and transistor at the output side and so the name diode transistor logic. It has more advantages than resistor transistor logic(RTL).
11. The Transistor-Transistor Logic (TTL) is a logic family made up of BJTs (bipolar junction transistors). As the name suggests, the transistor performs two functions like logic as well as amplifying. The best examples of TTL are logic gates namely the 7402 NOR Gate & the 7400 NAND gate.
12. The TTL "totem-pole" output structure often has a momentary overlap when both the upper and lower transistors are conducting, resulting in a substantial pulse of current drawn from the power supply.
13. Logic gates are the main components of any digital circuit. Almost all digital circuits are made of logic gates. These logic gates are made of transistors and other active or passive components. Here, transistors work as electronic solid-state switches. There are different classes or types of logic families according to which logic gates are built. The comparison between three common logic families that are RTL, DTL, and TTL.

8.13 Terminal Questions:

1) What do you mean by Introduction of logic families? Explain in detail classification of logic families.
2) Explain the Input and output characteristics of logic families.
3) Discuss and write the Fan-in and fan-out of logic families.
4) Explain the Noise margin and noise immunity.
5) What do you understand by the Rise and fall time?
6) Explain the working of the RTL circuit analysis and write its applications.
7) Explain the working of the DTL circuit analysis and write its advantages, disadvantages and applications.

8) Explain the working of TTL circuit analysis and write its applications.

9) What do you understand by the totem-pol output?

10) Write the Comparison between of the RTL, DTL and TTL.

Unit 9 Sequential circuits

9.1 Introduction

Combinational Circuit is the type of circuit in which output is independent of time and only relies on the input present at that particular instant. On other hand Sequential circuit is the type of circuit where output not only relies on the current input but also depends on the previous output.

RS flip flops find uses in many applications in logic or digital electronic circuitry. They provide a simple switching function whereby a pulse on one input line of the flip flop sets the circuit in one state. Further pulses on this line have no effect until the R-S flip flop is reset. This is accomplished by a pulse on the other input line. In this way the R S flip flop is toggled between two states by pulses on different lines.

A D-type flip-flop is a clocked flip-flop which has two stable states. A D-type flip-flop operates with a delay in input by one clock cycle. Thus, by cascading many D-type flip-flops delay circuits can be created, which are used in many applications such as in digital television systems.

The JK flip flop is basically a gated SR flip-flop with the addition of a clock input circuitry that prevents the illegal or invalid output condition that can occur when both inputs S and R are equal to logic level "1".

The purpose of the Master-Slave is to overcome from "Race-around condition". Repetition of toggle for a single clock pulse in the input of J-K FF is known as the Race-around condition. The Race-around condition will occur when J=K=1 and the time period of the clock pulse is greater than the propagation delay of FF. The master-slave configuration has the advantage of being edge-triggered, making it easier to use in larger circuits, since the inputs to a flip-flop often depend on the state of its output.

Registers are a type of computer memory used to quickly accept, store, and transfer data and instructions that are being used immediately by the CPU. The registers used by the CPU are often termed as Processor registers.

It is used to hold the memory addresses of data and instructions. It accesses data and instructions from memory during the execution phase of an instruction. For example, the CPU wants to store some data in the memory or to read the data from the memory. It places the address of the-required memory location in the MAR.

What is the use of counter in microcontroller?

A counter is a device that stores (and sometimes displays) the number of times a particular event or process occurred, with respect to a clock signal. It is used to count the events happening outside the microcontroller.

Counters are used not only for counting but also for measuring frequency and time ; increment memory addresses . Counters are specially designed synchronous sequential circuits, in which , the state of the counter is equal to the count held in the circuit by the flip flops.

Memory is the power of the brain to recall past experiences or information. In this faculty of the mind, information is encoded, stored, and retrieved. In the broadest sense, there are three types of memory: sensory memory, short-term memory, and long-term memory. Memory has three basic functions: encoding, storing, and retrieving information.

Examples of non-volatile memory include read-only memory, flash memory, most types of magnetic computer storage devices (e.g. hard disk drives, floppy disks and magnetic tape), optical discs, and early computer storage methods such as paper tape and punched cards.

A converter that is used to change the analog signal to digital is known as an analog to digital converter or ADC converter. This converter is one kind of integrated circuit or IC that converts the signal directly from continuous form to discrete form. This converter can be expressed in A/D, ADC, A to D. The inverse function of DAC is nothing but ADC.

A D/A Converter (DAC) convert a digital input signal into an analog output signal. The digital signal is represented with a binary code, which is a combination of bits 0 and 1.

They are also used in televisions and mobile phones to convert digital video data into analog video signals. These two applications use DACs at opposite ends of the frequency/resolution trade-off.

9.2 Objectives:

After studying this unit you should be able to

- Study and identify Difference from combinational circuit.
- Explain and identify Flip-flops (RS, D, JK) master slave.
- Study and identify Register (function and types).
- Explain and identify Counter (function and types).
- Study and identify Memory (function and types).
- Explain and identify Convertors (A/D and D/A).

9.3 Difference from combinational circuit:

Definition of Combinational Circuit: A combinational circuit is built up of the linked collection of several gates which produces output specific to the input at that instant. The basic AND, OR and NOT or universal gates NAND and NOR are the fundamental building blocks of the combinational circuit. As shown in the below-given diagram of the combinational circuit the output lines immediately follow the input lines. One of the typical examples of the combinational circuit is the decoder which is used to transform the binary code data into the decimal code data.

Fig.9.1 Combinational circuit

In these circuits, the output generated at that time would depend on the input of the particular time. There are three variations of the combinational logic circuits – arithmetic and logical functions, data transmission and code converters. The circuits included in the arithmetic and logic circuits are the adders, subtractors, comparators, PLDs, etcetera. Similarly, data transmission circuits are multiplexers, demultiplexers, encoders, decoders and so on. BCD and 7 segments are the code converter circuits.

Commonly, a combination circuit comprised of n number of binary inputs and m number of binary outputs. It implements the crucial functions of a digital computer.

The significant characteristics of the combinational circuit

- **Truth table**: It generates the 'm' number of binary output signals for the set of 2^n input signals.
- **Graphical symbol**: Shows the linked layout of the gates.
- **Boolean equations**: The output signals are expressed in the form of the boolean function of input signals.

Definition of Sequential Circuit: A sequential circuit is a class of circuits where the outcome depends on both the present input and past outputs. The characteristic of this circuit is that the state of output changes according to the sequence of the input has been inserted. This means that the sequential circuits contain an amount of memory to store the immediate results.

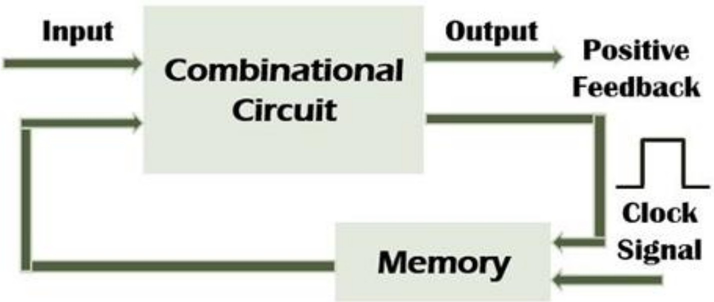


Fig.9.2 Sequential circuit

For example, it can remember which logic level 0 or 1 is connected to its input and employ the fact at the output also. This memory device can be made up of the simple OR gate.
There are various devices which implement sequential circuits such as latches, flip-flop, and registers. These inputs changes from one of the two states. The sequential circuits are categorized in two ways, **synchronous** and **asynchronous** circuits. The circuit is considered as synchronous when the internal state of the machine alters at the particular time driven by a clock.

What are the different types of sequential logic circuits?

1. **Synchronous sequential circuits**– The same clock input synchronize all the memory elements as in synchronous counters.

2. **Asynchronous sequential circuits**– An external clock is absent. However, the clock inputs receive pulse inputs from other sources/elements in the circuit — for example, Asynchronous counters.

Difference between Combinational and Sequential Circuit:

The crucial difference between combinational and sequential circuit is that combinational circuit result only relies on the input present at that instant while in the sequential circuit the output of the logic not just depends on the latest input but also on the earlier outputs. There is no feedback is provided to the combinational logic circuit but when it comes to the sequential logic circuit the feedback is the essential part of the circuit. So, the output generated considers both the present and past outputs.

Comparison Chart:

Basis For Comparison	Combinational Circuit	Sequential Circuit
Basic	The output is discovered by the present state of the inputs.	Both the present input and past state output are used to identify the output.
Storage capability	Does not store data.	Can store a small amount of data.
Application	Used in adders, encoders, multiplexer, etcetera.	Flip-flop and latches.
Clock	Circuits do not rely on the clock.	Clock is utilized for performing triggering functions.
Feedback	No requirement of the feedback.	Feedback is required.

9.4 Flip-flops (RS, D, JK) master slave:

RS Flip Flop:

A Flip Flop is a bi-stable device. There are three classes of flip flops they are known as Latches, pulse-triggered flip-flop, Edge- triggered flip flop. In this set word means that the output of the circuit is equal to 1 and the word reset means that the output is 0. The RS Flip Flop is considered as one of the most basic sequential logic circuits. The Flip Flop is a one-bit memory bi-stable device.

It has two inputs, one is called “SET” which will set the device (output = 1) and is labeled S and another is known as “RESET” which will reset the device (output = 0) labeled as R. The RS stands for SET/RESET.

The flip-flop is reset back to its original state with the help of RESET input and the output is Q that will be either at logic level “1” or logic”0”. It depends upon the set/reset condition of the flip-flop. Flip flop word means that it can be “FLIPPED” into one logic state or “FLOPPED” back into another.

The basic NAND gate RS Flip Flop circuit is used to store the data and thus provides feedback from both of its outputs again back to its inputs. The RS Flip Flop actually has three inputs, SET, RESET and its current output Q relating to its current state.

The symbol of the RS Flip-Flop is shown below:

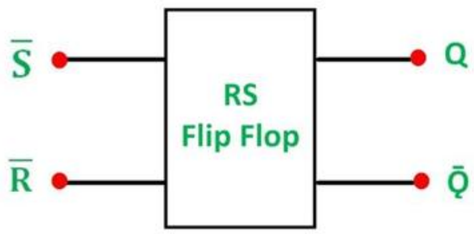


Fig.9.3 Symbol of the RS Flip-Flop

The NAND Gate RS Flip Flop:

A pair of cross-coupled 2 unit NAND gates is the simplest way to make any basic one-bit set/reset RS Flip Flop. It forms Set/Reset bi-stable or an active LOW RS NAND gate latch. The feedback is fed from each output to one of the other NAND gate input.

The device consists of two inputs; one is known as SET, (S) and the other is called as RESET, (R).

The two outputs are Q and Q bar as shown in the figure below:

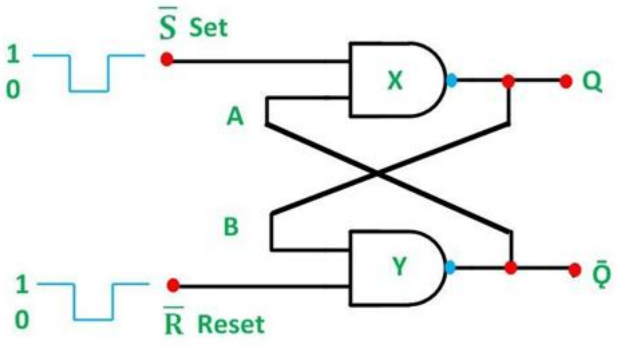


Fig.9.4 NAND Gate RS Flip Flop circuit

The Set State:

Considering the above circuit, if the input R is at logic level "0" (R = 0) and input S is at the logic "1" (S = 1), the NAND gate Y has, at least, one of its inputs at a logic "0". Therefore, its output Q must be at a logic level "1" (NAND gate principles). The Output (Q) is fed back to the input "A". Both the inputs of the NAND gates X are at logic "1", and therefore, its output Q must be at the logic level"0".

The reset input R changes its state, and goes HIGH to logic "1" with S constant at logic "1". The NAND gate Y input are now (R = 1) and (B = 0). The output at Q remains at HIGH or at logic level "1" as one of its inputs is still at logic level "0".

As a result, there is no change in state. Therefore, the flip-flop circuit is said to be "LATCHED" or "SET" with Q = 1 and Q̄ = 0.

The Reset State: In this second stable state, Q is at logic level '0" and its inverse output Q is at logic level "1". And is given by (R = 1) and (S = 0). As gate X has one of its inputs at logic "0" its output Q must equal logic level "1" (According to the NAND gate principle). The output Q is fed to input B, so both the inputs to NAND gate Y are at logic "1"., therefore, Q = 0.

If the set input S now changes the state to logic "1" with the input R remaining at logic "1", the output Q still remains LOW at logic level "0". And there is no change in the state.

Therefore, the flip-flop circuits "RESET" state has been latched.

The truth table of the Set/Reset is given below:

State	**S**	**R**	**Q**	**Q̄**	**Description**

SET	1	0	1	0	Set Q >>1
	1	1	1	0	No Change
RESET	0	1	0	1	Reset Q >>0
	1	1	0	1	No Change
INVALID	0	0	0	1	Memory with Q = 0
	0	0	1	0	Memory with Q = 1

From the truth table, it is clear that when both the inputs S = 1 and R =1 the outputs Q, and $\bar{Q}$ can be at either logic level '1' or "0" depending upon the state of the inputs.

When the input state R = 0 and S = 0 is an invalid condition and must be avoided because this will give both outputs Q and $\bar{Q}$ at logic level "1" at the same time and the necessary condition is that Q to be the inverse of $\bar{Q}$.

The flip-flop goes to an unstable state as both the output goes LOW. This unstable condition arises when the LOW input is switched to HIGH. The flip-flop switches to one state or the other and any one output of the flip-flop switches faster than the other. This unstable condition is known as Meta- stable state.

The bistable RS flip flop is activated or set at logic "1" applied to its S input and deactivated or reset by a logic "1" applied to R. The RS flip-flop is said to be in an invalid condition if both the set and reset inputs are activated simultaneously.

The NOR Gate RS Flip Flop: The circuit diagram of the NOR gate flip-flop is shown in the figure below:

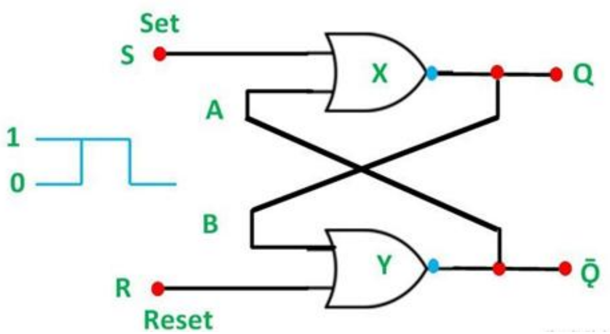


Fig.9.5 NOR Gate RS Flip Flop circuit

A simple one bit RS Flip Flops are made by using two cross-coupled NOR gates connected in the same configuration. The circuit will work similar to the NAND gate circuit.

The truth table of the NOR gate RS Flip Flop is shown below:

S	R	Q	Q̄
0	0	No Change	No Change
0	1	0	1
1	0	1	0
1	1	0	0

The inputs are active HIGH and the invalid condition exists when both its inputs are at logic level ‘1’.

D Flip Flop: The D Flip Flop is by far the most important of all the clocked flip-flops. By adding an inverter (NOT gate) between the Set and Reset inputs, the S and R inputs become complements of each other ensuring that the two inputs S and R are never equal (0 or 1) to each other at the same time allowing us to control the toggle action of the flip-flop using one single D (Data) input.

Then this Data input, labelled “D” and is used in place of the “Set” signal, and the inverter is used to generate the complementary “Reset” input thereby making a level-sensitive D-type flip-flop from a level-sensitive SR-latch as now S = D and R = not D as shown.

D-type Flip-Flop Circuit:

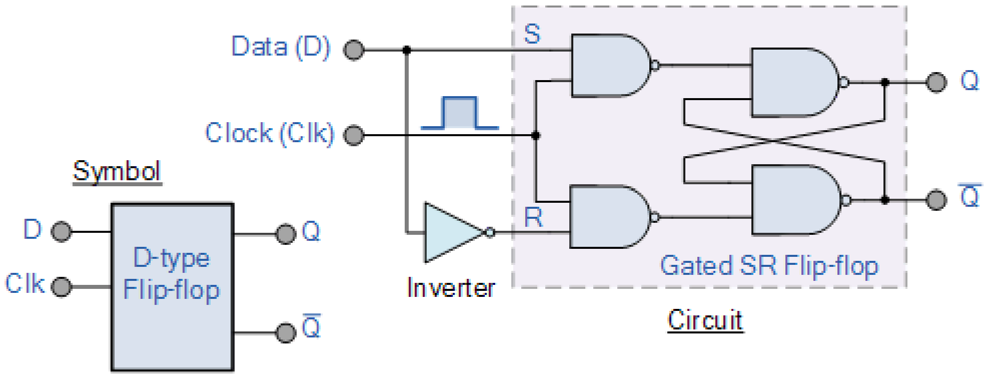


Fig.9.6 D-type Flip-Flop Circuit

We remember that a simple SR flip-flop requires two inputs, one to “SET” the output and one to “RESET” the output. By connecting an inverter (NOT gate) to the SR flip-flop we can “SET” and “RESET” the flip-flop using just one input as now the two input signals are complements of each other. This complement avoids the ambiguity inherent in the SR latch when both inputs are LOW, since that state is no longer possible.

Thus this single input is called the “DATA” input. If this data input is held HIGH the flip flop would be “SET” and when it is LOW the flip flop would change and become “RESET”.

However, this would be rather pointless since the output of the flip flop would always change on every pulse applied to this data input.

To avoid this an additional input called the "CLOCK" or "ENABLE" input is used to isolate the data input from the flip flop's latching circuitry after the desired data has been stored. The effect is that D input condition is only copied to the output Q when the clock input is active. This then forms the basis of another sequential device called a D Flip Flop**.**

The "D flip flop" will store and output whatever logic level is applied to its data terminal so long as the clock input is HIGH. Once the clock input goes LOW the "set" and "reset" inputs of the flip-flop are both held at logic level "1" so it will not change state and store whatever data was present on its output before the clock transition occurred. In other words the output is "latched" at either logic "0" or logic "1".

Truth Table for the D-type Flip Flop:

Clk	D	Q	Q	Description
↓ » 0	X	Q	Q	Memory no change
↑ » 1	0	0	1	Reset Q » 0
↑ » 1	1	1	0	Set Q » 1

Note that: ↓ and ↑ indicates direction of clock pulse as it is assumed D-type flip flops are edge triggered

JK Flip Flop: The JK Flip Flop is the most widely used flip flop. It is considered to be a universal flip-flop circuit. The sequential operation of the JK Flip Flop is the same as for the RS flip-flop with the same SET and RESET input.

The difference is that the JK Flip Flop does not the invalid input states of the RS Latch (when S and R are both 1). The JK Flip Flop name has been kept on the inventor name of the circuit known as Jack Kilby.

The basic symbol of the JK Flip Flop is shown below:

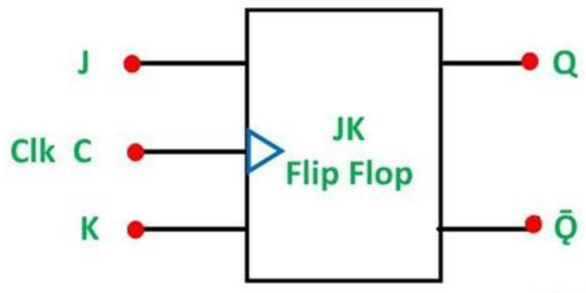


Fig.9.7 Symbol of the JK Flip Flop

The basic NAND gate RS flip-flop suffers from two main problems.

- Firstly, the condition when S = 0 and R = 0 should be avoided.
- Secondly, if the state of S or R changes its state while the input which is enabled is high, the correct latching action does not occur.

Thus to overcome these two problems of the RS Flip-Flop, the JK Flip Flop was designed.

The JK Flip Flop is basically a gated RS flip flop with the addition of the clock input circuitry. When both the inputs S and R are equal to logic "1", the invalid condition takes place.

Thus, to prevent this invalid condition, a clock circuit is introduced. The JK Flip Flop has four possible input combinations because of the addition of the clocked input. The four inputs are "logic 1", 'logic 0". "No change' and "Toggle".

The circuit diagram of the JK Flip Flop is shown in the figure below:

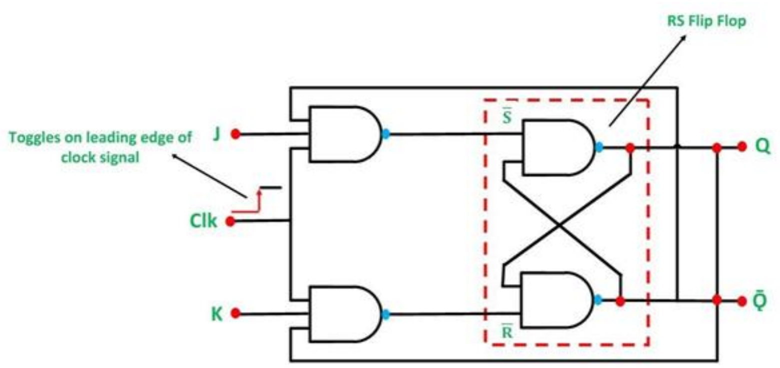


Fig.9.8 Circuit diagram of the JK Flip Flop

The S and R inputs of the RS bistable have been replaced by the two inputs called the J and K input respectively.

Here J = S and K = R. The two-input AND gates of the RS flip-flop is replaced by the two 3 inputs NAND gates with the third input of each gate connected to the outputs at Q and $\bar{Q}$. This cross-coupling of the RS Flip-Flop is used to produce toggle action. As the two inputs are interlocked.

If the circuit is in the “SET” condition, the J input is inhibited by the status 0 of Q through the lower NAND gate. Similarly, the input K is inhibited by 0 status of Q through the upper NAND gate in the “RESET” condition.

When both J and K are at logic “1”, the JK Flip Flop toggle.

The Truth Table of the JK Flip Flop is shown below.

	J	**K**	**Q**	**$\bar{Q}$**	**Description**
Same as for the RS Latch	0	0	0	0	Memory No Change
	0	0	0	1	
	0	1	1	0	Reset Q >> 0
	0	1	0	1	
	1	0	0	1	Set Q >> 1
	1	0	1	0	
Toggle	1	1	0	1	Toggle
	1	1	1	0	

JK Flip Flop is similar to RS flip flop with the feedback which enables only one of its input terminals. It eliminates the invalid condition which arises in the RS flip flop and put the input terminal either to set or reset condition one at a time.

When both the J and K inputs are at logic "1" at the same time and the clock input is pulsed HIGH, the circuit toggle from its SET state to a RESET or visa versa. When both the terminals are HIGH the JK flip-flop acts as a T type toggle flip-flop.

JK flip-flop has a drawback of timing problem known as "RACE". The condition of RACE arises if the output Q changes its state before the timing pulse of the clock input has time to go in OFF state.

The timing pulse period (T) should be kept as short as possible to avoid the problem of timing.

This condition is not possible always thus a much-improved flip-flop named Master Salve JK Flip Flop was developed. This eliminates all the timing problems by using two RS flip-flop connected in series. One is for the "MASTER" circuit, which triggers on the leading edge of the clock pulse. The other is called the "SLAVE" circuit, which triggers when the clock pulse is at the falling edge.

Master Slave Flip Flop Definition: Master-slave is a combination of two flip-flops connected in series, where one acts as a master and another act as a slave. Each flip-flop is connected to a clock pulse complementary to each other, i.e., if the clock pulse is in high state, the master flip-flop is in enable state, and the slave flip-flop is in disable state, and if clock pulse is low state, the master flip-flop is in disable state, and the slave flip flop is enable state.

Master Slave Flip Flop Diagram: Assume that in the initial state Y=0 and Q=0, the next input is S=1 and R=0; during that transition, the master flip-flop is set and Y=1, there is no change in slave flip-flop as slave flip-flop is disabled by the inverted clock pulse, when the clock pulse of master changes to '0', then the information of Y passes through slave and Q=1, in this clock pulse the slave flip-flop is active and master flip-flop gates deactivated.

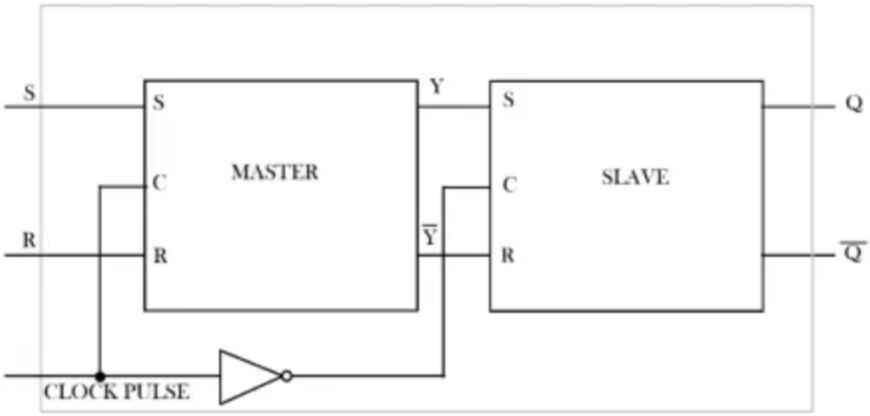


Fig.9.9 Master Slave flip flop logic diagram.

Master Slave Flip Flop Circuit | Master Slave Flip Flop Circuit Diagram:

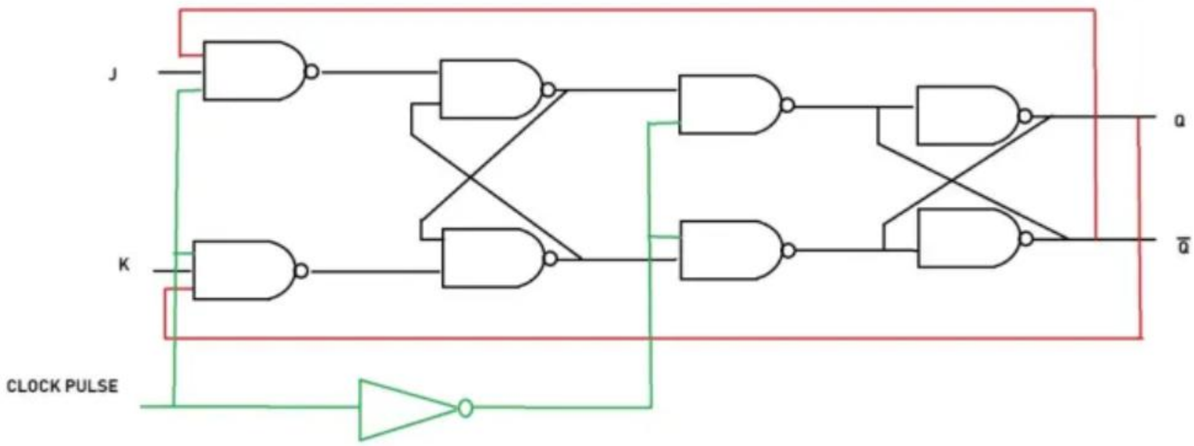


Fig.9.10 Clocked master slave JK flip flop

Master Slave Flip Flop Timing Diagram: The changes in input and output with respect to time can be defined in the timing diagram.

The behavior of a master-slave flip flop can be determined through a timing diagram. For example, in the given figure below, we can see a signal of the clock pulse, S is the input signal to the master flip flop, Y is the O/P signal of the master flip flop, and Q is the output signals of slave flip flop.

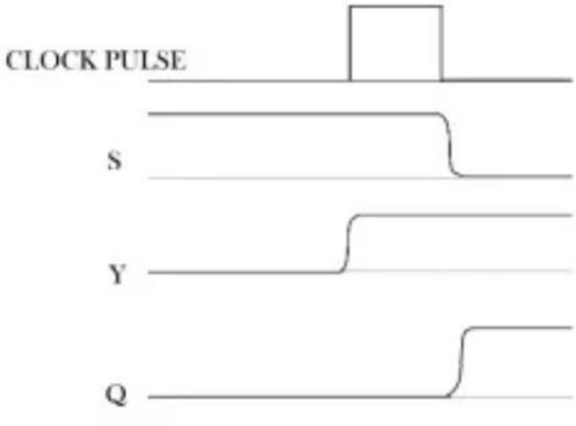


Fig.9.11 Time relationship of master slave flip-flop

Master Slave Flip Flop Truth Table: The truth table is a description of all possible output with all possible input combinations. In the master slave flip flop, there are two flip flops connected with inverted clock pulse to each other, so in the master slave truth table in addition to flip flop states, there must be an additional column for clock pulse so that the relationship between the input and output with the clock pulse can be determined.

Application of Master Slave Flip Flop:

Mater slave configuration is mainly used to eliminate the race around the condition and get rid of unstable oscillation in the flip flop.

Advantages of Master Slave Flip Flop:

Master slave can be operated on level triggered or edge triggered clock pulse; it can be used in various ways.

- A sequential circuit with an edge-controlled flip flop is straightforward to design rather than a level-triggered flip flop.
- By using the Master slave configuration, we also can eliminate the race around the condition.

RS Master Slave Flip Flop: Master slave is a configuration to prevent the unstable behavior of a flip flop; Here in RS master slave flip flop, two RS flip flop are connected to form master slave configuration, here flip flop is connected to a clock pulse inverted to each other; when the positive half of the clock pulse arrives the master flip flop is activated, and during negative clock pulse the slave flip flop is activated. Each flip flop works at different time interval.

In master salve configuration of RS flip flop, an unsalable oscillation cannot take place, because at a time master flip flop is in hold state or the slave flip flop is in hold state. For proper working of mater salve flip flop, we must consider hold time and setup time which can vary from one circuit to another; it depends on the design of the circuit.

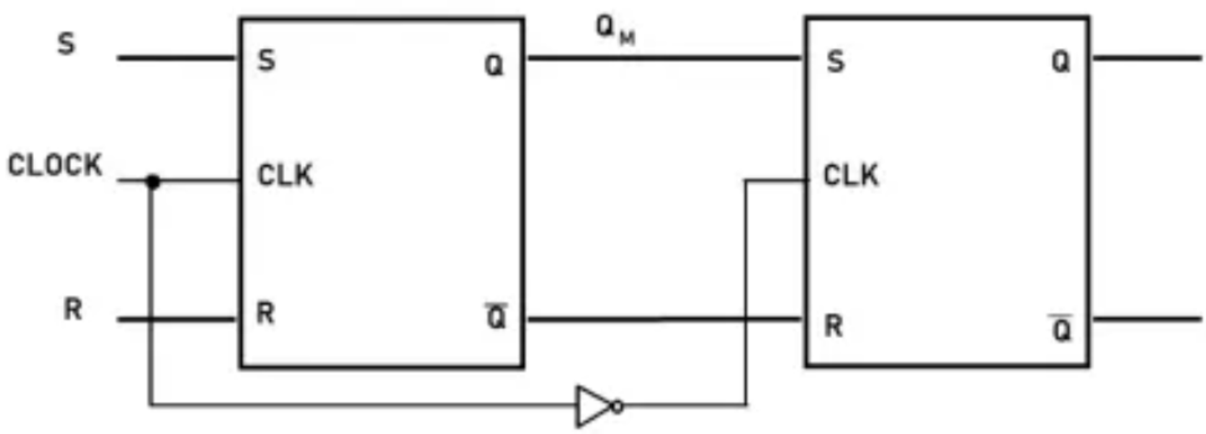


Fig.9.12 Block Representation of RS master slave flip flop

Master Slave SR Flip Flop Timing Diagram: Here, there is one clock signal, S is the input signal to the master flip flop, R is also an I/p signal to the master flip-flop, Qm is the O/P of the master flip-flop, Q if the O/P signal of the slave flip-flop.

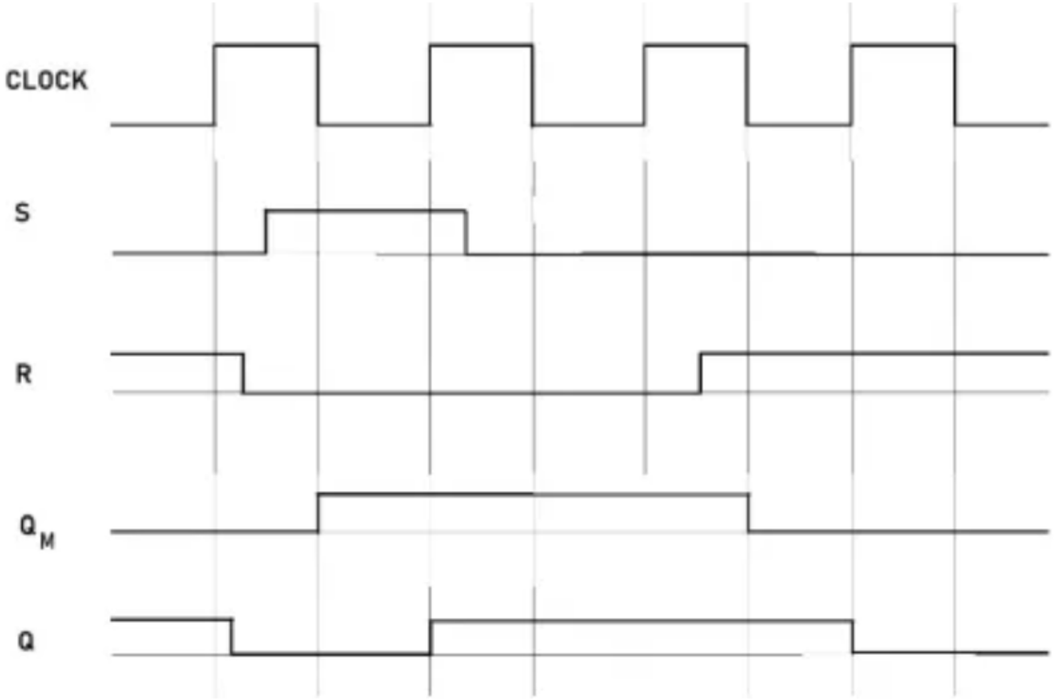


Fig.9.13 Timing Diagram of master slave SR flip flop.

Master Slave D Flip Flop: In this master slave also, two D flip flop connected to each other in series with clock pulse invited to each other. The basic mechanism of this master slave is also similar to other master slave flip-flops. D master slave flip flop can be level triggered, or edge triggered.

Master Slave D Flip Flop Circuit Diagram:

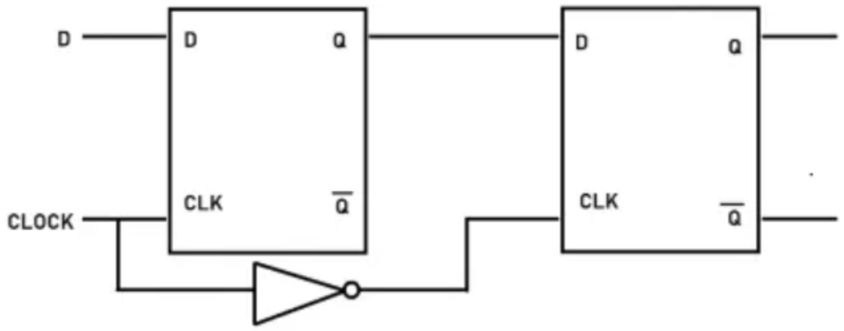


Fig.9.14 Block representation of master slave D flip flop circuit.

Master Slave D Flip Flop Timing Diagram: In the diagram, one signal of the clock pulse, one is D, the i/p to the master flip flop, Qm is the o/p of the master flip flop, and Q is the o/p of the slave flip flop.

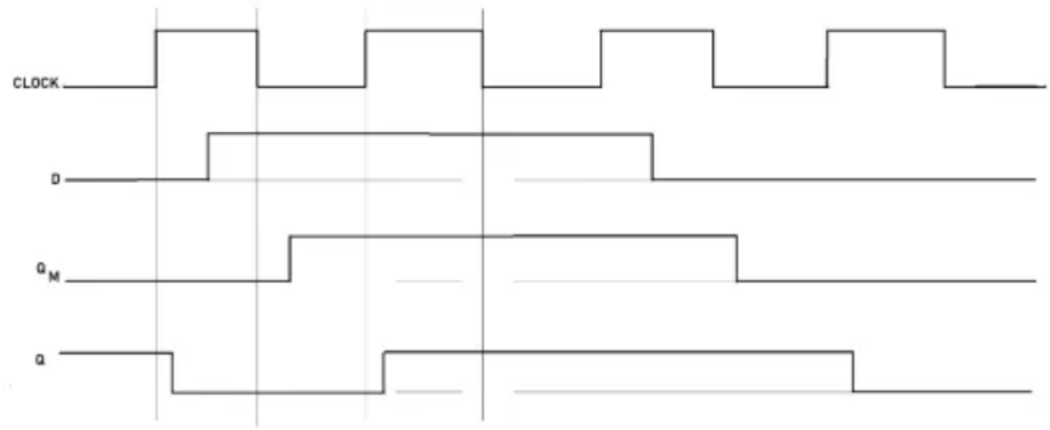


Fig.9.15 Master Slave D flip flop timing diagram

Master Slave JK Flip Flop: Master slave JK flip-flop could have been designed utilizing 2 JK flip-flops, in that each flip-flop is connected to CLK pulse complementary to each other, and the first flip flop is the master flip-flop which works when the CLK pulse is high state. And at that time the slave flip flop is in the hold state and if the CLK pulse is low state, then the slave flip-flop works, and the master flip-flop stays in the hold state.
The JK flip-flop characteristic is more or less similar to the SR flip-flop, but in SR flip flop, there is one uncertain output state when the S=1 and R =1, but in JK flip flop, when the J=1 and K=1, the flip flop toggles, that means the output state changes from its previous state.

JK Master Slave Flip Flop Circuit Diagram:

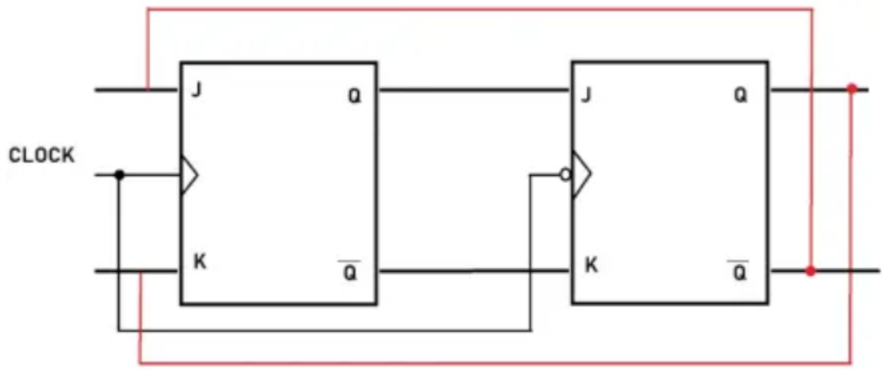


Fig.9.16 JK master salve block circuit diagram

JK Flip Flop Master Slave Timing Diagram:

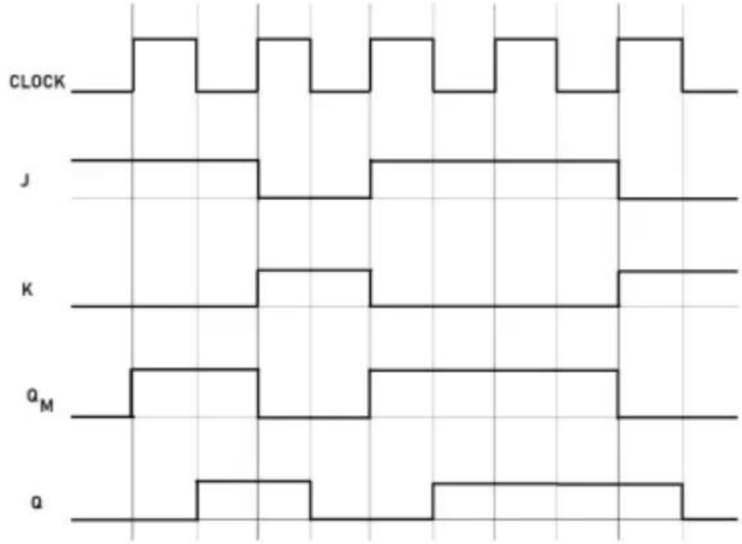


Fig.9.17 Timing Diagram for JK Master Slave flip flop

9.5 Register (function and types):

Register are used to quickly accept, store, and transfer data and instructions that are being used immediately by the CPU, there are various types of Registers those are used for various purpose. Among of the some Mostly used Registers named as AC or Accumulator**,** Data Register or DR, the AR or Address Register**,** program counter (PC), Memory Data Register (MDR), Index register**,** Memory Buffer Register.

These Registers are used for performing the various Operations. While we are working on the System then these Registers are used by the CPU for Performing the Operations**.** When We Gives Some Input to the System then the Input will be Stored into the Registers and When the System will gives us the Results after Processing then the Result will also be from the Registers. So that they are used by the CPU for Processing the Data which is given by the User. Registers Perform:-

1) **Fetch**: The Fetch Operation is used for taking the instructions those are given by the user and the Instructions those are stored into the Main Memory will be fetch by using Registers.

2) **Decode**: The Decode Operation is used for interpreting the Instructions means the Instructions are decoded means the CPU will find out which Operation is to be performed on the Instructions.

3) **Execute**: The Execute Operation is performed by the CPU. And Results those are produced by the CPU are then Stored into the Memory and after that they are displayed on the user Screen.

Types of Registers are as Followings:

MAR stand for Memory Address Register: This register holds the memory addresses of data and instructions. This register is used to access data and instructions from memory during the execution phase of an instruction. Suppose CPU wants to store some data in the memory or to read the data from the memory. It places the address of the-required memory location in the MAR**.**

Program Counter: The program counter (PC)**,** commonly called the instruction pointer (IP) in Intel x86 microprocessors, and sometimes called the instruction address register, or just part of the instruction sequencer in some computers, is a processor register It is a 16 bit special function register in the 8085 microprocessor. It keeps track of the the next memory address of the instruction that is to be executed once the execution of the current instruction is completed. In other words, it holds the address of the memory location of the next instruction when the current instruction is executed by the microprocessor**.**

Accumulator Register: This Register is used for storing the Results those are produced by the System. When the CPU will generate Some Results after the Processing then all the Results will be Stored into the AC Register.

Memory Data Register (MDR): MDR is the register of a computer's control unit that contains the data to be stored in the computer storage (e.g. RAM), or the data after a fetch from the computer storage. It acts like a buffer and holds anything that is copied from the memory ready for the processor to use it. MDR hold the information before it goes to the decoder.

MDR which contains the data to be written into or readout of the addressed location. For example, to retrieve the contents of cell 123, we would load the value 123 (in binary, of course) into the MAR and perform a fetch operation. When the operation is done, a copy of the contents of cell 123 would be in the MDR. To store the value 98 into cell 4, we load a 4 into the MAR and a 98 into the MDR and perform a store. When the operation is completed the contents of cell 4 will have been set to 98, by discarding whatever was there previously.

The MDR is a two-way register. When data is fetched from memory and placed into the MDR, it is written to in one direction. When there is a write instruction, the data to be written is placed into the MDR from another CPU register, which then puts the data into memory.

The Memory Data Register is half of a minimal interface between a micro program and computer storage, the other half is a memory address register.

Index Register: A hardware element which holds a number that can be added to (or, in some cases, subtracted from) the address portion of a computer instruction to form an effective address. Also known as base register**.** An index register in a computer's CPU is a processor register used for modifying operand addresses during the run of a program.

Memory Buffer Register: MBR stand for Memory Buffer Register**.** This register holds the contents of data or instruction read from, or written in memory. It means that this register is used to store data/instruction coming from the memory or going to the memory.

Data Register: A register used in microcomputers to temporarily store data being transmitted to or from a peripheral device.

Shift register: The group of flip-flops, which are used to hold store the binary data is known as register. If the register is capable of shifting bits either towards right hand side or towards left hand side is known as shift register. An 'N' bit shift register contains 'N' flip-flops. Following are the four types of shift registers based on applying inputs and accessing of outputs.

- Serial In − Serial Out shift register
- Serial In − Parallel Out shift register
- Parallel In − Serial Out shift register
- Parallel In − Parallel Out shift register

Serial In − Serial Out SISO Shift Register: The shift register, which allows serial input and produces serial output is known as Serial In – Serial Out SISO shift register. The block diagram of 3-bit SISO shift register is shown in the following figure.

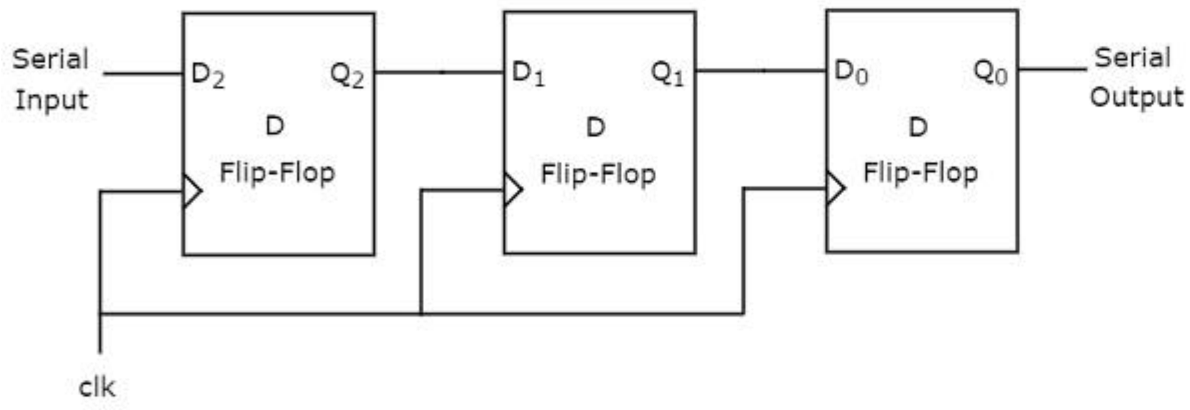


Fig.9.18 Block diagram of 3-bit SISO shift register

This block diagram consists of three D flip-flops, which are cascaded. That means, output of one D flip-flop is connected as the input of next D flip-flop. All these flip-flops are synchronous with each other since, the same clock signal is applied to each one.

In this shift register, we can send the bits serially from the input of left most D flip-flop. Hence, this input is also called as serial input. For every positive edge triggering of clock signal, the data shifts from one stage to the next. So, we can receive the bits serially from the output of right most D flip-flop. Hence, this output is also called as serial output.

Serial In - Parallel Out SIPO Shift Register: The shift register, which allows serial input and produces parallel output is known as Serial In – Parallel Out SIPO shift register. The block diagram of 3-bit SIPO shift register is shown in the following figure.

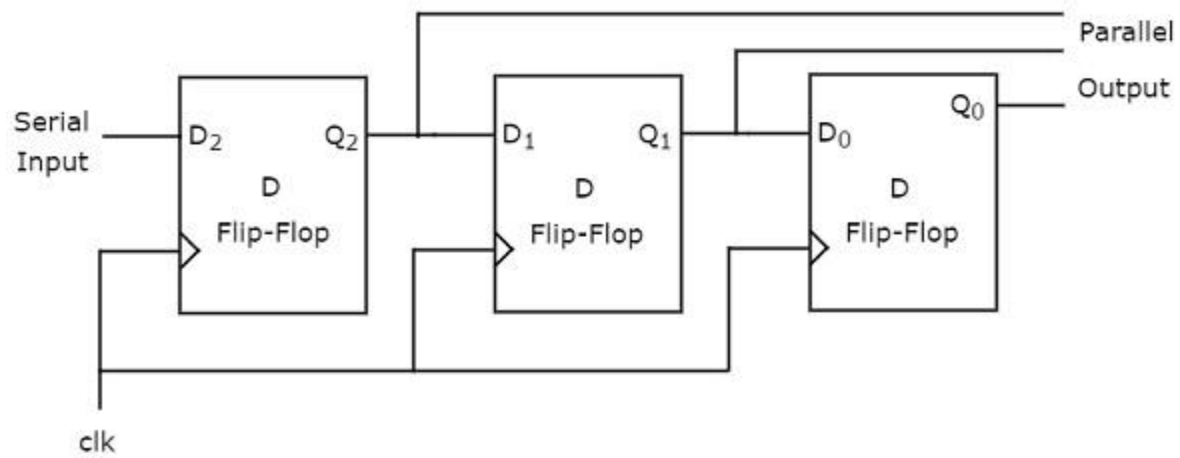


Fig.9.19 Block diagram of 3-bit SIPO shift register

This circuit consists of three D flip-flops, which are cascaded. That means, output of one D flip-flop is connected as the input of next D flip-flop. All these flip-flops are synchronous with each other since, the same clock signal is applied to each one.

In this shift register, we can send the bits serially from the input of left most D flip-flop. Hence, this input is also called as serial input. For every positive edge triggering of clock signal, the data shifts from one stage to the next. In this case, we can access the outputs of each D flip-flop in parallel. So, we will get parallel outputs from this shift register.

Parallel In − Serial Out PISO Shift Register: The shift register, which allows parallel input and produces serial output is known as Parallel In − Serial Out PISO shift register. The block diagram of 3-bit PISO shift register is shown in the following figure.

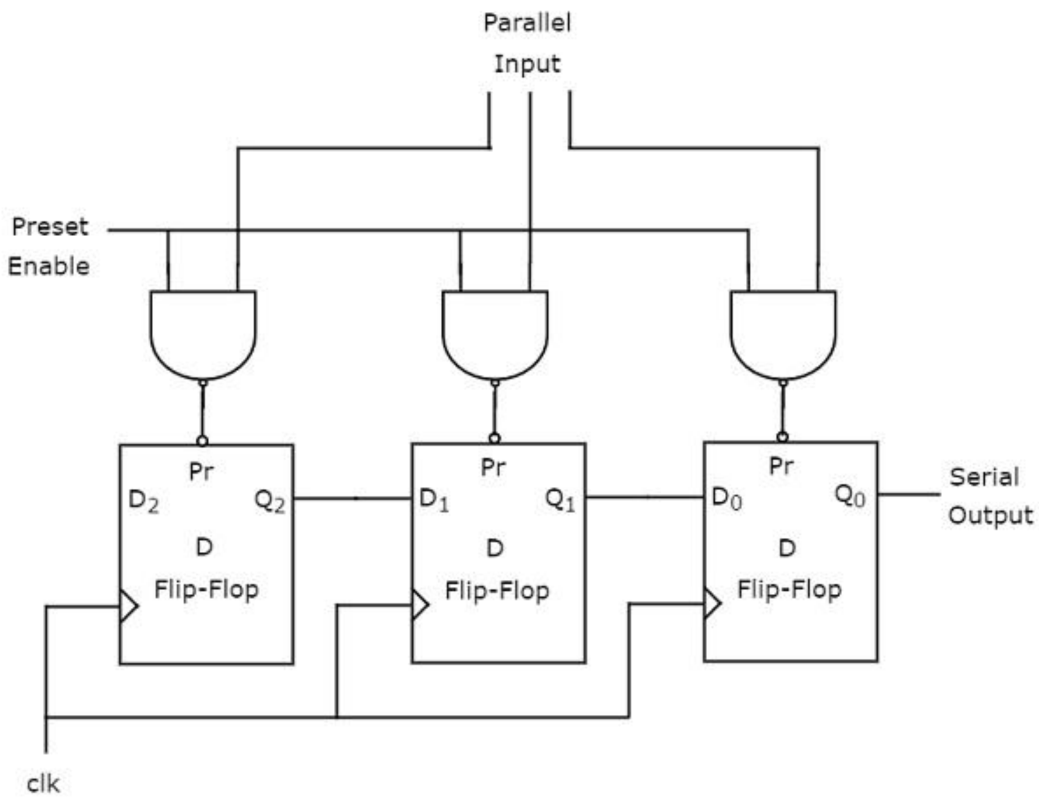


Fig.9.20 Block diagram of 3-bit PISO shift register

This circuit consists of three D flip-flops, which are cascaded. That means, output of one D flip-flop is connected as the input of next D flip-flop. All these flip-flops are synchronous with each other since, the same clock signal is applied to each one.

In this shift register, we can apply the parallel inputs to each D flip-flop by making Preset Enable to 1. For every positive edge triggering of clock signal, the data shifts from one stage to the next. So, we will get the serial output from the right most D flip-flop.

Parallel In - Parallel Out PIPO Shift Register:

The shift register, which allows parallel input and produces parallel output is known as Parallel In − Parallel Out PIPO shift register. The block diagram of 3-bit PIPO shift register is shown in the following figure.

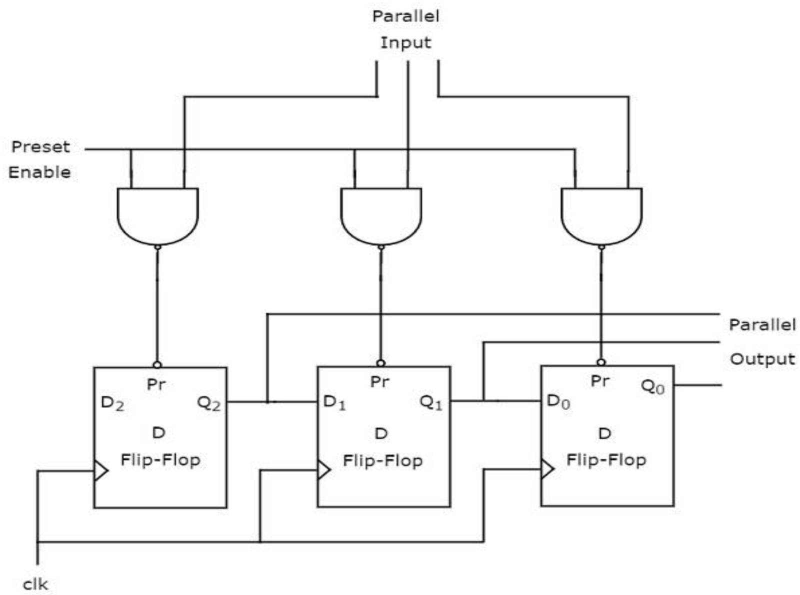


Fig.9.21 Block diagram of 3-bit PIPO shift register

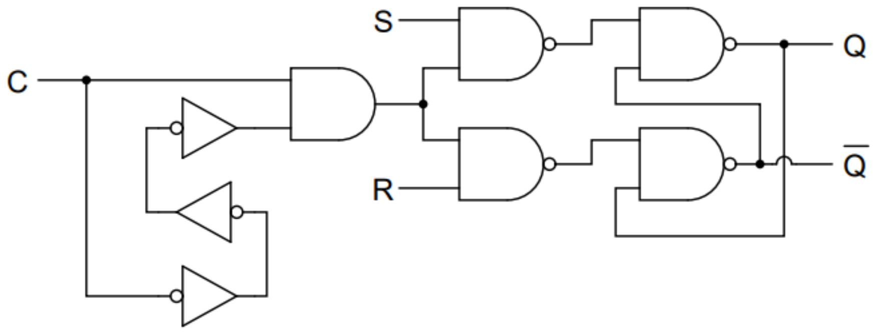
This circuit consists of three D flip-flops, which are cascaded. That means, output of one D flip-flop is connected as the input of next D flip-flop. All these flip-flops are synchronous with each other since, the same clock signal is applied to each one.

In this shift register, we can apply the parallel inputs to each D flip-flop by making Preset Enable to 1. We can apply the parallel inputs through preset or clear. These two are asynchronous inputs. That means, the flip-flops produce the corresponding outputs, based on the values of asynchronous inputs. In this case, the effect of outputs is independent of clock transition. So, we will get the parallel outputs from each D flip-flop.

SAQ.1

a) What do you mean by difference between Combinational circuit and Sequential circuits?
b) Discuss and explain Flip-flops RS, D and JK.
c) What do you mean by master slave in flip flop?
d) What do you mean by function of Register and its types?

e) Explain how the addition of a propagation-delay-based one-shot circuit to the enable input of an S-R latch changes its behavior:



Specifically, reference your answer to a truth table for this circuit.

9.6 Counter (function and types):

Introduction to Counters:

Counter is a digital device and the output of the counter includes a predefined state based on the clock pulse applications. The output of the counter can be used to count the number of pulses. Generally, counters consist of a flip-flop arrangement which can be synchronous counter or asynchronous counter. In synchronous counter, only one clock i/p is given to all flip-flops, whereas in asynchronous counter, the o/p of the flip flop is the clock signal from the nearby one. The applications of the microcontroller need counting of exterior events such as exact internal time delay generation and the frequency of the pulse trains. These events are frequently used in digital systems & computers. Both these events can be executed by software techniques, but software loops for counting will not give the exact result slightly more important functions are not done. These problems can be rectified by timers and counters in the microcontrollers which are used as interrupts.

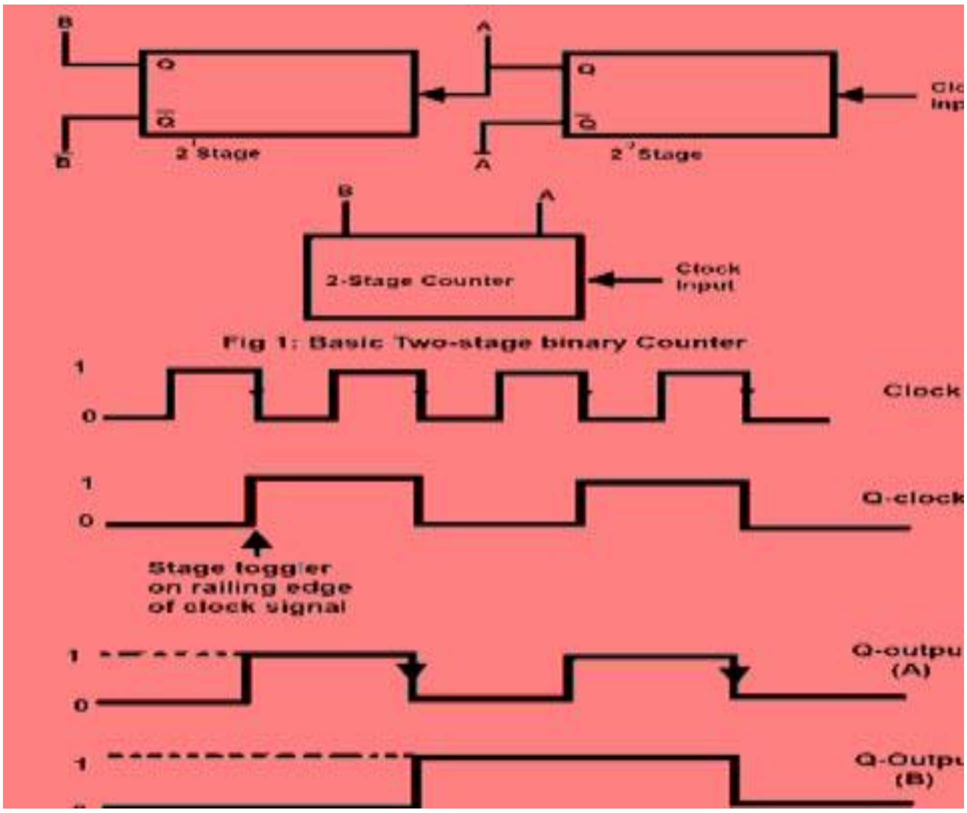


Fig.9.22 Output of the counter

Types of Counters: Counters can be categorized into different types according to the way they are clocked. They are:

- Asynchronous Counters
- Synchronous Counters
- Asynchronous Decade Counters
- Synchronous Decade Counters
- Asynchronous Up-Down Counters
- Synchronous Up-Down Counters

For better understanding of this type of counters, here we are discussing some of the counters.

Asynchronous Counters: The diagram of a 2-bit asynchronous counter is shown below. The exterior clock is connected to the clock i/p of the FF0 (first flip-flop) only. So, this FF changes the state at the decreasing edge of every clock pulse, but FF1 changes only when activated by the decreasing edge of the Q o/p of FF0. Because of the integral propagation delay through a FF, the change of the i/p clock pulse and a change of the Q o/p of FF0 can never occur at precisely the same time. So, the FF's cannot be activated concurrently, generating an asynchronous operation.

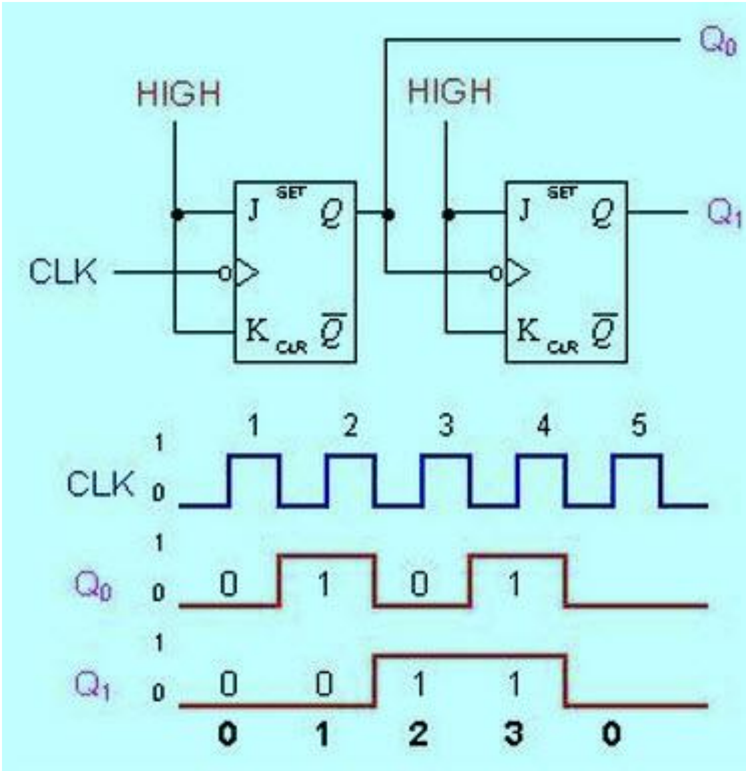


Fig.9.23 Asynchronous Counter

Note that for ease, the changes of Q0, Q1 & CLK in the above diagram are shown as concurrent, even though this is an asynchronous counter. Actually, there is a small delay b/n the Q0, Q1 and CLK changes.

Generally, all the CLEAR i/ps are connected together, so before counting starts then that a single pulse can clear all the FFs. The clock pulse fed into FF0 is rippled through the new counters after propagation delays, such as a ripple on the water, hence the term Ripple Counter.

The circuit diagram of the two bit ripple counter includes four different states ,each one consisting with a count value. Likewise, a counter with n FFs can have 2N states. The number of states in a counter is called as its mod number. Therefore a two-bit counter is a mod-4 counter.

Asynchronous Decade Counters: In the previous counter have 2n states. But, counters with states less than 2n is also possible. These are designed to have the no. of states in their series.These are called shortened sequences which are accomplished by driving the counter to recycle before going through all of its states. A common modulus for counters with shortened sequence is 10. A counter with 10-states in its series is called a decade counter.The implemented decade counter circuit is given below.

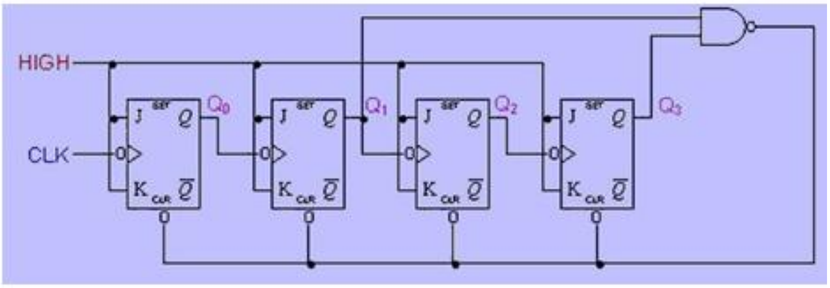


Fig.9.24 Asynchronous Decade Counter Circuit Diagram

When the counter counts to ten, then all the FFs will be cleared. Notice that only Q1&Q3 both are used to decode the count of 10, that is called partial decoding. At the same time one of the other states from 0-9 have both Q1&Q3 will be high. The series of the decade counter table is given below.

Clock Pulse	Q3	Q2	Q1	Q0
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1

Fig.9.25 Sequence of the Decade Counter

Asynchronous Up-Down Counters: In particular applications, a counter must be capable to count both up & down. The below circuit is a three bit up & down counter, that counts UP or DOWN based on the control signal status. When the UP i/p is at 1 & the DOWN i/p is at 0, the NAND gate between FF0 & FF1 will gate the non-inverted o/p (Q) of flip flop (FF0) into the clock i/p of flip flop (FF1). Likewise, the non-inverted o/p of Flip Flop1 will be gated through the other NAND gate into the clock i/p of flip-flop2. Therefore the counter will count up.

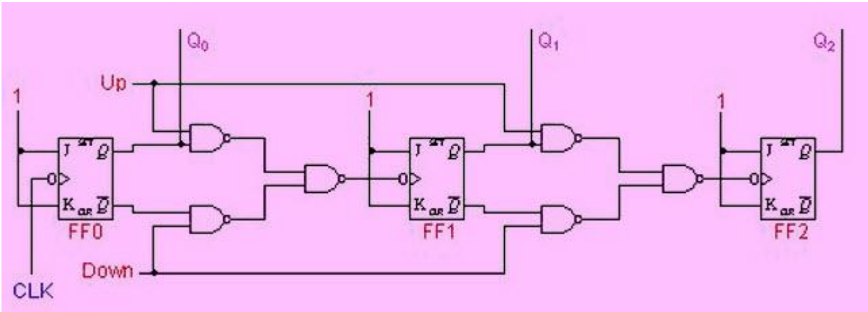


Fig.9.26 Asynchronous Up-Down Counter Circuit Diagram

Once the control i/p (UP) is at 0 & DOWN is at 1, the inverted o/ps of flip-flop0 (FF0) and flip-flop1 (FF) are gated into the clock i/ps of FF1 & FF2 separately. If the FFs are initially changed to 0's, then the counter will go through the below series as i/p pulses are applied. Notice that an asynchronous up-down counter is slower than an UP counter/down counter because of an extra propagation delay introduced by the NAND gates.

FF2	FF1	FF0
0	0	0
1	1	1
1	1	0
1	0	1
1	0	0
0	1	1
0	1	0
0	0	1

Fig.9.27 Sequence of the Asynchronous Up-Down Counter

Synchronous Counters: In this type of counters, the CLK i/ps of all the FFs are connected together and are activated by the i/p pulses. So, all the FFs change states instantaneously. The circuit diagram below is a three bit synchronous counter. The inputs J and K of flip-flop0 are connected to HIGH. Flip-flop 1 has its J &K i/ps connected to the o/p of flip-flop0 (FF0), and the inputs J & K of flip-flop2 (FF2) are connected to the o/p of an AND gate that is fed by the o/ps of flip-flop0 and flip-flop1. When the both the outputs of FF0 & FF1 are HIGH. The positive edge of the fourth CLK pulse will cause FF2 to alter its state because of the AND gate.

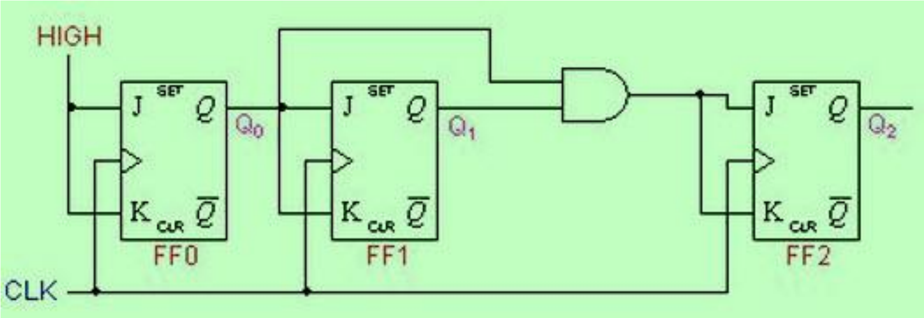


Fig.9.28 Synchronous Counter Circuit Diagram

The series of the three bit counter table is given below. The major advantage of these counters is that there is no increasing time delay due to all FFs are activated in parallel. Thus, the max operating frequency of this synchronous counter will be considerably higher than for the equivalent ripple counter.

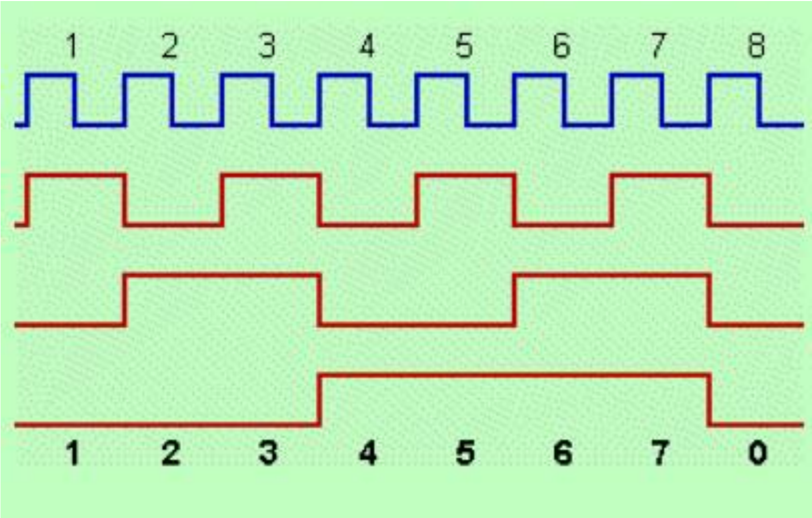


Fig.9.29 CLK Pulses of the Synchronous Counters

Synchronous Decade Counters: Synchronous counter counts from 0-9 similar to asynchronous counter and then again recycles zero. This process is done by driving the 1010 states back to the 0000 state. This is termed as truncated sequence, that can be designed by the below circuit.

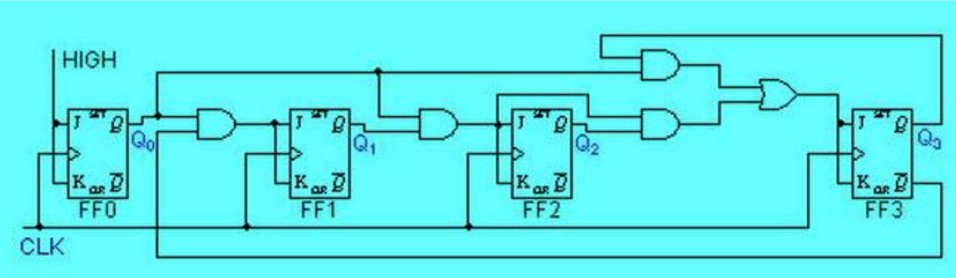


Fig.9.30 Synchronous Decade Counter Circuit Diagram

From the series on the left table, we can observe that

- Q0 ties on each and every CLK pulse
- Q1 alters on the next clock pulse every time when Q0=1 & Q3=0.
- Q2 alters on the next clock pulse every time when Q0=Q1=1.
- Q3 alters on the next CLK pulse each and every time when Q0=1, Q1=1 & Q2=1 (count 7), or when Q0=1 & Q3=1 (count 9).

Clock Pulse	Q3	Q2	Q1	Q0
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1

Fig.9.31 Sequence of the Synchronous Decade Counter

The above characteristics are employed with the AND gate or OR gate. The logic diagram of this is shown in the above diagram.

Synchronous Up-Down Counters: A three bit synchronous Up-Down counter, tabular form and series are given below. This type of counter has an up-down control i/p similar to asynchronous up-down counter, that is used to control the counter's direction through a certain series.

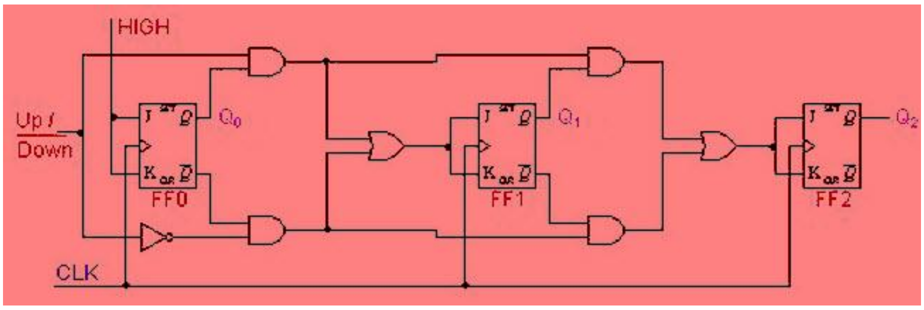


Fig.9.32 Synchronous Up-Down Counters Circuit Diagram

The series of the table shows

- Q0 ties on each CLK pulse for both up & down series
- When Q0=1 for the up series, then the state of the Q1 changes on the next CLK pulse.
- When Q0=0 for the down series, then the state of the Q1 changes on the next CLK pulse.
- When Q0=Q1=1 for the up series, then the state of the Q2 changes on the next CLK pulse.
- When Q0=Q1=0 for the down series, then the state of the Q2 changes on the next CLK pulse.

Clock Pulse	Q3	Q2	Q1	Q0
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1

Fig.9.33 Sequence of the Synchronous Decade Counters

The above characteristics are employed with the AND gate, OR gate and NOT gate. The logic diagram of this is shown in the above diagram.

Applications of Counters:

The applications of the counters mainly involve in digital clocks and in multiplexing. The best example of the counter is parallel to serial data conversion logic discussed below.

A set of bits, performing concurrently on parallel lines is called parallel data. A set of bits, performing on a single line in a time series is called serial data. The Parallel-to-serial data conversion is normally is done by using a counter to afford a binary series of the data, select i/ps of a MUX, as explained in the circuit below.

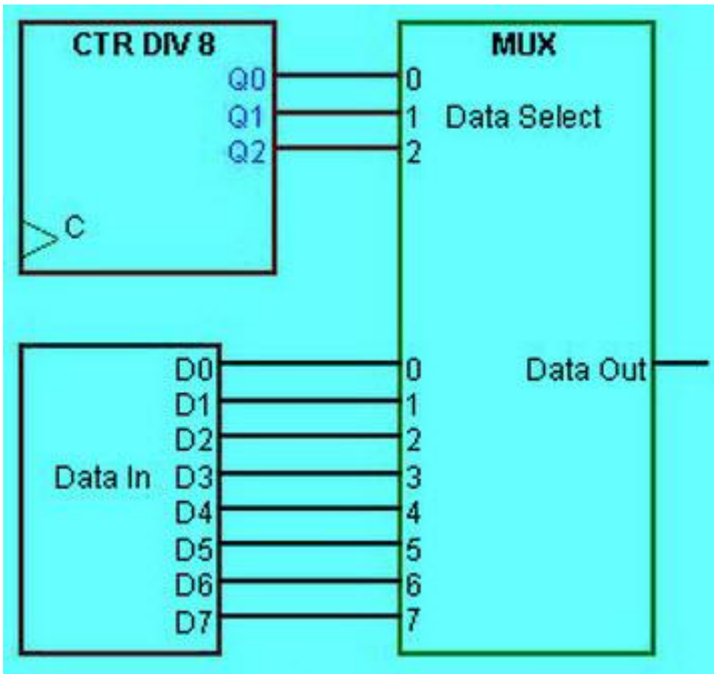


Fig.9.34 Parallel-to-Serial Data Conversion

In the above circuit, modulo-8 counter consist of Q o/ps, that are connected to the data, select i/ps of an 8-bit MUX. The first 8-bit group of parallel data is applied to the inputs of the MUX. As the counter goes through a binary series from 0-7, each bit starts with D0, is serially selected & passed through the MUX to the o/p line. After 8-CLK pulses, the data byte has been changed to a serial format & sent out through the transmission line. Then, the counter reprocesses back to 0 and changes another parallel byte serially again in the similar process.

9.7 Memory (function and types):

Just like a human brain, the digital devices such as computer, microcontroller & smartphone needs a space to store the information & instruction, this storage space is called memory or "computer memory" and the building block of memory is called a memory cell. In simple words, memory is an electronic circuit or device capable of storing information temporarily or permanently.

In computers, memory is the most essential component for the normal functioning of any system to store data, to perform calculations, to do complex operations, etc. We know that almost all our information and data is stored in the Hard Disk within the CPU. The Hard Disk/Hard Drive/Hard Disk Drive has the most extensive memory in the computer system.

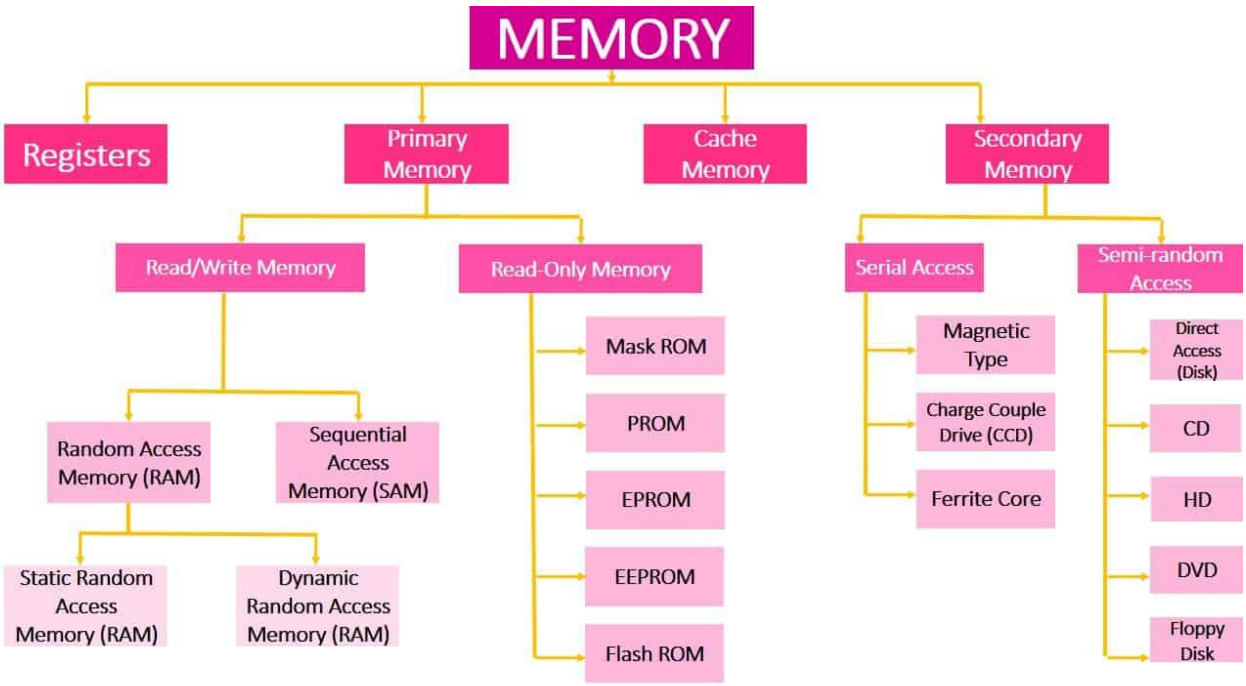


Fig.9.35 Memory Hierarchy

Types of Computer Memories:

The computer memory is classified into three main types; Cache Memory, Primary Memory and Secondary Memory.

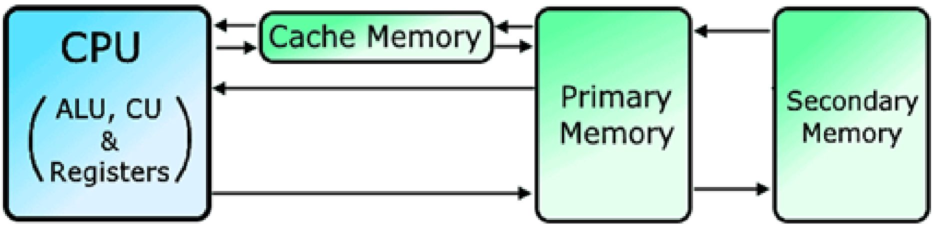


Fig.9.36 Cache Memory, Primary Memory and Secondary Memory

Cache Memory: The Cache memory is the fastest type of memory that acts as a buffer to store temporary data between the processor & data memory. It stores the necessary data frequently used by the CPU so that it can easily access it. It is the most expensive type of memory & it is integrated inside the CPU chip.

Primary Memory: Primary memory is also known as main memory or may also refer to "Internal memory" and primary storage. All those types of computer memories that are directly accessed by the processor using data bus are called primary memory. That allows a processor to access stores running programs and currently processed data that stored in a memory location.

The use of memories is therefore mandatory in all systems using a microprocessor, including computers. An example of Primary memory is RAM and ROM that store programs. These memories are limited in capacity and manufactured by using integrated circuits (IC) or semiconductor device. Its speed of Data accessing is faster than secondary memory. It is more expensive than secondary memory.

When you turn on the computer, Generally CPU searches for essential codes in RAM to get it. Otherwise, it goes to ROM. Yes, they both chips collectively called primary memory in a computer system.

The Primary memory is further classified into the following two types RAM and ROM.

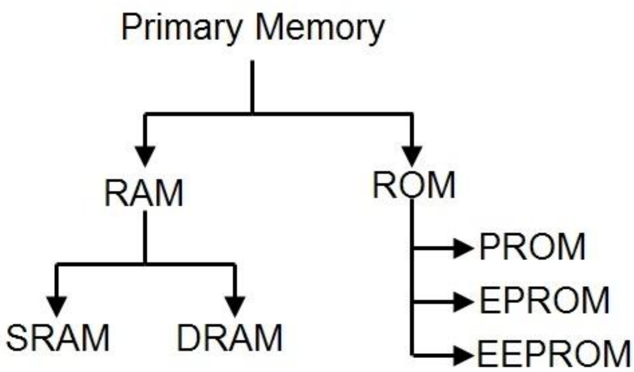


Fig.9.37 Classification of Primary memory

RAM (Random Access Memory): The RAM is an acronym for Random Access Memory & it is a type of primary memory that stores the information or data temporarily to be used by the processor while the system is running. It is also called volatile memory because the content of the data stored in its modules is erased when the power supply is interrupted. Thus they need a constant supply of power to retain that information.

DRAM (Dynamic RAM): DRAM or Dynamic RAM (Dynamic Random Access Memory) is a type of RAM where the computer memory cell used for storing a bit of data is made of capacitor & a transistor. Due to the leakage current of the capacitor, the cell cannot retain the information for too long. Therefore, the DRAM needs to be refreshed or recharge after a specific time to maintain the data.

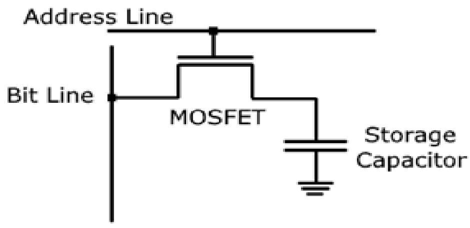


Fig.9.38 Dynamic Random Access Memory cell

Since the memory cell of DRAM is made from one transistor, more memory cells can be designed in a single chip to increase the storage density of DRAM. And the price per cell is reduced thus they are inexpensive.

Types of DRAM:

Synchronous DRAM (SDRAM): The SDRAM "synchronizes" the speed of the memory along with the CPU clock speed. By doing so, the memory controller (which is a digital circuitry managing the flow of data from and to the main memory) is aware of the exact clock cycle by which the demanded data will be ready. Thus, the CPU's efficiency is improved, and it can do many more instructions at a given time. A typical SDRAM works at speeds of up to 133 MHz.

Rambus DRAM (RDRAM): The Rambus DRAM is named after the company that introduced it, Rambus. It was mainly used for video game devices and on-computer graphics cards, having transfer speeds running up to 1 GHz.

Double Data Rate SDRAM (DDR SDRAM): This memory has nearly double the bandwidth of a single data rate (SDR) SDRAM. It works on the principle of "double pumping" – this permits data to be transferred on both the rising & falling edges of the clock. This type of memory has been succeeded by the DDR2, DDR3, DDR4 and most recently, the DDR5 SDRAM.

Advantages of DRAM:

Here are some of the advantages of DRAM

- Each memory cell is made of only one transistor, so its design is simple
- The bit per chip density is higher.
- The Cost per bit is relatively low
- The DRAM is inexpensive.

- Power consumption is low because the data is store in a capacitor.

Disadvantage of DRAM

- Here are some of the Disadvantages of DRAM;
- The memory is volatile.
- The cell needs to be refreshed at a constant rate to retain data.
- The refreshing consumes extra power.
- The circuitry required for refreshing is complex
- The DRAM is relatively slower in operation than SRAM

The DRAM is used as the main memory. It is an off-chip memory, placed externally on the motherboard.

Applications of a DRAM:

The DRAM is the main memory in computers and graphics cards. It is also used in many portable devices and video game consoles.

SRAM (Static RAM): The SRAM or Static RAM is another type of primary memory that stores it data using flip flop & latches. The flip flop is a memory cell of SRAM & it is made of multiple logic gates consisting transistors. There is no leakage current thus it does not need to be refreshed, unlike DRAM.

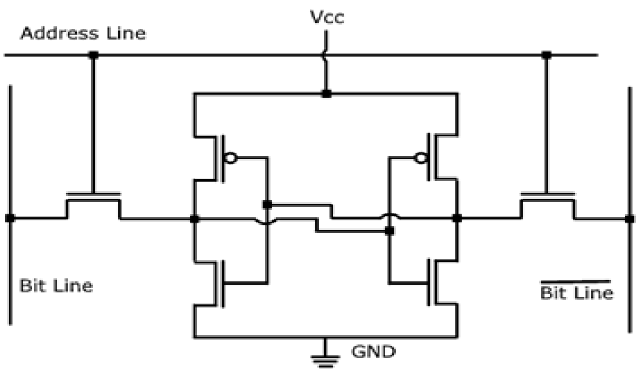


Fig.9.39 Static RAM memory cell

Each memory cell used for storing a single bit of data is made of 6 transistors. Thus price per bit of SRAM is more than DRAM & it requires more physical space on-chip to store the same amount of data. Therefore, the storage capacity of SRAM is lower than DRAM.

Advantage of SRAM:

Here are some of the advantages of SRAM over DRAM:

- There is no leakage current, thus it does not need to be refreshed.
- Due to no need for refreshing. The access time is faster.

- It is on-chip memory which also increases its operation speed
- It does not need extra circuitry for refreshing.
- It is relatively very faster than DRAM.

Disadvantages of SRAM:

- Each cell is made of 6 transistors which require more space than DRAM
- It is more expensive than DRAM.
- Due to large physical space for each bit, it has very low storage capacity.
- The memory is volatile i.e. the data is erased when there is no power.
- It consumes more power than DRAM due to a large number of transistors & constant power supply.

Applications of the SRAM:

Due to the high speed of operation, SRAM is used for cache memory and as part of the digital-to-analog converter on video cards. It is also found in CDs, printers, routers, DVDs and digital cameras.

ROM (Read-Only Memory): The ROM or Read-only Memory is a non-volatile type of primary memory i.e. the content of information stored does not get erased when the power is off. This data is stored during its manufacturing & it cannot be altered by any user. That is why it is called a Read-Only Memory**.**

The ROM is slower than RAM because the CPU cannot access its data. The data first need to be transferred to RAM where the CPU can access it. It only contains the necessary instruction regarding the startup of the system.

Characteristics of ROM:

- It is a non-volatile memory or permanent memory.
- It stores the instructions regarding the startup (& program code) of a microcontroller or computer.
- The CPU cannot directly access its data.
- It can be write once & read multiple times.
- It does not need the power to maintain its data.
- It has a very low capacity.
- It is slower than RAM
- It is comparatively cheaper than RAM

The ROM is classified into the following types are:

PROM (Programmable Read-Only Memory): PROM stands for Programmable Read-only Memory**,** which is a type of ROM where each memory cell is programmed after its manufacturing. Black PROMs are made during manufacturing which is then programmed during the production of devices they are being used for. The difference between ROM & PROM is that ROM is programed during manufacturing & PROM is programmed after its manufacturing.

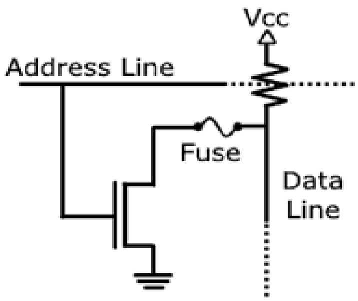


Fig.9.40 PROM Memory cell

The data in each cell is stored & maintained by a fuse. During manufacturing, all the fuses are intact & read "0". The fuse is blown using a high voltage pulse which then reads "1". In such a way the PROM is programmed by blowing the fuses for selected bit locations that should be "1" using a special programmer called PROM Programmer.

Since the blown or burned fuse cannot be restored, they can be programmed only once. And they can be read multiple times which is why it is a Read-Only Memory.

EPROM (Erasable Programmable Read-Only Memory): EPROM or also known as EROM is a modified type of PROM which is a non-volatile memory. Unlike PROM, the EPROM data can be erased once it's programmed by exposing it to a UV (ultraviolet) light source.

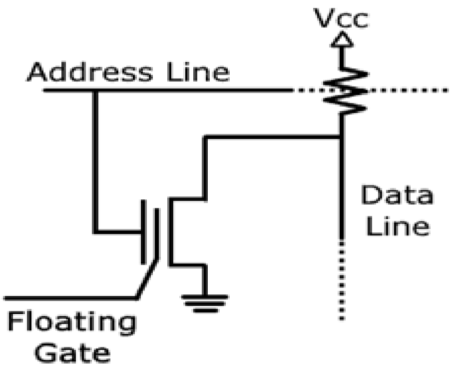


Fig.9.41 EPROM memory cell

The cells are made from a floating gate transistor which is programmed by supplying high voltage that creates an avalanche discharge of electrons to store it in the gate electrode. Due to the gate insulation, the charge does not leak & is stored permanently. This process cannot be reversed using electricity but exposing it to the UV light can dissipate the charge & erase the data inside.

The EPROM chips have a quartz window for UV light penetration & they are easily recognizable. Since the whole die is exposed, there is no way to erase any specific byte of data but the whole memory is wiped clean. It needs to be removed from the circuit & place under a UV lamp. It takes several minutes of exposure to completely erase the memory.

EEPROM (Electrically Erasable Programmable Read-Only Memory): EEROM stands for Electrically Erasable Programmable Read-Only Memory & it is a type of PROM whose individual bytes can be erased & reprogrammed. It is developed to replace the UV erase function of the EPROM & it can be erased & reprogrammed inside the circuit using a special programming circuit.

Due to single-byte operation, they are relatively slower while the flash memory which is a type of EEPROM is specially designed to have high speed. The flash memory has limited reprogramming cycles of around 10k while the latest models of EEPROM can have 1million cycles. The EEPROM is expensive than the flash memory but the flash memory does have a disadvantage of a large erase block.

Secondary Memory: Secondary memory is a broad term used for the computer memory that is used for storing data on a long-term basis. It is commonly known as the storage of a system. It is a non-volatile memory that can store data permanently without the need of power supply. The CPU does not have direct access to secondary memory. The data is first stored in primary memory RAM which is then accessed by CPU.

The secondary memory is also known as external memory. It has a huge amount of storage space with relatively low speed & cheaper than primary memory. It is used for backup storage.

The secondary memory is classified into fixed & removable memory;

Fixed Memory: This type of secondary memory is fixed internally in the system. Fixed memory means not be removed during the operation, it does not mean it is literally fixed. It can be removed using a proper tool kit for upgrading or replacing purposes. Here are some examples of fixed memories; HDD (Hard disk), SSD (Solid State Disk), etc.

Fig.9.42 Fixed secondary memory

Removable Memory: This type of Secondary memory is often referred to as external or portable memory. This type of memory or storage device can be removed even when the system is running. It does not need any special tool to remove them.

They offer a portable form of memory that can be transferred from one system to another. They are mostly used as backup storage devices. Examples of removable memories are optical disk such as Compact disk, DVDs & Blu-ray disk, etc, that require optical drives to read them & pen drives such as USB drives that can be connected to any USB port.

Fig.9.43 Removable Secondary memory

The secondary memory can be classified into the following types of memories are:

Magnetic Memory: Magnetic memory or magnetic storage media is a form of non-volatile secondary memory that uses any type of magnetized medium to store data. The two magnetic polarities are used to represent the “1” & “0” of the digital data.

The head moves over a circular rotating magnetic disk & reads the data in the form of magnetic polarities. Since the data is stored in a magnetic medium, the data can easily get damaged near any external magnetic field. It also cannot withstand any shock from falling. It is relatively slower & cheaper than solid-state computer memory.

Fig.9.44 Magnetic memory

This is the most common type of secondary storage memory used in almost every computer & laptop. Examples of magnetic memory are Hard disk, floppy disk & magnetic tapes.

Magnetic memory can be fixed internally such as hard disks & portable or removable such as floppy disk & external hard drives which can be connected through a USB port.

Optical Memory: Optical memory or optical storage as the name suggests is a type of memory that store & read the information using light. The data is stored in a removable disk called optical disk & the device used for storing & reading the data is called optical drive.

The optical drive is equipped with a sharp laser beam that is used for storing & reading the data in the form of marks on an optical disk that is spinning inside it. Back in the days, the optical memories used were read-only memories i.e. we could only access it & cannot modify the content of the disk. Nowadays, we can write them using special drives.

Fig.9.45 Optical memory

They are cheap & portable & lightweight. Examples of optical memories are CD, DVD, Blu-ray disc, etc.

Solid-State Memory SSD: The solid-state memory or solid-state drive (SSD) is an electronic storage device that permanently stores data in an electronic circuit. There is no moving part such as a spinning disk or magnetic disks or read-write heads. So they are resistant to any shock from falling.

The solid-state computer memory uses flash memory which is very fast but expensive than magnetic or optical memory. They are used as a backup storage device or to transfer data from one system to another. They are compact & small in size having very low power consumption.

Fig.9.46 Solid state memory

The SSDs are available as fixed internal storage used in computers & smart phones & as external portable memory such as pen or USB drives & memory cards.

Flash Memory:

- Flash memory is a non-volatile memory used for storage purposes and transferring information between devices.
- It is a developed version of the EEPROM memory, where data can be electronically programmed as well as erased.
- In Flash memory, the data is block-erasable. Whereas in EEPROM, the data is byte-erasable.
- Additionally, in practice, flash memories are rewritten constantly, whereas EEPROM memories are seldom rewritten.
- Like the EEPROM, flash memory implements floating-gate transistors or floating-gate MOSFET (FGMOS) in the internal architecture too.
- The Flash memory is similar to the standard MOSFET, except that the transistor has two gates and not one.
- Flash memory is reasonable to purchase than conventional EEPROM memories and does not require batteries for solid-state storage.

- It has high-speed access time.
- It also has excellent resistance to kinetic shock.
- These memory sticks are incredibly durable, with the ability to withstand intense pressure or extreme temperatures.
- Flash memory is often found in USB drives, iPods, MP3 players, and many more portable electronics.

Types of Flash memories:

- NOR flash memory
- NAND flash memory

The NOR flash connects individual memory cells parallelly, allowing data to be accessed randomly. The NAND flash instead has fewer but lines with the FGMOS stringed together to increase the storage density of the memory cell. Thus, NAND is better suited for random accessing of data in a serial fashion.

Flash memories based on the NOR architecture are good in reading data quickly, but are slower in writing and erasing functions when compared to the NAND flash.

The NOR flash memory programs data byte-by-byte, whereas the NAND memory programs data in pages, where each page might be having a memory of 4KB.

Since NOR deals with a small amount of memory, it consumes more power than the NAND flash memory for writing data. Sometimes devices such as digital cameras incorporate both the NAND and NOR flash memories.

What is the difference between RAM and ROM?

Factors	**RAM**	**ROM**
Nature of working	Volatile memory – needs power supply	Non-volatile memory – doesn't need power
Speed	Fast	Not as fast as RAM
Storage Capacity	High Capacity, ranging from 1 to 256 GB	Low Capacity of about 4-8MB
Space it occupies	RAM data takes a lot of space	RAM takes up less space
Cost	More expensive	Affordable
Data	Can be altered anytime	Cannot be modified/limited

		changes permitted but not done easily
Application	Used to store program codes which CPU needs immediately	Used to store booting instructions or firmware coding

What is the difference between SRAM and DRAM?

Factors	**SRAM**	**DRAM**
Number of transistors	6 Transistors	1 Transistor
Charge Leakage	Not evident	A lot of discharge happens, thus there is a refresh circuitry
Speed	Quite fast	Comparatively slower
Power Consumption	Low	High
Space it occupies	Less space	More space
Cost	Expensive	Cheap
Density	Less dense	Denser
Application	Cache Memory	Main memory

What is the difference between EPROM and EEPROM?

Factors	**EPROM**	**EEPROM**
The medium of erasure of data	UV Light is used to erase data	An electrical signal is used to delete data
Updating is done by	Need to eject the EPROM chip to update the data	No ejection is required for reading or erasing the data
Timeline of technology	This is an older technology.	This is a newer technology.
Design on the case	Has a transparent quartz crystal window on the top	The memory is entirely enclosed in an opaque case.
Preceded by	EPROM is the updated version of the PROM.	EEPROM is the updated version of the EPROM.

Time for erasing	Erasing the contents takes about 15-20 minutes.	Erasing the contents takes about 5ms only.
Programming Technique	A hot electron injection technique is used for reprogramming the EPROM.	The tunnel effect is used for programming the EEPROM.

Convertors (A/D and D/A):

<u>Analogue-to-Digital Converter:</u> A converter that is used to change the analog signal to digital is known as an analog to digital converter or ADC converter. This converter is one kind of integrated circuit or IC that converts the signal directly from continuous form to discrete form. This converter can be expressed in A/D, ADC, A to D. The inverse function of DAC is nothing but ADC. The analog to digital converter symbol is shown below.

The process of converting an analog signal to digital can be done in several ways. There are different types of ADC chips available in the market from different manufacturers like the ADC08xx series. So, a simple ADC can be designed with the help of discrete components.

The main features of ADC are sample rate and bit resolution.

- The sample rate of an ADC is nothing but how fast an ADC can convert the signal from analog to digital.
- Bit resolution is nothing but how much accuracy can an analog to digital converter can convert the signal from analog to digital.

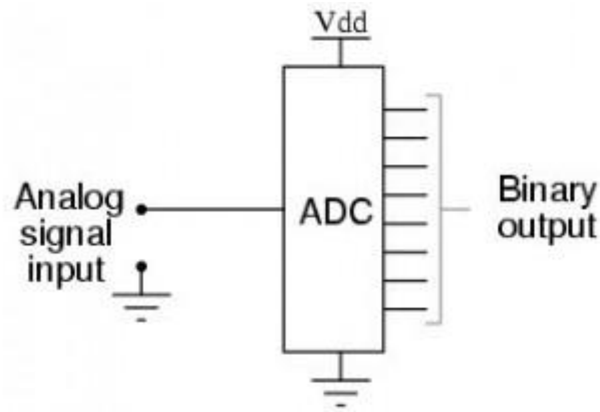


Fig.9.47 Analog to Digital Converter

One of the major benefits of ADC converter is the high data acquisition rate even at multiplexed inputs. With the invention of a wide variety of ADC integrated circuits (IC's), data acquisition from various sensors becomes more accurate and faster. Dynamic characteristics of the high-performance ADCs are improved measurement repeatability, low power consumption, precise throughput, high linearity, excellent Signal-to-Noise Ratio (SNR), and so on.

A variety of applications of the ADCs are measurement and control systems, industrial instrumentation, communication systems, and all other sensory-based systems. Classification of ADCs based on factors like performance, bit rates, power, cost, etc.

ADC Block Diagram: The block diagram of ADC is shown below which includes sample, hold, quantize, and encoder. The process of ADC can be done like the following.

First, the analog signal is applied to the first block namely a sample wherever it can be sampled at an exact sampling frequency. The amplitude value of the sample like an analog value can be maintained as well as held within the second block like Hold. The hold sample can be quantized into discrete value through the third block like quantize. Finally, the last block like encoder changes the discrete amplitude into a binary number.

In ADC, the conversion of the signal from analog to digital can be explained through the above block diagram.

Sample: In the sample block, the analog signal can be sampled at an exact interval of time. The samples are used in continuous amplitude and hold real value however they are discrete with respect to time. While converting the signal, the sampling frequency plays an essential role. So it can be maintained at a precise rate. Based on the system requirement, the sampling rate can be fixed.

Hold: In ADC, HOLD is the second block and it doesn't have any function because it simply holds the sample amplitude till the next sample is taken. So the value of hold doesn't change until the next sample.

Quantize: In ADC, this is the third block which is mainly used for quantization. The main function of this is to convert the amplitude from continuous (analog) into discrete. The value of continuous amplitude within hold block moves throughout quantize block to turn into discrete in amplitude. Now, the signal will be in digital form because it includes discrete amplitude as well as time.

Encoder: The final block in ADC is an encoder that converts the signal from digital form to binary. We know that a digital device works by using binary signals. So it is required to change the signal from digital to binary with the help of an encoder. So this is the entire method to change an analog signal to digital using an ADC. The time taken for the entire conversion can be done within a microsecond.

Analog to Digital Conversion Process: There are many methods to convert analog signals to digital signals. These converters find more applications as an intermediate device to convert the signals from analog to digital form, display output on LCD through a microcontroller. The objective of an A/D converter is to determine the output signal word corresponding to an analog signal. Now we are going to see an ADC of 0804. It is an 8-bit converter with a 5V power supply. It can take only one analog signal as input.

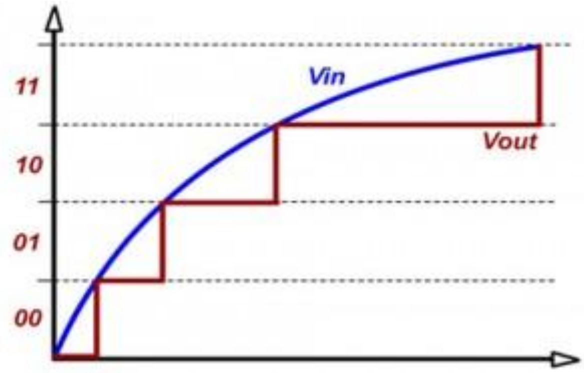


Fig.9.48 Analog to Digital Converter for Signal

The digital output varies from 0-255. ADC needs a clock to operate. The time taken to convert the analog to digital value depends on the clock source. An external clock can be given to CLK IN pin no.4. A suitable RC circuit is connected between the clock IN and clock R pins to use the internal clock. Pin2 is the input pin – High to low pulse brings the data from the internal register to the output pins after conversion. Pin3 is a Write – Low to high pulse is given to the external clock. Pin11 to 18 are data pins from MSB to LSB.

Analog to Digital Converter samples the analog signal on each falling or rising edge of the sample clock. In each cycle, the ADC gets the analog signal, measures it, and converts it into a digital value. The ADC converts the output data into a series of digital values by approximates the signal with fixed precision.

In ADCs, two factors determine the accuracy of the digital value that captures the original analog signal. These are quantization level or bit rate and sampling rate. The below figure depicts how analog to digital conversion takes place. Bit rate decides the resolution of digitized output and you can observe in the below figure where 3-bit ADC is used for converting the analog signal.

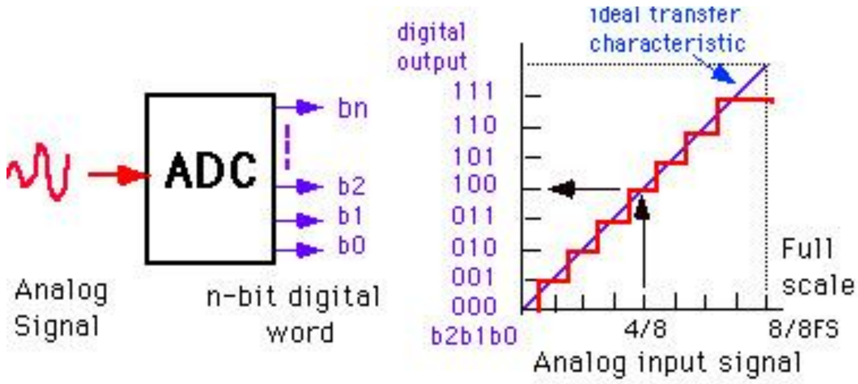


Fig.9.49 Analog to Digital Conversion Process

Assume that one-volt signal has to be converted from digital by using 3-bit ADC as shown below. Therefore, a total of 2^3=8 divisions are available for producing 1V output. This results 1/8=0.125V is called as minimum change or quantization level represented for each division as 000 for 0V, 001 for 0.125, and likewise upto 111 for 1V. If we increase the bit rates like 6, 8, 12, 14, 16, etc. we will get a better precision of the signal. Thus, bit rate or quantization gives the smallest output change in the analog signal value that results from a change in the digital representation.

There is an absolute chance of misrepresenting the input signal on the output side if it is sampled at a different frequency than the desired one. Therefore, another important consideration of the ADC is the sampling rate. The Nyquist theorem states that the acquired signal reconstruction introduces distortion unless it is sampled at (minimum) twice the rate of the largest frequency content of the signal as you can observe in the diagram. But this rate is 5-10 times the maximum frequency of the signal in practice.

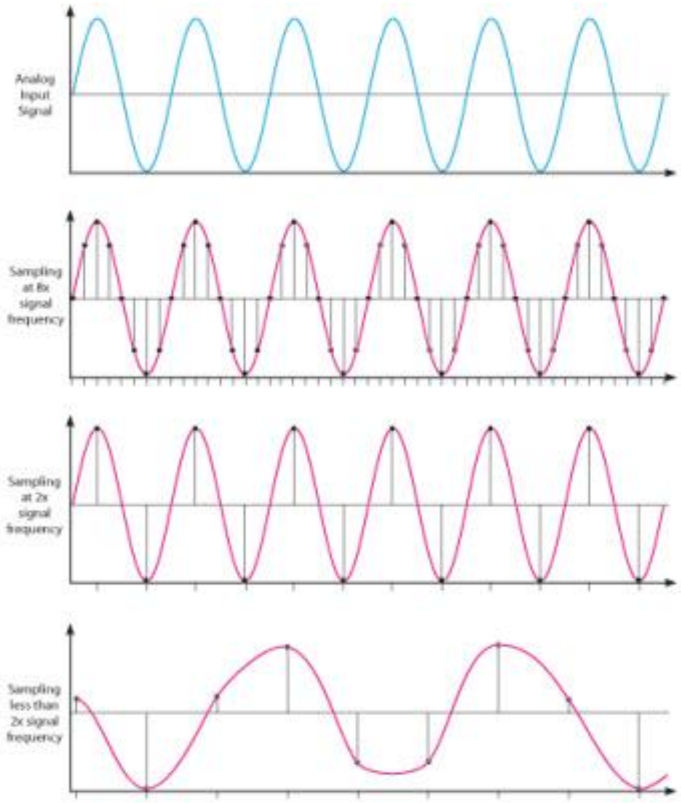


Fig.9.50 Sampling Rate of Analog to Digital Converter

Factors: The ADC performance can be evaluated through its performance based on different factors. From that, the following two main factors are explained below.

SNR (Signal-to-Noise Ratio): The SNR reflects the average number of bits without noise in any particular sample.

Bandwidth: The bandwidth of an ADC can be determined by estimating the sampling rate. The analog source can be sampled per second to produce discrete values.

Types of Analog to Digital Converters:

ADC is available in different types and some of the types of analog to digital converters include:

- Dual Slope A/D Converter
- Flash A/D Converter
- Successive Approximation A/D Converter
- Semi-flash ADC
- Sigma-Delta ADC
- Pipelined ADC

Dual Slope A/D Converter: In this type of ADC converter, comparison voltage is generated by using an integrator circuit which is formed by a resistor, capacitor, and operational amplifier combination. By the set value of Vref, this integrator generates a sawtooth waveform on its output from zero to the value Vref. When the integrator waveform is started correspondingly counter starts counting from 0 to 2^n-1 where n is the number of bits of ADC.

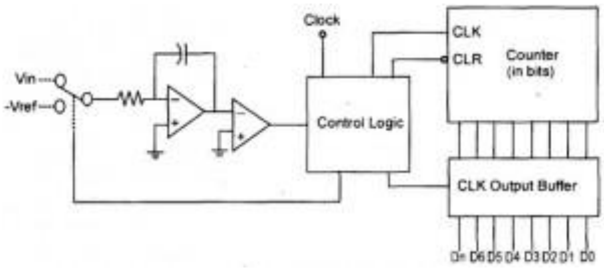


Fig.9.51 Dual Slope Analog to Digital Converter

When the input voltage Vin equal to the voltage of the waveform, then the control circuit captures the counter value which is the digital value of the corresponding analog input value. This Dual slope ADC is a relatively medium cost and slow speed device.

Flash A/D Converter: This ADC converter IC is also called parallel ADC, which is the most widely used efficient ADC in terms of its speed. This flash analog to digital converter circuit consists of a series of comparators where each one compares the input signal with a unique reference voltage. At each comparator, the output will be a high state when the analog input voltage exceeds the reference voltage. This output is further given to the priority encoder for generating binary code based on higher-order input activity by ignoring other active inputs. This flash type is a high-cost and high-speed device.

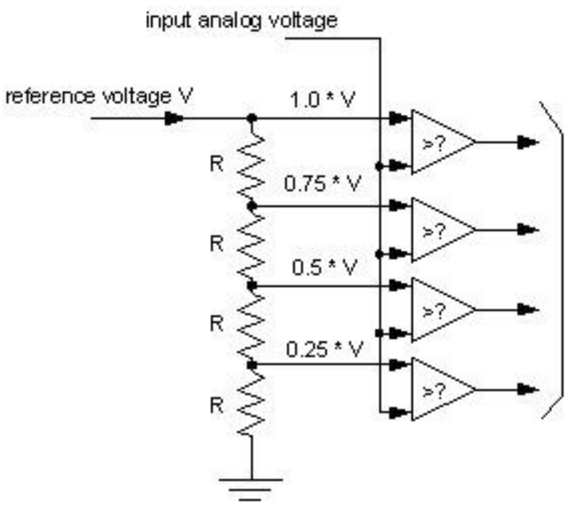


Fig.9.52 Flash A/D Converter

Successive Approximation A/D Converter: The SAR ADC a most modern ADC IC and much faster than dual slope and flash ADCs since it uses a digital logic that converges the analog input voltage to the closest value. This circuit consists of a comparator, output latches, successive approximation register (SAR), and D/A converter.

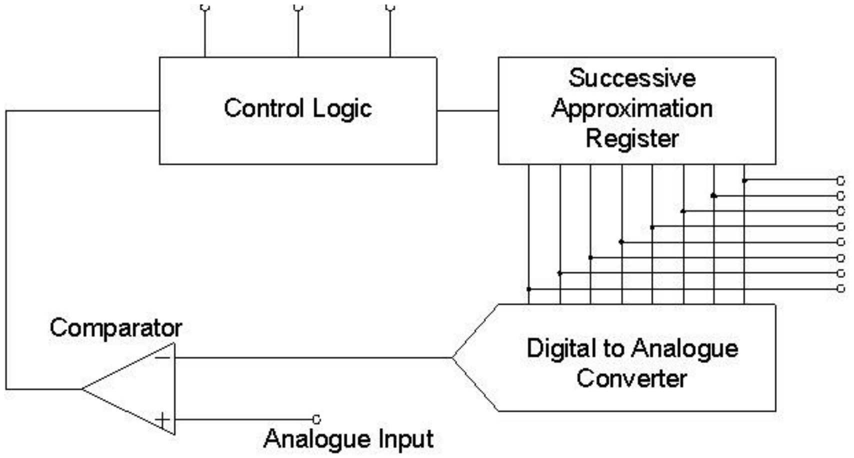


Fig.9.53 Successive Approximation A/D Converter

At the start, SAR is reset and as the LOW to HIGH transition is introduced, the MSB of the SAR is set. Then this output is given to the D/A converter that produces an analog equivalent of the MSB, further it is compared with the analog input Vin. If comparator output is LOW, then MSB will be cleared by the SAR, otherwise, the MSB will be set to the next position. This process continues till all the bits are tried and after Q0, the SAR makes the parallel output lines to contain valid data.

Semi-flash ADC: These types of analog to digital converts mainly works approximately their limitation size through two separate flash converters, where each converter resolution is half of the bits for the semi-flush device. The capacity of a single flash converter is, it handles the MSBs (most significant bits) whereas the other handles the LSB (least significant bits).

Sigma-Delta ADC: Sigma Delta ADC (ΣΔ) is fairly a recent design. These are extremely slow as compared to other kinds of designs however they offer the maximum resolution for all kinds of ADC. Thus, they are extremely compatible with high-fidelity based audio applications, however, they are normally not utilizable wherever high BW (bandwidth) is required.

Pipelined ADC: Pipelined ADCs are also known as sub ranging quantizers which are related in concept to successive approximations, even though more sophisticated. While successive approximations grow through every step by going to the next MSB, this ADC uses the following process.

- It is used for a coarse conversion. After that, it evaluates that change toward the input signal.
- This converter acts as a better conversion by allowing for a temporary conversion with a range of bits.
- Usually, pipelined designs offer a center ground among SARs as well as flash analog to digital converters by balancing its size, speed & high resolution.

Digital to Analog Converter: Digital to Analog Converter (DAC) basically converts digital code that represents digital value to analog current or voltage.

In any communication system, the analog signal which is in the form of physical variables is converted to digital value by ADC. This digital signal is converted back to Analog signal for further processing i.e. Analog signal is necessary to drive Motors, temperature controller etc.

Fig. shows a block diagram of DAC circuit which shows n-bit digital inputs converted to Analog Signal. The digital inputs are d_0, d_1, d_{n-2}, d_{n-1} and V_a is the output Analog Voltage.

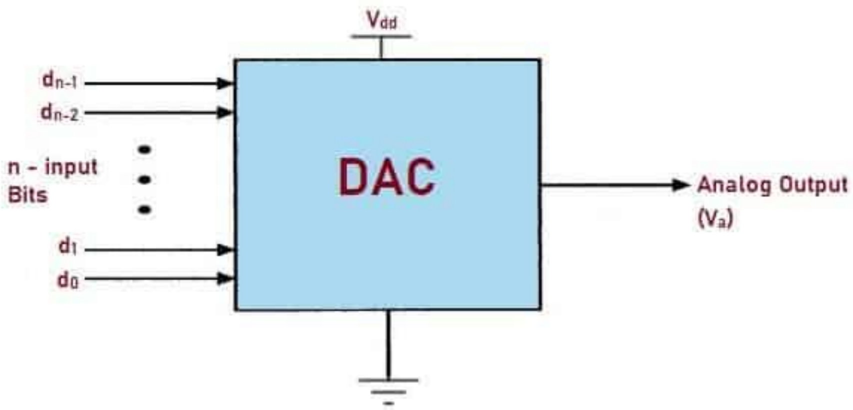


Fig.9.54 Block Diagram of Digital to Analog Converter (DAC)

Types of Digital to Analog Converter (DAC):

There are basically four types of Digital to Analog Converter circuits namely:

- Binary Weighted Resistor D/A Converter Circuit
- Binary ladder or R–2R ladder D/A Converter Circuit
- Segmented DAC
- Delta-Sigma DAC

Binary Weighted Resistor D/A Converter Circuit: In this type of Converter, every digital input bit that needs to be converted, requires one resistor or current source. These resistors are connected across the inputs and the summing point. The output is generated through this Summing Amplifier Circuit. Fig. below shows a typical Binary Weighted Resistor Converter Circuit which consists of an Op-Amp, four resistors which are connected at the input terminal of Op-amp along with the Feedback Resistor.

These Resistors at the input terminal are called as Variable Resistors. Here, A, B, C, D are the digital inputs where 'D' is at MSB and 'A' is at LSB. and V is the output Analog Voltage. The output of the Summing Amplifier circuit is given by the equation:

$$V_o = -R.\left(\frac{D}{R} + \frac{C}{2R} + \frac{B}{4R} + \frac{A}{8R}\right)$$

The output voltage is obtained by substituting different values of inputs in the equation as shown in the table below:

Digital inputs				Analog output Voltage (V_a)
D	C	B	A	
0	0	0	0	**0V**
0	0	0	1	**–1V**
0	0	1	0	**–2V**
0	0	1	1	**–3V**
0	1	0	0	**–4V**
0	1	0	1	**–5V**
0	1	1	0	**–6V**
0	1	1	1	**–7V**
1	0	0	0	**–8V**
1	0	0	1	**–9V**
1	0	1	0	**–10V**
1	0	1	1	**–11V**
1	1	0	0	**–12V**
1	1	0	1	**–13V**
1	1	1	0	**–14V**
1	1	1	1	**–15V**

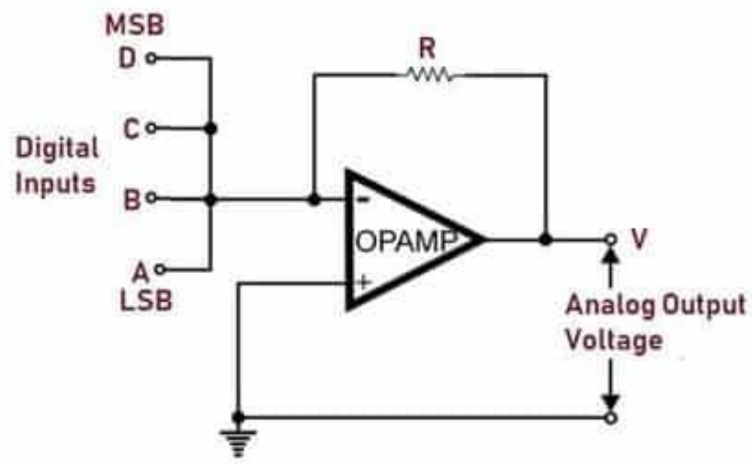


Fig.9.55 Binary Weighted Resistor Circuit

Binary Ladder or R–2R Ladder D/A Converter Circuit: This type of Converter has only two values of Resistors, R and 2R. The conversion speed reduces in this type of DAC due to parasitic capacitance. It is the simplest type of DAC where the switch between ground and inverting input of the Op-amp is controlled by the input bit.

Fig. shows Binary Ladder or R–2R Ladder DAC. To understand its working, let us consider only the Resistors in the network, omitting Op-amp and assuming the input to be DCBA = 1000. Now the output for the reduced network is given by-

$$\text{output } = \left(\frac{2R}{2R+2R}\right)\times(+V) = \frac{V}{2}$$

Similarly, with the change in the input bits, the voltage will be V/4, V/2, V/16…… etc. Now, adding the Op-amp to the circuit, V_4 becomes the input to the non-inverting Amplifier whose Gain is defined by the equation:

$$A_V = -R/2R = -1/2$$

Hence the output voltage of R–2R Ladder DAC is given by the equation:

$$V_O = -R\times\left(\frac{D}{2R}+\frac{C}{4R}+\frac{B}{8R}+\frac{A}{16R}\right)$$

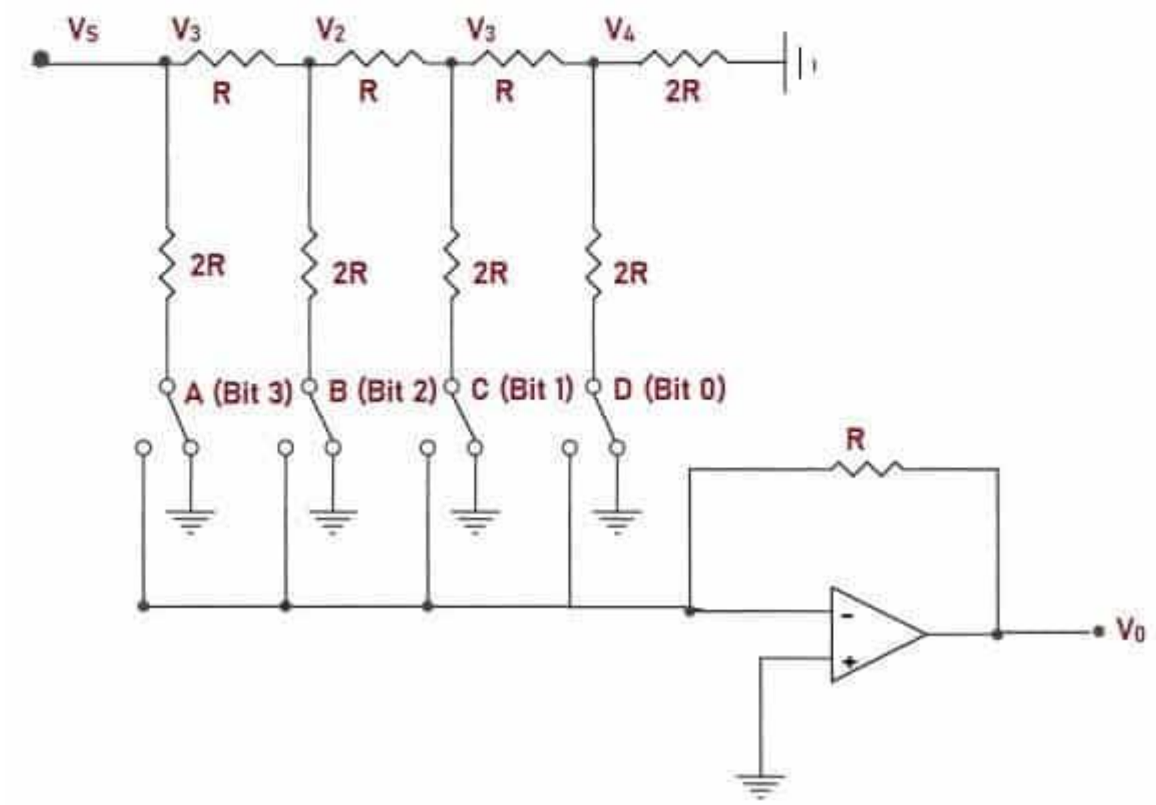


Fig.9.56 Binary Ladder or R–2R Ladder Converter Circuit

Segmented DAC: Segmented DAC's are designed according to the specifications based on performance. In such cases no architecture is ideal and hence two or more DAC's are combined and designed. Binary Weighted and Thermometer-coded DAC are combined. The input binary code is separated into 2 segments.

Thermometer-coding is used for MSB and Binary-weighted structure is applied for LSB. This is done to shrink the size of the chip. The size of the encoder grows exponentially with the increase in input bits, and more switches and interconnection is required.

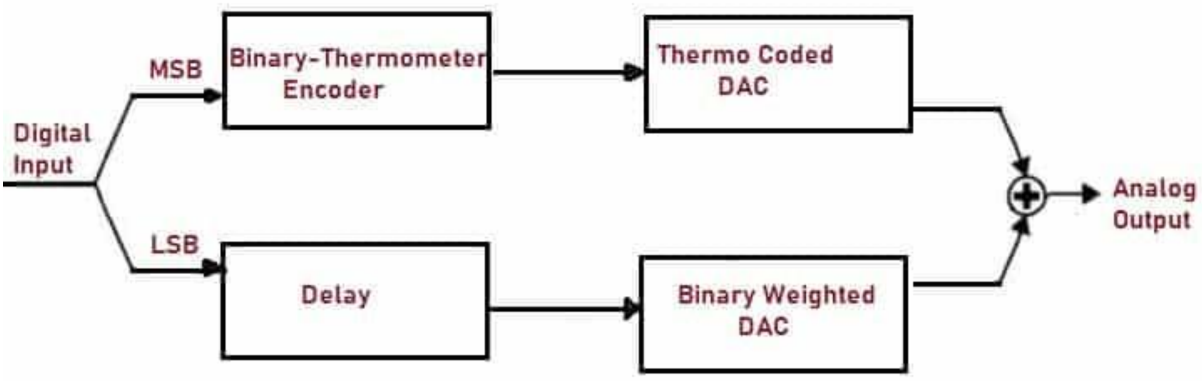


Fig.9.57 Segmented DAC

Delta-Sigma DAC: It is the fastest and highest precision DAC. Fig. shows Delta-Sigma DAC which consists of different blocks namely:

- Interpolation Filter
- Delta-Sigma Modulator
- 1-Bit DAC
- Analog Output Filter

Interpolation Filter: It increases the sample rate and reduces the time and thus, sampling frequency is increased by four times. The data from the Filter is the input to the Modulator block.

Delta-Sigma Modulator: Delta-Sigma Modulator acts as a High-Pass Filter to the quantization noise and Low-Pass Filter to the signal. It converts data into high speed bit-stream.

1-Bit DAC: Digital samples are converted back into Analog form for further amplification. Each bit of the sample is converted serial.

Analog Output Filter: The output from DAC is filtered out to produce Analog signal.

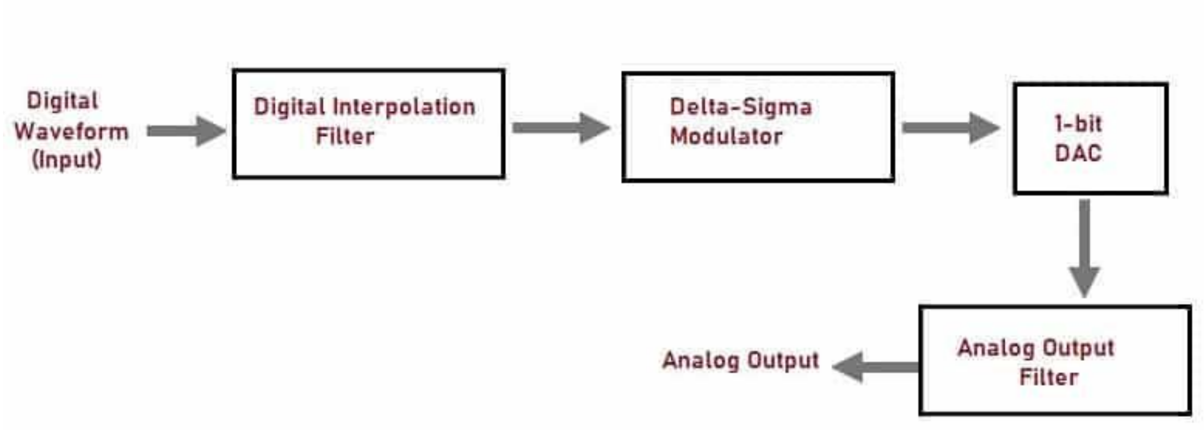


Fig.9.58 Delta-Sigma Digital to Analog Converter

How does Digital to Analog Converter (DAC) Work: To understand the working principle of Digital to Analog Converter, let us imagine a User on a computer recording an audio through a microphone. Here the speech or the audio signal is a physical variable and this signal has to be converted in to digital format. This is done by ADC (Analog to Digital Converter). This bit stream is processed by Digital Signal Processing techniques.

If the recorded audio has to be played through speakers, then the Digital signal has to be converted back to Analog signal with the help of Digital to Analog Converter.

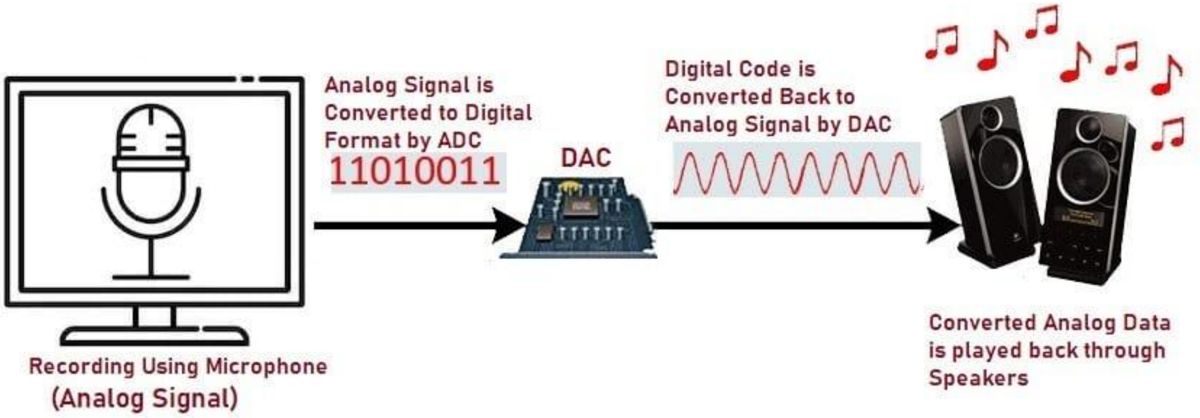


Fig.9.59 Working Principle of DAC

Applications of Digital to Analog Converter (DAC): The applications of Digital to Analog Converter include:

- DAC's are used in Digital Signal Processing.
- They are also used in digital power supplies for Micro-controller.
- DAC's are used in digital potentiometers.
- They are used in all digital data acquisition systems.

Advantages of Digital to Analog Converter (DAC): The advantages of Digital to Analog Converter are:

- Weighted Resistor Circuit is the fastest converting circuit compared to other methods.
- High resolution and accuracy can be achieved.
- DAC's are simple to implement.

Disadvantages of Digital to Analog Converter (DAC): The disadvantages of DAC are:

- Voltage levels must be exactly the same for all inputs in Weighted Resistors DAC. E.g. 4-bit Converter requires 4 resistors.
- Binary weighted Resistor circuit that require Op-Amps are expensive.
- Power dissipation of Binary weighted Resistors Circuit is very high.
- In R-2R Ladder converters, Delay is caused as the circuit needs switching based on the inputs.
- Gain Error, Offset Error, Non-linearity is commonly caused by Resistors used in the circuit.

SAQ.2

a) What do you mean by function of Counter and its types?

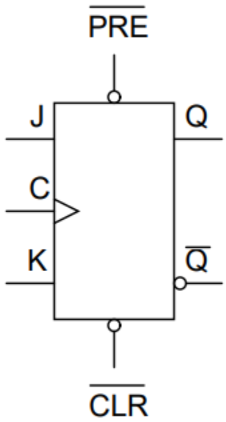
b) Discuss and explain the function of Memory and write its types.

c) What do you mean by Analog to Digital convertor and write its types?

d) Discuss and explain the Digital to Analog convertor and write its advantages.

Examples:

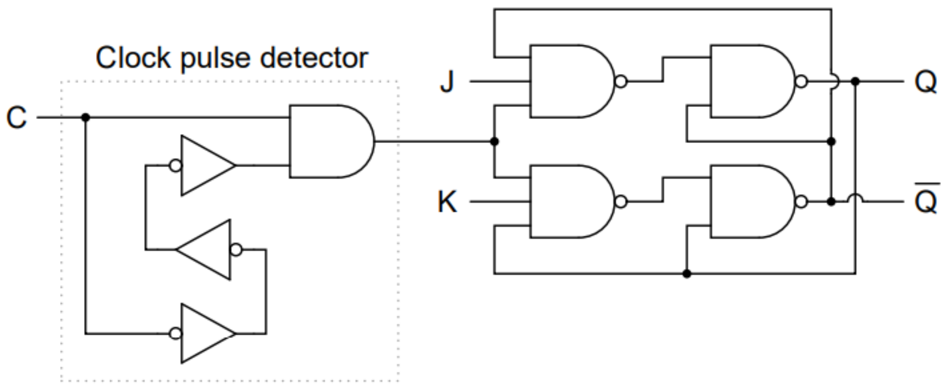
Q.1 Flip-flops often come equipped with asynchronous input lines as well as synchronous input lines. This J-K flip-flop, for example, has both “preset” and “clear” asynchronous inputs:



Describe the functions of these inputs. Why would we ever want to use them in a circuit? Explain what the “synchronous” inputs are, and why they are designated by that term.

Solution: “Asynchronous” inputs force the outputs to either the “set” or “reset” state independent of the clock. ”Synchronous” inputs have control over the flip-flop’s outputs only when the clock pulse allows.

Q.2 An extremely popular variation on the theme of an S-R flip-flop is the so-called J-K flip-flop circuit shown here:



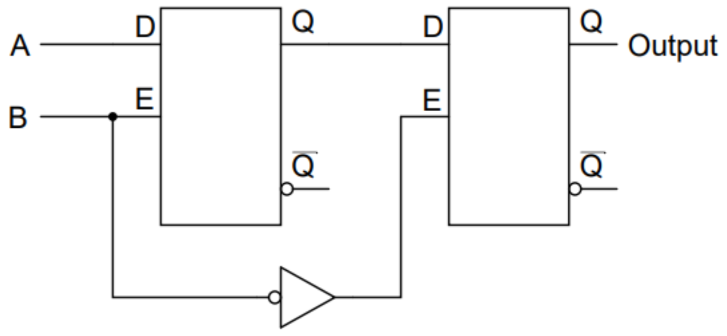
Note that an S-R flip-flop becomes a J-K flip-flop by adding another layer of feedback from the outputs back to the enabling NAND gates (which are now three-input, instead of two-input). What does this added feedback accomplish? Express your answer in the form of a truth table.

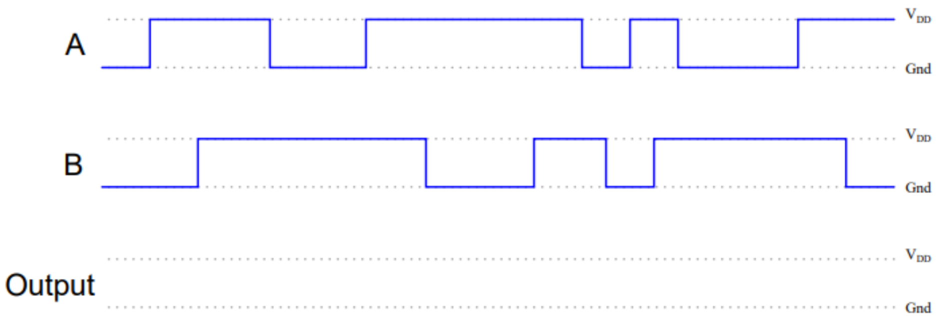
Solution:

C	J	K	Q	$\overline{Q}$
↑	0	0	*Latch*	
↑	0	1	0	1
↑	1	0	1	0
↑	1	1	*Toggle*	

I have found that J-K flip-flop circuits are best analyzed by setting up input conditions (1's and 0's) on a schematic diagram, and then following all the gate output changes at the next clock pulse transition. A technique that really works well in the classroom for doing this is to project a schematic diagram on a clean whiteboard using an overhead projector or computer projector, then writing the 1 and 0 states with pen on the board. This allows you to quickly erase the 1's and 0's after each analysis without having to re-draw the schematic diagram. As always, I recommend you have students actually do the writing, with you taking the role of a coach, helping them rather than simply doing the thinking for them.

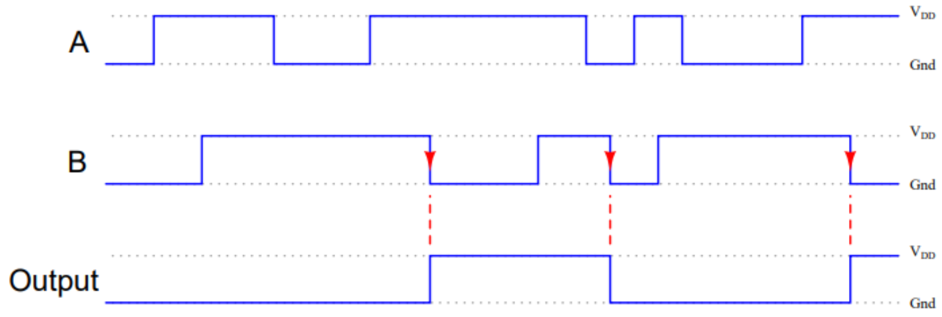
Q.3 Determine the final output states over time for the following circuit, built from D-type gated latches:





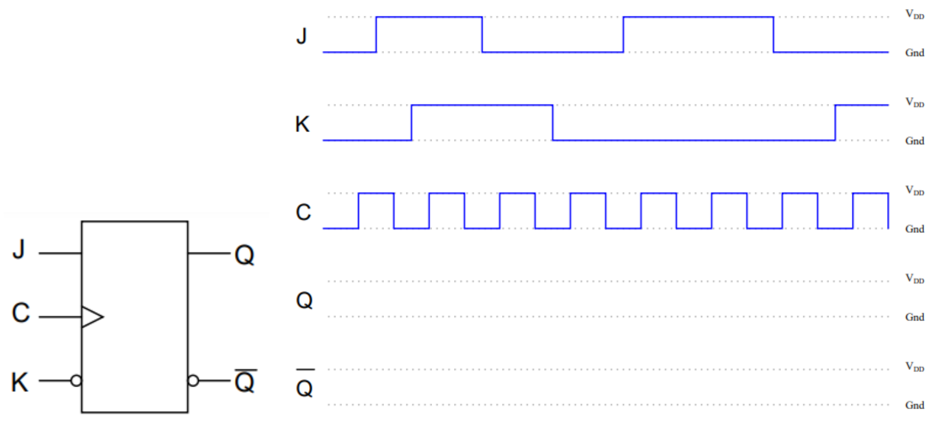
At what specific times in the pulse diagram does the final output assume the input's state? How does this behaviour differ from the normal response of a D-type latch?

Solution:

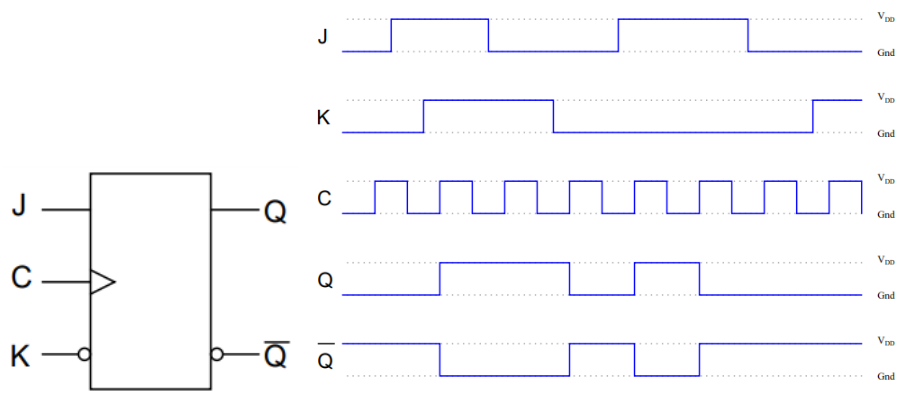


The final output assumes the same logic state as the input only when the enable input signal (B) transitions from "high" to "low".

Q.4 Determine the output states for this J-K flip-flop, given the pulse inputs shown:



Solution:



Q.5 Let us see the working of 3-bit SISO shift register by sending the binary information "011" from LSB to MSB serially at the input.

Solution: Assume, initial status of the D flip-flops from leftmost to rightmost is Q2Q1Q0=000. We can understand the working of 3-bit SISO shift register from the following table.

No of positive edge of Clock	**Serial Input**	**Q_2**	**Q_1**	**Q_0**
0	-	0	0	0
1	1LSB	1	0	0
2	1	1	1	0
3	0MSB	0	1	1LSB
4	-	-	0	1
5	-	-	-	0MSB

The initial status of the D flip-flops in the absence of clock signal is Q2Q1Q0=000. Here, the serial output is coming from Q0. So, the LSB 11 is received at 3rd positive edge of clock and the MSB 00 is received at 5th positive edge of clock.

Therefore, the 3-bit SISO shift register requires five clock pulses in order to produce the valid output. Similarly, the N-bit SISO shift register requires 2N-1 clock pulses in order to shift 'N' bit information.

Q.6 Let us see the working of 3-bit SIPO shift register by sending the binary information "011" from LSB to MSB serially at the input.

Solution: Assume, initial status of the D flip-flops from leftmost to rightmost is Q2Q1Q0=000. Here, Q2& Q0 are MSB & LSB respectively. We can understand the working of 3-bit SIPO shift register from the following table.

No of positive edge of Clock	**Serial Input**	**Q_2**MSB	**Q_1**	**Q_0**LSB
0	-	0	0	0
1	1LSB	1	0	0
2	1	1	1	0
3	0MSB	0	1	1

The initial status of the D flip-flops in the absence of clock signal is Q2Q1Q0=000. The binary information **"011"** is obtained in parallel at the outputs of D flip-flops for third positive edge of clock.

So, the 3-bit SIPO shift register requires three clock pulses in order to produce the valid output. Similarly, the N-bit SIPO shift register requires **N** clock pulses in order to shift 'N' bit information.

Q.7 Let us see the working of 3-bit PISO shift register by applying the binary information "011" in parallel through preset inputs.

Solution: Since the preset inputs are applied before positive edge of Clock, the initial status of the D flip-flops from leftmost to rightmost will be Q2Q1Q0=011. We can understand the working of 3-bit PISO shift register from the following table.

No of positive edge of Clock	**Q_2**	**Q_1**	**Q_0**
0	0	1	1LSB
1	-	0	1
2	-	-	0LSB

Here, the serial output is coming from Q0. So, the LSB 11 is received before applying positive edge of clock and the MSB 00 is received at 2^{nd} positive edge of clock.

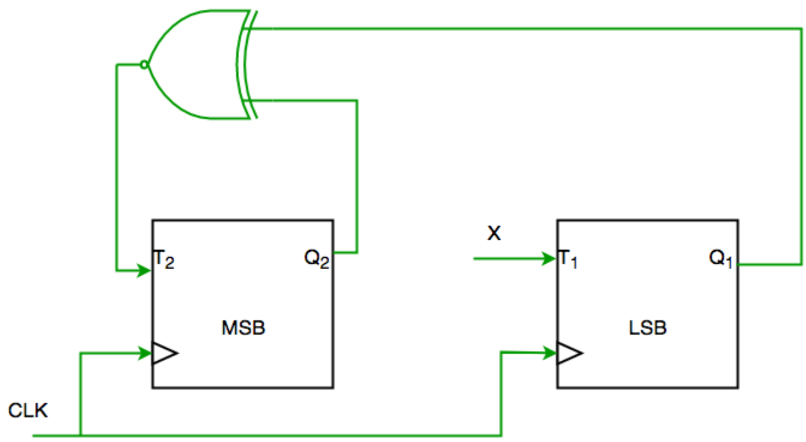
Therefore, the 3-bit PISO shift register requires two clock pulses in order to produce the valid output. Similarly, the N-bit PISO shift register requires N-1 clock pulses in order to shift 'N' bit information.

Q.8 Let us see the working of 3-bit PIPO shift register by applying the binary information "011" in parallel through preset inputs.

Solution: Since the preset inputs are applied before positive edge of Clock, the initial status of the D flip-flops from leftmost to rightmost will be Q2Q1Q0=011. So, the binary information "011" is obtained in parallel at the outputs of D flip-flops before applying positive edge of clock.

Therefore, the 3-bit PIPO shift register requires zero clock pulses in order to produce the valid output. Similarly, the N-bit PIPO shift register doesn't require any clock pulse in order to shift 'N' bit information.

Q.9 Consider the partial implementation of a 2-bit counter using T flip-flops following the sequence 0-2-3-1-0, as shown below



To complete the circuit and what is the input X?

Solution:

From circuit we see

T1=XQ1'+X'Q1—-(1)

AND

T2=(Q2 ? Q1)'—-(2)

And Desired Output IS 00->10->11->01->00

So, X should be **Q1Q2'+Q1'Q2** satisfying 1 and 2.

9.9 Summary:

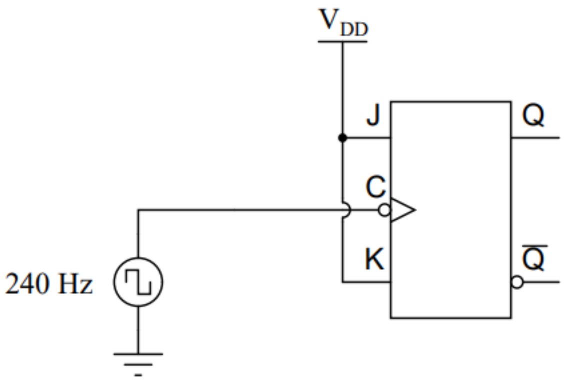
1. The main difference between them is that the Combinational circuit generates output depending only on the present given input, and sequential circuits use present input and present state (previous output) to generate the present output.
2. The R-S flip-flop is used to temporarily hold or store information until it is needed. A single R-S flip-flop will store one binary digit, either a 1 or a 0. Storing a four-digit binary number would require four R-S flip-flops.
3. The D flip-flop is a clocked flip-flop with a single digital input 'D'. Each time a D flip-flop is clocked, its output follows the state of 'D'. The D Flip Flop has only two inputs D and CP. The D inputs go precisely to the S input and its complement is used to the R input.
4. J represents SET, and 'K' represents CLEAR. In the JK flip-flop, the 'S' input is known as the 'J' input, and the 'R' input is known as the 'K' input. The output of the JK flip-flop does not modify if both 'J' and 'K' are '0'.

5. The Master-Slave Flip-Flop is basically a combination of two JK flip-flops connected together in a series configuration. Out of these, one acts as the "master" and the other as a "slave".
6. A Register is a collection of flip flops. A flip flop is used to store single bit digital data. For storing a large number of bits, the storage capacity is increased by grouping more than one flip flops. If we want to store an n-bit word, we have to use an n-bit register containing n number of flip flops.
7. In digital logic and computing, a counter is a device which stores (and sometimes displays) the number of times a particular event or process has occurred, often in relationship to a clock. The most common type is a sequential digital logic circuit with an input line called the clock and multiple output lines.
8. Computer memory is of two basic types – Primary memory (RAM and ROM) and Secondary memory (hard drive, CD, etc). Random Access Memory (RAM) is primary-volatile memory and Read-Only Memory (ROM) is primary-non-volatile memory.
9. An analog-to-digital converter (ADC) is used to convert an analog signal such as voltage to a digital form so that it can be read and processed by a microcontroller. Most microcontrollers nowadays have built-in ADC converters.
10. Applications of analog-to-digital converter are: Used in computer to convert the analog signal to digital signal. Used in cell phones. Used in microcontrollers. Used in digital signal processing. Used in digital storage oscilloscopes.
11. D/A converters are often used to convert finite-precision time series data to a continually varying physical signal. An ideal D/A converter take abstract numbers from a sequence of impulses that are then processed by using a form of interpolation to fill in data between impulses.

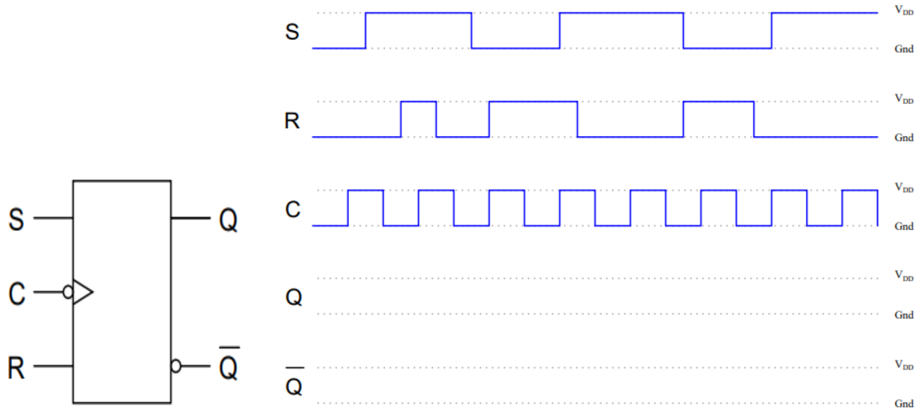
9.10 Terminal Questions:

1) Explain in detail difference between Combinational and Sequential Circuit.
2) What do you mean by Flip-flops? Explain working of the RS flip flop and write in brief its types.
3) Explain the working principle of the D type flip flop and write in brief its types.
4) What do you mean by the working of the JK flip flop and write in brief its types?
5) Explain the working of the master slave flip- flop and write in brief its types.
6) Explain the working of the function of the Register and also write its types.
7) What do you mean by the function of the Counter and explain the brief its types?

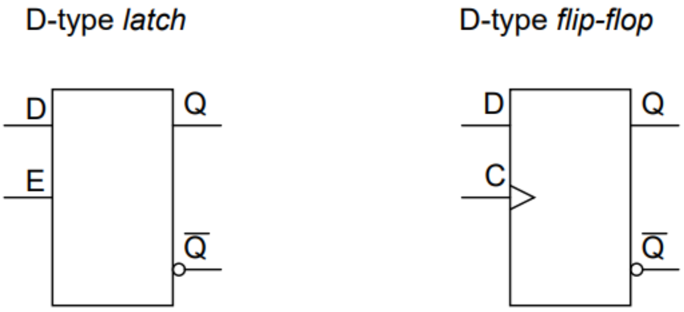
8) Explain the working of the working of the Memory and write the its function and types.
9) Explain the working principle of the Dual Slope A/D Converter.
10) What do you mean by Flash A/D Converter?
11) Explain the working principle of Successive Approximation A/D Converter.
12) Explain the working principle of the Digital to Analog Converter and discuss in brief its types.
13) If the clock frequency driving this flip-flop is 240 Hz, what is the frequency of the flip-flop's output signals?



14) Determine the output states for this S-R flip-flop, given the pulse inputs shown:

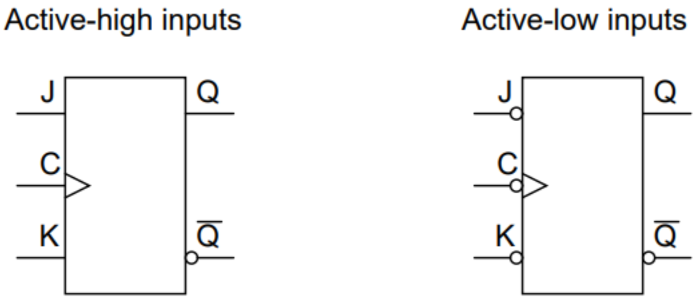


15) Shown here are two digital components: a D-type latch and a D-type flip-flop:



Other than the silly name, what distinguishes a "flip-flop" from a latch? How do the two circuits differ in function?

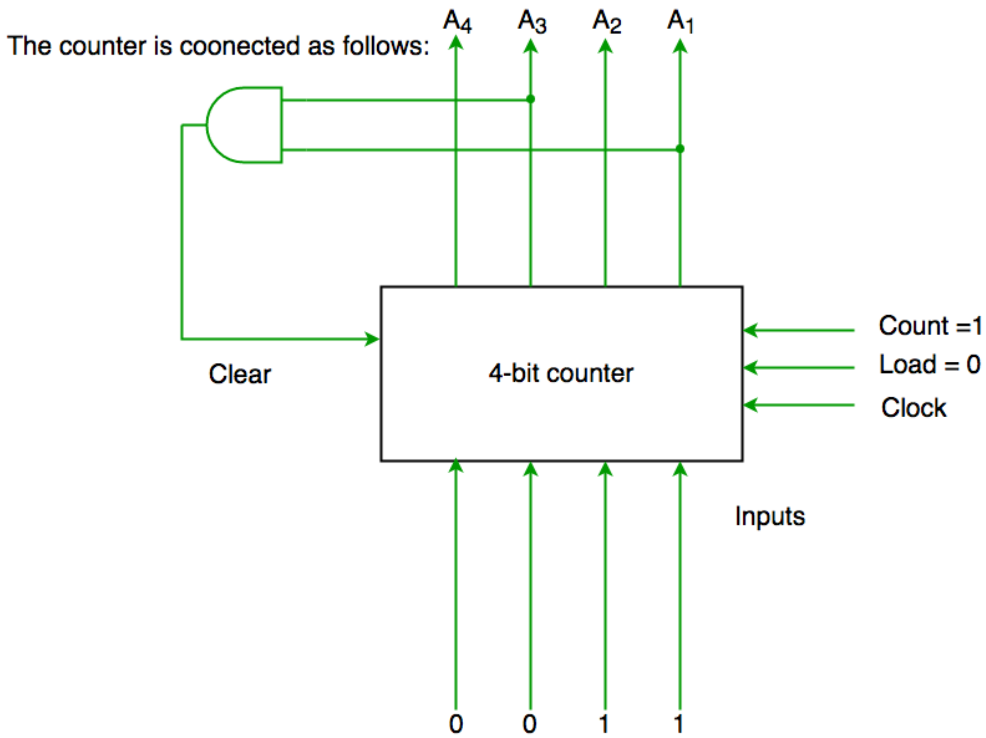
16) Determine what input conditions are necessary to set, reset, and toggle these two J-K flip-flops:



17) Let us see the working of 3-bit PISO shift register by applying the binary information **"111"** in parallel through preset inputs.

18) The control signal functions of a 4-bit binary counter are given below (where X is "don't care"). The counter is connected as follows:

Clear	Clock	Load	Count	Function
1	X	X	X	clear to 0
0	X	0	0	No change
0	↑	1	X	Load Input
0	↑	0	1	Count next



Assume that the counter and gate delays are negligible.

Unit 10 Integrated circuits and devices

Structure:

10.1 Introduction:

Integrated Circuits or an IC is defined as a microchip on which thousands and hundreds of electrical components, such as resistors, capacitors and transistors, are fabricated. An IC functions as an oscillator, amplifier, microprocessor, timer or as computer memory.

The importance of the integrated circuit (IC) is in its ability to allow for the miniaturization of components. Smaller components mean smaller overall size, which allowed for the creation of almost every device we use today from smart TVs to cell phones.

A discrete circuit is constructed of components which are manufactured separately. Later, these components are connected together by using conducted wires on a circuit board or a printed circuit board.

Discrete circuits are individual components such as resistors, capacitors, inductors and transistors mounted independently usually on a PCB (printed circuit board). An integrated circuit has all these components etched directly into layers of silicon.

Integrated circuits can be broadly classified into analog, digital and mixed signal, consisting of analog and digital signaling on the same IC. Digital integrated circuits can contain billions of logic gates, flip-flops, multiplexers, and other circuits in a few square millimeters.

An integrated circuit consists of many extremely thin layers of P- and N-type material arranged in configurations such as transistors, diodes, resistors, and capacitors. A single chip may contain millions of transistors and occupy less than one square inch of area.

he basic structure of a monolithic IC will have 4 layers of different materials. The base layer will be a P-type silicon layer and is named as the substrate layer. This layer will have a typical thickness of 200 micrometer.

The monolithic fabrication process consists of wafer preparation, epitaxial growth, diffused isolation, base and emitter diffusions, pre-ohmic etch, metallization, circuit probing, dicing, mounting and packaging, wire bonding, encapsulation and final testing.

A cathode ray oscilloscope is an instrument based upon the cathode ray tube, that provides a visible image of one or more rapidly varying electrical quantities. A cathode ray oscilloscope may be used to display the variations in voltage signals.

Cathode ray oscilloscope works on the the principle of deflection of the electron beam in the horizontal and vertical directions and creates the trace on the fluorescent screen.

The main applications of CRO are to observe the properties of electrical and electronic signals, oscillation distortion testing, the frequency response of signals, etc. The important uses of CRO are the measurement of voltage, current, frequency, etc.

A multimeter is basically a PMMC meter. To measure dc current the meter acts as an ammeter with a low series resistance. Range changing is accomplished by shunts in such a way that the current passing through the meter does not exceed the maximum rated value.

A multimeter or a multitester, also known as a volt/ohm meter or VOM, is an electronic measuring instrument that combines several measurement functions in one unit. A typical multimeter may include features such as the ability to measure voltage, current and resistance.

Ultrasonic waves are produced when an electrical signal generator sends a burst of electrical energy to a piezoelectric crystal in the transducer causing the crystal to vibrate and convert the electrical pulses into mechanical vibrations (sound waves).

Ultrasonic waves can be detected using Radiometer. In this method ultrasonic beam is made to fall on a thin mica fan suspended by a thin wire carrying a small mirror from one end of a light rod. Due to pressure exerted by ultrasonic waves the fan gets detected along with the mirror.

The velocity of ultrasonic waves in liquid hydrogen depends upon the ortho-para concentration in the liquid. Between 14 and 20°K the velocity in a 99.8% para mixture is about 8 m/sec or about 0.6% smaller than in the 25% mixture. In the frequency range from 1 to 5 MHz no influence of frequency could be detected.

In industry, ultrasound is used in a wide range of processes, such as cleaning, welding of plastics and metals, cutting, forming, testing of materials, separating, mixing, de-gassing, atomising, localising, measuring and many others.

Hydrogen has been the fuel of choice for hypersonic ight because of its rapid burning rate, high speci c energy content, and good heat transfer properties for active cooling and recuperation. A hypersonic aero-ballistic system is dropped from an aircraft, accelerated to hypersonic speed using a rocket and then follows a ballistic, meaning unpowered, trajectory. The system Russian forces used to attack Ukraine, the Kinzhal, is an aero-ballistic missile.

10.2 Objectives:

After studying this unit you should be able to

- Explain and identify Introduction of integrated circuit and its comparison with discrete circuits.
- Study and identify Classification of IC on the basis of construction and operation.
- Explain and identify Monolithic IC (basics structure and fabrication).
- Study and identify Cathode ray oscilloscope (principle, construction, block diagram, working and application).
- Explain and identify Multimeter (principle, types, construction and function).
- Study and identify Ultrasonics (production, detection, velocity measurements and applications), Hypersonics and ultrasonics.

10.3 Introduction of integrated circuit and its comparison with discrete circuits:

Introduction of integrated circuit:

An integrated circuit is a microscopic array of electronic circuits and electronic components (resistors, capacitors, inductor) that are diffused or implanted into the surface of semiconductor material wafer such as silicon. Integrated Circuit invented by Jack Kilby in the 1950s. A chip is commonly termed as Integrated circuits (IC).

The concept of IC was first introduced in the year 1958. Since then this concept has reached great technological heights than any other concepts and has helped in the miniaturization of a lot of components like mobiles, computers, laptops, and many more devices in the digital world.

The digital era started with the invention of vacuum tubes. Vacuum based computers were rare and expensive. This was then replaced by transistors, which were faster in use and smaller in size, cost effective, less power consuming and reliable. Then came the invention of integrated circuits which just revolutionized the use of computers. Due to its small dimension, low cost, and very high reliability even the common man is familiar with its applications like smart phones and laptops.

The IC's also found its way in military applications, state of the art communication systems, and industrial applications due to its high reliability and compact size. Nowadays, an IC that has the size of a fingernail consists of more than a million transistors and other discrete components embedded into it. Thus an integrated circuit can also be called a microchip and is basically a collection of some discrete circuits on a small chip that is made of a semiconductor material like silicon.
The use of discrete circuits was replaced by IC's due to two factors. One is space consumption. A discrete circuitry consists of transistors, resistors, diodes, capacitors, and many other discrete devices. Each of them is soldered on to printed circuit boards (PCB) according to the need of circuitry. In the end PCB will occupy a large space. Another drawback is that the soldered components will show less reliability due to the use of many components. Both these factors urged engineers to invent microcircuits that have more reliability and consume less space.

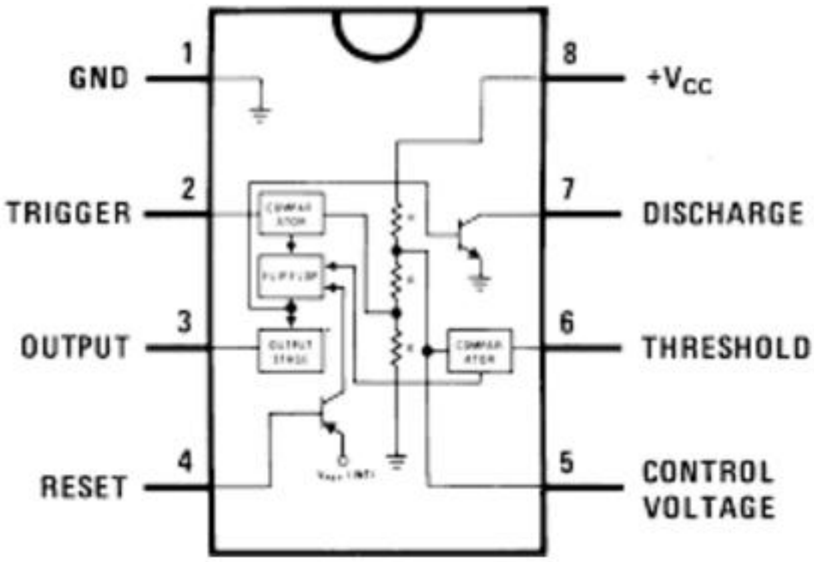


Fig.10.1 Basic structure of an IC

These ICs are packed in a solid outer cover which can be made of an insulating material with high thermal conductivity and with contact terminals (also called pins) of the circuit coming out from the body of the IC.

Integrated Circuit or IC is a very small array of electronic circuits and components like capacitor, resistor, transistor and diode which are diffused into the surface of semiconductor wafer such as silicon. Usually comes in chip package and it represents a circuit system (not single element) and a single IC can be represent as up to thousands of elements brought together into a single chip.

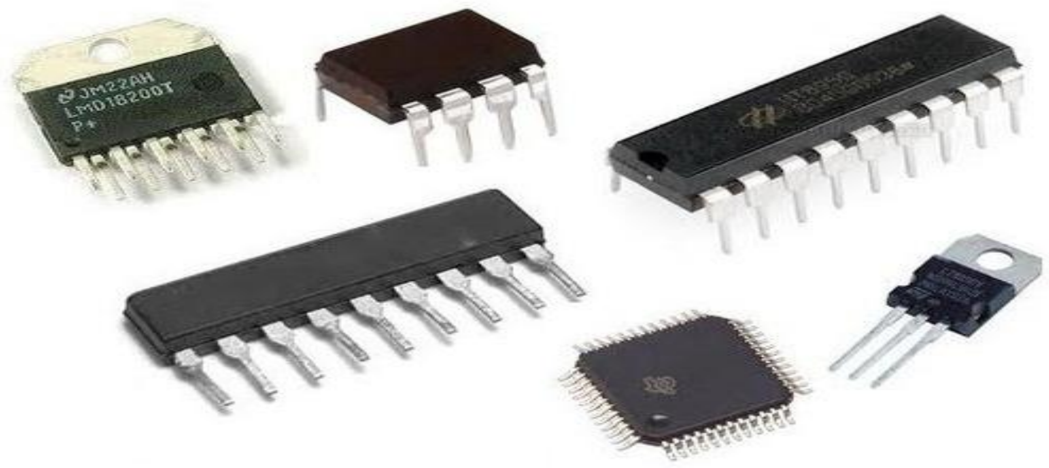


Fig.10.2 Integrated Circuit

IC's are enclosed with a solid outer cover made from insulating material with contact or terminals called as pins. Also, the solid outer cover was designed to have a high thermal conductivity for the devices not to melt during runtime.

Based on pin configuration different types of IC`s packaging are available.

- Dual In-line Package (DIP)
- Plastic Quad Flat Pack (PQFP)
- Flip-Chip Ball Grid Array (FCBGA)

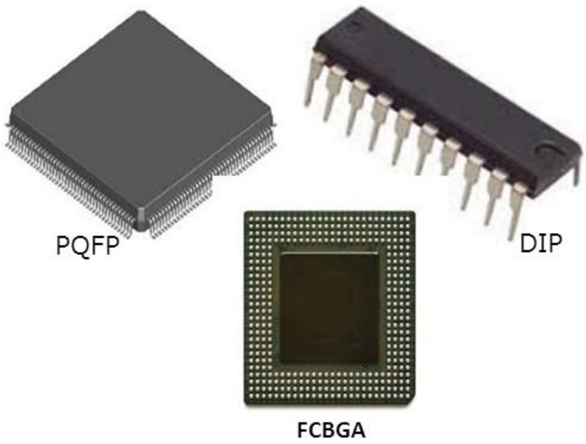


Fig.10.3 Types of ICs Packaging

The transistors are the main components in IC manufacturing. These transistors may be Bipolar Transistors or a field effect transistor depends upon the application of ICs. As the technology is growing day by day, the number of transistors incorporated in an IC is also increasing. Depending upon the number of transistors in an IC or Chip, the ICs are categorized into five types given below.

S.No	**IC category**	**Number of transistors incorporated in a single IC chip**
1	Small Scale Integration (SSI)	Up to 100
2	Medium Scale Integration(MSI)	From 100 to 1000
3	Large Scale Integration(LSI)	From 1000 to 20K
4	Very Large Scale Integration (VLSI)	FROM 20K to 1000000
5	Ultra Large Scale Integration (ULSI)	From 10,00,000 to 1,00,00,000

Advantages of IC:

1. Small in size: IC's are small in size but, practically as equivalent as around 20,000 components that are incorporated in a single chip.

2. Simplifies Design: On the end, complex circuits are fabricated with in a single chip. Therefore, designing of a complex circuit will be simplified into a single chip.

3. High Reliability: Complex circuit are incorporated on a single chip making it with lesser number of connections. Thus, soldering process will be minimize to gain high reliability.

4. Low cost: Production of IC's are done in bulk or mass production. Therefore, the price of IC per unit will be reduce due to high supply. It will be much inexpensive rather than creating equivalent discrete circuit of an IC.

5. Low power consumption: Because IC's are made of semiconductor material wafer, it consumes very little power or less power. As we know that semiconductors draw lesser power than other elements.

6. Easy to replace component: In integrated circuit, a circuit system will be considered a single element. So, replacing the entire circuit is easy rather than finding the specific component and replace it in discrete circuit. Also, there is no need of repairing. Replacement is economical and reliable than repair.

7. Increase operating speed: Integrated circuits can increase the operation speed because parasitic elements such as capacitors are omitted.

8. Small Signal Suitability: IC components are designed with very compact arrangement within the wafer. So, these components are help close to each other and there is no chance of stray electrical pick ups.

Disadvantages of Integrated Circuits:

1. Impossible to modify the parameters: After the IC was fabricated, it is impossible to modify its parameters in which it will operate.

2. Cannot be repair: When the component of an IC is damaged, the whole IC should be replaced and it is impossible the repair the component inside.

3. Cannot be integrated with high capacitance: With the capacitance value higher than 30pF, it is impossible to fabricate in the semiconductor chip. To make it possible, should have connected to discrete component externally.

4. Low power production: It is impossible for a IC to produce a high power output. Usually it can not produce power greater than 10W. So, it is impossible to feed directly the output from IC to actuators like motors. Discrete component is needed to drive the actuators.

5. Exterior connection of inductor and transformer: Discrete components such as inductors and transformers are impossible to fabricate in semiconductor chip. If the circuit needs inductor, connecting it external inductor is the only way.

Discrete Circuit: Refers to the type of circuit construction in which the components are manufactured separately. The component are connected using a conducting wires, breadboard or a printed circuit board (PCB). These components can be resistor, diodes, transistors and inductors.

Fig.10.4 Discrete Circuit

Components of this circuit type are combined to performed a particular output. For example, we can use transistors to create a logic gates. Discrete circuits can be designed to use in low voltage or high voltage, it depends on the acceptability and compatibility of components. Nowadays, most of electronics devices are mixture of discrete component and integrated circuits.

Advantages of Discrete circuit:

1. Handles high power input.

2. Produces high power output. It can produce high power output that can drive the actuators directly.

3. Achieves low temperature coefficient.

4. Can change parameters of the circuit. Changing the parameter of the circuit can be achieved here as same as changing the components and values of components.

Disadvantages of Discrete circuits:

1. Assembling and Wiring: Assembling and wiring of separate components will take time and it usually occupies space that makes the circuitry bulky.

2. Designing the Circuitry: In case of designing, it takes patience and creativity. Usually discrete circuits have many components to combining. So, this will test your patience in designing a circuit in which components and wires will not messed up.

3. Replacement of failed components: It is complicated to replace failed components in this type of circuit construction. Some components are connected to each other and the other components will be affect when removing the other. For the intention of making the circuit as compact as possible, replacing single component is fragile.

4. Less Relibility: Wires and components are usually connected with soldering process. Soldering process by hand affects reliability due to solder joint behavior, component damage and cleanliness. Solder joint behavior can affect because it can add resistances on the circuit specially when you had bad solder joint. Usually, soldering process would reached to 400 degrees celsius and this high temperature can damage the component.

Advantages of an Integrated Circuit over Discrete Circuits:

- An integrated circuit quite small in size practically around 20,000 electronic components can be incorporated in a single square inch of IC chip.
- Many complex circuits are fabricated on a single chip and hence this simplifies the designing of a complex circuit. And also it improves the performance of the system.
- IC’s will give high reliability. A lesser number of connections.
- These are available at low cost due to bulk production.
- IC’s consume very tiny power or less power.
- It can easily replaceable from the other circuit.

A discrete circuit is constructed of components which are manufactured separately. Later, these components are connected together by using conducted wires on a circuit board or a printed circuit board. The transistor is one of the primary components used in discrete circuits, and combinations of these transistors can be used to create logic gates. These logic gates can be used to obtain the desired output from an input. Discrete circuits can be designed to operate at higher voltages.

Disadvantages of Discrete Circuits:

- Assembling and wiring of all individual discrete components take more time and occupies a larger space required.
- Replacement of a failed component is complicated in an existed circuit or system.
- Actually, the elements are connected using soldering process so, that may have caused less reliability.
- To overcome these problems of reliability and space conservation, integrated circuits are developed.

10.4 Classification of IC on the basis of construction and operation:

All the IC's have interconnected discreet devices inside the chip and the corresponding external connecting terminals outside. Each pin may have each function and may vary according to the manufacturer's design. In order to make the circuit fully operative, the pins in the IC must be used for supply voltage, input and output connections, and also some external components according to the needs of the manufacturer.

ICs can be classified on the basis of their chip size as given below:

Small scale integration (SSI)—3 to 30 gates/chip.
Medium scale integration (MSI)—30 to 300 gates/chip.
Large scale integration (LSI)—300 to 3,000 gates/chip.
Very large scale integration (VLSI)—more than 3,000 gates/chip.

IC Types:

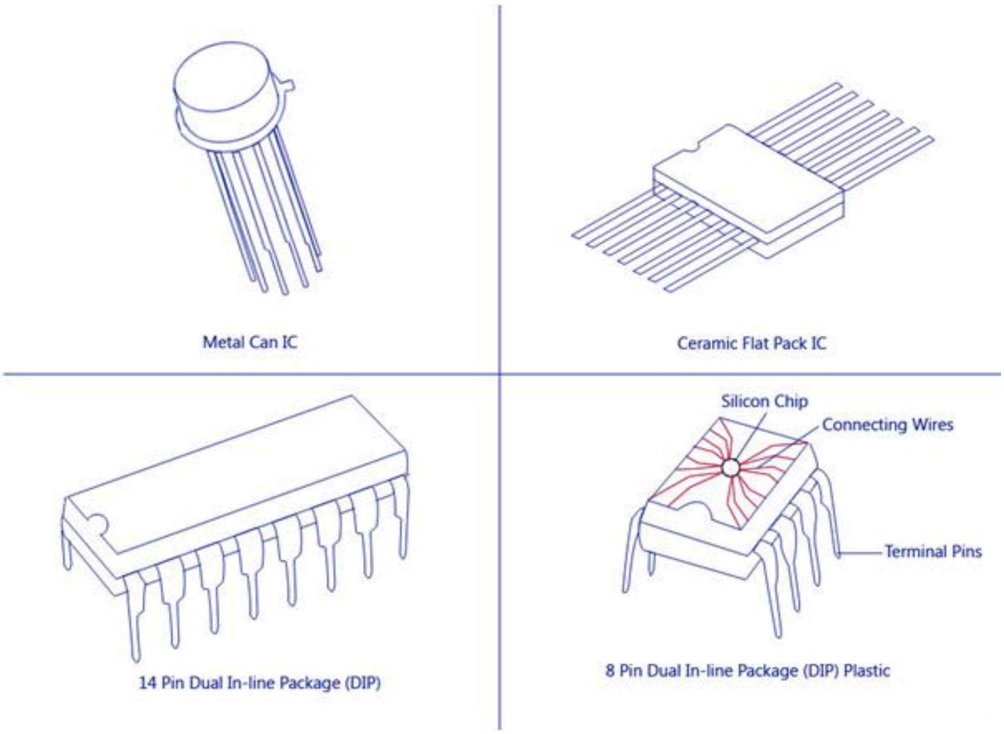


Fig.10.5 Types of ICs

On the basis of applications ICs are of two types namely: Linear Integrated Circuits and Digital Integrated Circuits.

Linear IC's are used in cases when the relationship between the input and output of a circuit is linear. An important application of linear IC is the operational amplifier commonly referred to as op-amp.

When the circuit is either in on-state or off-state and not in between the two, the circuit is called a digital circuit. IC's used in such circuits are called digital IC's. They find wide applications in computers and logic circuits.

Here are some further classification of integrated circuits based on the fabrication techniques used.

1. Monolithic Integrated Circuits: The word 'monolithic' comes from the Greek words 'monos' and 'lithos' which means 'single' and 'stone'. As the name suggests, monolithic IC's refer to a single stone or a single crystal. The single crystal refers to a single chip of silicon as the semiconductor material, on top of which all the active and passive components needed are interconnected. This is the best mode of manufacturing IC' as they can be made identical, and produces high reliability. The cost factor is also low and can be manufactured in bulk in very less time. They have been found applicable for C's used for AM receivers, TV circuits, computer circuits, voltage regulators, amplifiers and so on.

A detailed article explaining the concept and fabrication process of different components and monolithic IC production process is explained here Monolithic Integrated Circuit.

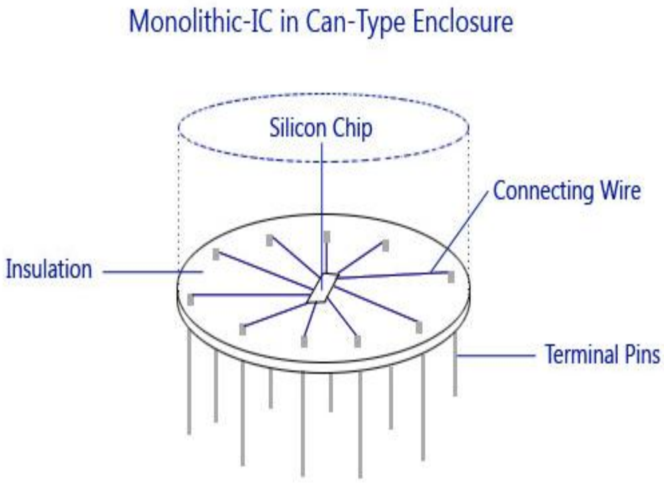


Fig.10.6 Monolithic IC – Can Type

Being as it is, monolithic IC's have some limitations as well.

1. Monolithic IC's have low power rating. They cannot be used for low power applications as they cannot have a power rating of more than 1 watt.

2. The isolation between the components inside the IC is poor.

3. Components like inductor cannot be fabricated to the IC.

4. The passive components that are fabricated inside the IC will be if small value. For higher values they have to be connected externally to the IC pins.

5. It is difficult to make a circuit flexible for any kind of variation; a new set of masks is required.

2. Thin and Thick Film Integrated Circuit: Thick and thin film IC's are comparatively larger than monolithic IC's and smaller than discrete circuits. They find their use in high power applications. Though it is a little large in size, these IC's cannot be integrated with transistors and diodes. Such devices have to be externally connected on to its corresponding pins. Passive components like resisters and capacitors can be integrated.

Both thick and thin film IC’s are explained in detailed below. Though both the IC’s have similar appearance, properties, and general characteristics, the main difference between the two of them is the manner in which the film is deposited on to the IC.

Thin Film Integrated Circuits: This IC is fabricated by depositing films of conducting material on the surface of a glass or ceramic base. The resistors are fabricated by controlling the width and thickness of the films and by using different materials selected for their resistivity. For capacitors, a film of insulating oxide is sandwiched between two conducting films. A spiral form of film is deposited onto the IC to create an inductor.

Mainly two methods are used for producing thin films. One method, called vacuum evaporation is used in which vaporized material is deposited on a substrate contained in a vacuum. The other method is called cathode sputtering in which atoms from a cathode made of the desired film material are deposited on a substrate located between a cathode and an anode.

Thick Film Integrated Circuits: They are also commonly called as printed thin film circuits. The desired circuit pattern is obtained on a ceramic substance by using a manufacturing process called silk-screen printing technique.

The inks used for printing are usually materials that have resistive, conductive, or dielectric properties. They are selected accordingly by the manufacturer. The screens are actually made of fine stainless steel wire mesh. The films are fused to the substrate after printing by placing them in hot high temperature furnaces.

The fabrication techniques used for thin film passive components are adopted for thick films as well. As with thin-film circuits, active components are added as separate devices. A portion of thick-film circuit is given in the figure below.

Thick-Film IC

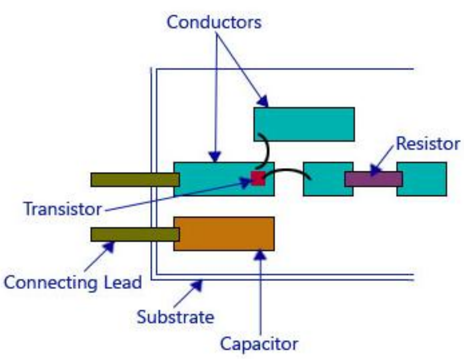


Fig.10.7 Thick Film IC

When compared to monolithic IC's, thick and thin film IC's do have some advantages. They have the advantage of better tolerance, better isolation between components, and greater flexibility in circuit design that further helps in providing high frequency performance. But, these are the only factors that must be considered for the application of such IC's as they are costly in making, and has higher dimensions than monolithic IC's. They also cannot be used to fabricate active components which further increase the size.

3. Hybrid or Multi-chip Integrated Circuits: As the name suggests, the circuit is fabricated by interconnecting a number of individual chips. Hybrids ICs are mostly used for high power audio amplifier applications from 5 Watts to more than 50 Watts. The active components are diffused transistors or diodes. The passive components may be group of diffused resistors or capacitors on a single chip, or they may be thin-film components. Interconnection between the individual chips is made by wiring process or a metallized pattern.

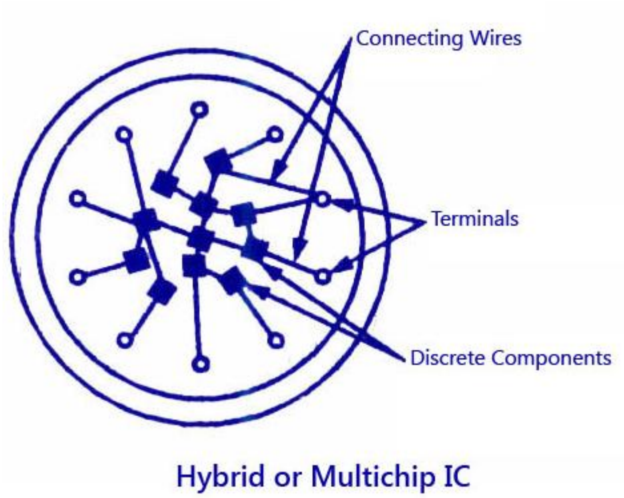


Fig.10.8 Diagram of a hybrid or multi-chip IC

The diagram of a hybrid or multi-chip IC is shown in the figure above. Hybrid IC's are also known to provide a better performance than monolithic IC's. Although the process is too expensive for mass production, multi-chip techniques are quite economical for small quantity production and are more often used as prototypes for monolithic ICs.

Based upon the active devices employed the ICs can be classified as bipolar ICs using bipolar active devices (BJT) and unipolar IC's using unipolar active devices like FET.

10.5 Monolithic IC (basics structure and fabrication):

Monolithic IC's basics structure: We have already discussed the basics of Integrated Circuits in our previous post. The concepts of a basic monolithic IC will be discussed here.

To know the basics a sample circuit must be considered to be converted to its monolithic form. With basic components like resistor, diode, and transistor a basic circuit is first made.

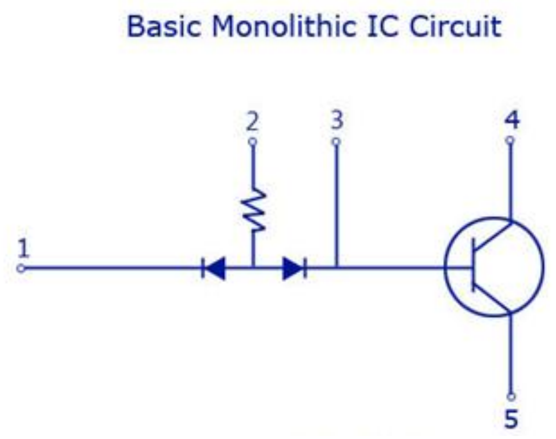


Fig.10.9 Basic Monolithic IC Circuit

With the basic circuit, the different layers for the monolithic IC are then considered. The basic structure of a monolithic IC will have 4 layers of different materials. The base layer will be a P-type silicon layer and is named as the substrate layer. This layer will have a typical thickness of 200 micrometer. Silicon is the preferred semiconductor for the P-type and N-type layer because of its favourable characteristics for the manufacturing of an IC.

The layer above the substrate P-type silicon layer is the N-type layer. All the active and passive components required for the circuit are fabricated onto this layer. This layer has a typical thickness of 25 micrometer. The N-type silicon material is grown as a single crystal extension of the P-layer and the components are required are fabricated using series of P-type and N-type impurity diffusions. The N-type layer becomes the collector for the transistor or an element for a diode or a capacitor.

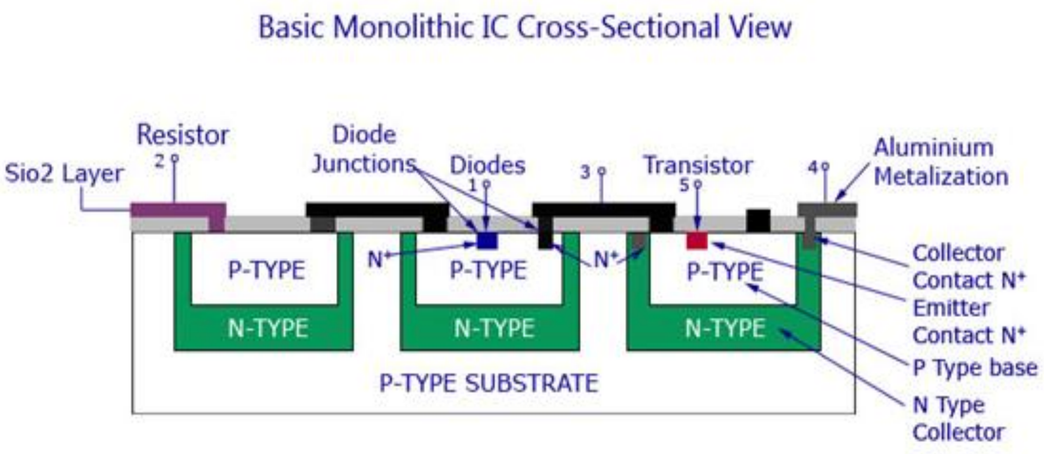


Fig.10.10 Basic Monolithic IC Cross-Sectional view

The layer above N-type is made of silicon dioxide (SiO2) material. Since there is a selective P-type and N-type impurity diffusion going on in the second layer, this layer acts as a barrier in the process. This layer is etched away from the region where diffusion is desired to be permitted with photolithographic process. The rest of the wafer remains protected against diffusion. This layer also protects the silicon layer from contamination.

The up-most layer is that made of aluminium. This metallic layer is used to provide interconnections between the different components used in the IC.

Monolithic IC Manufacturing Process: For the manufacture and production of the monolithic IC, all circuit components and their interconnections are to be formed in a single thin wafer. The different processes carried out for achieving this are explained below.

1. P-layer Substrate Manufacture: Being the base layer of the IC, the P-type is silicon is first built for the IC. A silicon crystal of P-type is grown in dimensions of 250mm length and 25mm diameter. The silicon is then cut into thin slices with high precision using a diamond saw. Each wafer will precisely have a thickness of 200 micrometer and a diameter of 25 mm. These thin slices are termed wafers. These wafers may be circular or rectangular in shape with respect to the shape of the IC. After cutting hundreds of them each wafer is polished and cleaned to form a P-type substrate layer.

2. N-type Epitaxial Growth: The epitaxial groth process of a low resistive N-type over a high resistive P-type is to be carried out. This is done by placing the n-type layer on top of the P-type and heating then inside a diffusion furnace at very high temperature (nearly 1200C). After heating, a gas mixture f Silicon atoms and pentavalent atoms are also passed over the layer. This forms the epitaxial layer on the substrate. All the components required for the circuit are built on top of this layer. The layer is then cooled down, polished and cleaned.

3. The Silicon Dioxide Insulation Layer: As explained above, this layer is required contamination of the N-layer epitaxy. This layer is only 1 micrometer thin and is grown by exposing the epitaxial layer to oxygen atmosphere at 1000C. A detailed image showing the P-type, N-type epitaxial layer and SiO2 layer is given below.

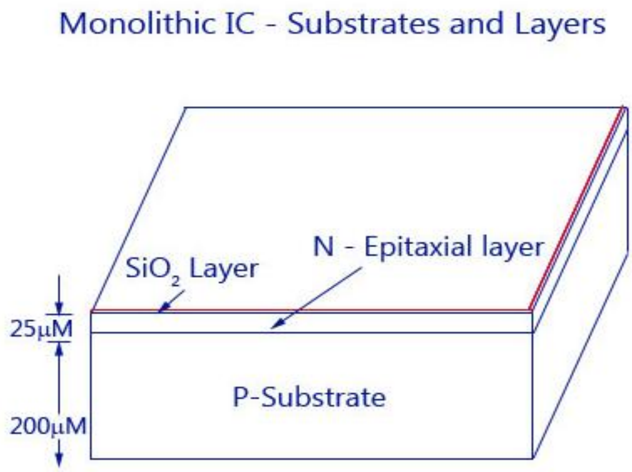


Fig.10.11 Monolithic IC-Substrates and Layers

4. Monolithic IC - Photolithographic-Process (Photolithographic Process for SiO2): To diffuse the impurities with the N-type epitaxial region, the silicon dioxide layer has to be etched in selected areas. Thus openings must be brought at these areas through photolithographic process. In this process, the SiO2 layer is coated with a thin layer of a photosensitive material called photoresist. A large black and white pattern is made in the desired patter, where the black pattern represents the area of opening and white represents the area that is left idle. This pattern is reduced in size and fit to the layer, above the photoresist. The whole layer is then exposed to ultraviolet light. Due to the exposure, the photoresist right below the white pattern becomes polymerized. The pattern is then removed and the wafer is developed using a chemical like trichloroethylene. The chemical dissolves the unpolymerized portion of the photoresist film and leaves the surface. The oxide not covered by polymerised photoresist is then removed by immersing the chip in an etching solution of HCl. Those portions of the Si02 which are protected by the photoresist remain unaffected by the acid. After the etching and diffusion process, with the help of chemical solvents like sulphuric acid, the resist mask is then removed by mechanical abrasion. The appropriate impurities are then diffused through oxide free windows.

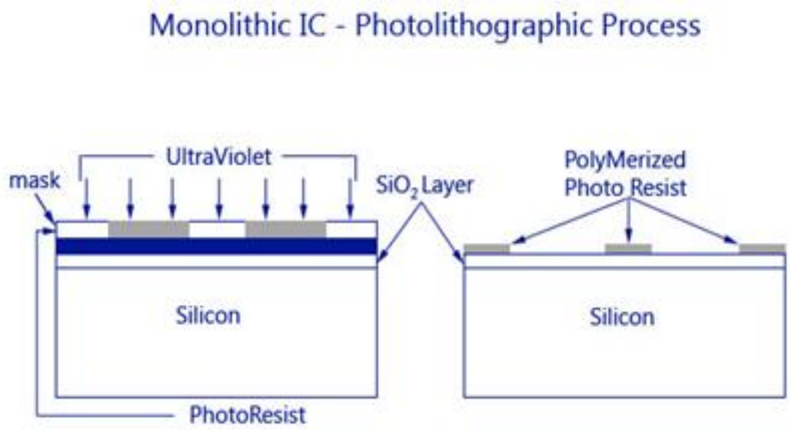


Fig.10.12 Monolithic IC - Photolithographic-Process

5. Monolithic IC - Isolation Diffusion: After photolithographic process the remaining SiO2 layer serves as a mask for the diffusion of acceptor impurities. To get a proper time period for allowing a P-type impurity to penetrate into the N-type epitaxial layer, isolation diffusion is to be carried out. By this process, the P-type impurity will travel through the openings in SiO2 layer, and the N-type layer and thus reach the P-type substrate, Isolation junctions are used to isolate between various components of the IC. The temperature and time period of isolation diffusion should be carefully monitored and controlled. As a result of isolation diffusion, the formation of N-type region called Isolation Island occurs. Each isolated island is then chosen to grow each electrical component. From the figure below you can see that the isolation islands look like back-to-back P-N junctions. The main use if this is to allow electrical isolation between the different components inside the IC. Each electrical element is later on formed in a separate isolation island. The bottom of the N-type isolation island ultimately forms the collector of an N-P-N transistor. The P-type substrate is always kept negative with respect to the isolation islands and provided with reverse bias at P-N junctions. The isolation will disappear if the P-N junctions are forward biased.

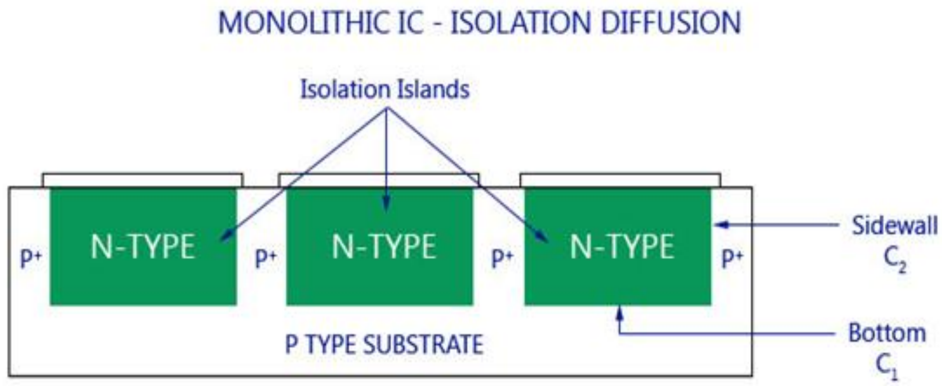


Fig.10.13 Monolithic IC - Isolation Diffusion

An effect of capacitance is produced in the region where the two adjoining isolation islands are connected to the P-type substrate. This is basically a parasitic capacitance that will affect the performance of the IC. This kind of capacitance is divided into two. As shown in the figure C1 is one kind of capacitance that forms from the bottom of the N-type region to the substrate and capacitance C2 from the sidewalls of the isolation islands to the P-region. The bottom component C1 is essentially due to step junction formed by epitaxial growth and, therefore, varies as the square root of the voltage V between the isolation region and substrate. The sidewall capacitance C2 is associated with a diffused graded junction and so varies as (-1/2) exponential of V. The total capacitance is of the order of a few picoFarads.

6. Monolithic IC - Base Diffusion: The working of base diffusion process is shown in the figure below. This process is done to create a new layer of SiO2 over the wafer. P-regions are formed under regulated environments by diffusing P-type impurities like boron. This forms the base region of an N-P-N transistor or as well as resistors, the anode of diode, and junction capacitor. In this case, the diffusion time is so controlled that the P-type impurities do not reach the substrate. The resistivity of the base layer is usually much higher than that of the isolation regions. The isolation regions will have a lot lesser resistivity than that of the base layer.

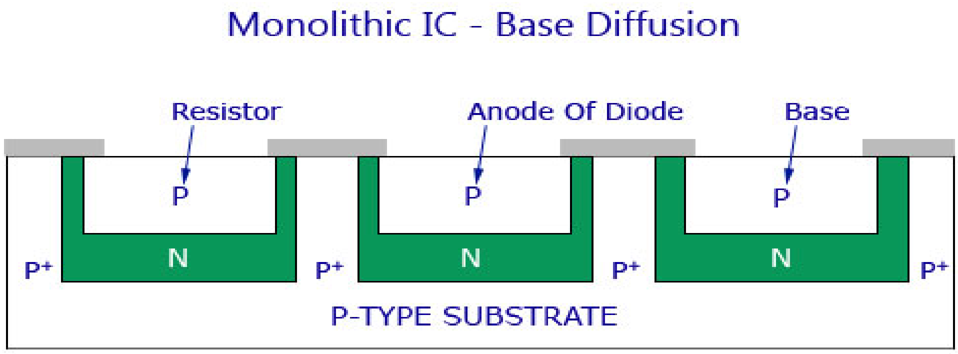


Fig.10.14 Monolithic IC - Base Diffusion

7. Emitter Diffusion: Masking and etching process is again carried out to form a layer of silicon dioxide over the entire surface and opening of the P-type region. The transistor emitters, the cathode regions for diodes, and junction capacitors are grown by diffusion using N-type impurities like phosphorus through the windows created through the process under controlled environmental process. As shown in the figure below there are two additional windows: W1 and W2. These windows are made in the N-region to carry an aluminium metallization process.

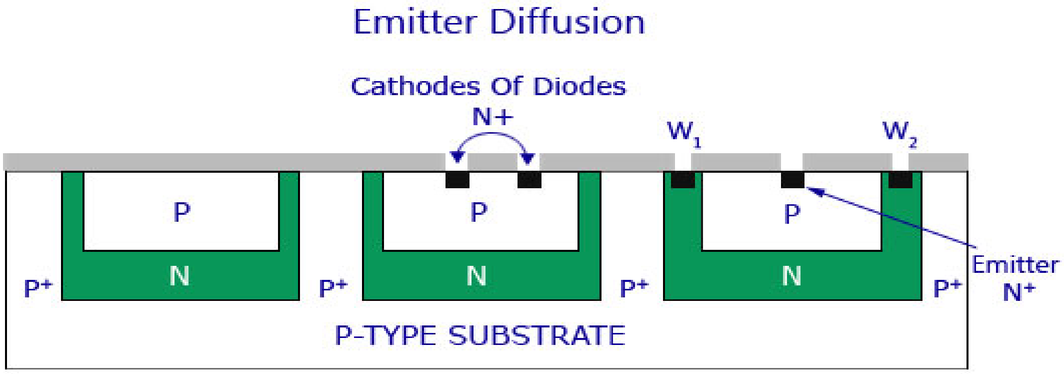


Fig.10.15 Emitter Diffusion

8. Aluminium Metallization: The windows made in the N-region after creating a silicon dioxide layer are then deposited with aluminium on the top surface. The same photoresist technique that was used in photolithographic process is also used here to etch away the unwanted aluminium areas. The structure then provides the connected strips to which the leads are attached. The process can be better understood by going through the figure below.

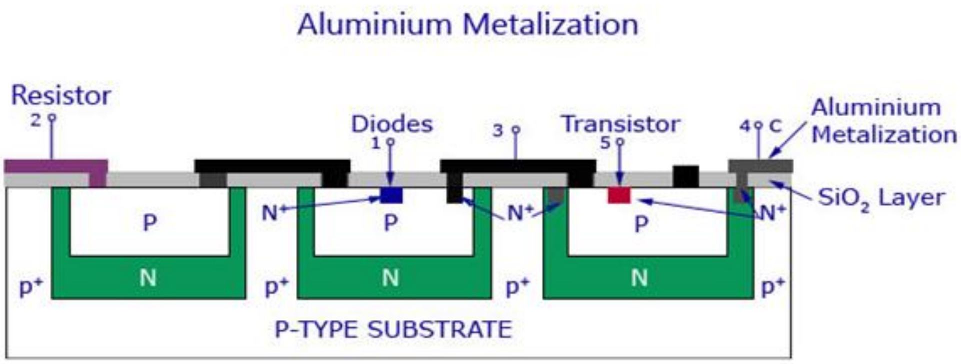


Fig.10.16 Aluminium Metalization

9. Scribing and Mounting: This is the final stage of the IC manufacturing process. After the metallization process, the silicon wafer is then scribed with a diamond tipped tool and separated into individual chips. Each chip is then mounted on a ceramic wafer and is attached to a suitable header. Next the package leads are connected to the IC chip by bonding of aluminium or gold wire from the terminal pad on the IC chip to the package lead. Thus the manufacturing process is complete. Thu, hundreds of IC's is manufactured simultaneously on a single silicon wafer.

Monolithic IC – Component Fabrication

Now we shall discuss in detail how different circuit elements like capacitors, transistors, diodes, and resistors are fabricated into an IC. Please note that it is practically impossible to fabricate an inductor into an IC. It is thus added externally by connecting it to the corresponding IC pin as designed by the manufacturer.

Monolithic IC - Transistor Fabrication: The fabrication process of a transistor is shown in the figure below. A P-type substrate is first grown and then the collector, emitter, and base regions are diffused on top of it as shown in the figure. The surface terminals for these regions are also provided for connection.

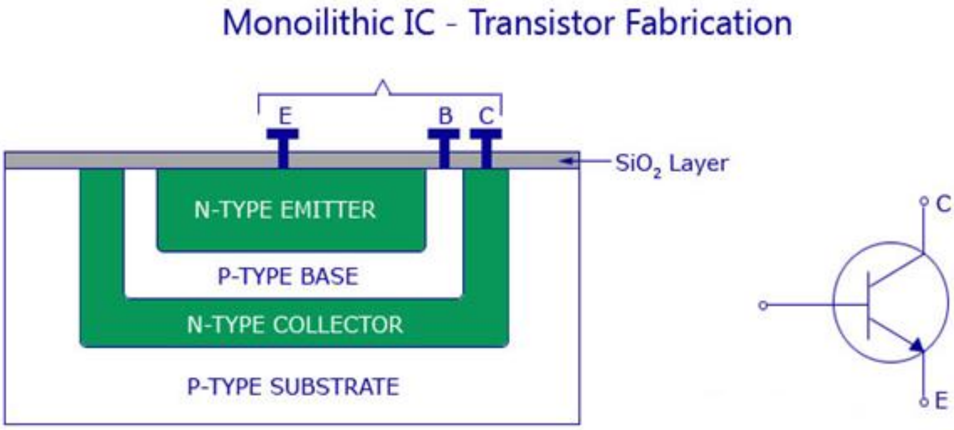


Fig.10.17 Monoilithic IC - Transistor Fabrication

Both transistors and diodes are fabricated by using the epitaxial planar diffusion process that is explained earlier. In case of discrete transistors, the P-type substrate is considered as the collector. `But this is not possible in monolithic IC's, as all the transistors connected on one P-type substrate would have their collectors connected together. This is why separate collector regions are diffused into the substrate.

Even though separate collector regions are formed, they are not completely isolated from the substrate. For proper functioning of the circuit it is necessary that the P-type substrate is always kept negative with respect to the transistor collector. This is achieved by connecting the substrate to the most negative terminal of the circuit supply. The unwanted or parasitic junctions, even when reverse-biased, can still affect the circuit performance adversely. The junction reverse leakage current can cause a serious problem in circuits operating at very low current levels. The capacitance of the reverse-biased junction may affect the circuit high-frequency performance, and the junction break down voltage imposes limits on the usable level of supply voltage. All these adverse effects can be reduced to the minimum if highly resistive material is employed for the substrate. If the substrate is very lightly doped, it will behave almost as an insulator.

Monolithic IC - Diode Fabrication: They are also fabricated by the same diffusion process as transistors are. The only difference is that only two of the regions are used to form one P-N junction. In figure, collector-base junction of the transistor is used as a diode. Anode of the diode is formed during the base diffusion of the transistor and the collector region of the transistor becomes the cathode of the diode. For high speed switching emitter base junction is used as a diode.

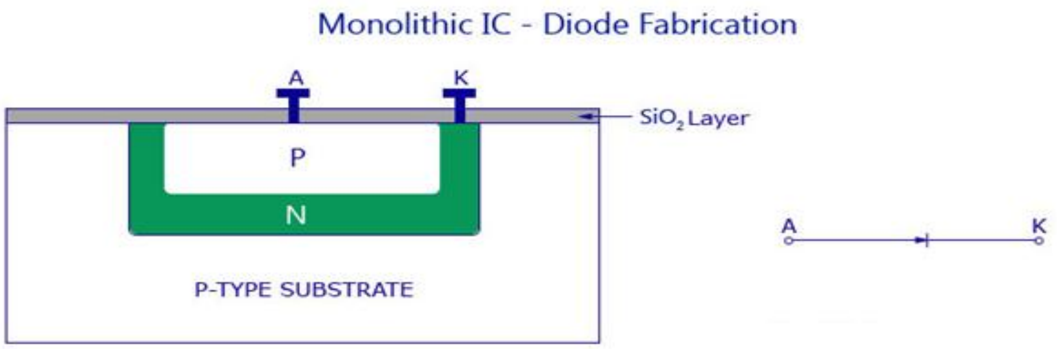


Fig.10.18 Monolithic IC - Diode Fabrication

Monolithic IC - Resistor Fabrication: The resistors used in IC's are given their respective ohmic value by varying the concentration of doping impurity and depth of diffusion. The range of resistor values that may be produced by the diffusion process varies from ohms to hundreds of kilohms. The typical tolerance, however, may be no better than ± 5%, and may even be as high as ± 20%. On the other hand, if all the resistors are diffused at the same time, then the tolerance ratio may be good. Most resistors are formed during the base diffusion of the integrated transistor, as shown in figure below. This is because it is the highest resistivity region. For low resistance values, emitter region is used as it has much lower resistivity.

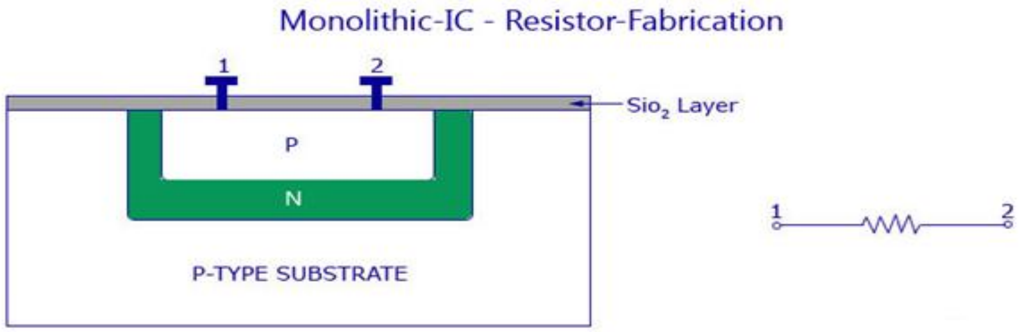


Fig.10.19 Monolithic IC - Resistor Fabrication

Another diffusion technique is also used for the growth of IC resistors. It is basically a thin-film technique. In this process a metal film is deposited on a glass or Si02 surface. The resistance value can be controlled by varying thickness, width and length of the film. Since diffused resistors can be processed while diffusing transistors. This technique is more economic and less time consuming and therefore, the most widely used.

Monolithic IC - Diffused Capacitor Fabrication: The figure below shows the P and N-regions forming the capacitor plates. The dielectric of the capacitor is the depletion region between them.

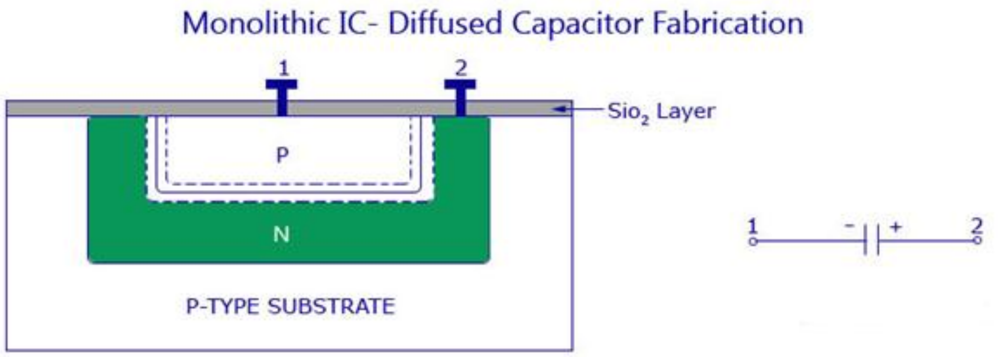


Fig.10.20 Monolithic IC - Diffused Capacitor Fabrication

All P-N junctions have capacitance so capacitors may be produced by fabricating junctions. The amount of change in the reverse bias varies the value of junction capacitance and also the depletion width. The value may be as less as 100 picoFarads.

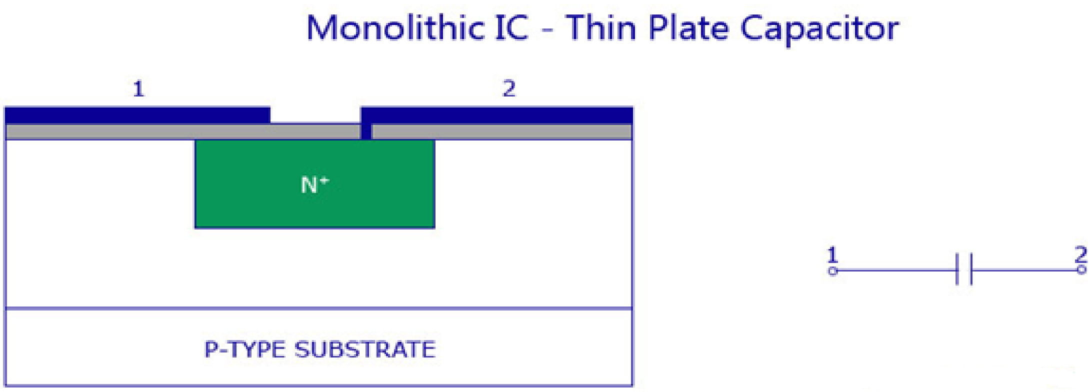


Fig.10.21 Monolithic IC - Thin Plate Capacitor

Using the silicon dioxide as a dielectric may also be a way to fabricate capacitors. One plate of the capacitors is formed by diffusing a heavily doped N-region. The other plate of the capacitor is formed by depositing a film of aluminium on the silicon dioxide dielectric on the wafer surface. For such a capacitor, a voltage of any polarity can be used, and when comparing a diffused capacitor with such a capacitor the diffused capacitor may have very small values of breakdown voltage.

Advantages of Monolithic IC:

- Miniature to the small in size. The IC becomes a lot smaller, when compared to the discrete circuit, it may be at least a thousand times smaller.
- All IC is tested for operating ranges in very low and very high temperatures.
- It Improved the functional performance of the more complex circuits. It can be fabricated for achieving better characteristics.
- Due to its small size, the weight of the IC also reduces when compared to the discrete circuit.
- The small size of the IC causes lesser power consumption and lesser power loss.
- All components are fabricated inside the using chip, so there will not be any external projection ratings.
- All the components are fabricated is very close to each other in an Integrated circuit, they are highly suitable for small-signal operation of a device, as there won't be any stray electrical pickup.
- The PCB consisting of soldered joins will be less reliable. This problem is omitted in IC because of no soldered joints, with fewer interconnections and thus highly reliable.
- As the IC is produced in bulk the temperature coefficients & some other parameters will be closely matching.
- Increased operating speed because of the absence of parasitic capacitance effect.

Disadvantages of monolithic ICs:

- Low-power
- Poor isolation between components.
- NO possibility of fabrication of inductors.
- A small range of values of passive components used in the ICs.

- In this IC's there is a lack of flexibility in circuit design as for making any variation in the circuit, a new set of masks is required.

SAQ.1

a) Discuss about the Introduction of integrated circuit
b) Write the comparison between integrated circuit with discrete circuits.
c) What do you mean by Classification of IC?
d) Discuss about the basics structure of the Monolithic IC.
e) Write the types of fabrication of Monolithic IC.

10.6 Cathode ray oscilloscope (principle, construction, block diagram, working and application):

Principle of Cathode Ray Oscilloscope:

A Cathode Ray Oscilloscope (CRO) is an instrument generally used in a laboratory to display measure and analyze various waveforms of electrical circuits. A cathode ray oscilloscope is a very fast X-Y plotter that can display an input signal versus time or another signal.

Cathode ray oscilloscopes use luminous spots which are produced by striking the beam of electrons and this luminous spot moves in response variation in the input quantity. At this moment one question must be arise in our mind that why we are using only an electron beam? The reason behind this is low effects of beam of electrons that can be used for following the changes in the instantaneous values of rapidly changing input quantity. The general forms of cathode ray oscilloscope operate on voltages.

So the input quantity that we have talked above is voltage. Nowadays, with the help of transducers it is possible to convert various physical quantities like current, pressure, acceleration etc to voltage thus it enable us to have a visual representations of these various quantities on cathode ray oscilloscope**.** Now let us look at the constructional details of the cathode ray oscilloscope.

It is an integral part of modern-day oscilloscopes that are used for measuring various waveforms of electrical circuits. An X-Y plotter plots an input signal against another signal or against time, enabling the study of waveforms, transients, time-based or frequency-based analysis.

Cathode Ray Tube:

A CRO is a complex device consisting of many parts and components. Therefore, before going in depth into it, let's first understand the Cathode Ray Tube (CRT).

A CRT is a vacuum tube that acts as the main part through which the functionalities of a CRO is carried out.

It consists of several parts:

- **Heater** – This part is responsible for heating the cathode.
- **Cathode** – On being sufficiently heated, electrons are emitted from the cathode. In order to achieve this, a layer of barium oxide is coated on one end of the cathode.
- **Grid** – It is kept at a negative potential and helps control the intensity of the electron beam moving towards the anode. The entire structure is made from nickel.
- **Pre-accelerating anode** – Provides acceleration to the anode before entering the Focusing Anode.
- **Focusing Anode** – It helps align the incoming electron beam.
- **Accelerating anode** – Its job is to again accelerate the newly aligned electron beam. Note here that both pre-accelerating and accelerating anodes are connected to a common positive potential of 1500 volts.

All these components collectively form the electron gun.

The beam then moves towards a set of horizontal and vertical deflection plates. These provide an electric field that acts on the electron beam, rendering them a vertical and horizontal motion.

After passing through this final stage, the electron beam strikes the fluorescent screen to give a glow. CRTs generally find applications in forming television screens.

Construction of Cathode Ray Oscilloscope:

The main part of cathode ray oscilloscope is cathode ray tube which is also known as the heart of cathode ray oscilloscope.

It consists of:

- **Cathode Ray Tube** – Through this mechanism, electrons are emitted and controlled to form the desired signal image on the fluorescent screen.
- **Vertical Amplifier** – It amplifies the input signal for display on the CRT screen.

- **Delay Line** – It provides a certain signal delay that is applied to vertical deflection plates of CRT.
- **Trigger Circuit** − It produces a triggering signal for synchronizing both horizontal and vertical deflections of the electron beam.
- **Time base Generator** − It produces a sawtooth signal for horizontal deflection of the electron beam.
- **Horizontal Amplifier** − It amplifies the sawtooth signal and then connects it to the horizontal deflection plates of CRT.
- **Power supply** − It produces both high and low voltages. The negative high voltage and positive low voltage are applied to CRT and other circuits respectively.

Working of Cathode Ray Oscilloscope: The electron gun generates the beam of electrons; this electron beam consists of several electrons moving towards phosphor screen. The control grids are also used in CRT (Cathode Ray Tube) to control the intensity of electrons.

The accelerating anodes are used to increase the velocity of electrons so that they strike the phosphor screen with high speed and thus form a bright spot. The beam creates the luminous spot at the different points on the screen. This becomes easy with the help of deflection plates which deflects the electron beam through various angles.

Through this procedure, a CRO displays the applied input signal on the screen of CRT, providing signals in the time domain.

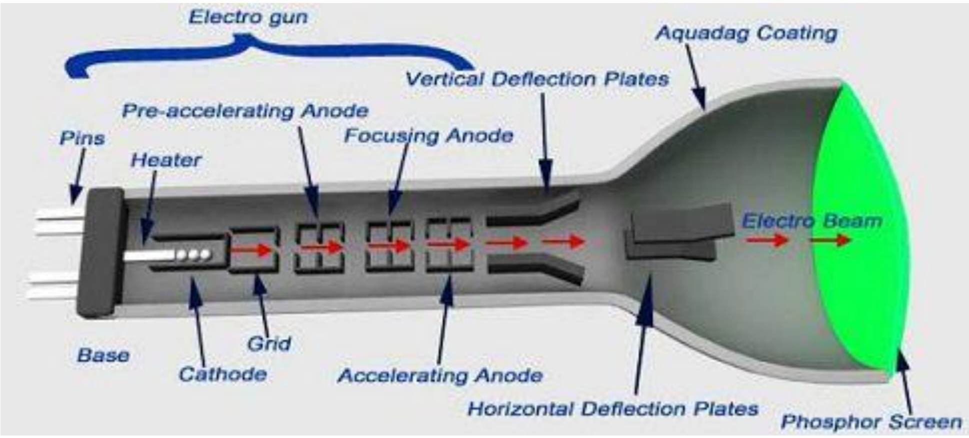


Fig.10.22 Internal structure of CRT

Let us discuss the construction of cathode ray tube in order to understand the construction of cathode ray oscilloscope. Basically the cathode ray tube consists of five main parts:

1. Electron gun
2. Deflection plate system
3. Fluorescent screen
4. Glass envelope
5. Base

You will need all 5 of these components to build your own DIY oscilloscope. We'll now discuss these 5 components in detail:

Electron Gun: It is the source of accelerated, energized and focused beam of electrons. It consists of six parts namely heater, a cathode, a grid, a pre-accelerating anode, a focusing anode and an accelerating anode. In order to obtain the high emission of electrons the layer of barium oxide (which is deposited on the end of cathode) is indirectly heated at moderate temperature. The electrons after this passes through a small hole called control grid which is made up of nickel. As the name suggests the control grid with its negative bias, controls the number of electrons or indirectly we can say the intensity of emitted electrons from cathode. After passing through the control grid these electrons are accelerated with the help of pre-accelerating and accelerating anodes. The pre-accelerating and accelerating anodes are connected to a common positive potential of 1500 volts.

Now after this the function of the focusing anode is to focus the beam of the electrons so produced. The focusing anode is connected to adjustable voltage 500 volts. Now there are two methods of focusing the electron beam and are written below:

1. Electrostatic focusing.
2. Electromagnetic focusing.

Here we will discuss electrostatic focusing method in detail.

Electrostatic Focusing: We know that the force on an electron is given by – qE, where q is the charge on electron ($q = 1.6 \times 10^{-19}$ C), E is the electric field intensity and negative sign shows that the direction of force is in opposite direction to that of electric field. Now we will this force to defect the beam of electrons coming out of electron gun. Let us consider two cases:

Case One:

In this case we are having two plates A and B as shown in the figure.

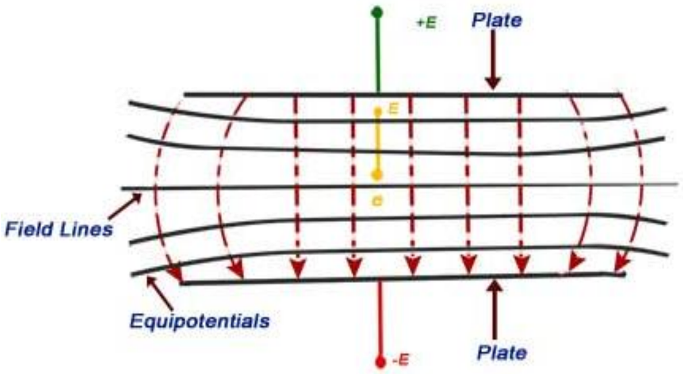


Fig.10.23 Electric field between parallel plates

The plate A is at potential +E while the plate B is at potential –E. The direction of electric field is from A plate to plate B at right angle to the surfaces of the plate. The equipotential surfaces are also shown in the diagram which is perpendicular to the direction of electric field. As the beam of electron passes through this plate system, it deflects in the opposite direction of electric field. The deflection angle can be easily varied by changing the potential of the plates.

Case Second: Here we have two concentric cylinders with a potential difference applied between them as shown in the figure.

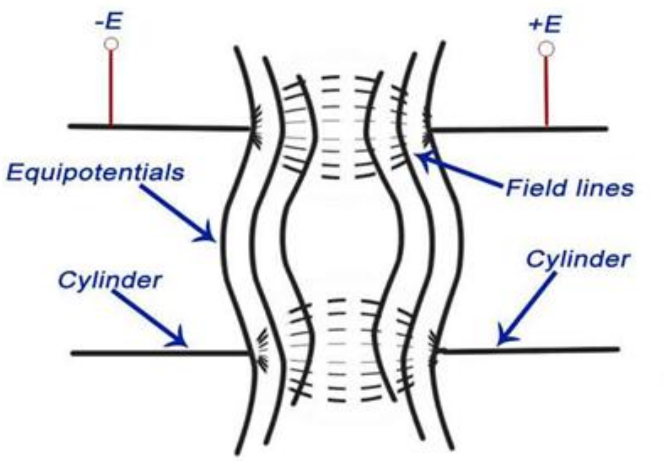


Fig.10.24 Field between two co-axial cylinders

The resultant direction of electric field and the equipotential surfaces are also shown in the figure. The equipotential surfaces are marked by the dotted lines which are curved in shape. Now here we are interested in calculating the deflection angle of electron beam when it passes through this curved equipotential surface. Let us consider the curved equipotential surface S as shown below. The potential on the right side of the surface is +E while the potential on the left side of the surface –E. When a beam of electron is incident at angle A to the normal then it deflects by angle B after passing through the surface S as shown in the figure given below. The normal component of velocity of the beam will increase as force is acting in s direction normal to the surface. It means that the tangential velocities will remain same, so by equating the tangential components we have $V_1 \sin(A) = V_2 \sin(B)$, where V_1 is the initial velocity of the electrons, V_2 is the velocity after passing through the surface. Now we have relation as $\sin(A)/\sin(B)=V_2 / V_1$.

We can from the above equation see that there is bending of the electron beam after passing through the equipotential surface. Therefore this system is also called focusing system.

Electrostatic Deflection:

In order to find out the expression for the deflection, let us consider a system as shown below:

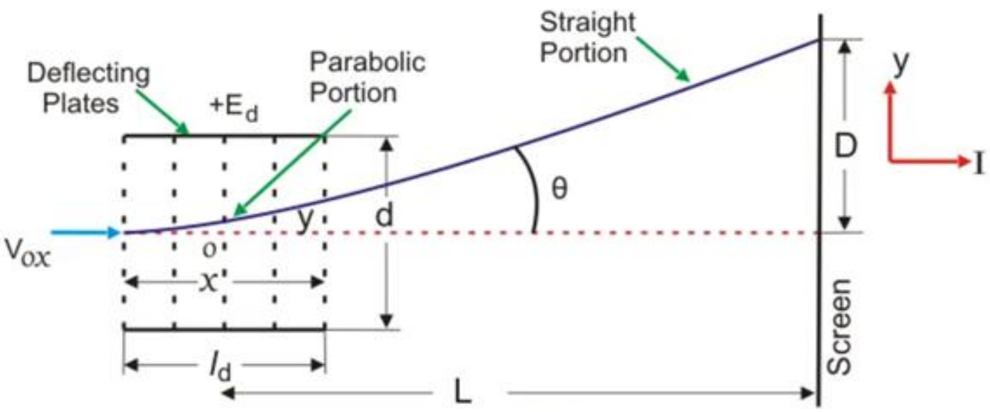


Fig.10.25 Electrostatic Deflection

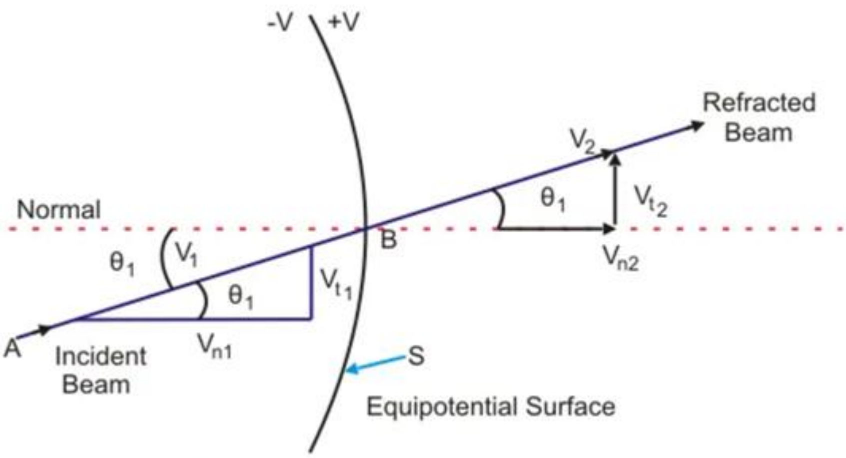


Fig.10.26 Refraction of an Electron Beam

In the above system we have two plates A and B which are at potential +E and 0 respectively. These plates are also called deflection plates. The field produced by these plates is in the direction of positive y axis and there is no force along the x-axis. After deflection plates we have screen through which we can measure net deflection of the electron beam. Now let us consider a beam of electron coming along the x-axis as shown in the figure. The beam deflects by angle A, due presence of electric field and deflection is in the positive direction of y axis as shown in the figure. Now let us derive an expression for deflection of this beam. By the conservation of energy, we have loss in potential energy when the electron moves from cathode to accelerating anode should be equal to gain in kinetic energy of electron. Mathematically we can write,

$$eE = \frac{1}{2}mv^{\frac{1}{2}} \quad \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

Where, e is the charge on electron, E is the potential difference between the two plates, m is the mass of electron, and v is the velocity of the electron. Thus, eE is loss in potential energy and $1/2mv^{1/2}$ is the gain in kinetic energy. From equation (1) we have velocity $v = (2eE/m)^{1/2}$. Now we have electric field intensity along the y axis is E/d, therefore force acting along the y axis is given by F = eE/d where d is the separation between the two deflection plates. Due to this force the electron will deflect along the y axis and let the deflection along y axis be equal to D which is marked on the screen as shown in the figure. Due to the force F there is net upward acceleration of the electron along positive y axis and this acceleration is given by Ee/(d × m).Since the initial velocity along positive y direction is zero therefore by equation of motion we can write the expression of displacement along y axis as,

$$y = \frac{1}{2}\left(\frac{Ee}{m \times d}\right) \times t^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots (2)$$

As the velocity along the x direction is constant therefore we can write displacement as,

$$x = u \times t \cdots\cdots\cdots\cdots\cdots\cdots (3)$$

Where, u is velocity of electron along x axis. From equations 2 and 3 we have,

$$y = \frac{1}{2}\left(\frac{eE}{mu^2}\right) \times x^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots (4)$$

Which is the equation of trajectory of the electron. Now on differentiating the equation 4 we have slope i.e.

$$\frac{dy}{dx} = \frac{eEl}{mu^2}$$

Where, l is the length of the plate. Deflection on the screen can be calculated as,

$$D = L \times \frac{dy}{dx}$$

Distance L is shown in the above figure. Final expression of D can be written as,

$$D = \frac{LlE}{2dE}$$

From the expression of deflection, we calculate deflection sensitivity as,

$$\frac{D}{E} = \frac{Ll}{2dE}$$

Graticule: These are the grid of lines whose function is to serve as a scale when the cathode ray oscilloscope is used for the amplitude measurements. There are three types of graticules and they are written below:

1. **Internal Graticule:** Internal graticule as name suggests deposited on the internal surface of the cathode ray tube face plate. There is no problem of parallax errors but we cannot change internal graticules as they are fixed.
2. **External graticule:** Given below is the circuit diagram of cathode ray oscilloscope.

Block Diagram of Cathode Ray Oscilloscope:

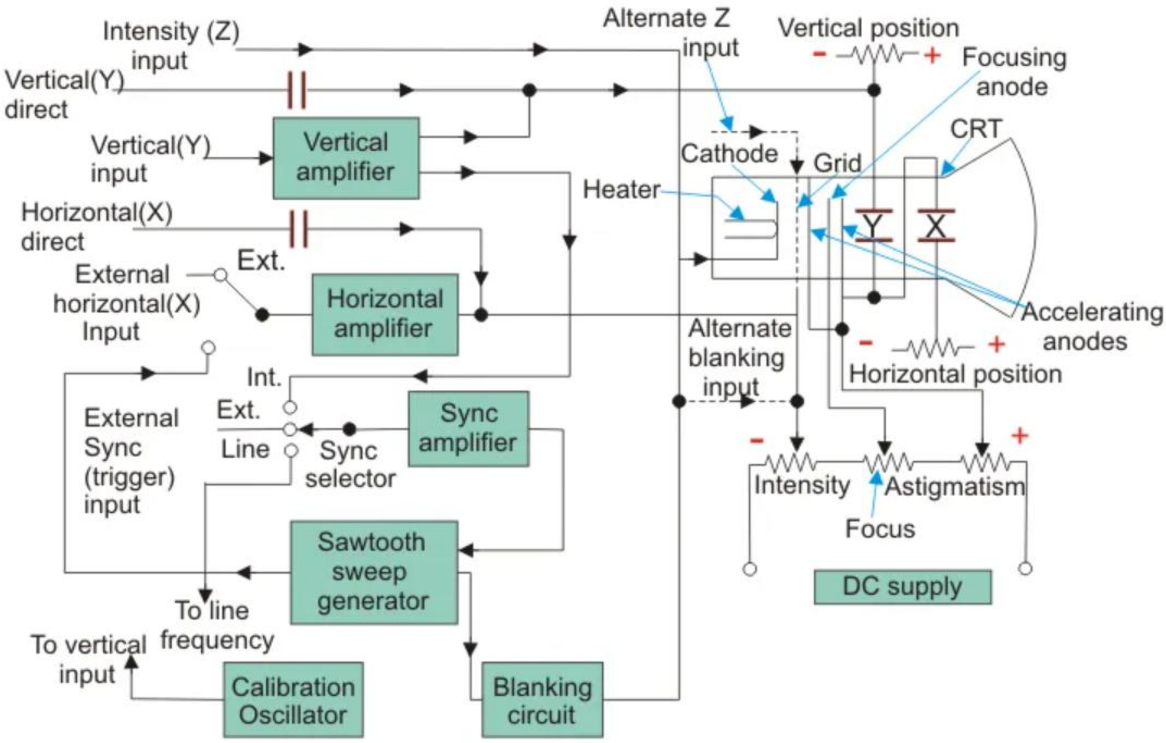


Fig.10.27 Basic Block Diagram of Cathode Ray Oscilloscope

Now we will study the basic circuit diagram of cathode ray oscilloscope under the following main parts.

1. **Vertical Deflection System:** The input signal for examining are fed to the vertical deflection system plates with the help of input attenuator and a number of amplifier stages. The main function of these amplifiers is to amplify the weak the weak signals so that the amplified signal can produce the desirable signals.
2. **Horizontal Deflection System:** Like the vertical system horizontal system also consists of horizontal amplifiers to amplify the weak input voltage signals but in contrast to vertical deflection system, horizontal deflection plates are fed by a sweep voltage that provides a time base as shown above. As shown in the circuit diagram, the saw tooth sweep generator is triggered by the synchronizing amplifier when the sweep selector switch is in the internal position and thus the triggered saw tooth generator gives input to the horizontal amplifier by following this mechanism. Now there are four types of sweeps:
   - Free Running or Recurrent Sweep: As the name suggests, the saw tooth waveform is repetitive i.e. a new sweep is started immediately after the previous sweep.
   - Triggered Sweep: Some time the waveform to be observed may not be periodic so it is desired that the sweep circuit remain inoperative and the sweep be initiated by the waveform under examination. In such cases we use triggered sweep.

- Driven Sweep: Generally a driven sweep is used where the sweep is free running but triggered by the signal under test.
- Non Saw Tooth Sweep: This is used for finding the phase difference between the two voltages. Another important application is that we can compare frequency of input voltages using non saw tooth sweep.

3. Synchronization: There must be synchronization between the sweep and the signal being measured. Synchronization is done to produce stationary pattern. There are three sources of synchronization which can be selected by synchronization selector and they are written below:

1. Internal: In this trigger is obtained from the signal being measured through vertical amplifier.
2. External: In this trigger an external trigger source is required.
3. Line: In this method trigger is obtained power supply.
4. Intensity Modulation: Intensity modulation can be done by inserting the signal between the ground and the cathode. Intensity modulation causes the brightening of the display.
5. Positioning Controls: Position can be control by applying small independent internal direct voltage sources to the deflecting plates and with the help of potentiometer (using it as voltage divider) we can control the position of signal.
6. Focus Control: Focus can be controlled by changing the focal length of the focusing electrode which acts like a lens and focal length can be changed by the changing potential of the focusing anode.
7. Intensity Control: The intensity can be varied by changing the grid potential with respect to cathode.
8. Calibration Circuit: Calibrating voltage has a square shape which is usually internally generated of known amplitude.
9. Astigmatism: By adjusting the focus the spot can be made sharp in order to avoid the problem of astigmatism.

Applications:

- While numerous, a CRO can be used for the following purposes:
- To determine the amplitude of a waveform.
- Comparison between the phases and frequencies of electrical signals.

- Help measure capacitance and inductance values.
- In the medical field and medical trials, it can be used for monitoring various body parameters like heartbeat rates and nervous reactions.

10.6 Multimeter (principle, types, construction and function): A Multimeter is an electronic instrument, every electronic technician and engineer's widely used piece of test equipment. A multimeter is mainly used to measure the three basic electrical characteristics of voltage, current, and resistance. It can also be used to test continuity between two points in an electrical circuit. This post mainly introduces the basic information of multimeters, applications, and types of multimeters are in. Let's see all of these.

The multimeter has multi functionalities like, acts like ammeter, voltmeter, and ohmmeter. It is a handheld device with positive and negative indicator needles over a numeric LCD digital display. Multimeters can be used for testing batteries, household wiring, electric motors, and power supplies.

The essential parts of the multimeter mainly include a display, power source, probes, and controls.

How to use a Multimeter?

The function and operation of a multimeter are similar for both analog and digital types. This instrument includes two leads or probes namely red and black & three ports. The black color lead is used to plug into the common port, whereas the red color leads plug into other ports based on the requirement.

Once the leads are plugged in, the knob can be switched ON in the center of the instrument so that the appropriate function can be done for the specific component test. For instance, once the knob is situated to 20V DC, then the multimeter will notice DC voltage up to 20V. To calculate low voltages, then set the knob in the multimeter to the 2V/200mV range.

To obtain a reading from the meter, you need to touch the end of each probe to the end of the terminals of components. Types of multimeter devices are very safe to utilize on devices and circuits to provide the current or voltage that does not go above the highest rating of the meter.

While measuring, we must be very cautious so don't touch the bar ends of the metal in the tester when activated otherwise you will get an electrical shock.

Functions of Multimeters:

These instruments are capable of different readings based on the model. So basic types of multimeter are mainly used to measure amperage, resistance, voltage, checks continuity and a complete circuit can be tested like the following.

- Resistance in Ohms
- Capacity in Farads
- The temperature in Fahrenheit/ Celsius
- AC Voltage & Amperage
- Inductance Henrys
- DC Voltage & Amperage
- Frequency in Hz
- Conductance in Siemens
- Decibels
- Duty Cycle

To some types of multimeters, special sensors or accessories can be attached for extra readings like acidity, light level, alkalinity, wind speed & relative humidity.

Multimeter Operating Instructions: The combination volt-ohm-milliammeter is a basic tool in any electronic laboratory. The proper use of this instrument increases its accuracy and life. The following precautions should be observed.

1. To prevent meter overloading and possible damage when checking voltage or current, start with the highest range of the instrument and move down the range successively.
2. For higher accuracy, the range selected should be such that the deflection falls in the upper half on the meter scale.
3. For maximum accuracy and minimum loading, choose a voltmeter range such that the total voltmeter resistance (ohms per volt x full scale voltage) is at least 100 times the resistance of the circuit under test.
4. Make all resistance readings in the uncrowded portion on the meter scale, whenever possible.
5. Take extra precautions when checking high voltages and checking current in high voltage circuits.
6. Verify the circuit polarity before making a test, particularly when measuring dc current or voltages.
7. When checking resistance in circuits, be sure power to the circuit is switched off, otherwise the voltage across the resistance may damage the meter.

8. Renew ohmmeter batteries frequently to insure accuracy of the resistance scale.
9. Re-calibrate the instrument at frequent intervals.
10. Protect the instrument from dust, moisture, fumes and heat.

Types of Multimeter: There are different types of multimeters like Analog, Digital, and Fluke multimeters.

Analog Multimeter: The Analog Multimeter or VOM (Volt-Ohm-Milliammeter) is constructed using a moving coil meter and a pointer to indicate the reading on the scale. The moving coil meter consists of a coil wound around a drum placed between two permanent magnets.

As current passes through the coil, the magnetic field is induced in the coil which reacts with the magnetic field of the permanent magnets and the resultant force causes the pointer attached to the drum to deflect on the scale, indicating the current reading. It also consists of springs attached to the drum which provides an opposing force to the motion of the drum to control the deflection of the pointer.

Fig.10.28 Analog Multimeter

For the measurement of DC, the D Arsonval movement described above can be directly used. However, the current to be measured should be lesser than the full-scale deflection current of the meter. For higher currents, the current divider rule is applied. Using different values of shunt resistors, the meter can also be used for multi-range current measurements. For current measurement, the instrument is to be connected in series with the unknown current source.

For measurement of DC voltage, a resistor is connected in series with the meter, and the meter resistance is taken into account such that the current passing through the resistor is the same as the current passing through the meter and the whole reading indicates the voltage reading. For voltage measurement, the instrument is to be connected in parallel with the unknown voltage source. For multirange measurement, different resistors of different values can be used, which are connected in series with the meter.

For measurement of resistance, the unknown resistance is connected in series with the meter and across a battery, such that the current passing through the meter is directly proportional to the unknown resistance. For AC voltage or current measurement, the same principle is applied, except for the fact that the AC parameter to be measured is first rectified and filtered to get the DC parameter and the meter indicates the RMS value of the AC signal.

Advantages of an Analog Multimeter are that it is inexpensive, doesn't require a battery, can measure fluctuations in the readings. The two main factors affecting the measurement are sensitivity and accuracy. Sensitivity refers to the reciprocal of the full-scale deflection current and is measured in ohms per volt.

Digital Multimeters: We mostly used a multimeter is a digital multimeter (DMM). The DMM performs all functions from AC to DC other than analog. It has two probes positive and negative indicated with black and red color is shown in the figure. The black probe connected to COM JACK and the red probe connected by user requirement to measure ohm, volt, or amperes.

The jack marked VΩ and the **COM** jack on the right of the picture are used for measuring voltages, resistance, and for testing a diode. The two jacks are utilized when an LCD shows what is being measured (volts, ohms, amps, etc.). Overload protection prevents damage to the meter and the circuit and protects the user.

Fig.10.29 Digital Multimeter

The Digital Multimeter consists of an LCD, a knob to select various ranges of the three electrical characteristics, an internal circuitry consisting of a signal conditioning circuitry, an analog to digital converter. The PCB consists of concentric rings that are connected or disconnected based on the position of the knob. Thus as the required parameter and the range are selected, the section of the PCB is activated to perform the corresponding measurement.

To measure the resistance, current flows from a constant current source through the unknown resistor, and the voltage across the resistor are amplified and fed to an Analog to Digital Converter and the resultant output in form of resistance is displayed on the digital display. To measure an unknown AC voltage, the voltage is first attenuated to get the suitable range and then rectified to DC signal and the analog DC signal is fed to an A/D converter to get the display, which indicates the RMS value of the AC signal.

Similarly to measure an AC or DC, the unknown input is first converted to a voltage signal and then fed to an analog to digital converter to get the desired output(with rectification in case of AC signal). Advantages of a Digital Multimeter are its output display which directly shows the measured value, high accuracy, ability to read both positive and negative values.

Block Diagram of Digital Multimeter: The following figure represents the digital multimeter block diagram with all the functional blocks.

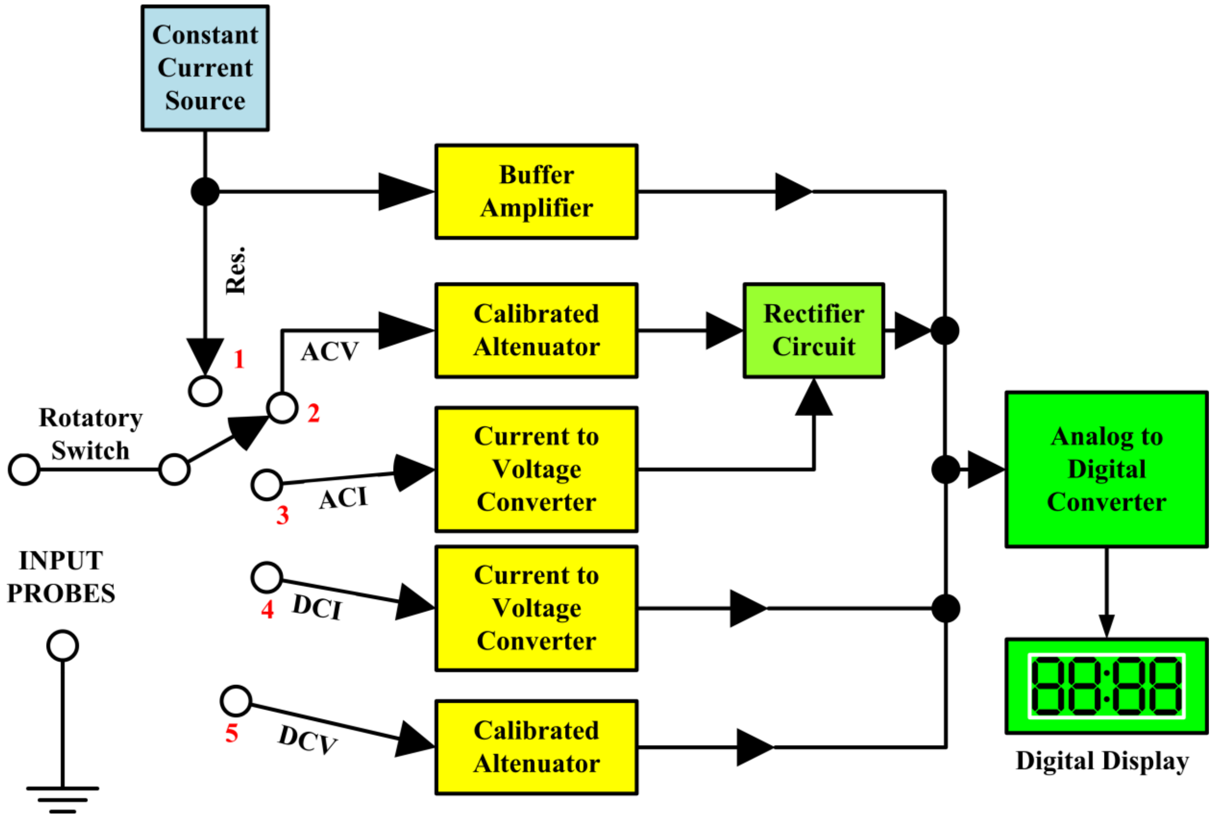


Fig.10.30 Digital Multimeter Block Diagram

The auto-ranging digital multimeter (DMM) only requires you to choose your electrical quantity you are attempting to measure, make sure you are properly placing your leads into the correct terminals and then reading the LCD display. Auto-ranging digital multimeters allow technicians to spend more time getting to the root of a problem instead of switching and calculating.

Step by Step Guide on how to use a multimeter to test for voltage: Testing for voltage is carried out to ensure the effectiveness of the electrical system. Loads (for example lights or motors) that are designed to do the work need a nominal voltage to operate. Overvoltage will result in equipment failure and not enough voltage will result in the load not turn on. When testing voltage there is an expected voltage reading to look for. If the load is rated at 120 volts then the expected reading from the outlet needs to be 120 volts plus or minus 10%. If the voltage reading is out of specifications then the problem can be found using the voltmeter to isolate the load and find if there is a problem with the source or the load.

Here is a step by step guide on how to use a multimeter to test for voltage:

1. First, figure out whether the application being testing utilizes AC or DC voltage. Afterward, adjust the meter dial to the suitable function to DC Voltage or AC voltage.
2. Adjust the range to the number little higher than the predictive value. If the value being measured is unknown, then set the range to the maximum available number.
3. Plug in the test leads into the common (black) and voltage (red) terminals.
4. Apply the leads to the test circuit.
5. Position and reposition the test till a dependable reading appears on the meter LCD.
6. While measuring AC voltage, variations may happen in the reading. As the test continues the measurement will steady.

Step by Step Guide on how to use a multimeter to test for Current: Testing for current is used when there is no physical way to tell if a load is doing its job because there are no indicators or the load is located in a hazardous area. When the voltage is tested and found to be present at the load, it doesn't tell the whole story until a current is measured. It is important to understand a load consumes power which is measured in watts. Watts is calculated by multiplying volts by the amps. A digital multimeter is used to measure or give a good indication of current flowing.

Testing current with a Digital Multimeter:

1. Plug the leads into the terminals marked mA for low current or A for currents over 500mA.
2. Set the dial to AC or DC current depending on the circuit being measured.
3. Apply the leads to the open circuit current and observe the measurement.

To test resistance with a Digital Multimeter:

1. Turn the power off in the circuit being tested.
2. Adjust the meter dial to the resistance mode.
3. Choose the suitable range on the dial.
4. Plug in your test leads into the suitable terminals.
5. Connect the leads to the component being tested and note a reading.

Test continuity with a Digital Multimeter:

1. Adjust the dial to the meter continuity (the little speaker) function.

2. Plug the test leads into the suitable terminal.
3. Touch the component under test using the leads

The DMM beeps under good continuity that allows the flow of current. If no continuity exists, the DMM does not beep.

Types of Digital Multimeter: Digital types of multimeter are available in three types.

Fluke Multimeter: The fluke digital multimeter can be designed with various collaboration functions. Generally, it includes a large display and this instrument is used to measure the voltage as well as electrical resistance. Some kinds of devices are available with advanced features to measure humidity, duty cycle, pressure, frequency temperature, etc. The fluke multimeter is one of the most frequently and famous instruments. This kind of multimeter is mainly used for calibration efforts and used to calibrate currents, volts & other electrical units.

Fig.10.31 Fluke Multimeter

The fluke multimeters are protected against the transient voltage. It is a small portable device used to measure voltage, current, and test diodes. The multimeter has multi selectors to select the desired function. The fluke MM automatically ranges to select most measurements. This means the magnitude of the signal does not have to be known or determined to take an accurate reading, it directly moved to the appropriate port for the desired measurement. The fuse is protected to prevent damage if connected to the wrong port.

Clamp Digital Multimeter: *The clamp digital multimeter is used to measure the electricity flow. As the name suggests, this multimeter includes the feature namely clamp which measures the amps whenever the probes measure the volts. The adjustment of power utilization otherwise watts can be done through multiplying the reading of voltage with the amps. This multimeter also includes an additional feature that is different kinds of settings. The appropriate feature is used while measuring.*

Fig.10.32 Clamp Type

This kind of multimeter includes fixed tools for measuring the current flow. This device extremely changes from the fluke type because, in the fluke multimeter, it utilizes a clamp to measure the flow of current. So, this instrument is usually recommended for professionals only.

Current can be tested in several ways; the most reliable procedure is using a clamp meter, shown in fig.

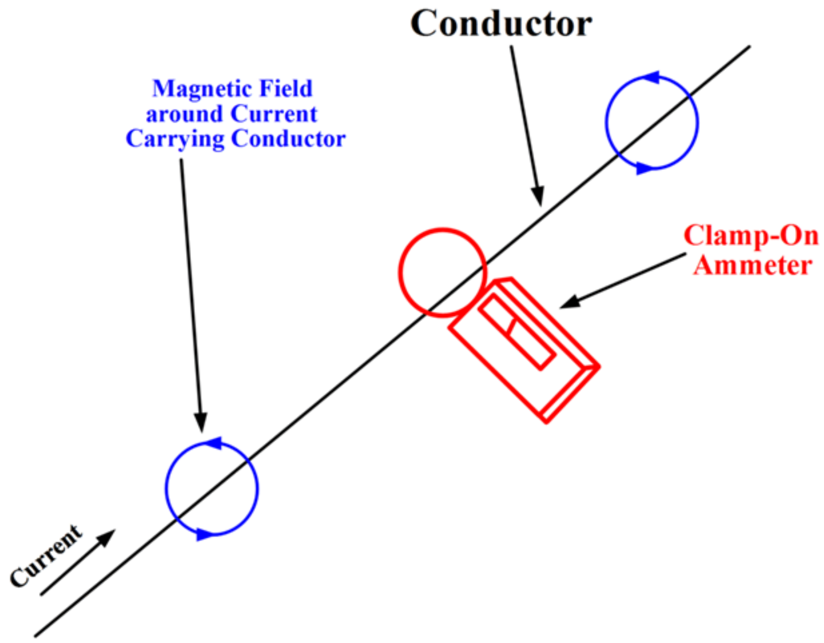


Fig.10.33 Current tested by Clamp Meter

The advantage of using clamp meter is that measurements can be obtained even without opening the test circuit. Proper protective equipment must be worn before testing can be done.

1. To test for current, determine the type of current if it is AC or DC.
2. Afterward, adjust the meter dial to the suitable function to DC current or AC current.
3. Adjust the range on the dial except it is an auto-ranging meter.

Testing current with a clamp meter

1. Press the thumb lever to open the clamp meter head
2. Close the head once it is around a single conductor and then release the metering lever.
3. Now observe the reading.

Autoranging Multimeter: *The auto-ranging multimeter is a simple multimeter to utilize even though it is similarly the most costly of all kinds of digital multimeters. This multimeter includes a knob in the center and has less position. So it doesn't switch automatically to measure. This instrument is applicable in simple projects. For beginners as well as electricians at home, this instrument is highly recommended. Generally, it measures a single component at a time.*

Fig.10.34 Autoranging Type

Types of Multimeter Probes: A multimeter includes different test probes and the main function of these probes is to connect to the circuit under test. The most common types of probes are retractable hook clips, pointed probes & crocodile clips.

Generally, a multimeter includes two-color wires like black and red, known as leads or probes. One end of the probe is called a banana jack that is plugged into a multimeter, whereas the remaining end is known as the probe tip, used to test the circuit. The red probe is used for +ve whereas the black probe is used for –Ve.

These probes include a probe tip on one end whereas the other end includes banana plugs. Most of the multimeters include fuses to guard them against the extremely high current. When too much current supplies through the multimeter, this fuse will restrict the flow of current to prevent the damage. Some kinds of multimeters include fuses based on the measurement of low current or high current and they determine where you have to place the probes.

Working of Multimeter: *Types of multimeter include two probes like red and black & two or three ports. From them, one of the ports is labeled.COM for common which is used for black probe whereas the remaining ports are labeled A used for amps and mA/µA (milliamps/microamps). The final port is labeled VΩ used for ohms & volts. Sometimes, this port is integrated into the 3rd one, which is next labeled mAVΩ.*

If the multimeter includes four ports, then the red probe can be plugged into the VΩ port for measuring resistance as well as voltage. When the red probe is inserted into the mA port then the current can be calculated & plugged into the A port then the current can be measured in amps. For instance, the port used to test a diode using a multimeter is the VΩ port and this port can also be used to test a transistor.

Difference between Analog Multimeter and Digital Multimeter:

The main difference between analog and digital multimeters include the following.

Analog Multimeter	**Digital Multimeter**
Analog Multimeter is used to gauge restricted electrical quantities like resistance, voltage & current.	Digital Multimeter is used to compute various electrical quantities like voltage, current, capacitance, resistance, values of diode and impedance, etc.
The size of the analog multimeter is larger	The size of the digital multimeter is smaller
This meter provides the reading on a scale next to the pointer.	This meter provides the reading in the form of numeric on an LCD.
These are calibrated manually.	These are calibrated automatically.
Its construction is simple	Its construction is complicated because of the involvement of components like electronics and logic.
Analog multimeters are less accurate because of the parallax errors & readings of the wrong pointer	Digital multimeters are very accurate
It doesn’t need ADC to show reading.	It needs ADC to exhibit the reading.
Input resistance is not stable	Input resistance is stable
The pointer of this multimeter tries to turn aside to the left in reverse polarity.	This multimeter shows a negative quantity once the polarity is reversed.
These are less cost	These are expensive
The o/p of this meter cannot be interfaced through exterior equipment.	The o/p of these meters can be interfaced through exterior equipment.
The frequency range is up to2kHZ.	The frequency range is high as compared to analog
Analog Multimeter measures the current with the help of a Galvanometer.	Digital Multimeter measures voltage with ADC
It has less electric noise	It has more electric noise

It allows simply one i/p signal for each operation.	It allows several input signals & consumers can select the required signal on the variable display.
The maximum AC frequency which can be calculated is lesser	The maximum AC frequency which can be calculated is high than its counter element

Advantages and Disadvantages of Digital Multimeters:

The advantages of a digital multimeter include the following.

- It gives an automatic o/p display.
- The measurement results of the meter can record and store in memory and synchronizes through a PC
- It includes auto polarity functions
- The meter reading accuracy cannot depend on the charging of the battery
- It ensures accuracy
- Resistance toward mechanical damage.
- Multifunctionality
- Zero adjustments cannot be required
- Measurement accuracy is high
- Measuring ranges can be selected through manual or automatically

The disadvantages of digital multimeter include the following

- As compared to analog, it is expensive
- This multimeter does not work properly through measurement fluctuations. It can be tricky to discover one for your exact needs.

Advantages and Disadvantages of Analog Multimeter:

The advantages of an analog multimeter include the following.

- Possibility of achieving measurements at below-30 ° C temperature
- Power utilization is not required from the fixed power supply while measuring current and voltage
- When high precision is not necessary, then quick operation through a large amount of measurement can be done.
- By using this instrument, all measurements can be done simply.

- The signal level can be observed.

The disadvantages of analog multimeter include the following:

- These meters are large
- These are expensive
- Voltage polarity cannot be recognized
- They are susceptible to vibration or shock.
- The movement of the pointer is slow and it cannot be utilized to measure voltages through frequencies above 50 HZ.
- Incorrect because of the earth's magnetic field effect.
- An unexpected change in the signal can notice through an analog multimeter more quickly as compared with a digital multimeter.
- These are sensitive to vibration, mechanical damage.
- Input resistance is less, thus a high error while measuring less voltage

Types of Multimeter Applications:

The applications of types of multimeter mainly involve various electrical and electronic projects for components testing and also used in different measurement applications in the multimeter.

Temperature and Environmental Applications:

- Low-cost weather station
- DMM internal temperature

Voltage Measurements:

- High and low-value DC measurement
- Peak to Peak and DC average measurement

Current Measurements:

- DC measurement
- True RMS AC

Resistance Measurement:

- Micro ohmmeter
- Measuring resistance with constant voltage
- Measuring resistance with constant current

Time and Frequency measurement:

- Fast frequency

- Time measurement

10.8 Ultrasonics (production, detection, velocity measurements and applications), Hypersonics and ultrasonics:

Ultrasonic Waves:

- The sound waves of frequency greater than 20 KHz are called Ultrasonic Waves.
- These sound waves are inaudible to the human ear.
- The ultrasonic waves due to their shorter wavelength have a greater penetrating power.
- Ultrasonic waves are widely used in medical diagnosis, marine application, NDT, etc.

Important Properties of Ultrasonic Waves:

1. Ultrasonic waves vibrate at a frequency greater than the audible range for humans (20 kilohertz).
2. They have smaller wavelengths. As a result, their penetrating power is high.
3. They cannot travel through vacuum.
4. Ultrasonic waves travel at the speed of sound in the medium. They have maximum velocity in a denser medium.
5. In a homogeneous medium, they travel at a constant velocity.
6. In low viscosity liquids, ultrasonic waves produce vibrations.
7. They undergo reflection, refraction and absorption.
8. They have high energy content. They can be transmitted over a large distance without much loss of energy.
9. They produce intense heat when they are passed through objects.
10. Like sound waves, ultrasonic waves are longitudinal waves that produce alternate compressions and rarefactions.

Production of Ultrasonic Waves: Based on frequency range and power output, the ultrasonic wave generators are divided into two groups.

1. Mechanical Generator
2. Electrical Generator

In this section, the electrical generator methods of producing ultrasonic waves are explained.

The electrical generators are subdivided into two categories.

(i) Magnetostriction generator

(ii) Piezoelectric generator

1. Magnetostriction method to produce ultrasonic waves:

Principle of Magnetostriction effect:

- When a ferromagnetic material in the form of a rod is subjected to an alternating magnetic field parallel to its length, the rod undergoes alternate contractions and expansions at a frequency equal to the frequency of the applied magnetic field. This phenomenon is known as the magnetostriction effect.
- Due to resonance, the rod is thrown into longitudinal vibrations, thereby producing ultrasonic waves in the surrounding medium. Such ferromagnetic materials which are used for the production of ultrasonic waves are called magnetostriction materials.

Merits of Magnetostriction method:

1. The design of this generator is very simple and the production cost is low.
2. At low ultrasonic frequencies, large power output is possible without the risk of damage to the circuit.
3. Frequencies ranging from 100Hz to 3000KHz can be produced.

Demerits of Magnetostriction method:

1. It cannot generate ultrasonic of frequency above 3000KHz.
2. The frequency of oscillations depends greatly on temperature.
3. There will be losses of energy due to hysteresis and eddy current.

2. Production of Ultrasonic Waves by Piezoelectric method:

The piezoelectric method of producing ultrasonic waves is based on the principle of inverse piezoelectric effect.

Principle of Inverse piezoelectric effect:

- If an alternating voltage is applied to one pair of opposite faces of the crystal, alternatively mechanical contractions and expansions are produced in the crystal and the crystal starts vibrating. This phenomenon is known as the inverse piezoelectric effect or electrostrictive effect.
- If the frequency of the applied alternating voltage is equal to the vibrating frequency of the crystal, then the crystal will be thrown into resonant vibration producing ultrasonic waves.

Detection of Ultrasonic Waves:

The presence of ultrasonic waves can be detected by using any one of the following methods:

1. Quartz crystal method
2. Thermal detection method
3. Sensitive flame method
4. Kundt's tube method

1. Quartz Crystal Method:

- This method of detecting ultrasonic waves is based n the principle of the piezoelectric effect.
- When one pair of opposite faces of the quartz crystal is exposed to ultrasonic waves, in the other pair of opposite faces, charges get developed.
- These charges are amplified and detected using suitable electronics circuits.

2. Thermal Detection Method:

- When the ultrasonic waves are propagated through a medium the temperature of the medium changes due to alternate compressions and rarefactions.
- In the case of stationary waves, there is a change in temperature at nodes and no changes in temperature at antinodes.
- based on this principle, by sensing the changes in temperature using suitable components, ultrasonic waves can be detected.

3. Sensitive Flame Method:

- This is a qualitative method to detect ultrasonic waves. When a narrow sensitive flame is moved in a medium of ultrasonic waves, the flame remains steady at the antinodes but flickers at nodes due to maximum changes in pressure.
- Thus, by observing the behaviors of the flame, the presence of the ultrasonic waves can be detected.

4. Kundt's Tube Method:

- This method is used to detect ultrasonic waves of low frequency.
- A Kundt's tube apparatus, shown below Figure, consists of a long glass tube of more than 1m in length and 5 cm in diameter kept horizontally with two supports on a

wooden base board. One end of the tube is fitted with an adjustable piston rod with cork. A quartz crystal placed in between the two metal plates is placed at the mouth of the other end of the tube.

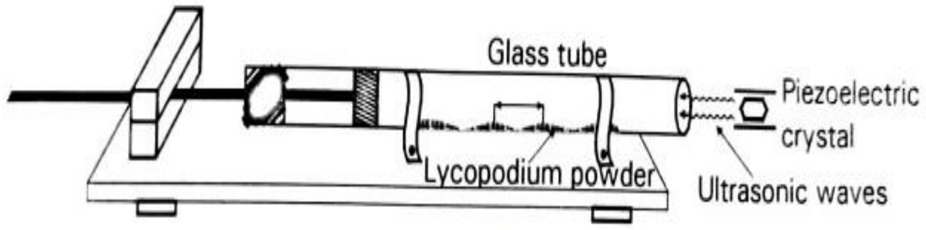


Fig.10.35 Kundt's Tube Method

- The glass tube is thoroughly dried by passing through in a hot blast of air. A thin layer of dry lycopodium or cork powder is spread along the entire length of the glass tube. When stationary ultrasonic sound waves generated by using the piezoelectric method are passed in the air contained in the long glass tube, the lycopodium powder gets collected in the form of heaps at the nodes and is blown off at the antinodes. Thus, the formation of heaps at nodes confirms the nature of the transmitted sound wave to be ultrasonic waves.
- The average distance between two consecutive heaps gives the value of half the wavelength. Thus using the relation $v = f\lambda = 2fd$, the velocity of the ultrasonic waves in the medium can also be calculated.

Application of Ultrasonic Waves:

SONAR:

- One of the major application of ultrasonic waves is used in **SONAR**. SONAR is a device which stands for Sound Navigation and Ranging.
- It is based on the principle of echo sounding. In this accoistical technique high freqeuncy ultrasonic waves are used.
- Using SONAR, the distance and direction of submarines, depth of sea, depth of rocks in the sea, the shoal of fish in the sea etc, can be determined.

Cavitation:

- One of the major application of high power and low freqeuncy ultrasonic sound waves id ultrasonic processor and cleaner. In a processsor and cleaner, the principle of cavitation is used.
- Ultrasonic Cleaning

- Ultrasonic Drilling and Cutting
- Ultrasonic Welding
- Ultrasonic Soldering

Other application of Ultrasonic Waves:

1. It is used to detect flaws or cracks in metals.
2. It is used to detect ships, submarines, iceberg, etc., in the ocean.
3. It is used for soldering aluminium coil capacitors, aluminium wires and plates without
4. It is used to weld some metals which cannot be welded by electric or gas welding.
5. It is used for cutting and drilling holes in metals.
6. It is used to form stable emulsions of even immiscible liquids like water and oil or water and mercury which finds application in the preparation of photographic films, face creams, etc.
7. It acts like a catalytic agent and accelerates chemical reactions.

Application of Ultrasonic Waves in Medical field:

1. It is used to remove kidney stones and brain tumours without shedding any blood.
2. It is used to remove broken teeth.
3. It is used for sterilizing milk and for killing bacteria.
4. It is used to study the blood flow velocities in blood vessels of our body.
5. It is used as a diagnostic tool to detect tumours, breast cancer and also the growth of foetus.

Ultrasonic Detection: Ultrasonic detection is most commonly used in industrial applications to detect hidden tracks, discontinuities in metals, composites, plastics, ceramics, and for water level detection. For this purpose, the laws of physics which are indicating the propagation of sound waves through solid materials have been used since ultrasonic sensors using sound instead of light for detection.

Principle of Ultrasonic Detection:

Defining sound wave: Sound is a mechanical wave traveling through the mediums, which may be a solid, or liquid or gas. Sound waves can travel through the mediums with specific velocity depends on the medium of propagation. The sound waves which are having high frequency reflect from boundaries and produce distinctive echo patterns.

Laws of physics for sound waves: Sound waves are having specific frequencies or number of oscillations per second. Humans can detect sounds in a frequency range from about 20Hz to 20 KHz. However, the frequency range normally employed in ultrasonic detection is 100 KHz to 50MHz. The velocity of ultrasound at a particular time and temperature is constant in a medium.

W = C/F (or) W = CT

Where W = Wave length

C = Velocity of sound in a medium

F = Frequency of wave

T=Time Period

The most common methods of ultrasonic examination utilize either longitudinal waves or shear waves. The longitudinal wave is a compression wave in which the particle motion is in the same direction of the propagation wave. The shear wave is a wave motion in which the particle motion is perpendicular to the direction of propagation. Ultrasonic detection introduces high-frequency sound waves into a test object to obtain information about the object without altering or damaging it in any way. Two values are measured in ultrasonic detection.

The amount of time, taking for the sound to travel through the medium and amplitude of the received signal. Based on velocity and time thickness can be calculated.

The thickness of material = Material sound velocity X Time of Fight

Transducers for Wave Propagation and particle detection: For sending sound waves and receiving an echo, ultrasonic sensors, normally called transceivers or transducers will be used. They work on a principle similar to radar that will convert electrical energy into mechanical energy in the form of sound, and vice versa.

The commonly used transducers are contact transducers, angle beam transducers, delay line transducers, immersion transducers, and dual element transducers. Contact transducers are typically used for locating voids and cracks to the outside surface of a part as well as measuring thickness. Angle beam transducers use the principle of reflection and mode conversion to produce refracted shear or longitudinal waves in the test material.

Delay line transducers are single element longitudinal wave transducers used in conjunction with a replaceable delay line. One of the reasons for choosing a delay line transducer is that near-surface resolution can be improved. The delay allows the element to stop vibrating before a return signal from the reflector can be received.

The major advantages offered by immersion transducers over contact transducers are Uniform coupling reduces sensitivity variations, Reduction in scan time, and increases sensitivity to small reflectors.

Operation of ultrasonic sensors:

When an electrical pulse of high voltage is applied to the ultrasonic transducer it vibrates across a specific spectrum of frequencies and generates a burst of sound waves. Whenever any obstacle comes ahead of the ultrasonic sensor the sound waves will reflect back in the form of echo and generates an electric pulse. It calculates the time taken between sending sound waves and receiving the echo. The echo patterns will be compared with the patterns of sound waves to determine the detected signal's condition.

Applications involving Ultrasonic detection: The distance of obstacle or discontinuities in metals is related to the velocity of sound waves in a medium through which waves are passed and the time taken for echo reception. Hence the ultrasonic detection can be used for finding the distances between particles, for detecting the discontinuities in metals and for indicating the liquid level.

- **Ultrasonic Distance Measurement:** Ultrasonic sensors are used for distance measuring applications. These gadgets regularly transmit a short burst of ultrasonic sound to a target, which reflects the sound back to the sensor. The system then measures the time for the echo to return to the sensor and computes the distance to the target using the speed of sound within the medium.

  Different sorts of transducers are utilized within industrially accessible ultrasonic cleaning devices. An ultrasonic transducer is affixed to a stainless steel pan which is filled with a solvent and a square wave is applied to it, conferring vibration energy on the liquid.

Fig.10.36 Ultrasonic Distance Sensor

The ultrasonic distance sensors measure distance using sonar; an ultrasonic (well above human hearing) beat is transmitted from the unit and distance-to-target is determined by measuring the time required for the echo return. The output from the ultrasonic sensor is a variable-width beat that compares to the distance to the target.

Features of the Ultrasonic Distance Sensor:

1. Supply voltage: 5V (DC).
2. Supply current: 15mA.
3. Modulation frequency: 40Hz.
4. Output: 0 – 5V (Output high when obstacle detected in range).
5. Beam Angle: Max 15 degrees.
6. Distance: 2cm – 400cm.
7. Accuracy: 0.3cm.
8. Communication: Positive TTL pulse.

- **Operation of Ultrasonic distance Sensor:** The ultrasonic sensor module comprises of one transmitter and one receiver. The transmitter can deliver 40 KHz ultrasonic sound while the maximum receiver is designed to accept only 40 KHz sound waves. The receiver ultrasonic sensor that is kept next to the transmitter shall thus be able to receive reflected 40 KHz, once the module faces any obstacle in front. Thus, whenever any obstacles come ahead of the ultrasonic module it calculates the time taken from sending the signals to receiving them since time and distance are related for sound waves passing through air medium at 343.2m/sec. Upon receiving the signal MC program while executed displays the data i.e., the distance measured on an LCD interfaced with the microcontroller in cm.

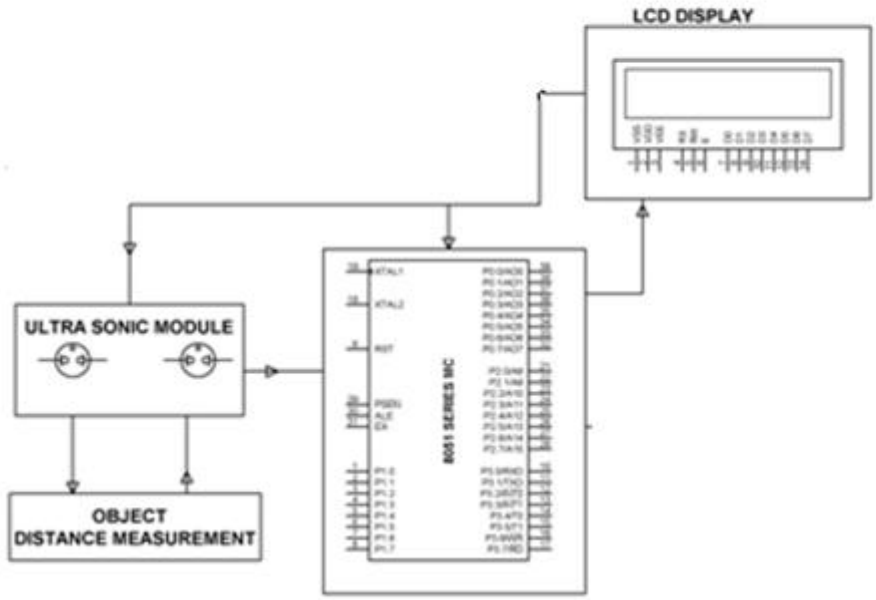


Fig.10.37 Ultrasonic Distance Sensor Circuit

Characteristically, robotics applications are very popular but you'll also find this product to be useful in security systems or as an infrared replacement if so desired.

- **Ultrasonic transducer for water level detection:**

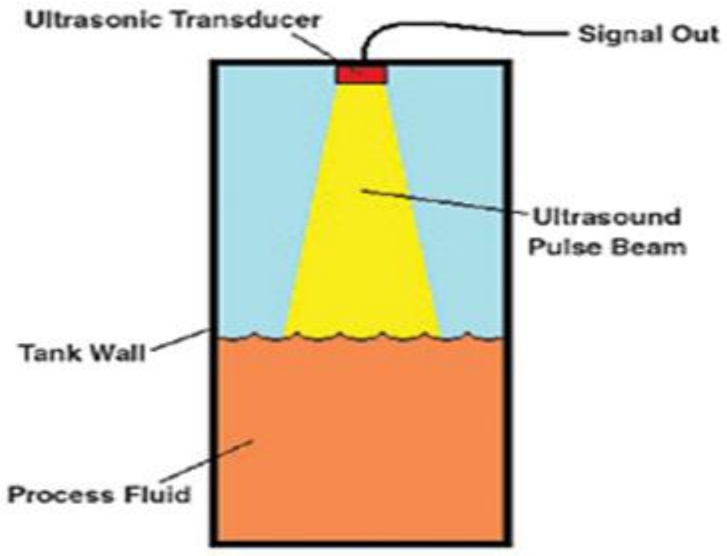


Fig.10.38 Ultrasonic Detection

The contactless liquid level controller is the ultrasonic sensor module is interfaced with the microcontroller. Whenever level distance measured in cm falls down below a set point the pump starts by sensing the signal coming out and receiving level coming to the ultrasonic transducer which is fed to the microcontroller. When the microcontroller receives the signal from the ultrasonic transducer it activates the relay through a MOSFET that operated the pump ON or OFF.

- **Ultrasonic Obstacle Detection**: Ultrasonic sensors are used to detect the presence of targets and to measure the distance to targets in many robotized processing plants and process plants. Sensors with an ON or OFF digital output are available for detecting the presence of objects and sensors with an analog output which changes relative to the sensor to target separation distance are commercially available.

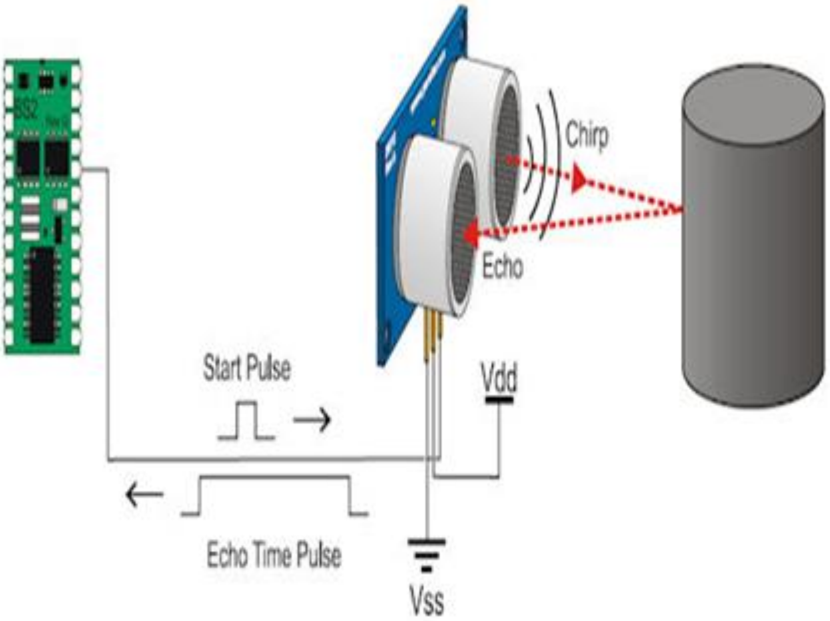


Fig.10.39 Circuit of Ultrasonic Obstacle Detection

Ultrasonic obstacle sensor consists of a set of ultrasonic receiver and transmitter which operate at the same frequency. The point when something moves in the zone secured the circuit's fine offset is aggravated and the buzzer/alarm is triggered.

Fig.10.40 Ultrasonic Obstacle Sensor

Features:

- Power consumption of 20mA.
- Pulse in/out communication.
- Narrow acceptance angle.
- Provides exact, non-contact separation estimations within 2cm to 3m.
- The explosion points LED shows estimations in the advancement.
- 3-pin header makes it simple to connect utilizing a servo development link.

Specifications:

- Power supply: 5V DC
- Quiescent current: <15mA
- Effectual angle: <15°
- Ranging distance: 2cm – 350 cm
- Resolution: 0.3 cm
- Output cycle: 50ms

The sensor detects objects by emitting a short ultrasonic burst and then listening for the eco. Under the control of a host microcontroller, the sensor emits a short 40 kHz explosion. This explosion ventures or travels through the air hits an article and after that bounces once again to the sensor.

The sensor provides an output pulse to the host that will terminate when the echo is detected; hence the width of one pulse to the next is taken into the calculation by a program to provide results in a distance of the object.

Ultrasonic Velocity Measurement: Ultrasonic velocity measurement using Ultrasonic Flaw detector is as simple as measuring the time it takes for a pulse of ultrasound to travel from one transducer to another (pitch-catch) or return to the same transducer (pulse-echo).

Changes in ultrasonic wave propagation speed, along with energy losses, from interactions with materials microstructures are often used to nondestructively gain information about a material's properties. Measurements of sound velocity and ultrasonic wave attenuation can be related to the elastic properties that can be used to characterize the texture of polycrystalline metals. These measurements enable industry to replace destructive microscopic inspections with nondestructive methods.

Applications of ultrasonic imaging system: Ultrasound imaging is a medical tool that can help a physician evaluate, diagnose and treat medical conditions. Common ultrasound imaging procedures include: Abdominal ultrasound (to visualize abdominal tissues and organs) Bone sonometry (to assess bone fragility)

Medical applications of the ultrasonic in industries:

Therapy and surgery: Because ultrasound is a mechanical vibration and can be well focused at high frequencies, it can be used to create internal heating of localized tissue without harmful effects on nearby tissue. This technique can be employed to relieve pains in joints, particularly in the back and shoulder.

Ultrasonic waves in the wide applications in engineering and industry as follows:

- Non destructive testing (detection of flaws in metals)
- Ultrasonic drilling.
- The ultrasonic welding.
- Ultrasonic drilling.

- A ultrasonic soldering.
- ultrasonic cutting and machinery.
- A ultrasonic cleaning.
- Sonar.

Hypersonic Sound System Technology:

Fig.10.41 Hypersonic Sound System

Hypersonic Sound (HSS) is one of the most revolutionary digital sound reproduction technologies of this century**.** It is considered one of the most innovative approaches to sound engineering. This technology was developed by American Technology Corporation. The biggest advantage of this technology is that it does not include any complex electronic circuits having enclosures, crossovers, woofers, and midrange or tweeter elements as compared to a conventional loudspeaker system. This technology also provides ideal characteristics to almost all audio frequencies (normal loudspeakers is not ideal for all frequencies such as 20Hz to 20 KHz).

Fig.10.42 HSS Audio System

The Hypersonic system produces highly directional audio without any disturbance for outside. By using this technology, we can create virtual real audio. This technology works with the help of modulated ultrasonic waves (above 40KHz) and these waves itself create audio frequency waves according to the input audio signal (This property of creating new audio waves within the ultrasonic wave is called non-linearity of air) .

Working of HSS System:

HSS technology works on the property of non-linearity. Non-linearity is a property of air that creates new waves within another wave by the continuous expansion and contraction of waves. Here ultrasonic waves are used as modulated waves according to the input audio signal. This wave, when passing through the air, creates new sound waves within the range of human hearing (20Hz-20KHz). We can't get any disturbance from the ultrasonic sound (since it can't be heard by the human ear), we can only hear the newly generated audio waves (that same as the input audio).

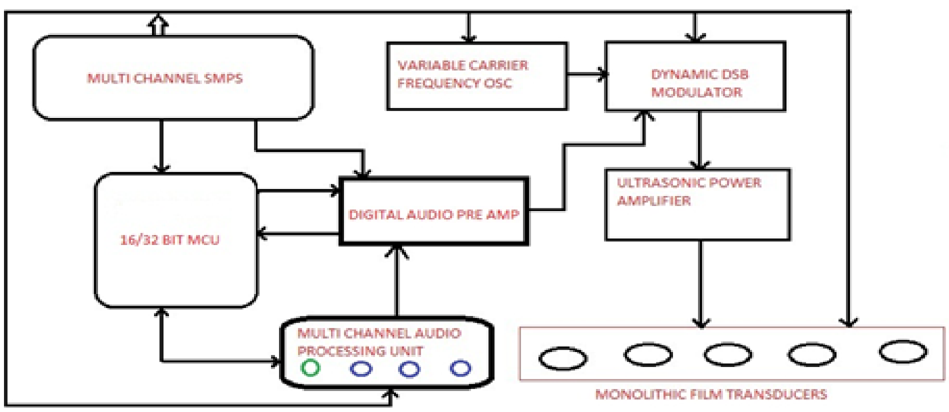


Fig.10.43 HSS System Architecture

The architecture of an ultrasonic transducer is given above. This structure consists of multiple sections as explained below

- **SMPS**: As like all electronic equipment, The HSS system is also powered with a highly filtered SMPS. Normally multi-voltage power supplies are used as SMPS. According to the power at the output, the supply voltage may vary from 12V to 60V.
- **Microcontroller Unit**: This unit can perform all the audio adjustments. Usually, 16 to 32-bit microcontrollers are used as the microcontroller.
- **Audio Signal Processing Unit**: Audio signal processing unit is an advanced audio circuit that consists of multichannel-Multi format High Definition Audio processor. This circuit detects and decodes all audio inputs and are filtered according to the requirement and sent to a digital pre-amp unit. We can provide several types of inputs to this unit like single channel, multichannel audio up to 11.2 channels, and also all the DTS and DOLBY digital true audio formats, etc. This section decodes almost all type of digital audios present. This unit also controls equalization, other dynamic controls, etc
- **Digital Audio Pre-amp**: Digital audio pre-amp is used for the pre-amplification of the decoded audio from the audio signal processor. After pre-amplification is completed, the pre-amplified audio is sent to the ultrasonic modulator
- **Variable Carrier Frequency Oscillator (VCFO)**: This is an ultrasonic oscillator circuit used for the generation of ultrasonic carrier waves for the modulation. Usually ultrasonic crystal oscillators are used for making these oscillations and its frequency is above 40 KHz.
- **Dynamic DSB Modulator**: Dynamic DSB modulator performs the Ultrasound modulation with the audio signal. In this unit, the input digital audio is modulated with the high-frequency ultrasonic waves and produces an HF modulated wave. This unit also filters all other HF disturbances too. In DSB system, the modulation index can be reduced to suppress all the distortions.
- **Ultrasonic Power Amplifier**: An ultrasonic power amplifier is an HF Power amplifier that increases the power at the output stage. This is the final stage of HSS amplifier unit. After the HF power amplification, the signal is fed to the monolithic film transducers, which transmit the HSS audio

- **Monolithic Film Transducers**: Monolithic Film Transducers are used for the emission of HSS waves into the air. Usually, piezoelectric polyvinylidene transducers are used as a transducer. The signal from the UF power amplifier is emitted into the air by a transducer. The ultrasonic waves are highly directional and transmit like a flashlight. Inside this range of HSS sound beam, the audio can be heard by a human. Outside this range, there is nothing to hear.

The characteristics and frequency response of HSS sound system is given below

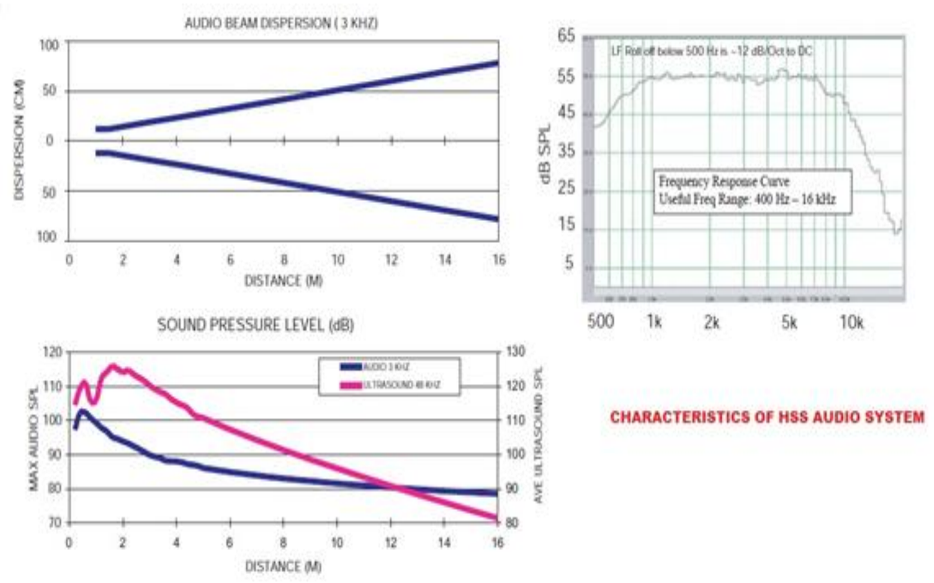


Fig.10.44 HSS Characteristics

Advantages:

1) Highly focused directional audio
2) Very thin and lightweight transducer used as the ultrasonic emitter.
3) The sound only can hear within a particular angle.
4) Provide HD real audio for virtual reality (virtual real audio processing is possible).
5) Support multi-channel audio.

Disadvantages:

1) High operating voltage required as compared to a normal audio amplifier.
2) Direct analog amplification is not possible.
3) Minimum stereo quality sound input required for the proper working.
4) High manufacturing cost.

Applications of Hypersonic Sound Systems:

Hypersonic sound systems have a wide range of applications. The great advantage of such a directional sound system is that it is audible only to those in its focus; which means they won't annoy any passersby. This has a massive scope in art, entertainment and hospitality industries. Imagine restricting the audio explanations of paintings in an art gallery only to those in front of that particular painting? Same goes for restaurants, retail, airports and, my favorite, video conferencing. These sound systems basically help to create a personal zone in a crowded place.

SAQ.2

a) What do you mean by working principle of the Cathode ray oscilloscope?
b) Discuss about the block diagram of the CRO and write its application.
c) What do you mean by function of the Multimeter?
d) Discuss about the working principle of the Multimeter.
e) What do you mean by the production of the Ultrasonics?
f) Write and discuss about the detection of the Ultrasonics.
g) What do mean by the Hypersonics?

10.9 Summary:

1. An integrated circuit (IC), sometimes called a chip, microchip or microelectronic circuit, is a semiconductor wafer on which thousands or millions of tiny resistors, capacitors, diodes and transistors are fabricated.
2. Integrated circuits are frequently used in electronics design and can be categorized as analog, digital, or a combination of the two. ICs can be used for a variety of purposes including amplifiers, video processors, computer memory, switches, and microprocessors.
3. A discrete circuit is constructed of components which are manufactured separately. Later, these components are connected together by using conducted wires on a circuit board or a printed circuit board.
4. Any system can be constructed by using discrete components and also by an IC. We cannot physically put all multiple discrete circuits on a plate of silicon and simply call

it an integrated circuit. Integrated circuits are made up of silicon wafers, not inserted (or placed) onto silicon wafers.

5. IC's are classified into two types based on the Mode of Operation or the signal processed. They are: Digital IC or Non-Linear IC. Analog IC or Linear IC.
6. An integrated circuit (IC), sometimes called a chip or microchip, is a semiconductor wafer on which a thousand or millions of tiny resistors, capacitors, and transistors are fabricated. An IC can be a function as an amplifier, oscillator, timer, counter, computer memory, or microprocessor.
7. A monolithic integrated circuit (IC) is a complete circuit or group of circuits manufactured in a single piece of silicon, a typical physical size being 1.25 mm square (or about fifty thousandths of an inch square). Such a circuit may contain fifty or more components such as transistors or resistors.
8. For the manufacture and production of the monolithic IC, all circuit components and their interconnections are to be formed in a single thin wafer. The different processes carried out for achieving this are explained below. Being the base layer of the IC, the P-type is silicon is first built for the IC.
9. The CRO working principle depends on the electron ray movement because of the electrostatic force. Once an electron ray hits a phosphor face, then it makes a bright spot on it. A Cathode Ray Oscilloscope applies the electrostatic energy on the electron ray from two vertical ways.
10. The applications of CRO are mainly involved in the radio, TV receivers, also in laboratory work involving research and design. In modern electronics, the CRO plays an important role in the electronic circuits.
11. An oscilloscope is a laboratory instrument commonly used to display and analyze the waveform of electronic signals. In effect, the device draws a graph of the instantaneous signal voltage as a function of time.
12. The multimeter basically consists of a bridge DC amplifier, rectifier, PMMC meter, function switch, internal battery and an attenuator. The function of the attenuator is that it helps to select a particular range of voltage values.
13. A multimeter is a measuring instrument that can measure multiple electrical properties. A typical multimeter can measure voltage, resistance, and current, in which case it is also known as a volt-ohm-milliammeter (VOM), as the unit is equipped with voltmeter, ammeter, and ohmmeter functionality.

14. The term that is ultrasonic devices that are used to detect objects and measure distances. By imagining the ultrasound or sonography is often used in medicine. In the nondestructive testing which is of structures and products.
15. Ultrasonic waves can be detected using Radiometer. In this method ultrasonic beam is made to fall on a thin mica fan suspended by a thin wire carrying a small mirror from one end of a light rod. Due to pressure exerted by ultrasonic waves the fan gets detected along with the mirror.
16. Ultrasonic velocity measurements are found to be useful for on-line assessment of the extent of degradation of mechanical properties associated with precipitation of intermetallic in Inconel 625.
17. Ultrasonic are used for making holes in very hard materials such as glass, diamond etc., When ultrasonic are passed through these materials it creates air bubbles. This air bubbles collapses within short span of time, thereby a larger amount of pressure and temperature which are used for cutting and drilling.
18. They're so fast**,** their speed can change the surrounding air molecules. They can carry a nuclear warhead, fly low and be hard to detect.
19. A hypersonic aero-ballistic system is dropped from an aircraft, accelerated to hypersonic speed using a rocket and then follows a ballistic, meaning unpowered, trajectory.

10.10 Terminal Questions:

1) Explain the Introduction of integrated circuit and write the comparison between of the integrated circuit and discrete circuits.
2) Discuss and explain the Classification of IC and also explain the basis of construction and operation of IC.
3) Explain the working principle of the Monolithic IC with the help of the basics structure.
4) Discuss and explain the types of the Monolithic IC fabrication in detail.
5) Explain the working principle of the Cathode ray oscilloscope with the help of the block diagram also write its application.
6) Explain the construction and working principle of the Multimeter.
7) Discuss and explain the types of the Multimeter and also write the its function.
8) What do you mean by the production of the Ultrasonics? Explain the Ultrasonics detection in detail.

9) Explain the velocity measurements of the Ultrasonics and write its applications.
10) Explain the working principle of the Hypersonic.